Doctor Thesis

Shibaura Institute of Technology

QUICK IDENTIFICATION OF ARRHYTHMIA SYMPTOMS USING EMPIRICAL APPROACH IN LONG SEQUENCE OF HEART CYCLES

2019 March

MOHAMAD SABRI BIN SINAL

Declaration of Authorship

I, Mohamad Sabri bin Sinal, declare that this thesis titled, “Quick Identification of Arrhythmia Symptoms using Empirical Approach in Long Sequence of Heart Cycles,” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at Shibaura Institute of Technology.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at Shibaura Institute of Technology or any other institution, this has been clearly stated.
- Where I have consulted the published work of other, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with other, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

(Mohamad Sabri bin Sinal)

Date: _____

“I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.”

-Isaac Newton-

This thesis is dedicated to my parents, Nor Hasimah Zaim and Professor Dr. Halil Zaim.

ABSTRACT

Computational electrocardiogram (ECG) analysis is one of the most crucial topics in cardiovascular research domain especially in identifying the abnormalities of heart conditions through cardiac arrhythmia symptom. Each symptom consists of its own unique characteristic and the complexity to characterize various types of abnormalities is one of the big challenges in this study. Hence, the difficulties in identifying an early stage of heart diseases symptom due to random behaviour and rare appearance in time series has proven a challenge to create the solution towards the detection of the symptom. Previous studies have tried to solve this issue and some of them achieved the discrimination with a high degree of accuracy. However, the accuracies never reach 100%. Hence, the complexity in identifying the disease is huge. In addition, no research has achieved it in a long duration time frame, e.g. 12 hours, of ECG data. Therefore, this dissertation tries to deal with that problem constructively. In this dissertation, an efficient, quick and highly sensitive computational intelligence to accurately detect abnormalities of a heart condition based on Arrhythmia symptom is proposed. The proposed mechanism consists of two primary components, namely an efficient Arrhythmia detection using autocorrelation and statistical approach, and hybrid mechanism to detect Paroxysmal stage of Atrial fibrillation using adaptive threshold-based algorithm with Artificial Neural. There are two main concerns for each study. For the first study, the detection of heart condition abnormalities should be simple and with the capability to detect abnormalities regardless of the symptom's origin. Next, for the second study, the focus is more on the design of a very sensitive mechanism to detect the abnormalities at early stage namely the Paroxysmal stage of the Atrial Fibrillation. How to deal with the complexity of the disease behavior at an early stage and the

visual representation of the outcome to classify the disease are the two of main concerns in this study.

In the first study, an Efficient Arrhythmia Detection Using Autocorrelation and Statistical Approach is proposed. This study proposes an autocorrelation method with K-Nearest Neighbor (KNN) classifier method to accurately and robustly detect 14 types of Arrhythmia symptom regardless of the origin of the symptom in a long hour data. Moreover, variability analysis based on periodic autocorrelation result is proposed and used for the classification procedure. 1 minute and 12 hours duration data are chosen to compare and signify the most suitable time duration to detect the Arrhythmia symptom. As a result of the proposed method performance evaluation, it is revealed that the accuracy of 95.5% in discriminating Arrhythmia from Normal Sinus data is achieved. Furthermore, it is confirmed that by utilizing the autocorrelation result in long hour data can help generalize the abnormalities' characteristic of heart condition like Arrhythmia symptom. It is concluded that the proposed method can be useful to diagnose abnormalities of a heart condition at any stage.

Secondly, the Hybrid Mechanism to Detect Paroxysmal Stage of Atrial Fibrillation using Adaptive Threshold-based Algorithm with Artificial Neural Network is proposed. In this study, a new mechanism called "Door-to-Door" algorithm is introduced to accurately and quickly detect the significant peaks of heart cycle in 12 hours ECG data and to discriminate obvious Preliminary stage of Atrial Fibrillation rhythms from Normal Sinus rhythms. In addition, a quantitative method using Artificial Neural Network (ANN), which discriminates unobvious Paroxysmal stage of Atrial Fibrillation rhythms from Normal Sinus rhythms is investigated. As a result of Door-to-Door algorithm performance evaluation, it is revealed that the

Door-to-Door algorithm achieves the accuracy of 100% in detecting the significant peaks of heart cycle in 17 Normal Sinus ECG data. Furthermore, it is verified that the ANN-based method achieves the accuracy of 100% in discriminating the Paroxysmal stage of 15 Atrial Fibrillation data from 17 Normal Sinus data. Therefore, it is confirmed that the computational time to perform the proposed mechanism is less than the half of the previous study.

As concretely presented in this study, the proposed mechanism not only accurately detects the abnormalities of the heart condition based on Arrhythmia symptom but also reduces the complexity in identifying the symptom with small and simple parameter. Consequently, it is concluded that this research can contribute to the medical field as one of the best technologies in diagnosing abnormalities of heart condition as early stage as Paroxysmal stage of Atrial fibrillation.

ACKNOWLEDGEMENT

Even though only my name appears on the cover of this dissertation, many people have contributed to make this dissertation possible. To them, I am eternally grateful and will always cherish this experience forever. My deepest gratitude is to my advisor, Professor Dr Eiji Kamioka whom allowed me the freedom to explore on my own while guiding me when my steps faltered. This amazing Professor Dr Eiji Kamioka had also taught me how to question thoughts and express ideas. His patience and support helped me overcome many crisis in enabling me to finish this dissertation. I am deeply grateful to him for the long discussions that helped me sort out the technical details of my work. I am also thankful to him for encouraging me to use the correct grammar and consistent notation in my writings and for carefully reading and commenting on countless revisions of my works. Professor Dr Eiji Kamioka insightful comments and constructive criticisms at different stages of my research were thought-provoking and they helped me to focus on my ideas. I am grateful to him for holding me to a high research standard and enforcing strict validations for each research result, and thus teaching me on the writing of a good research. I hope that one day I will become as good an advisor to my students as Professor Dr Eiji Kamioka had and has been to me.

My gratitude is also to Associate Professor Dr Atsushi Saito. He is one of the best teachers in my life. He sets high standards for his students and encourages and guides them to meet those standards. He introduced me to Mathematical logic, sensor and signal processing method and his teachings inspired me to work hard in this dissertation. I am indebted to him for his continuous encouragement and guidance as well as his encouragement and practical advice regarding cardiovascular studies.

I am also thankful to Associate professor Ryota Horie for reading my reports and my articles, commenting on my views and helping me understand while enriching my ideas in bioinformatics engineering area. Furthermore, I would like to acknowledge Professor Dr Koichi Gyoda, Professor Dr Hiroaki Morino, Dr Khaironi, Muhammad Luqman bin Mahamad Zakaria, Teh Boh Tsong, Saiful Zulkifli, Muhammad Aiman, Professor Dr Halil Zaim, Professor Dr Noor Azizi Ismail for numerous discussions and lectures on related topics that helped me improve my knowledge in the those areas. I am also thankful to the graduate school and international department staff who assisted me during my stay in Japan' making my life here less worrisome. I feel that they are the most friendly and nice staff that I have ever met in Japan.

I am moreover indebted to Dr. Tachibana Masahiko, Ms Kageyama, for his assistance and kindness since the day I applied to further study in Shibaura Institute of Technology (SIT) and during my stay in Japan until the end of my Doctorate studies. I still remember how Dr. Tachibana Masahiko helped me settled the apartment arrangement when I first came to Japan. I am nobody without Dr. Tachibana Masahiko's guidance. I would like to greatly honour him for his kindness.

I am grateful to the Graduate School Section who always kindly support and advice the procedures related to my research duty at SIT. I would like to thank Prof Masato Murakami, board of director and the management team for their incredible Human Resource Fostering Program which positively influences the way I do this research and also giving me full scholarships for my doctorate study.

I also wish to thank SIT's Librarian Staff, Vice Chancellor of Univerisiti Utara Malaysia, Deputy Vice Chancellor of Universiti Utara Malaysia and his team, Universiti Utara Malaysia as a whole, Dean of School of Computing and School of

Computing (Universiti Utara Malaysia) for their supports while I pursue my studies in Japan. Thank you to the whole team!

Many friends have helped me stay sane through these difficult years. Their support and care helped me overcame setbacks so that I was able to stay focused on my study. I greatly value their friendship and I deeply appreciate their belief in me. I am also grateful to the Malaysian and Japanese families and friends that helped me adjust to a new country. Most importantly, none of this would have been possible without the love and patience of my family. My immediate family to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heartfelt gratitude to my family and my elder sister in Kuwait, Nor Hasimah binti Zainal for her patience and kindness in guiding and teaching me on how to write a good research paper and thesis. My extended family has aided and encouraged me throughout this endeavour.

I also would like to give a special mention for the support given by Shinryo Kurata. I warmly appreciate the generosity and patience in helping me out to go through difficult time dealing with my nerve disease during my last year as a doctorate student. Moreover, I am really grateful to Teh Boh Tsong, Muhammad Aiman and Mizan for his generosity and assistance during my hard time dealing with my serious illness while staying in Japan. Moreover, a special appreciation to Orthopaedic Doctor, Dr Yukawa, Doctor Pain, Dr Jin at Pain Clinic Department and the rest of the team at Jikei Hospital for giving me a full support to treat my nerve illness and chronic pain illness. I am indebted to you all. Finally, I none of this is possible without the financial support from Shibaura Institute of Technology and Universiti Utara Malaysia which funded my studies in Japan for 3 years. THANK YOU.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ParAF	Paroxysmal stage of Atrial Fibrillation
NS	Normal Sinus
KDD	Knowledge Discovery process in Database

CHAPTER 1

INTRODUCTION

This chapter presents an overview of the research in this thesis. Firstly, it describes the background of the research area and the motivation in detail. Next, the research problems are explained as well as the research objectives. Finally, the research contribution of this thesis is summarized.

1.1 Motivation

Heart disease has been the leading global cause of death around the world for the last 15 years. In 2016, over 15.6 millions out of 56.9 millions deaths were observed as heart disease deaths [1]. It is anticipated that the number will continue to grow. The World Health Organization (WHO) has reported that the biggest killer diseases in the world are Ischemic heart disease and stroke [1] as shown in Figure 1.1. Heart disease, which is also known as the Heart and Blood Vessel disease, is defined as a range of problematic conditions that affects the heart. It includes various types of diseases such as coronary heart disease, strokes, rheumatic heart disease, cardiomyopathy, and other heart diseases.

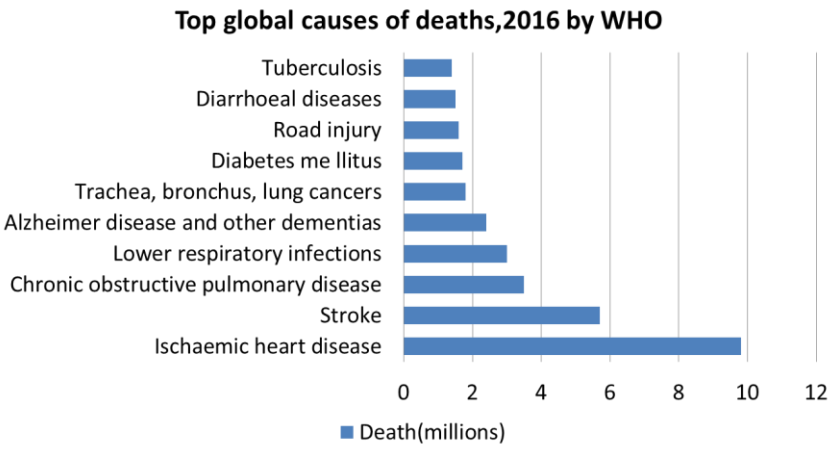

Figure 1.1: Global health estimation 2016

The impact of heart disease to the world population is undeniable. A drastic lifestyle change in the society, the increment of smokers and obesity, the urbanization in the country, the westernization of diet and post-infectious illnesses have all contributed to this issue [1]. Moreover, the lack of government support in providing screening, prevention and treatment service together with the lack of health workforce are still roots to the growth of illnesses around the world, particularly in developing countries [1]. Therefore, a new holistic solution is needed to fix the issue rather than relying solely on the conventional ways of treating diseases at hospitals. It is very important to understand and identify the real constraints of the current problems before designing the right solution. The insight on how the world deals with the heart disease should be taken into consideration. The details will be discussed next.

With the growth of heart disease cases around the world, plenty of treatments with high quality like cardiopulmonary resuscitation (CPR), high-tech surgeries like Angioplasty, stents or even heart bypass surgery are offered in hospitals. In reality, diagnosing or detecting heart disease is a crucial task for cardiologists. Different

diseases consist of different symptoms and each heart disease requires different treatment. Some of diseases can be cured only at the early stage. The urgency of giving effective treatment at a different stage of disease requires the expertise and the availability of cardiologists to accurately analyse the symptoms and properly decide the treatments. However, the small number of cardiologists availability to the large number of patients who need the detection procedures and the treatments cause longer time to completion. As a result, the cost for treatment will increase when the doctors are unable to detect the symptoms as early as they are supposed to be.

With the continuous trend of high medical cost and high demand for the treatment around the world, an alternative solution to improve the condition should be provided. One of the possible ways is to develop a mechanism that can detect symptoms of disease at the early stage, and thus, a better treatment could be given as soon as the disease is identified. As a result, the cost for the treatment could be lower and more affordable to everyone.

Conventionally, the most common way to measure heart condition is to use the index of BPM, which is the count of heart beat per minute. Measuring heartbeats per minute indicates how fast or slow the heart is beating up. Normally, the heart beats 60 to 90 times per minute and it varies from one person to another. The second most common way is based on photoplethysmography (PPG). It uses a light-based technology and it measures blood flow rate controlled by the heart's pumping action. PPG uses the electrical signals derived from reflected light due to changes in the blood flow during heart activities. However, the significant findings provided by both methods are very limited and insufficient to describe the details of a heart

condition. Therefore, electrocardiogram (ECG) examination is widely used in order to measure heart condition.

The ECG data is a record of the heart's electrical activities described as wave-forms in chronological order and it shows the depolarization and repolarization processes in the myocardial cells. In a normal heart cycle, ECG consists of five different characteristic wave-forms and each wave includes a peak, which is a spike or a dip. ECG contains a larger number of cardiovascular specifics. Hence, it plays a significant role in guiding clinical diagnoses of heart diseases and their symptoms. This examination includes two important steps. The first step is the recording of the heart activities on the chest. The second step is a deep cardiac behaviour analysis to the collected data by cardiologists. This procedure takes a day or more depending on the heart condition. In addition, the procedure itself is very complicated, time consuming, and thus, only experts in the medical field are capable of doing such a task.

With the emergence of echocardiography technology, the accessibility to the subclinical results has gone through to the new whole level [1]. The availability of the technology and the portability feature have become the stronger point in utilizing it for further use. Yet, the lack of standard echocardiography criteria to diagnose the result and the practicalities of delivering echocardiographic screening program have become the drawbacks [1]. Nevertheless, as previously mentioned, one of the biggest issues in dealing with this matter is the limited availability of medical officers. Therefore, the utilization of the current technology can help to reduce the burden of limitation in solving the issue.

There are plenty of advanced technologies available for diagnosing heart condition. One of the most suitable technologies for medical diagnosis is machine learning approach. Machine learning has the capability to imitate the human brain and performs specific tasks precisely. This can provide a huge positive impact in analysing hidden pattern of biomedical data. Moreover, the mechanisms itself may provide an insight about the patient's condition and the symptom which are not literally evident. Consequently, better diagnosis will lead to better treatment. Therefore, to strengthen up the solution to deal with the heart disease diagnosis constructively, a mechanism using computational intelligence for heart disease diagnosis (CIHDD) is needed. In lieu of that, this thesis proposes an efficient, quick and highly sensitive mechanism to accurately detect abnormalities of heart condition based on Arrhythmia symptom.

Computational intelligence is a branch of computer science that studies adaptive mechanisms to enable or facilitate intelligent behaviour in complex and changing environments. It utilizes data mining and machine learning approaches to solve specific problems based on a large dataset. By definition, data mining is also known as "knowledge discovery in database" which is a process of discovering hidden or complex pattern in a large dataset. By utilizing the capability of machine learning, statistics and database system, raw data can be transformed into comprehensible structure for further use. Conventionally, data mining approach has several stages such as data selection, pre-processing, transformation, data mining and interpretation. The examples of data mining approach are decision tree learning, K-Nearest neighbours, association rule learning, artificial neural networks, deep learning, inductive logic programming, support vector machines, clustering, Bayesian networks, genetic algorithms, rule-based machine learning and so forth.

Data mining approach is commonly used for specific tasks such as anomaly detection, association rule learning, clustering, classification, regression and summarization [2].

In computational intelligence study for diagnosing heart disease, mainly clustering, classification and anomaly detection of biomedical data are considered. The study focuses on how to develop the most suitable intelligent mechanism to classify disease and non-disease data based on hidden pattern in biomedical data. The accuracy of detection, the complexity of the method and how fast it performs to produce the result are the most important discussion topics in this area. Computational intelligence for heart disease prediction and detection has many benefits.

In general, computational intelligence methods have the capability to discover hidden pattern in a large dataset. It extracts useful information from the dataset using advanced search techniques and algorithms to discover patterns and correlation among the data. In computational heart disease diagnosis context, the capability to understand complex biomedical data may help doctors give new insights about their patients' conditions. There is the potential that doctors overlook some important information with a complex behaviour existing in ECG data. Therefore, intelligent mechanisms are necessary to handle this situation. The existence of computational based diagnosis is not to replace or replicate a doctor's duty but to work as a supporting tool, providing new insights on the symptoms and assisting in decision making. In short, utilising data mining approach helps doctors in providing better perspectives on the patient's heart condition in various points of view. Subsequently, it assists doctors to make better and more accurate decisions in diagnosis.

Moreover, this study tries to improve healthcare quality and patient outcomes. This method also benefits the end user in various ways particularly in providing a better preliminary diagnosis technology enable to diagnose heart conditions without relying too much on the doctor. Nonetheless, there are some challenges in identifying heart disease symptoms computationally using biomedical data to model and detect the heart disease accurately. In the next section, they will be discussed.

1.1.1 Issues

To design and create an early abnormalities detection model and mechanism for heart disease come with a lot of challenges. They are;

1) The decision on which disease symptom for computational modelling of heart disease symptom detection is to be used.

There are various kinds of illnesses under these categories. In general, heart disease or cardiovascular disease includes a series of conditions that affect the structures or functions of the heart such as coronary artery disease, heart attack, abnormal rhythms or also known as arrhythmias, heart failure and vascular disease. Each disease consists of very complex behaviours and it is not easy to characterize the entire symptoms, and then, it is necessary to detect them effectively by using an autonomous mechanism like a computer based approach. To do that, it is important to well understand the heart diseases including their characteristics in order to propose a right mechanism.

Deciding the most suitable heart disease symptom as the target is the most important because it influences how far the proposed mechanism will provide a good impact to the issue. The sensitivity of the mechanism to detect the symptom can lead to the early disease detection for a better treatment.

Stroke, for example, is a disease where the best and effective treatments can be given only at the first stage of the symptom. At this point, stroke disease can be treated permanently. However, the symptom itself rarely appears in time series domain of cardiac electrogram. Generally, including this case, it is difficult to detect diseases at early stages since it is invisible to the normal human's eyes and miss detection by sophisticated medical devices occurs stochastically. The lack of significant evidence or information may result in inaccurate treatments by doctors.

In this study, the focus is the detection of heart condition abnormalities based on arrhythmia symptom. The detection mechanism will have a strong sensitivity in identifying the earlier stage of the disease. In general,a heart condition abnormality also known as arrhythmia is defined as the heart beating too fast, too slow or irregularly. Detecting abnormalities is a crucial point in this study since arrhythmia symptom can lead to more critical heart disease such as stroke or heart attack. Moreover, the preliminary sign of stroke can be identified only at the early stage of Atrial Fibrillation. Atrial Fibrillation is described as a symptom of heart disease in which the heart activity irregularly behaves in random period of time. It consists of three stages, which are Paroxysmal stage, persistent stage and permanent stage. The symptom of Atrial Fibrillation falls under Arrhythmia group.

A recent study has shown that Atrial Fibrillation is often overlooked after interventional therapies when the standard strategy for treatment evaluation is used [3]. Since stroke, which is the second top killer around the world [1], is initiated by Arrhythmia symptom which is the early stage of Atrial Fibrillation, the early detection is vital to prevent the interconnecting diseases.

Therefore, there are two important points that need to be highlighted here. Firstly, this study mainly focuses on detecting abnormalities of heart condition based on Arrhythmia symptom. This is because the abnormalities are the gateway to the other chronic heart disease. Secondly, the detection of preliminary stage of Atrial Fibrillation with enough sensitivity is also focused on in this study. This is because stroke disease can be identified at the early stage of Atrial Fibrillation. In addition, this study proposes the most suitable attribute to accurately detect Arrhythmia symptoms. For that, the effectiveness of developing very high sensitivity to detect abnormalities can be proven.

2) The complexity in designing computational model with biomedical data

In general, biomedical data consists of distinctive features that are different from other types of data in different disciplines. Technically, biomedical data has possibilities of being affected by several resource of uncertainty such as measurement error, missing data or error in coding. In the context of identifying Arrhythmia symptom, for instance, the mixing of signal noise and disease symptoms in biomedical data may increase the complexity in detecting Arrhythmia symptom itself. As a result, it will affect the overall detection performance.

The variability of individual or the unique patient characteristics from one to another is the second biggest challenge with biomedical data. Due to the uniqueness of individual characteristics, the difficulty to set a specific standard to detect a disease by using computational approach is challenging. As a result, additional information is required to aid the computational approach to detect the disease accurately. Additional information such as weight, gender, treatment information, time duration of having the disease are prospect to boost the detection performance. Moreover, detecting the heart disease based on each symptom comes with a set of difficulties.

As mentioned previously, the heart disease consists of various characteristics and each characteristic is unique in its own way. To define those characteristics individually is challenging since there is no specific guideline to follow in computational approach. There are many existing related works focusing on using different kinds of attribute to detect the same disease. However, the performance like detection accuracy fluctuates from one work to another, and the complexity of the mechanism itself increases with the usage of different techniques.

One of the best approaches to develop a computational intelligence mechanism with high accuracy is to have a good training model. A training model here refers to the knowledge that is extracted from the raw data in a large dataset and is used as a reference to classify the disease. The larger the number of dataset is, the better the accuracy can be achieved. However, biomedical data is costly and limited in number. There may be researchers or outsiders who misuse the biomedical data for different purposes if it is publicly shared. Therefore, development of a good and balance disease model for the identification and detection is a big issue.

As the conclusion, a good computational model with minimal attribute but high accuracy to detect Arrhythmia symptom is needed. However, the lack of data availability to support the model as well as the lack of specific guidelines to identify the disease has been the disadvantage for this research area. Still, there are various data mining techniques introduced so far for heart disease detection. For that, this study will try to investigate the possibility behind this theory.

3) The implementation of the model for high accuracy detection rate

Machine learning model is well known as one of the most common tools used to diagnose heart disease symptom. This is due to the advancement of the algorithm and current technology to process numerous patterns of data and then find the precise correlation between each of them. However, finding the right implementation mechanism for the well-designed model is a challenging task. The non-linear of high dimensionality properties in the medical dataset has made it difficult to determine the most suitable implementation mechanism for disease detection. Moreover, there is no specific machine learning considered as the best machine learning for all dataset as it is depended on the problem and characteristic of the data. Therefore, it is important to identify what is the best implementation mechanism for the proposed computational model to detect an early abnormalities symptom in ECG data with high accuracy rate. This is one of the main concerns in this study.

4) The Model Output for disease classification

Most learning technique output machines are different in many ways. It is important to understand what the most suitable output is to represent or visualize the solution based on well-defined problem. Neural network for instance, the output generated by it is hard to interpret because it is a black box approach. Therefore, in this study, several models will be utilized to deal with the complexity of the dataset to detect disease symptom in long time duration.

Conclusion

As conclusion, the effectiveness of designing autonomous heart disease detection comes with a huge computational cost and high complexity. Selecting the right number of attributes which carry significant values in biology perspective while using suitable data mining approach to classify the disease may lead to high accuracy in disease detection and increase the significant value in the computational finding. Therefore, every aspect must be considered in order to create the most balanced formula for computational mechanism of the abnormalities of heart disease detection. It is highlighted in yellow in Table 1.1. The balanced heart disease computational modelling is also shown in Figure 1.2.

Table 1.1: Challenge in designing autonomous heart disease detection mechanism

Effectiveness **(Low)**	Number of symptom cover **(High)**	Complexity in characterizing and extracting the symptom quantitatively **(High)**	Detection accuracy **(High)**
Effectiveness **(Low)**	Number of symptom cover **(Low)**	Complexity in characterizing and extracting the symptom quantitatively **(Low)**	Detection accuracy **(High)**
Effectiveness **(High)**	Number of symptom cover **(High)**	Complexity in characterizing and extracting the symptom quantitatively **(Low)**	Detection accuracy **(High)**

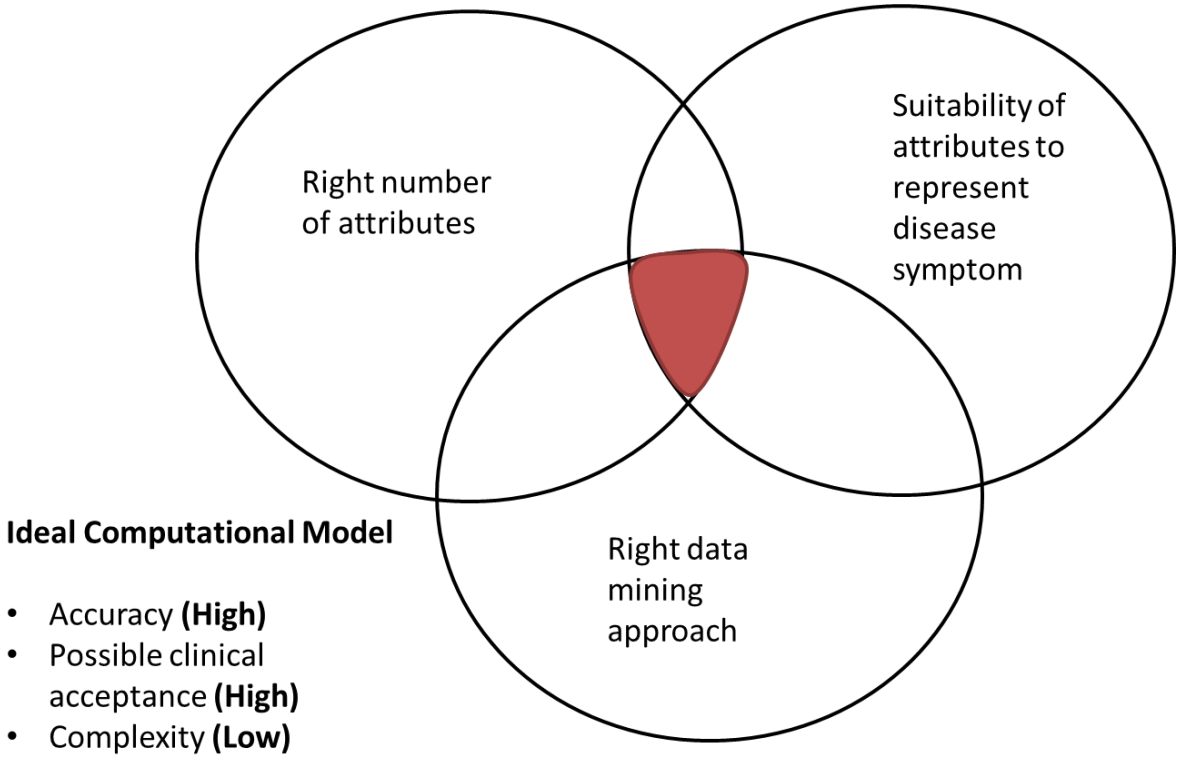

Figure 1.2: Ideal computational modelling for heart disease detection

1.1.2 Research question

Based on the problem statement mentioned above, several issues related to the design of an effective computational model with the right number of attribute must be solved. Moreover, in order to achieve high sensitivity in detecting abnormal heart condition but less complexity during the procedure, the effectiveness of utilization of machine learning output will be proposed. Hence, three research questions are addressed here:

1) Is it possible that the minimal number of attributes used in data mining approach may influence the detection rate in detecting abnormalities of heart condition?

2) How can the utilization of machine learning output assisst in improving the sensitivity and accuracy in detecting the early stage of Atrial Fibrillation?

3) What is the most suitable machine learning approach to deal with different scales of dataset, different numbers of attributes and different time duration of ECG data to detect abnormalities of heart condition?

1.1.3 Research Goals

The key contribution of this thesis is proposing an efficient, quick and highly sensitive mechanism to accurately detect abnormalities of heart condition based on Arrhythmia symptom. To serve the entire research question here, this study is done to provide constructive evidence to the issue. For that, the research goals of this study are described below:

- To construct an effective and simple computational model based on the ECG segment to detect abnormalities of heart condition regardless the origin of the symptom.
- To clarify the effectiveness of the proposed computational model based on minimal number of attributes used to detect abnormal heart disease symptom through statistical and data mining approach.
- To clarify the most effective data mining approach used to deal with the ECG data with abnormalities symptom based on different scale of dataset and difference time duration.
- To clarify the effectiveness of utilizing the neural network output to improve the accuracy of detecting early stage of heart disease symptom.

1.1.4 Database for Evaluation

All the ECG data in this study are recorded at the Arrhythmia Laboratory, Beth Israel Hospital. The recorded data which are then downloaded to the Physionet [4] website acts as an open research resource, providing a large number of recorded physiological data to users worldwide. It has over 90,000 recordings of over 4 terabytes of digitized physiologic signals and time series, organized in over 80 databases. Three different types of databases which are “MIT-BIH Normal Sinus”, “MIT-BIH Atrial Fibrillation” and “MIT-Arrhythmia” are utilized in order to evaluate the performance of the proposed work. All selected databases are recorded under proper condition without putting any stress or pressure towards patients. The detailed descriptions about the database are shown below:

1) MIT-BIH Normal Sinus Database

MIT-BIH Normal Sinus database includes 18 long-term ECG recordings of human subjects. No significant arrhythmias are found. The subjects are 5 men (aged 26 to 45) and 13 women (aged 20 to 50).

2) MIT-BIH Atrial Fibrillation Database

MIT-BIH Atrial Fibrillation database includes 25 long-term ECG recordings of human subjects with atrial fibrillation and mostly Paroxysmal stage of Atrial Fibrillation. It includes two ECG signals with 12-bit resolution over a range of ±10 millivolts. The original analog recordings were made at Boston's Beth Israel using ambulatory ECG recorders with a typical recording bandwidth of approximately 0.1 Hz to 40 Hz. These ECG data contain the rhythm annotation type such as (AFIB

(atrial fibrillation), (AFL (atrial flutter), (J (AV junctional rhythm), and (N (used to indicate all other rhythms).

3) MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database contains 48 and a half-hours excerpts of two-channels ambulatory ECG recordings obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979. Twenty-three recordings are chosen at random from a set of 4000 24-hour ambulatory ECG recordings collected from a mixed population of inpatients (about 60%) and outpatients (about 40%) at Boston's Beth Israel Hospital; the remaining 25 recordings are selected from the same set to include less common but clinically significant arrhythmias that are not well represented in small random samples. The recordings are digitized with 11-bit resolution over a 10mV range. The subjects are 25 men aged 32 to 89 years and 22 women aged 23 to 89 years.

1.1.5 Knowledge discovery with Computational Intelligence for heart disease detection

Time series data analyses are the most conventional method used in diagnosing heart condition. Examining the whole time series of data is needed to accurately review the heart status. However, identifying abnormalities based on time series domain is complex and also time consuming especially for long hour duration data.

Some symptoms are overlooked or detected wrongly due to the human error or weak appearance in the disease symptom itself. Therefore, it is important to introduce different contexts in analysing heart condition. To ensure the development of the proposed method, which works accordingly and has the capability to overcome such a shortcoming, a knowledge discovery in databases (KDD) method [5] is used for this study.

KDD is an interdisciplinary area utilizing data mining approach to extract useful information from raw data and exploring unknown patterns of a large dataset. The core of this framework does not only focus on utilizing data mining approach but also involve in the understanding, preparation and interpretation of the data. The analysis of time series data for knowledge discovery is performed in its own unique way and it requires a specific application to accomplish such a task. Moreover, in this method, the key information of interest is concentrating on a particular time series region in ECG data, known as abnormal heart events and not just continuous behaviour in time series data only.

There are huge potentials for analysing such large volume of biomedical data at a specific segment or event to discover new knowledge. Shadabi and Sharma [6] showed the success stories utilizing this model constructively. From the literature perspective, the evidence shows that KDD application has been applied to many domains of medicine in different contexts. From the expert knowledge based application [7-12], KDD application based on previous study [13, 14] to KDD application to support in decision making [15].

Next, the overview of KDD framework is explained. Conventionally, KDD framework includes six stages [16], which are:

1) **Domain and data understanding:** General characteristic of the data will be analysed in detail. It will involve in finding any relevant knowledge for proposed developing system.

2) **Data selection:** Process of determining the most suitable source of data interest in target dataset is done properly here.

3) **Data pre-processing:** Once the most suitable source of data of interest is decided, the process of ensuring quality of data in the best condition is conducted on the target dataset. Procedures such as Dimension Reduction (Feature selection and Sampling), Data cleansing (Removal of Noise or outliers and handling missing values) and data transformation (Attribute Extraction) are implemented.

4) **Data transformation and reduction:** In this phase, the process of checking if the data format is suitable for running data mining algorithm is done properly.

5) **Data mining:** In this phase, selecting the proper data mining technique to search hidden pattern is done. After that, the process of employing selected algorithm for further analysis is executed.

6) **Knowledge interpretation:** In the last stage of KDD, the process of evaluation, assessment and interpretation of the knowledge inferred from the models is conducted.

1.2 Contribution and dissertation organization

The key contribution of this thesis is the proposal of an efficient, quickly and highly sensitive mechanism to accurately detect abnormalities of heart condition based on Arrhythmia symptom. In this study, new mechanisms which accurately detect abnormalities of heart condition based on statistical and data mining approach is proposed. Arrhythmia symptom is mainly selected as a reference to represent the abnormalities of heart condition. Arrhythmia symptom consists of more than 10 types of symptoms and each symptom has its own unique characteristics. To ensure the sensitivity of the proposed mechanism is addressed accurately, an early stage of Atrial Fibrillation database is selected with 12 hours duration period to evaluate the proposed mechanism.

As a conclusion, the main contributions of this dissertation can be summarized as follows:

- Proposal of a constructive framework to detect abnormalities of heart condition regardless the origin of the symptom based on Arrhythmia symptom by using statistical and data mining approach.
- Proposal of simple computational modelling to detect Arrhythmia symptom based on two parameters.
- Proposal of a constructive framework to detect an early stage of Atrial Fibrillation by using custom algorithm and neural network approach.

- Proposal of simple and accurate algorithm to detect healthy heart condition for long hour duration ECG data.
- Proposal of quantitative classification procedure to classify an early stage of Atrial Fibrillation symptom based on error output in neural network.
- Proposal of simple computational modelling to detect an early stage of Atrial Fibrillation symptom based on 5 parameters.

The organization of this thesis is graphically illustrated in Figure 1.3 and is described as follow:

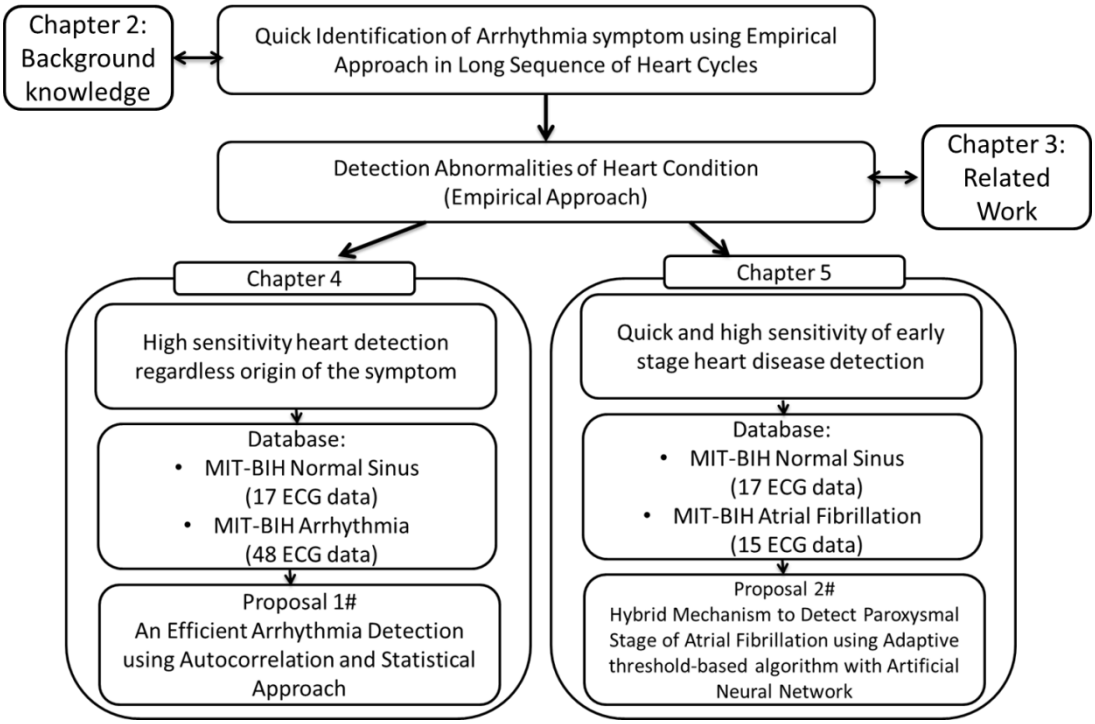

Figure 1.3: Overview of the study

Chapter 1: Introduction. The background and the motivation of this research were described in detail in this chapter. The challenges and shortcoming in dealing with the conventional method and the reason why the computational intelligence is needed were also explained in this chapter. Next, an introduction to Knowledge Based Discovery model was discussed in detail for development of the computational intelligent mechanism for this study. Finally, the primary contributions of this research were also concretely summarized in this chapter.

Chapter 2: Background knowledge (Principle). The background knowledge of the current research in the ECG data and heart disease is discussed here. The ECG segment, the heart disease and the non-disease symptom are covered in this chapter.

Chapter 3: Related Works. This chapter presented some works related to this study based on three perspectives. The first is the detection mechanism and their performance in detecting Arrhythmia symptom. The second is the impact of heart disease detection based on the number of attributes and the third is the accuracy of detecting Atrial Fibrillation symptom.

Chapter 4: An Efficient Arrhythmia Detection Using Autocorrelation and Statistical Approach. This chapter proposes an autocorrelation method with K-Nearest Neighbour (KNN) classifier method to accurately and robustly detect 14 types of Arrhythmia symptoms regardless of the origin of the symptoms in a long hour data. Moreover, variability analysis based on periodic autocorrelation result is proposed and used for the classification procedure. 1 minute and 12 hours duration data are chosen to compare and signify the most suitable time duration to detect Arrhythmia

symptom. In addition, the effectiveness of the proposed method in comparison with other methods is discussed. Moreover, complete result can be seen in this chapter.

Chapter 5: Hybrid Mechanism to Detect Paroxysmal Stage of Atrial Fibrillation using Adaptive Threshold-based Algorithm with Artificial Neural Network. In this chapter, a new mechanism called “Door-to-Door” algorithm is introduced to accurately and quickly detect significant peaks of heart cycle in 12 hours of ECG data and to discriminate obvious Preliminary stage of Atrial Fibrillation rhythms from Normal Sinus rhythms. In addition, a quantitative method using Artificial Neural Network (ANN), which discriminates unobvious Paroxysmal stage of Atrial Fibrillation rhythms from Normal Sinus rhythms is investigated. The effectiveness of the proposed method with the detail evidence are provided and discussed in this chapter.

Chapter 6: Discussion. This chapter discusses the work investigated and the solution proposed in this dissertation, and the significance of each result is summarized in detail. In addition to that, the difficulty and challenge in detecting the symptom is discussed.

Chapter 7: Summary. This chapter summarizes and concludes the dissertation, stating the benefits, limitations and difficulties. Finally, the research direction for possible future works is suggested.

CHAPTER 2

BACKGROUND KNOWLEDGE

In this chapter, the background knowledge of this study is discussed in detail. It includes the general knowledge of the principle of electrocardiogram, the ECG segment, the heart disease and the non-disease symptom. The details are discussed in the next section.

2.1 Principle of Electrocardiogram

Electrocardiography is considered the most conventional procedure to measure heart condition. It records the heart's electrical activities in time domain to check the abnormalities that may exist. It also provides information of the patient's heart rate and rhythm and shows if there is enlargement in the heart due to diseases such as high blood pressure (hypertension) or evidence of a previous heart attack (myocardial infarction). Once all the analysis is done, a proper treatment can be made accordingly based on the findings. A typical ECG waveform periodically consists of 5 main waves; P, Q, R, S and T waves. The P wave represents the depolarization of the right and left atria. The P, Q, R and the S waves depict the activation of the right and left ventricles, while the T wave shows the repolarization of the ventricular. It is indispensable to detect periodic heart cycle in order to identify any disease inside the heart. In this study, three types of symptoms are mainly focused on which are Normal Sinus, Arrhythmia and early stage of Atrial Fibrillation.

2.1.1 Normal Sinus Rhythm

Sinus rhythm is the set of heart's normal regular rhythms by the heart's natural pacemaker called sinoatrial node. Normal cardiac impulses start at the right atrium wall and are transmitted to the atria, then down to the ventricles. Additionally, it is a reflection of normally functioning conduction system in the body. This electrical current is following the normal conduction pathway without interference from other bodily system or disease processes [17]. The Normal Sinus signal pattern can be seen in Figure 2.1.

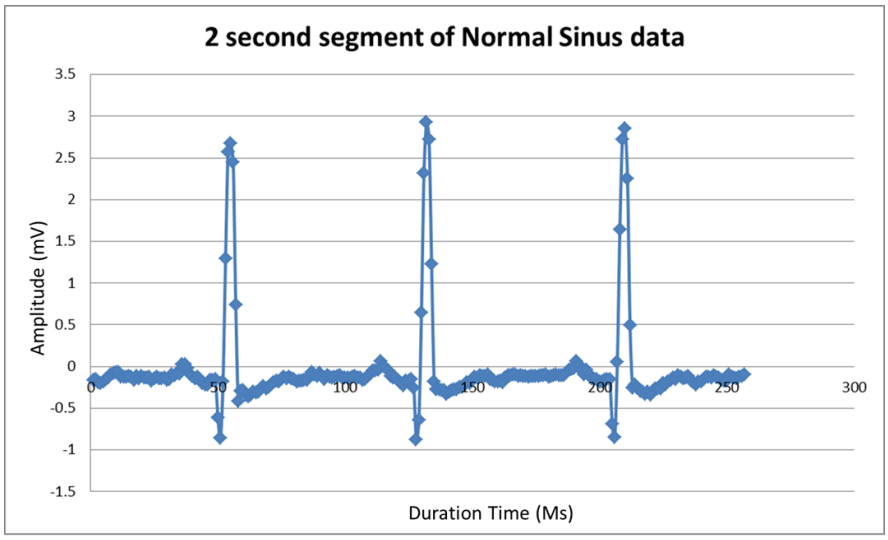

Figure 2.1: ECG data with Normal Sinus symptoms

2.1.2 Arrhythmia

Arrhythmia refers to any irregular change from the normal sequence of electrical impulses of the heart, that is to say, the electrical impulses could be too fast, too slowly or erratic. If the heartbeat is too fast, it is called tachycardia, while the heartbeat is too slow, it is called bradycardia. Arrhythmia consists of more than 10 various types of abnormality symptoms. For each symptom, it consists of it`s own unique identification characterization. One example of Arrhythmia signal pattern can be seen in Figure 2.2.

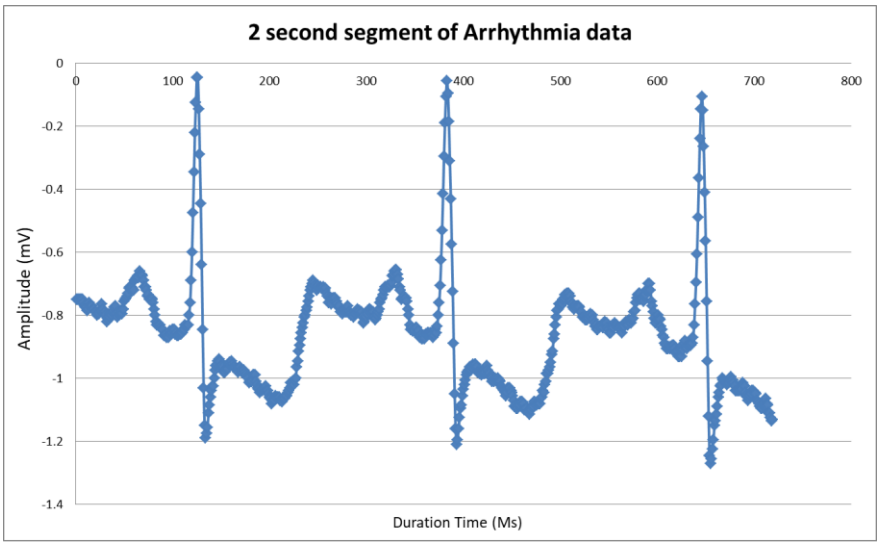

Figure 2.2: ECG data with Arrhythmia symptoms

2.1.3 Atrial Fibrillation Rhythm

Atrial Fibrillation rhythm is a situation where many different impulses rapidly fire at once, causing an unstable rhythm in the atria. Due to these unstable electrical impulses, the atria cannot contract or squeeze blood effectively into the vertical. Atrial fibrillation is the most common irregular heart rhythm that starts in the atria area. One example of Arrhythmia signal pattern can be seen in Figure 2.3.

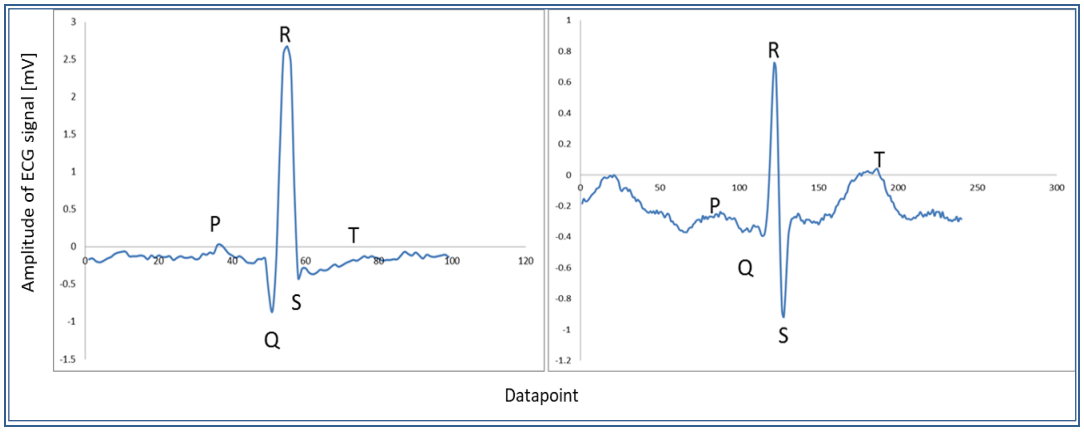

Figure 2.3: Normal Sinus Rhythm pattern (left) and Premature Atrial Fibrillation Rhythm pattern (right) with P, Q, R, S and T peaks

2.1.4 P, Q, R, S, T Wave Morphology

A normal ECG signal is considered as a periodic signal. This electrical signal of the heart consists of a sequence waves, conventionally labelled as P, Q, R, S and T. Each wave basically has a peak (hereafter P, Q, R, S or T peak), which is a spike or a dip. This sequence constitutes the sinus waveform of the heart signal. In the P, Q, R, S and T wave morphology, the process of depolarization and repolarization of the atrial and the ventricular occur. It represents a series periodical waves which start with the P wave and then followed by the QRS complex and T wave.

The first deflection is the P wave which is associated with the right and the left atrial depolarization. The second wave is the QRS complex. The first deflection in the complex with the negative wave is called the Q wave. Followed by R wave, it is represented as the first positive deflection in the complex. Next, the S wave is represented in a negative deflection after the R wave. The T wave represents ventricular repolarization. A normal ECG signal is considered as a periodic signal. However, it is regularly irregular wave if there is a sign of a heart problem.

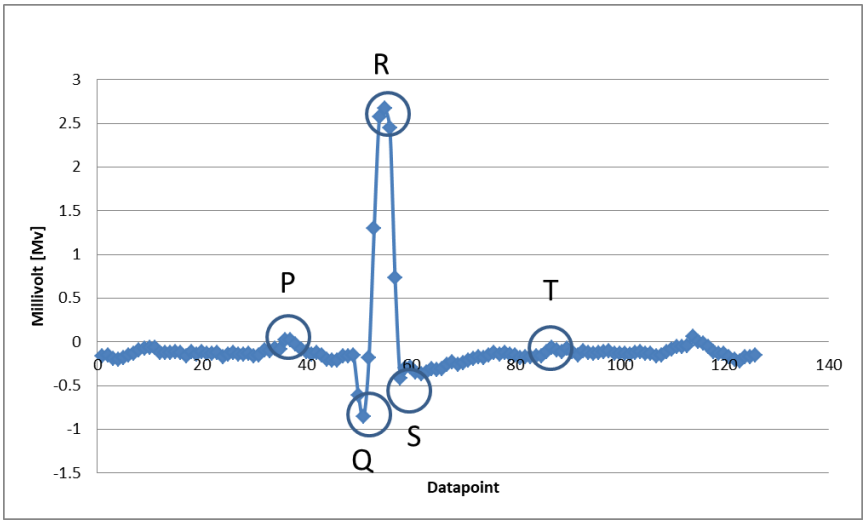

Figure 2.4: P, Q, R, S, T morphology

CHAPTER 3

RELATED WORK

In this chapter an extensive literature review is done to get a clear overview of the available solution in this study. For that purpose, in the next section, the list of related works with data mining approach and database will be shown and discussed.

3.1 Literature Review Overview

In this section, an extensive review based on three factors is considered.

- The detection mechanism and the performance of detecting Arrhythmia symptom.
- The accuracy of Atrial Fibrillation symptom detection.
- The impact of heart disease detection based on the number of attributes.

3.1.1 Arrhythmia symptom detection mechanism and its performance

Uday Maji et al. [18] proposed an ECG signal analysis by using Variational Mode Decomposition (VMD). This method is used to characterize atrial and ventricular Arrhythmia simultaneously and independently from each heart cycle segment. VMD is a non-recursive signal decomposition method proposed by

Dragomiretskiy et al. [19]. It is able to decompose the input signal into the desired number of mode N.

The experimental result revealed that the Arrhythmia detection accuracy of 99.1% and 99.8% at ventricular segment and at atrial segment, respectively, has been achieved. However, this method relies on two class multi-stage classifications. The two stage classifications are based on ventricular segment and atrial segment. The complexities in the classification process has become the main disadvantage for this study because it requires to separate the ECG segment into two parts for further step in abnormalities analysis. Therefore, this characterization of abnormalities under Arrhythmia symptom is less efficient and may consume more computational cost to reach the purpose.

Rekha Rajagopal et al. [20] reviewed the performance of five different types of unsupervised dimensionality reduction (DR) techniques in classifying cardiac Arrhythmias. DR technique is a method representing high-dimensional data in lower-dimensional spaces. Since interpretation of high-dimensional data remains a difficult task, DR technique is considered as one of the best alternative solutions to overcome the issue. In this study, they include principal component analysis (PCA), probabilistic Neural Networks (PNN) classifier, fast independent component analysis (fastICA) with tangential, kurtosis and Gaussian contrast function, Kernel PCA (KPCA) with polynomial kernel, hierarchical nonlinear PCA (hNLPCA) and principal polynomial analysis (PPA) for comparative study.

The overall result demonstrates that the combination of PNN classifier and fastICA DR technique with tangential contrast function shows a high accuracy of 99.3% to classify Arrhythmia. However, most of these approaches requires a huge amount of time even for low dimensional mapping. Moreover, they are complex in terms of implementation. The procedure involves an R peak detection using Pan Tompkins algorithm, segmentation and feature extraction, dimensionality reduction using three linear dimensionality reduction techniques with dimensions from 1 to 10 before using probabilistic classifier for classifications. Therefore, it is expected that the number of parameters should be minimal.

Likewise, Chia et al. [21] proposed hybrid adaptive feature selection mechanism for detection of Arrhythmia in ECG data. A combination of k-mean clustering and support vector machine was introduced and tested with more than 100,000 samples of Arrhythmia symptoms. The accuracy of 98.92% was achieved to detect Arrhythmia. However, the proposed method relies on three feature extraction processes, which are screening feature sample, partitioning the right sample and balancing the number of samples, before the classification in order to maintain a high detection rate.

3.1.2 Detection of Atrial Fibrillation

In this section, the comparison of four existing works in detecting Atrial Fibrillation symptom is made. Felix et al. [22] proposed an automatic multiscale-based peak detection algorithm in noisy periodic and quasi-periodic ECG signals. The achievement of this study is that the accuracy of R peak detection is 100% for 200 seconds of the ECG data. However, there is sceptism that the proposed algorithm can be applied to long hours ECG data as long hour ECG data may trigger miss detection to the ECG segment if there is no proposal to overcome or deal with such event.

In general, there are many ways to detect peaks for signal processing such as the use of wavelet transform [23]-[28], artificial neural network [29], [30], nonlinear filtering [31], linear prediction analysis [32], hidden Markov models [33], momentum [34] and window-threshold techniques [35]-[38]. The disadvantage with most of the existing peak detection algorithms is that different algorithms require different parameters to detect peaks. Therefore, these methods cannot be easily combined to achieve a high performance. Furthermore, the issue of signal noise is one of the biggest challenges.

Shadnaz et al. [39] proposed an automatic detection of Atrial Fibrillation using stationary wavelet transform and support vector machine. The proposed method achieved a sensitivity of 97.0% and specificity of 97.1% without relying on the detection of P peak, R peak and heartbeat. However, this method induces a computational complexity. During ECG feature extraction stage, data with a different frequency band requires a different stationary wavelet transform process. Moreover, the proposed method does not dynamically choose the most effective wavelet scale for noise reduction, resulting in less flexibility in implementation.

Likewise Andrius et al. [40] designed a low-complexity method to detect Atrial Fibrillation by observing the irregularity of R to R interval and associating this value with the increase of heart rate. The proposed method achieved a sensitivity of 97.1% and specificity of 98.3 %. The proposed method relies on several processes which are data pre-processing, R to R interval irregularity analysis, bigamy suppression analysis, signal fusion and detection analysis. There are two drawbacks in this method. Firstly, it has not been fully automated to analyse the ECG data. Secondly, this method is impractical because it requires a lot of time to analyse a long duration of ECG data.

Sujit et al. [41] reviewed several techniques for detecting Atrial Fibrillation from Non-Episodic ECG data. Several features have been defined to describe the behavior of Atrial Fibrillation by focusing on P wave, QRS waves cycle and R to R interval. However, most of the reviewed techniques do not focus on the data distribution model of Normal Sinus heart cycle to distinguish early stage of Atrial Fibrillation. Since the early stage of Atrial Fibrillation symptom has similarities in rhythm to Normal Sinus symptom. Therefore, a constructive data distribution model for Normal Sinus is required to classify the diseases. This data distribution model is basically regarded as a series of P, Q, R, S and T values in ECG data.

Some of the techniques to detect Atrial Fibrillation are K-nearest neighbour (KNN) [42], [43], Bayer Optimal classier [42], Artificial Neural Network (ANN) [42], [44], [45], Linear Discrimination Analysis [41] and Empirical Detector [46]. However, there exist a few drawbacks. It needs to define too many ECG parameters such as P,Q,R,S,T wave segment, the time difference between each wave, age, gender, blood type and so on to characterize Atrial Fibrillation, resulting in the increase of computational complexity. Although some methods [42], [43] attempted

to overcome this issue by reducing the number of parameters in characterizing the Atrial Fibrillation, the relation of the trade-off between the computational complexity and the classification performance has not been solved.

In this study, to overcome all the listed issues above, a suitable mechanism with a computational intelligence approach for heart disease diagnosis is proposed. The details of the proposed mechanism will be explained in the chapter 4 and chapter 5.

3.1.3 Comparison of data mining approach based on number of attribute used in classification.

In this section, 29 related works is provided in Table 3.1 to support the evidence on how the number of attributes can influence the accuracy of heart disease detection. Based on 29 related works, the most minimal number of attributes used to detect a heart disease is 6 and the highest number of attributes is 17. The most minimal number of attributes used has achieved 99.2%, 96.5%, and 88.3% accuracy with genetic decision tree, genetic Naïve Bayes and classification via clustering. On the other hand, the highest number of attributes used for detecting the disease only achieved 55.23%, 52% and 45.67% accuracy with Naïve Bayes, Decision Tree and KNN.

The comparisons of computational approaches to detect heart disease with different number of attributes are shown in Table 3.1.

Table 3.1: Computational approaches to detect heart disease with different number of attributes

Author, Year	Technique	No. of Attributes	Accuracy (%)
Resul et al. [47],2009	Neural Network	13	89.01
Anbarasi et al.[48],2010	Genetic with Decision Tree	6	99.2
	Genetic with Naïve Bayes		96.5
	Genetic with Classification via Clustering		88.3
Rajkumar et al.[49],2009	Naïve Bayes	17	52.33
	Decision Tree		52
	KNN		45.67
Kumari et al.[50],2011	Decision Tree	14	79.05
	Artificial Neural Network		80.06
	Support Vector		84.12
	RIPPER		81.08
Sundar,et al.[51],2018	WAC	15	84
	Naïve Bayes		78
Chaitrali et al. [52],2012	Artificial Neural Network	13	99.25
		15	99.9
John, et. al [53],2012	Naïve Bayes	14	85.18
	Multilayer		78.88
	J48		85.18
	KNN		85.55
Shouman et al. [54],2012	K-Nearest Neighbour	13	97.4
Nidhi et al. [55],2012	Naïve Bayes	15	90.74
		13	94.44
		6	96.5
	Decision Trees	15	99.62
		13	96.66
		6	99.2
	Neural Network	15	96.5
		13	99.2
		6	88.3
Pethalakshmi et al. [56],2012	Fuzzy Decision Tree	13	90.06
	Fuzzy Naïve Bayes		89.62
	Fuzzy Neural Network		91.09
	Fuzzy K-means		99.49
Abhishek et al.[57],2013	J48	15	95.56
	J48		94.85
	Naïve Bayes		92.42
Chitra et al [58],2013	Artificial Neural Network	14	85
	K-Means Clustering		88
	Fuzzy C Means Clustering		92
Dessai [59],2013	PNN	14	94.6
	Decision Tree		84.2
	Naïve Bayes		84
	BNN		84.6
Patel el al.[60],2013	Decision Tree	14	99.2
	Naïve Bayes		96.5
	Classification Clustering		88.3
Vikas and Pal [61],2013	CART Classification	11	84.49

Methaila et al. [62],2014	Naïve Bayes	15	96.53
	Decision Tree		99.2
	Classification via Clustering		88.3
Wisaeng [63],2014	K-Nearest Neighbor	14	93
		8	90
Waghulde et al.[64],2014	Neural Network & Genetic Algorithm	13	98
Rupali et al. [65],2014	Classification using Naïve Bayes	14	78
	Classification using Laplace Smoothing		86
Venkatalakshmi et al. [66],2014	Naïve Bayes	13	85.03
	Decision Tree		84.01
Jarad [67],2015	Naïve Bayes	14	85.03
	Decision Tree		52
	KNN		45.67
D`Souza [68],2015	ANN	14	79.38
	K-Mean Clustering		63.299
Baiju and Janet [69],2015	Naïve Bayesian classification Technique	13	81
Adbar et al. [70],2015	C 5.0	14	93.02
	NN		89.4
	SVM		86.05
	KNN		80.23
Kaur and Kaut [71],2015	SVM Classifier with Genetic Algorithm	12	95
Swati et al. [72],2015	Naïve Bayes	13	84
	KNN		76
Patel et al. [73],2016	J48	13	56.76
Rajalakshmi et al.[74],2016	K-Means clustering	14	93.89
	WAC		92.84
Suganya et al. [75],2016	CART Classifier	14	83
Karthikeyan et al. [76],2017	Deep Belief Network	16	90
Wadhawan [77],2017	K-Means Using Apriori Algorithm	7	74

Different data mining methods have their strengths and weaknesses in dealing with biomedical data. There is huge concern in this domain for more robust algorithm which can minimise the noise in the dataset because the dataset may consist various types of redundancy and noise. In order to determine the right method to deal with the problem, it is important to compare major data mining capability so that the best proposed method is used to solve the main issue. Therefore, the comparison table for

major machine learning algorithm is prepared to give a new insight on these matters [77]. Table 3.2 provides the detail of major machine learning algorithm based on five criteria.

Table 3.2: Comparison major machine learning based on 5 criteria

Technique	Outlier	Over fitting and under fitting	Parametric	Accuracy	Execution on technique
Support Vector Machine	It can handle outlier properly	Perform better than over fitting and under fitting	Non parametric model	Higher than other parametric model	Depend upon dataset used, generally quite slow NLP operation
Decision Tree	Outlier does not critical role in interoperation of dataset by decision tree.	It suffer over fitting and under fitting	Non parametric model	Accuracy depend on the dataset, hybrid technique used with decision tree have higher accuracy than SVM	Require less time than other parametric model if not suffering from over fitting where hybrid technique need higher execution than decision tree
Naïve Bayes	It is less pruned to outlier	It does not suffer over fitting and under fitting	It is parametric model	High with limited dataset	Low with limited dataset
Artificial	It is pruned to	It is more	It is parametric	Higher than	Execution time

Neural Network	outlier	pruned to over fitting than support vector machine	model	other parametric model	depend upon number of layer declared and number of epochs need for testing
Linear regression model	It is less pruned to outlier because it strong probabilistic background	It does not suffer from under fitting and over fitting	It is parametric model	Higher for linear dataset	Require less execution time than other model.

CHAPTER 4

AN EFFICIENT ARRHYTHMIA DETECTION USING AUTOCORRELATION AND STATISTICAL APPROACH

This chapter will explain in detail about an efficient Arrhythmia detection using Autocorrelation and Statistical Approach. As mention in the introduction section that the overall focus of this study is to design computational intelligence mechanism to detect abnormalities of heart condition accurately. The focus was divided into two sub objectives. First is to propose mechanisms which have the capability to detect abnormalities of heart symptom regardless the origin of the symptom in ECG data. Second is to propose mechanisms which strongly have the sensitivity in detecting abnormalities of heart condition as early stage as Paroxysmal stage of Atrial Fibrillation.

For this chapter, the first objective will be served first. However, to design a very high accuracy mechanism to detect abnormalities of heart condition regardless the origin of the symptom in ECG segment, there are two factors influential has been identified in the literature review in chapter 3. One of them is the number of attributes used to classify the heart disease. Based on the literature review has proven that most of the existing work need to rely on more than 5 attributes to achieve good accuracy in disease classification. However, there is a trade-off between complexity of classification procedure and the number of attributes used for classification.

Therefore, in this chapter, the first research question will be answered in detail with strong evidence. The first research question is addressed here. The research questions whether there are some influential factors toward the accuracy of detection if the minimal number of attributes is used. Moreover, one of the concerns is this study is to reduce the complexity in detecting the abnormalities of heart condition computationally. Therefore, the detail of this study will be explained in the next section.

4.1 First Proposal: Detection of Arrhythmia Symptoms Using Autocorrelation and KNN Classifier

In this section, an autocorrelation-based approach with KNN classifier is proposed to classify the Arrhythmia from Normal Sinus. Autocorrelation is a statistical method that can measure internal correlation within a time series domain. It is defined based on the concept of time lag. Performing autocorrelation of a time series data is beneficial especially to identify signal stationary condition, measure variability level of continuous data or even indicate quantitative relation of some previous data points occurring with a time lag. Conventionally, ECG data is a time series data and it provides very useful information of heart condition. There are various types of heart disease symptoms and the characteristic differs from one symptom to another.

However, there is a common characteristic shared among heart disease symptoms, which is inconsistent irregularity in shape that appears in time series domain. In this study, a quantitative analysis is proposed to numerically characterize the two symptoms. Variability analysis based on the first autocorrelation periodic cyclic result is introduced. Two parameters, which come from the first periodic slope of autocorrelation result, will be used for analysis. Based on the two parameters, a classification procedure to discriminate the two different groups of data will be performed by using KNN classifier. KNN is a common classification method used in pattern recognition. Without relying on any specific segment in ECG cycle to classify the disease accurately, an autocorrelation function can simplify the detection mechanism by relying on variability behaviour of large group data from two symptoms only. The details of procedure for autocorrelation, KNN and the analysis will be stated in Section. 4.1.1 and 4.1.2.

4.1.1 Autocorrelation Coefficient

Autocorrelation also known as “serial correlation” or “lagged correlation” is a statistical method that measures dependency of variables arranged in time. There are three tools for assessing the autocorrelation of a time series data which are time series plot, the lagged scatterplot and the autocorrelation function. In this study, autocorrelation function will be used to measure variability level of long hour serial correlation of ECG data. The ECG data with 12 hours duration time frame is proposed to evaluate the correlation of Arrhythmia disease with time via autocorrelation function. It is assumed that time can influence the characteristic of

Arrhythmia. A 12 hours duration of ECG data compared to a short duration of ECG data can have an impact on the result. Here in this section, an autocorrelation coefficient function will be explained in detail.

Let yi (i=1, n) be a time series data at the data point i, and let α be the average value of all the data.

The autocorrelation function at lag k is defined as:

$$r_k = \frac{\sum_{i=1}^{n-k}(y_i - \alpha)(y_{i+k} - \alpha)}{\sum_{i=1}^{n}(y_i - \alpha)^2}$$

As the autocorrelation function shows, r_k describes the correlation between the two data which are located with the lag *k* each other. The score ranges from -1.0 (perfect negative relation) to +1.0 (perfect positive relation). If there is no correlation between the two variables, the score is zero.

Next, the procedure to classify the two symptoms is mentioned below shown in Figure 4.1.

(1) Perform an autocorrelation to 65 ECG data with 12 hours duration and describe the score with lag *k* as the autocorrelation result.

(2) Investigate the first periodic slope segment in autocorrelation result based on two parameters. The two parameters are:

(a) Peak values of the first periodic slope.

(b) Time length of the first periodic slope as shown in Figure 4.2.

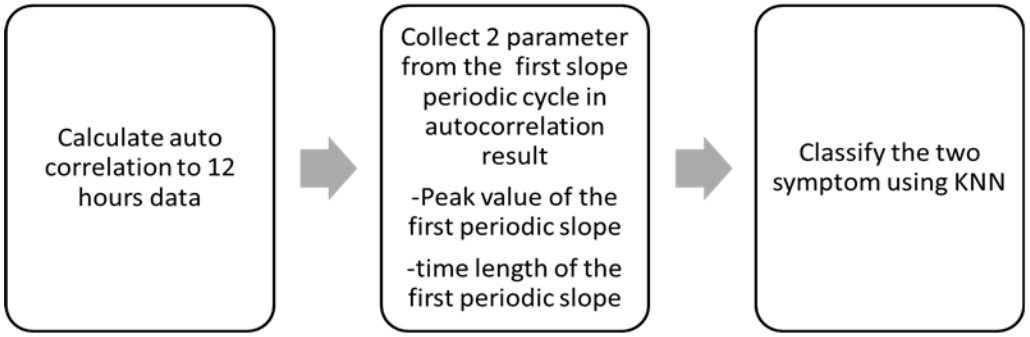

Figure 4.1: Proposed automatic mechanism to detect Arrhythmia symptom

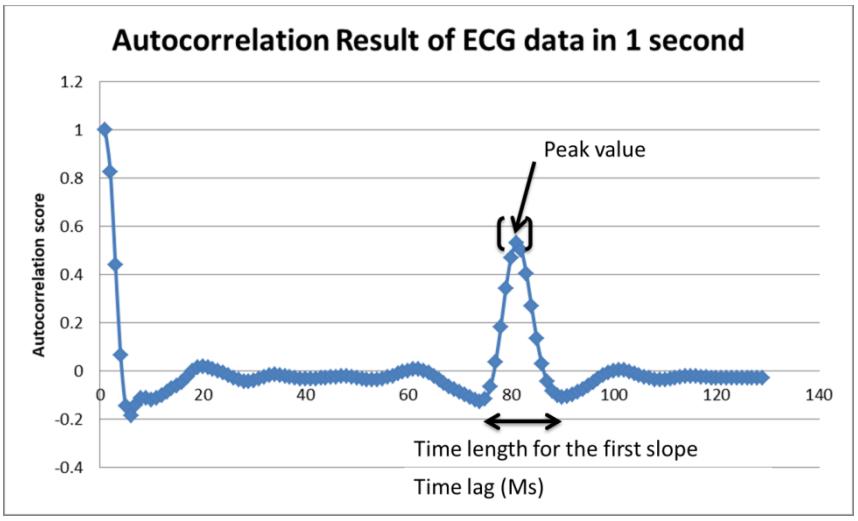

Figure 4.2: Two important parameters for classification at the first periodic slope

Figure 4.3 represent 2 second segment of Normal Sinus autocorrelation result and Figure 4.4 represent 2 second segment of Arrhythmia autocorrelation result in 1 minute duration. For Figure 4.5, it represent 2 second segment of Normal Sinus autocorrelation result in 12 hours duration while Figure 4.6 represent Arrhythmia autocorrelation result with the same duration. In these figures, the horizontal axis indicates the time lag of time series data while the vertical axis indicates the autocorrelation score of serial correlation output. Note that the first periodic slope

peak value of autocorrelation result for Normal Sinus is much lower than the one for Arrhythmia.

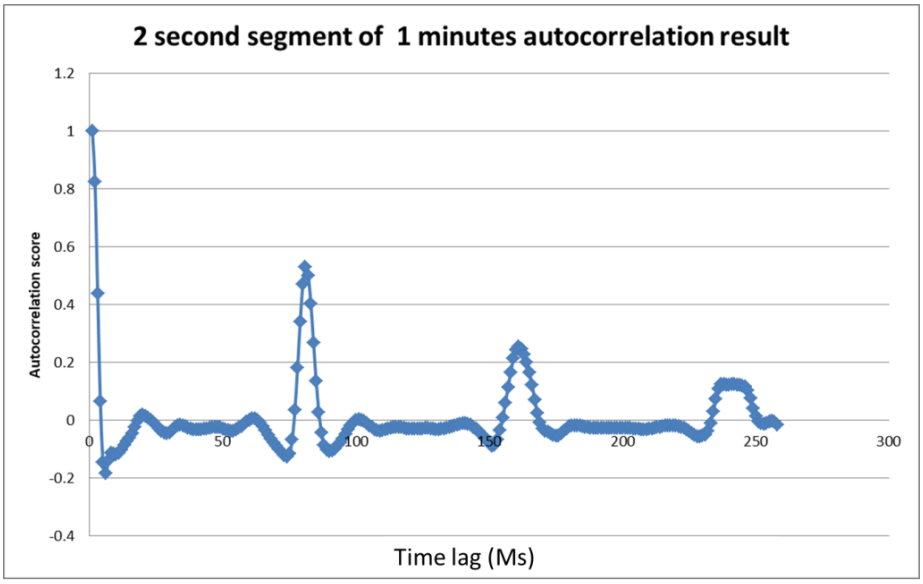

Figure 4.3: Example of 2 second segment in 1 minute autocorrelation result with Normal Sinus symptoms

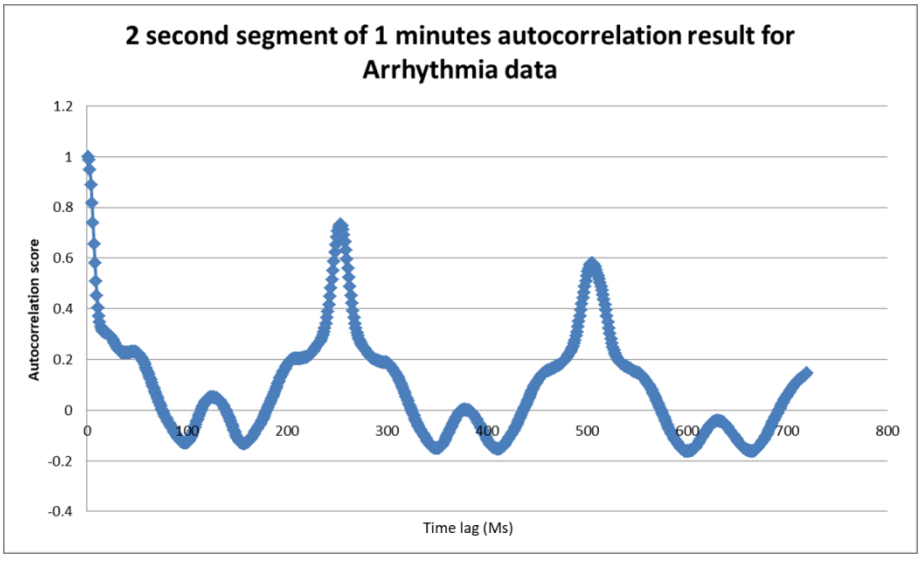

Figure 4.4: Example of 2 second segment in 1 minutes autocorrelation result with Arrhythmia symptoms

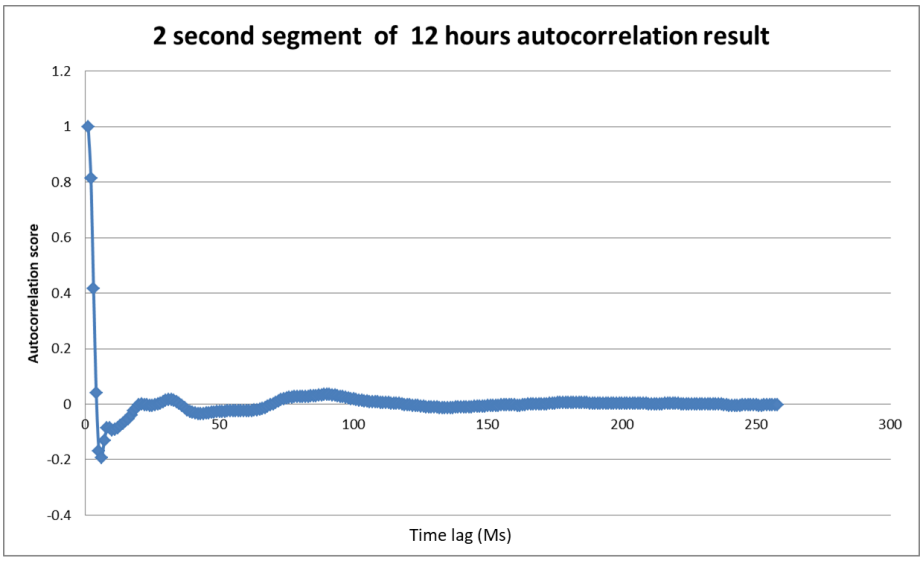

Figure 4.5: Example of 2 second segment in 12 hours autocorrelation result with Normal Sinus symptoms

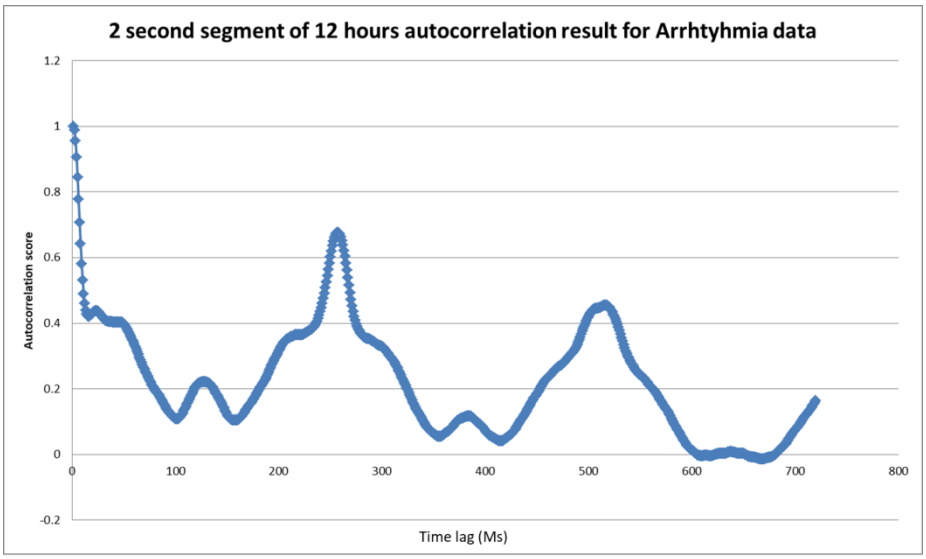

Figure 4.6: Example of 2 second segment in 12 hours autocorrelation result with Arrhythmia symptoms

4.1.2 Discrimination of Arrhythmia from Normal Sinus with supervised machine learning classifier

In this research, the values of significant first peak and time length of each first periodic slope in autocorrelation result are investigated in detail. KNN classifier is introduced to discriminate Arrhythmia symptoms from Normal Sinus. KNN is an instance-based learning method which keeps all available cases and classifies new cases based on a similarity measurement using distance function. KNN is a non-parametric approach and performs very fast. Also it makes no assumption about the data distribution before classification. As a result, the flexibility of KNN's decision boundary is a huge advantage compared to other approaches which rely on linear, elliptic or parabolic decision boundaries to classify specific group of data. Therefore, this method has been chosen for this study to classify the Arrhythmia symptoms from Normal Sinus.

4.2 Performance Evaluation

In this section, the performance of the proposed method to discriminate Arrhythmia symptoms from Normal Sinus symptom is evaluated. In the following subsections, the database used for the evaluation experiment and the analytical results are discussed.

4.2.1 Arrhythmia Database for Evaluation

Two types of database were introduced in this study, which are “MIT-BIH Arrhythmia” and “MIT-BIH Normal Sinus”. They have been provided by Physionet [4]. Physionet is an open source database which provided a large number of recorded physiological data and the related open-source software. 48 records from MIT-BIH Arrhythmia and 17 records from MIT-BIH Normal Sinus were selected to evaluate the performance of the proposed mechanism. MIT-BIH Arrhythmia includes 14 types of abnormalities as shown in Table 4.1. The sampling frequency for each record is 360Hz. MIT-BIH Normal Sinus consists of ECG data with healthy heart condition. The sampling frequency for each record is 129 Hz. The amplitude range of typical ECG signal is from -5 mV to 5 mV. The time duration for each record is approximately 12 hours.

As mention in Section 4.1, the autocorrelation was performed for all the ECG data; MIT-BIH Arrhythmia and MIT-BIH Normal Sinus, selected from the above mentioned database. The input parameters for KNN classifier were the autocorrelation score and the time length of the first periodic slope for each ECG data. Since there are various types of Arrhythmia symptoms covered in this research, the effectiveness of the proposed mechanism was evaluated based on the capability to detect the symptom regardless the origin and classifying accuracy.

Table 4.1: Detail of information for MIT-BIH Arrhythmia database

MIT-BIH Arrhythmias Database Detail information	
Heartbeat Type	**Total**
Normal rhythm	74607
Left bundle branch block	8069
Right bundle branch block	7250
Atrial premature contraction	2514
Premature ventricular contraction	7127
Paced beat	7020
Aberrated atrial premature beat	150
Ventricular flutter wave	472
Fusion of ventricular and normal beat	802
Non-conducted P-wave (Blocked APC)	193
Nodal (Junctional) escape beat	229
Fusion of paced and normal beat	986
Ventricular escape beat	106
Nodal (Junctional) escape beat	83
Atrial escape beat	16
Unclassified beat	35
Total	109655

4.2.2 Result of Experiment

Quantitative analyses were performed to evaluate the proposed method. To demonstrate the performance improvement with the proposed Arrhythmia detection mechanism, the performance evaluation was divided into two sections: (1) Accuracy, sensitivity and specificity of the proposed method to discriminated Arrhythmia from Normal Sinus, (2) Statistical analysis of the proposed approach.

4.2.3 Accuracy, Sensitivity and Specificity Evaluation

To evaluate the classification performances of the proposed mechanism, “Accuracy”, “Sensitivity” and “Specificity” were selected as the evaluation metrics. The sensitivity and specificity are considered as the best paired performance metrics to evaluate the classification accuracy of heart disease [78]. The true positive rate of sensitivity represents the proportion of actual positives correctly identified as Arrhythmia data as having the Arrhythmia [79] .On the other hand, the true negative rate of specificity represents the proportion of actual negatives correctly identified as Normal Sinus and not having the Arrhythmia condition [79]. Accuracy rate represent the overall ratio of the proposed method to differentiate Normal Sinus and Arrhythmia correctly.

The discrimination accuracy, sensitivity and specificity are defined as follows:

Sensitivity = True positives / (True positive + False negative)

Specificity = True negatives / (True negative + False positives)

Accuracy = (True negatives + True positive)/ (True negatives + True positive+ False negative + False positive)

where,

True positive: The number of Arrhythmia data correctly identified as Arrhythmia data.

False negative: The number of Arrhythmia data incorrectly identified as Normal Sinus data.

False positive: The number of Normal Sinus data incorrectly identified as Arrhythmia data.

True negative: The number of Normal Sinus data correctly identified as Normal Sinus data.

Figure 4.8 and Figure 4.9 represent the distribution of the peak value and the time length of the first periodic slope with 66 patients' data with 1 minute and 12 hours duration, respectively. The result in Figure 4.8 shows that the Arrhythmia data and Normal Sinus data are not well discriminated. On the other hand, the result in Figure 4.9 shows that almost all the Arrhythmia data is scattered without overlapping the Normal Sinus data. Only three Arrhythmia data are overlapped with the region of Normal Sinus data. In order to classify these two symptoms computationally, a KNN classifier was used.

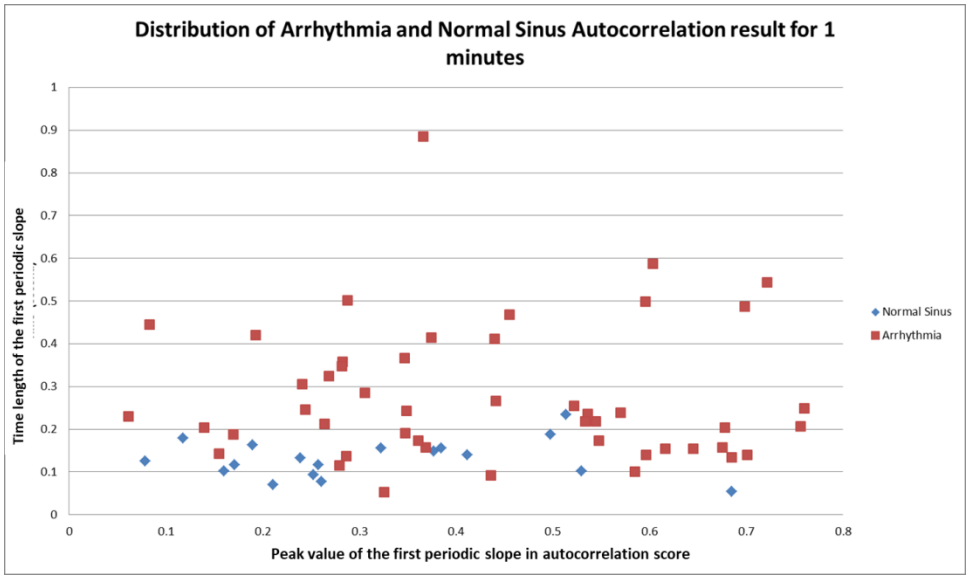

Figure 4.7: Distribution of autocorrelation result based on first periodic slope peak value with time length for 1 minute data

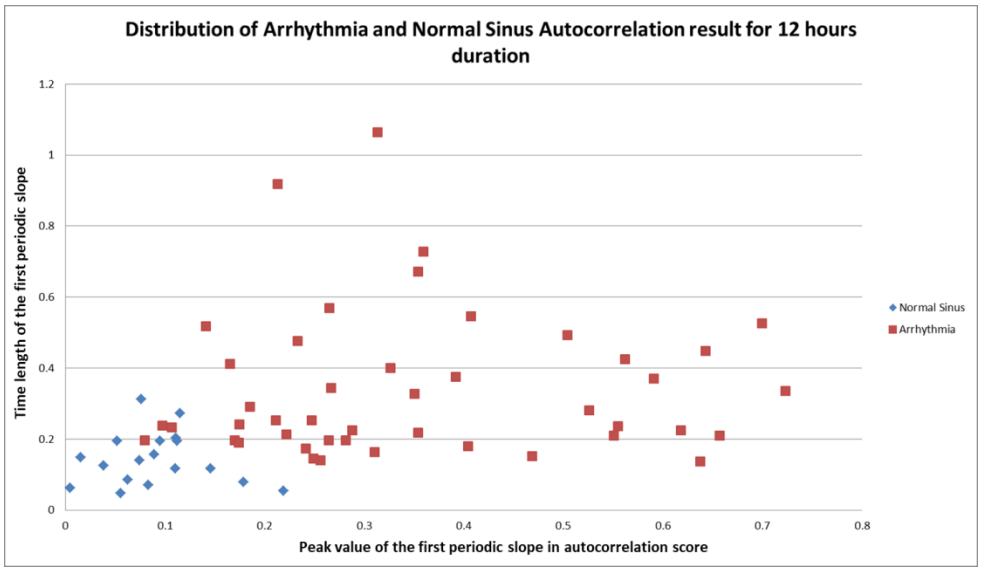

Figure 4.8: Distribution of autocorrelation result based on first periodic slope peak value with time length for 12 hours data

In this evaluation, to classify Arrhythmia from Normal Sinus, KNN is used. 83.3% sensitivity of Arrhythmia detection and 55.5% specificity of Normal Sinus detection were achieved for 1 minute's duration data. However, 97.9% sensitivity of Arrhythmia detection and 88.8 % specificity of Normal Sinus detection were achieved for 12 hours duration data as shown in Table 4.2. The higher rate of Arrhythmia detection was achieved even all 14 various types of symptoms were normalized into a series of correlation using autocorrelation. It is confirmed that the peak value and the time length of the first periodic slope in the autocorrelation result for 12 hours duration data are effective parameters to classify Arrhythmia from Normal Sinus with high detection accuracy.

Table 4.2: Sensitivity, Specificity and Accuracy of Arrhythmia and Normal Sinus symptom detection using KNN with the proposed method

Performance of proposed method (Fine KNN + Autocorrelation)		
	1 minute	12 hours
Accuracy	72.7%	95.5%
Sensitivity	83.3%	97.9%
Specificity	55.5%	88.8%

In order to compare the KNN classifier used here with others in the proposed method, other 17 types of supervised machine learning classifiers were selected and evaluated. Table 4.3 represent the performance comparison using 17 type of classifier using same dataset with 1 minute duration while Table 4.4 is focusing on the performance comparison using 17 type of classifier with 12 hours duration. Figure 4.9 and Figure 4.10 represent the overall performance of 1 minute and 12 hours data based on 17 various types of classifier in histogram plot. The evaluation metrics for the performance comparison are accuracy, sensitivity and specificity.

Table 4.3: Performance comparison using 17 types of classifiers using the same dataset with 1 minute duration

Overall performance of 17 various classifier			
Classifier Name	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Linear SVM	78.8	89.5	50
Quadratic SVM	75.8	85.4	55.5
Cubic SVM	80.3	91.6	72.2
Fine Gaussian SVM	78.8	77.0	55.5
Medium Gaussian SVM	75.8	85.4	50
Coarse Gaussian SVM	72.7	100	0
Fine KNN (Proposed method)	72.7	83.3	55.5
Medium KNN	81.8	89.5	61.1
Coarse KNN	72.7	100	0
Cosine KNN	74.2	83.3	50
Cubic KNN	83.3	89.5	66.6
Weighted KNN	78.8	87.5	55.5
Linear Discrimination	78.8	91.6	44.4
Quadratic Discrimination	81.8	83.3	77.7
Complex Tree	80.3	81.25	77.7
Medium Tree	80.3	81.25	77.7
Simple Tree	78.8	83.3	66.6

Table 4.4: Performance comparison using 17 types of classifiers using the same dataset with 12 hours duration

Overall performance of 17 various classifier			
Classifier Name	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Linear SVM	86.4	91.6	83.3
Quadratic SVM	86.4	91.6	83.3
Cubic SVM	87.9	89.5	83.3
Fine Gaussian SVM	87.9	97.9	61.1
Medium Gaussian SVM	86.4	91.6	83.3
Coarse Gaussian SVM	72.7	100	0
Fine KNN (Proposed method)	95.5	97.9	88.8
Medium KNN	86.4	93.7	77.7
Coarse KNN	72.7	100	0
Cosine KNN	68.2	83.3	33.3
Cubic KNN	86.4	93.7	77.7
Weighted KNN	95.5	97.9	88.8
Linear Discrimination	87.9	93.7	83.3
Quadratic Discrimination	83.3	87.5	83.3
Complex Tree	89.4	95.8	72.2
Medium Tree	89.4	95.8	72.2
Simple Tree	89.4	95.8	72.2

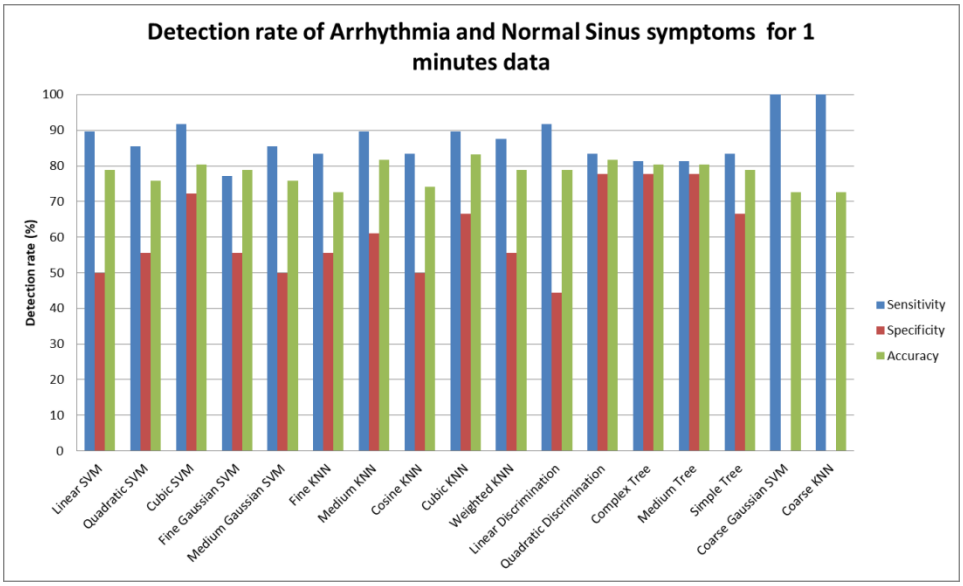

Figure 4.9: Sensitivity, Specificity and Accuracy detection rate of Arrhythmia and Normal Sinus symptoms for 1 minute's duration based on 17 various type of classifier

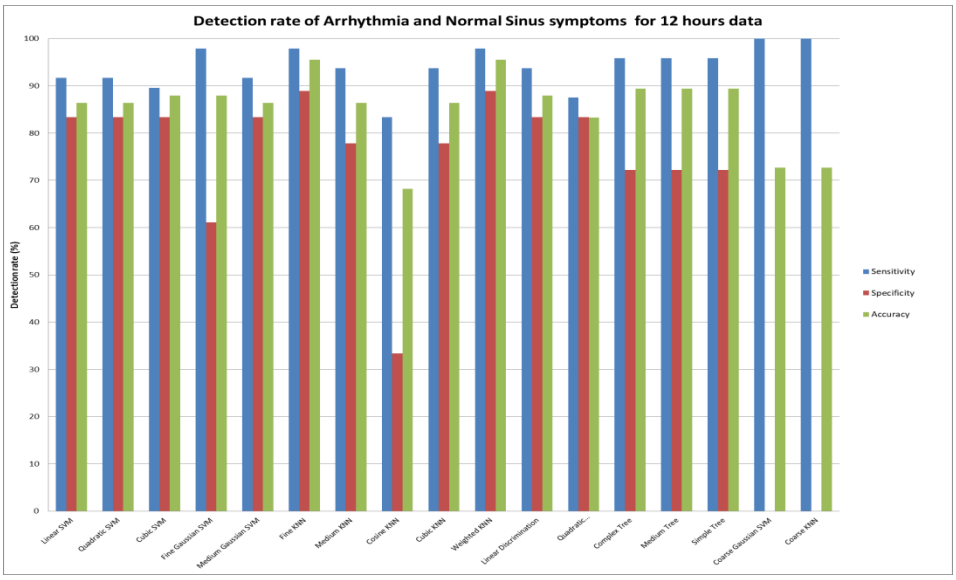

Figure 4.10: Sensitivity, Specificity and Accuracy detection rate of Arrhythmia and Normal Sinus symptoms for 12 hours duration based on 17 various type of classifier

4.3 Discussion

To complete this study, an analytical result and discussion was conducted. Based on the result with 12 hours duration data, the distributions pattern of the first periodic slope peak value with the time length are very identical for both symptoms. Correlation coefficient peak value of the first periodic slope in Arrhythmia shows higher score tendency compared to Normal Sinus. Although correlation is conventionally being interpret based on the score size of the relationship, but it is very important to understand the influential factor affecting the size of the score before interpreting it.

First, the nature of raw data itself and second is the implementation mechanism of autocorrelation method towards the data. Based on related finding have shown that there are 6 factors that can influence the correlation coefficient score [80]. For this study, there are three factors are correlated with the finding especially on why autocorrelation score for the first periodic slope are much higher with Arrhythmia compared to Normal Sinus. Those factors are amount of variability in the dataset [80][81], the size of the sample [80-82] and the characteristic of the sample data itself [80-83].

Typically, Normal Sinus and Arrhythmia are difference in shape and time. Normal Sinus conditions are more stable and consistent. In Arrhythmia, the symptoms are irregular heartbeat in time series. The heart may beat too fast, too slowly, too early, or irregularly. It was assumed that the amounts of variability in Normal Sinus are much lower compared to Arrhythmia.

Although Normal Sinus data may include some noise due to unexpected patient behaviour or miss conducted procedure during ECG recording but those events are small compared to the nature of Arrhythmia symptom itself. Hence, the presence of an outlier in a dataset can only influencing the correlation score based on the two factors [80]. First, the location of the outlier itself and second, the size of the sample is small enough. In this experiment, each record consists of approximately more than 700,000 data point. Even if there is a small group of outliers included in Normal Sinus data, the overall autocorrelation score will not be affected by it.

In this study, it is confirmed that the time length of the data is considered as the biggest influential factor towards the overall classification performance. Due to fact that Arrhythmia randomly appear in time series domain, longer duration is required to accurately detect the disease. In summary, the longer the ECG data is used, the more explicit the behaviour of the data can be seen. For that, high accuracy of detecting Arrhythmia can be achieved.

To compare the discrimination performance of the proposed method, Table 4.5 shows the result of sensitivity and specificity of related studies with the same focus with this research using the same database. Most of the related works are focusing on specific segment in ECG data to classify the symptom. Without relying on any specific feature like the proposed method, the possibilities to discriminate Arrhythmia from Normal Sinus can be done accurately. Best of our knowledge, there is no existing work has proposed as simple method as this study to detect and classify arrhythmia with high accuracy. It is revealed that the proposed mechanism has overcome the other studies in term of simplicity of the mechanism to identify Arrhythmia and abnormalities symptoms regardless the origin of the symptom, and the proposed approach to classify the two symptoms with high accuracy.

Table 4.5: Overall performance evaluation result comparing this research with other studies

Ventricular Arrhythmia analysis		
Method	**Sensitivity (%)**	**Specificity (%)**
Parham et al. [84]	99.5	99.66
Yun-Chi et al. [85]	98.28	-
Shing et al. [86]	98.75	99
Yun-Chi Yeh et al. [87]	98.28	-
Joo S. Lim [88]	99.02	96.67
Uday Maji et al.[89]	99.3	99.2
Atrial Arrhythmia analysis		
Jinseok Lee et al. [90]	94.7	94.4
Uday Maji et al.[89]	99.8	98.4
Arrhythmia analysis at both segment		
Proposed method	97.9	88.8

4.4 Conclusion

In this study, a novel approach, which statistically detects abnormalities of heart condition based on Arrhythmia symptoms using autocorrelation functions and KNN classifier, was proposed. A variability analysis based on periodic cycle in autocorrelation result was done. It is based on two parameters at first periodic slope of autocorrelation result. In order to discriminate the two symptoms, KNN classifier was used. The effectiveness of the proposed method was evaluated based on 3 performance evaluations metric which are accuracy, sensitivity and specificity. From the result, the overall accuracy was 95.5% with sensitivity of detecting Arrhythmia was 97.5% and the specificity of detecting Normal Sinus was 88.8%.

The comparison approach between this research and other studies shows that the proposed mechanism are robust, flexible and less complexity in detecting abnormalities symptoms like Arrhythmia. In this study, 17 different type of supervised machine learning classifier was used to compare with the proposed classifier. It is proven that fine KNN has outperforms the others classifier for 12-hour duration segment. 14 types of symptoms had cover in this study and there is no dependency towards any specific characteristic and feature segment in ECG data to identify each Arrhythmia symptom. It is justified the robustness of the proposed method in discriminating abnormalities of heart condition. It is confirmed based on this study that the time length of data is consider the biggest influential factor towards overall classification performance. Therefore, the longer time duration of ECG data is used, the higher accuracy the classifier can be achieved. It is concluded that this research finding can contribute to the medical field to identify Arrhythmia symptom with less complexity procedure in long hour duration.

CHAPTER 5

HYBRID MECHANISM TO DETECT PAROXYSMAL STAGE OF ATRIAL FIBRILLATION USING ADAPTIVE THRESHOLD-BASED ALGORITHM WITH ARTIFICIAL NEURAL NETWORK

This chapter will describe the detail of Hybrid Mechanism to Detect Paroxysmal Stage of Atrial Fibrillation using Adaptive Threshold-based Algorithm with Artificial Neural Network. Automatic detection of heart cycle abnormalities in a long duration of ECG data is a crucial technique for diagnosing an early stage of heart diseases. The necessities to detect an early stage of heart disease are important due to the fact that it can leads to more chronic illness like Stroke. Concretely, Paroxysmal stage of Atrial Fibrillation rhythms (ParAF) must be discriminated from Normal Sinus rhythms (NS) since it is the first stage of Stroke symptom. The both of waveforms in ECG data are very similar, and thus it is difficult to completely detect the Paroxysmal stage of Atrial Fibrillation rhythms. Previous studies have tried to solve this issue and some of them achieved the discrimination with a high degree of accuracy. However, the accuracies of them do not reach 100%. In addition, no research has achieved it in a long duration, e.g. 12 hours, of ECG data.

In the fourth chapter, the first research question was address and answered in detail. The number of attributes used to detect abnormalities of heart condition based on Arrhythmia symptom can be less than 5 attributes but still maintain high accuracy. Moreover, the number of attributes used for that study relying on 2 attributes only. This attribute was taken from autocorrelation result. Although the accuracy is high but the time used to auto-correlate the ECG data takes too long. As a result, the complexity procedure to classify the disease is huge. Moreover, the main goal of this study is to develop quick and effective intelligence computational mechanism in detecting heart disease symptom but the previous works in chapter 4 doesn't serve that purpose completely.

To reduce the complexity in classifying the disease and to get high sensitivity in detecting abnormalities as earlier as early stage, therefore those issues will be delivered completely in this chapter. A new mechanism to tackle with these issues is proposed: "Door-to-Door" algorithm is introduced to accurately and quickly detect significant peaks of heart cycle in 12 hours of ECG data and to discriminate obvious Paroxysmal stage of Atrial Fibrillation rhythms from Normal Sinus rhythms. In addition, a quantitative method using Artificial Neural Network (ANN), which discriminates unobvious Paroxysmal stage of Atrial Fibrillation rhythms from Normal Sinus rhythms, is investigated. In order to understand the detail of the proposed method, it will be explained in the next section.

5.1 Second Proposal: Automatic Detection of Paroxysmal Stage of Atrial Fibrillation Symptoms

In this section, an automatic mechanism to detect Paroxysmal stage of Atrial Fibrillation symptoms is proposed. Door-to-Door algorithm, which is a new algorithm with the capability to accurately extract heart cycle in ECG data, is proposed in this research. This algorithm captures normal heart cycle episodes even in a noise environment like base line wander. “Door-to-Door” is derived from a continuous process of finding the right heart cycle among a series of local maximums in the ECG data. The word “Door” refers to the highest local maximum detected among a group of data for local search.

Each “Door” represents the entrance or starting point for the deep investigation, which is performed to the surrounding data of the local maximum. Once the investigation is done to one local maximum, this “Door” will be used again as an exit to search for another “Door”. This process will be continued until the end of data. Therefore, “Door-to-Door” was named after this investigation process. The flowchart of the mechanism is described in Figure 5.1.

The five significant peaks in ECG data, which are P, Q, R, S and T peaks, are extracted using “Door-to-Door” algorithm and the number of detected peaks is counted. If the number of detected peaks is smaller than a certain threshold value (e.g. when a 12 hours ECG data is utilized, the threshold value is set to 46,000), the ECG data is regarded as an obvious Paroxysmal stage of Atrial Fibrillation rhythm or the advanced one. This is because it shows that the five significant peaks are not fully detected due to the unstable heart cycle. On the other hand, if the number of detected peaks is larger than the threshold value, the ECG data may be Normal Sinus

rhythm, but there is still the possibility that it is an unobvious Paroxysmal stage of Atrial Fibrillation rhythm. In this case, it must be discriminated from Normal Sinus rhythm.

In this research, therefore, the values of significant five peaks in each heart cycle are investigated using Artificial Neural Network (ANN). ANN outputs a numerical value to each heart cycle which indicates how much tendency of Atrial Fibrillation rhythm or Normal Sinus rhythm the heart cycle has. Based on the numerical values for whole heart cycles of the ECG data, the unobvious Paroxysmal stage of Atrial Fibrillation rhythm is detected. The details of Door-to-Door algorithm and ANN will be stated in Section 5.1.1 and 5.2.

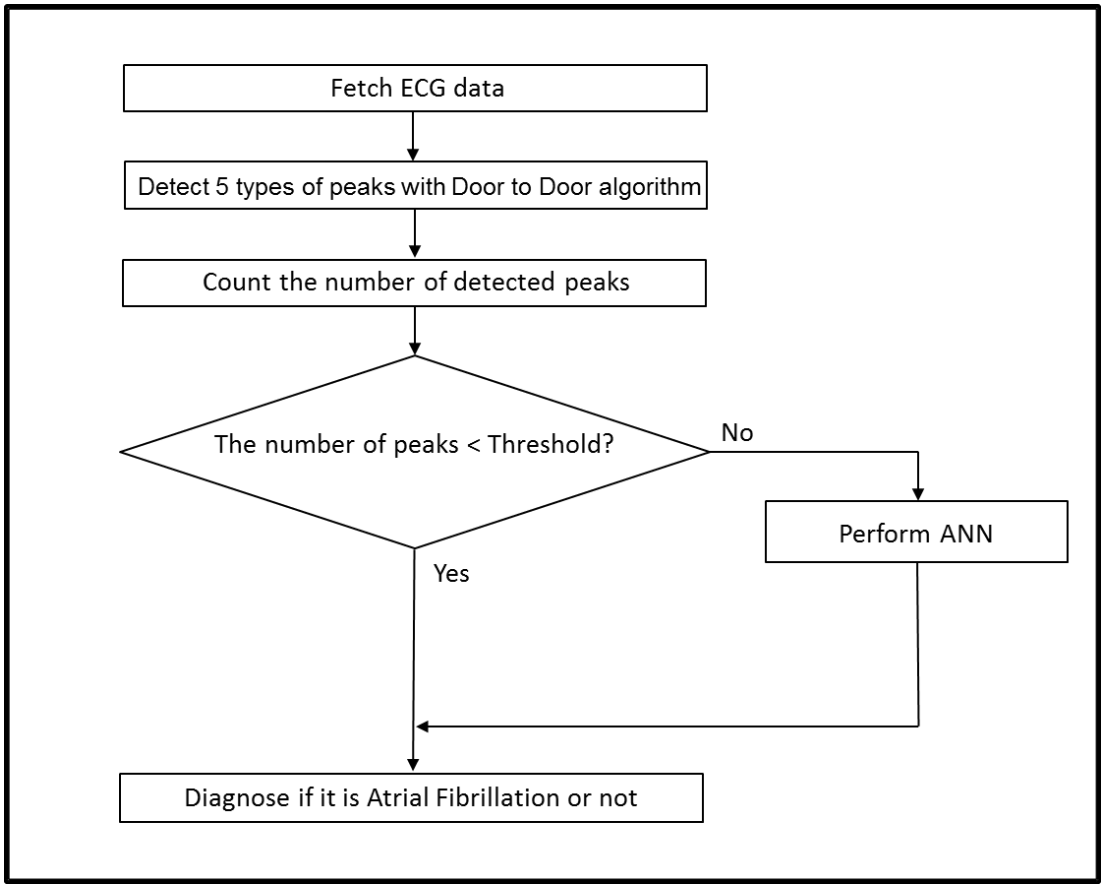

Figure 5.1: Proposed automatic mechanism to detect Paroxysmal stage of Atrial Fibrillation

5.1.1 Door-to-Door algorithm for five significant peaks detection

Figure 5.2 illustrates the procedure of Door-to-Door algorithm. R peak, Q peak, P peak, S peak and T peak are detected in a sequential order. When R peak is searched for, adaptive thresholds to the neighbouring data points are introduced. Here, let D_1, D_2, D_3 and D_4 be the relative horizontal distances from R peak to the direction of Q peak, from Q peak to the direction of P peak, from R peak to the direction of S peak and from S peak to the direction of T peak, respectively.

Note that the relative horizontal distance is expressed as data point unit. Q peak and S peak are detected as the minimum values in the range of D_1 and D_3, respectively. In the same way, P peak and T peak are detected as the maximum values in the range of D_2 and D_4, respectively. The values of D_1, D_2, D_3 and D_4 are determined by investigating the MIT-BIH Normal Sinus rhythm database [4]. The detailed procedure of each peak detection is described in the following sections.

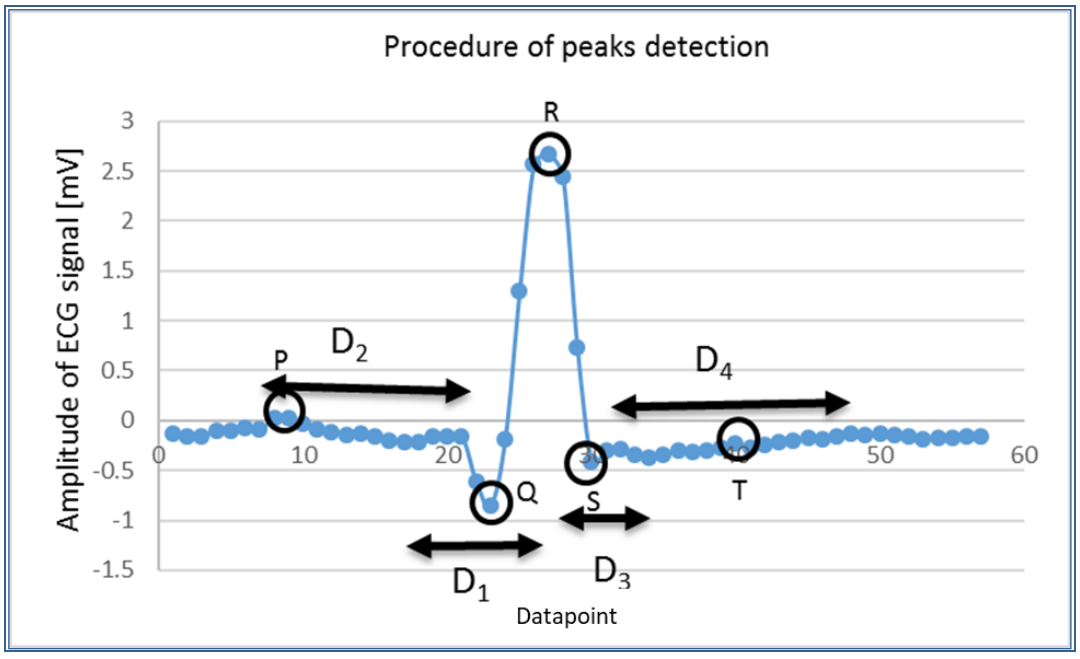

Figure 5.2: P, Q, S, T peaks detection mechanism based on R peak position

5.1.2 R Peak Detection with Adaptive Thresholds

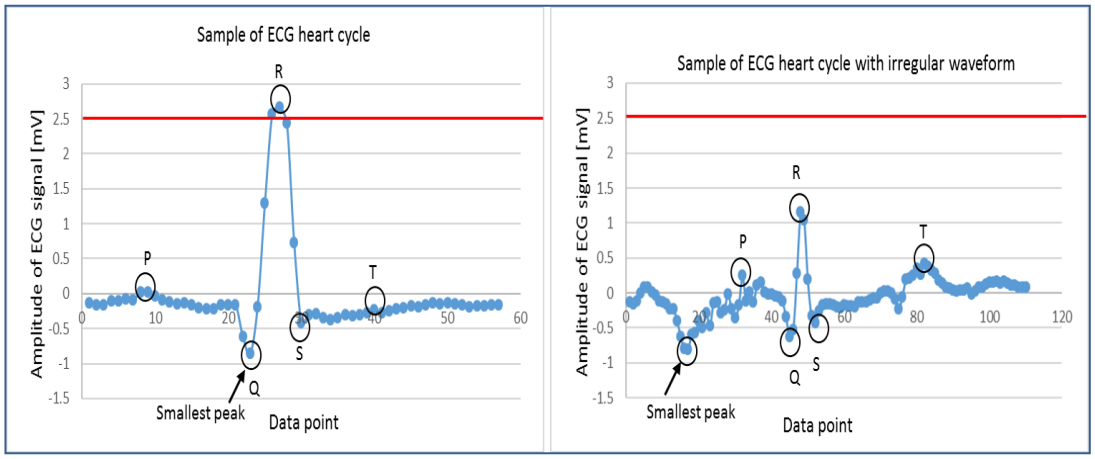

Figure 5.3: Different ECG heart cycle behaviours in Normal Sinus rhythm

R peak detection is a fundamental pre-requisite for the detection of other peaks and its detection accuracy is crucial for diagnosing the Paroxysmal stage of Atrial Fibrillation. R peak detection based on an absolute threshold has a significant weak point. Figure 5.3 depicts two different waveforms of Normal Sinus rhythm in ECG data. As seen from this comparison, it is not appropriate to use the absolute threshold value of 2.5mV since the value of R peak fluctuates so much. In contrast, R peak detection by Door-to-Door algorithm relies on the adaptive thresholds to overcome the difficulty.

In this study, three types of heart cycles in ECG data are considered to detect R peak in different ways. These three types of heart cycles correspond to normal rhythm scenario, irregular rhythm scenario and off-the-baseline scenario. The detail of each procedure to detect R peak is mentioned below.

(1) Detection of Local Maximum Point

When Door-to-Door algorithm starts, it searches for the local maximum point in the range of successive seven data points from the beginning of ECG data. If the local maximum point is not detected, the considered successive seven data points are shifted one data point onward and the local maximum point is searched for again. This procedure is repeated until the local maximum point is detected.

Let V_i denote the value of the detected local maximum point at *i*-th data point. Hereafter, the local maximum value is described as V_i for the sake of simplicity.

(2) Identification of R Peak

In this study, 6 adaptive threshold values have been introduced to accurately extract all heart cycles in ECG data. False Acceptance Rate (FAR) analysis has been done to determine the most optimal value to be used as each adaptive threshold value for this algorithm.

(a) Normal Rhythm

When ECG data shows a normal rhythm, the five significant peaks can easily be seen at the standard positions as shown in the figure on the left side of Figure 5.3. In this case, no abnormal waveforms such as noisy or unexpected signals are observed. However, there need some restrictions to V_i in order to regard the local maximum point as R peak. Here, three types of adaptive threshold values are introduced, which restrict the relative position of the local maximum point to the neighbouring data points. The values of five data points, V_i, V_{i-2}, V_{i-4}, V_{i+2} and V_{i+4}, are utilized to identify R peak as shown in Figure 5.4.

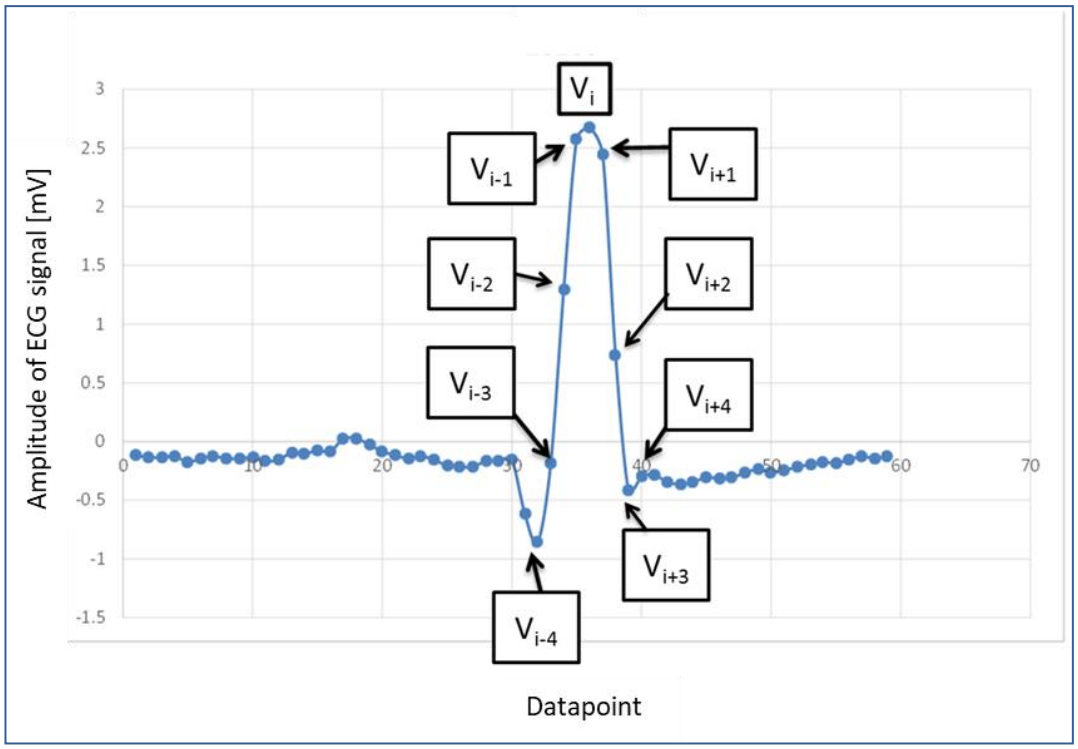

Figure 5.4: R peak detection in a normal rhythm scenario

When all the following conditions are fulfilled, V_i is identified as the value of R peak:

(1) $V_{i+2} < V_{i+1} < V_i$

(2) $V_{i-2} < V_{i-1} < V_i$

(3) $V_{th1} \leq V_i - V_{i+2}$

(4) $V_{th2} \leq V_i - V_{i-2}$

(5) $V_{i+k\ (k=1,15)} < V_i$

For this research, V_{th1} and V_{th2} are 0.35. $V_{i+k\ (k=1,15)}$ represents the range of data which is used to check if other local maximums close to the current local maximum exist or not. k=15 is determined by using False Acceptance Rate (FAR) analysis where the most optimal value to be used is selected. This process is very important to ensure the only the right heart cycle detection, not a noise signal.

(b) Irregular Rhythms

When ECG data shows an irregular rhythm, it is difficult to identify the local maximum point as R peak only by comparing the value to the neighbouring data points. This is because the values of data points around the baseline fluctuate so much, and thus it is necessary to confirm that V_i is only the local maximum value within a certain range of data points. Therefore, in addition to the value differences of V_i-V_{i+2} and V_i-V_{i-2}, V_i must be compared to the other data points values onward as shown in Figure 5.5

When all the following conditions are fulfilled, V_i is identified as the value of R peak:

(1) $V_{i+2} < V_{i+1} < V_i$

(2) $V_{i-2} < V_{i-1} < V_i$

(3) $V_{th3} \leq V_i - V_{i+2}$

(4) $V_{th4} \leq V_i - V_{i-2}$

(5) $V_{i+k\,(k=1,15)} < V_i$

For this research, V_{th3} is 0.35 and V_{th4} is 0.4. $V_{i+k\,(k=1,15)}$ represent the range of data to be check for any possible of other local maximum value exist that close to the current local maximum value location. k=15 is determined by using False Acceptance Rate (FAR) analysis where the most optimal value to be used is selected. This process is very important to ensure the only right heart cycle is detected and not a noise signal.

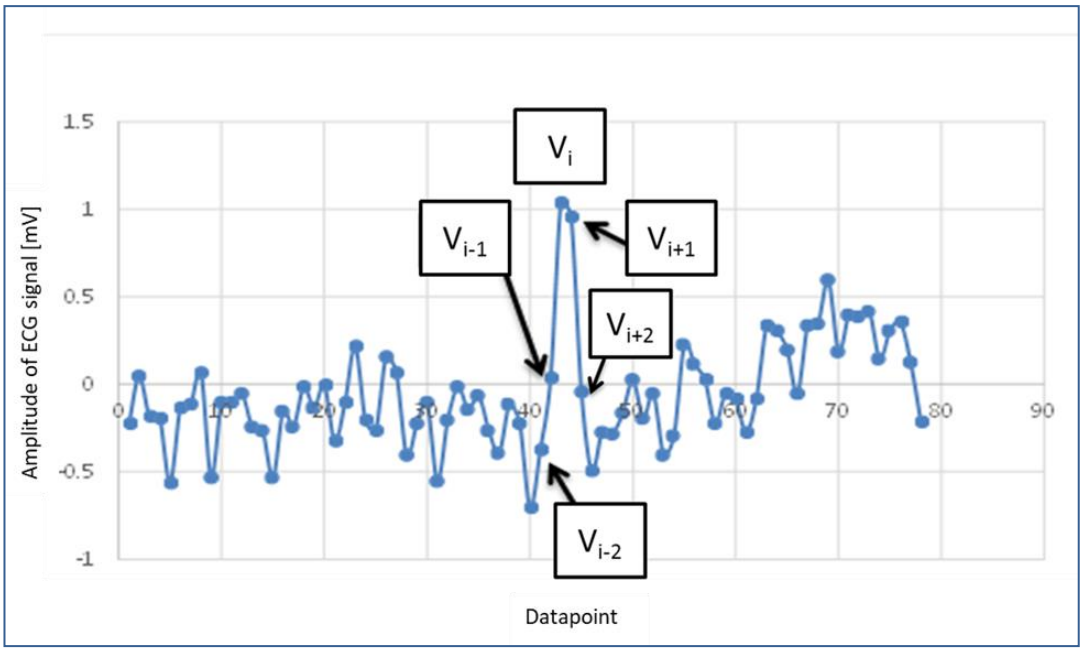

Figure 5.5: R peak detection in an irregular rhythm scenario

(C) Off-the-baseline Rhythms

When the heart rhythm is off the baseline of ECG data, it is also difficult to identify the local maximum point as R peak only by comparing the value to the neighbouring data points. This is because the absolute value of V_i is meaningless due to the offset of the baseline. To avoid the miss-detection of R peak, the values of adjacent data points around the local maximum point must be carefully investigated, considering the offset of the baseline. Therefore, in addition to the value differences of V_i-V_{i+2} and V_i-V_{i-2}, it must be confirmed that V_i is the local maximum value at the centre of the successive seven data points and each data point value decreases from V_i to V_{i-3} and from V_i to V_{i+3} as shown in Figure 5.6.

When all the following conditions are fulfilled, V_i is identified as the value of R peak:

(1) $V_{i+3} < V_{i+2} < V_{i+1} < V_i$

(2) $V_{i-3} < V_{i-2} < V_{i-1} < V_i$

(3) $V_{th5} \leq V_i - V_{i+2}$

(4) $V_{th6} \leq V_i - V_{i-2}$

(5) $0 < V_{i+k\ (k=1,6)}$

(6) $0 < V_{i-k\ (k=1,6)}$

For this research, V_{th5} and V_{th6} are 0.2. $V_{i+k\ (k=1,6)}$ and $V_{i-k\ (k=1,6)}$ represent the range of data to be check for any possible of other local maximum value exist that close to the current local maximum value location. k=6 is determined by using False Acceptance Rate (FAR) analysis where the most optimal value to be used is selected. This process is very important to ensure the only right heart cycle is detected and not a noise signal in the off-the-baseline rhythms.

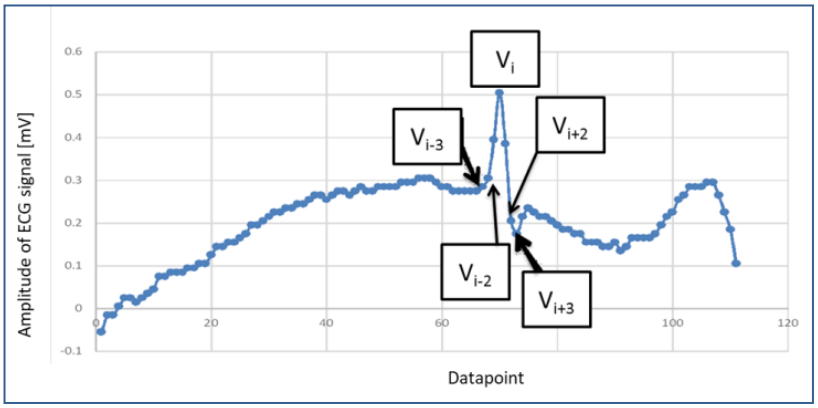

Figure 5.6: R peak detection in an off-the-baseline rhythm scenario

5.1.3 Q Peak Detection

After R peak in a heart cycle is detected, Q peak detection is performed. Q peak is searched for in the range of D_1. The data point, which has the minimum value in the range, is identified as Q peak as shown in Figure 5.2.

5.1.4 P Peak Detection

After Q peak in the heart cycle is detected, P peak detection is performed. P peak is searched for in the range of D_2. The data point which has the maximum value in the range is identified as P peak as shown in Figure 5.2.

5.1.5 S Peak Detection

After P peak in the heart cycle is detected, S peak detection is performed. S peak is searched for in the range of D_3. The data point which has the minimum value in the range is identified as S peak as shown in Figure 5.2.

5.1.6 T Peak Detection

After S peak in the heart cycle is detected, T peak detection is performed. T peak is searched for in the range of D_4. The data point which has the maximum value in the range is identified as T peak as shown in Figure 5.2.

5.2 Artificial Neural Network (ANN) Classifier

Artificial Neural Network (ANN) is one of the machine learning methods which imitate human's way of thinking to decide the most suitable solution to an issue. In this research, ANN is used to classify an ECG data into unobvious Paroxysmal stage of Atrial Fibrillation rhythm or Normal Sinus rhythm based on the values of P, Q, R, S and T peaks.

When ANN is performed with the values of P, Q, R, S and T peaks in a heart cycle as an input dataset, a numerical value is output. In this research, the output values of 0 and 1 indicate a typical Normal Sinus rhythm and an unobvious Paroxysmal stage of Atrial Fibrillation rhythm, respectively. Therefore, the middle value of 0.5 can be the border. Note that the output value is sometimes larger than 1 or smaller than 0 since it can be overestimated or underestimated by ANN, depending on the training dataset. When ANN is performed for the whole ECG data, the same number of output values as the heart cycles is obtained. Figure 5.7 and Figure 5.8 show the output values obtained from a Normal Sinus data and an unobvious Paroxysmal stage of Atrial Fibrillation data, respectively. In these figures, the horizontal axis indicates data point for each heart cycle, however, it is arranged in ascending order.

Let N_{ns} and N_{af} be the number of data points that has the output value less than 0.5 and the number of data points that has the output value more than 0.5. In Figure 5.7, N_{ns}/N_{af} is obviously larger than 1. On the other hand, N_{ns}/N_{af} is much smaller than 1 in Figure 5.8. It revealed that the unobvious Paroxysmal stage of Atrial Fibrillation rhythm can clearly be discriminated from the Normal Sinus rhythm by using ANN.

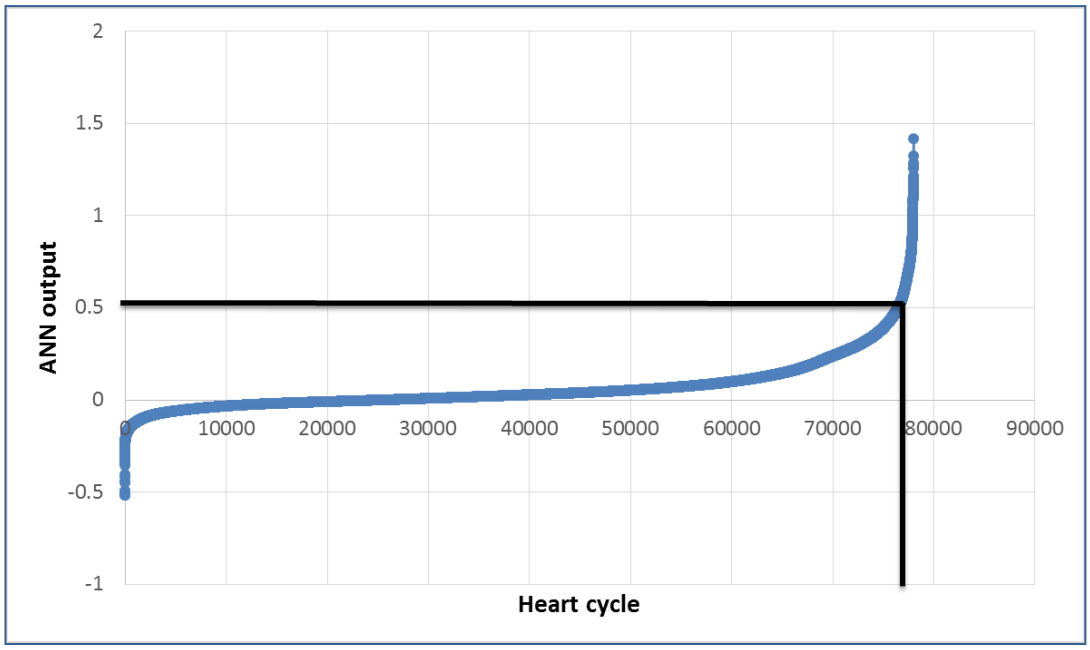

Figure 5.7: ANN output values obtained from an ECG data of a Normal Sinus patient

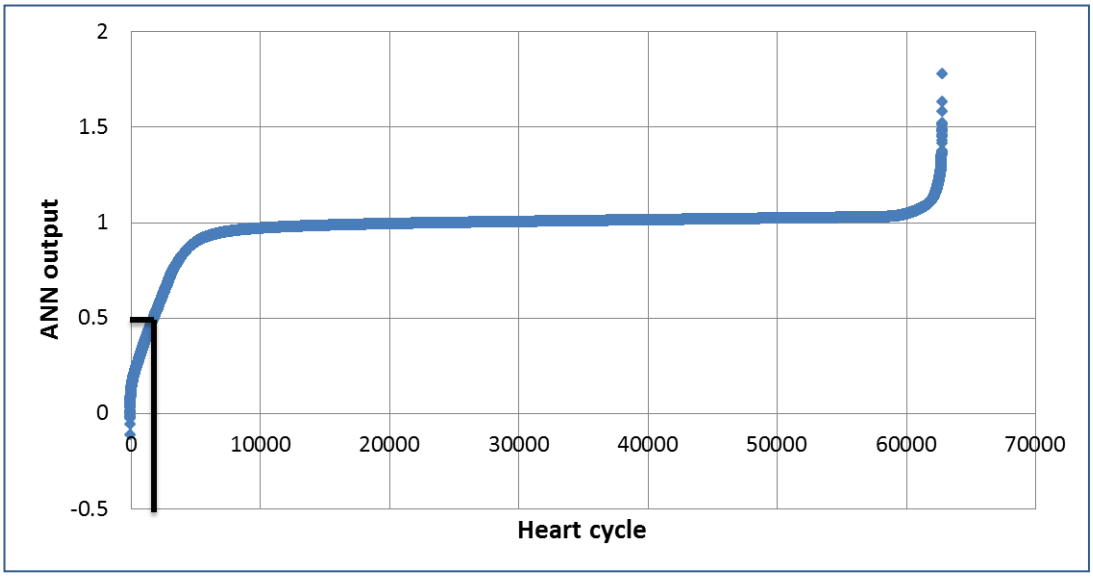

Figure 5.8: ANN output values obtained from an ECG data of an Atrial Fibrillation patient

5.3 Performance Evaluation

In this section, the performance of the proposed mechanism to detect five significant peaks and to discriminate Atrial Fibrillation data from Normal Sinus data is evaluated. In addition, the computational time of the proposed mechanism is also discussed. In the following subsections, the databases utilized for the evaluation, ANN setup, the performance metrics and the evaluation results will be presented.

5.3.1 Atrial Fibrillation Database for Evaluation

In this research, two types of databases, which are "MIT-BIH Normal Sinus" and "MIT-BIH Atrial Fibrillation", were utilized in order to evaluate the performance of the proposed mechanism. These two databases have been provided by PhysioNet [4]. PhysioNet is a research resource, providing a large number of recorded physiological data and the related open-source software.

17 patients' ECG data from the Normal Sinus database and 15 patients' ECG data from the Atrial Fibrillation database were selected for the evaluation. Each patient's data consists of a time series ECG data for 12 hours, which is one of the most important criteria to select the data. Note that the 15 patients' data from the Atrial Fibrillation database show the Paroxysmal stage of Atrial Fibrillation. Therefore, the characteristics of their ECG waveforms are quite similar to the ones of Normal Sinus. The proposed mechanism was performed to discriminate these 15 Paroxysmal stage of Atrial Fibrillation data from the 17 Normal Sinus data. The sampling frequency of all the ECG data utilized in this evaluation is 129 Hz.

The amplitude range of typical ECG signals is from - 5 mV to 5 mV. For a routine recording, most electro cardiographers agree that visual diagnostic accuracy can be maintained with a high frequency specification between 50 and 100 Hz [91]. In this experiment, the sampling frequency of 129 Hz is acceptable for measuring the consecutive R-peak of electrocardiogram. A small difference in sampling frequency does not influence the adaptive threshold setting and the detection performance itself. Therefore, in this research, ECG data with sampling frequency of 129 Hz, which is the original data configuration provided by Physionet, was utilized.

5.3.2 ANN Setup

As mentioned in Section 5.1, the obvious Paroxysmal stage of Atrial Fibrillation data were discriminated from Normal Sinus data by counting the number of detected five significant peaks using Door-to-Door algorithm. After that, ANN was performed to discriminate the unobvious Paroxysmal stage of Atrial Fibrillation data from the Normal Sinus data. The input parameters to ANN were the voltage values of P, Q, R, S and T peaks in each heart cycle. These five values constitute a dataset for a heart cycle. In this experiment, a conventional two-layered neural network with a single output neuron was used for ANN model development. As a result of network training, a decision function is chosen from the family of functions represented by the network architecture. This function family is defined by the complexity of the neural network: number of hidden layers, number of neurons in these layers, and topology of the network.

The decision function is determined by choosing the appropriate weights for the neural network. Optimal weights usually minimize an error function for the particular network architecture. The error function describes the deviation of predicted target values from the observed or desired values. In this research, class/non-class classification problem the target values is 1 for class (Atrial Fibrillation) and 0 for non-class (Normal Sinus). The number of hidden layers is 20. Training of neural network is performed on variations of ECG peak value based on Levenberg-Marquardt algorithms by trying to minimize an error function with 60% of dataset is allocated from the whole dataset. To avoid over fitting and under fitting, cross validation is used to find an earlier point of training by providing about 5% of validation data from the whole dataset. Finally, 35 % dataset is allocated to provide an unbiased evaluation of a final model fit on the training dataset.

As shown in Table 5.2, the data divisions for each set are stated in this table. The main aim of the separation into 5% validation and 35% testing is to create a balance prediction model between over fitting and under fitting. Since the prediction model may perfectly predict predefine training data, but it is very unlikely to perfectly predict any other data, a balance prediction model is required to support the case. Hence, the proposal is used to get the balance prediction model. In order to provide a balance prediction model, the average mean square errors of training, testing and validation values are used to evaluate. Mean square errors represent the average square difference between output and targets. The lowers values of mean square errors are better for prediction. The equation for Mean square errors (MSE) of the predictor are shown below.

$$MSE = \frac{1}{n}\sum_{i=0}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$\hat{Y}$ is a vector of n predictions, and Y is the vector of observed value of the variable being predicted.In order to get the best ratio, a series of experiments are conducted where 10 different ECG datasets with 3 different ratios for testing, validation and training data were investigated. Moreover, 270 times of training model was tested to find the best ratio to be used in this research. As the result, over fitting occurred when mean square errors values are very low for training data compared to the others two, while under fitting occurred when Mean Square Error value are very high for testing and validation data. Here, the investigation result was described in Table 5.1. As seen in Table 5.1, the best ratio is the one with validation 5 %, testing 35 % and training 60% comparatively.

Table 5.1: Average Mean Square error for training, testing and validation for different ratio

Training dataset (%)	Testing dataset (%)	Validation dataset (%)	Average Training Mean Square error	Average Testing Mean Square error	Average Validation Mean Square error
100	40	0	0.001109	0.001175	0.001222
60	15	5	0.001660	0.001874	0.001640
55	15	10	0.001696	0.001923	0.001983

Table 5.2: Data division for testing, training and validation for ANN experiment

Validation data (5%)	Testing data (35%)	Training data (60%)
3,094	21,662	37,135
2,458	15,732	29,499
2,828	19,781	33,940
3,065	22,428	36,775
3,204	22,428	38,445
2,954	20,678	35,449
2,704	18,928	32,449
2,798	19,591	33,584
2,827	19,791	33,927
3,187	22,312	38,249
3,439	24,072	41,266
3,107	21,751	37,288
3,251	22,754	39,007
2,611	18,276	31,330
2,340	16,379	28,078
3,069	21,488	36,829
3,543	24,804	42,521
3,142	21,996	37,708
2,867	20,067	34,400
2,791	19,535	33,489
3,289	23,023	39,467

5.4 Results of Performance Evaluation

The performance of Door-to-Door algorithm was evaluated in quantitative ways. The performance evaluation was divided into two: (1) How correctly five significant peaks in 12 hours ECG data of Normal Sinus is detected, (2) How correctly the obvious Paroxysmal stage of Atrial Fibrillation data is discriminated from the Normal Sinus data.

To evaluate the peaks detection performance, *"Sensitivity"* was selected as the evaluation metric. It indicates how correctly each peak can be detected. The detection sensitivity of the five peaks (expressed as *"Sensitivity"*) is defined as follows:

Sensitivity = True Positive / (True Positive + False Negative)

where,

True Positive: The number of actual peaks that are correctly detected as peaks.

False Negative: The number of actual peaks that are not detected as peaks.

Since there were a huge number of heart cycles in a 12 hours ECG data for 17 Normal Sinus data, 1000 heart cycles data randomly sampled from each Normal Sinus data were manually investigated. In other words, 1000 (heart cycles) x 5 (peaks) x 17 (data) = 85,000 (peaks) were validated. As the result, the detection sensitivity of the five significant peaks for the sampled data was 100%. This surprising accurate sensitivity concludes that Door-to-Door algorithm works very well to detect heart cycles of ECG data with adaptive thresholds as shown in Table 5.3.

Table 5.3: The sensitivity of heart cycle detection on MIT-BIH Normal Sinus database with Door-to-Door algorithm

Peak	P	Q	R	S	T
True Positive	100%	100%	100%	100%	100%
False Negative	0%	0%	0%	0%	0%
Sensitivity	100%	100%	100%	100%	100%

In this research, it was assumed that when the number of detected peaks (including P, Q, R, S and T peaks) is smaller than 46,000 in a 12 hours ECG data, it is identified as an obvious Paroxysmal stage of Atrial Fibrillation data. Therefore, 32 ECG data (15 Atrial Fibrillation data and 17 Normal Sinus data) were investigated if each data is an obvious Paroxysmal stage of Atrial Fibrillation data or not. As the result, 11 out of 32 ECG data were regarded as obvious Paroxysmal stage of Atrial Fibrillation data and actually they were correctly identified.

At this moment, the rest of 21 ECG data have not as yet identified with either unobvious Paroxysmal stage of Atrial Fibrillation data or Normal Sinus data since the number of detected peaks was larger than 46,000. Then, subsequently, ANN was performed to the 21 ECG data as stated in Section 5.2. The value of N_{ns}/N_{af} for each ECG data was obtained as the result as shown in Table 5.4.

Table 5.4: Nns/Naf for the 21 ECG data

ECG data number	N_{ns}/N_{af}
NS16265	1048.000
NS16272	246.075
NS16273	1177.458
NS16420	424.638
NS16483	2287.536
NS16539	808.328
NS16773	1039.019
NS16786	3997.143
NS16795	276.181
NS17453	826.896
NS18177	437.063
NS18184	738.845
NS19088	138.209
NS19090	599.195
NS19093	1455.75
NS19140	229.759
NS19830	119.320
AF04043	0.0566
AF05261	0.153
AF06995	0.049
AF08455	0.091

As clearly seen from Table 5.4, N_{ns}/N_{af} is much larger than 1 in the first 17 ECG data, and in the other 4 ECG data, N_{ns}/N_{af} is much smaller than 1. Therefore, the 17 ECG data and the 4 data were regarded as Normal Sinus data and unobvious Paroxysmal stage of Atrial Fibrillation data. As a matter of fact, these 21 ECG data were correctly identified.

To compare the discrimination performance of the proposed mechanism with other existing studies, *"Sensitivity"* and *"Specificity"* were selected as the evaluation metrics. The sensitivity and specificity are considered as the best paired performance metrics to evaluate the discrimination accuracy of Atrial Fibrillation from Normal Sinus [92]. The discrimination sensitivity of Atrial Fibrillation indicates the true positive rate in identifying Atrial Fibrillation. On the other hand, the discrimination specificity of Atrial Fibrillation indicates the true negative rate in identifying Atrial Fibrillation, which means the rate how correctly Normal Sinus is identified. The discrimination sensitivity and specificity (expressed as *"Sensitivity"* and *"Specificity"*, respectively) are defined as follows:

Sensitivity = True positives / (True positive + False negative)

Specificity = True negatives / (True negative + False positives)

where,

True positive: The number of Atrial Fibrillation data correctly identified as Atrial Fibrillation data.

False negative: The number of Atrial Fibrillation data incorrectly identified as Normal Sinus data.

False positive: The number of Normal Sinus data incorrectly identified as Atrial Fibrillation data.

True negative: The number of Normal Sinus data correctly identified as Normal Sinus data.

In this evaluation, to discriminate Paroxysmal stage of Atrial Fibrillation data from Normal Sinus data, two steps were taken, that is to say, the steps using Door-to-Door algorithm and ANN. Based on the definitions of sensitivity and specificity for the discrimination performance, the overall results for both are 100%. In order to ensure the proposed method are the most suitable model for classification, 3 different classification model have been selected to compare with, which are conventional Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbour (KNN) with 5 folds cross-validation. Since all those techniques are highly effective in data classification, hence, this result may describe the significance of the ANN and the study itself. The same dataset is used to test the performance of these 3 models. However, ANN has shown more advantages in predicting medical outcome compared to the other 3 classification model. The results are shown in Table 5.5.

Table 5.5: Overall performance evaluation results comparing this research with conventional SVM, Decision Tree and KNN using same dataset

ECG data number	SVM Accuracy (%)	Proposed method Accuracy (%)	Decision Tree Accuracy (%)	KNN Accuracy (%)
NS16265	99.9	100	99.9	99.9
NS16272	96.7	100	98.0	99.4
NS16273	99.8	100	99.9	99.9
NS16420	99.5	100	99.8	99.8
NS16483	99.9	100	99.9	99.9
NS16539	100	100	100	100
NS16773	99.8	100	99.9	99.9
NS16786	100	100	100	100
NS16795	98.3	100	99.2	99.4
NS17453	99.8	100	99.9	99.9
NS18177	99.5	100	99.7	99.7
NS18184	99.7	100	99.8	99.9
NS19088	96.2	100	97.5	97.1
NS19090	99.2	100	99.8	99.8

NS19093	99.6	100	99.9	99.9
NS19140	99.0	100	99.3	99.2
NS19830	97.8	100	98.0	98.4
AF04043	85.6	100	95.1	96.6
AF05261	58.4	100	86.4	90.7
AF06995	86.1	100	94.9	96.6
AF08455	61.8	100	89.9	94.7

Table 5.6 shows the results of the sensitivity and specificity of other studies with this research using the same database with duration more than 10 hours long. It is revealed that the proposed mechanism in this research outperformed the other studies.

Table 5.6: Overall performance evaluation result comparing this research with other studies. The bold font shows the results re-evaluated by Larbruru et al. [93]

Algorithm name	Sensitivity (%)	Specificity (%)
Slocum et al. [94]	**62.8**	**77.5**
Babaeizadeh et al.[95]	**87.3**	**95.5**
Tateno et al. [96]	**91.2**	**96.1**
Couceiro et al. [97]	**93.8**	**96.1**
Dash et al.[98]	94.4	95.1
Huang et al.[99]	96.1	98.1
Sarkar et al.[100]	97.5	99
Lee et al.[101]	98.2	97.7
Jiang et al.[102]	98.2	97.5
Shadnaz et al.[103]	97.0	97.1
Zhou et al. [104]	96.9	98.3
Carvalho et al. [105]	93.8	96.1
Huang et al. [99]	96.1	98.1
Lake et al. [106]	91	94
Lian et al. [107]	95.8	96.4
Dash et al. [108]	94.4	95.1
Tateno & Glass [96]	94.4	97.2
Proposed method	100	100

To complete research, the computational time of detecting heart cycle was evaluated. The computational time of Door-to-Door algorithm implemented on a personal computer (with Intel® Core™ i7 2.50 GHz, 16 GB RAM, 64 bit OS) was approximately 15 ms for a 30 second of ECG data. To the best of our knowledge, few research has mentioned the computational time of heart cycle detection for a long duration of ECG data. One research [109] stated that the computational time to detect heart cycle for a 30 second of ECG data was 40ms, which is more than two times longer than Door-to-Door algorithm.

The computational time of Door-to-Door algorithm for a 12 hour of ECG data varied from 30 seconds to 360 seconds. Even in the longest case, it takes only six minutes, which is clinically acceptable for diagnosis. Given this short computational time, the proposed mechanism can effectively be used in diagnosing a long duration of ECG data, especially to detect Paroxysmal stage of Atrial Fibrillation.

5.5 Conclusion

In this study, a novel and hybrid mechanism, which automatically detects Paroxysmal stage of Atrial Fibrillation symptom using Door-to-Door algorithm and ANN classifier, was proposed. To show the effectiveness of the proposed mechanism, the performance was thoroughly evaluated. The sensitivity of peaks detection in Normal Sinus data by Door-to-Door algorithm was 100% and the obvious Paroxysmal stage of Atrial Fibrillation data were perfectly discriminated from Normal Sinus data based on the number of detected heart cycles. By performing ANN, the overall unobvious Paroxysmal stage of Atrial Fibrillation data were discriminated from Normal Sinus data with the accuracy of 100%.

The comparison result between this research and other studies shows that the proposed mechanism outperformed in sensitivity and specificity of the discrimination performance. Moreover, the proposed mechanism holds strong advantages, that is to say, the computational cost and time are less than the other studies. It is concluded that this research can contribute to the medical field as one of the best technologies in diagnosing Paroxysmal stage of Atrial Fibrillation symptoms.

CHAPTER 6

DISCUSSION

6.1 Introduction

This chapter will address 3 issues related to the proposed study for detail discussion. An analytical discussion is done in this section to answer all the addressed issue. The issues are shown below:

The segmentation and analysis of the ECG data

i. The variability analysis with autocorrelation ECG data result.
ii. The significant point with the P, Q, R, S, T peak millivolt value focus for classification complex disease symptom
iii. The effectiveness of using quantitative approach to discriminate Paroxysmal stage of Atrial Fibrillation and Normal Sinus symptom

6.2 The segmentation and analysis of the ECG data

The biological signal analyses involve various human body's physiological process to extract relevant information to analyse specific condition of the body. This study was used by several interdisciplinary topics to provide cost effective diagnosis and treatment solution. However, these signals, in their rawest form do not provide a good insight and sufficient information for use to predict human's conditions.

Moreover, the ECG data represented are discrete in time and they usually contain two types of noise inherent in the signal which are baseline and power line interference. The mixing of heart symptom and noise interference may make the heart disease analysis more complex. However, with the advance computational power and the existence of complex analysis methodologies, possibilities to counter such issues properly are available. In this thesis, the two noble proposed methods are introduced to handle such issue. The significance in the identification of heart condition abnormalities based on statistical approach and data mining will be justified in this section. In addition, the details of the first proposed work on the usage of variability analysis of autocorrelation result to characterize the abnormalities will be explained here to signify the finding, the significance of using P, Q, R, S and T peak millivolt value will be analysed.

6.2.1 The variability analysis of Autocorrelation results from ECG data to detect abnormalities.

The variability analysis based on autocorrelation result is proposed in this study. This is mainly to detect heart condition abnormalities regardless of the symptom origin. Arrhythmia is selected as the symptom reference for abnormalities that occurred in the collected ECG data. Two types of attributes are used in autocorrelation result to classify the two symptoms precisely. Next, KNN classifier is used to classify the disease based on two selected attributes taken from the autocorrelation result.

In order to understand the significant point of using autocorrelation result and having variability analysis to identify the abnormalities of a heart condition, it is important to explain the detail motivation for this study. First of all, this study is trying to design a new computational mechanism to detect the various types of heart condition abnormalities regardless of the origin and then representing it in the simplest form which is the numerical value. The representation of numerical value is motivated by the idea that the preliminary diagnosis of long hour data should be simple and easy to understand.

Further diagnostic may require more complex analysis if the abnormalities symptom is detected accurately at the early stage. For that, simple representation of the abnormalities condition in numerical value may help reduce the complexity of diagnosis. The procedure may take shorter time and better view on the next possible treatment. It is expected in the future that if the mechanism performs well, the possibility of the technology to be used for the public can be realized. Therefore, designing computational mechanisms with high sensitivity level of detecting abnormalities of heart condition in the simplest form is vital.

Normally, abnormality identification of a heart condition can be done quantitatively by using heartbeat per minute. The key important information to measure is the measurement of the inconsistency of number of heartbeat per minute. This method relies on R-R intervals in the ECG data. If the duration of these intervals is constant, the heartbeat is regular. However, if the interval is fluctuated, abnormalities may occur. Nevertheless, this method comes with a series of challenges. First, the analysis of a heart condition based on R-R interval does not reflect the overall abnormalities of a heart disease symptom.

There are several types of symptom that may occur due to the chaotic behaviour of the signal and can only be seen at the interval between R-R. Moreover, it happens in random times in time series domain. Symptoms like Atrial fibrillation, Atrial Flutter, Multifocal atrial tachycardia, Wolff-Parkinson White syndrome, Long Q-T syndrome or much worse are Premature ventricular contractions where the disease symptoms overshadow the normal beats in the ECG data and are among the unexpected disease that occur not at the main peak of each cycle. Moreover, the practicality of monitoring heart disease symptom per minute each is complicated. The necessity to overcome such issue via early detection is compulsory. Therefore, the first proposed study is mainly trying to serve that issue constructively. In the first study, 14 types of symptom abnormalities have been covered as shown in Table 6.1.

Table 6.1: Detail of information for MIT-BIH Arrhythmia database

MIT-BIH Arrhythmias Database Detail information	
Heartbeat Type	**Total**
Normal rhythm	74607
Left bundle branch block	8069
Right bundle branch block	7250
Atrial premature contraction	2514
Premature ventricular contraction	7127
Paced beat	7020
Aberrated atrial premature beat	150
Ventricular flutter wave	472
Fusion of ventricular and normal beat	802
Non-conducted P-wave (Blocked APC)	193
Nodal (Junctional) escape beat	229
Fusion of paced and normal beat	986
Ventricular escape beat	106
Nodal (Junctional) escape beat	83
Atrial escape beat	16
Unclassified beat	35
Total	**109655**

Based on the list, there are plenty of abnormal heart disease symptoms available in the data. It is important for this study, to define the right definition of abnormalities of a heart condition. By definition, irregularity words is synonym to variability. In statistics, there is a method to measure the variability behaviour to the large group of data. It is call autocorrelation method. By utilizing autocorrelation method, the representation of the variety level of the overall data can be represented quantitatively and precisely.

As stated in chapter 4, it is hypothesized that the amounts of variability in Normal Sinus are much lower compared to Arrhythmia. Although Normal Sinus data may include some noise due to the unexpected patient behaviour or miss conducted procedure during ECG recording, those events are small compared to the nature of Arrhythmia symptom itself. This statement is supported by the fact that all arrhythmia symptoms are grounded by one common behaviour which is the irregularity behaviour in time series domain. In addition to that, irregularity is different from one symptom to another in Arrhythmia symptom. Since this study is grounded by empirical study and the assumption is founded by the variability of arrhythmia symptom are huge compared to Normal Sinus symptom, this hypothesis requires testing. It is important to have the preliminary test to ensure that the overall long hour performance' prediction feasible.

At the preliminary stage of the experiment, a very short duration of ECG data is selected randomly and tested with the autocorrelation function. The data consist of two different phenomena and are divided into two small segments for autocorrelation. The two phenomena are the strong fluctuation segment and the other one is the normal ECG symptom that behaves in time series domain as shown in Figure 6.1. The results of the two segments can be seen in Figure 6.2 and Figure 6.3

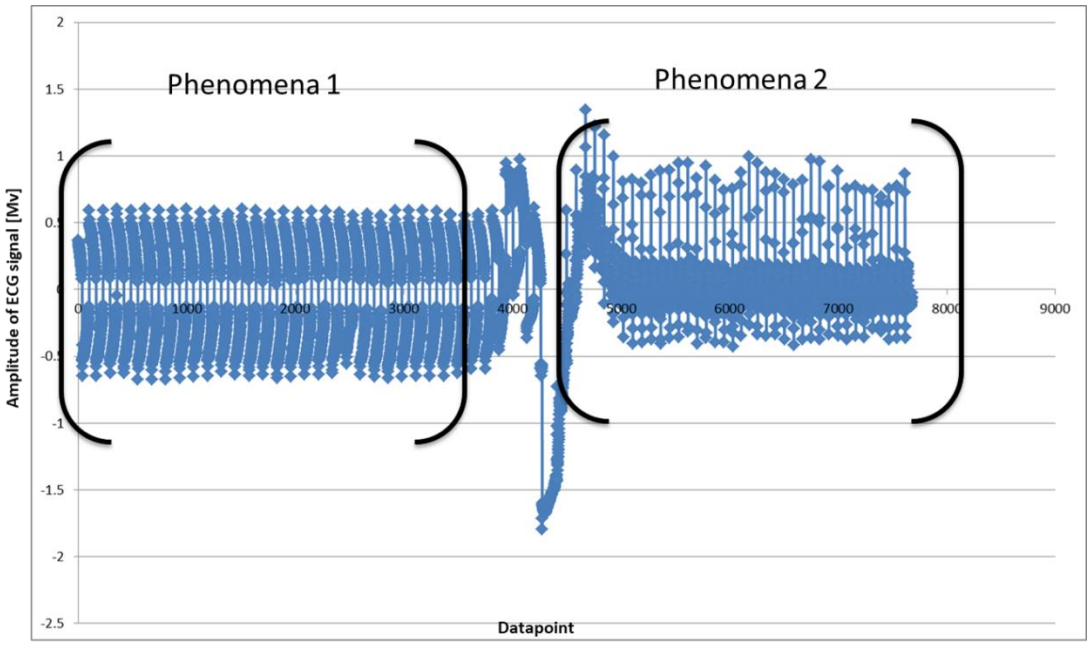

Figure 6.1: Two phenomena in ECG data with different behaviour

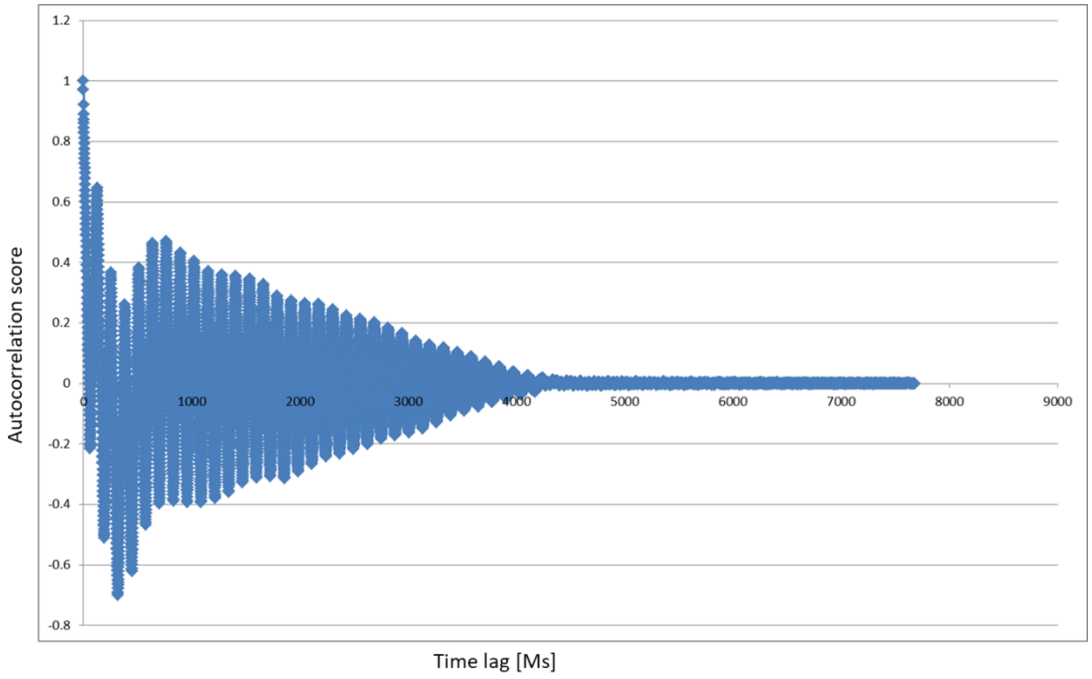

Figure 6.2: First phenomena after autocorrelation

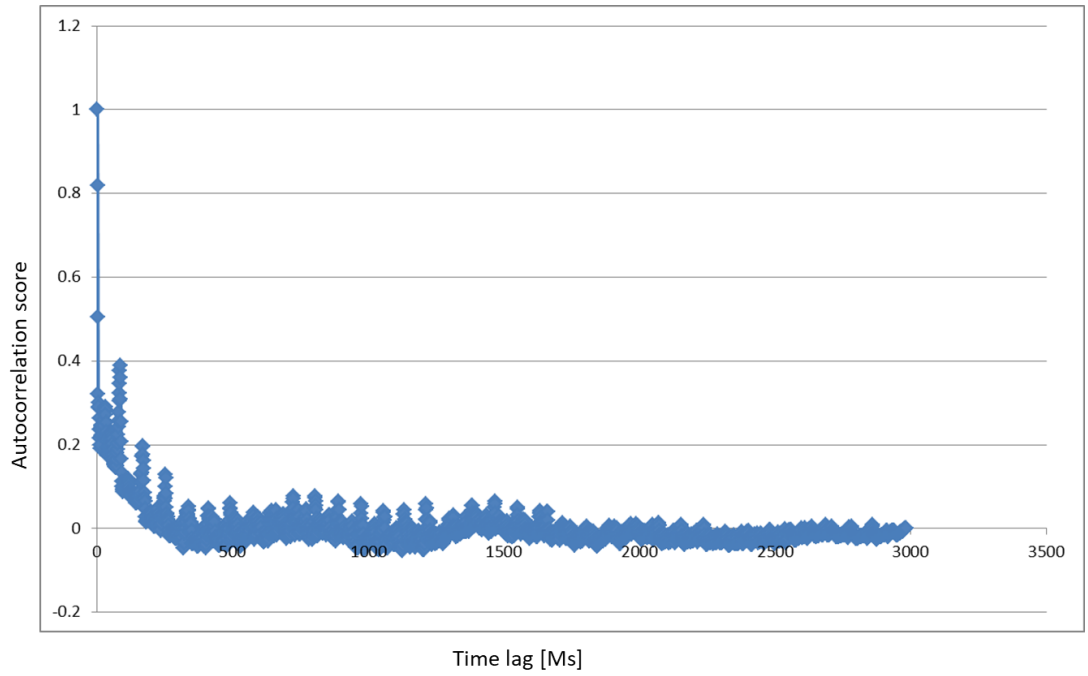

Figure 6.3: Second phenomena after autocorrelation

The main objective of this study is to characterize the abnormalities of a heart condition regardless of the origin of the symptom by using autocorrelation method and then utilize this characterize to classify the symptom by using classifier method. The result shows that the difference level of tendency in two different segments can be seen and identified clearly. The periodic slope for the first phenomena shows very high autocorrelation score compared to the second phenomena. Based on the study cases here, it is expected that the anomalies of symptom can be done even for long hour data. It is shows that the signal with strong uncertainty has the strong tendency to produce high autocorrelation score in time series domain. As a result, the evidence provided by the real experiment conducted with long hour data has shown that the hypothesis is true.

The result shows that the accuracy of detection of abnormalities can be achieve at a very high level which is 95%. Therefore, the significance of using autocorrelation result to characterize the anomalies symptom of the heart is well justified and supported. It is critical to emphasize that the medical data analysis is based on hypothesis generating and not hypothesis testing as conventional in classical statistical analysis as the utilization of the biomedical data is mainly to find the correlations of each data and not the causal evidence of it.

Moreover, the significance of using autocorrelation method to characterize anomalies symptom in ECG data is contributed by the point where biomedical data are commonly with high dimensionality characteristic. The effectiveness to extract the right information and not having multiple types of testing to discriminate the anomalies is a difficult task. However, this study has shown that the proposed method in anomalies detection in long duration of data is possible.

In addition to that, an additional experiment on 66 patient data from arrhythmia and Normal Sinus symptom has been conducted. This experiment has been tested with autocorrelation function mainly to study and identify the behaviour of autocorrelation result of both symptoms in different duration. In this experiment, 1 minutes,1 hour and 12 hours duration of data are used. In conclusion, a new insight has been identified and concluded from this experiment. The finding is summarize in Table 6.4.

Table 6.2: Summarization of finding

Time vs. Symptoms	Symptom
• The mechanism/procedure to identify the abnormalities in autocorrelation is relying on the sample size of the data. • The larger the sample size, the stronger characteristic of variability appear to arrhythmia data compared to Normal Sinus as shown in Figure 4.7 and Figure 4.8.	• This study shows that most arrhythmia data consist of high autocorrelation score compared to Normal Sinus. • Due to the nature of the Arrhythmia symptoms that weak appearance in time series, therefore, few arrhythmia data fall into low autocorrelation score and share same score with Normal Sinus data.

6.2.2 The significant point focus at P, Q, R, S, T peak millivolt value

In the second study, the focus is on designing a mechanism which has the capability to detect heart disease abnormalities as early as the Paroxysmal stage of Atrial Fibrillation. The key indicator point to measure the success of the proposed mechanism is based on how accurate the proposed mechanism can distinguish an early stage of Atrial Fibrillation from Normal Sinus symptom. As mentioned in chapter 5, the Normal Sinus symptom and Paroxysmal stage of Atrial fibrillation look similar in behaviour but different in various ways. Therefore, it is very crucial to define suitable attribute or parameter to distinguish the two symptoms accurately. In this study, 5 parameters are introduced which are the P,Q, R, S, and T peak millivolt value. They represent the characteristics of the two symptoms based on periodic cyclic signal in the ECG data. For this study, 12 hours duration data is used to measure the sensitivity of the proposed method to detect abnormalities of a heart condition based on atrial fibrillation symptom.

Theoretically, the ECG procedure is about recording the electrical activities of the heart on the chest. It will reflect the heart behaviour precisely based on the electrical flow of the heart. If abnormality is present, the electrical signal may show some disturbance like fluctuation or irregularity rhythm in the ECG data. This kind of event may occur at very specific segments in the ECG data in various ways. Selecting the right segment at the very specific event in the ECG data can help give new insight on the condition and the symptom. However, raw ECG data is a mix between heart cycle and noise. Therefore, specifying the right parameter to represent the disease symptom may trigger a lot of issues and discussions such as how relevant this parameter will signify the relationship between the parameter and the disease symptom itself. The complexity in defining such relationship in computational approach is addressed by many researchers [110].

By utilizing correlation coefficient method for instance to describe the relationship of one or more parameter with specific disease has been debating by many researchers around the world [1]. It is very important to emphasize here that the correlation can only represent the associations of the data with the other data and not the causal relationships. Therefore, the result represented by the computational approach can only be the supporting evidence in diagnosing the symptom. Consequently, it is a vital criterion to propose such parameters grounded by any relevant theory in biology to support computational evidence. Lack of support in biological perspective can decrease the value of the evidence provided by the computational approach. Hence, by only autonomous conventional method to diagnose the heart conditions at hospitals do not provide a new insight to tackle the big issue. For that, the necessity to look from different perspectives in the ECG data must be considered.

In this study, one of the main concerns is the proposal of the right parameter to detect an early stage of heart condition abnormalities by using the ECG data. The Paroxysmal stage of Atrial Fibrillation and Atrial Fibrillation symptom is chosen as a study case to test and signify the findings in detail. In this study, 5 parameters are introduced to present the symptom. P, Q, R, S and T peak millivolt value is used. The peak value of each data is extracted carefully from each heart cycle and further investigation is done after that. There are two main reasons to use this specific point in heart cycle to detect Atrial Fibrillation and abnormalities of heart condition. First, the empirical study is conducted mainly to find a new possible way to characterize the abnormalities of heart symptom by using raw ECG data. The simplification in representing the abnormalities in numerical ways is one of the main concerns for this study. It is expected that the irregularity of 5 peaks will be varied from one to another for the heart disease data compared to the Normal Sinus data.

Since the ECG data consist of series of numerical value in time series domain, it is important to utilize this data originally and study the tendency of the symptoms at very specific event. It is expected that the noise included in the ECG data may not overshadow the real symptom. By understanding the heart symptom based on numerical value at very specific segment may help to justify the possibility of this study to go beyond hypothesis. Without relying on any filtering method towards the ECG data, this study is trying to see how significant the proposed five parameters is in distinguishing the abnormalities of a heart condition data from a healthy patient data. Therefore, the preliminary study is conducted and Artificial Neural Network is used to classify the disease and to confirm the preliminary hypothesis. This preliminary experiment is conducted mainly to test if the 5 peaks can be used as a reference to represent the two symptoms.

In the preliminary stage of the experiment, 6 patient data is selected randomly from Normal Sinus database and Atrial Fibrillation database. Three patients data consisting of Normal Sinus symptom and another 3 patients data consisting of Atrial Fibrillation symptom are used. The duration of the data is 1 minutes. All the peak millivolt value is collected manually. By utilizing activation function in neural network, the possibility of classifying the two symptoms is shown in Figure 6.7. It has been decided in this procedure that the categorization for atrial fibrillation symptom must fall in value 1 and the categorization for Normal Sinus symptom should fall in value 0.

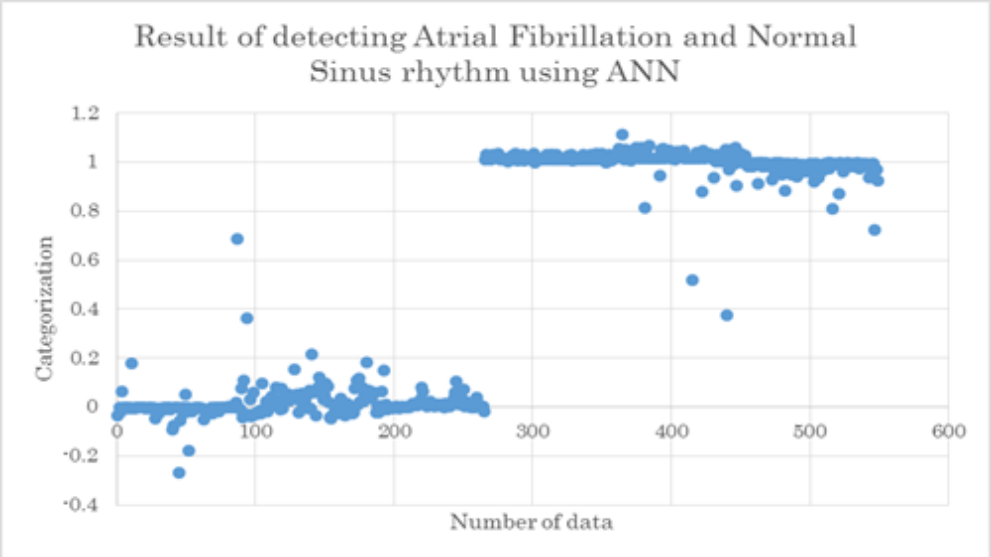

Figure 6.4: Result of detecting Atrial Fibrillation and Normal Sinus rhythm using ANN

Collecting the ECG peak value manually consumes a lot of time. Accordingly, it is important to design a very sensitive feature extraction mechanism to automatically detect normal heart cycle in the ECG data.

Hence, this study is proposing the Door-to-Door algorithm to serve that purpose. Presently, there is no existing works focusing on studying the tendency of millivolt value in ECG data and distinguish abnormalities of heart condition from healthy heart condition using artificial neural network. Based on the preliminary result with the short duration ECG data, the evidence provides some possibilities to distinguish the two symptoms.

However, it only responds to small numbers of sample. The result may not represent perfectly the overall picture of the study if it is tested to a large sample dataset. Therefore, in chapter 5, the study is trying to prove the possibility of dealing with a large dataset. Based on the result of long hour data, the evidence signify the finding constructively. Therefore, the first reason is justified clearly here.

The second justification in the usage of the P,Q,R,S and T peak value is inspired by a few related works which provide strong evidence in justifying the atrial fibrillation symptom detection based on the P-wave characteristic [111-112]. The increasing detection rate performance at detecting the atrial fibrillation is proven when they are focusing on the P wave segment [111-112]. Since this study is trying to detect as early stage as Paroxysmal stage of atrial fibrillation and arrhythmia symptom, this information is a crucial reference to decide which parameter is suitable for the proposed method. Moreover, the atrial fibrillation and the atrial flutter are symptoms under the Arrhythmia category.

The Atrial Fibrillation symptom and atrial flutter symptom normally consist of many types of P wave behaviour. It is expected that the P peak value may vary strongly from one to another based on these symptoms. Thus, by selecting the P peak value in each heart cycle, a new insight on the normal heart condition and abnormalities of heart symptom based on distribution of P peak value can be made.

However, the significance of using the other peak value in the ECG segment is because of the abnormality of a heart condition based on arrhythmia symptom is covered from the atrial segment to the ventricular segment in the heart. This segment is reflected by the other peak behaviour like Q, R, S and T peak value.

As a result, the utilization of those peaks may help detect the abnormalities of the heart condition in general and are highly sensitive in the detection level to detect atrial fibrillation regardless of the stage of the symptom. The use of artificial neural network may help distinguish those characteristics precisely as shown in chapter 5.

6.3 Clarification of using activation function to discriminate Paroxysmal stage of Atrial Fibrillation and Normal Sinus symptom

In the second study of this thesis, the detection of Paroxysmal stage of atrial fibrillation is relying on the activation function with the neural network. This method is used to describe the tendency between Normal Sinus and Paroxysmal stage of atrial fibrillation in the micro perspective. Since the similarity of behaviour between the two symptoms is huge and the Paroxysmal stage of the atrial fibrillation rarely appear in time series, the possibilities to detect this small tendency is difficult.

The small group of data may fall into categories of outliers' data if the situation is not handled properly. Therefore, the quantification method is proposed to describe the tendency precisely and classify the two symptoms accurately as described in detail in chapter 5. Utilizing the activation functions with neural network may help in the visualization of the output of the neural network and provide a good insight on the data. Moreover, it may help give the best way to predict the probability of the output. Since the probability of the disease can only be

between range 0 and 1, even a small difference in the disease symptom can be described by the activation function quantitatively. Moreover, with the capability of nonlinear function, it may help to learn complex functional mapping from the data and deliver the right outcome accurately. Therefore, based on the proposed method, the significance of this proposed method has been described and explained here.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Introduction

This chapter concludes the dissertation and figure out open research direction for possible future works.

7.2 Conclusion

This dissertation proposed an efficient, quickly and highly sensitive computational intelligence to accurately detect abnormalities of a heart condition based on Arrhythmia symptom. The proposed mechanism consists of two primary components, namely the efficient Arrhythmia detection using autocorrelation and statistical approach, and the hybrid mechanism to detect Paroxysmal stage of Atrial fibrillation using adaptive threshold-based algorithm with Artificial Neural network.

In the first proposed mechanism, a feasibility study on the effectiveness of using autocorrelation function and KNN to detect heart condition abnormalities regardless of the origin of the symptom based on Arrhythmia symptom is presented. A variability analysis based on periodic cycle in autocorrelation result is done. It is based on two parameters at first periodic slope of autocorrelation result. In order to discriminate the two symptoms, KNN classifier is used. The effectiveness of the proposed method is evaluated based on 3 performance evaluations metric which are accuracy, sensitivity and specificity. From the result, the overall accuracy is 95.5%

with the sensitivity level of detecting Arrhythmia at 97.5% and the specificity of detecting Normal Sinus at 88.8%.

Then, the effectiveness of this study is compared with other studies. The comparative studies have shown that the proposed mechanisms are robust, flexible and less complex in detecting abnormality symptoms like Arrhythmia. In this study, 17 different types of supervised machine learning classifiers are used to compare with the proposed classifier. It is proven that fine KNN has outperforms other classifiers for a 12 hours duration segment. 14 types of symptoms are covered in this study and there is no dependency towards any specific characteristic and feature segment in the ECG data to identify each Arrhythmia symptom. With a very minimal number of attribute used to characterize the abnormalities of heart condition based on arrhythmia symptom, it is justified of the robustness of the proposed method in discriminating heart condition abnormalities.

In addition to that, the first research question is completely answered in this study. The key important point to highlight here is that a small number of attribute can provide high detection accuracy if the attribute can represent the overall characteristic of the symptom and it should be proven in time. It is confirmed based on this study that the time length of data is consider the biggest influential factor towards the overall classification performance. Therefore, the longer the time duration of ECG data is used, the higher accuracy the classifier can be achieved.

Although the effectiveness of using minimal number of attributes to detect arrhythmia symptom is shown here, it requires a lengthy period to complete the process. Therefore, a mechanism that can serve the overall objective of this study to design an effective, quick and highly sensitive computational intelligence approach to detect abnormalities of heart condition as early stage as Paroxysmal stage of atrial fibrillation is proposed.

In the second proposed mechanism, an empirical study is done to show the effectiveness of using a new algorithm called “Door-to-Door” algorithm and ANN to detect abnormalities of a heart condition as early stage as Paroxysmal stage of Atrial Fibrillation. The motivation of this study is to deliver completely the main objective of this thesis that is to design a very sensitive mechanism to detect the abnormalities as early the symptoms can be detected. Therefore, a novel and hybrid mechanism, which automatically detects an early stage of Atrial fibrillation using Door-to-Door algorithm and ANN classifier, is proposed.

First, the effectiveness of the proposed mechanism is shown in this study based on the sensitivity of peaks detection in Normal Sinus data by Door-to-Door algorithm as 100% and the obvious early stage of Atrial Fibrillation data is perfectly discriminated from the Normal Sinus data based on the number of detected heart cycles. Due to the fact that the Paroxysmal stage of the atrial fibrillation and Normal Sinus look similar in the ECG data and rarely appear in time, the complexity to identify such symptom is a crucial task. By performing ANN and quantitative evaluation based on the error output, the overall unobvious early stages of Atrial Fibrillation data are discriminated from Normal Sinus data with the accuracy of 100%.

The comparison result between this research and other studies show that the proposed mechanism outperformed in sensitivity and specificity of the discrimination performance. Moreover, the proposed mechanism holds strong advantages, that is to say, the computational cost and time are less than the other studies. Moreover, the second research question is address and justified with strong evidence. Based on the finding, in order to determine the most suitable machine learning approach to detect heart disease symptom, there are two criteria that must be considered so that classification can be made. The first is the length of the duration of data and second is the size of the dataset used for classification. For the small duration of time and the small number of data, the decision tree is considered to be the best machine learning approach to classify the heart disease based on this study.

However, it is different for long hour duration of ECG data. For long hour duration and large number of data, the Artificial Neural Network and K-nearest Neighbour is considered the best machine learning approach for that kind of specification to detect a heart disease. Moreover, a constructive conclusion has been made based on computational modelling for heart disease detection. It is represented in Figure 7.1 until Figure 7.4. Figure 7.1, Figure 7.2 and Figure 7.3 represent the impact of imbalance design in computational modelling based on 3 main components. They are the data mining approach, the suitability of attributes to represent disease symptom and the number of attributes used for classification. Figure 7.1 represents the impact of imbalance design for computational modelling where the model only focuses on the suitability of attributes to represent disease symptom and right data mining approach. Figure 7.2 represents the impact of imbalance design for computational modelling where the model only focuses on the suitability of attributes to represent the disease symptom and the right number of attributes. Figure 7.3

represents the imbalance computational modelling where the model only focuses on the right number of attributes and the right data mining approach. There is always a trade-off in achieving a balanced computational modelling while having quick, accurate and efficient mechanism to detect heart disease symptom. However, this study shows that an ideal model can be achieved as described in detail in chapter 4 and chapter 5. The ideal models are shown in Figure 7.4.

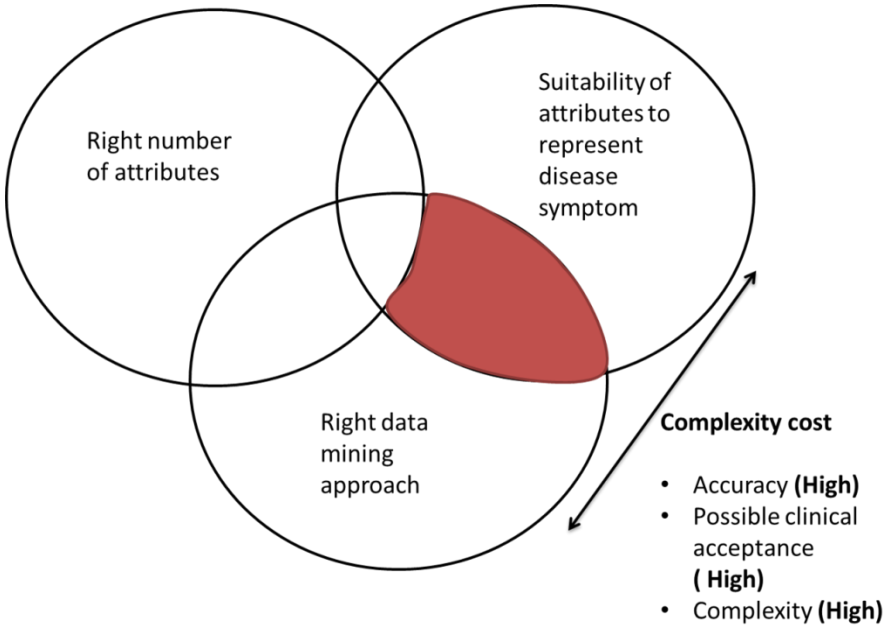

Figure 7.1: Complexity cost in computational modelling for heart disease detection

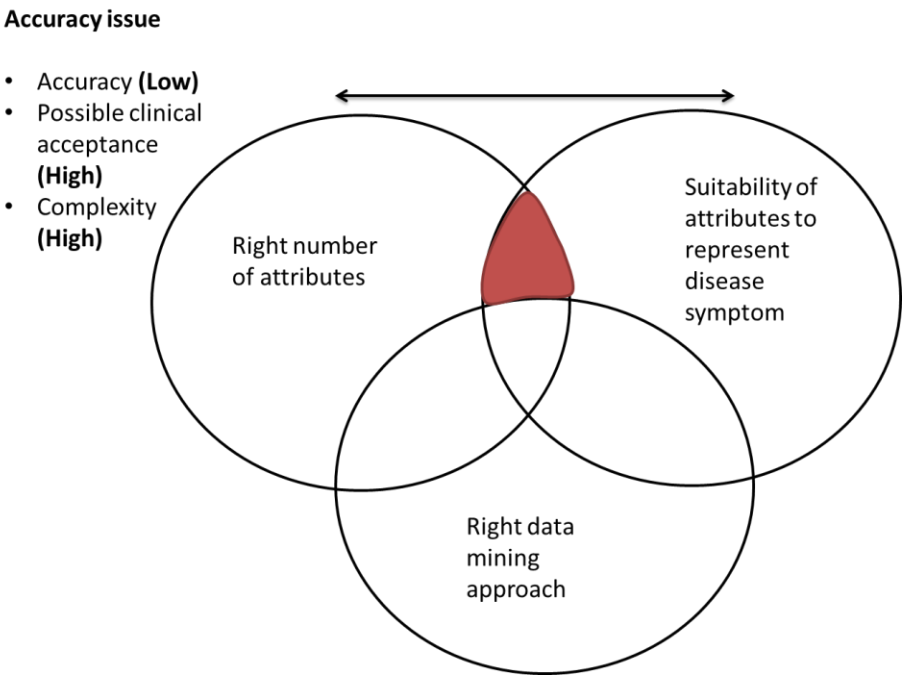

Figure 7.2: Accuracy issue in computational modelling for heart disease detection

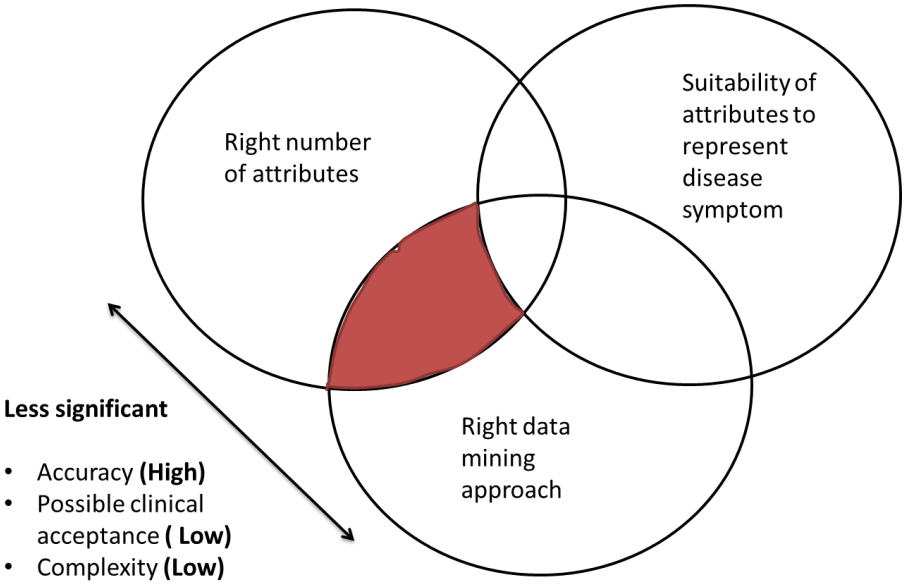

Figure 7.3: Significant issue in computational modelling for heart disease detection

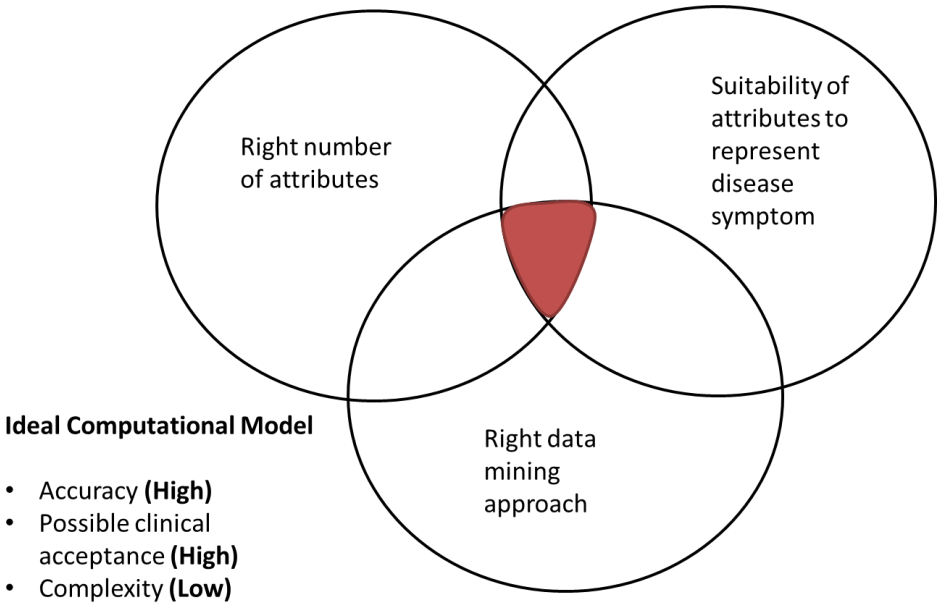

Figure 7.4: Ideal computational modelling for heart disease detection

Although noise filtering process in signal data is considered an important process in the signal processing analysis and without having such kind of procedure may strongly affect the accuracy of heart disease detection, this study has proven otherwise. It is proven that without having such kind of procedure, by relying on an effective computational model to detect abnormalities of heart condition and robust machine learning approach, such issue can be dealt with constructively. These empirical studies have signified the finding based on the proposed framework, proposed computational model and technique that detecting abnormalities of heart condition with high accuracy is possible.

It is important to highlight that the result produced by the proposed method in this study should be interpreted and considered as an association relationship to determine the condition of the heart and not the causal relationship of it. Therefore, these studies are trying to provide a new insight and possibility in analysing the heart disease based on the proposed computational model to detect any abnormalities regardless of the origin of the symptom and at any stage of the symptom.

As a summary, the overall performance and comparison of some related works are shown in Table 7.1.

Table 7.1: Comparative study of computational heart disease detection

Author	Technique	No. of Attributes	Accuracy (%)
Resul et al. [47],2009	Neural Network	13	89.01
Anbarasi et al.[48],2010	Genetic with Decision Tree	6	99.2
	Genetic with Naïve Bayes		96.5
	Genetic with Classification via Clustering		88.3
Rajkumar et al.[49],2009	Naïve Bayes	17	52.33
	Decision Tree		52
	KNN		45.67
Kumari et al.[50],2011	Decision Tree	14	79.05
	Artificial Neural Network		80.06
	Support Vector		84.12
	RIPPER		81.08
Sundar,et al.[51],2018	WAC	15	84
	Naïve Bayes		78
Chaitrali et al. [52],2012	Artificial Neural Network	13	99.25
		15	99.9
John, et. al [53],2012	Naïve Bayes	14	85.18
	Multilayer		78.88
	J48		85.18
	KNN		85.55
Shouman et al. [54],2012	K-Nearest Neighbour	13	97.4
Nidhi et al. [55],2012	Naïve Bayes	15	90.74
		13	94.44
		6	96.5
	Decision Trees	15	99.62
		13	96.66
		6	99.2
	Neural Network	15	96.5
		13	99.2
		6	88.3
Pethalakshmi et al. [56],2012	Fuzzy Decision Tree	13	90.06
	Fuzzy Naïve Bayes		89.62
	Fuzzy Neural Network		91.09
	Fuzzy K-means		99.49
Abhishek et al.[57],2013	J48	15	95.56
	J48		94.85
	Naïve Bayes		92.42
Chitra et al [58],2013	Artificial Neural Network	14	85
	K-Means Clustering		88
	Fuzzy C Means Clustering		92
Dessai [59],2013	PNN	14	94.6
	Decision Tree		84.2
	Naïve Bayes		84
	BNN		84.6
Patel el al.[60],2013	Decision Tree	14	99.2
	Naïve Bayes		96.5
	Classification Clustering		88.3
Vikas and Pal [61],2013	CART Classification	11	84.49
	Naïve Bayes		96.53

Methaila et al. [62],2014	Decision Tree	15	99.2
	Classification via Clustering		88.3
Wisaeng [63],2014	K-Nearest Neighbor	14	93
		8	90
Waghulde et al.[64],2014	Neural Network & Genetic Algorithm	13	98
Rupali et al. [65],2014	Classification using Naïve Bayes	14	78
	Classification using Laplace Smoothing		86
Venkatalakshmi et al. [66],2014	Naïve Bayes	13	85.03
	Decision Tree		84.01
Jarad [67],2015	Naïve Bayes	14	85.03
	Decision Tree		52
	KNN		45.67
D`Souza [68],2015	ANN	14	79.38
	K-Mean Clustering		63.299
Baiju and Janet [69],2015	Naïve Bayesian classification Technique	13	81
Adbar et al. [70],2015	C 5.0	14	93.02
	NN		89.4
	SVM		86.05
	KNN		80.23
Kaur and Kaut [71],2015	SVM Classifier with Genetic Algorithm	12	95
Swati et al. [72],2015	Naïve Bayes	13	84
	KNN		76
Patel et al. [73],2016	J48	13	56.76
Rajalakshmi et al.[74],2016	K-Means clustering	14	93.89
	WAC		92.84
Suganya et al. [75],2016	CART Classifier	14	83
Karthikeyan et al. [76],2017	Deep Belief Network	16	90
Wadhawan [77],2017	K-Means Using Apriori Algorithm	7	74
1st Proposed method	Autocorrelation function and KNN	2	95.5
2nd Proposed method	ANN	5	100

As concretely presented in this dissertation, the proposed mechanism not only accurately detects the abnormalities of a heart condition based on Arrhythmia symptom but also reduces the complexity in identifying the symptom with small and simple parameter. The performance of this works is confirmed by using real patient data and the quality of the evidence is supported by theoretical approach in biology

for heart disease diagnosis and computational approach itself. The difficulty in identification of the very complex symptom like Paroxysmal stage of Atrial fibrillation is dealt with via activation functions in the neural network. The visualization of the prediction output has simplified the procedure in classifying the disease computationally and statistically. Therefore, it is concluded that this research may lead to one of the best technologies in diagnosing abnormalities of heart condition as early stage as Paroxysmal stage of Atrial fibrillation.

7.3 Future work

This section presents an open research direction for the future works. Several of them are the extension issues concretely discussed in this dissertation.

7.3.1 The suitable Data transformation, discretization and reduction method for the prediction

In the first proposed method model, the process of converting raw ECG data into a series of correlation coefficient score with autocorrelation method requires long hour period. In this study, each 12 hours of ECG data needs more than 5 hours long to complete the autocorrelation process. Although this work is trying to investigate the variability level of the larger group of ECG data and then classify the two different symptoms, the first process before classification takes too long. It is less effective in time for real application. Therefore, the necessity to find the right method to

transform such data faster may help in providing a new perspective in analyzing ECG data constructively.

7.3.2 Interpretability in prediction result

In many related works in this domain, there are huge difficulties in understanding such machine learning outcome. It is suggested that it would be better if the level of understanding and insight provided by the model can be simpler to understand and interpret. The improvement in visualizing the outcome to serve different purposes in this study may help simplify the process in identifying or detecting complex disease more accurately and precisely using computational approach. Even though this study is trying to represent the result in the simplest way by using sigmoid function, this method should be tested unto various kinds of disease. This process is very important to measure how reliable the proposed method is on order to handle more complex behaviour in the heart disease. The extension of this study is one of the top priorities for the future work.

7.3.3 Prediction accuracy is low with reduced number of attribute

Based on the first work of this dissertation, it is proven that the capability to utilize such a small number of attributes to reach very high accuracy in heart disease detection is possible. However, there is a trade-off of between having a certain amount number of attributes and complexities in the classification process. Therefore, a deep study on this issue is needed to find the most optimal point in reducing the complexities in the procedure while providing effective disease detection method through computational approach.

REFERENCES

[1] D. S. Celermajer, C. K. Chow, E. Marijon, N. M. Anstey, and K. S. Woo, "Cardiovascular disease in the developing world: Prevalences, patterns, and the potential of early disease detection," *J. Am. Coll. Cardiol.*, vol. 60, no. 14, pp. 1207–1216, 2012.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, pp. 37–54, 1996.

[3] A. Petrenas, V. Marozas, and L. Sörnmo, "Low-complexity detection of atrial fibrillation in continuous long-term monitoring," *Comput. Biol. Med.*, vol. 65, pp. 184–191, 2015.

[4] PhysioNet. Detrended Fluctuation Analysis (DFA). http://www.physionet.org/ (accessed August 31, 2018).

[5] Oded Maimon and Lior Rokach. 2005. Data Mining and Knowledge Discovery Handbook. Springer-Verlag, Berlin, Heidelberg

[6] Shadabi F, Sharma D. Artificial Intelligence and Data Mining Techniques in Medicine – Success Stories. Int Conf BioMed Eng Inf 2008.

[7] Zhou Z-G, Liu F, Li L-L, Jiao L-C, Zhou Z-J, Yang J-B, et al. A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer. Knowledge-Based Systems; 2015[in press].

[8] Ge L, Kristensen AR, Mourits MC, Huirne RB. A new decision support framework for managing foot-and-mouth disease epidemics. Ann Oper Res 2014;219(1):49–62.

[9] Raghu A, Praveen D, Peiris D, Tarassenko L, Clifford G. Lessons from the Evaluation of a Clinical Decision Support Tool for Cardiovascular Disease

Risk Management in Rural India, Technologies for Development. Part V Springer International Publishing;2015 199–209.

[10] Gil D, Soriano A, Ruiz D, Montejo CA. Embedded systems for diagnosing dysfunctions in the lower urinary tract. Proceedings of the 22nd Annual ACM Symposium on Applied Computing (SAC'07); 2007.

[11] Waring S, Sharland M, Bianco J, Boyce M, Quinlan S. PS2–8: Development and Implementation of Clinical Decision Support Tools in Epic to Standardize Dementia Diagnosis and Care at Essentia Health. Clin Med Res 2014;12(1–2):88.

[12] Bourouis A, Feham M, Hossain MA, Zhang L. An intelligent mobile based decision support system for retinal disease diagnosis. Decis Support Syst 2014;59:341–50.

[13] Rosenblum H, Radcliffe N. Case-based approach to managing angle closure glaucoma with anterior segment imaging. Can J Ophthalmol 2014;49(6):512–8.

[14] Siva A, Lampl C. Case-Based Diagnosis and Management of Headache Disorders.Springer International Publishing; 2015.

[15] Horn M, Glauche I, Müller MC, Hehlmann R, Hochhaus A, Loeffler M, et al. Model-based decision rules reduce the risk of molecular relapse after cessation of tyrosine kinase inhibitor therapy in chronic myeloid leukemia. Blood J 2013;121(2):378–84.

[16] Fayyad UM, Piatetsky-Shapiro G, Smyth P. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R, editors. From Data Mining To Knowledge Discovery: An Overview, Advances In Knowledge Discovery And Data Mining. AAAI Press/The MIT Press; 1996. p. 1–34.

[17] Uchida S, Mori A, Kurazume R, Taniguchi R-I, Hasegawa T. Logical DP Matching for Detecting Similar Subsequence. Computer Vision – ACCV 2007 Lecture Notes in Computer Science:628–37. doi:10.1007/978-3-540-76386-4_59.

[18] Maji U, Mitra M, Pal S. Characterization of cardiac arrhythmias by variational mode decomposition technique. Biocybernetics and Biomedical Engineering 2017;37:578–89. doi:10.1016/j.bbe.2017.04.007.

[19] Dragomiretskiy K, Zosso D. Two-Dimensional Variational Mode Decomposition. Lecture Notes in Computer Science Energy Minimization Methods in Computer Vision and Pattern Recognition 2015:197–208. doi:10.1007/978-3-319-14612-6_15.

[20] Rajagopal R, Ranganathan V. Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification. Biomedical Signal Processing and Control 2017;34:1–8. doi:10.1016/j.bspc.2016.12.017.

[21] Shen C-P, Kao W-C, Yang Y-Y, Hsu M-C, Wu Y-T, Lai F. Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines. Expert Systems with Applications 2012;39:7845–52. doi:10.1016/j.eswa.2012.01.093.

[22] F. Scholkmann, J. Boss, and M. Wolf, “An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals,” Algorithms, vol.5, no. 4, pp.588–603, 2012

[23]P. Du, W. A. Kibbe, and S. M. Lin, “Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching,” Bioinformatics, vol.22, no. 17, pp.2059–2065, Apr. 2006.

[24] K. R. Coombes, S. Tsavachidis, J. S. Morris, K. A. Baggerly, M.-C. Hung, and H. M. Kuerer, “Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform,” Proteomics, vol.5, no. 16, pp.4107–4117, 2005.

[25] O. Singh, R.K. Sunkaria, “ A robust R-peak detection algorithm using wavelet transform”. Int. J. Comput. Appl. Vol.36, pp.37-43, 2011.

[26] Z. Nenadic and J. Burdick, “Spike Detection Using the Continuous Wavelet Transform,” IEEE Transactions on Biomedical Engineering, vol.52, no. 1, pp.74–87, 2005.

[27] C. Li, C. Zheng, and C. Tai, “Detection of ECG characteristic points using wavelet transforms,” IEEE Transactions on Biomedical Engineering, vol.42, no. 1, pp.21–28, 1995.

[28] J. M. Gregoire, D. Dale, and R. B. V. Dover, “A wavelet transform algorithm for peak detection and application to powder x-ray diffraction data,” Review of Scientific Instruments, vol.82, no. 1, pp.1-8, 2011.

[29] Q. Xue, Y. Hu, and W. Tompkins, “Neural-network-based adaptive matched filtering for QRS detection,” IEEE Transactions on Biomedical Engineering, vol.39, no. 4, pp.317–329, 1992.

[30] G. Vijaya, V. Kumar, and H. K. Verma, “ANN-based QRS-complex analysis of ECG,” Journal of Medical Engineering & Technology, vol.22, no. 4, pp.160–167, 1998.

[31] Y. Ferdi, J. Herbeuval, A. Charef, and B. Boucheham, “R wave detection using fractional digital differentiation,” Itbm-Rbm, vol.24, no. 5-6, pp.273–280, 2003.

[32] K.-P. Lin and W. Chang, “QRS feature extraction using linear prediction,” IEEE Transactions on Biomedical Engineering, vol.36, no. 10, pp.1050–1055, 1989.

[33] D.A. Coast, R. Stern, G. Cano, and S. Briller, “An approach to cardiac arrhythmia analysis using hidden Markov models,” IEEE Transactions on Biomedical Engineering, vol.37, no. 9, pp.826–836, 1990.

[34] K. Harmer, G. Howells, W. Sheng, M. Fairhurst, and F. Deravi, “A Peak-Trough Detection Algorithm Based on Momentum,” 2008 Congress on Image and Signal Processing, 2008.

[35] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” IEEE Transactions on Biomedical Engineering, vol.BME-32, no. 3, pp.230–236, 1985.

[36] A. Jacobson, “Auto-threshold peak detection in physiological signals,” 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

[37] G. Vivó-Truyols, J. Torres-Lapasió, A. V. Nederkassel, Y. V. Heyden, and D. Massart, “Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals,” Journal of Chromatography A, vol.1096, no. 1-2, pp.133–145, 2005.

[38] J. L. Excoffier and G. Guiochon, “Automatic peak detection in chromatography,” Chromatographia, vol.15, no. 9, pp.543–545, 1982

[39] Asgari, S., Mehrnia, A., & Moussavi, M. “Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine,” Computers in Biology and Medicine, vol.60, pp.132-142, 2015.

[40] Petrėnas, A., Marozas, V., & Sörnmo, L. “Low-complexity detection of atrial fibrillation in continuous long-term monitoring,” Computers in Biology and

Medicine, vol.65, pp.184-191,2015.

[41] S. K. Sahoo, W. Lu, S. D. Teddy, D. Kim, and M. Feng, “Detection of Atrial fibrillation from non-episodic ECG data: A review of methods,” 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011

[42] E. Ros, S. Mota, F. Fernandez, F. Toro, and J.Bernier, “ECG characterization of Paroxysmal atrial fibrillation: Parameter extraction and automatic diagnosis algorithm,” Computer in Biology and Medicine, vol. 4, no. 8, pp. 679-696, 2004.

[43] B. Pourbabaee and C. Lucas, “Automatic Detection and Prediction of Paroxysmal Atrial Fibrillation based on Analyzing ECG Signal Feature Classification Methods,” 2008 Cairo International Biomedical Engineering Conference, 2008.

[44] N. Kikillus, M. Schweikert, and A. Bolz, “Identifying Patients Suffering From Atrial Fibrillation During Atrial Fibrillation and Non-Atrial Fibrillation Episodes,” IFMBE Proceedings 4th European Conference of the International Federation for Medical and Biological Engineering, pp.1349–1352, 2009.

[45] Y. Chesnokov, A. Holden, and H. Zhang, “Screening patients with Paroxysmal atrial fibrillation (PAF) from non-PAF heart rhythm using HRV data analysis,” 2007 Computers in Cardiology, pp.459–462, 2007.

[46] B. Hickey, C. Heneghan, and P. D. Chazal, “Non-Episode-Dependent Assessment of Paroxysmal Atrial Fibrillation Through Measurement of RR Interval Dynamics and Atrial Premature Contractions,” Annals of Biomedical Engineering, vol.32, no. 5, pp.677–687, 2004.

[47] R. Das, I. Turkoglu, and A. Sengur, “Diagnosis of valvular heart disease

through neural networks ensembles," *Comput. Methods Programs Biomed.*, vol. 93, no. 2, pp. 185–191, 2009.

[48] N. C. S. N. I. M Anbarasi, E Anupriya, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 10, pp. 5370–5376, 2010.

[49] A. Rajkumar and M. G Sophia Reena, "Diagnosis Of Heart Disease Using Datamining Algorithm," *Glob. J Comput Sci Technol*, vol. 10, 2009.

[50] M. Kumari and S. Godara, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction," *Int. J. Comput. Sci. Trends Technol.*, vol. 2, no. 2, pp. 304–308, 2011.

[51] N. Aditya Sundar, P. Pushpa Latha, and R. Chandra, "Performance analysis of classification data mining techniques over heart disease database," *Int. J. Eng. Sci. Adv. Technol.*, vol. 2, 2018.

[52] C. S. Dangare and S. S. Apte, "a Data Mining Approach for Prediction of Heart Disease Using Neural Networks," *Int. J. Comput. Eng. Technol.*, vol. 3, no. 3, pp. 30–40, 2012.

[53] T. J. Peter and K. Somasundaram, "Study and Development of Novel Feature Selection Framework for Heart Disease Prediction," *Int. J. Sci. Res. Publ.*, vol. 2, no. 10, pp. 1–7, 2012.

[54] M. Shouman, T. Turner, and R. Stocker, "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients," *Int. J. Inf. Educ. Technol.*, vol. 2, no. 3, pp. 220–223, 2012.

[55] N. Bhatla and K. Jyoti, "An analysis of heart disease prediction using different data mining techniques," *Int J Eng Res Technol*, vol. 1, pp. 1–4, 2012.

[56] A. Pethalakshmi and A. Anushya, “Effective Features Selection via Futuristic Genetic on Heart Data,” *Int. J. Comput. Intell. Informatics*, vol. 2, no. 1, pp. 23–27, 2012.

[57] A. Taneja, “ORIENTAL JOURNAL OF Heart Disease Prediction System Using Data Mining Techniques,” 2013.

[58] C. R, “Heart Disease Prediction System Using Supervised Learning Classifier,” *Bonfring Int. J. Softw. Eng. Soft Comput.*, vol. 3, no. 1, pp. 01–07, 2013.

[59] I. S. F. Dessai, “Intelligent Heart Disease Prediction System Using Probabilistic Neural Network,” pp. 38–44, 2013.

[60] S. B. Patel, P. K. Yadav, and D. P. Shukla, “Predict the Diagnosis of Heart Disease Patients Using Classification Mining Techniques,” *J. Agric. Vet. Sci.*, vol. 4, no. 2, pp. 61–64, 2013.

[61] V. Chaurasia, “Early Prediction of Heart Diseases Using Data Mining,” *Caribb. J. Sci. Technol.*, vol. 1, pp. 208–217, 2013.

[62] A. Methaila, P. Kansal, H. Arya, and P. Kumar, “Early heart disease prediction using data mining techniques,” *Comput. Sci. Inf. Technol.*, pp. 53–59, 2014.

[63] K. Wisaeng, “Predict the diagnosis of heart disease using feature selection and k-nearest neighbor algorithm,” *Appl. Math. Sci.*, pp. 4103–4113, 2014.

[64] N. P. Waghulde and N. P. Patil, “Genetic Neural Approach for Heart Disease Prediction,” *Int. J. Adv. Comput. Res.*, vol. 4, no. 3, pp. 778–784, 2014.

[65] M. R. R. Patil, “Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 3, no. 5, pp. 6787–6789, 2014.

[66] B. Venkatalakshmi and M. V Shivsankar, “Heart Disease Diagnosis using Predictive Data Mining,” *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 3, no. 3, pp. 1873–1877, 2014.

[67] A. Jarad, R. Katkar, A. R. Shaikh, and A. Salve, “INTELLIGENT HEART DISEASE PREDICTION SYSTEM WITH MONGODB relative to rest,” vol. 4, no. 1, pp. 4–7, 2015.

[68] Andrea D. Souza, “Heart Disease Prediction using Data Mining Techniques,” *Int. J. Res. Eng. Sci.*, vol. 3, no. 3, pp. 74–77, 2015.

[69] M. P. Scholar, “a Survey on Heart Disease Prediction Using Data Mining Techniques,” vol. 5, no. 6, pp. 14–17, 2015.

[70] M. Abdar, S. R. N. Kalhori, T. Sutikno, I. M. I. Subroto, and G. Arji, “Comparing performance of data mining algorithms in prediction heart diseses,” *Int. J. Electr. Comput. Eng.*, vol. 5, no. 6, pp. 1569–1576, 2015.

[71] R. Kaur, “Prediction of Heart Disease Based on Risk Factors Using Genetic SVM Classifier,” vol. 5, no. 12, pp. 205–208, 2015.

[72] P. Shinde, S. B & Amrit, “Decision Support System on Prediction of Heart Disease Using Data Mining Techniques,” *Int. J. Eng. Res. Gen. Sci.*, vol. 3, no. 2, pp. 1453–1458, 2015.

[73] J. Patel, T. Upadhyay, and S. Patel, “Heart Disease Prediction Using Machine Learning and Data Mining Technique,” *Ijcsc*, vol. 7, no. March, pp. 129–137, 2016.

[74] K. Rajalakshmi and K. Nirmala, “Heart Disease Prediction with MapReduce by using Weighted Association Classifier and K-Means,” *Indian J. Sci. Technol.*, vol. 9, 2016.

[75] P. T. Selvy, “A proficient heart disease prediction method using fuzzy-cart

algorithm," no. 1, pp. 1–6, 2016.

[76] T. Karthikeyan and Vak. Associate Professor, "Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network," *Int. J. Adv. Res. Sci. Eng. Technol.*, vol. 4, no. 1, pp. 3194–3201, 2017.

[77] Himanshu Sharma and M.A. Rizvi. 2017. Prediction of Heart Disease using Machine Learning Algorithms: A Survey. International Journal on Recent and Innovation Trends in Computing and Communication 5, 8 (August 2017), 99–104.

[78] Vijaya G, Kumar V, Verma HK. ANN-based QRS-complex analysis of ECG. Journal of Medical Engineering & Technology 1998;22:160–7. doi:10.3109/03091909809032534.

[79] Florkowski, C. M. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *The Clinical Biochemist Reviews*, *29*(Suppl 1), S83–S87;2008.

[80] Goodwin LD, Leech NL. Understanding Correlation: Factors That Affect the Size of r. The Journal of Experimental Education 2006;74:249–66. doi:10.3200/jexe.74.3.249-266.

[81] Glass GV, Hopkins KD. Statistical methods in education and psychology. Boston: Allyn & Bacon; 2008.

[82] Lockhart RS. Introduction to statistics and data analysis for the behavioral sciences. New York: W.H. Freeman; 1998.

[83] Sprinthall RC. Basic statistical analysis. Pearson; 2014.

[84] Ghorbanian P, Ghaffari A, Jalali A, Nataraj C. Heart arrhythmia detection using

continuous wavelet transform and principal component analysis with neural network classifier. Computing in Cardiology 2010.

[85] Yeh Y-C, Lin H-J. Cardiac arrhythmia diagnosis Method using Fuzzy C-Means algorithm on ECG signals. 2010 International Symposium on Computer, Communication, Control and Automation (3CA) 2010. doi:10.1109/3ca.2010.5533831.

[86] Pan S-T, Chiou Y-J, Hong T-P, Chen H-C. Automatic recognition for arrhythmias with the assistance of Hidden Markov model. 2013 9th International Conference on Information, Communications & Signal Processing 2013. doi:10.1109/icics.2013.6782934.

[87] Yeh Y-C. An Analysis of ECG Beats by Using the Mahalanobis Distance Method. 2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC) 2009. doi:10.1109/icicic.2009.75.

[88] Shyu L-Y, Wu Y-H, Hu W. Using Wavelet Transform and Fuzzy Neural Network for VPC Detection From the Holter ECG. IEEE Transactions on Biomedical Engineering 2004;51:1269–73. doi:10.1109/tbme.2004.824131.

[89] Maji U, Mitra M, Pal S. Characterization of cardiac arrhythmias by variational mode decomposition technique. Biocybernetics and Biomedical Engineering 2017;37:578–89. doi:10.1016/j.bbe.2017.04.007.

[90] Lee J, Reyes BA, Mcmanus DD, Maitas O, Chon KH. Atrial Fibrillation Detection Using an iPhone 4S. IEEE Transactions on Biomedical Engineering 2013;60:203–6. doi:10.1109/tbme.2012.2208112.

[91] G. D. Pinna, R. Maestri, A. D. Cesare, R. Colombo, and G. Minuco, “The accuracy of power-spectrum analysis of heart-rate variability from annotated RR lists generated by Holter systems,” Physiological Measurement, vol.15,

no. 2, pp.163–179, Jan. 1994.

https://physionet.org/physiobank/database/, accessed Jul. 2. 2016.

[92] G. Vijaya, V. Kumar, and H. K. Verma, “ANN-based QRS-complex analysis of ECG,” Journal of Medical Engineering & Technology, vol.22, no. 4, pp.160–167, 1998.

[93] Larburu,N.,Lopetegi,T., and Romero,I. “Comparative study of algorithms for Atrial Fibrillation detection,” Computing in Cardiology,Vol. 38,pp.265-268, 2011.

[94] J. Slocum, A. Sahakian, and S. Swiryn, “Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity,” Journal of Electrocardiology, vol.25, no. 1, pp.1–8, 1992.

[95] S. Babaeizadeh, R. E. Gregg, E. D. Helfenbein, J. M. Lindauer, and S. H. Zhou, “Improvements in atrial fibrillation detection for real-time monitoring,” Journal of Electrocardiology, vol.42, no. 6, pp.522–526, 2009.

[96] K. Tateno and L. Glass, “Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals,” Medical & Biological Engineering & Computing, vol.39, no. 6, pp.664–671, 2001.

[97] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, and J. Habetha, “Detection of Atrial Fibrillation using model-based ECG analysis,” 2008 19th International Conference on Pattern Recognition, 2008.

[98] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, “Automatic Real Time Detection of Atrial Fibrillation,” Annals of Biomedical Engineering, vol.37, no. 9, pp.1701–1709, 2009.

[99] C. Huang, S. Ye, H. Chen, D. Li, F. He, and Y. Tu, “A Novel Method for

Detection of the Transition Between Atrial Fibrillation and Sinus Rhythm," IEEE Transactions on Biomedical Engineering, vol.58, no. 4, pp.1113–1119, 2011.

[100] S. Sarkar, D. Ritscher, and R. Mehra, "A Detector for a Chronic Implantable Atrial Tachyarrhythmia Monitor," IEEE Transactions on Biomedical Engineering, vol.55, no. 3, pp.1219–1224, 2008.

[101] J. Lee, Y. Nam, D. D. Mcmanus, and K. H. Chon, "Time-Varying Coherence Function for Atrial Fibrillation Detection," IEEE Transactions on Biomedical Engineering, vol.60, no. 10, pp.2783–2793, 2013.

[102] K. Jiang, C. Huang, S.-M. Ye, and H. Chen, "High accuracy in automatic detection of atrial fibrillation for Holter monitoring," Journal of Zhejiang University SCIENCE B, vol.13, no. 9, pp.751–756, 2012.

[103] Asgari, S., Mehrnia, A., & Moussavi, M. "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," Computers in Biology and Medicine, vol.60, pp.132-142, 2015.

[104] X. Zhou, H. Ding, B. Ung, E. Pickwell-Macpherson, and Y. Zhang, "Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy," BioMedical Engineering OnLine, vol.13, no. 1, p.18, 2014.

[105] P. D. Carvalho, J. Henriques, R. Couceiro, M. Harris, M. Antunes, and J. Habetha, "Model-Based Atrial Fibrillation Detection," ECG Signal Processing, Classification and Interpretation, pp.99–133, Nov. 2011.

[106] D. E. Lake and J. R. Moorman, "Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices," AJP: Heart and Circulatory Physiology,

vol.300, no. 1, 2010.

[107] J. Lian, L. Wang, and D. Muessig, “A Simple Method to Detect Atrial Fibrillation Using RR Intervals,” The American Journal of Cardiology, vol.107, no. 10, pp.1494–1497, 2011.

[108] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, “Automatic Real Time Detection of Atrial Fibrillation,” Annals of Biomedical Engineering, vol.37, no. 9, pp.1701–1709, 2009.

[109] R. Das, I. Turkoglu, and A. Sengur, “Diagnosis of valvular heart disease through neural networks ensembles,” *Comput. Methods Programs Biomed.*, vol. 93, no. 2, pp. 185–191, 2009.

[110] Mukaka M. A guide to appropriate use of Correlation coefficient in medical research. Malawi Medical Journal : The Journal of Medical Association of Malawi. 2012;24(3):69-71.

[111] V. R. Zurro, a. L. Stelle, and J. Nadal, “Detection of atrial persistent rhythm based on P-wave recognition and RR interval variability,” Comput. Cardiol. 1995, pp. 185–188, 1995.

[112] H. Pürerfellner et al., “P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors,” Hear. Rhythm, vol. 11, no. 9, pp. 1575– 1583, 2014.

APPENDICES

PUBLICATION

Journal

1) Mohamad Sabri bin Sinal and Eiji Kamioka, “Hybrid Mechanism to Detect Paroxysmal Stage of Atrial Fibrillation Using Adaptive Threshold-Based Algorithm with Artificial Neural Network,” IEICE Transactions on Information and Systems, Vol. E101-D, No. 6, pp. 1666–1676, Jan. 2018
2) Mohamad Sabri Bin Sinal and Eiji Kamioka, "An Efficient Arrhythmia Detection Using Autocorrelation and Statistical Approach," International Journal of Computer and Communications, Vol.6, No.10, 2018, pp.63-81.
3) Mohamad Sabri Bin Sinal and Eiji Kamioka, “Effective Computational Modeling for Early Arrhythmia Symptom Classification by using Decision Tree Approach," International Journal of Bioscience, Biochemistry and Bioinformatics, 2018,(Accepted).

International Conference

4) Mohamad Sabri Bin Sinal and Eiji Kamioka, "ADAPTIVE THRESHOLD BASED APPROACH TO PERFECTLY DETECT HEART CYCLE IN ECG DATA," Proceedings of the 6th International Conference on Computing and Informatics 2017 (ICOCI2017), Selangor, Malaysia, April 26, 2017, pp. 492-498.
5) Mohamad Sabri Bin Sinal and Eiji Kamioka, " EARLY ABNORMAL HEARTBEAT MULTISTAGE CLASSIFICATION BY USING DECISION TREE AND K-NEAREST NEIGHBOR," Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference(AICCC2018),Tokyo, Japan, (Accepted).

Book chapter

6) M. S. B. Sinal and E. Kamioka. “Improvement of heart cycle detection accuracy for long duration ECG data,” The Landscape of Computing and Information research, 2016-2017, pp. 143-156, UUM publication.