FEATURE EXTRACTION AND CLASSIFICATION OF ELECTROCARDIOGRAM SIGNAL TO DETECT ARRHYTHMIA AND ISCHEMIA DISEASE

NOR HAFFEEZAH BINTI KAMARUDIN

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
MAY 2010
FEATURE EXTRACTION AND CLASSIFICATION OF
ELECTROCARDIOGRAM SIGNAL TO DETECT
ARRHYTHMIA AND ISCHEMIA DISEASE

NOR HAFEEZAH BINTI KAMARUDIN

DISSEYATATION SUBMITTED IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF COMPUTER SCIENCE

FACULTY OF COMPUTER SCIENCE AND INFORMATION
TECHNOLOGY
UNIVERSITY OF MALAYA
MAY 2010
ACKNOWLEDGEMENTS

In the name of Allah, the Beneficent, the Merciful

I dedicate this work to my parents, Mr. Kamarudin Hamzah and Madam Azimah Jusoh who have always encouraged, motivated and given me the moral values to lead a peaceful life. Special thanks forwarded to my beloved husband, Mr. Muhad Rozi Mat Nawi for his patience and support in my life. I would like to express my sincere thanks to my supervisor, Mr. Emran Mohd Tamil for his tremendous guidance, advice, assistance and encouragement throughout my postgraduate candidature period. Also, special thanks to my co-supervisor, Dr. Rosli Salleh, Mr. Zaidi Razak, Mr. Yamani Idna Idris, Mr. Noorzaily Md Noor and also Prof Dato’ Dr. Zulkifli Mohd Yusoff for their guidance, help and advice to handle the research works.

Apart from them, I would like to express my appreciation to Dr. Azmi Mohd Tamil from Medical Faculty, National University of Malaysia (UKM) for his tremendous help and guidance on academic and professional developments with impromptu discussions or exchange of ideas. I would like to forward gratitude to my fellow colleagues, for their moral supports and help during my research works.

I also would like to express my gratitude to my siblings, and other family members for their continuous support in my life. Their support allowed me to concentrate fully in my studies without hesitation. I am gratitude to the University of
Malaya for supporting me the financials of research and life, under the University Malaya Scholarship. Last but not least, my sincere thanks, compliments and regards to anyone who had helped and supported me in one way or another.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>GLOSSARY</td>
<td>xiii</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>xiv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xvi</td>
</tr>
</tbody>
</table>

## CHAPTER 1 - INTRODUCTION

1.0 Overview 1

1.1 Electrocardiogram Analysis: Background to Research 1

1.2 Problem Definitions 4

1.3 Significance of Research 5

1.4 Research Objectives 6

1.5 Scope of Works 7

1.6 Thesis Outlines 8
CHAPTER 2 - LITERATURE REVIEW

2.0 Overview  
2.1 Wireless Technology Developments  
2.2 Wireless ECG  
2.3 Heart Disease  
2.4 The Electrocardiogram Diagnosis Techniques  
2.5 ECG Features Extraction Algorithms: A Review of Previous Studies  
2.6 ECG Training and Classification Algorithms: A Review of Previous Studies  
2.7 Summary of an Approach of Electrocardiogram Analysis Algorithms  
2.8 Summary  

CHAPTER 3 - RESEARCH METHODOLOGY AND SYSTEM DESIGN

3.0 Overview  
3.1 Overview of Electrocardiogram Analysis System  
3.2 System Requirement  
3.3 Design and Architecture  
3.4 Electrocardiogram Signal Analysis Procedures  

3.4.1 Signal Data Preparation  
3.4.1.1 Signal Data Characteristics  

3.4.2 Preprocessing  
3.4.2.1 Databases and Data Sources  
3.4.2.2 Normalization Data  
3.4.2.3 Data Selection
3.4.2.4 Data Processing 47

3.4.3 Features Extraction using Discrete Wavelet Transform 47

3.4.3.1 Discrete Wavelet Transform 49

3.4.3.2 DWT Implementation 52

3.4.3.3 Features Extraction Procedures 53

3.4.3.4 Coefficients Extraction 56

3.4.4 Classification using Neuro Fuzzy 59

3.4.4.1 Neuro Fuzzy Approach 60

3.4.4.2 ANFIS Model 61

3.4.4.3 ANFIS Implementation in Classifying Heart Disease 65

3.4.4.3.1 ECG Signals Dataset 66

3.4.4.3.2 Fuzzy Inference System 67

3.4.4.3.3 Rule Base Identification 70

3.4.4.4 Training and Testing Datasets 74

3.5 ECG Analysis Performances 75

CHAPTER 4 - RESULTS AND DISCUSSION

4.0 Overview 77

4.1 Introduction 77

4.2 Pre Processing 78

4.2.1 Normal Signal 78

4.2.2 Bradycardia Signal 80

4.2.3 Tachycardia Arrhythmia Signal 82

4.2.4 Ischemia Signal 84
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Summary of an Approach of Electrocardiogram Analysis</td>
<td>33</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Three Main Phases Exist: P, QRS, and T</td>
<td>43</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Amplitudes Values for Normal ECG Signal</td>
<td>45</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Durations Values for Normal ECG Signal</td>
<td>45</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Sample Data of each ECG Signal</td>
<td>47</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>The Extracted Features of Four Exemplary from Four Classes</td>
<td>58</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Set of ECG Signals Class</td>
<td>66</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>Statistic Properties of the ECG Datasets</td>
<td>67</td>
</tr>
<tr>
<td>Table 3.8</td>
<td>Created Rule Base by Expert Knowledge</td>
<td>71</td>
</tr>
<tr>
<td>Table 3.9</td>
<td>Class Distribution of the Sample in the Training and Testing Datasets</td>
<td>74</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>The Extracted Features of Five Exemplary Records from Four Classes</td>
<td>94</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Statistic of Correct and Incorrect Heart Disease Classification</td>
<td>99</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>ANFIS Training Performance</td>
<td>101</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>ANFIS Classification Performance</td>
<td>102</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Diagram of the Human Heart and an Example of Normal ECG Trace (Otto et al., 2006)</td>
<td>2</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Simple Feed Forward Neuron Model by Gao et al. (2004)</td>
<td>21</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Overall Techniques for Electrocardiogram Analysis</td>
<td>40</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Electrocardiogram Analysis</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Preprocessing Step</td>
<td>46</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Diagrams of Sinusoidal Signal and a Daubechies Wavelet (Mahmoodabadi et al., 2005)</td>
<td>48</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Filter Banks Signal Decomposition</td>
<td>49</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Three-Level Wavelet Decomposition Trees</td>
<td>50</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Feature Extraction Techniques</td>
<td>53</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>DWT Decomposition Step in ECG Analysis</td>
<td>55</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>Flowchart of DWT Coefficient Calculation</td>
<td>57</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>Analysis Framework</td>
<td>59</td>
</tr>
<tr>
<td>Figure 3.11</td>
<td>Structure of Feedforward Neuro Fuzzy</td>
<td>60</td>
</tr>
<tr>
<td>Figure 3.12</td>
<td>Basic Structure of ANFIS Model</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3.13</td>
<td>Block Diagram of Heart Disease Classification through Adaptive Neural Fuzzy Inference System (ANFIS)</td>
<td>65</td>
</tr>
<tr>
<td>Figure 3.14</td>
<td>Desired Output Target for Each Class</td>
<td>66</td>
</tr>
</tbody>
</table>
Figure 3.15  Fuzzy Inference System for Heart Disease Classification  68
Figure 3.16  Initial Membership Functions for Each Input Dimensions  69
Figure 3.17  243 Rule-base ANFIS Structure  72
Figure 3.18  A Fuzzy Rule Base for a Trained 243 ANFIS Rule Base  73
Figure 4.1  Normal ECG Signal  78
Figure 4.2  Normalized Normal ECG Signal for the First Sample  79
Figure 4.3  Frequency Component of Normal ECG Signal  79
Figure 4.4  Bradycardia Arrhythmia ECG Signal  80
Figure 4.5  Normalized Bradycardia Arrhythmia ECG Signal for the  
            First Sample  80
Figure 4.6  Frequency Component of Bradycardia Arrhythmia ECG Signal  81
Figure 4.7  Tachycardia Arrhythmia ECG Signal  82
Figure 4.8  Normalized Tachycardia Arrhythmia ECG Signal for the  
            First Sample  82
Figure 4.9  Frequency Component of Tachycardia Arrhythmia ECG Signal  83
Figure 4.10  Ischemia ECG Signal  84
Figure 4.11  Normalized Ischemia ECG Signal for the First Sample  84
Figure 4.12  Frequency Component of Ischemia ECG Signal  85
Figure 4.13  The Vector, C and Length, L of Normal Signal  86
Figure 4.14  The Approximation, A and Detail, D Coefficients of  
            Level 1 Decomposition  86
Figure 4.15  The Detail Coefficients of D_1 to D_5 and Approximation  
            Coefficient of A_5  87
Figure 4.16  The Vector, C and Length, L of Bradycardia Arrhythmia Signal  88
Figure 4.17 The Approximation, A and Detail, D Coefficients of
Level 1 Decomposition 88

Figure 4.18 The Detail Coefficients of $D_1$ to $D_5$ and Approximation
Coefficient of $A_5$ 89

Figure 4.19 The Vector, C and Length, L of Tachycardia Arrhythmia Signal 90

Figure 4.20 The Approximation, A and Detail, D Coefficients of
Level 1 Decomposition 90

Figure 4.21 The Detail Coefficients of $D_1$ to $D_5$ and Approximation
Coefficient of $A_5$ 91

Figure 4.22 The Vector, C and Length, L of Ischemia Signal 92

Figure 4.23 The Approximation, A and Detail, D Coefficients of
Level 1 Decomposition 92

Figure 4.24 The Detail Coefficients of $D_1$ to $D_5$ and Approximation
Coefficient of $A_5$ 93

Figure 4.25 ANFIS Network Error Convergence 96

Figure 4.26 Plot for 243 Rule-base ANFIS Training Performance 97

Figure 4.27 ANFIS Model Testing Performance 98

Figure 4.28 Statistic of Heart Disease Classification 99

Figure 4.29 Statistic of Heart Disease Classification according to Class 100
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANFIS</td>
<td>Adaptive Neuro Fuzzy Inference System</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transform</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>FL</td>
<td>Fuzzy Logic</td>
</tr>
<tr>
<td>FIS</td>
<td>Fuzzy Inference System</td>
</tr>
<tr>
<td>FFNF</td>
<td>Feedforward NeuroFuzzy</td>
</tr>
<tr>
<td>FNN</td>
<td>Fuzzy Neural Network</td>
</tr>
<tr>
<td>MF</td>
<td>Membership Function</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi Layer Perceptron</td>
</tr>
<tr>
<td>NF</td>
<td>Neuro Fuzzy</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Networks</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>TS</td>
<td>Takagi Sugeno</td>
</tr>
</tbody>
</table>
## GLOSSARY

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac arrhythmia</td>
<td>A group of conditions in which the muscle contraction of the heart is irregular or is faster or slower than normal.</td>
</tr>
<tr>
<td>DWT coefficient</td>
<td>Parameter extracted from the wavelet analysis using the filtering process.</td>
</tr>
<tr>
<td>Defuzzification output</td>
<td>Fuzzy sets is taking as input, defuzzification outputs a crisp value which suitable for analysis.</td>
</tr>
<tr>
<td>Electrocardiogram</td>
<td>An electrical recording of the heart and is used in the investigation of heart disease.</td>
</tr>
<tr>
<td>Fuzzy rule</td>
<td>A consequent of a rule of fuzzy set that represent by a membership function. The consequent is reshaped using a function associated with the antecedent.</td>
</tr>
<tr>
<td>Ischemia</td>
<td>A restriction in blood supply, generally due to factors in the blood vessels, with resultant damage or dysfunction of tissue. ‘Isch’ means reduced and ‘emia’ means blood. Myocardial ischemia is caused due to lack of sufficient blood supply to the cells, and may lead to myocardial infarction and in extreme cases, even death.</td>
</tr>
<tr>
<td>Input fuzzification</td>
<td>The step to take the inputs and determine the degree to which they belongs to each of the appropriate fuzzy sets via membership functions.</td>
</tr>
<tr>
<td>Mean</td>
<td>Provides a score that each case would have if the variable were distributed equally among all observations. The arithmetic average.</td>
</tr>
<tr>
<td>Median</td>
<td>A score that separates all of the observations (frequencies) into two equal-sized groups.</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>A measure of dispersion, based on deviations from the mean, calculated by taking the square root of the average squared deviations of each score from the mean.</td>
</tr>
</tbody>
</table>
PENGEKSTRAKAN DAN PENGKELASAN CIRI-CIRI ISYARAT ELEKTROKARDIOGRAM UNTUK MENENTUKAN PENYAKIT ARRHYTHMIA DAN ISCHEMIA

ABSTRAK

ANFIS telah berjaya mengkelaskan isyarat Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia dan Ischemia dengan kadar ketepatan yang tinggi iaitu 97.67%. Sistem ini juga mampu mencapai kadar kepekaan yang tinggi iaitu 97.73%, 95.24%, 100% dan 97.62%, untuk setiap kelas yang diuji. Nilai kekhususan yang didapati untuk kelas isyarat Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia dan Ischemia yang dilaksanakan oleh sistem ANFIS ialah 100%, 99.23%, 99.22% dan 98.46%. 
FEATURE EXTRACTION AND CLASSIFICATION OF
ELECTROCARDIOGRAM SIGNAL TO DETECT ARRHYTHMIA AND
ISCHEMIA DISEASE

ABSTRACT

Electrocardiogram (ECG) is a method used to measure the rate and regularity of heartbeats. Comparison of overall ECG waveform pattern and shape enables doctors to diagnose possible diseases. Currently there is computer based analysis which employs certain signal processing to diagnose a patient based on ECG recording. Signal processing usually takes the form of a transformation of a signal into another signal that is in some sense more desirable than the original. The purpose of this research is to address in identifying the Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia signal using the method of Discrete Wavelet Transform (DWT) and Neuro Fuzzy classifier, a hybrid method of Artificial Neural Networks and Fuzzy Logic. DWT coefficients are used to extract the relevant information from the ECG input data which are Energy, Maximum, Minimum, Mean and Standard Deviation value. Then the extracted features data is analyzed and classified using Adaptive Neuro Fuzzy Inference System (ANFIS) as a Neuro Fuzzy classifier. The proposed algorithm is implemented and also tested in MATLAB software. The ECG signal are being selected and tested from PhysioNet Database using MIT-BIH Arrhythmia Database, and Intracardiac Atrial Fibrillation Database and also from National University Hospital of Malaysia (HUKM) database. The ANFIS system successfully classifies the Normal, Bradycardia
Arrhythmia, Tachycardia Arrhythmia and Ischemia signal with the rate of accuracy is 97.67%. The analysis system also can achieved the sensitivity up to 97.73%, 95.24%, 100% and 97.62%, respectively for each class tested and the specificity value of Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia class proposed by ANFIS are 100%, 99.23%, 99.22% and 98.46%, respectively.
CHAPTER 1
INTRODUCTION

1.0 Overview

In this chapter, brief information about the principles, benefits and problems in electrocardiogram analysis is presented. This part discusses the electrocardiogram analysis problems concerning health issue which encourage the present research. Then, the problem definitions from the previous studies, the research objectives, scope of present works and thesis outlines are presented.

1.1 ECG Analysis: Background to Research

Electrocardiogram (ECG) is the electrical manifestation of the contractile activity of the heart that can be recorded fast and automatically. It is a noninvasive diagnostic tool, meaning that ECG signal can be measured without entering the body at all. Electrodes are placed on the user’s skin to detect the bioelectric potentials given off by the heart that reach the skin’s surface in order to measure the rate and regularity of heartbeats, the position of the chambers, the presence of any damage to the heart and shows the information of the cardiovascular condition or devices used to regulate the heart such as a pacemaker as described by Gandelman (2006). A natural electrical system causes the heart muscle to contract and pump blood through the heart to the lungs and the rest of
the body. The electrical potential generated by electrical activity in the cardiac tissue calls biosignal.

An ECG translates the heart electrical activity into line tracings on paper which shown in Figure 1.1. The spikes and dips in the line tracings are called waves which are P wave, QRS complex, T wave and ST segment.

![Diagram of the Human Heart and an Example of Normal ECG Trace (Otto et al., 2006)](image)

**Figure 1.1** Diagram of the Human Heart and an Example of Normal ECG Trace (Otto et al., 2006)

The heart is a muscular pump made up of four chambers as shown in Figure 1.1. The two upper chambers are called atria, and the two lower chambers are called ventricles. A natural electrical system causes the heart muscle to contract and pump blood through the heart to the lungs and the rest of the body as carried out by Melnyk and Silberman (2004), Gandelman (2006) and Otto et al. (2006). The P wave is a record of the electrical activity through the upper heart chambers (atria). The QRS complex is a record of the movement of electrical impulses through the lower heart chambers (ventricles). The ST segment corresponds to the time when the ventricle is contracting but no electricity is
flowing through it. The ST segment usually appears as a straight, level line between the QRS complex and the T wave. Then, the T wave corresponds to the period when the lower heart chambers are relaxing electrically and preparing for their next muscle contraction.

In the medical test using ECG, the heart disease detection is based on the difference wave signals that appear on the screen during the ECG test. The detection of the pulse is usually detected on the basis of the largest in a signal of PQRST of ECG signal. The normal heart beat in a regular rhythm will show the line tracing of the PQRS and T wave looks normal. If there any obvious changes of the PQRST line tracing, it shows that the heart may having a problems. Comparison of overall ECG waveform pattern and shape enables doctors to identify diseases. The ECG remains the simplest noninvasive diagnostic method for various heart diseases.

In order to accurately characterize heart rate analysis, so a precise and reliable ECG waveform recognition procedure is necessary. The information and measurement waveform of the ECG signal is to extract features (characteristic) from the signal. The features are sufficiently representative of the physical process and the heart disease problem. Previous studies prove that frequency analysis sufficiently represents ECG waveforms; therefore frequency analysis is utilized to extract features from the ECG signal. The shapes of the ECG waveforms of different persons are different, so the differences of the waveform can be used to identify the different individual’s characteristic. The ECG signal can varies from person to person due to the differences in
position, size anatomy of the heart, age, sex relative body weight and also chest configuration.

In the past, many works of using algorithms for ECG analysis have been investigated and the successful results are achieved. However, mostly the works are regarding to the specific disease checking system with so many limitations. Nowadays, many researches of the ECG waveforms detection methods such as Artificial Neural Networks, Genetic Algorithms, Pattern Comparison, Wavelet Transform and other methods has been done in order to get the accurate, fast and better classification signal.

1.2 Problem Definitions

As noticed, the ECG analysis system is quite complicated due to the system needs a proper algorithm to make sure the level of accuracy is very high. From the past studies, most systems of ECG analysis system are characterized based on the heart disease that will be measured by the system and each types of heart disease is evaluated based on the characteristic of the heart signal. It is quite huge of signal types since the ECG signal itself can represent hundred of heart disease. The system really needs a high efficiency and safe because the health care is the most important in our life.

A trained people can easily recognizes the heart failure by manually searching thousands of heartbeats. But the several important heart cycle movements are also very small and rapid and cannot be able to catch by the human vision. It cause the level of classifying in order to detect the heart disease is not accurate. So it needs an excellent algorithm that
every single event of heart signal cycle can be catch correctly and each characteristic of each heart disease signal must be studied to make sure the signals represent the correct disease. This characteristic needs critical evaluation to detect the suitable heart disease. Therefore, the algorithms proposed must be high precision of detection and exact classifiers are needed to obtain a successful operation and can give earlier notification to the patients. But it is very hard to choose the algorithms that can suite with all of the disease. At least an algorithm with high levels of accuracy and reduced level of false are really needed to be approved. In order to minimize such limitations of the ECG analysis system, a system of ECG system analysis incorporates with Discrete Wavelet Transform and Neuro Fuzzy has been developed which can detect the disease properly.

1.3 Significance of Research

The ECG analysis system is one of the biosignal processing areas that involve the application of computer science and engineering to detect and visualize the biological processes. It is important tools and knowledge to the study of diseases to apply advanced technology to the complex problems of medical care that essential to enhance the patient living quality and appropriate treatment.

This research is important because it can be used by the other health care professionals including physicians, nurses, therapists and technicians to bring together knowledge from many technical sources to develop new procedures, or to solve clinical problems.
The ECG analysis system can bring the possibility to record the heart condition at early stage, which the problem is being hard interpretation for non-trained people. Therefore the importance in developing the system that makes this interpretation easier for non-trained people and the system could detect the disease with high levels of accuracy because many people who died cause of heart disease showed no outward symptoms.

The results of ECG analysis using the proposed algorithm are able to be used in the patient monitoring system that can be used in a hospital, transport, or emergency response environment.

1.4 Research Objectives

This research aims to design an ECG analysis system that will measure the rate and regularity of heartbeats. This system need a good quality and accurate of analysis output to make sure the result of the heart problem are correct.

Basically the goals of this research as follows.

1.4.1 To determine the viability of ECG signal and characterization of ECG waveform in classifying the heart disease problem.

1.4.2 To implement an analysis system for ECG signals using Discrete Wavelet Transform (DWT) and Adaptive Neuro Fuzzy Inference System (ANFIS) as a Neuro Fuzzy classifier.

1.4.3 To evaluate the performance of ECG analysis using DWT and ANFIS that can allow for more accurate diagnoses in classifying the heart disease.
1.4.4 To analyze the ECG signal waveform in classifying Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia of ECG signals.

1.5 Scope of Works

In this research, a system of ECG analysis which operates by Discrete Wavelet Transform and Adaptive Neuro Fuzzy Inference System is developed. During operation, the analysis from the variation input signal based on the P, Q, R, S and T wave of ECG waveform are evaluated. The systems are able to recognize the different types of input wave from different people. The varying input signals based on few peoples are being tested on the same instruction. Then ECG module will determine the different data information from different people. In this study, the system has been conducted using MATLAB software.

The MATLAB simulation will be used to determine the output from the various ECG signals. Four parameters are considered in the present study based on the P, Q, R, S and T wave signals which are Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia signals. These parameters are completely patient independent which means the learning data is independent of the testing data. The data used is from the PhysioNet Database.

The signal parameters being stimulated by using the algorithms of DWT and the DWT coefficient from the values of Energy, Maximum, Minimum, Mean and Standard Deviation is fed into the ANFIS classifier and then the coefficients will be used to
determine either the event is caused by heart disease or not. The results from the simulation are compared with each group of training and testing in order to measure the level of accuracy between the different heart signals. The analysis of the results is concentrated on the classification of Normal heart signal, Bradycardia Arrhythmia signal, Tachycardia Arrhythmia signal and Ischemia signal.

1.6 Thesis Outlines

This thesis is organized in such a way that it provides a continuous and smooth flow of information to the reader, regarding the development and analysis of ECG analysis system. There are total of five major chapters which are subdivided into suitable sections. The major five chapters in this thesis are Introduction, Literature Review, Research Methodology and System Design, Result and Discussion and finally Conclusion and Recommendation of the project. The content of each chapters are outlined as follows.

Chapter 1 is an introduction of the project. This chapter gives brief information about the background, problems of the analysis system and the proposed solution for the ECG analysis system. The overall overview of the project scopes and objectives of the project are also presented in this chapter.

Chapter 2 focuses on the literature review and the methodologies of the previous studies of ECG analysis using the various applications and algorithms of features extraction and
classifier are reviewed. This chapter deals with the past and current trends of the ECG analysis study.

Chapter 3 will discuss the research methodology and system design of the project. This chapter explains how the project is organized and the flow of all the project operation part. This chapter discusses about the ECG analysis design and the implementation of DWT and ANFIS. It also discusses MATLAB software of the system.

Chapter 4 discusses the features extraction and classification result of ECG analysis using DWT and ANFIS methods. This part includes the MATLAB software development for the both methods. All discussions that concentrate on the result and performance of the ECG analysis using DWT and ANFIS are presented. It gives a brief review the correlation of all methods.

Chapter 5 discusses the conclusion and further development of the project. This chapter also presents and describes the problems, limitations and the recommendation for this project and overall ECG analysis for the future development or modification.
CHAPTER 2
LITERATURE REVIEW

2.0 Overview

The improvement of Electrocardiogram (ECG) analysis which is part of biosignal processing to obtain the heart disease classification has been studied by many researchers from the past decades up to now. Their studies have been carried out through experimental and numerical works. In ECG analysis, the main idea is to make the analysis methods enhancement in the degree of accuracy in classifies the disease. By filling the suitable analysis methods, the heart disease classification can be calculated accurately at a fast rate through the analysis process. Therefore, the past studies of ECG analysis algorithms enhancement are an important topic that should be reviewed. This chapter is aimed at providing some of related information regarding the research carried out pertaining to the improvement of heart disease classification with the important roles played by ECG analysis, from different researchers across the globe.

This literature begins by reviewing some of the previous studies of wireless technology in medical application. It follows by reviewing the studies of heart disease analysis that divided into features extraction and classification techniques. This literature also highlights the limitation of existing ECG analysis system pertaining to current work.
2.1 Wireless Technology Developments

The development of the wireless technology in medical application will increase the human healthcare because, this technology can be implement in the medical care instrumentation to develop a system that can assist people to check their health condition especially their heart in the good condition with fast and accurate. Based on the research, many people who died cause of Coronary Heart Failure because this disease does not showed outward symptoms. The invention of wireless ECG can help to increase the human healthcare and improve the efficiency and powerful medical applications.

2.2 Wireless ECG

Recent technological advances in integrated circuits, wireless communications allow miniature, lightweight, ultra-low power and intelligent monitoring devices. World health benefits from many forms of technology used for diagnosis, treatment, and the general management of medical care. A wealth of research in wireless technology has begun in pervasive healthcare which, as it develops, will allow patients to lead an independent lifestyle in their own homes and produce maximum benefits to society stated by Centers for Disease Control and Prevention (2007).

A number of these devices can be integrated into a Wireless Body Area Network (WBAN), a new enabling technology for health monitoring. A WBAN include a number of physiological sensors such as ECG sensor for monitoring heart activity, EMG (electromyography) sensor for monitoring muscle activity, an EEG

11
(electroencephalography) sensor for monitoring brain electrical activity and other sensors for their own application.

Recently, NASA Space Technology in ScienceDaily (2007) has developed ArterioVision software at NASA's Jet Propulsion Laboratory, Pasadena, Calif along with a standardized, painless, non-invasive ultrasound examination of the carotid artery, which carries blood from the heart to the brain to help doctors diagnose and monitor treatments for hardening of the arteries in its early stages, before it causes heart attacks and strokes. A carotid ultrasound used with the software to test and measure the thickness of the inner two layer of the carotid artery which is intima and media. Arterial thickening provides the earliest evidence in detect the disease process that leads to heart disease and stroke.

The needs of wireless technology in medical application also encourage an Academic chief fellow of the Division of Cardiology at the University of Medicine and Dentistry of New Jersey (UMDNJ), New Jersey Medical School, in ScienleDaily (2007) to invent a Wireless Technology System Developed to Speed Care Of Heart Attack Patients. This wireless system allows the cardiologists and nurses to receive their patient’s ECGs signals from home, another hospital, cars or anywhere they are. It will enables on-call cardiologists to view full ECGs signal on smart phones by simply pressing a button that sends an ECG over a wireless network, then paramedics mobilizing which are a team of cardiologists and nurses within minutes will arriving at their patient home who is having a heart attack, so it will cut in half the time it takes to begin the treatment.
The wireless telemedicine system also has been developed by Qiang and Mingshi (2005). It is a flexible system to transfer the biological signal to the remote hospital central to get the professional instructions from doctors. This wireless ECG transmission is integrates current personal digital assistant (PDA) technology and general packet radio service (GPRS) technology. This system is able to acquire ECG signals, display the ECG waveform on the PDA LCD screen and wirelessly transmit them to an authorized remote medical ECG management server. It is facilitate accurate system, reliable monitoring and transmission of patient’s ECG signal.

From the research that have been done in all over the world shows that the medical treatment also need to be improve using latest technology to get the better medical equipment to make sure the people can improve theirs life style with healthy condition. The faster technology in medical can make sure the fastest treatment to the patient. So, this research area must always be improved to give benefits to all people in the world. The wireless ECG which has a snapshot of what the heart pattern was during the heart attack, the doctor can almost immediately know what direction to go in when treating the patient. Since every moment is critical when dealing with heart attacks, the minute or two saved could be life-saving.
2.3 Heart Disease

In the earlier 1980, according to the Centers for Disease Control and Prevention, United States (2007), heart disease is the leading cause of death for both women and men almost in the world and it is also a major cause of disability. In the worldwide, coronary heart disease kills more than 7 million people each year. Heart disease is a broad term that includes several more specific heart conditions which are Coronary Heart Disease, Heart Attack, Ischemia, Arrhythmias, Cardiomyopathy, Congenital Heart Disease, Peripheral Arterial Disease (PAD). The most common heart condition is coronary heart disease, which can lead to heart attack and other serious conditions and the research from PubMed Central Journals (2007) shows that the Ischemia is the most common cause of death in the industrialized countries. So the earliest diagnosis and treatment using electrocardiography (ECG) has been developed to observe the disease signal. Papaloukas et al. (2003) has indicated that the development of suitable automated analysis techniques can make this method very effective in supporting the physician's diagnosis and in guiding clinical management.

2.4 The Electrocardiogram Diagnosis Techniques

In recent year, numerous research and algorithm have been developed for the work of analyzing and classifying the ECG signal. The classifying method which have been proposed during the last decade and under evaluation includes digital signal analysis, Fuzzy Logic methods, Artificial Neural Network, Hidden Markov Model, Genetic
Algorithm, Support Vector Machines, Self-Organizing Map, Bayesian and other methods with each approach exhibiting its own advantages and disadvantages.

Researchers including Li et al. (1995) and Hu et al. (1997) have studied that the ECG features can be extracted in time domain and Minami et al. (1999), Moraes et al. (2002) and Papaloukas et al. (2003) have studied in frequency domain. Some of the features extraction methods implemented in previous research include Discrete Wavelet Transform (Li et al., 1995), Karhunen-Loeve Transform (Jager, 2002), Hermitian Basis (Ahmadian et al., 2007) and other methods (Lin et al., 1989), (Frau et al., 2002).

2.5 ECG Features Extraction Algorithms: A Review of Previous Studies

The ECG analysis technique required the feature extraction and classifier stage. In the previous ECG analysis research, the feature extraction methods include Discrete Wavelet Transform has been discussed by Thakor et al. (1993), Li et al. (1995) and Clarek (1995), Optimal Mother Wavelet by Castro et al. (2000), Karhunen-Loeve Transform method by Jager (2002), Hermitian Basis functions by Ahmadian et al. (2007) and other features extraction methods by researchers Lin and Chang (1989) and Cuesta-Frau et al. (2002) and other features extraction methods.

Douglas et al. (1990) described an approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models. This technique classified by detecting and analyzing QRS complex and determining the R-R intervals to determine the ventricular arrhythmias. The Hidden Markov modeling approach combines structural and statistical knowledge of
the ECG signal in a single parametric model. The Hidden Markov modeling addresses the problem of detecting low amplitude P waves in typical ambulatory ECG recordings.

Zigel et al. (2000) presented the method of Synthesis Coding in their paper. The synthesis ECG compressor algorithm is based on analysis by synthesis coding, and consists of a beat codebook, long and short-term predictors, and an adaptive residual quantizer. Predetermined distortion level is used in feature extraction of ECG signal. Their algorithm uses a defined distortion measure in order to efficiently encode every heartbeat, with minimum bit rate, while maintaining a predetermined distortion level. Their proposed compression algorithms were found to have the best performances at any bit rate as stated in their paper.

Researcher Li et al. (1995) use the Wavelet transforms method including Thakor et al. (1993), Frau et al. (2002), Pretorius and Nel (2002) and Mahmoodabadi et al. (2005) because the results indicated that the DWT-based feature extraction technique yields superior performance. Li et al. (1995) has done the ECG analysis using wavelet transform. This method can distinguish the between the QRS wave and P, T wave. This technique also can distinguish noise, baseline drift and artifacts. So its can characterize the signal information analysis very well and suitable to process time-varying biomedical signals. The wavelet transform also capable of representing signals in different resolutions by dilating and compressing its basis functions as explain by Clarek (1995).
Park et al. (2008) applied two morphological feature extraction methods which are higher-order statistics and Hermite basis functions. Their research results showed that hierarchical classification method gives better performance than the conventional multiclass classification method. They used the support vector machines to compare the feature extraction methods and classification methods to evaluate the generalization performance. But the use of higher order models need more computation cost and caused over fitting problem in generalization performance. In term of accuracy, they found that their hierarchical classification method showed better classification performance than the conventional multiclass classification method with despite the loss in accuracy and sensitivities certain classes. The hierarchical classification improved the mean values of sensitivity mean. It agreed that their classification method can distinguish the multiclass heartbeats with the unbalanced data distribution.

Researcher Jager (2002) developed a new approach to feature extraction which is Karhunen Lo`eve Transform (KLT) which is an attractive and powerful approach to the feature-extraction and shape representation process. It has the solution if the probability densities of population of pattern vectors of a problem domain are unknown. The problem about this method is it is too sensitive to noisy pattern of ECG signal.

According to Ranjith et al. (2003) which used wavelet transforms to detect myocardial ischemia, the wavelet transform is obtained using the quadratic spline wavelet. These correspond to the detection of T wave and P wave. Their methods shown this method is having a comparatively higher sensitivity and nominal positive predictivity value. It is also can be easily extended to detect other abnormalities of the ECG signal. But this
method also has the limitation of this method is that the computations required are higher than those required by other methods. This is mainly because of the calculation of Wavelet Transform.

According to Kadbi et al. (2004) in their paper highlighted those three features for features extraction stage which are time-frequency features, 2-time domain features and 3-statistical feature. These three features have been used in their project because these features can overcome the limitations of other methods in classifying multiple kinds of arrhythmia with high accuracy at once. These methods have been combined with PCA method to reduce the redundancy caused by the frequency coefficient in the feature dimension to make sure the average of the classification accuracy can be increased.

Tinati et al. (2006) in the studies used wavelet-transform based search algorithm to use the energy of the signal in different scales to isolate baseline wander from the ECG signal. They first remove the artifacts which is the noise that induced in ECG signals that result from movements of electrodes. The baseline wanders that are considered as an artifact can affect inaccurate data when measuring the ECG parameters. In their study using the presented algorithm; it can eliminate the baseline drifts from the ECG signals without introducing any deformation to the signal and also from losing any clinical information of the signal.

Herrero et al. (2006) used the independent component analysis and matching pursuits for the features extraction for extracting additional spatial features from multichannel electrocardiographic recordings. It test the classification performance of 5 largest classes.
of heartbeats in the MITBIH arrhythmia database which are Normal Sinus Beats (NSB), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Premature Ventricular Contraction (PVC) and Paced Beats (PB). The performance of the system is remarkably good, with specificities and sensitivities for the different classes. They have a problem because the complicated separation between ventricular PBs and PVCs because of the inverted T wave.

Ahmadian et al. (2007) proposed a new piecewise modeling for approximation of ECG signal using Hermitian Basis. This method uses only the 5th order Hermitian basis functions. This method yields to weighting the approximation error of each segment based on its importance throughout the ECG complex. This method shows the total error obtained in this method is almost halved in comparison with similar non-segmented method. The disadvantage of this method is a small error could mislead the diagnosis.

2.6 ECG Training and Classification Algorithms: A Review of Previous Studies

The pattern recognition of the type of ECG waveform, there are different solutions presented in the literature have been proposed during the last decade and are under evaluation. In ECG training and classification analysis stages, some researchers have tried to maximize the detection level of accuracy in many different ways such as digital signal analysis (Papaloukas et al., 2003), Fuzzy Logic methods (Bortolan et al., 1989; Zong and Jiang, 1998; Lei et al., 2002), Artificial Neural Network (Yang et al., 1997; Silipo and Marchesi, 1998; Pretorius and Nel, 2002; Papaloukas et al., 2002; Gao et al., 2004), Hidden Markov Model (Hughes et al., 2003; Graja and Boucher, 2005), Genetic
Algorithm (Goletsis et al., 2004), Support Vector Machines (Osowski et al., 2004), Self-Organizing Map (Lagerholm et al., 2000), Bayesian and other method with each approach exhibiting its own advantages and disadvantages. But the most recent systems employ artificial neural networks (Papaloukas et al., 2003) to perform diagnoses since they have demonstrated great consistency in producing accurate results. The performance of the developed detection systems is very promising but they need further evaluation. The automatic detection of ECG waves is important to cardiac disease diagnosis. A good performance of an automatic ECG analyzing system depends heavily upon the accurate and reliable detection of the disease.

2.6.1 Neural Network

The classification of the ECG using Neural Networks (NNs) has become a widely used method in recent years. The network architectures for modeling process modeling in NNs include the feedforward network, the radial basis function (RBF) network, recurrent network, and other advanced network architecture as explained by Centre for Process Analytics and Control Technology (1999) and Sjoberg (2004). The efficiency of these classifiers depends upon a number of factors including network training. It has the inputs models in the training parameters and the output indicated the point at which training should stop. Simple feed forward Neuron Model shows in Figure 2.1 by Gao et al. (2004). Researcher from Harvard University, Sordo (2002) indicated that the training and testing of the models was based on the results from the signal database of the normal patient and heart disease patient.
In designing an ECG classifier based on Neural Network (NN), the normal procedure is to firstly train the network by presenting it with training data that is representative of the unknown data it is likely to experience during the classification process. A well-chosen training algorithm, results in a NN which is capable of generating a non-linear mapping function with the capability of representing relationships between given ECG features and heart disease disorders. A well designed NN will exhibit good generalization when a correct input-output mapping is obtained even when the input is slightly different from the examples used to train the network.

![Simple Feed Forward Neuron Model](image)

**Figure 2.1** Simple Feed Forward Neuron Model by Gao et al. (2004)

Researchers Silipo and Marchesi (1998) also developed an Automatic ECG Analysis based on Artificial Neural Network. This project present the result by carrying out the classification tasks for the most common features of ECG analysis which are arrhythmia, myocardial ischemia and chronic alterations and achieve high classification accuracy. Another researches Papaloukas et al. (2002) developed an automated technique for the ischemic detection based on the recordings from European Society of Cardiology (ESC)
ST-T database in order to train the network for beat classification also achieve high accuracy rate.

According to researchers Gao et al. (2004) from the National University of Ireland which has developed a diagnostic system for cardiac arrhythmias from ECG data, using an Artificial Neural Network (ANN) classifier based on a Bayesian framework. The Bayesian ANN Classifier is built by the use of a logistic regression model and the back propagation algorithm. A dual threshold method is applied to determine the diagnosis strategy and suppress false alarm signals. This system consists of three basic modules which are a Server, multiple Client Machines and BAN-Hubs which used real time patient biosignal data provides earlier information and high classification accuracy.

2.6.2 Neuro-Fuzzy Approach

The idea of the ECG analysis and classification using Neuro Fuzzy has been start around 1990, yet it remains one of the most important indicators of proper heart disease classification today. The most difficult problem faced by an automatic ECG analysis is the large variation in the morphologies of ECG waveforms, it happens not only for different patients or patient groups but also within the same patient. So the Neuro Fuzzy is the most suitable technique because it is more tolerance to morphological variations of the ECG waveforms.

Researcher Linh et al. (2003) have studied in depth on the Neuro-Fuzzy approach to the recognition and classification of heart rhythms on the basis of ECG waveforms. It uses
the new approach of heart beat recognition. This project is the resolution for the problem of less sensitivity to the morphological variation of the ECG. It combining two techniques which are characterization of the QRS complex of ECG by Hermite polynomials and using the coefficients of Hermite kernel expansion as the features of the process and the application of the modified neuro-fuzzy TSK network for ECG pattern recognition and classification.

The performance enhancement using proposed method in Neuro-Fuzzy using autoregressive model coefficients, higher-order cumulant and wavelet transform variances as features by Engin (2004) and Papaloukas et al. (2003) and can solve the problem to detect more heart disease types in high accuracy.

The Neuro fuzzy Techniques which refers to the combination of fuzzy set theory and neural networks with the advantages of both which are handle any kind of information, numeric, linguistic, logical, imperfect information, resolve conflicts by collaboration and aggregation, self-learning, self-organizing and self-tuning capabilities, no need of prior knowledge of relationships of data, mimic human decision making process, and fast computation using fuzzy number operations in order to do the classification task.
2.6.3 Hidden Markov Models

This technique was successfully used since the mid 1970s to model speech waveforms for automatic speech recognition. The hidden Markov modeling approach combines structural and statistical knowledge of the ECG signal in a single parametric model. The model constructed contains multiple states per extraction field, model parameter and training algorithms as explained by Seymore et al. (1999).

Researchers Douglas et al. (1990) described an approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models. This technique classified by detecting and analyzing QRS complex and determining the R-R intervals to determine the ventricular arrhythmias. The Hidden Markov modeling approach combines structural and statistical knowledge of the ECG signal in a single parametric model. Model parameters are estimated from training data using an iterative, maximum likelihood reestimation algorithm. This method has ability of beat detection, segmentation and classification, with highly suitable to the ECG problem. Its approach addresses a waveforms modeling, multichannel beat segmentation and classification, and unsupervised adaptation to the patient’s ECG.

Cheng and Chan (1998) have discovered the method of Hidden Markov Model in classifying arrhythmia. They have developed a fast and reliable method of QRS detection algorithm based on a one-pole filter which is simple to implement and insensitive to low noise levels. The disadvantages are that the observations are very sensitive to baseline wander, DC drift and heart rate variation. The HMM method also is
not sufficient to represent one particular type of beat. This is because some beats exhibit large variations in the morphologies of their ECG signals. Therefore, several HMMs are needed for certain some beats.

### 2.6.4 Support Vector Machine

The Support Vector Machine-Based Expert System that have been described by Burges (1998), Osowski et al. (2004) and Walt et al. (2006) also the best method to apply in ECG analysis. The recognition system that uses the support vector machine (SVM) working in the classification mode. Support vector machines map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separates the data. The separating hyperplane shows the maximize distance. The larger the distance between these parallel hyperplanes, the better the generalisation error of the classifier.

According to Osowski et al. (2004) have performed their studies of Heartbeat Recognition using Support Vector Machine-Based Expert System. This recognition system has used the different preprocessing methods for generation of features which are higher order statistics (HOS) while the second is the Hermite characterization of QRS complex for the registered ECG waveform. Their paper presented the combination of multiple classifiers by the weighted voting principle. In their studies, stated that a good recognition system should depend on the features representing the ECG signals in such a way, that the differences among the ECG waveforms are suppressed for the waveforms of the same type but are emphasized for the waveforms belonging to different types of
beats. It is an important item, since the observed signal is a high variation of signals among the same types of beats. These two different preprocessing methods of the data, cooperating with SVM classifier, that have been integrated into one expert system have proven in improve the overall accuracy of heartbeat recognition.

In previous project of the classification of Myocardial Ischemia has been developed by Zimmerman and Povinelli (2004) to improve an algorithm for myocardial ischemia classification created by Langley et al. of specificity value. The proposed algorithms have been implemented with the support vector machine classifier. The radial basis function (RBF) has been used as the kernel of the support vector machine as a featured vector. But this proposed algorithm for myocardial ischemia classification resulted that is not able to improve the method developed by Langley et al. It is happened because the tradeoff that occurs between specificity and sensitivity is too great. Increasing the specificity caused the sensitivity to drop, and automatically decreased the overall accuracy. So completely independent of the ST deviation values must be done in order to increase specificity and accuracy.

Mehta and Lingayat (2007) indicate the application of Support Vector Machine (SVM) in QRS detection using entropy and combined entropy criterion. The advantages of using SVM are the ability to find a hyperplane that divides samples in to two classes with the widest margin between them, and the extension of this concept to a higher dimensional setting using kernel function to represent a similarity measure on that setting. This algorithm performs better as compared with published results of other QRS detectors tested on the same database and depends strongly on the selection and the variety of the
ECGs included in the training set, data representation and the mathematical basis of the classifier.

### 2.6.5 Self-Organizing Map

Meanwhile, in the ECG analysis of the Ischemia Detection with a Self-Organizing Map Supplemented by Supervised Learning has been developed in 2001 by Papadimitriou et al. It is to solve the problem of maximizing the performance of the detection of ischemia episodes. The basic self-organizing map (SOM) algorithm modified with a dynamic expansion process controlled with entropy based criterion that allows the adaptive formation of the proper SOM structure. This extension proceeds until the total number of training patterns that are mapped to neurons with high entropy reduces to a size manageable numerically with a capable supervised model. Then, a special supervised network is trained for the computationally reduced task of performing the classification at the ambiguous regions only. The utilization of sNet-SOM with supervised learning based on the radial basis functions and support vector machines has resulted in an improved accuracy of ischemia detection.

### 2.6.6 Fuzzy logic

Zong and Jiang (1998) described the method of fuzzy logic approach single channel ECG beat and rhythm detection. The method summarized and makes use of the medical knowledge and diagnostic rules of cardiologists. Linguistic variables have being used to represent beat features and fuzzy conditional statements perform reasoning. The
algorithm can identified rhythms as well as individual beats. This method also handling the beat features and reasoning process is heuristic and seems more reasonable as stated in their paper. It also presented that this method may be of great utility in clinical applications such as multi-parameter patient monitoring systems, where many physiological variables and diagnostic rules exist.

2.6.7 Bayesian

According to Gao et al. (2004) point out that the Bayesian network are improved methods in determining the arrhythmia diagnosis system. This method is able to deal with nonlinear discrimination between classes, incomplete or ambiguous input patterns, and suppression of false alarms. It develops new detection schemes with a high level of accuracy. This approach have been find out as a potentially useful for generating a pattern recognition model to classify future input sets for arrhythmia diagnosis. The Bayesian network have been proof that it have a capability of uncertainty management to work with the dual threshold method that could be used to control a diagnosis strategy and suppress false alarm signals in future improvement.

Wiggins et al. (2005) had evolving a Bayesian classifier for ECG classification. The patients classification was according to statistical features extracted from their ECG signals using a genetically evolved Bayesian network classifier and the identification depend on the variables of interest. The Bayesian Network has an ability to handle missing data points and its lower requirement of information based on a priori knowledge of the system’s variable dependencies is its major benefits. It is a relatively
new tool that identifies probabilistic correlations in order to make predictions or assessments of class membership that could solve many complex problems exist in identifies the data for the variables of interest. This method shown it is very easy to implement, and one of the research area that are good to be discovered. The limitation of their studies has been the method for binary discretization used after feature extraction because of small size of the data set.

2.6.8 Genetic Algorithms

ECG analysis researchers including Nugent et al. (2002) have studied in depth on the Prediction models in ECG classifiers using genetic programming approach. In their studies they developed the prediction models to indicate the point at which training should stop for Neural Network based ECG classifiers in order to ensure maximum generalization. According to them, this good wave prediction could exhibit good generalization. They found that it could give benefit to developers of Neural Networks, not only in the presented case of Neural Network based ECG classifiers, but indeed any classification problems.

2.6.9 Autoregressive Model

Ge et al. (2002) have extended the study of Cardiac arrhythmia classification using autoregressive modeling. This Computer-assisted arrhythmia recognition have been proposed to classify normal sinus rhythm (NSR) and various cardiac arrhythmias including atrial premature contraction (APC), premature ventricular contraction (PVC),
superventricular tachycardia (SVT), ventricular tachycardia (VT) and ventricular fibrillation (VF). Their studies have shown the AR coefficients were classified using a generalized linear model (GLM) based algorithm in various stages. From their study, they found that the AR modeling is useful for the classification of cardiac arrhythmias, with reasonably high accuracies. From the study, they found that AR modeling based classification algorithm has demonstrated good performance in classification. The algorithms are also easy to implement and the AR coefficients can be easily computed. AR modeling can lead to a low cost, high performance, simple to use portable telemedicine system for ECG offering a combination of diagnostic capability with compression. Therefore, it revealed that enhancement is suitable for real-time implementations and can be used for compression as well as diagnosis.

2.6.10 Other methods

Bousseljot and Kreiseler (1998) have introduced a new technique for the ECG interpretation by waveform recognition without feature extraction process. Their studies found that the method for ECGs computer-aided interpretation by signal pattern comparison have presented many advantages which are it is no feature extraction for the ECG and also no limitation of the diagnostic statements. Their system also shown that it inclusion of rare diseases by specialized ECG databases and it have a possibility of extending the databases as bases of knowledge without changes of the algorithm. The method also performed robustness with respect to disturbances or signal failures in a lead and has learning ability by inclusion of the results of current patient examinations.
Also, the study showed cost advantages by use of existing PCs and inclusion of further patient information from the database for making diagnoses.

Minfen et al. (1998) defines a novel method for extracting time-varying rhythms using Multiresolution decomposition to investigate the transition of clinical EEG signals. The method proposed in their paper is more flexible and accurate due to the better matching in time-frequency characteristics of EEG signal for extracting 4 kinds of EEG rhythms. The results of have demonstrated the superior performance of the new wavelet packet analysis algorithm. But this method still has problems which are the optimal segmentation length resolution for such analysis, which is obviously related to the time-varying characteristics of the EEG signals observed. So they need build an optimal segmentation-based adaptive algorithm to improve the result of the signal analysis.

A Novel Digital Filter Based Enhancement of QRS Complex of ECG for Improved Arrhythmia Classification has been discussed by Mahesh and Lavanavarjit (2002). The filter is used in this method to enhance the QRS complex of the ECG wave. This algorithm is tried to prevent the problems that are normally encountered in detecting ECG characteristic points which are noise in the process of measurement, non-zero baseline, baseline drift, high P and T waves and artifacts. The algorithm gives remarkable improved the results of against noise, base-line drift and other problems encountered in the detection process of the QRS complex of ECG. Novel Digital Filter is also a simple algorithm that makes the method fast and suitable in any real time application.
Based on Tsipouras et al. (2005), a knowledge-based method for arrhythmic beat classification and arrhythmic episode detection and classification using only the R-interval signal extracted from ECG recordings can also be done to get the highest accuracy in the ECG signal processing. The method is advantageous because the signal can be extracted with high accuracy even for noisy or complicated ECG recordings, while the extraction of all other ECG features or any other type of ECG analysis is seriously affected by noise and the processing time is reduced since only one feature is required compared to other methods that use more features or other types of ECG analysis.

Olvera (2006) has reviewed the Matched Filter algorithms for their feature extraction method in ECG analysis. The matched filter was used to detect different signal features of interest were the QRS Complex, the R-R intervals, and the ST segments on a human heart ECG signal. The matched filters will maximize the signal-to-noise ratio for a noisy signal so that the signal of interest that can be extracted. It is used to extract the ST segment which is the most important section to be extracted. The matched filter outputs were better than expected for the Normal and Long Term ST ECGs but for the Arrhythmia and Sudden Cardiac Death (Cardiac Arrest) ECGs, the results were not as good. Unfortunately, this method was not the case with the Arrhythmia and Cardiac Arrest ECGs when using R-wave peak detection. The problem was due to the detection threshold used, not the matched filter implementation. The peak detection resolution was not sensitive enough to distinguish between the R-wave peaks and the ST segment peaks due to the use of a feature-length, sliding detection window. The use of QRS complex detection had much better results. The QRS complex matched filter detection with the
Cardiac Arrest ECG was much better in detection, especially in the noisy areas where the R-wave peak matched filter failed. A better method for QRS complex detection along with R wave peak and ST segment detection would involve using an adaptive filter to whiten the noise in the QRS complex. This method can be improved by using different form of matched filter and better threshold detection, and then the matched filter ECG feature extraction could be made more successful.

### 2.7 Summary of an Approach of Electrocardiogram Analysis Algorithms

Table 2.1 shows the summary of the approach of ECG Analysis system that has been done in previous research.

**Table 2.1 Summary of an Approach of Electrocardiogram Analysis**

<table>
<thead>
<tr>
<th>Method</th>
<th>Researcher</th>
<th>Comment and Description</th>
<th>Overall performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karhunen Lo`eve Transform</td>
<td>Jager (2002)</td>
<td>Attractive and powerful approach to the feature-extraction and shape representation process. It has the solution if the probability densities of pattern vectors population of a problem domain are unknown.</td>
<td>Most of accuracy 80%</td>
</tr>
<tr>
<td>Wavelet Transform</td>
<td>(Thakor et al., 1993; Li et al., 1995; Frau et al., 2002)</td>
<td>Time frequency analysis. Ability to reconstruct the signal from the wavelet decomposition and to preserve the energy under</td>
<td>Accuracy more than 90%</td>
</tr>
<tr>
<td>Pretorius et al., 2002; Mahmoodabadi et al., 2005</td>
<td>the transformation. By decomposing signals into elementary building blocks that are well localized both in time and frequency, the WT can characterize the local regularity of signals.</td>
<td>Sensitivity attained over 90%</td>
<td></td>
</tr>
<tr>
<td>Artificial Neural Network</td>
<td>Heden et al. (1995)</td>
<td>ANN method can enhance the sensitivity of the conventional P waves detection. The conventional criteria had a much lower sensitivity in the absence of P wave data which is 30.9%. The corresponding sensitivity for the ANN is 94.5%.</td>
<td>Accuracy 80% to 99% Sensitivity more than 90%</td>
</tr>
<tr>
<td>Silipo and Marchesi (1998)</td>
<td>This approach capable of dealing with the ambiguous nature of ECG signal.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papaloukas et al. (2002)</td>
<td>Obtained 80% to 90% accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gao et al. (2004)</td>
<td>The results presented in this project shows that more than 90% prediction accuracy may be obtained.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et al. (2006)</td>
<td>Exhibited the best performance and excellent model for the</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Hidden Markov  | Douglas et al. (1990) | P-wave detection results show an accurate detection of the low amplitude P wave in ECG recordings. | Specificity
90%

Accuracy
50% to 80%

Sensitivity
80% to 95% |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bardonova (2000)</td>
<td>The Hidden Markov models which applied to wavelet transform can improve the biosignal analysis.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hughes et al. (2003)</td>
<td>It is significantly better than similar models trained on the raw time ECG series data. Hidden Markov modeling can addresses the problem of detecting low amplitude P waves in typical ambulatory recordings.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodrigo et al. (2006)</td>
<td>This system combining Wavelet Transform and Hidden Markov Models, it obtained high beat detection performance with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Author(s) and Year</td>
<td>Details</td>
<td>Results/Notes</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>Osowski et al. (2004)</td>
<td>The results of the performance recognition of heart rhythm types confirmed the reliability for the approach.</td>
<td>Accuracy attained over 80%</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>Tsiouras et al. (2007)</td>
<td>Performed a good efficiency of the recognition of the normal and different types of beats representing the arrhythmias.</td>
<td>Specificity more than 90%</td>
</tr>
<tr>
<td>Fuzzy Hybrid Neural Network</td>
<td>Linh and Osowski (2001)</td>
<td>This method shows the simplicity, good recognition rate, and fast performance. Perform a good efficiency of the recognition for the different types of heart beats</td>
<td>Sensitivity 60% to 70%</td>
</tr>
<tr>
<td>Bayesian</td>
<td>(Wiggins et al., 2005; Popescu et al., 1998)</td>
<td>A new tool to make predictions or assessments of class membership. The BN is an excellent method for making decisions based on collected information and makes those decisions in a very similar way to that of a physician: by taking each individual piece of information and assessing probabilities of how it affects the final diagnosis.</td>
<td>Accuracy above 80%</td>
</tr>
<tr>
<td>Method</td>
<td>Performance</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Genetic Algorithms (Goletsis et al., 2004; Nugent et al., 2002)</td>
<td>It obtained performance 91% in terms of both sensitivity and specificity. Need to combine different classifiers to obtain a better result.</td>
<td>more than 90%</td>
<td>90%</td>
</tr>
<tr>
<td>Self-Organizing Maps (Papadimitriou et al., 2001)</td>
<td>The average beat classification accuracy is 76.51%.</td>
<td></td>
<td>more than 70%</td>
</tr>
</tbody>
</table>

2.8 **Summary**

In the literatures, most researchers have developed the system based on the various techniques and algorithms. Each technique presented in the previous project of ECG analysis has their advantages and disadvantages. The performance of the developed detection systems is very promising but they need the further evaluation. The automatic detection of ECG waves is important to cardiac disease diagnosis. A good performance of an automatic ECG analyzing system depends heavily upon the accurate and reliable detection of the QRS complex, as well as the T and P waves and most of the researches only depend on certain disease.
From the literatures, it is indicated that the ECG analysis systems developed by using hybrid algorithms are too complex. But the hybrid techniques that have been implemented in the project recently shown that it yield the better result analysis of heart disease classification. In the systems developed, such technique needs a very precise code in order to achieve good efficiency and accuracy. Therefore, the present work in this thesis is an attempt to simplify and maximize the accuracy of the algorithms.

From the reviewed, there are improvement for ECG analysis in feature extraction and classification techniques, it is found that Artificial Neural Network and hybrid methods is one of the latest ECG analysis techniques researches particularly in biosignal processing for medical application which are being carried out by biomedical researchers. Therefore, this type of research is definitely worth for further study. Our research mainly aimed to use the selected algorithms for feature extraction and classification task to enhance the result of accuracy and expand the types of heart disease that can be classify. An ECG analysis system with fast and easily will be developed. This study will be carried out by the simulation works. In the simulation, MATLAB code will be used due to its capability to give good predictions signal processing. This study is expected to be an initial attempt to the development of ECG analysis module.
CHAPTER 3
RESEARCH METHODOLOGY AND SYSTEM DESIGN

3.0 Overview

This chapter describes the electrocardiogram (ECG) signals analysis modeling by using MATLAB. Here, the methods of analysis are discussed. The analysis system based technique in most ECG analysis was performed in three stages: (1) division of the preprocessing, (2) feature extraction by computation of Discrete Wavelet Transform coefficient (DWT) and (3) classification procedures using Adaptive Neuro Fuzzy inference System (ANFIS) trained with the backpropagation gradient descent method in combination with the least squares method for analysis purposes will be explained. Also, a brief simulation procedure is presented in this chapter.

3.1 Overview of Electrocardiogram Analysis System

The health of a population is a fundamental element contributing to progressive sustainable development in all regions of the world. Virtually all sciences contribute to the maintenance of human health and the practice of medicine. The development and implementation of science and technology in the medical application tools such as ECG will help to enhance the human healthcare and can assist people to check their health condition with fast and accurate. The invention of the new analysis method of medical
instrumentation also can help to improve the efficiency and powerful medical applications.

The analysis of ECG is widely used for diagnosing many cardiac diseases, which are the main cause of mortality in developed countries. Biosignal processing techniques such as ECG analysis system offer a powerful tool to simulate the human heart signal. The performance of such detection systems relies heavily on the accuracy and reliability in the detection of the signals, which is necessary to determine the heart disease. The arrhythmia and ischemia detection of ECG wave is an important topic.

This section explores the methods used to collect data for analysis, the isolation of the necessary data, and the experiments used to analyze the biosignal data. The overall technique of Electrocardiogram Analysis is shown in Figure 3.1 below.

![Figure 3.1 Overall Techniques for Electrocardiogram Analysis](image-url)
3.2 System Requirement

This project is developed by using MATLAB (MathWorks Inc.) software tool, the numerical computing environment and programming language software for modeling the heart signal in complex algorithms. The ANFIS were also used as classifier tools in Fuzzy Logix Toolbox. MATLAB is a high-level language and interactive environment that enables to perform computationally intensive tasks faster than with traditional programming languages such as C, and C++. This software is among the most commonly used development languages. MATLAB codes also being used because it could read the raw data of ECG signal easily. The input ECG signal are imported from the data files .dat and also the excel file .xls.

3.3 Design and Architecture

In the previous ECG analysis research, numerous research and algorithm have been developed for the work of analyzing and classifying the ECG signal. The ECG analysis techniques are reviewed in and evaluate proposed methods of the classification methods. The ECG analysis techniques have been identified and it required several stages which are shown in the Figure 3.2.
3.4 Electrocardiogram Signal Analysis Procedure

The methods presented here are divided into three pieces of work. Firstly, procedures to identify and annotate of ECG signal for Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia characteristic. Secondly, a strategy is presented for extracting the features vector for each sample of selected heart disease using an algorithm that exploits the coefficient derived from Discrete Wavelet Transform. Lastly, this part presented the procedures of classification process using Adaptive Neural Fuzzy Inference System modeling.

3.4.1 Signal Data Preparation

The ECG recording signals data are partitioning into cardiac cycles, and detection of the main events and intervals in each cycle have been done. The ECG signals which consist of P,Q,R,S and T wave have been detect based on their wave characteristic such as
position, amplitude and intervals are shown in Table 3.1. The major features such as the QRS amplitude, R-R intervals, and wave’s slope of ECG signal can be used as features to create the mapping structure are also identified.

Table 3.1 Three Main Phases Exist: P, QRS, and T

<table>
<thead>
<tr>
<th>Section of Electrocardiogram</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Wave</td>
<td>Record the electrical activity through the upper heart chambers (Atria Excitation)</td>
</tr>
<tr>
<td>QRS-Complex</td>
<td>Record the movement of electrical impulses through the lower heart chambers. (Atria repolarization + Ventricle depolarization)</td>
</tr>
<tr>
<td>T- Wave</td>
<td>Corresponds to the period when the lower heart chambers are relaxing electrically and preparing for their next muscle contraction. (Ventricle repolarization)</td>
</tr>
<tr>
<td>ST segment</td>
<td>Corresponds to the time when the ventricle is contracting but no electricity is flowing through it.</td>
</tr>
</tbody>
</table>

3.4.1.1 Signal Data Characteristics

The characteristic for each sample of heart disease must be studied to make sure the characteristic is correct with the exact characteristics that have been identified by the doctor. The characteristics of each disease are described below.
Ischemia template

- Inversion of T wave
- Decrease in amplitude or disappearance of R wave
- Shift of ST segment

Tachycardia Arrhythmia template

- Tachyarrhythmias are accelerated atrial or ventricular rates that exceed what is considered normal
- Beat too fast
- Regular
- Presence or absence of atrial depolarization (P wave, flutter waves).

Diagnosis of cardiac arrhythmia cannot be considered complete without accounting for atrial activity.

Bradycardia Arrhythmia template

- Rhythms producing cardiac slowing are grouped together as bradyarrhythmias
- Beat too slow
- No P wave
The standard value of normal signal characteristics for Amplitudes and Durations of ECG Parameters are shown in Table 3.2 and Table 3.3 below.

### Table 3.2 Amplitudes Values for Normal ECG Signal

<table>
<thead>
<tr>
<th>Wave</th>
<th>Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>P Wave</td>
<td>0.25mV</td>
</tr>
<tr>
<td>R Wave</td>
<td>1.60mV</td>
</tr>
<tr>
<td>Q Wave</td>
<td>25% of R wave</td>
</tr>
<tr>
<td>T Wave</td>
<td>0.1 to 0.5 mV</td>
</tr>
</tbody>
</table>

### Table 3.3 Durations Values for Normal ECG Signal

<table>
<thead>
<tr>
<th>Wave</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-R interval</td>
<td>0.12 to 0.20 sec</td>
</tr>
<tr>
<td>Q - T interval</td>
<td>0.35 to 0.44 sec</td>
</tr>
<tr>
<td>S-T segment</td>
<td>0.05 to 0.15 sec</td>
</tr>
<tr>
<td>P wave interval</td>
<td>0.11 sec</td>
</tr>
<tr>
<td>QRS interval</td>
<td>0.09 sec</td>
</tr>
</tbody>
</table>

### 3.4.2 Preprocessing

Signal data acquisition is the first stage of record and capture data from the patient. Data files of the ECG recordings were imported into MATLAB software where all computations were carried out. Initially, long sequences of data were processed to obtain the discrete wavelet transform of the entire recording of each subject. Figure 3.3 shows the preprocessing step of ECG analysis.
3.4.2.1 Databases and Data Sources

The ECG signals were downloaded and recorded from the PhysioBank database using MIT-BIH Arrhythmia Database and Intracardiac Atrial Fibrillation Database which are generally recognized as a standard test bench for the evaluation of arrhythmia detectors and basic research of cardiac dynamics. The databases were also referred to a Hospital Universiti Kebangsaan Malaysia (HUKM) cardiology test unit to investigate the ECG signal by patient monitors in real ICU settings.

3.4.2.2 Normalization Data

In this study, a rectangular window, which was formed by 1200 discrete data, was selected so that it contained a stationary ECG signal frame. One frame from each of the recorded ECG signals was used for extracting the inputs of the classifiers.
3.4.2.3 Data Selection

Based on Table 3.4, the ECG recordings consist of 352 subjects, 180 data samples were used for training and 172 data samples were used for testing. The samples belong to four categories: Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Class</th>
<th>Training set</th>
<th>Testing set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1</td>
<td>45</td>
<td>44</td>
<td>89</td>
</tr>
<tr>
<td>Bradycardia Arrhythmia</td>
<td>2</td>
<td>45</td>
<td>42</td>
<td>87</td>
</tr>
<tr>
<td>Tachycardia Arrhythmia</td>
<td>3</td>
<td>45</td>
<td>44</td>
<td>89</td>
</tr>
<tr>
<td>Ischemia</td>
<td>4</td>
<td>45</td>
<td>42</td>
<td>87</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>172</td>
<td>352</td>
<td></td>
</tr>
</tbody>
</table>

3.4.2.4 Data Processing

The Class 1 consists of 89 Normal subjects, Class 2 contains 87 subjects suffering from Bradycardia Arrhythmia, Class 3 has 89 subjects suffering from Tachycardia Arrhythmia disease and Class 4 is being obtained for 87 subjects from Ischemia disease as shown in Table 3.4 above.

3.4.3 Features Extraction using Discrete Wavelet Transform

Features extraction is extracting and converting the input data information into a set of features which called feature vector, by reducing the data representation pattern. The features set will extract the relevant information from the input data in order to perform
the classification task. The transform of a signal is just another form of representing the signal. It does not change the information content present in the signal.

Wavelet theory is the mathematics associated with building a model for a signal, system, or process. A wavelet is a wave which has its energy concentrated in time. It has an oscillating wavelike characteristic but also has the ability to allow simultaneous time and frequency analysis and it is a suitable tool for transient, non-stationary or time-varying phenomena. WT has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time-frequency resolution in all frequency ranges.

From the Figure 3.4 above, the signals with sharp changes might be better analyzed with an irregular wavelet than with a smooth sinusoid, as quoted in Mahmoodabadi et al. (2005). Also, local features can be described better with wavelets that have local extent.

The Wavelet Transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions. It is capable of representing signals in different resolutions by dilating and compressing its basis functions. The basis function in wavelet analysis is defined by two parameters which are scale and translation. A basis function...
which is mother wavelet is used in wavelet analysis. For a wavelet of order $N$, the basis function can be represented in Equation 3.1:

$$\psi (n) = \sum_{j=0}^{N-1} (-1)^j c_j (2^n + j - N + 1)$$  \hspace{1cm} (3.1)

### 3.4.3.1 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT), which is a time-scale representation of the digital signal is obtained using digital filtering techniques, is found to yield a fast computation of Wavelet Transform. It is easy to implement and adopts dyadic scales and translations in order to reduce the amount of computation time, which results in better efficiency of calculation.

The DWT which also referred to as decomposition by wavelet filter banks is computed by successive low pass filter (LPF) and high pass filtering (HPF) of the discrete time-domain signal as the process shown graphically in Figure 3.5 below.

![Figure 3.5 Filter Banks Signal Decomposition](image)

Figure 3.5 Filter Banks Signal Decomposition
The different cutoff frequencies are used for the analysis of the signal at different scales to measure the amount of detail information in the signal and the scale is determined by upsampling and downsampling process where D and A denoting for details and approximations, while c representing coefficients. The approximations of the signal are what define its identity while the details only imparts nuance.

![Three-level Wavelet Decomposition Trees](image.png)

**Figure 3.6** Three-level Wavelet Decomposition Trees

Figure 3.6 show the decomposition process is iterative. It connects the continuous-time multiresolution to discrete-time filters. The signal is denoted by the sequence input signals \( x[n] \), where \( n \) is an integer passed through a series of high-pass filters to analyze the high frequencies, and through a series of low-pass filters to analyze the low frequencies. Each stage consists of two digital filters and two downsamplers by 2 to produce the digitized signal. The low pass filter is denoted by \( G_o \) while the high pass filter is denoted by \( H_o \). At each level, the high pass filter produces detail information; \( d[n] \), while the low pass filter associated with scaling function produces coarse approximations, \( a[n] \). The downsampled outputs of first high pass filters and low-pass
filters provide the detail, $D_1$ and the approximation, $A$. The first approximation, $A_1$ is decomposed again and this process is continued. The filtering and decimation process is continued until the desired level is reached. The maximum number of levels depends on the length of the signal. Only the last level of approximation is save among all levels of details, which provides sufficient data. The DWT of the original signal is then obtained by concatenating all the coefficients, $a[n]$ and $d[n]$, starting from the last level of decomposition. The signal decomposition can mathematically be expressed in Equation 3.2 and Equation 3.3:

\[ y_{hl}[k] = \sum x[n] g[2k - n] \]  
\[ y_{lo}[k] = \sum x[n] h[2k - n] \]

(3.2)  
(3.3)

With this approach, the time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies.

In DWT the signals can be represented by approximations and details. The detail at level $j$ is defined as Equation 3.4:

\[ D_j = \sum_{k \in Z} a_{j,k} \psi_{j,k}(t) \]

(3.4)

Where, $Z$ is the set of positive integers.
Then, the approximation at level \( J \) is defined as Equation 3.5:

\[
A_i = \sum_{j > J} D_i
\]  

(3.5)

Finally, the signal \( f(t) \) can be represented by Equation 3.6:

\[
f(t) = A_j = \sum_{j=j} D_j
\]  

(3.6)

In DWT where a scaling function is used, which are related to low-pass and high-pass filters, respectively. The scaling function can be represented as Equation 3.7 and Equation 3.8:

\[
\Phi(n) = \sum_{j=0}^{N-1} c_j \Phi(2n - j)
\]  

(3.7)

\[
\Phi_{jk}(t) = 2^{j/2} \Phi(2^j t - k)
\]  

(3.8)

3.4.3.2 DWT Implementation

In the scope of this thesis, feature extraction was conducted by applying wavelet analysis techniques to patient data, thus providing ECG characteristic point detection capabilities. Since most recently published detectors are based on standard database libraries and limited wave detection, this application is an attempt to expand the horizons of current research efforts.
The input selection of feature extraction methods applied in this thesis has to select well to make sure which components of a inputs best represent the given pattern of ECG signals. Since the details wavelet coefficients contain a significant amount of information about the signal, the detail wavelet coefficients of ECG signal of each subject were computed. The procedures of DWT implementation is describe as follow in Figure 3.7.

![Figure 3.7 Feature Extraction Techniques](image)

### 3.4.3.3 Features Extraction Procedures

Selection of appropriate wavelet and the number of decomposition level is very important in DWT. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients.
The general wavelet decomposition of DWT procedure involves three steps. The result of decomposed signal will show the important details and approximation coefficients which represent the original signal. The basic version of the procedure follows the steps described below.

- Choose a wavelet types.
- Choose a wavelet name.
- Choose a level $N$ which will compute the wavelet decomposition of the signal $s$ at level $N$.

The dwt wavelet types have been chosen in this features extraction method and the ECG signals were decomposed into time-frequency representations using single-level one-dimensional wavelet decomposition. The wavelet names of Daubechies wavelet filters db4 have been choosing and the number of decomposition levels was chosen to be 5. Thus, the ECG signals were decomposed into the details coefficients $D_1$-$D_5$ and one final approximation coefficient, $A_5$.

The result of applying the Daubechies wavelet of order 4 (db4) which is more suitable to detect changes of ECG signal is evaluated. The wavelet filter with scaling function more closely similar to the shape of the ECG signal achieved better detection. Db wavelet family is similar in shape to ECG signal and their energy spectrums are concentrated around low frequencies the signal is approximated by omitting the signals high frequency components.
The ECG signal and the details for five wavelet scales are schematically shown for better illustration in Figure 3.8.

**Figure 3.8** DWT Decomposition Step in ECG Analysis
3.4.3.4 Coefficients Extraction

The computed wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. Therefore, the computed details and approximation wavelet coefficients of the ECG signal were used as the features vector representing the signals.

In this study, from the original intervals of ECG signal, five standard measures parameters used are used. A signal of 75 discrete data was selected as considered ECG signals data. For each ECG signals, the detail wavelet coefficients of fourth level (75 coefficients) were computed. In order to reduce the dimensionality of feature vectors, statistics over the set of the wavelet coefficients were used. The following statistical features were used to represent the time–frequency distribution of the ECG signals: The flows of the calculated DWT coefficient are shown in Figure 3.9 below.

1. Energy of the wavelet coefficients of each ECG signals sample.
2. Maximum of the wavelet coefficients of each ECG signals sample.
3. Minimum of the wavelet coefficients of each ECG signals sample.
4. Mean of the wavelet coefficients of each ECG signals sample.
5. Standard deviation of the wavelet coefficients of each ECG signals sample.
Figure 3.9 Flowchart of DWT Coefficient Calculation

Start to calculate dB4 coefficient

Read data for each samples - Class 1, Class 2, Class 3 and Class 4

DWT Filtering Subband and Coefficient

Calculate the ‘Energy’ of $D_4$ coefficient for each sample

Calculate the ‘Maximum’ of $D_4$ coefficient for each sample

Calculate the ‘Minimum’ of $D_4$ coefficient for each sample

Calculate the ‘Mean’ of $D_4$ coefficient for each sample

Calculate the ‘Standard Deviation’ of $D_4$ coefficient for each sample

End
These features vector which were calculated for the D₄ frequency band, were used in classifying the ECG signals as shown in Table 3.5.

**Table 3.5** The Extracted Features of Four Exemplary from Four Classes

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Extracted features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal beat</td>
<td>Energy</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Bradycardia Arrhythmia</td>
<td>Energy</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Tachycardia Arrhythmia</td>
<td>Energy</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Ischemia</td>
<td>Energy</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
</tr>
</tbody>
</table>

The subband 4, D₄ of details coefficients from the wavelet decomposition structure has been extracted. These vectors are extracted at each scale without scale one, two and three. It is ignoring the higher levels of decomposition because it contains high frequency details and noise. These details are insignificant information that will not affect the classification accuracy and signal quality.
This implies that it is possible to delete information of very small magnitude in each subspace, resulting in much less data information being needed to reconstruct a very good approximation of the original signal as defined by Daubechies (1990).

Then, the output of the detail coefficients extracted from the signal will be defined as the input of ANFIS classifier as systematically shown in Figure 3.10.

![Figure 3.10 Analysis Framework](image)

3.4.4 Classification using Neuro Fuzzy

Decision making of classification was performed in two stages: selection of coefficients computing by DWT and the ANFIS classifiers. Four types of ECG beats (Normal, Tachycardia Arrhythmia, Bradycardia Arrhythmia, and Ischemia) obtained from the PhysioBank databases will be classified by ANFIS classifiers.
3.4.4.1 Neuro Fuzzy Approach

Neuro Fuzzy is a hybrid of artificial neural networks and fuzzy logic. Neuro Fuzzy networks are the realizations of the functionality of fuzzy systems using neural techniques. Neuro Fuzzy Network incorporates the human-like reasoning style of fuzzy systems through the use of fuzzy sets and a linguistic model consisting of a set of IF-THEN fuzzy rules as shown in Figure 3.11.

![Figure 3.11 Structure of Feedforward Neuro Fuzzy](image-url)

**Figure 3.11** Structure of Feedforward Neuro Fuzzy
The important part of fuzzy layer, it is responsible to analyze the distribution of data and group the data into the different membership values. This membership value is applied as the input vector to the Multi Layer Perceptron Neural Network classifier. The membership value also representing the parameter of each heart beat class.

This research use the output of DWT technique as features vector and ANFIS as a Neuro Fuzzy classifier for the ECG analysis, because based on the previous research; the accuracy rates achieved by the combined neural network model presented for classification of the ECG beats were to be higher than stand alone classifier model. The Neuro Fuzzy network is also more tolerant to the noise and less sensitive to the morphological changes of the ECG characteristic and ANFIS also plays an important role in dealing with uncertainty when making decisions in medical application.

3.4.4.2 ANFIS Model

ANFIS stands for Adaptive Neuro-Fuzzy Inference System. This technique brings the learning capabilities of neural networks to fuzzy inference systems. In ANFIS, Takagi-Sugeno type Fuzzy Inference System (FIS) is used. The learning algorithm tunes the membership functions Takagi-Sugeno type using the training input-output data. The output of each rule can be a linear combination of input variables. The final output is the weighted average of each rule’s output. The embedded fuzzy system in a neural fuzzy network can self-adjust the parameters of the fuzzy rules using neural network learning algorithms to achieve the desired results.
The ANFIS learning techniques provide a method for the fuzzy modeling procedure to learn information about data set, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input-output data. ANFIS constructs an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and simulated input output data pairs. It serves as a basis for building the set of fuzzy if-then rules with appropriate membership functions to generate the input output pairs.

The parameters associated with the membership functions are open to change through the learning process. The computation of these parameters (or their adjustment) is facilitated by a gradient vector, which provides a measure of how well the ANFIS is modeling the input output data for a given parameter set. Once the gradient vector is obtained, backpropagation or hybrid learning algorithm can be applied in order to adjust the parameters. Basic ANFIS architecture that has two inputs \( x \) and \( y \) and one output \( z \) is shown in Figure 3.12.

![Figure 3.12 Basic Structure of ANFIS Model](image-url)
Two Takagi-Sugeno if-then rules are shown in Equation 3.9 and Equation 3.10:

Rule1: If $x$ is $A_1$ and $y$ is $B_1$ THEN $f_1 = p_1x + q_1y + r_1$  

(3.9)

Rule2: If $x$ is $A_2$ and $y$ is $B_2$ THEN $f_2 = p_2x + q_2y + r_2$  

(3.10)

The nodes functions of ANFIS architecture in the same layer are described below:

Layer 1: Every node $I$ in this layer is a square node with a node function as in Equation 3.11 and Equation 3.12:

\[ 0_{1,i} = \mu_{A_i}(x), \text{ for } I = 1, 2 \]  

(3.11)

\[ 0_{1,i} = \mu_{B_{i-2}}(y), \text{ for } I = 3, 4 \]  

(3.12)

where $x$ is the input to node $I$, and $A_i$ (or $B_{i-2}$) is a linguistic label (such as “small”, “medium”, “large”) associated with this node. The $0_{1,I}$ is the membership function of a fuzzy set $A_i$ and it specifies the degree to which the given input $x$ satisfies the quantifier $A_i$. Usually is chosen $\mu_{A_i}(x)$ to bell-shaped with maximum equal to 1 and minimum equal to 0, such as the generalized bell function in Equation 3.13:

\[ \mu_{A_i}(x) = \frac{1}{1 + \left(\frac{(x-c_i)}{a_i} \right)^{2b_i}} \]  

(3.13)

or the Gaussian function represent in Equation 3.14:

\[ \mu_{A_i}(x) = e^{-\left((x-c_i)/a_i \right)^{2}} \]  

(3.14)

where $a_i$, $b_i$, $c_i$ is the parameter set. The membership function for $A_i$ can be any appropriate membership function, such as the Bell-shaped, Triangular or Gaussian.
When the parameters of membership function changes, chosen membership function varies accordingly, thus exhibiting various forms of membership functions for a fuzzy set $A_i$. Parameters in this layer are referred to as “premise parameters”.

**Layer 2:** Every node in this layer is a fixed node labeled as $\Pi$, whose output is the product of all incoming signals defined by Equation 3.15:

$$0_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \text{ for } I = 1, 2$$

Each node output represents the firing strength of a fuzzy rule.

**Layer 3:** Every node in this layer is a fixed node labeled $N$. The $i$th node calculates the ratio of the rule’s firing strength to the sum of all rules’ firing strengths as represent by Equation 3.16:

$$0_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \text{ for } I = 1, 2$$

Outputs of this layer are called “normalized firing strengths”.

**Layer 4:** Every node $I$ in this layer is an adaptive node with a node function in Equation 3.17:

$$0_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$$

where $\bar{w}_i$ is a normalized firing strength from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node. Parameters in this layer are referred to as “consequent parameters”.

**Layer 5:** The single node in this layer is a fixed node labeled $\Sigma$ that computes the overall
output as the summation of all incoming signals represent in Equation 3.18:

\[
\text{Overall output} = 0_{s_j} = \sum w_i f_i = \frac{\sum iw_i f_i}{\sum iw_i} \quad (3.18)
\]

Thus an adaptive network, which is functionally equivalent to the Takagi-Sugeno type fuzzy inference system, has been constructed.

3.4.4.3 ANFIS Implementation in Classifying Heart Disease

The classification was performed using the ANFIS classifiers in Fuzzy Logic Toolbox. ANFIS were trained with the backpropagation gradient descent method in combination with the least squares method. The block of featured processed in ANFIS were shown in Figure 3.13.

![Block Diagram of Heart Disease Classification through Adaptive Neural Fuzzy Inference System (ANFIS)](image)

**Figure 3.13** Block Diagram of Heart Disease Classification through Adaptive Neural Fuzzy Inference System (ANFIS)

65
Based on the Figure 3.13, the featured vector that are being computing from the DWT coefficient which are Energy, Maximum, Minimum, Mean and Standard Deviation were defined as extracted features for ANFIS inputs and Normal, Bradycardia Arrhythmia, Tachycardia and Ischemia are defined as ANFIS outputs.

3.4.4.3.1 ECG Signals Dataset

The datasets with target outputs Class 1 (Normal), Class 2 (Bradycardia Arrhythmia), Class 3 (Tachycardia Arrhythmia) and Class 4 (Ischemia) was given the target values of 1, 2, 3 and 4 respectively as shown in Figure 3.14 and Table 3.6.

**Table 3.6** Set of ECG Signals Class

<table>
<thead>
<tr>
<th>Heart disease type</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1</td>
</tr>
<tr>
<td>Bradycardia Arrhythmia</td>
<td>2</td>
</tr>
<tr>
<td>Tachycardia Arrhythmia</td>
<td>3</td>
</tr>
<tr>
<td>Ischemia</td>
<td>4</td>
</tr>
</tbody>
</table>

![Figure 3.14 Desired Output Target for each Class](image-url)
The statistic properties of the raw data are presented in Table 3.7. There are five features that represent the ECG signals were used by ANFIS classifiers to predict the Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>2.1403</td>
<td>0.0179</td>
<td>6.9619</td>
<td>0.0146</td>
</tr>
<tr>
<td>Maximum</td>
<td>2.2859</td>
<td>0.5251</td>
<td>2.9569</td>
<td>0.6690</td>
</tr>
<tr>
<td>Minimum</td>
<td>-2.8638</td>
<td>-1.1435</td>
<td>-2.3872</td>
<td>-0.9561</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.0094</td>
<td>-0.00457</td>
<td>0.0611</td>
<td>-0.0118</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.5737</td>
<td>0.1477</td>
<td>0.8639</td>
<td>0.2376</td>
</tr>
</tbody>
</table>

The number of data samples that represent each feature for Class 1, Class 2, Class 3 and Class 4 are 89 subjects, 87 subjects, 89 subjects and 87 subjects, respectively.

3.4.4.3.2 Fuzzy Inference System

ANFIS required a predefined network structure and its membership function as well as other parameters can be trained during the learning process. The system is first designed using Sugeno Fuzzy Inference System (FIS). There are two types of FIS namely Grid Partition and Subtractive Clustering.

The ANFIS Grid Partition was adopted in this study because this system required the number of membership functions for each input. This system uses the gbell shaped
membership function to characterize the fuzzy sets input and Sugeno output membership functions are linear types. In the Layer 1, there are five nodes have been used for each input dimension $X_i$ where $i = 1, 2, \ldots, d$ and $d$ is the number of input dimensions.

The ANFIS which constructs a FIS, whose membership function parameters are tuned using a backpropagation algorithm in combination with a least squares type of method, will allows fuzzy systems to learn from the data that they are modeling. The FIS of heart disease classification is shown in Figure 3.15.

![Figure 3.15 Fuzzy Inference System for Heart Disease Classification](image)

Based on five input-one output systems, the five variables were used which are Energy, Maximum, Minimum, Mean and Standard Deviation of DWT coefficients and the output class either Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia or Ischemia is taken as the output variable. The input parameters are represented by fuzzy set or
linguistic variables. The membership functions for input variables are shown in Figure 3.16 (a-e).

(a) Energy coefficients  (b) Maximum coefficients

(c) Minimum coefficients  (d) Mean coefficients

(e) Standard deviation coefficients

**Figure 3.16** Initial Membership Functions for Each Input Dimensions
Based on Figure 3.16, the membership function of each input parameter was divided into three regions, which are, small, medium and large. The examination of initial and final membership functions indicates that there are considerable changes in the final membership functions of the features.

3.4.4.3.3 Rule Base Identification

Based on the membership functions, then the fuzzy IF-THEN rules that have a fuzzy antecedent and constant consequence are constructed. The rule base was created according the expert knowledge using MATLAB rule base editor. Based on the three membership function (small, medium, large) that being used in this project, the number of rule base created by Equation 3.19:

\[ a \land b = c \]  \hspace{1cm} (3.19)

where; \( a \) is membership function
\( b \) is number of input nodes
\( c \) is number of rules output

Therefore, for 3 membership functions and 5 input nodes,

\[ a = 3 \text{ membership function, for small, medium and large} \]
\[ b = 5 \text{ input nodes, for energy, maximum, minimum, mean, standard deviation} \]

\[ a \land b = c \]

\[ 3 \land 5 = 243 \text{ rules are generated} \]
As discussed, some of the rules generated are as in Table 3.8.

**Table 3.8** Created Rule Base by Expert Knowledge. S=small, M=medium, L=large

<table>
<thead>
<tr>
<th>Rule</th>
<th>Antecedent</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feature 1 (Energy)</td>
<td>Feature 2 (Maximum)</td>
</tr>
<tr>
<td>1</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>L</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>6</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>7</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>8</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>9</td>
<td>L</td>
<td>M</td>
</tr>
<tr>
<td>10</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>11</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>S</td>
</tr>
</tbody>
</table>
There are totally 243 rules nodes are generated by the FIS structure. The 243-rule ANFIS structure is shown in Figure 3.17.

Figure 3.17 243 Rule-base ANFIS Structure

There are 5 input nodes for ANFIS structure with 3 inputs of membership functions that processed by 243 rules to identify the desired output of heart disease either in Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia or Ischemia class.
The membership functions in ANFIS fuzzy rule base can be modify manually for each input dimension to improve the FIS output system. The graphical interfaces of fuzzy rules created in rule based layer are shown in the Figure 3.18.

![Figure 3.18 A Fuzzy Rule Base for a Trained 243 ANFIS Rule Base](image)

The figure above shows the fuzzy rule base and the value of membership functions that represent the rules regions either in small, medium or large regions. It can be change as our desired output which suitable for heart disease characteristic of each class.
By viewing the fuzzy rule base tool, it shown that some features dominates the classification result which is some features have more weights in determining which class a data sample belongs to.

### 3.4.4.4 Training and Testing Datasets

The training dataset was used to train the ANFIS model, whereas the testing dataset was used to verify the accuracy and the effectiveness of the trained ANFIS model for classification of the four classes of ECG signals.

In this study, training and testing sets were formed by 352 data samples. The 180 data samples were used for training and 172 data samples were used for testing. The class distribution of the samples in the training and validation data set is summarized in Table 3.9. In order to improve the generalization capability of the ANFIS, training and testing were performed by data obtained from different samples.

<table>
<thead>
<tr>
<th>Heart disease class</th>
<th>Training set</th>
<th>Testing set</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>45</td>
<td>44</td>
<td>89</td>
</tr>
<tr>
<td>Bradycardia</td>
<td>45</td>
<td>42</td>
<td>87</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tachycardia</td>
<td>45</td>
<td>44</td>
<td>89</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemia</td>
<td>45</td>
<td>42</td>
<td>87</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>172</td>
<td>352</td>
</tr>
</tbody>
</table>

**Table 3.9 Class Distribution of the Sample in the Training and Testing Datasets**
All of the features used in the training data have different levels of relevancy. So, the final (after training) membership functions with respect to the initial (before training) membership functions of the input parameters were examined. After training, 172 testing data were used to validate the accuracy of ANFIS model for the detection of heart disease.

### 3.5 ECG Analysis Performances

The ANFIS classifier was trained so that they are likely to be more accurate for each class of ECG signals. The correct classification rates of the proposed ANFIS model were examined and the performance of the ANFIS model was reported.

The classification performance of the ANFIS model was measured to evaluate the performance of the classifiers. It was calculated based on the statistical parameters such as sensitivity, specificity and accuracy. The sensitivity, specificity and accuracy are defined as Equation 3.20, Equation 3.21 and Equation 3.22.

(i) *Sensitivity* is a test to know how good the test is at detecting disease.

\[
\text{Sensitivity} = \frac{\text{number of True Positives}}{\text{number of True Positives} + \text{number of False Negatives}} \quad (3.20)
\]

\[
= \frac{TP}{TP + FN}
\]
(ii) *Specificity* is a test to know how good the test is at identifying normal.

\[
\text{Specificity} = \frac{\text{number of True Negatives}}{\text{number of True Negatives} + \text{number of False Positives}} \tag{3.21}
\]

\[
= \frac{TN}{TN + FP}
\]

(iii) *Total Classification Accuracy* is a number of correct decisions cases/total numbers of cases. \tag{3.22}

Where;

True positive (TP): Sick people correctly diagnosed as sick

False positive (FP): Healthy people wrongly identified as sick

True negative (TN): Healthy people correctly identified as healthy

False negative (FN): Sick people wrongly identified as healthy
CHAPTER 4
RESULT AND DISCUSSION

4.0 Overview

This chapter contains the results and discussion of the Discrete Wavelet Transform (DWT) and Adaptive Neuro Fuzzy Inference System (ANFIS) model from the work that have been carried out in this thesis are presented and discussed. This chapter begins with an introduction to the analysis that has been investigated. Next, this chapter covers the results of DWT for features extraction which will be parametric studies to be analyzed. Some conclusions concerning the rational of features on classification ECG signals that were obtained through the ANFIS. The performance of the ANFIS model was evaluated in terms of training performance and classification accuracies in classifying the Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia signal. The results confirmed that the proposed ANFIS model has potential in classifying the ECG signals.

4.1 Introduction

The ECG analysis system which is one of the expert system that can help human experts in medicine field to monitor and detect variations or actions outside the norm, can be implemented in designing the fast and accurate system. Based on the Webopedia (2004), the Artificial Intelligent which is one of the fields of study in the computer science
includes the expert systems that use the programming computers to make decisions in real-life situations, for example, in helping doctors to diagnose diseases based on disease symptoms.

4.2 Pre Processing

The Figure 4.1 displays below represent the preprocessing process for each sample of dataset which are Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia. It is consist of original signal before processing and the detail wavelet coefficients of Discrete Wavelet Transform.

4.2.1 Normal Signal

![Original Normal ECG signal](image)

**Figure 4.1** Normal ECG signal
Figure 4.2 Normalized Normal ECG signal for the first sample

Figure 4.1 and 4.2 show the original data of Normal ECG signal, the design for wavelet analysis were formed by first 1200 packets data, the collected data are shown in Figure 4.2. This provides a meaningful data for processing in defining detail wavelet coefficients of Normalized Normal ECG signals.

Figure 4.3 Frequency component of Normal ECG signal
Figure 4.3 show the frequency component for first 1200 data processing of the Normal ECG signal.

4.2.2 Bradycardia Signal

Figure 4.4 Bradycardia Arrhythmia ECG signal

Figure 4.5 Normalized Bradycardia Arrhythmia ECG signal for the first sample
Figure 4.4 and 4.5 show the original data of Bradycardia Arrhythmia ECG signal, the design for wavelet analysis were formed by first 1200 packets data, the collected data are shown in Figure 4.5. This provides a meaningful data for processing in defining detail wavelet coefficients of Normalized Bradycardia Arrhythmia ECG signals.

![Original Signal](image)

**Figure 4.6** Frequency component of Bradycardia Arrhythmia ECG signal

Figure 4.6 show the frequency component for first 1200 data processing of the Bradycardia Arrhythmia ECG signal.
4.2.3 Tachycardia Arrhythmia Signal

**Figure 4.7** Tachycardia Arrhythmia ECG signal

**Figure 4.8** Normalized Tachycardia Arrhythmia ECG signal for the first sample

Figure 4.7 and 4.8 show the original data of Tachycardia Arrhythmia ECG signal, the design for wavelet analysis were formed by first 1200 packets data, the collected data are
shown in Figure 4.8. This provides a meaningful data for processing in defining detail wavelet coefficients of Tachycardia Arrhythmia ECG signals.

![Figure 4.9](image)

**Figure 4.9** Frequency component of Tachycardia Arrhythmia ECG signal

Figure 4.9 show the frequency component for first 1200 data processing of the Tachycardia Arrhythmia ECG signal.
4.2.4 Ischemia Signal

Figure 4.10 Ischemia ECG signal

Figure 4.11 Normalized Ischemia ECG signal for the first sample

Figure 4.10 and 4.11 show the original data of Ischemia ECG signal, the design for wavelet analysis were formed by first 1200 packets data, the collected data are shown in
Figure 4.11. This provides a meaningful data for processing in defining detail wavelet coefficients of Ischemia ECG signals.

Figure 4.12 Frequency component of Ischemia ECG signal

Figure 4.12 show the frequency component for first 1200 data processing of the Ischemia ECG signal.
4.3 Features Extraction

The figures below show the overall process of DWT decomposition of each signal.

4.3.1 Normal Signal

**Figure 4.13** The Vector, C and Length, L of Normal signal

Figure 4.13 show the vector coefficient and the length of the Normal signal for 1200 data processing.

**Figure 4.14** The Approximation, A and Detail, D Coefficients of Level 1 Decomposition
Figure 4.14 show the approximation and detail coefficient for the level 1 DWT decomposition of Normal signal. The filtering process of detail coefficient of level 1 is continued until the desired level up to 5 levels is reached.

Figure 4.15 The Detail coefficients of D₁ to D₅ and Approximation coefficient of A₅

Figure 4.15 show the original Normal ECG signals and detail coefficients of each level and also approximation coefficient of A₅. These wavelet coefficients were used as ANFIS inputs.
4.3.2 Bradycardia Arrhythmia Signal

Figure 4.16 The vector, C and length, L of Bradycardia Arrhythmia signal

Figure 4.16 show the vector coefficient and the length of the Bradycardia Arrhythmia signal for 1200 data processing.

Figure 4.17 The Approximation, A and Detail, D Coefficients of Level 1 Decomposition
Figure 4.17 show the approximation and detail coefficient for the level 1 DWT decomposition of Bradycardia Arrhythmia signal. The filtering process of detail coefficient of level 1 is continued until the desired level up to 5 levels is reached.

**Figure 4.18** The Detail coefficients of $D_1$ to $D_5$ and Approximation coefficient of $A_5$

Figure 4.18 show the original Bradycardia Arrhythmia ECG signals and detail coefficients of each level and also approximation coefficient of $A_5$. These wavelet coefficients were used as ANFIS inputs.
4.3.3 Tachycardia Arrhythmia Signal

Figure 4.19 The vector, C and length, L of Tachycardia Arrhythmia signal

Figure 4.19 show the vector coefficient and the length of the Tachycardia Arrhythmia signal for 1200 data processing.

Figure 4.20 The Approximation, A and Detail, D Coefficients of Level 1 Decomposition
Figure 4.20 show the approximation and detail coefficient for the level 1 DWT decomposition of Tachycardia Arrhythmia signal. The filtering process of detail coefficient of level 1 is continued until the desired level up to 5 levels is reached.

**Figure 4.21** The Detail coefficients of $D_1$ to $D_5$ and Approximation coefficient of $A_5$

Figure 4.21 show the original Tachycardia Arrhythmia ECG signals and detail coefficients of each level and also approximation coefficient of $A_5$. These wavelet coefficients were used as ANFIS inputs.
4.3.4 Ischemia Signal

Figure 4.22 The vector, C and length, L of Ischemia signal

Figure 4.22 show the vector coefficient and the length of the Ischemia signal for 1200 data processing.

Figure 4.23 The Approximation, A and Detail, D Coefficients of Level 1 Decomposition
Figure 4.23 show the approximation and detail coefficient for the level 1 DWT decomposition of Ischemia signal. The filtering process of detail coefficient of level 1 is continued until the desired level up to 5 levels is reached.

![Wavelet Coefficients](image)

**Figure 4.24** The Detail coefficients of $D_1$ to $D_5$ and Approximation coefficient of $A_5$

Figure 4.24 show the original Ischemia ECG signals and detail coefficients of each level and also approximation coefficient of $A_5$. These wavelet coefficients were used as ANFIS inputs.
4.3.5 DWT Featured Vector

From the Table 4.1, the 20 extracted features vectors of four classes of ECG signal which were calculated from the $D_4$ frequency band was shows the different from each other, therefore, it is useful parameters in classifying the ECG signals.

**Table 4.1** The Extracted Features of Five Exemplary Records from Four Classes

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Extracted features</th>
<th>DWT Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$D_4$ subband</td>
</tr>
<tr>
<td>Normal beat</td>
<td>Energy</td>
<td>2.1403</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>2.2859</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-2.8638</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.0094</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.5737</td>
</tr>
<tr>
<td>Bradycardia Arrhythmia</td>
<td>Energy</td>
<td>0.0179</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.5251</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-1.1435</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.00457</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.1477</td>
</tr>
<tr>
<td>Tachycardia Arrhythmia</td>
<td>Energy</td>
<td>6.9619</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>2.9569</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-2.3872</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.0611</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.8639</td>
</tr>
<tr>
<td>Ischemia</td>
<td>Energy</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>0.6690</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>-0.9561</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>-0.0118</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.2376</td>
</tr>
</tbody>
</table>

The DWT features coefficient extracted form each ECG signal showing the different value for 352 samples for the training and testing datasets. All of the features coefficients were used by ANFIS as classifier input.
4.4 Classification using ANFIS

The classification of the ECG signals using the combination of DWT features coefficients and ANFIS that was trained with the backpropagation gradient descent method in combination with the least squares method has been made. The hybrid algorithm is composed of a forward pass and a backward pass. The least squares method (forward pass) is used to optimize the consequent parameters with the premise parameters fixed. Once the optimal consequent parameters are found, the backward pass starts immediately. The gradient descent method (backward pass) is used to adjust optimally the premise parameters corresponding to the fuzzy sets in the input domain. The output of the ANFIS is calculated by employing the consequent parameters found in the forward pass. The output error is used to adapt the premise parameters by means of a standard backpropagation algorithm.

The present research demonstrated that the wavelet coefficients are the features which well represent the ECG signals and the ANFIS trained on these features achieved high classification accuracies. In classification, the aim is to assign the input patterns to one of four classes, usually represented by outputs restricted to lie in the range from 1 to 4, so that they represent the probability of the class membership. While the classification is carried out, the specific pattern is assigned to specific class according to the characteristic features that represent the ECG signal.
In this study, training and test sets were formed by 352 data samples. The 180 data samples were used for training and 172 data samples were used for testing. The training dataset was used to train the ANFIS, whereas the testing dataset was used to verify the accuracy and the effectiveness of the trained ANFIS model for the detection of heart disease patients. The steps of parameter adaptation of the ANFIS are shown in Figure 4.25.

![ANFIS Network Error Convergence](image)

**Figure 4.25** ANFIS Network Error Convergence

Based on the Figure 4.25, the ANFIS used 180 training data in 100 training periods and the step size for parameter adaptation had an initial value of 0.011. At the end of 100 training periods, the network error convergence curve of ANFIS had the final error convergence value which is 0.0083787.
Figure 4.26 below shows the plot for 243-rule base ANFIS training performance. This system shows it is sufficient to get zero training error since there are 243 rules trying to classify 180 data samples.

The red points represent in the figure is computed output from the ANFIS structure while the blue points are desired target for 180 training data. The figure also shows the small errors occurs for Class 2 whose the computed output not lies precisely in the Class 2, whereas the other computed output strictly tied to their class desired output. But, the errors spotted around Class 2 data samples still not across the decision boundary of 2.5, so it is still classify as Class 2.
Then, after training, 172 testing data was used to validate the accuracy of the ANFIS classifier for the detection of ECG signals. The Figure 4.27 shows the result of ANFIS model testing performance.

![ANFIS Model Testing Performance](image)

**Figure 4.27** ANFIS Model Testing Performance

The testing data distribution in Figure 4.27 shows that there are small errors occurs in classifying the heart disease since the average testing error is 0.31715. One error is spotted around the 20\(^{th}\) sample whose value is above 1.5, that is across the decision boundary and misclassified as Class 3. Two errors occur from Class 2 that spotted around the 60\(^{th}\) and 80\(^{th}\) samples of the class which are misclassified as Class 4. One error occurs from Class 4 since the data sample was classified as Class 2.
The classification results of the ANFIS model for 172 testing data were displayed in Table 4.2, Figure 4.28 and Figure 4.29.

Table 4.2 Statistic of Correct and Incorrect Heart Disease Classification

<table>
<thead>
<tr>
<th>Heart disease type</th>
<th>Class</th>
<th>Correct Classified</th>
<th>Misclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1</td>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>Bradycardia Arrhythmia</td>
<td>2</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>Tachycardia Arrhythmia</td>
<td>3</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Ischemia</td>
<td>4</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>168</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

Figure 4.28 Statistic of Heart Disease Classification
Table 4.2 and Figure 4.28 show the correct classified and misclassified data samples of heart disease for each class. 43 samples from Class 1 were classified correctly and 1 data sample is incorrect classified. There are 40 samples out of 42 data samples of Class 2 are classified correctly and 41 samples from 42 data samples from Class 4 were correctly classified. For Class 3, all of their 44 data samples were classified correctly. The ANFIS misclassified 4 samples out of 172 data samples.

The statistic of heart disease classification according to class also shows in percentage as displayed in Figure 4.29.

Figure 4.29 Statistic of Heart Disease Classification according to Class

Figure 4.29 shows the percentage for each data that classified correctly and percentage of misclassified data from the testing data. 25% from the testing data is classified as Normal signals, 23% as Bradycardia Arrhythmia signals, 26% of the data was classified as Tachycardia Arrhythmia signals and the data classified as Ischemia signals is 24%
from the testing data. There is only 2% from the testing data is misclassified by the ANFIS system.

The confusion matrix in Table 4.3 below showing the classification results of the ANFIS model used for classification of the ECG signals. This matrix can tell the frequency with which an ECG signals is misclassified as another. The confusion matrix is defined by desired classification on the rows and actual network outputs on the columns.

Table 4.3 ANFIS Training Performance

<table>
<thead>
<tr>
<th>Desired output</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class 2</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Class 3</td>
<td>1</td>
<td>0</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Class 4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>41</td>
</tr>
</tbody>
</table>

According to the confusion matrix, 1 Normal signal from Class 1 was classified incorrectly by the ANFIS model as Tachycardia Arrhythmia signal from Class 3. 2 Bradycardia Arrhythmia signals from Class 2 were classified as Ischemia signals from Class 4 and 1 Ischemia signal from Class 4 was incorrectly classified as Bradycardia Arrhythmia signal from Class 2. All of Tachycardia Arrhythmia signals from Class 3 were classified correctly.
4.5 Performance Analysis

The classification performance of the proposed ANFIS model was determined by the computation of statistical parameters such as sensitivity, specificity and accuracy as follows.

Table 4.4 ANFIS Classification Performance

<table>
<thead>
<tr>
<th>ECG Datasets</th>
<th>Statistical parameters</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
<td>Total classification accuracy (%)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>97.73</td>
<td>100</td>
<td>97.68</td>
<td></td>
</tr>
<tr>
<td>Bradycardia Arrhythmia</td>
<td>95.24</td>
<td>99.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tachycardia Arrhythmia</td>
<td>100</td>
<td>99.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemia</td>
<td>97.62</td>
<td>98.46</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4 shows the classification accuracy determined by ANFIS model. The total classification accuracy determined by ANFIS model was 97.68%. The classification specificity value of Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia signals proposed by ANFIS system are 100%, 99.23%, 99.22% and 98.46%, respectively. As seen from the Table 4.6, the ANFIS classified Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia signals with the classification sensitivity of 97.73%, 95.24%, 100% and 97.62%, respectively.
From the study, the ANFIS algorithm showed significant results of the accuracy of the classification which are above 90%. The accuracy rates presented are highly encouraging and suggest that adaptive neuro fuzzy approach is feasible in heart disease detection. That means, the ANFIS classification system is an excellent system for predicting and classifying. It is able to classify the extracted data from the DWT coefficient of patients ECG signals efficiently and the ANFIS model can improve the classification quality for any ECG signal analysis application. Thus, it can help in improving the life of heart disease patient.

4.5.1 Calculation of Sensitivity and Specificity

4.5.1.1 Normal Signal

(i) Sensitivity = Number of correct classified Normal
Number of total Normal beats

= \frac{TP}{TP + FN}

= \left(\frac{43}{44}\right) \times 100\%

= 97.73 \%

(ii) Specificity = Number of correct classified heart disease
Number of total heart disease

= \frac{TN}{TN + FP}

= \left(\frac{40 + 2 + 44 + 1 + 41}{40 + 2 + 44 + 1 + 41}\right) \times 100\%

= 100 \%
4.5.1.2 Bradycardia Arrhythmia Signal

(i) Sensitivity  = \frac{\text{Number of correct classified Bradycardia Arrhythmia}}{\text{Number of total Bradycardia Arrhythmia beats}}

= \frac{TP}{TP + FN}

= \frac{40}{42} \times 100\%

= 95.24 \%

(ii) Specificity  = \frac{\text{Number of correct classified heart disease}}{\text{Number of total heart disease}}

= \frac{TN}{TN + FP}

= \frac{43 + 1 + 44 + 41}{43 + 1 + 44 + 41 + 1} \times 100\%

= 99.23 \%

4.5.1.3 Tachycardia Arrhythmia Signal

(i) Sensitivity  = \frac{\text{Number of correct classified Tachycardia Arrhythmia}}{\text{Number of total Tachycardia Arrhythmia beats}}

= \frac{TP}{TP + FN}

= \frac{44}{44} \times 100\%

= 100 \%
(ii) Specificity \[= \frac{\text{Number of correct classified heart disease}}{\text{Number of total heart disease}}\]
\[= \frac{TN}{TN + FP}\]
\[= \left( \frac{43 + 40 + 2 + 1 + 41}{43 + 40 + 2 + 1 + 41 + 1} \right) \times 100\%\]
\[= 99.22\%\]

4.5.1.4 Ischemia Signal

(i) Sensitivity \[= \frac{\text{Number of correct classified Ischemia}}{\text{Number of total Ischemia beats}}\]
\[= \frac{TP}{TP + FN}\]
\[= \left( \frac{41}{42} \right) \times 100\%\]
\[= 97.62\%\]

(ii) Specificity \[= \frac{\text{Number of correct classified heart disease}}{\text{Number of total heart disease}}\]
\[= \frac{TN}{TN + FP}\]
\[= \left( \frac{43 + 1 + 40 + 44}{43 + 1 + 40 + 44 + 2} \right) \times 100\%\]
\[= 98.46\%\]
4.5.2 Calculation of Total Classification Accuracy

Total classification accuracy \( = \frac{\text{number of correct decisions cases}}{\text{Total numbers of cases}} \)

\[
= \left( \frac{168}{172} \right) \times 100\%
\]

\( = 97.68\% \)

4.6 Summary

The primary interest of this research, which is the classification of Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia heart disease using DWT and ANFIS as a Neuro Fuzzy classifier have been successfully investigated, and the results have been discussed in details in this chapter. Several parameters have been investigated in the study, which are the values of energy, maximum, minimum, mean and standard deviation of level 4 DWT detail coefficients. The results indicate that by using DWT and ANFIS, the classification of Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia signals can be classified; therefore the primary objective of this study is achieved. The simulation results show that the class of heart disease is well predicted using DWT and ANFIS system and the system working well since it achieve the 97.68\% of classification accuracy rate. This result indicates that it has some potential and had been found to be successful in heart disease detection.
CHAPTER 5
CONCLUSION AND RECOMMENDATION

5.0 Overview

This chapter discussed overall process and outcomes of Electrocardiogram Analysis using Discrete Wavelet Transform and Adaptive Neuro Fuzzy Inference System including the facing problems and project limitation during the projects.

5.1 Conclusion and Outcomes of Research

This thesis is an endeavor to suggest a solution utilizing the hybrids algorithms and to determine an optimum ECG classification scheme designed for the medical environment, where technological advancements have seen changes to many aspects of the daily lives, but there is still a significant gap between the existing solutions and the needs in the medical field. This system provides an analysis system that capable to identify the certain heart disease.

This analysis system is composed of three major components. Based on the preprocessing stage, it is responsible for gathering the database for patient from MIT-BIH Arrhythmia database, Intracardiac Atrial Fibrillation Database and also referred to Hospital Universiti Kebangsaan Malaysia (HUKM) cardiology test unit. This stage have
been done by divided each element of the heart disease phase into Normal signal, 
Tachycardia Arrhythmia signal, Bradycardia Arrhythmia signal and Ischemia signal. The 
signals are successfully can be evaluated and processed. The data gathered from the 
selected databases are connected to MATLAB software where the data are processed.

The second part of the analysis system is based on wavelet analysis theories. This takes 
Discrete Wavelet Transform as a medium to process the patient data. Feature extraction 
techniques are applied to the patient data and the characteristic points of interests are 
being extracted which are Energy, Maximum, Minimum, Mean and Standard Deviation 
values. These data provide meaningful information for the diagnosis of possible heart 
diseases. The data are successfully can be extracted from 172 subjects of patients signal 
in classifying the Normal signal, Tachycardia Arrhythmia signal, Bradycardia 
Arrhythmia signal and Ischemia signal.

Finally, the last part of this system involve the Neuro Fuzzy classifier which is Adaptive 
Neuro Fuzzy Inference System (ANFIS) in classifying the heart disease where the 
decision of heart disease is made based on the extracted features processed by Discrete 
Wavelet Transform. ANFIS plays an important role in dealing with uncertainty when 
making decisions in medical application. The ability to learn how to determine results 
from the sample data is its biggest asset. In ANFIS, the membership function parameters 
are extracted from dataset that describes the behavior of the ECG signals. ANFIS was 
used to detect ECG changes while the wavelet coefficients are defined as its inputs. The 
ANFIS presented in this study was trained with the backpropagation gradient descent 
method in combination with the least squares method. ANFIS classifier is able to get
total classification accuracy rate up to 97.68%, sensitivity rate are 97.73%, 95.24%, 100% and 97.62%, for Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia class. Then, the specificity rate that achieved in classifying the Normal, Bradycardia Arrhythmia, Tachycardia Arrhythmia and Ischemia class are 100%, 99.23%, 99.22% and 98.46%, respectively. The ANFIS model presented in this study prove that it achieved the higher rates of classification accuracy.

5.2 Project Limitation and Problems

5.2.1 There are many types of heart disease which their ECG signals vary closely in amplitude and time duration and represent the expected disease. So the signals must be understood and recognize clearly to make sure the signals are not misclassified.

5.2.2 The increased of input nodes used in ANFIS model will cause increasing of the number of rules, so that it will affect to increase the time to run the sample data used in the training process of the ANFIS system.

5.2.3 The increasing of input nodes also will cause the networks to learn more complex functions and relatively increase the number of training epochs to complete the learning process until the root mean square error close into zero error rates.
5.3 **Recommendations for Future Work**

5.3.1 Due to large number of patients in intensive care units and the need for continuous observation, the current technology development can help to develop the automated ECG monitoring system that allows the system for continuous heart signal monitoring capabilities. By automating the ECG monitoring process, the most updated information for all patients are made available at all times and avoided the delays treatments. It is also can intended to give support to the current health care environments.

5.3.2 The characteristics of the wave features for the ECG analysis can be extended to the other form by using a better or other hybrid algorithms to evaluate the selected features which suitable for many types of heart disease detection.

5.3.3 The quality of accuracy, sensitivity and specificity of ECG analysis can be improved by adding more input databases in the training samples, so that the system are able to learn more and train the system to identify the signal accurately.

5.3.4 The diagnostic accuracy of ANFIS model which combined the neural network adaptive capabilities and the fuzzy logic qualitative approach also can be improve by combining several ANFIS classifier in input data training stage.

5.3.5 The performance of accuracy and training time for classifying the heart disease of ECG analysis systems that widely done in MATLAB software can be improved by embeds the system in the Field Programmable Logic Arithmetic (FPGA). In the code development, more accurate algorithms rates should be used.
REFERENCES

Books


Journals and Conferences


Gao, D., Madden, M., Schukat, M., Chambers, D. and Lyons, G. 2004, ‘Arrhythmia Identification from ECG Signals with a Neural Network Classifier Based on a Bayesian Framework’, Department of Information Technology, National University of Ireland.


**Thesis**


**Website**

Centers for Disease Control and Prevention, 2007, Heart Disease, [Online], [Accessed 22\textsuperscript{nd} March 2008].
Available from World Wide Web:
http://www.cdc.gov/heartdisease/index.htm

Centers for Disease Control and Prevention, 2007, Heart Disease Facts and Statistics, [Online], [Accessed 6\textsuperscript{th} April 2008].
Available from World Wide Web:
http://www.cdc.gov/heartdisease/facts.htm

ScienceDaily, 2007, Research Identifies Major Challenges in Future of Pervasive Healthcare, [Online], [Accessed 6\textsuperscript{th} April 2008].
Available from World Wide Web:

ScienceDaily, 2007, NASA Technology Helps Detect and Treat Heart Disease and Strokes, [Online], [Accessed 19\textsuperscript{th} May 2008].
Available from World Wide Web:

Webopedia : Online Computer Dictionary for Computer and Internet Terms and Definitions, 2004, What is artificial intelligence, [Online], [Accessed 13\textsuperscript{th} March 2009].
Available from World Wide Web:
http://www.webopedia.com/TERM/A/artificial_intelligence.html
APPENDIX A

In this section, some of the input data in the preprocessing and DWT stage are provided for Normal signal, Bradycardia Arrhythmia signal, Tachycardia Arrhythmia signal and Ischemia signal.

1. Normal Signal
2. Bradycardia Arrhythmia Signal

- Bradycardia Arrhythmia ECG signal
- Level 1 Approximation, A1 coefficient for db4
- Level 1 Detail, D1 coefficient for db4
- Vector C
- Signal Length
3. Tachycardia Arrhythmia Signal
4. Ischemia Signal
APPENDIX B

Table shows some of the dataset for the extracted parameter from the DWT coefficient for Energy, Maximum, Minimum, Mean and Standard Deviation.

<table>
<thead>
<tr>
<th>Energy</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Mean</th>
<th>StdDev</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1253</td>
<td>2.2691</td>
<td>-2.8973</td>
<td>-0.0089</td>
<td>0.5578</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.1403</td>
<td>2.2859</td>
<td>-2.8638</td>
<td>-0.0094</td>
<td>0.5737</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.1087</td>
<td>2.3748</td>
<td>-2.8594</td>
<td>-0.0079</td>
<td>0.5409</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.0921</td>
<td>2.2756</td>
<td>-2.9134</td>
<td>-0.0071</td>
<td>0.5525</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.2307</td>
<td>2.2777</td>
<td>-2.9128</td>
<td>-0.0084</td>
<td>0.5588</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.3509</td>
<td>2.3143</td>
<td>-2.8845</td>
<td>-0.0089</td>
<td>0.5688</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.1587</td>
<td>2.2954</td>
<td>-2.8956</td>
<td>-0.0093</td>
<td>0.5736</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.142</td>
<td>2.2858</td>
<td>-2.9135</td>
<td>-0.0083</td>
<td>0.5825</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.0987</td>
<td>2.2896</td>
<td>-2.8859</td>
<td>-0.0091</td>
<td>0.5932</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>2.1186</td>
<td>2.2457</td>
<td>-2.8268</td>
<td>-0.0088</td>
<td>0.5748</td>
<td>&quot;Normal&quot;</td>
</tr>
<tr>
<td>0.0179</td>
<td>0.5251</td>
<td>-1.1435</td>
<td>-0.00457</td>
<td>0.1477</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0131</td>
<td>0.5282</td>
<td>-1.1437</td>
<td>-0.00462</td>
<td>0.1479</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0177</td>
<td>0.5236</td>
<td>-1.1428</td>
<td>-0.00438</td>
<td>0.1456</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0174</td>
<td>0.5275</td>
<td>-1.1432</td>
<td>-0.00421</td>
<td>0.1464</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0168</td>
<td>0.5263</td>
<td>-1.1436</td>
<td>-0.00459</td>
<td>0.1452</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0151</td>
<td>0.5229</td>
<td>-1.1444</td>
<td>-0.00477</td>
<td>0.1481</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0153</td>
<td>0.5237</td>
<td>-1.1424</td>
<td>-0.00458</td>
<td>0.1469</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0191</td>
<td>0.5232</td>
<td>-1.1435</td>
<td>-0.00464</td>
<td>0.1465</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0173</td>
<td>0.5231</td>
<td>-1.1439</td>
<td>-0.00446</td>
<td>0.1476</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0171</td>
<td>0.5255</td>
<td>-1.1451</td>
<td>-0.00451</td>
<td>0.1468</td>
<td>&quot;Bradycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>6.9132</td>
<td>2.9568</td>
<td>-2.3876</td>
<td>0.0635</td>
<td>0.8888</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>6.9459</td>
<td>2.9453</td>
<td>-2.3922</td>
<td>0.0644</td>
<td>0.8234</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>6.9752</td>
<td>2.9559</td>
<td>-2.3902</td>
<td>0.0699</td>
<td>0.8541</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>7.0114</td>
<td>2.9669</td>
<td>-2.3909</td>
<td>0.0676</td>
<td>0.8693</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>6.9383</td>
<td>2.9647</td>
<td>-2.3865</td>
<td>0.0642</td>
<td>0.8647</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>7.0179</td>
<td>2.9573</td>
<td>-2.3848</td>
<td>0.0653</td>
<td>0.8397</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>6.9687</td>
<td>2.9565</td>
<td>-2.3929</td>
<td>0.0642</td>
<td>0.8542</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>7.0421</td>
<td>2.9662</td>
<td>-2.3832</td>
<td>0.0648</td>
<td>0.8639</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>6.9567</td>
<td>2.9549</td>
<td>-2.3865</td>
<td>0.0683</td>
<td>0.8104</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>7.0243</td>
<td>2.9557</td>
<td>-2.3937</td>
<td>0.0676</td>
<td>0.8462</td>
<td>&quot;Tachycardia Arrhythmia&quot;</td>
</tr>
<tr>
<td>0.0146</td>
<td>0.669</td>
<td>-0.9561</td>
<td>-0.0118</td>
<td>0.2376</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0159</td>
<td>0.712</td>
<td>-0.9756</td>
<td>-0.0186</td>
<td>0.2426</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0155</td>
<td>0.686</td>
<td>-0.9646</td>
<td>-0.0163</td>
<td>0.2456</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0151</td>
<td>0.694</td>
<td>-0.9564</td>
<td>-0.0132</td>
<td>0.2647</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0151</td>
<td>0.728</td>
<td>-0.9684</td>
<td>-0.0136</td>
<td>0.2656</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0153</td>
<td>0.677</td>
<td>-0.9324</td>
<td>-0.0197</td>
<td>0.2123</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0139</td>
<td>0.654</td>
<td>-0.9678</td>
<td>-0.0163</td>
<td>0.2788</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0144</td>
<td>0.687</td>
<td>-0.9894</td>
<td>-0.0164</td>
<td>0.2896</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0143</td>
<td>0.643</td>
<td>-0.9947</td>
<td>-0.0124</td>
<td>0.2652</td>
<td>&quot;Ischemia&quot;</td>
</tr>
<tr>
<td>0.0152</td>
<td>0.711</td>
<td>-0.9536</td>
<td>-0.0131</td>
<td>0.2341</td>
<td>&quot;Ischemia&quot;</td>
</tr>
</tbody>
</table>


Chapter(s) in A Book : A Review on Feature Extraction & Classification Techniques for Biosignal Processing (Part I: Electrocardiogram)
Book Title : IFMBE Proceedings Series
Publisher : Springer-Verlag, Berlin Heidelberg
Year : 2008
Page : 107-112
Volume : 21