Status of thesis

Title of thesis	Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System

I, NOOR AKHMAD SETIAWAN ,

hereby allow my thesis to be placed at the Information Resource Centre (IRC) of Universiti Teknologi PETRONAS (UTP) with the following conditions:

1. The thesis becomes the properties of UTP.
2. The IRC of UTP may make copies of the thesis for academic purposes only.
3. This thesis is classified as

☐ Confidential

☑ Non-confidential

If this thesis is confidential, please state the reason:

____

The contents of the thesis will remain confidential for __________ years.

Remarks on disclosure:

____

Endorsed by

27/7/09

Signature of Author	Signature of Supervisor
Gambiran UH V 225 Yogyakarta Indonesia	Universiti Teknologi PETRONAS
Date : 27-7-2009	Date : 27-7-09

UNIVERSITI TEKNOLOGI PETRONAS

Approval by Supervisor (s)

The undersigned certify that they have read, and recommend to The Postgraduate Studies Programme for acceptance, a thesis entitled **"Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System"** submitted by (**Noor Akhmad Setiawan**) for the fulfilment of the requirements for the degree of Doctor of Philosophy in Electrical and Electronics Engineering.

____

Date

Signature : 27/7/09

Main supervisor : Prof. Dr. P.A. Venkatachalam

Date : 27-7-09

TITLE PAGE

UNIVERSITI TEKNOLOGI PETRONAS

Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System

By

Noor Akhmad Setiawan

A THESIS
SUBMITTED TO THE POSTGRADUATE STUDIES PROGRAMME
AS A REQUIREMENT FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY
ELECTRICAL AND ELECTRONIC ENGINEERING
BANDAR SERI ISKANDAR,
PERAK

JULY, 2009

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTP or other institutions.

Signature :

Name : NOOR AKHMAD SETIAWAN

Date : 27-7-2009

Acknowledgement

First and foremost, all praises and thanks are due to ALLAH, the almighty God, the source of my life and hope for giving me the strength and wisdom to succeed in life.

I am most grateful to my supervisor Professor Dr PA Venkatachalam for giving me an opportunity to pursue a PhD degree. Many times, his patience and constant encouragement has steered me to be a better person. I am also thankful to my co-supervisor Professor Ir Dr Ahmad Fadzil Mohd Hani.

I should also mention very kindly persons who contributed greatly in my study: Associate Professor Dr Mohd Noh Karsiti, Director of Postgraduate Studies and Dr Nor Hisham Hamid, Head of Electrical and Electronic Engineering Department, provided me resources to make this thesis completed. Special thanks to Dr Vijanth Sagayan Asirvadam for the helpful discussion and suggestion. I would like also express my gratitude to Dr. Hasral Noor Hasni from Ipoh Specialist Hospital, Dr. Choong Choon Hooi and Dr. Tan Huat Chai from Hospital Fatimah for their effort in helping and providing me with the all data and guidance that I need for this research.

Special thanks also to my postgraduate research fellows Hanung Adi Nugroho, Teguh Bharata Adji, Dani Adhipta and Agus Arif for helping me as proofreaders of this thesis and special thanks also to Intan, Dalila and Raja Syahira for the abstract in Bahasa Melayu. Additional thanks to Mr Fakhizan for giving me the opportunity as his tutorial assistant and laboratory assistant and thanks also to Mr Zurimi as the lab technologist. Special thank also to all my friends for their helpful and special friendships.

At last and most importantly, I would like to thank my family for their open-mindedness and endless support. They are always close to my heart.

This thesis is dedicated to my beloved parents Mr Wazir Nuri rahimahuLLAAH and Mrs Sri Suhartati, my beloved wife Etik Irijanti and my beloved children Ali, Ubai and Atika.

Abstract

Heart disease is any disease that affects the normal condition and functionality of heart. Coronary Artery Disease (CAD) is the most common. It is caused by the accumulation of plaques within the walls of the coronary arteries that supply blood to the heart muscles. It may lead to continued temporary oxygen deprivation that will result in the damage of heart muscles. CAD caused more than 7,000,000 deaths every year in the worldwide. It is the second cause of death in Malaysia and the major cause of death in the world. To diagnose CAD, cardiologists usually perform many diagnostic steps. Unfortunately, the results of the diagnostic tests are difficult to interpret which do not always provide definite answer, but may lead to different opinion. To help cardiologists providing correct diagnosis of CAD in less expensive and non-invasive manner, many researchers had developed decision support system to diagnose CAD.

A fuzzy decision support system for the diagnosis of coronary artery disease based on rough set theory is proposed in this thesis. The objective is to develop an evidence based fuzzy decision support system for the diagnosis of coronary artery disease. This proposed system is based on evidences or raw medical data sets, which are taken from University California Irvine (UCI) database. The proposed system is designed to be able to handle the uncertainty, incompleteness and heterogeneity of data sets. Artificial Neural Network with Rough Set Theory attribute reduction (ANNRST) is proposed is the imputation method to solve the incompleteness of data sets. Evaluations of ANNRST based on classifiers performance and rule filtering are proposed by comparing ANNRST and other methods using classifiers and during rule filtering process. RST rule induction is applied to ANNRST imputed data sets. Numerical values are discretized using Boolean reasoning method. Rule selection based on quality and importance is proposed. RST rule importance measure is used to select the most important high quality rules. The selected rules are used to build fuzzy decision support systems. Fuzzification based on discretization cuts and fuzzy rule weighing based on rule quality are proposed. Mamdani inference method is used to provide the decision with centroid defuzzification to give numerical results, which represent the possibility of blocking in coronary arteries.

The results show that proposed ANNRST has similar performance to ANN and outperforms k-Nearest Neighbour (k-NN) and Concept Most Common attribute value

Filling (CMCF). ANNRST is simpler than ANN because it has fewer input attributes and more suitable to be applied for missing data imputation problem. ANNRST also provides strong relationship between original and imputed data sets. It is shown that ANNRST provide better RST rule based classifier than CMCF and k-NN during rule filtering process. Proposed RST based rule selection also performs better than other filtering methods. Developed Fuzzy Decision Support System (FDSS) provides better performance compared to multi layer perceptron ANN, k-NN, rule induction method called C4.5 and Repeated Incremental Pruning to Produce Error Reduction (RIPPER) applied on UCI CAD data sets and Ipoh Specialist Hospital's patients. FDSS has transparent knowledge representation, heterogeneous and incomplete input data handling capability. FDSS is able to give the approximate percentage of blocking of coronary artery based on 13 standard attributes based on historical, simple blood test and ECG data, etc, where coronary angiography or cardiologist can not give the percentage. The results of FDSS were evaluated by three local cardiologists and considered to be efficient and useful.

Abstrak

Penyakit Jantung ialah mana-mana jenis penyakit yang mengganggu kelancaran fungsi jantung dan menyebabkan jantung berada pada keadaan tidak normal. Coronary Artery Disease (CAD) adalah contoh penyakit yang paling banyak dihidapi oleh orang ramai. Penyakit tersebut adalah disebabkan oleh terkumpulnya plak pada dinding arteri koronari yang berfungsi menyalurkan darah ke otot jantung. Hal ini boleh membawa kepada kekurangan oksigen untuk sementara waktu tetapi boleh menyebabkan kerosakan otot jantung. CAD telah menyebabkan kematian sebanyak 7,000,000 orang setiap tahun di seluruh dunia. Ia merupakan punca kematian kedua terbesar di Malaysia dan punca kematian terbesar di seluruh dunia. Untuk mengkaji tentang CAD, ahli kardiologi kebiasaannya menjalankan beberapa langkah ujian. Walau bagaimanapun, keputusan ujian diagnosis yang diperoleh adalah sukar untuk difahami dan tidak dapat memberikan keputusan ujian yang tepat yang boleh menyebabkan kepada tersalah tafsir. Oleh yang demikian, untuk membantu ahli kardiologi mendapatkan keputusan kajian CAD yang tepat tetapi dalam masa yang sama menggunakan kos yang rendah dan non-invasive, banyak ahli kaji telah menghasilkan system bantuan untuk membuat keputusan bertujuan untuk mendiagnosis CAD.

Thesis ini mengusulkan *fuzzy decision support system* (FDSS) untuk mendiagnosis CAD berdasarkan rough set theory. Objektifnya adalah untuk membangunkan FDSS berdasar fakta bagi mendiagnosis CAD. Sistem ini adalah diusulkan berdasarkan bukti-bukti atau data perubatan yang diambil daripada Universiti Calfornia Irvine (UCI). Ia diciptakan agar mampu mengatasi mana-mana keadaan yang tidak menentu, ketidaksempurnaan, dan ketidakseragaman daripada data. Artificial Neural Network (ANN) dengan Rough Set Theory (RST) attribute reduction (ANNRST) yang diusulkan adalah kaedah imputasi bagi menyelesaikan data yang tidak lengkap. Evaluasi untuk ANNRST berdasarkan prestasi pengelas dan juga penyaringan peraturan diusulkan. Iaitu membandingkan ANNRST dan kaedah lain dengan menggunakan pengelas dan juga pada saat proses penyaringan peraturan. Induksi peraturan RST diaplikasikan kepada data yang telah dilengkapi dengan ANNRST. Nilai selanjar adalah didiskritkan dengan kaedah Boolean reasoning. Pemilihan peraturan berdasarkan kepada kualiti dan kepentingan diusulkan. Pengukuran kepentingan peraturan berdasarkan RST digunakan untuk memilih peraturan yang paling penting dan berkualiti. Peraturan yang terpilih digunakan bagi membina FDSS.

Fuzzification berdasarkan titik potong pendiskritan dan pemberat peraturan berdasar kualiti diusulkan. Kaedah inferens Mamdani digunakan untuk menyediakan keputusan dengan *centroid fuzzification* untuk memberi keputusan selanjar yang mewakili kebarangkalian penyekatan dalam arteri koronari.

Hasil-hasil penyelidikan menunjukkan ANNRST serupa kepada ANN dan mengatasi k-Nearest Neighbour (k-NN) dan Concept Most Common nilai Filling (CMCF). ANNRST adalah lebih mudah daripada ANN kerana ia mempunyai input lebih sedikit dan lebih sesuai diaplikasikan untuk masalah data hilang. ANNRST juga menunjukkan hubungan erat antara data asal dan data terlengkapi. ANNRST menyediakan pengelas RST lebih baik daripada CMCF dan k-NN semasa proses penapisan peraturan. Pemilihan peraturan berdasarkan RST juga lebih baik daripada kaedah pemilihan yang lain. FDSS mempunyai prestasi lebih baik berbanding MLP-ANN, k-NN, C4.5 dan Repeated Incremental Pruning kepada Produce Error Reduction (RIPPER) yang digunakan di atas Data CAD UCI dan Ipoh Specialist Hospital. FDSS mempunyai pengetahuan boleh difahami, pengendalian data input heterogen dan imputasi data tidak lengkap. FDSS boleh memberikan peratusan penyekatan arteri koronari berdasarkan 13 keadaan standard berdasarkan data-data perubatan, di mana angiografi jantung atau pakar kardiologi tak boleh beri peratusan. FDSS dinilai oleh tiga orang pakar kardiologi tempatan dan dianggap menjadi cekap dan berguna.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Heart disease can be categorized as disorders of conduction and rhythm, disorders of myocardium (heart muscle) and disorder of heart's structure. Coronary Artery Disease (CAD), sometimes called as Coronary Heart Disease (CHD), cardiomyopathy, ischemic heart disease and heart failure are the most common heart diseases. They belong to disorders of myocardium category. CAD is caused by the accumulation of plaques within the walls of the coronary arteries that supply blood to the myocardium. CAD may lead to continued temporary oxygen deprivation that will result in the damage of myocardium. This disease is called ischemic heart disease. CAD causes more than 500,000 deaths every year in USA and more than 7,000,000 deaths every year in the worldwide. It is the second cause of death in Malaysia and the major cause of death in the world [1-8].

The primary symptom of CAD is chest pain or *angina*. This symptom of CAD always occurs during or after physical activity or emotional stress for most of the people. The presence of *angina* can be used to diagnose the presence of CAD. However, diagnosing CAD using only symptoms that occur may be difficult. Cardiologist must consider many factors in order to give the correct diagnosis of CAD. Usually medical history and physical examinations are used in the initial stage of diagnosis. Heart sound can be used to determine whether the blood flows to the heart normally. The advanced tests are conducted after initial diagnosis which can utilize electrocardiography (ECG), echocardiogram, stress test, nuclear imaging and coronary angiography. Coronary angiography is called the "gold standard" diagnosis of disease in the coronary arteries [1][2].

The difficulties of CAD diagnosis lead many researchers to develop computer methods for the diagnosis of heart disease. At the earlier time, the use of computer is to build knowledge based decision support system which uses knowledge from medical experts and transfers this knowledge into computer algorithms manually. This process can be time consuming and strongly depends on medical expert's opinion which may be subjective. To handle this problem, Decision Support System (DSS) based on intelligent machine learning techniques and knowledge discovery from data have been proposed to gain knowledge automatically from examples or raw data.

Several research papers on DSS for diagnosis of coronary artery disease have been proposed by many researchers. A Multilayer perceptron based medical decision support system has been developed to diagnose five types of heart diseases including coronary heart disease [9]. Bayesian network model of heart disease is proposed [10]. The system could predict the probabilities of heart diseases and dependency among attributes related to heart diseases. A set of machine learning methods was evaluated on the atherosclerotic coronary heart disease [11]. Prognosis of cardiac events (cardiac death or non-fatal myocardial infarction) was proposed in [12]. Diagnosis of ischemic heart disease using various machine learning techniques was found in [13]. Extension of multi layer perceptron for coronary heart disease diagnosis by making it interpretable was shown in [14]. A combination of data mining technique and fuzzy modelling was used to diagnose CAD [15-17]. Most of these research works uses large number of patients.

Rough set theory is a mathematical method to deal with imprecise, vague and uncertain data sets [18]. This theory was originally proposed by Pawlak [19][20]. A set is rough if its elementary set can not be clearly and precisely defined using the available knowledge. Medical data is usually uncertain, imprecise and vague. Thus, RST with other artificial intelligent techniques can be used to build decision support systems in medical diagnosis [17][21]. The application of rough set theory in practical problem is still need more development and improvement.

The quality of machine learning and knowledge discovery strongly depends on the data. Obtaining large amount of medical high quality data may not always be an easy task or even possible. For example, not all the laboratory tests are available in most hospitals. Incompleteness of data sets is the common issue in medical data set collection. Many research works in missing value prediction have been conducted in general data sets and specific area data sets. Grzysmala-Busse and Hu compare and present nine different approaches to missing attribute values [22]. Al Shalabi, Najjar and Al Kayed present a framework to deal with missing data [23]. Li and Cercone propose RSFit, a combination of rough set and distance method, to assign the missing data [24]. Wasito and Mirkin present nearest neighbour approach in the least square data imputation algorithms [25][26]. Ragel and Cremilleux build missing values completion based on association rules [27]. Troyanskaya, et al, compare missing data imputations methods for DNA microarrays [28]. Missing value estimation in DNA microarrays is found in [29]. Missing value estimation methods in other specific area are also discussed in [30-32]. In the above works, more comprehensive evaluation method of missing data imputation is not discussed. The evaluation should not be only based on accuracy and simulated missing values. The performance of classifiers that will use the imputation should be taken into account.

The knowledge discovery process extracts the knowledge in various forms such as decision tree, association rules, decision rules, sequential pattern, etc. The most comprehensive and interpretable extracted knowledge is in the form of rules. Some rule induction algorithm such as rough set theory results in large number of rules. This large number of rules makes interpretability of the knowledge low. Lacking of interpretability will cut down the advantages of rule based systems which should be interpretable and easy to understand. The resulting large number of rules is caused by noise, redundancy of input or training data sets and others. Large number of redundant rules can not provide efficient result. Rule pruning is the method to reduce the number of rules while maintaining the quality of the system. Agotnes investigates genetic algorithm and the use of quality measure for individual rule to filter the large number of rules [33]. Maddouri and Gammoudi conduct comparative study on semantic properties of rule interestingness measure [34]. Li and Cercone propose another measure of rules which is called rule importance measure [35]. This rule importance measure is applied on association rules. The combination of these rule measures is not yet implemented for rule selection problems. Existing works are using only individual measure for rule selection problems.

1.2 OBJECTIVES

The main objective of this research is to develop an evidence based fuzzy decision support system for the diagnosis of coronary artery disease. The system must be able to handle uncertain medical data, incomplete and different types of data which are heterogeneous. In order to reach the main goal, the following objectives should be achieved.

i. To develop the appropriate imputation method to predict the missing data in coronary artery disease data sets. The result of imputation should be not only accurate but must have strong relationship with original complete data. Artificial neural network can learn the relationship among attributes and predict the missing values based on this relationship. In order to simplify the artificial neural network without reducing its performance, rough set theory is embedded to reduce the number of input attributes.

ii. To discover and mine the knowledge from coronary artery disease in the form of classification rules and to develop rule selection method for filtering the generated large number of rules from coronary artery disease data sets. The quality and importance of rules is used as filtering criteria. A reduct of rough set theory of full data sets represents the important feature of data set which can be used to find the most important rule set by considering rules as attributes.

iii. To develop fuzzification method of crisp rules and to improve the fuzzy inference process for estimating the approximate percentage of blocking of coronary artery that indicates the presence of coronary artery disease. The fuzzification of numerical input attributes is proposed based on discretization results of those numerical attribute values. The improvement of fuzzy inference system is proposed by considering the weight of rules derived from rule supports.

1.3 THESIS CONTRIBUTIONS

The major contributions of this thesis are:

i. This thesis proposes a missing data imputation based on Artificial Neural Network (ANN) with Rough Set Theory (RST) attribute reduction, namely ANNRST, which provides similar performance to artificial neural network imputation but with simpler topology architecture with smaller number of input attributes.

ii. This research introduces a new evaluation method for real missing attribute values imputation. The evaluation is based on classifiers which classify complete training data set using imputed data set to find the relationship between imputed data set and complete data set. The high performance of classifiers indicates the high relationship between complete training data set and imputed data set. Almost of existing imputations are evaluated using simulation of missing values and accuracy of imputation.

iii. This work contributes another new evaluation method of imputation evaluations based on rule filtering. Rule filtering will affect the quality of rule based classifiers, the quality of imputation can be observed based on rule filtering. The imputation method can be considered better when the filtered rules are still has high performance during filtering process.

iv. The rule selection based on combined rule quality and importance proposes a new idea in the field of rule selection. The rule quality filtering based on support is fast and simple to filter very large number of rules. The rule importance filtering based on rough set theory is applied on the medium number of rules to select the most important rules from the filtered high quality rules. To maintain the generalization of rules, a decision table of rules is build based on testing data set instead on training data set. This is one of the differences from previous rough set based rule filtering.

v. The thesis contributes new ideas in fuzzy system development for the diagnosis of coronary artery disease. The membership function creation is based on Boolean reasoning discretization of numerical attributes, which is considered indirect learning from training decision table. The rule weighing based on rule support is proposed to assign the strength of fuzzy rules during inference process. The defuzzification of output provides numerical values that can be considered as percentage of blocking possibility of coronary artery. Comparable performance to cardiologists and other computer based methods is obtained even without adjusting and optimizing the membership functions.

1.4 THESIS ORGANIZATION

This thesis is organized in five chapters as follows:

Chapter 1 introduces the background and problems raised in development of decision support system for coronary artery disease diagnosis. It describes the important problems in coronary artery disease diagnosis and general issues in development of decision support system. It also explains the objective of the research and thesis major contributions.

Chapter 2 provides a literature review of problems in coronary artery disease diagnosis and recent research works in development of decision support system. The chapter starts from coronary artery disease and diagnosis. Relevant research works in computer methods applied on diagnosis of coronary artery disease are reviewed in this chapter. Review of problems and recent works in development of decision support system using artificial intelligence are also included in this chapter. At the end of this chapter, the research goal of this thesis is revisited to justify the position of the research among other relevant works.

Chapter 3 gives the detailed methodology to build the fuzzy decision support system and the experimental design. It consists of coronary artery disease data sets taken from University of California, Irvine, and Ipoh Specialist Hospital, Malaysia. This chapter describes the detailed methods and design of the missing data imputation and evaluations, the rule generation method based on rough set theory, the rule selection based on rule quality measure and importance, the fuzzy modeling and inference and the evaluation method of the developed fuzzy decision support system for coronary artery disease diagnosis. This chapter provides the major contributions of this thesis.

Chapter 4 is the results of experimental study of the missing data imputation and evaluation, the rule selection performance, the fuzzy decision support system and evaluation. This chapter provides benchmarking of the proposed methods with other existing methods using data from University of California, Irvine and Ipoh Specialist Hospital, Malaysia. The discussion of each experimental works is also provided.

Chapter 5 presents the concluding remark and suggested recommendations for future research work to improve or extend the outcome of current research.

CHAPTER 2: LITERATURE REVIEW OF CAD AND DSS

2.1 INTRODUCTION

Heart is the most amazing organ inside the human body. Like any other device, heart also can go wrong. The abnormal heart function can be less efficient in functionality or break down. These abnormal conditions of the heart are considered as heart diseases.

Heart disease is a term that covers any disease that affects the normal condition and functionality of heart. Heart disease is a type of Cardiovascular Disease (CVD). Generally, diseases of the heart can be divided into three main categories as explained by Randall and Romaine [6]. These categories are disorders of conduction and rhythm, disorders of myocardium (heart muscle) and disorder of heart's structure.

Disorders of conduction and rhythm category, associates the heart's electrical system. Arrhythmia (literally, "no rhythm") is included in this category. This disease is caused by the heart rate that is too slow, too fast or irregular. This arrhythmia can affect one atrium, both of atria, one ventricle, both of ventricles or the entire heart. Sometimes arrhythmia is caused by hypertension (high blood pressure). Another disease of this category is heart block. Heart block is conductive disorder of heart's electrical system. In this case, electrical impulses do not travel through the heart's electrical system correctly. Coronary Artery Disease (CAD), cardiomyopathy, ischemic heart disease and heart failure are the most common heart diseases of disorders of myocardium category. CAD is caused by the accumulation of plaques within the walls of the coronary arteries that supply blood to the myocardium. CAD may lead to continued temporary oxygen deprivation that will result in the damage of myocardium. This disease is called ischemic heart disease. Heart failure is the inability of the heart to pump enough blood to meet the body's need. This can be caused by a loss of myocardium strength or other causes such as long standing hypertension, narrowed exit valves of the heart, leaky heart valves, viral infections, alcohol and heart attack which lead to inefficiency of myocardium. Cardiomyopathy occurs when the heart's ability and capacity to pump blood is decreased. This decreased ability of the hearts is often accompanied by enlargement of the heart (hypertrophy). Cardiomyopathy may lead to heart failure. Congenital and valvular heart diseases belong to disorder of structure category of heart disease. Congenital heart disease presents at birth but not necessarily hereditary, which is acquired during foetal development. Valvular heart disease happens when heart's valves are failed to function properly [1][6].

Among heart diseases, coronary artery disease is the first killer and the most common heart disease in developed countries as well as in developing countries [1][2][5-7]. However, the diagnosis of coronary artery disease is difficult, especially when there is no symptom occurs. Much information from patients is needed in order to draw the correct diagnosis. It will be beneficial to use an advanced computer method such as artificial intelligence to build decision support system for the diagnosis of coronary artery disease. Therefore, diagnosis of coronary artery disease is chosen as the problem for this research.

2.2 CORONARY ARTERY DISEASE

One of the abnormal functions of the hearts is arteriosclerotic heart disease. Arteriosclerosis (hardening of the artery) may be caused by atherosclerosis (athero refers to the fatty substance) resulting from the buildup of fatty substances named as plaque on the wall of the arteries inside. This plaque can cause reduction or blocking in blood flow. Heart attack occurs when the blood flow is blocked by blood clot. The fat deposits within the wall of arteries are called atheroma. Atherosclerosis that forms plaque and the blocking of blood flow which causes heart attack can be illustrated in Figure 2.1 and Figure 2.2, respectively.

Atherosclerosis can be particularly dangerous when the location of plaque is in a coronary artery which supplies blood to the heart muscles. This condition is called Coronary Artery Disease (CAD). The coronary artery must be narrowed by 50 percent of its normal diameter before the reduction of blood flow to the heart can be considered serious and gives symptoms. This is because at that point, the supply of oxygen is not sufficient for myocardium to operate at resting condition. CAD caused almost 500,000 deaths every years in USA and it is the major cause of death in the world [1-7].

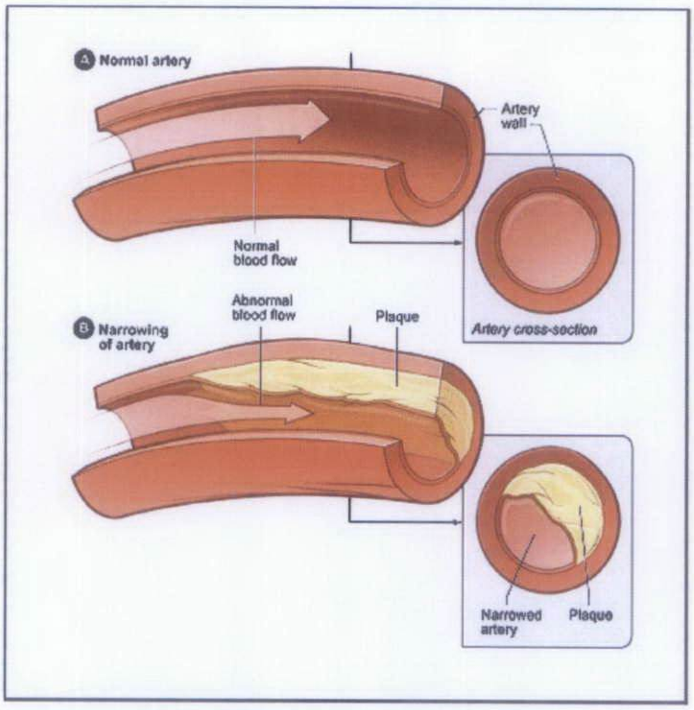


Figure 2.1 Atherosclerosis : (a) A normal artery with normal blood flow (b) An artery with plaque buildup [36].

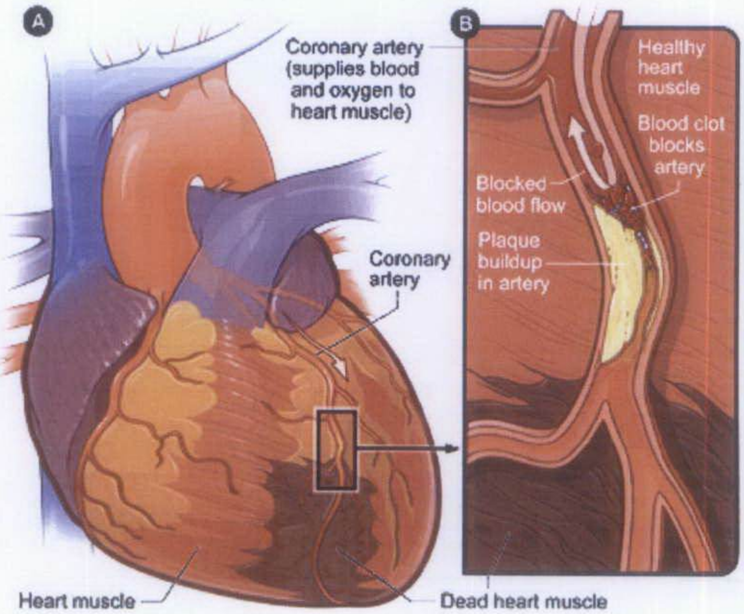


Figure 2.2 Heart with muscle damage and blocked artery: (a) An overview of a heart and coronary artery showing damage (dead heart muscle) caused by a heart attack (b) a cross-section of the coronary artery with plaque buildup and a blood clot [37].

The exact causes of build up of atheroma are not known. The long term studies have made the people to identify what are called *risk factors* that increase the possibility of atherosclerosis development. Some of these risk factors are controllable such as smoking, high blood pressure, cholesterol level, etc, while others are uncontrollable such as age, gender, family history and other similar factors [1][3][4]. During the development of atheroma, there will be signs and symptoms of CAD. When the blockage of coronary artery is very low, there is no sign and symptom occurs because heart muscles still have enough supply of oxygen, although there is small oxygen deprivation,. The increasing of blockage inside the coronary artery will give signs and symptoms like shortness of breath and swelling in feet and ankles. When the blockage is more than 50%, it will give chest pain (angina) as symptom because coronary artery can not supply enough blood to meet the oxygen demands. Finally, when coronary artery is totally blocked, heart attack will occur. These signs and symptoms can be illustrated in Figure 2.3. They start from silent ischemia, which is considered as no symptom, to heart attack, which is the worst.

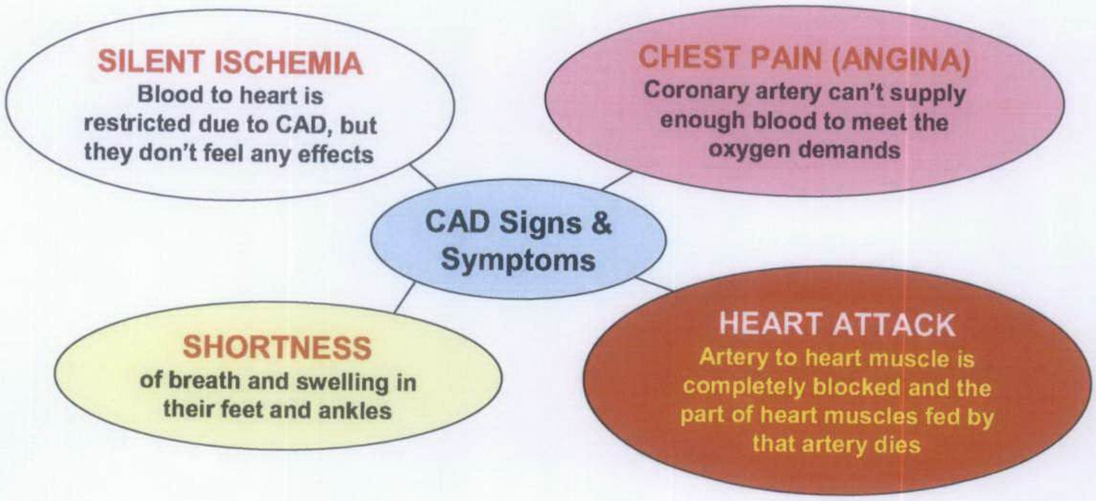


Figure 2.3 CAD signs and symptoms

The primary symptom of CAD is chest pain or *angina*. For most people, this symptom of CAD always occurs during or after physical activity or emotional stress. For some people this symptom does not occur. The presence of *angina* can be used to diagnose the presence of CAD. However, diagnosing CAD using only symptoms that occur may be difficult. It is important for the physicians to consider many factors in order to give the correct diagnosis of CAD. The initial diagnosis is usually by using medical history and physical examinations. The medical history provides information about symptoms, risk factors and data such as heart and lung problem, diabetic condition, chest pain, blood pressure and cholesterol. The physical exam is conducted by listening to the sound of the heart to determine the blood flow to the heart normally. After taking the initial diagnosis, the advanced test can be conducted. The advanced tests are electrocardiography (ECG), echocardiogram, stress test, nuclear imaging and coronary angiography. Coronary angiography provides the "gold standard" diagnosis of disease in the coronary arteries [1][2].

All the tests are difficult to interpret. The first problem is that the test does not always provide definite answer, but may lead to different opinion. The second problem is that every test occasionally provides incorrect information. Certain results expected to be present in some diseases may be not found (false negative), and inversely the test may provide the results that are not expected to be present (false positive). Physician may be confused because of the inconsistent results of the examinations and tests. The diagnosis of CAD usually needs many tests. Sometimes diagnosis needs more than what is necessary to obtain the good result. Unfortunately the cost-benefit relationship may restrict the physician and patient to do more tests where cardiological procedures are among the most expensive in medicine. The advanced tests such as nuclear imaging and coronary angiography are much more expensive than electrocardiogram or chest-X-ray [3]. In general the diagnosis of heart disease moves on in a step by step manner from the simplest, least invasive, least expensive and least risky method. As more information about a patient's condition can be obtained, proper decision can be made regarding the use of more advanced and more invasive diagnostic processes [1].

2.2.1 Conventional methods for the diagnosis of coronary artery disease

To diagnose coronary artery disease is not easy task. Usually several diagnosis steps should be performed. Many people are unaware that they have heart disease until heart attack or stroke occurs because there is no symptom occurs previously. Some people have symptoms such as angina, shortness of breath, etc. Usually cardiologist will diagnose coronary artery disease based on medical and family histories, risk factors and the results of a physical exam and diagnostic tests and procedures.

The medical and family histories provide information about person's risk of coronary artery disease. Family histories include all significant medical concerns for which the person has experienced symptoms or sought treatment of heart disease, kidney problems, cancer, blood disorders, diabetes, neurogical disorders, etc. Family histories also include any medical operations, medications, and symptoms such as angina, dizziness or shortness of breath, family history of heart attack or stroke in close blood relatives. To examine risk factors, cardiologist will ask the patient to identify relevant risk factors. Risk factors that cannot be changed are age, gender and heredity. Risk factors that can be changed are high blood pressure, elevated serum cholesterol, lipoprotein, cigarette smoking, obesity, glucose intolerance, diabetes, fibrinogen, left ventricular hypertrophy, cocaine and behavioral factors. There are opposite of risk factors called protective factors such as HDL cholesterol, exercise, estrogen [1][6].

After all the personal and family histories and risk factors are identified, cardiologist will decide whether physical exam and diagnostic tests and procedures should be performed or not. No single test can diagnose CAD. If the cardiologist thinks patient has CAD, he or she will probably do one or more of the following tests [1].

Electrocardiogram (ECG/EKG) is one of the simplest and most routine tests used by cardiologists. Using ECG, cardiologists measure and record the electrical activity of the heart which is plotted in the form of a series of wave representing each heartbeat. It provides initial evaluation of patient suspected heart disease. It can also detect the presence of heart attack old or current. ECG is noninvasive, safe, easy and relatively low cost but it has disadvantages such as often non specific, not sufficiently precise for detailed diagnosis. ECG is usually available in all health care facilities.

Stress test is performed by walking on treadmill or pedaling exercise bicycle at increasingly higher levels of exertion while heart rate and rhythm, blood pressure and sometimes oxygen consumption are monitored. This test can evaluate chest pain, establish severity of coronary disease, screen people at high risk of coronary disease, check effectiveness of antianginal drugs, screen other adults (especially males) before they begin strenuous exercise or activity programs. Very high safety rate, identification ability of cardiac problems that do not show up at rest or with moderate activity, reliable results, noninvasive, simulates stress to heart in everyday activity, easy to repeat, less expensive than thallium stress test are the advantages of stress test. This test cannot be used on patients with abnormal resting ECG. It has also relatively high false positive rate and high false negative rate and can only be used with individuals that are able to perform strenuous exercise on a treadmill or exercise by bicycle. It is available at hospitals, many cardiologists' offices and exercise training facilities.

Thallium stress test is performed after standard exercise stress test. Patient is injected with thallium radioisotope then laid on a table while a gamma-detection camera is used to track uptake of the thallium in heart muscle. It can diagnose coronary artery disease, determine extent of diagnosed coronary artery disease, assess effectiveness of angioplasty and evaluate patient with abnormal ECG. Thallium stress test is able to measure the percentage of heart muscle not receiving sufficient oxygen and identify problems with heart's blood supply during exercise for patients who have no ECG changes or symptoms. It has also low false positive and false negative rate and can identify more cases of undetected heart disease using standard stress test. Beside these advantages, thallium stress test also has disadvantages. It is time consuming, expensive and needs injection of the thallium radioisotope.

Other noninvasive test is echocardiography which is obtained by reflecting high frequency sound waves off various structures of the heart, then transforming the reflected waves into two dimensional images. This test provides information about the size and shape of heart and how well heart chambers and valves are working. A diagnosis of coronary artery disease usually can be made without the use of this test. Other tests are based on imaging technique such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) but they are rarely used for diagnosis of coronary artery disease.

If the cardiologist suspects the patients to have coronary artery disease based on aforementioned diagnostic information, he or she will perform invasive test called coronary angiography. This test is considered as "gold standard" compared to other methods of diagnosing coronary artery disease are compared. To perform this diagnosis, a small tube (catheter) is brought forward into and around the heart through an artery or vein in the groin, arm or neck in order to measure pressures within the heart and produce angiogram or the coronary arteries. It provides precise anatomic information and it is reliable but invasive. It still has small risk of artery blockage, embolism or heart attack.

2.2.2 Computer methods for the diagnosis of coronary artery disease

Development of computer methods for the diagnosis of heart disease attracts many researchers. At the earlier time, the use of computer is to build knowledge based decision support system, which uses and transfers knowledge from medical experts into computer algorithms manually. This process is time consuming and really depends on medical expert's opinion which may be subjective. To handle this problem, machine learning techniques have been developed to gain knowledge automatically from examples or raw data. Machine learning can be defined as a study of computer algorithms that improve automatically through analysis of data [38].

Statistical and probabilistic methods have been used for machine learning purpose. One of the examples of statistical method is Bayesian model. Artificial Intelligence (AI) methods attracted many researchers as machine learning method to develop medical decision support system. Another emerging method namely soft computing which has overlap definition with AI also gains much attention of many researchers. In this section, research works on the application of machine learning to the diagnosis of CAD are reviewed.

Detrano, et al, built a new discriminant function model for estimating probabilities of angiographic coronary disease [39]. Their model is derived from 303 patients undergoing angiography at Cleveland Clinic in Cleveland, US, and applied to 425 patients undergoing angiography at the Hungarian Institute of Cardiology in Budapest, Hungary, 200 patients undergoing angiography at the Veterans Administration Medical Center in Long Beach, US and 143 patients from the University Hospital in Zurich and Basel, Switzerland. 74 to 82 percent of accuracy is obtained. This discrimination function operates based on logistic regression which is not easily interpretable.

Modeling of heart disease using Bayesian network (also called belief network) is proposed by Jayanta and Marco [10][40]. The limitation of this model is that the model should be treated as prototype. The model does not consider any interactions, which may be synergistic or antagonistic between variables, as well as any unknown variables that might influence the presence of CAD. The data consists of 167 objects, in which more data should be obtained and used. In spite of its advantages, the Bayesian approach has also several drawbacks. The difficulty to find accurate conditional probabilities is one of its drawbacks. With the condition of unavailability of enough data, the missing probabilities sometimes must be predicted by experts subjectively. Another drawback is that the approach can be computationally expensive, particularly when the variables being examined are unconditionally independent of one another.

Aha, et al, proposed instance based learning algorithms [41]. Their method generates classification predictions using only specific instances. This approach improved nearest neighbor algorithm which has large storage requirements. Reduced storage requirement and better accuracy has been obtained. Their method has 78 and 80.5 percent of accuracy for Cleveland and Hungarian data sets from University California Irvine [42].

The use of Artificial Neural Network (ANN) to diagnose coronary heart disease has been proposed by Bologna, et al [14]. Their study is using non invasive medical information to build decision support system. One of the drawbacks of ANN based classifier is lack of interpretability. The new approach is by making Multi Layer Perceptron (MLP) type of interpretable ANN. The interpretability is by extracting some rules from the MLP using novel algorithm. Their method is called Interpretable Multi Layer Perceptron (IMLP). The input data consists of 12 continuous and 4 binary variables. The number of extracted rules from IMLP is 46 to 78 rules. C4.5, a rule induction algorithm developed by Quinlan [43], is compared to IMLP. Quantitatively, IMLP has better prediction accuracy but has bigger number of rules than that of C4.5. Normally, ANN based classifiers will have higher accuracy than that of induction rule based classifiers such as C4.5. One of the disadvantages of ANN is that ANN can not give interpretable or transparent knowledge directly. It needs some algorithm to convert the knowledge to become transparent or interpretable. The extracted rules are based on the architectures and parameters of ANN which are determined by user. Another disadvantage of ANN is that the result of training processes of ANN is not the same for every different weight initialization. Qualitatively, C4.5 is simpler than IMLP. The evaluation of IMLP is based on prediction accuracy and rule number. More comprehensive evaluation should be carried out.

Another work is proposed by Yan, et al, by using MLP to build decision support system for the diagnosis of five major heart diseases [9]. CAD is one of them. The MLP system uses 40 input variables that are classified into four groups. The number of neurons in the hidden layer is determined using cascade learning process. Those five categories of heart diseases are represented by five output neurons. The missing data is imputed using mean method. Their system is trained and tested using 352 medical data records collected from the patients suffering from five heart diseases. Cross validation, hold-out and bootstrapping techniques are used to evaluate generalization of their system. Although their results give very high diagnosis accuracy, the system acts like black box. The knowledge is not transparent. The transparency of knowledge is very important for physician in order to understand the relationship between input and output variables. Decision support system should be able to tell the reasoning of making decision so that junior physician can learn the knowledge from it. The way of handling missing data also has drawback. Imputing the mean would make bias in original data set. The characteristics of original data can be changed because of inappropriate imputation method.

Research work on rough set theory (RST) to model prognostic power of cardiac tests has been proposed by Komorowski and Ohrn [12]. The work explores and identifies the need of a scintigraphic scan of a patient group using rough set approach. The work is not aimed to diagnosis the presence of CAD but it is aimed to predict cardiac death or infarction (whether the patient will die or have myocardial infarction) without the need of scintigraphic scan. Standard RST method is used with dynamic reduct algorithm for reduct calculation. The rule extraction process is not fully performed by them. Their work needs further study to obtain the useful and compact rules automatically. There is a disadvantage using only RST. The extracted rules are in crisp form which can not handle numerical input without discretization. The number of rules that is proportional to the number of data is another disadvantage of RST. The number of rules will go to unreasonable number when the training data is large. A big number of rules are difficult to understand and interpret. Rule filtering or pruning should be undertaken to overcome the problem.

Hybrid system of rough set and neural network has been proposed by Liping and Lingyun [44]. Their work uses rough set theory to reduce the number of input to ANN. Then, ANN is used as a classifier to diagnose the presence of CAD. Cleveland CAD data set from UCI data repository [42] is used as training and testing process. The aim of combining one or more AI techniques is to overcome the disadvantage of individual AI. ANN based decision support system usually has high prediction accuracy but it can not give interpretable knowledge or it acts like black box. On the other side, rule induction based decision support systems, such as rough set, have interpretable and explainable knowledge but their accuracy is lower than that of the ANN. They use ANN as the classifier. It means the knowledge is not transparent. The drawback of ANN which is a black box is not covered by rough set. Rough set only covers the input dimensionality reduction of ANN. Their work also uses mean imputation for numerical missing data and most common value for categorical missing data. These imputation methods will make the characteristic of original data altered.

Gamberger, et al, proposed Inductive Learning by Logic Minimization (ILLM) [11]. The aim of using machine learning technique is to find the important and useful information extracted from medical data. In this case, the accuracy of diagnosis is not taken into account. ILLM is proposed for this purpose. ILLM generates rules based on minimization with two classes and propositional rules. It starts from a set of literals that is implicitly defined and generated from the example set. Unlike other rule inductive algorithms, ILLM does not use statistic or probability concept. Every confirmation rule must be a simple conjunction of literals to form a set of confirmation rules with the condition that every rule must be true for examples of the target class only, and every confirmation rule must be true for at least predefined number of target class examples defined by the selectable support level. Support level is used to determine the number of induced rules in the set, so that all confirmation rules that meet this condition are included into the set [45]. Since the aim is only rule extraction, the quality of prediction is not covered in the study. In medical decision support system, the quality of diagnosis or prediction is important as well as usefulness of the generated knowledge.

There is another work on the diagnosis of CAD proposed by Polat, et al [46]. The work uses Artificial Immune Recognition System (AIRS) combined with fuzzy resource allocation mechanism and k-NN (nearest neighbor) based weighting preprocessing. AIRS is based on resource-limited Artificial Immune Systems. It uses k-NN method to calculate the value of parameter called affinity. The work improves AIRS by weighting preprocessing method based on fuzzy resource allocation and k-NN. The system is then improved by fuzzy weighted preprocessing and able to gain better accuracy [47]. Although these AIRS based diagnosis can give very high prediction accuracy, their transparency is not as well as rule based classifiers. The work does not cover the method of missing data handling. Hence, only Cleveland CAD data set is used despite of other sources of CAD data sets taken from UCI [42].

Most of the rule based diagnosis uses decision rules. There is a research work proposed by Ordonez that uses association rules instead of decision rules for heart disease prediction [48]. The drawback of association rules is the extremely large number of rules when they are applied on medical data set. Rarely association rules are validated using independent sample. His work uses training and testing approach to discover association rule for the prediction of heart disease. Rules evaluation metrics such as support, confidence and lift are used to find the medical significance of discovered rules. Two phases rule constraints are proposed in the study to limit the number of rules. The first phase is applying filtering constraints based on predictive goal and size to frequent itemsets. The second phase is applying antecedent-consequent constraint on rules. Like other rule based classifiers, this method still produces crisp rules that will produce crisp output prediction which can not give the possibility degree of disease. The missing data issue is not covered in his study assuming that all the data are complete. It is very difficult to find the large number of complete medical data. The machine learning system will be robust enough when the large number of data is available.

Most of knowledge discovery and data mining techniques based on rule induction result in crisp mode rules. Fuzzy logic method is known as a method to deal with uncertainty of variables. Fuzzy logic can also deal with numerical or continuous value directly. However the crisp rule extracted from data set by rule induction algorithm can be fuzzified. This fuzzified version of extracted rules can deal with numerical data directly. This hybrid method can cover drawback of single method. Fuzzy logic obtains its knowledge from experts. Expert will provide the knowledge to fuzzy system in order to produce appropriate decision making. This is one of the drawbacks of original fuzzy system.

Fuzzy logic alone can not learn directly from raw data. This disadvantage can be covered by rule induction method that can learn directly from raw data. The drawback of rule induction which need discretization and always gives crisp rules is covered by fuzzy logic.

The combination of rough set and fuzzy set has been proposed by Anderson, et al, to perform prognostic modeling of progression of coronary artery disease [17]. Their system is to predict progression of atherosclerosis in the Left Anterior Descending (LAD) artery. Using stepwise logistic regression, four input variables are chosen from 27 variables. The input attributes are type of myocardial infarction, fasting glucose level, total glucose level, and maximum percent of stenosis of LAD. The output is the maximum percent stenosis in the LAD had increased 5% or more. Fuzzy rules are generated using LERS (Learning from Examples based on Rough Sets) algorithm [49]. Compared to rough set classifier, neural network and logistic regression, their system performs better. Fuzzification of crisp rough set rules, has improved the accuracy. Their system uses limited number of cases. There is no guarantee that the system will work for every coronary artery disease problems. Beside, the system does not consider missing attribute values.

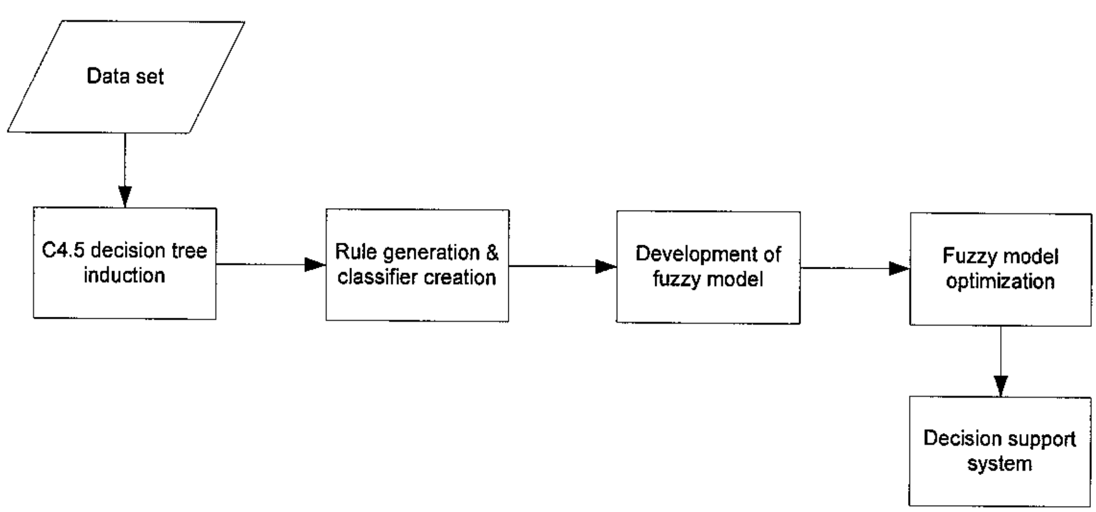


Figure 2.4 Methodology to create a system for diagnosis of CAD

There is a research work on automated diagnosis on CAD based on rule induction and fuzzy modeling which is proposed by Tsipouras, et al [15][16]. The rule induction method that is used to extract rules indirectly is C4.5 algorithm. First, decision tree induction method namely C4.5 [43] is used to form decision tree from training data. Then decision rules generated based on the decision tree. The rules are in a disjunctive normal form. Based on the rules, fuzzy model is developed by transforming the crisp decision rules into fuzzy. Their methodology is illustrated in Figure 2.4.

The used data set has nineteen attributes which are age, sex, family history, smoking, diabetes mellitus, hypertension, hyperlipidemia, creatinine, glucose, total cholesterol, high density lipoprotein (HDL), triglyceride, body mass index (BMI), waist, heart rate, systolic and diastolic blood pressure, carotid femoral pulse wave velocity and augmentation index. The overall accuracy from ten fold cross validation is 73.4 percent. The fuzzification of crisp rules can increase the accuracy from 61.8 to 65.3 percent.

The system uses C4.5 algorithm which is classified as decision tree algorithm. This decision tree algorithm has drawbacks. The approach of decision tree method is top-down and recursive partitioning where the number of records becomes smaller as the tree is traversed down. The small number of records may be too small to make a statistically significant decision about the class representation of the nodes. Another drawback of decision tree method is that a sub tree can be reproduced more than once. This phenomenon can lead to complexity and redundant trees. Hence, the tree will be difficult to interpret [50].

The rule generation from raw data using C4.5 is indirect. The C4.5 produces decision tree and not rules. It must be converted into rules in order to be used by fuzzy system classifier. The decision tree produced in the study uses binary split method to deal with numerical attributes instead of multi-way split or discretization of numerical attributes. This method will make the rules converted from tree difficult to read. This binary split can occur several times along a path in a tree and may give different split. The fuzzy rules have only binary split for numerical attributes. Thus the membership function will only have two conditions, e.g. low and high.

The system uses high quality data set but the availability of this high quality data set is not always easy to obtain. The study does not consider missing data handling. In the real practice, it is difficult to collect. Complete medical data are still the issue of the quality of medical decision support system.

However, the system is very good in artificial intelligent point of view. The system learns directly from raw medical data using C4.5 decision tree. The extracted knowledge from the tree is converted to rules. Fuzzy modeling of the rules make the decision support system can handle the vagueness of medical data. The drawback of fuzzy system which can not learn directly from raw data is covered by C4.5 algorithm. The drawback of C4.5 which is built in a crisp manner is covered by fuzzy system.

The role of C4.5 can be replaced by Rough Set Theory (RST). RST proposed by Pawlak can learn from raw data and generate rules based on reducts [19][20]. The theory is based on indiscernibility relation. The way of RST generates rules is different from underlying theory of C4.5 which is based on entropy and information gain. Unlike statistical approach which needs probability from sufficient samples, RST does not need any preliminary or additional information about data [51]. The resulting rules by RST can be used to model the fuzzy decision support system.

2.3 ROUGH SET THEORY

Rough set theory (RST) is a mathematical and artificial intelligent technique developed by Zdzislaw Pawlak, Warsaw University of Technology, in the early 1980 [19].

RST is especially useful to discover relationships in data. The discovering of relationship in data is called knowledge discovery or data mining. The result of knowledge discovery is understandable and meaningful knowledge extracted from data. RST method emerged as mathematical tool to manage uncertainties, ambiguity and vagueness from incomplete, inexact and noisy information [52]. Rough sets mean approximate sets. Rough set theory deals with approximation of sets. It uses upper and lower approximations when the available information insufficiently can not define the exact value of the set. Based on approximation of sets, RST is suitable to deal with uncertainties of decision making [21].

Rough set theory has been applied to various areas of knowledge for example expert system, machine learning, image processing, pattern recognition, knowledge discovery and control systems [21]. Because RST has not widely known yet, the basic concept of RST is explained in this following section.

2.3.1 Basic concept of rough set theory

This section gives an overview of the basic concepts and basic understanding of rough set theory for the subsequent discussion about application of rough set to build fuzzy decision support system [51]. The reader should have basic knowledge about classical set theory in order to understand the underlying concept of rough set theory.

Rough set theory proposes a novel mathematical approach to deal with incomplete knowledge [19]. The approach of rough set theory to vagueness is expressed by boundary region of a set.

Suppose U represents a finite set of objects and $R \subseteq U \times U$ represents a binary relation. The sets U is called *universe* and relation R is called *indiscernibility relation*. R is assumed to be an equivalence relation. *Approximation space* is represented as $S = (U, R)$.

Let X be a subset of U which is represented as $X \subseteq U$. X will be characterized with respect to R. The equivalence class of R determined by element x is denoted as $[x]_R$. This class represents an elementary component of knowledge that can be perceived due to R. This elementary component is called *granule*. This indiscernibility relation can not be used to observe individual objects from U but it can be used to observe the granules of knowledge described by this relation.

The set of all objects that are certainly classified as members of X with respect to R is called *R-lower approximation* of a set X with respect to R, and denoted by $\underline{R}X$ which is defined as:

$$\underline{R}X = \{x \mid [x]_R \subseteq X\}\,. \qquad (2.1)$$

The set of all objects that are possibly classified as members of X with respect to R is called *R-upper approximation* of a set X with respect to R, and denoted by $\overline{R}X$ which is defined as:

$$\overline{R}X = \{x \mid [x]_R \cap X \neq \varnothing\} \tag{2.2}$$

The set of all objects that can not be classified as members of X or non member of X denoted by *U-X or –X* with respect to R is called *boundary region* of a set X with respect to R, and denoted by $BR_R(X)$ which is defined as:

$$BR_R(X) = \overline{R}X - \underline{R}X \tag{2.3}$$

Using the above representation and notion, rough set definition can be made. A set X is called *crisp* with respect to R if and only if the boundary region of X is empty. Otherwise, a set X is called *rough* with respect to R if and only if the boundary region of X is non-empty.

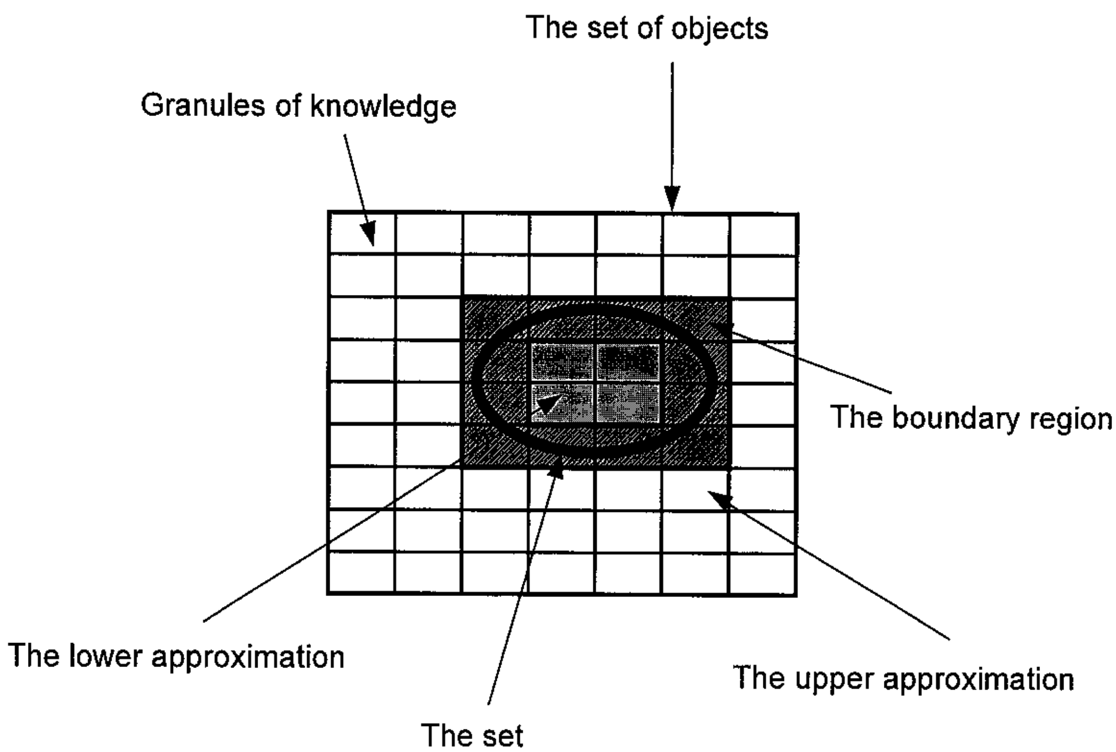


Figure 2.5 Graphical illustrations of the set approximations

The above definition of set approximations can be illustrated into graphical representation as shown in Figure 2.5 [19].

For $X \subseteq U$ and $Y \subseteq U$, the approximations have the following properties [51]:

i. $\underline{R}X \subseteq X \subseteq \overline{R}X$

ii. $\underline{R}\varnothing = \overline{R}\varnothing = \varnothing$; $\underline{R}U = \overline{R}U = U$

iii. $\overline{R}(X \cup Y) = \overline{R}X \cup \overline{R}Y$

iv. $\underline{R}(X \cap Y) = \underline{R}X \cap \underline{R}Y$

v. $\underline{R}(X \cup Y) \supseteq \underline{R}X \cup \underline{R}Y$

vi. $\overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y$

vii. $X \subseteq Y \rightarrow \underline{R}X \subseteq \underline{R}Y \;\&\; \overline{R}X \subseteq \overline{R}Y$

viii. $\underline{R}(U - X) = U - \overline{R}X$

ix. $\overline{R}(U - X) = U - \underline{R}X$

x. $\underline{R}\underline{R}X = \overline{R}\underline{R}X = \underline{R}X$

xi. $\overline{R}\overline{R}X = \underline{R}\overline{R}X = \overline{R}X$

Based on the above concepts, four classes of rough sets as four categories of vagueness can be defined using following conditions [52]:

i. $\underline{R}X \neq \varnothing$ and $\overline{R}X \neq U$, then a set X is *roughly R-definable.*

ii. $\underline{R}X = \varnothing$ and $\overline{R}X \neq U$, then a set X is *internally R-undefinable* or *externally R-definable.*

iii. $\underline{R}X \neq \varnothing$ and $\overline{R}X = U$, then a set X is *externally R-undefinable* or *internally R-definable.*

iv. $\underline{R}X = \varnothing$ and $\overline{R}X = U$, then a set X is *totally R-undefinable* or *totally R-nondefinable.*

The coefficient defined as

$$\alpha_R(X) = \frac{|\underline{R}X|}{|\overline{R}X|} \tag{2.4}$$

can be used to characterize numerically a rough set X. This coefficient is called the *accuracy of approximation,* where $|X|$ denotes the cardinality of $X \neq \varnothing$.

The value of $\alpha_R(X)$ is in the range of 0 to 1. A set X is crisp with respect to R if $\alpha_R(X) = 1$, and otherwise a set X is rough if $\alpha_R(X) < 1$.

Rough set can be also defined by using a rough membership function instead of approximations [53].

The *rough membership function* measures the degree of relative overlap between the set X and the equivalence class $[x]_R$ to which x belongs which is defined as follows:

$$\mu_X^R : U \to \langle 0,1 \rangle \tag{2.5}$$

where

$$\mu_X^R(x) = \frac{|X \cap [x]_R|}{|[x]_R|}$$

The rough membership function can be used to define the approximations and boundary region of a set as shown in the following equations:

$$\underline{R}X = \{x \mid \mu_X^R(x) = 1\} \tag{2.6}$$

$$\overline{R}X = \{x \mid \mu_X^R(x) > 0\} \tag{2.7}$$

$$BR_R(X) = \{x \mid 0 < \mu_X^R(x) < 1\}. \tag{2.8}$$

For $X \subseteq U$ and $Y \subseteq U$, the rough membership function has the following properties

i. $\mu_X^R(x) = 1$ iff $x \in \underline{R}X$

ii. $\mu_X^R(x) = 0$ iff $x \in U - \overline{R}X$

iii. $0 < \mu_X^R(x) < 1$ iff $x \in BR_R(X)$

iv. $\mu_{U-X}^R(x) = 1 - \mu_X^R(x)$ for any $x \in U$

v. $\mu_{X \cup Y}^R(x) \geq \max(\mu_X^R(x), \mu_Y^R(x))$ for any $x \in U$

vi. $\mu_{X \cap Y}^R(x) \leq \min(\mu_X^R(x), \mu_Y^R(x))$ for any $x \in U$

Rough sets can thus approximately depict sets of patients, events, outcomes and others that may be difficult to confine.

2.3.2 Rough set data analysis

In this section, rough set theory is applied to data analysis based on the concepts of section 2.3.1.

Data sets are represented in a tabular form as rows and columns. Each element of a row may represent an object or an instance which can be an event, a patient, etc. Every column represents an attribute which can be a variable, an observation, a property, etc. that can be measured for each object. Such table can be called *information system*. In a formal manner, an information system is a pair $S = (U, A)$ where U is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes such that $a: U \rightarrow V_a$ for every $a \in A$. The set V_a is called the value set of a.

In practical use of information system, there is an outcome of classification that is given as decision and expressed by single special attribute. This type of information system is called *decision system*. Formally, a decision system sometimes called decision table is information system with the form $S = (U, A \cup D)$ where $D = \{d\}$ and $d \notin A$ is decision attribute or simply decision. The elements of A are called conditional attributes or simply conditions.

Data set usually contains large part of objects with their attributes values. There are two issues in decision system. The first issue is that the same indiscernible objects may be represented more than one time. The second issue is that the attributes may be redundant or superfluous.

Let $S = (U, A)$ be an information system, and $B \subseteq A$. A binary relation denoted by $IND_S(B)$ is defined as

$$IND_S(B) = \{(x, x') \in U \times U \mid \forall a \in B, a(x) = a(x')\} \tag{2.9}$$

This binary relation is called the *B-indiscernibility relation* which is an equivalence relation. If $(x, x') \in IND_S(B)$ then objects x and x' are indiscernible or undistinguishable from each other using only attributes from B. The equivalence classes of this relation are denoted by $[x]_B$.

For $X \subseteq U$, a set X can be approximated using only information contained in the set of attributes B. This is can be done by using approximation methods as shown in (2.1) to (2.3). Then, there are *B-lower approximation* and *B-upper approximation* of X which are denoted by $\underline{B}X$ and $\overline{B}X$ respectively, where $\underline{B}X = \{x \mid [x]_B \subseteq X\}$ and $\overline{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$. Using the same analogy as in general case in (2.1) to (2.3), the objects which are members of $\underline{B}X$ can be certainly classified as members of X on the basis of knowledge in B, while the objects which are members of $\overline{B}X$ can be possibly classified as members of X using only the knowledge in B. The set $BR_B(X) = \overline{B}X - \underline{B}X$ which is called *B-boundary region of X*, consists of those objects which can not be classified as members of X or non members of X using only the knowledge in B. The set $U - \overline{B}X$ which is called *B-outside region of X*, consists of those objects which can be certainly classified as non-members of X using only knowledge in B. This set is said to be rough if the boundary region is non-empty.

For $X \subseteq U$ and $Y \subseteq U$, the approximations have the following properties:

i. $\underline{B}X \subseteq X \subseteq \overline{B}X$

ii. $\underline{B}\emptyset = \overline{B}\emptyset = \emptyset$; $\underline{B}U = \overline{B}U = U$

iii. $\overline{B}(X \cup Y) = \overline{B}X \cup \overline{B}Y$

iv. $\underline{B}(X \cap Y) = \underline{B}X \cap \underline{B}Y$

v. $\underline{B}(X \cup Y) \supseteq \underline{B}X \cup \underline{B}Y$

vi. $\overline{B}(X \cap Y) \subseteq \overline{B}X \cap \overline{B}Y$

vii. $X \subseteq Y \rightarrow \underline{B}X \subseteq \underline{B}Y \;\&\; \overline{B}X \subseteq \overline{B}Y$

viii. $\underline{B}(U - X) = U - \overline{B}X$

ix. $\overline{B}(U - X) = U - \underline{B}X$

x. $\underline{B}\underline{B}X = \overline{B}\underline{B}X = \underline{B}X$

xi. $\overline{B}\overline{B}X = \underline{B}\overline{B}X = \overline{B}X$

Based on the above concepts, four classes of rough sets as four categories of vagueness can be defined using following conditions:

i. $\underline{B}X \neq \varnothing$ and $\overline{B}X \neq U$, then a set X is *roughly B-definable.*

ii. $\underline{B}X = \varnothing$ and $\overline{B}X \neq U$, then a set X is *internally B-undefinable* or *externally B-definable.*

iii. $\underline{B}X \neq \varnothing$ and $\overline{B}X = U$, then a set X is *externally B-undefinable* or *internally B-definable.*

iv. $\underline{B}X = \varnothing$ and $\overline{B}X = U$, then a set X is *totally R-undefinable* or *totally B-nondefinable.*

The coefficient defined as

$$\alpha_B(X) = \frac{|\underline{B}X|}{|\overline{B}X|} \tag{2.10}$$

can be used to characterize numerically a rough set X. This coefficient is called the *accuracy of approximation,* where $|X|$ denotes the cardinality of $X \neq \varnothing$.

The value of $\alpha_B(X)$ is in the range of 0 to 1. A set X is crisp with respect to B if $\alpha_B(X) = 1$, and otherwise a set X is rough if $\alpha_B(X) < 1$.

Rough set can be also defined by using a rough membership function instead of approximations.

The *rough membership function* measures the degree of relative overlap between the set X and the equivalence class $[x]_B$ to which x belongs which is defined as follows:

$$\mu_X^B : U \to [0,1] \tag{2.11}$$

where

$$\mu_X^B(x) = \frac{|X \cap [x]_B|}{|[x]_B|}$$

The rough membership function can be used to define the approximations and boundary region of a set as shown in the following equations:

$$\underline{B}X = \{x \mid \mu_X^B(x) = 1\} \tag{2.12}$$

$$\overline{B}X = \{x \mid \mu_X^B(x) > 0\} \tag{2.13}$$

$$BR_B(X) = \{x \mid 0 < \mu_X^B(x) < 1\}. \tag{2.14}$$

For $X \subseteq U$ and $Y \subseteq U$, the rough membership function has the following properties

i. $\mu_X^B(x) = 1$ iff $x \in \underline{B}X$

ii. $\mu_X^B(x) = 0$ iff $x \in U - \overline{B}X$

iii. $0 < \mu_X^B(x) < 1$ iff $x \in BR_B(X)$

iv. $\mu_{U-X}^B(x) = 1 - \mu_X^B(x)$ for any $x \in U$

v. $\mu_{X \cup Y}^B(x) \geq \max(\mu_X^B(x), \mu_Y^B(x))$ for any $x \in U$

vi. $\mu_{X \cap Y}^B(x) \leq \min(\mu_X^B(x), \mu_Y^B(x))$ for any $x \in U$

One of the important problems in data analysis is to find dependency among attributes. A set of attributes D is said to be dependent on a set of attributes C if all values of attributes D are uniquely determined using values of attributes C. D depends on C denoted by $C \Rightarrow D$.

The dependency D on C, denoted by $\gamma_C(D)$, is a plausible measure of the dependency degree D on C and defined as follows:

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|} \tag{2.15}$$

where $POS_C(D) = \bigcup_{X \in U/D} \underline{C}X$ called a *positive region* of partition U/D with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition U/D using knowledge from C. The value of dependency is between 0 and 1.

i. If $\gamma_C(D) = 1$: D totally dependent on C, i.e., C functionally determines B.
ii. If $\gamma_C(D) = 0$: C and D are totally independent of each other.
iii. If $\gamma_C(D) < 1$: D is roughly dependent on C.

In general, the dependency of D on C in a degree of γ can be denoted by $C \Rightarrow D_\gamma$. The concept of dependency of attributes is used to find the redundant or superfluous attributes. For information system $S=(U,A)$ and $B \subseteq A$, and $a \in B$, it can be said that a is dispensable in B if $IND_S(B) = IND_S(B - \{a\})$, otherwise a is indispensable. A set B is said to be independent if all its attributes are indispensable. Any subset B' of B is called a *reduct* of B if B' is independent and $IND_S(B') = IND_S(B)$.

Reduct can be defined as the minimal subset of attributes that have the same classification of elements as the universe which is the whole set of attributes. In other words, attributes that is not the element of reduct are redundant with respect to classification of elements of the universe. There are several important properties of reduct. The core of attributes is one of the important properties of reduct defined by:

$$CORE(B) = \bigcap RED(B), \tag{2.16}$$

where $B \subseteq A$ and *RED*(*B*) is the set of all reducts of *B*. The elements of core are always in every reducts. Hence, the core is the most important subset of attributes. None of core's elements can be removed without affecting the classification ability of attributes.

There is a dependency $C \Rightarrow D$ given. It could happen that the set *D* depends not on the whole attributes of set *C* but only on the subset of *C* which is denoted by *C*'. In this case finding *C*' is interesting.

To find *C*', suppose $C, D \subseteq A$. Evidently, if $C' \subseteq C$ is *D*-reduct of *C*, then *C*' is minimal subset of *C* such that $\gamma_C(D) = \gamma_{C'}(D)$. The attribute $a \in C$ is *D*-dispensable in *C* if $POS_C(D) = POS_{C-\{a\}}(D)$. Otherwise, the attribute *a* is *D*-indispensable in *C*. If all attributes $a \in C$ are *D-indispensable* in *C*, then *C* is *D*-independent.

A subset $C' \subseteq C$ is a *relative reduct*, in this case called *D-reduct* of *C* if and only if *C*' is *D*-independent and $POS_C(D) = POS_{C'}(D)$.

The term *D-core* of *C* is the set of all *D*-indispensable attributes in *C*. It is denoted by $CORE_D(C)$ which has properties:

$$CORE_D(C) = \bigcap RED_D(C), \tag{2.17}$$

where $RED_D(C)$ represents *D*-reduct of *C* which may be more than one reduct.

Reduct and core are very important concept for feature selection, attribute reduction and rule generation.

2.3.3 Applications and tools of rough set theory

Pawlak, the founder of rough set theory, wrote a review on rough set approach to knowledge-based decision support [20]. The underlying operations of rough set theory are used to discover the essential pattern in data. Thus, rough set methodology has strong relationships with machine learning, knowledge discovery, statistics and inductive reference.

In his paper, Pawlak mentioned that rough set theory has found many interesting applications. Rough set theory has the main advantage that it does not need any external information about data. Rough set theory lets the data speak about themselves. Successful applications of rough set theory in many real life problems in medicine, engineering, banking, financial and market analysis and others have been found in many technical papers [20].

Ziarko's study found that the primary challenge dealing with rough set community today is the development of real life industrial applications of rough set method. Most of rough set papers are theoretical works with a few papers describing practical implementation. Rough set methodologies also have been used in medical domain [54].

Wu, et al, show the recent development of research works using rough set methodology. Recently, many research works deal with combination the rough set theory with other artificial intelligence methods such as fuzzy logic, neural networks and expert systems. Promising results have been achieved [21]. In their paper, rough set theory applications in a number of challenging fields such as pattern recognition and information processing, business and finance, industry and environmental engineering, system faults diagnosis and monitoring, intelligent control systems and other areas are shown and reviewed. One of the challenging areas is medical diagnosis and medical data analysis. Research work on rough set/fuzzy logic application coronary artery disease by Anderson, et al, is found and reviewed in their paper [17][21].

Research using rough set theory in many interesting fields requires experimental tools and verifications. Focusing on main problems of research needs powerful tools and software that automate rough set data analysis and operations. Several tools have been made by many researchers. A brief review of main software systems of rough set data analysis which is freely distributed is introduced here.

Ohrn developed a software ROSETTA that is freely distributed and can be downloaded [55]. ROSETTA software supports complete steps of data mining and knowledge discovery process based on rough set theory. All the steps of data mining such as data preprocessing, missing data handling, discretization, reducts and rules generation, classification and evaluation. Discretization and reducts generation packages take from RSES library are also available [56].

Bazan, et al, developed a software RSES which stands for Rough Set Exploration System [56][57]. RSES software is also downloadable. Like ROSETTA, the system has complete data mining steps and provides other data mining methods which instance based method or k-NN, Local Transfer Function Classifier (LTF-C) and decision tree algorithm.

ROSETTA and RSES are used in rough set data analysis in this thesis because they provide complete processes of data mining, user friendly, freely distributed and can handle large data set with different format.

Another rough set theory software is ROSE which stands for Rough Sets Data Explorer developed by Predki, et al [58][59]. ROSE is not used in this thesis because it has limitation in reduct generation for large data set.

There are other software systems based on rough set theory available such as GROBIAN [60] and LERS [49].

2.4 ARTIFICIAL NEURAL NETWORK

Artificial neural network or simply ANN, is a computational model of the brain. Artificial neurons are interconnected by the link, forming a network called artificial neural network. ANN works similar to the work of the brain. The network received input and internal processes take place. These internal processes are activation of neurons. The network results in output. The use of ANN provides many properties and capabilities such as nonlinearity, input-output mapping, adaptivity, evidential response, contextual information, fault tolerance and uniformity of analysis and design [61]. Although many ANNs models have been proposed, the multi layer perceptron or backpropagation is the most widely used model in terms of practical applications [52]. MLP is also widely used in medical decision support system [9][61]. It has been successfully applied to many difficult problems in many areas. Figure 2.6 shows the topology of MLP.

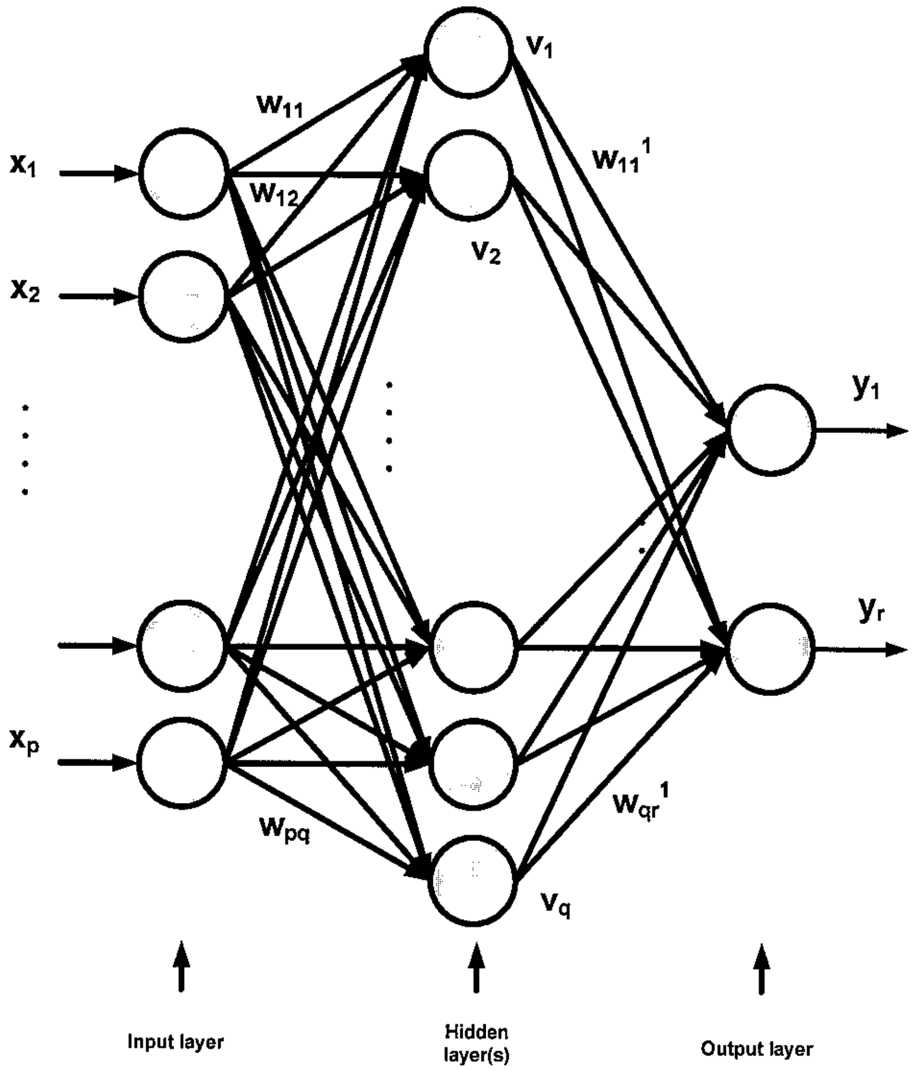


Figure 2.6 Topology of multi-layer perceptron

An MLP is a network of nodes or neurons as processing elements which are arranged into layers. A typical MLP usually consists of three layers: input layer, hidden layer and output layer. Single hidden layer is used to make MLP act as a universal classifier. The neurons are connected between layers by their corresponding weights. The propagation of the signals that pass from input to output can be modelled as:

$$v_k = f_k(\sum_{i=1}^{p} w_{ik} x_i) \tag{2.18}$$

$$y_j = f_j(\sum_{k=1}^{q} w_{kj}{}^1 v_k) \tag{2.19}$$

where $x_1,...,x_p$ are input signal including the bias. w_{ik} is the connection weight between the input x_i and the neuron k. $v_1,...,v_q$ are hidden layer's output signals that will become the input of the output layer $y_1,...,y_r$. w_{kj}^{1} is the connection weight between the hidden layer's output v_k and the neuron j. $f_k(.)$ and $f_j(.)$ are the activation functions of the kth and jth neuron respectively. The sigmoid function is usually used in MLP. One of sigmoid functions is hyperbolic tangent function defined as

$$f(a) = \tanh(a). \tag{2.20}$$

Learning or training in MLP is updating the weight values based on the mean squared error (MSE) between target and output, which are described in the following equations:

$$E = (\tfrac{1}{2})\sum_j (t_j - y_j)^2 \tag{2.21}$$

$$\Delta w_{ij}^{(n)} = -\eta\left(\frac{\partial E}{\partial w_{ij}}\right), \; or \tag{2.22}$$

$$\Delta w_{ij}^{(n)} = -\eta\left(\frac{\partial E}{\partial w_{ij}}\right) + \alpha \Delta w_{ij}^{(n-1)} \tag{2.23}$$

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} + \Delta w_{ij}^{(n)} \tag{2.24}$$

where t_j is the target of learning at neuron j, η is a learning rate parameter and α is a momentum factor. There are many learning methods based on backpropagation (BP). Resilient Backpropagation (Rprop) is a modified BP by considering only the sign of error gradient to indicate the direction of the weight update. The size or weight change is determined by a certain value Δ_{ij}. The typical values of parameters are $\eta^- = 0.5$, $\eta^+ = 1.2$ and $\Delta_{ij}^{(0)} = 0.07$ [62].

Hence, the weight increment update becomes:

$$\Delta w_{ij}^{(n)} = \begin{cases} -\Delta_{ij}^{(n)}, if \frac{\partial E^{(n)}}{w_{ij}} > 0 \\ +\Delta_{ij}^{(n)}, if \frac{\partial E^{(n)}}{w_{ij}} < 0 \\ 0, else \end{cases} \tag{2.25}$$

The new update value $\Delta_{ij}^{(n+1)}$ can be computed as:

$$\Delta_{ij}^{(n+1)} = \begin{cases} \eta^{+} \times \Delta_{ij}^{(n)}, if \frac{\partial E^{(n)}}{w_{ij}} \times \frac{\partial E^{(n+1)}}{w_{ij}} > 0 \\ \eta^{-} \times \Delta_{ij}^{(n)}, if \frac{\partial E^{(n)}}{w_{ij}} \times \frac{\partial E^{(n+1)}}{w_{ij}} < 0 \\ \Delta_{ij}^{(n)}, else \end{cases} \tag{2.26}$$

where $0 < \eta^{-} < 1 < \eta^{+}$.

2.5 MISSING ATTRIBUTE VALUES IN THE FIELD OF DATA MINING

Data mining or Knowledge discovery from data (KDD) processes usually encounter missing data problem. The source of problem may be from collecting processes due to the device and human error. The quality of data is very important factor for KDD to reach good quality. There are two approaches to solve this problem. The first is to impute the missing values before using them in knowledge discovery process. The second is modifying the knowledge discovery methods in order to be able to handling the missing data without any imputation [63].

Many methods to deal with missing data have been proposed. Grzysmala-Busse and Hu compared nine approaches to deal with missing data in data mining [22]. Their work has used ten input data files.

The nine approaches are:

i. Most common attribute value.
ii. Concept or decision based most common attribute value.
iii. Decision tree, namely C4.5.
iv. Method of assigning all possible values of the attribute.
v. Method of assigning all possible values of the attribute restricted to the given concept (decision).
vi. Method of ignoring examples with unknown attribute values.
vii. Event-covering method.
viii. A special LEM2 (Learning from Example Module version 2) algorithm.
ix. Method of treating missing attribute values as special values.

All the nine approaches are used to impute all ten data sets. Then the original data sets and imputed data sets, except C4.5 methods, are sampled into ten pairs of training and testing data. The sampled files then are used as the input of classifiers to generate classification rules. Ten fold cross validation is used to evaluate the classifications. The average error of different methods is calculated for comparison. Their experiment can not draw a conclusion which method is the best imputation because the best imputation only occurs at specific data set and specific classifiers.

The comparison on approaches to assign missing attribute values was also studied by Li and Cercone [24][64]. They proposed a new approach RSFit on processing data with missing attribute values based on rough set theory and distance based method. The core of the data set is calculated except when there is no core then a reduct is used. The new decision table based on core or reduct can be built. To impute the target of missing attribute value, the non target attributes contains missing values are included. Two approaches based on decision class are used. The first approach is that all the data instances are involved during distance calculation. The second approach is that only the data that belong to the same decision class is used. The distance is calculated where the missing values distance is assigned as 1 which is the maximum difference between unknown values. The selected value which will be assigned to the target of missing value has the smallest difference. RSFit can not give high accuracy imputation because it does not consider the relationship among attributes inside the data set. Li and Cercone then proposed a combination between RSFit and frequent itemset of association rule technique named ItemRSFit [64]. It uses frequent itemset to find relationship among attributes and impute the missing values. The frequent itemset technique does not cover all the missing values because there is no possible match can be found to impute the missing values. For this aforementioned case, RSFit is used. The frequent itemset technique needs predefined value of support to impute the missing values. Smaller value of support gives better accuracy of imputation. The use of RSFit in ItemRSFit means that not all imputation is based on relationship among attributes.

The use of RST for missing data imputation is also proposed by Nelwamondo and Marwala [65]. By finding relationships between attribute values based on RST without extracting any rules, the missing data can be estimated. This method can only impute the missing case which is similar to or related to another case. If there is no similar or related case then no imputation can be made.

Fessant and Midenet proposed imputation method using Self-Organising Map (SOM) [66]. The method is applied on data in surveys. They use the ability of SOM handling the missing input to build map similar to the map that uses complete input. The resulting map is used to impute the missing input values. The designed SOM is tested using transport survey and compared to Multi Layer Perceptron (MLP) imputation, hotdeck imputation which uses nearest neighbor approach and mean imputation. Their results show that SOM has accuracy that is similar to MLP and hotdeck imputations. In the case of categorical missing values, SOM is better than MLP. However, MLP is better than SOM for numerical missing values. Unlike MLP, SOM does not need complete input attributes to impute the missing values. SOM needs only one model to impute can missing values in every attributes while MLP needs many models to impute missing values in every attributes. SOM needs predefined parameters such as learning steps and the size of maps before applying to impute the missing values. The correct size of map should be determined. The evaluation method is mostly based on error and mean squared error. There is no study the effect of imputation to the quality of knowledge discovery from data in their work. The similar work proposed by Wei and Tang [67]. They proposed a generic framework for missing value imputation based on SOM. The model is applied on credit card company data set. They used two stage imputation algorithms. Mean or most common value filling is applied before using the second stage which is SOM. Since SOM is used then it has similar limitation as Fessant and Midenet work.

Siripitayananon, et al, proposed imputation of missing data in wind speeds using neural network [32]. The neural network used is multi layer perceptron with five input, four hidden nodes and single output. The model is used to estimate the missing wind speeds data at one of the station based on the other stations. Z-score is applied to normalize the input data. The MLP model is compared to nearest neighbor imputation method. Their results show that MLP is better than nearest neighbor method. The input attributes used in their work are only four. The model is not proven to be robust enough for many input attributes. Similar work is done by Bhattacharya, et al [31]. They used MLP to predict missing wave data in sedimentation modeling which is time series data. The model has seven inputs and single output. Their results are not compared to other methods. Bhattacharya and Solomatine compared the MLP missing data imputation to M5 model tree and found that MLP is slightly better than M5 for missing data in sedimentation modeling [68].

Wu, et al, proposed the use of association rules for completing missing data [69]. Association rules from complete data set are extracted. Then the extracted association rules are applied to impute missing values in incomplete data set. The attributes that contain missing values must be rule consequent of corresponding association rules. Their results are compared to Robust Bayesian Estimator (RBE) imputation method and give better results. Only the discrete data assumed to be correct and noise-free are used in their experiment. Predefined parameters which are minimum support and confidence must be determined. The way to determine these parameters is one of their problems. And also not all extracted association rules can cover the missing data. These are limitations of their method. Ragel and Cremilleux proposed similar method to impute missing data based on association rules [27][70]. They split incomplete data set into several valid databases (vdb) instead of using complete data set. Then similar method as what proposed by Wu, et al, is applied.

Wasito and Mirkin combined least squares method and nearest neighbor imputations [25][26][63]. The least squares method used are iterative majorization least squares (IMLS) developed by Kiers based on weighted least squares [71]. Wasito and Mirkin proposed three stages of imputation method. The stages are applying IMLS, then nearest neighbor and then IMLS. The system is tested using multiple type of missing data. Their method is more in generic solution than specific one. Practically the missing data problem is specific in particular type of data.

In the field of bioinformatics, Wang, et al, proposed imputation method based on Support Vector Regression (SVR) and orthogonal coding scheme to estimate missing values for DNA microarray gene expression data [29]. SVR is based on Support Vector Machines (SVM). Radial basis function is used as kernel function. Similar to MLP base imputation, the attribute that contains missing values is assigned as output and the other attributes are assigned as inputs. In case of more than one missing attributes, simple imputation schemes such as zero value, mean and orthogonal coding scheme imputations are applied to the input attributes that contain missing values. Their proposed method is compared to k-NN, Bayesian Principal Component Analysis (BPCA) and Local Least Squares (LLS). The results show that SVR based imputation with orthogonal coding scheme performs better than the others.

Troyanskaya, et al, compared k-NN, Singular Value Decomposition and row average methods to impute missing values in DNA microarrays [28]. Variations of parameters are evaluated over different data sets using 1-20% of simulated missing values. The evaluation results show that k-NN has better performance than other methods.

Similar comparison study over different methods of imputations on medical data is conducted by Giardina, et al [72]. Five missing values imputations which are k-NN and correlation-based models are applied to diabetes data sets. Their results show that correlation-based model namely EMImpute_Columns which is based on correlation between attributes gives assuring method to estimate missing values.

Missing data also occurs in the field of environmental sciences. Junninen, et al, compared several methods which are regression bases imputation, k-Nearest Neighbour (k-NN), SOM, MLP and hybrid combinations of the aforementioned methods [30]. The methods are applied on simulated missing data in air quality data sets. SOM and MLP gave better performance in term of accuracy than other methods. k-NN has better computational speed.

There is other strategy to handle missing values without imputation by modifying the data mining methods to be able to deal with incomplete data sets without any imputation needed. Grzymala-Busse proposed rough set strategies to data with missing entries[73]. With Siddhaye, Grzymala-Busse improved the previous rough set approach for rule induction using data with missing entries [74].

Most of the research works on missing data imputation use simulated values and evaluates their results on accuracy of imputations. The practical use of imputation is on knowledge discovery problem which the quality classification should be taken into account more than just imputation accuracy. The choice of imputation methods should depend on knowledge discovery methods which will be used. In this type of problems, research work by Al Shalabi, et al, is proposed [23]. Their research did not propose imputation methods but they proposed framework to deal with missing data when the knowledge discovery method is induction of classification rules base on rough set theory. The coverage percentage, number of rules and number of reducts are used to evaluate the four missing data handling methods. They draw a conclusion that the best method to deal with missing data depends on the task of knowledge discovery.

Based on aforementioned literature reviews, ANN based imputation gives good performance on accuracy of imputations. RST can be used to reduce the number of input attributes when there are many input attributes for ANN based imputation [75]. The combination of ANN and RST to build classifiers for many domains of applications is found to be success [76-78]. The type of incompleteness of medical data is not missing in random. Usually the missing data occurs from one to several rows of attributes not in random. ANN based imputation is appropriate for this type of missing data. The rough set classifier will be used in this research. Hence evaluation based on classifiers especially RST classifier is suitable. The hybrid ANN and RST for missing data imputation namely ANNRST is proposed in this work. The evaluation of imputation methods using novel evaluation framework based on RST is also discussed.

2.6 RULE FILTERING AND SELECTION

The result of knowledge discovery process can be decision tree, association rules, decision rules, sequential pattern, etc. The most comprehensive and interpretable extracted knowledge is in the form of rules. Some rule induction algorithm such as rough set theory results in large number of rules. This large number makes interpretability of the knowledge becomes low. Lacking of interpretability will cut down the advantages of rule based systems which should be interpretable and easy to understand. The resulting large number of rules is due to noise, redundancy in input and/or training data sets. Rule pruning is the method to reduce the number of rules while maintaining the quality of the system.

Agotnes investigated genetic algorithm and the use of quality measure for individual rule to filter the large number of rules [33]. Predefined function is used to identify "good" rules. Two properties that are commonly used to define the function are accuracy and coverage [79]. Accuracy is used to measure the reliability or accuracy of single rule. Coverage is used to measure the importance or powerfulness of single rule. In general, a rule is said to be good if it is accurate and powerful. The performance is investigated using Receiver Operating Characteristic (ROC) and statistical hypothesis testing. Ten different formula of rule quality are used. Experiments using real data of acute appendicitis and coronary artery disease are conducted. For acute appendicitis, six to twelve rules filtered from 400 to 500 rules are obtained without significantly reducing the performance of the classifiers.

Maddouri and Gammoudi conducted comparative study on semantic properties of rule interestingness measure [34]. More than forty measures of interestingness are investigated. These interestingness measures are then combined with semantic properties of rules to find which of these measures have good semantic properties. They found Zhang measure [34] has most of the semantic properties. There is no measure that can perform constantly better than others in all fields of applications.

Li and Cercone proposed another measure of rules which is called rule importance measure [35]. Based on medical cases, there are some routine exams that must be conducted by the doctor such as age, blood pressure, body temperature, sex, etc. There are some examinations or symptoms that are not always considered for patients by the doctor. Li and Cercone assumed that the most important examinations or symptoms should be included more frequently than the less important ones. Based on these and in term of rule generation, they defined that if a rule is generated more frequently across different rule sets, it can be said that this rule is more important than rules generated less frequently across those same rules. In term of rough set theory, rule importance measure equals the number of times a rule appears in all the generated rules from the reduct sets divided by number of reduct sets. In order to select important rules from large number of rules automatically, the concept of reduct is introduced. The concept of reduct is used by considering extracted rules as attributes of new reconstructed decision table. Because reduct represents important features of decision table, the attributes of reduct can be considered as important rules. If the new reconstructed decision table has a core, then the attributes in the core are considered as the most important rules [64][80][81]. The applications of rule importance measure on different data sets is found in [64][80-82].

Rule importance measure based on reduct calculation is not feasible when the number of rules is very large. Finding the reduct will take high computation resource especially for large number of attributes with large number of objects. To overcome this problem, two stages of rule selection are proposed. The first stage is by using rule quality measure which is support filtering. After getting acceptable number of rules with acceptable accuracy and coverage, rule importance measure is applied. It is worthless to apply rule importance measure on very low quality of rule sets. Li and Cercone applied rule importance measure on association rules and using training data to reconstruct new decision table. In this thesis, modified rule importance measure is applied on decision rules and testing data is used instead of training data to increase the generalization.

2.7 FUZZY DECISION SUPPORT

The types of applications for which fuzzy systems are particularly useful are mainly difficult cases where traditional techniques do not work well. Medical problem is one of such difficult problem. Some people consider field of medicine is more an "art" than a "science" [39]. Fuzzy system approaches also allow us to represent descriptive or qualitative expression such as "high", "medium" or "low' instead of quantitative expression. These qualitative expressions are more natural than mathematical equations for many decisions making.

Fuzzy theory began with paper written by Zadeh in 1965 entitled "Fuzzy Sets" [83]. He named fuzzy sets for sets or classes where their boundaries can not be defined precisely, such as "a class of beautiful women", "a class of tall men" and "a class of real number that is much greater than 1". He stated that imprecisely defined sets play an important role in human reasoning in pattern recognition, semantics communication and abstraction.

Some essential characteristics of fuzzy sets are: exact thinking is considered as special case of approximate reasoning, everything is a matter of degree and every logic system can be fuzzified.

The key concept of fuzziness is that it allows the gradual transition from 0 to 1 rather than crisp value 0 or 1. As an example, ordinary set only allow elements belong to or does not belong to the set. In fuzzy set, everything is a matter of degree which means some elements may belong to the set in a degree of membership values between 0 and 1. In classical set, an element belongs to the set if the membership degree is 1, and does not belong to the set if the membership degree is 0. It does not accept other than 0 or 1. Based on this concept the fuzzy inference system is developed and gained tremendous success of application on various domains.

Siu, et al, developed a fuzzy expert system for turbomachinery diagnosis [84]. Incremental forward chaining is performed to identify and rank the causes of vibration. The rules for fuzzy inference are not obtained automatically. This fuzzy expert system still uses complex analysis of domain experts to build the rules. The system is implemented in FuzzyCLIPS. Similar works in different domain is proposed by Koutsojannis and Hatzilygeroudis by using fuzzy expert system for diagnosis and treatment of male impotence [85]. The same as Siu's work, Koustsojannis and Hatzilygeroudis use expert's knowledge statistical analysis from 70 patients to model the diagnosis process of male impotence, linguistic variables and their values. FuzzyCLIPS is also used in their model. Their diagnosis accuracy is below the diagnosis of expert although it is still better than resident's (trainee specialist) diagnosis. In the field of risk assessment of insurance, fuzzy expert system is proposed by Carreno and Jani [86]. The same as Siu's fuzzy expert system, their system derived its knowledge form expert's investigation of factors that affect risk assessment of life insurance. The results of fuzzy expert system using FuzzyCLIPS is compared to the results obtained from the solution using traditional numerical equations.

Pure fuzzy system can not obtain its knowledge from raw data automatically. This is one of the drawbacks of fuzzy system. Because the knowledge is obtained from domain experts, the subjectivity of experts may give influence to the system built using fuzzy concept. In order to cover this drawback, fuzzy system should be incorporated with other method that can learn directly from raw data and discover knowledge in the form of rules. Then the fuzzy system can use the rules to form inference system. One of the rule discovery methods is rough set theory. Rough set theory can discover rules from raw data without any external parameter and let the data speak for themselves.

Cho, et al, proposed autogeneration of fuzzy rules and membership function to fuzzy system model using rough set theory [87]. The system does not really autogenerate rules from raw data because the data only consist of twenty objects. They used rough set theory concept for inconsistency handling by introducing occupancy degree. Although no optimization is applied, complicated method is still used in their research work to model the fuzzy system. The results are compared to those obtained by Lin and Cunningham [88]. The performance is slightly lower than that is obtained by Lin and Cunningham in their work after optimization.

Drwal and Sikora, proposed rough set based rule generation to build fuzzy decision support system [89]. Tolerance mode rough set theory is used to generate rules from decision table. Pearson rule quality measure was applied to filter the generated rules. Fuzzy decision support was then built and applied to digital fundus eye images to diagnose normal and glaucomatous patients. The result gave 65% classification accuracy and 72% after fuzzification while pure rough set (ROSETTA) gave 60% and C5 (decision tree based classifier) gave 64%. Their works are based on complete data. The filtering method used only Pearson rule quality measure while the importance of rules is not considered. No optimization scheme is applied in their work. Sikora improved his previous work by improving rule generation procedures namely RMatrix and MODLEM including optimization using genetic algorithm [90]. His system was tested in benchmark and industrial data sets. The optimization scheme failed to give significant improvement of accuracy.

In the field of medical diagnosis, the combinations of fuzzy set and rule induction methodology where rule induction method is used to generate fuzzy rule sets are found in [15-17][91][92] . Some of them are explained and reviewed in section 2.2.3.

Based on aforementioned review, it is found that fuzzy system attracts many researchers to build their decision support systems. Although fuzzy system has disadvantage, by combining it with rule induction methods such as rough set theory and decision tree method (C5), the fuzzy based decision support system are successfully applied in many areas with good performance. The combination of fuzzy system and rough set based rule generation is potentially useful for solving the problem in many fields such as coronary artery disease diagnosis.

2.8 SUMMARY

This chapter has reviewed the important problem of coronary artery disease and its diagnosis. Diagnosis of coronary artery disease needs much information in order to draw the correct decision. Practically the step of gathering complete information such as historical data, symptoms and medical test data for the diagnosis are not always easy and economically efficient. Usually the medical test starts from cheap to expensive one. Even though the complete historical data, symptoms and medical test data from patient are available, cardiologist may still get confused to draw conclusion of the presence of coronary artery disease. The angiography is the choice for the cardiologist to make sure that the diagnosis is correct. Although the angiography has risk and is expensive, it is still the standard for the diagnosis of coronary artery disease.

Alternative methods to diagnose coronary artery disease are emerging using advanced computer methods such as artificial intelligence, soft computing and expert systems. These computer based methods are used as decision support system for cardiologists to help them giving the correct diagnosis of coronary artery disease. The decision support system for diagnosis is based on information from patients such as historical data, symptoms and medical tests data. The computer program gives inference method to draw conclusion from patient's data. One of the decision support systems is called evidence based system. This type obtains the knowledge directly from the patient's cases and encompasses medical experts. Rule induction based decision support is example of this method. Numerous research works in this field have been reviewed.

Rough set theory is introduced as the new emerging method and reviewed. It can be applied on attributes reduction as well as rule induction.

To build evidence based decision support system, high quality data sets are needed. Missing data occur in data sets will degrade the quality of data. Artificial neural network based missing data imputations have been proven to be successful for estimating missing values in input attributes. The rough set concept of attribute reduction can be combined with artificial neural network to impute missing attribute values in data sets as proposed in this thesis.

Rough set theory can be used to measure the importance of rules and then may be applied in rule selection problem. Rule quality and importance measure should be used in rule selection problem to give the high quality and important selected rules.

The use of fuzzy theory is very useful for decision making problems. Research works based on fuzzy theory and combined fuzzy and rough set theory for decision support system have been studied. The combination of fuzzy theory and rough set theory to build decision support system for the diagnosis of coronary artery disease is proposed in this thesis.

CHAPTER 3: METHODOLOGY

3.1 INTRODUCTION

The aim of this chapter is to describe and propose a method to develop evidence based decision support system for the diagnosis of coronary artery disease. The system must be able to handle the incomplete and continuous data with high accuracy at acceptable coverage. The system must have transparent knowledge. This chapter starts from the data sets that used in this research. The method to find the appropriate method to impute the missing data is explained. Knowledge extraction method from raw data in the form of acceptable number of decision rules with high accuracy and coverage is proposed. An appropriate method to evaluate the imputation method and extracted knowledge is also introduced. Finally, the development and evaluation methods of decision support system based on the extracted knowledge are explained.

3.2 RELATED WORKS

There are several research papers on the diagnosis of coronary artery disease. A multilayer perceptron based medical decision support system has been developed to diagnose five types of heart diseases such as hypertension, coronary heart disease, rheumatic valvular heart disease, chronic cor pulmonale and congenital heart disease [9]. However artificial neural network itself cannot explain the knowledge inside it, even though the system has high accuracy. Bayesian network model of heart disease is proposed by Jayanta and Marco [10]. The system could predict the probabilities of heart diseases and dependency among attributes related to heart diseases. A set of machine learning methods was evaluated on the atherosclerotic coronary heart disease [11]. The objective was not to compare different machine learning results but to explore the possibilities of both machine learning and medical expertise improving the quality of regular medical practice. Prognosis of cardiac events (cardiac death or non-fatal myocardial infarction) was proposed by Komorowski and Ohrn [12]. Diagnosis of ischemic heart disease using various machine learning techniques was found in [13]. The data of 4000 patients were used. Extension of multi layer perceptron for coronary heart disease diagnosis by making it interpretable was shown in [14]. Fuzzy discrimination analysis for diagnosis of valvular heart disease was also proposed by Watanabe, et al [93]. Most of these research works used large number of patients. A combination of data mining technique based on rule induction, namely C4.5 [43] and fuzzy modelling was used to diagnose CAD [15][16]. This work emphasized on fuzzy modelling of C4.5 generated rules. This system used 19 attributes of 199 objects to predict the presence of 50% or more narrowing at least in one of coronary artery vessels.

In this thesis, the combination of data mining technique and fuzzy modelling is adopted and improved. RST technique is used to discover the knowledge from CAD data sets. The discovered knowledge is in the form of "IF-THEN" rules. Missing data imputation technique is based on ANN and RST with novel evaluation method. RST is used to select important and high quality rules from large number of generated rules after missing data imputation. Finally a fuzzy modelling based on the selected discovered rules is implemented to handle the heterogeneity of input and gives continuous output that represent the possibility of coronary artery blockage.

MLP is an ANN that is widely used in medical decision support system [9]. It has been applied successfully to many difficult problems in many areas [61]. MLP is also used for imputation purposes [30-32]. MLP can learn and find the complex and non-linear relationship among the variables. Thus it can be used to impute missing values because one variable can be predicted using other variables. These existing MLP imputation methods use all variables without feature selection. In the decision systems, not all of the variables may contribute to the decision. Only contributing variables should be selected to simplify the MLP topology. RST has the ability to select variables that have strong contribution to the decision by the concept of reduct [94]. A hybrid model of ANN and RST called ANNRST is used in this thesis as the method of imputation. Using RST, the complexity of ANN topology can be reduced by RST feature selection or attribute reduction. Hence, the topology will be simpler but still has the ability of imputing the missing attribute values with similar performance to ANN before attribute reduction. The required number of training samples usually increases with the dimensionality of the input space [95].

Many researchers have evaluated the performance of imputation using the accuracy and error based metrics on simulated missing data [24-26][28-32]. The comprehensive evaluation based on classifiers to measure the quality of imputation has also been conducted in [22][23][27]. An idea to use the most complete data set (Cleveland) to evaluate the quality of imputation is introduced in this study. The knowledge discovered of data sets that have many imputed missing values (Hungarian and Long Beach) is used to classify the most complete data set. This method is called "imputed classify complete" because the imputed data set is used to classify the complete data sets to know the correctness of imputation in term of extracted knowledge from the imputed data set .

The quality metrics of resulting RST rules as classifier, namely accuracy, coverage, sensitivity, specificity are calculated. Area under curve (AUC) of receiver operating characteristics (ROC) is also calculated [96]. Michalski rule quality is adopted and modified as metrics to measure the quality of classifier based on testing data [33]. This rule filtering method is proposed as an evaluator of the imputation performance.

Rules extraction by using RST will produce large numbers of rules. These rules need to be filtered in order to get a reliable number of rules. There are many studies that proposed various rule filtering methods [33-35][64][80-82]. In this section, RST based rule importance measure is adopted to select the rules by converting the rules to decision tables [64]. Filtering method based on rule support is applied for the first stage to select the rules. The second stage is applying rule importance measure to select the important rules from the rules which are filtered in the first stage.

These crisp rules are fuzzified, i.e. the crisp rules are transformed into fuzzy using fuzzy membership function. Different from fuzzy decision support system developed by Tsipouras, et al, fuzzy weighing method is introduced based on selected RST rules support of training data [15][16].

Coronary artery disease (CAD) data sets taken from University California, Irvine, U.S. are the most used data sets in research works on computer based diagnosis of CAD. Missing data imputation based on ANN and RST is proposed and discussed. Rule generation and selection based on rough set theory is also demonstrated. Fuzzy modeling of the decision support system is described at the end of this chapter. The research methodology flow is shown in Figure 3.1.

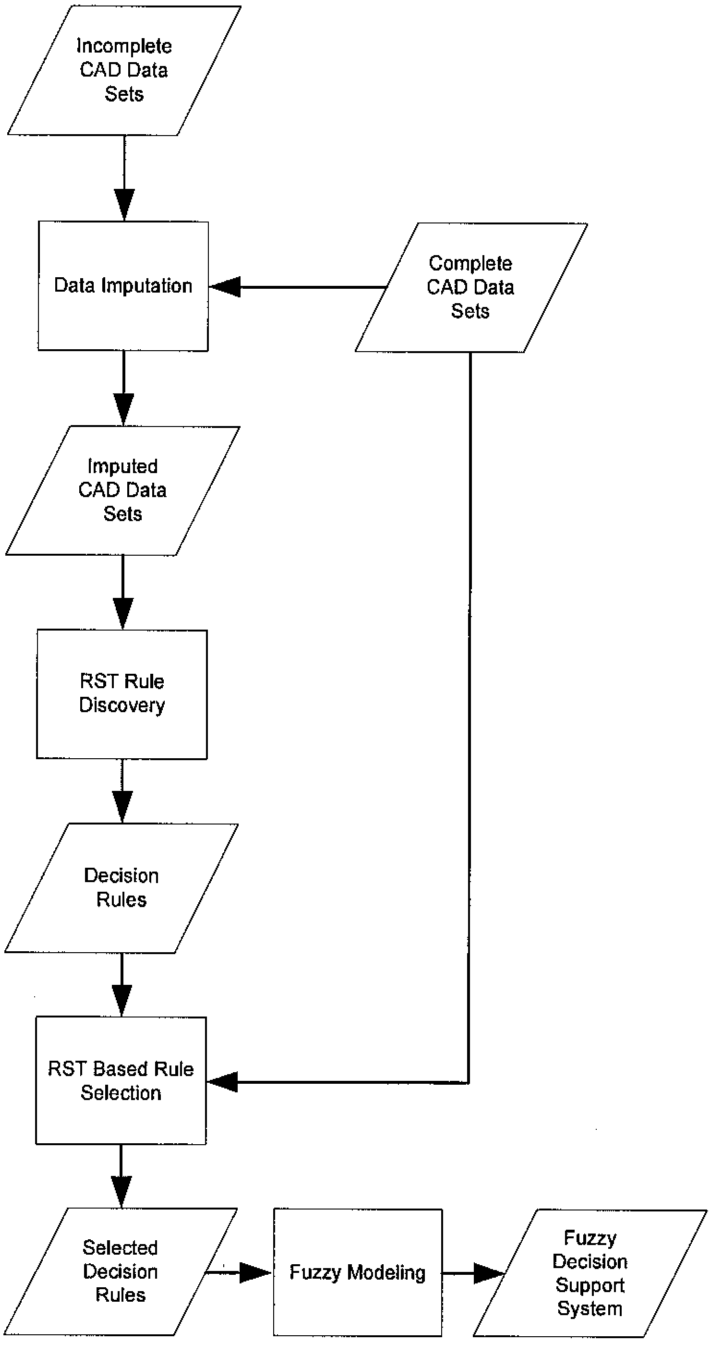


Figure 3.1 Research methodology

3.3 CORONARY ARTERY DISEASE DATA SETS

Coronary artery disease data sets from UCI database repository attract many researchers to build decision systems [39][44][46][47][97-100]. The UCI-CAD data sets contain 920 patients that are collected from Cleveland Clinic Foundation U.S. (303 patients); Hungarian Institute of Cardiology, Budapest, Hungary (294 patients); Veterans Administration Medical Center, Long Beach, California, U.S. (123 patients) and University Hospital, Zurich, Switzerland (200 patients). The data form Ipoh Specialist Hospital, Malaysia consists of 22 patients. The results of CAD disease are obtained by coronary angiography. There are 14 attributes of CAD data which are proved to be complete data set. These attributes can be seen in Table 3.1.

Table 3.1 Summary of attributes and decision (UCI Heart Disease Data Set) [42]

Attribute	Description	Value description
age	Age	Numerical
sex	Sex	1 if male; 0 if female
cp	Chest pain type	1 typical angina 2 atypical angina 3 non-anginal pain 4 asymptomatic
trestbps	Resting systolic blood pressure on admission to the hospital (mmHg)	Numerical
chol	Serum cholesterol (mg/dl)	Numerical
fbs	Fasting blood sugar over 120 mg/dl ?	1 if yes 0 if no
restecg	Resting electrocardiographic results :	0 normal 1 having ST-T wave abnormality 2 LV hypertrophy
thalach	Maximum heart rate achieved	Numerical
exang	Exercise induced angina?	1 if yes 0 if no
oldpeak	ST depression induced by exercise relative to rest	Numerical
slope	The slope of the peak exercise ST segment	1 upsloping 2 flat 3 downsloping
ca	Number of major vessels colored by fluoroscopy	Numerical
thal	Exercise thallium scintigraphic defects	3 normal 6 fixed defect 7 reversible defect
num (decision)	Diagnosis of heart disease (angiographic disease status / presence of coronary artery disease (CAD))	0 if less than 50% diameter narrowing in any major vessel (CAD no) 1 if more than 50% (CAD yes)

An example of data set in the form of decision table is shown in Table 3.2. The missing values are indicated by "?".

Table 3.2 Example of UCI Heart Disease Data set

No	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
1	47	1	3	138	257	0	2	156	0	0	1	0	3	0
2	53	0	3	128	216	0	2	115	0	0	1	0	?	0
3	53	0	4	138	234	0	2	160	0	0	1	0	3	0
4	58	0	4	100	248	0	2	122	0	1	2	0	3	0
5	48	1	3	124	255	1	0	175	0	0	1	2	3	0
6	57	1	4	132	207	0	0	168	1	0	1	0	7	0
7	52	1	3	138	223	0	0	169	0	0	1	?	3	0
8	50	1	3	129	196	0	0	163	0	0	1	0	3	0
9	51	1	4	140	298	0	0	122	1	4.2	2	3	7	1
10	43	1	4	132	247	1	2	143	1	0.1	2	?	7	1
11	57	1	2	154	232	0	2	164	0	0	1	1	3	1
12	58	0	4	130	197	0	0	131	0	0.6	2	0	3	0
13	57	1	4	110	335	0	0	143	1	3	2	1	7	1
14	47	1	3	130	253	0	0	179	0	0	1	0	3	0
15	55	0	4	128	205	0	1	130	1	2	2	1	7	1
16	35	1	2	122	192	0	0	174	0	0	1	0	3	0
17	61	1	4	148	203	0	0	161	0	0	1	1	7	1
18	58	1	4	114	318	0	1	140	0	4.4	3	3	6	1
19	58	0	4	170	225	1	2	146	1	2.8	2	2	6	1
20	58	1	2	125	220	0	0	144	0	0.4	2	?	7	0
21	63	1	4	140	187	0	2	144	1	4	1	2	7	1
22	63	0	4	124	197	0	0	136	1	0	2	0	3	1
23	41	1	2	120	157	0	0	182	0	0	1	0	3	0
24	59	1	4	164	176	1	2	90	0	1	2	2	6	1
25	57	0	4	140	241	0	0	123	1	0.2	2	0	7	1
26	45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
27	68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
28	57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
29	57	0	2	130	236	0	2	174	0	0	2	1	3	1
30	38	1	3	138	175	0	0	173	0	0	1	?	3	0

Most of the researchers use Cleveland data set that consists of 303 patients because it contains only six patients which have missing values on thallium scintigraphic defects or number of vessels colored by fluoroscopy. The number of disease prevalence is 139 out of 303 patients (45.87%). Hungarian data set has the disease prevalence of 106 out of 294 patients (36.05%). Long Beach data set has the disease prevalence of 149 out of 200 patients (74.5%). Switzerland data set has the disease prevalence of 115 out of 123 patients (93.5%). These three data sets have many missing values not just in thallium scintigraphic defects and number of vessels colored by fluoroscopy that appeared to contain calcium, but also in the slope of the peak exercise ST segment and serum cholesterol. A few missing values are also found in other attributes of these three data sets. There are two reported works that use Hungarian, Long Beach and Switzerland data sets which contain many missing values to develop CAD diagnosis system which are proposed by Detrano, et al [39] and Pedreira, et al [100]. However, Detrano, et al, have the complete version of data sets while Pedreira, et al, eliminates the objects that contain missing data. The three data sets that contain missing values are too valuable to ignore during the development of data driven decision support system. In this thesis, selected objects from Hungarian and Long Beach data sets which are incomplete are used and imputed using ANNRST imputation. This imputed data set then is considered as the training data of RST based rule generation. Cleveland data set is used as testing data and that of Switzerland is used as validation data. The data from Ipoh Specialist Hospital is used for fuzzy decision support system validation. To verify the performance of ANNRST imputation, 597 selected objects from Cleveland, Hungarian and Long Beach data sets that have complete attribute values are used.

3.4 ROUGH SET THEORY

Rough set theory, as described by Pawlak, deals with the analysis of classification of data tables that may consist of uncertainty.

3.4.1 Decision system and set approximation

The objects of RST are information system and decision system and set approximation. This study deals with decision system. Thus, only decision system will be covered in this brief explanation. The decision system is defined as:

$$DS = (U, C \cup D), \tag{3.1}$$

where $D \not\subset C$ is called decision attribute or simply as decision.

Table 3.3 Example of decision table

$x \in U$	c_1	c_2	d
x_1	0	3	1
x_2	0	0	0
x_3	1	1	0
x_4	1	1	1
x_5	2	2	0
x_6	0	3	1
x_7	2	2	0

Table 3.3 is an example of decision table with $D = \{d\}$ and $C = \{c_1, c_2\}$.

U and C are finite nonempty set called the universe and the set of condition attributes or simply conditions.

Let $A \subseteq C$, each subset defines an equivalence relation called indiscernibility relation which is defined as:

$$IND_C(A) = \{(x, x') \in U \times U \mid \forall c \in A, c(x) = c(x')\}. \tag{3.2}$$

Using only condition in A, the indiscernibility relation of equation (3.2) will induce a partition of U into sets. Each object in the set cannot be discerned from other object in same set. For example $A = \{c_2\}$, then the partition which is induced by equation (3.2): $U / IND_C(A) = \{\{x_1, x_6\}, \{x_2\}, \{x_3, x_4\}, \{x_5, x_7\}\}$. The sets of classified objects are called equivalence classes denoted as $[x]_A$.

Set approximation is used when a decision concept such as d cannot be defined in a crisp manner. It is shown in Table 3.3 of objects x_3 and x_4. For $A \subseteq C$, the approximations of $X \subseteq U$ using only information in A are a lower-approximation $\underline{A}X$ and an upper-approximation $\overline{A}X$ that are defined by:

$$\underline{A}X = \{x \mid [x]_A \subseteq X\} \tag{3.3}$$

$$\overline{A}X = \{x \mid [x]_A \cap X \neq \varnothing\} \tag{3.4}$$

The set of lower-approximation consists of objects which certainly belong to X. The set of upper-approximation consists of objects which possibly belong to X. The set that contains objects that cannot be classified as definitely inside X or outside X is called boundary region of X:

$$BR_A(X) = \overline{A}X - \underline{A}X \tag{3.5}$$

A set is said to be rough if $BR_A(X) \neq \varnothing$. Let $A = \{c_2\}$ and $X = \{x_1, x_4, x_6\}$ as a result of $d = 1$. Then $\underline{A}X = \{x_1, x_6\}$, $\overline{A}X = \{x_1, x_3, x_4, x_6\}$ and $BR_A(X) = \{x_3, x_4\}$. It follows that outcome of $d = 1$ is rough since the boundary region is not empty.
Let D as decision and $A \subseteq C$ as conditions, then A-positive region of D is defined as:

$$POS_A(D) = \bigcup_{X \in U/D} \underline{A}X \tag{3.6}$$

where U/D represents the partition of U according to decision D. For example in Table 3.2, with $A = \{c_2\}$ then $POS_A(D) = \bigcup\{\{x_1, x_6\}, \{x_2\}, \{x_5, x_7\}\} = \{x_1, x_2, x_5, x_6, x_7\}$.

3.4.2 Reduct

The condition attributes of decision system DS may be redundant so they can be reduced. The reduction of DS will result in reducts. A reduct is a minimal set of attributes $A \subset C$ such that $POS_A(D) = POS_C(D)$. A reduct is a combination of conditions that can discern among objects as well as all conditions. Reducts can be computed using discernibility matrix and discernibility function [101]. A discernibility matrix of decision system DS can be defined as $n \times n$ matrix M with its elements:

$$m_{ij} = \begin{cases} \{c \in C \mid c(x_i) \neq c(x_j)\}, & \text{if } d(x_i) \neq d(x_j), \\ \varnothing & \text{if } d(x_i) = d(x_j). \end{cases} \tag{3.7}$$

for $i, j = 1 \ldots n$ and $d \in D$.

The elements of M are the conditions that are needed to discern object i from object j relative to the decision. Table 3.4 shows an example of decision relative discernibility matrix of Table 3.3. Discernibility function can be derived based on a discernibility matrix. A discernibility function f_C for decision system DS consists of k Boolean variables of Boolean function which is defined as:

$$f_C(c_1^*,...,c_k^*) = \wedge\{\vee m_{ij}^* \mid 1 \le j \le i \le n, m_{ij} \ne \varnothing\} \quad (3.8)$$

where $m_{ij}^* = \{c^* \mid c \in m_{ij}\}$. All the minimal reducts of the decision system may be obtained by finding the set of all the prime implicants of discernibility function. According to Table 3.3, the discernibility function of decision system DS is $f_C(c_1,c_2) = \{c_1 \vee c_2\} \wedge \{c_2\}$. After simplification using Boolean algebra properties then the discernibility function becomes $f_C(c_1,c_2) = \{c_2\}$. Thus there is only single reduct $\{c_2\}$. The concept of reduct can be used as attribute reduction or feature selection.

Sometimes reducts computed based on relative discernibility of objects are more interesting than full discernibility reduct, especially for decision rule generation. Reducts that are relative to object $x \in U$ can be found modifying equation (3.8):

$$f_C(c_1^*,..c_2^*,x_i) = \underset{x_j \in U}{\wedge}\{\vee m_{ij}^* \mid m_{ij} = M(x_i,x_j), m_{ij} \ne \varnothing\} \quad (3.9)$$

Table 3.4 Example of decision relative discernibility matrix

$x \in U$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1							
x_2	c_2						
x_3	c_1,c_2						
x_4		c_1,c_2					
x_5	c_1,c_2			c_1,c_2			
x_6		c_2	c_1,c_2		c_1,c_2		
x_7	c_1,c_2			c_1,c_2		c_1,c_2	

For example object x_3 relative reducts can be calculated using object relative discernibility function shown in equation (3.14) $f_C(c_1,c_2,x_3) = \{c_1 \vee c_2\}$, hence relative reducts are $\{c_1\}$ and $\{c_2\}$. All relative reducts of x_1 to x_2 can be found.

3.4.3 Decision rule generation

Once all of the relative reducts are determined, a set of decision rules can be generated from those reducts. Various algorithms are available to generate rules from those reducts. Consider $DS = (U, C \cup D)$ as a decision table. $\forall x \in U$, series $c_1(x),...,c_k(x), d(x)$ can be defined, where $\{c_1,...,c_k\} = C$ and $\{d\} = D$. Hence the decision rules can be generated in the form of $c_1(x),...,c_2(x) \rightarrow d(x)$. C can be the condition attributes of reduced form of decision table (reduct). For example, using information from Table 3.3 by considering one of its relative reducts = $\{c_2\}$ as shown in Table 3.5,

Table 3.5 Relative reduct of Table 3.3

$x \in U$	c_2	d
x_1	3	1
x_2	0	0
x_3	1	0
x_4	1	1
x_5	2	0
x_6	3	1
x_7	2	0

the rules can be generated as antecedent consequent form as follows

IF $c_2 = 3$ THEN $d = 1$. (3.10)

IF $c_2 = 0$ THEN $d = 0$. (3.11)

IF $c_2 = 2$ THEN $d = 0$. (3.12)

IF $c_2 = 1$ THEN $d = 1$ or $d = 0$. (3.13)

The set of rules (3.10-3.13) is said to be deterministic. Rule (3.13) is indeterministic because its consequence is uncertain. The definition of rule support is described as how many objects match the corresponding rule. For example rule (3.12) has support = 2 because there are 2 objects that match the rule antecedent and consequent. Those objects are x_5 and x_7. Support can be used as rule filtering criterion when there are too many rules generated. Accuracy of certain rule can be calculated as its support divided by the number of objects that match its rule antecedent. Thus, rule (3.12) has accuracy $= \frac{2}{2}$. Coverage of a rule can also be calculated by dividing its support by the number of objects that match its rule consequence. Hence, rule (3.12) has coverage $= \frac{2}{4}$.

3.5 MISSING DATA IMPUTATION

Knowledge discovery from data (KDD) processes usually encounter missing data problems. The source of problems may be from collecting processes due to the device and human error. The quality of data is a very important factor for KDD to reach a quality discovery of knowledge.

A naïve technique such as to ignore the instances which contain missing attribute values, is usually used to process the data set. However, this technique will cause important information loss within the data. In the worst case, a significant amount of data will be discarded. Even the instances that only have one missing attribute value will be discarded. Some methods will assign the most common value or the average value to the missing data. These techniques will add noise to the data set because the imputed values may not come from the data originally derived from. These techniques also can not learn the nature and the property of the data set.

ANNRST attribute reduction is proposed here to impute the missing attribute values on CAD data from UCI data set. This system takes the ability of learning from ANN and the ability of removing redundancy of the data from RST. Function of ANNRST is discussed and explained to give the clear idea about the methods.

3.5.1 ANNRST data imputation

ANNRST is a hybrid method of MLP-ANN as a missing data estimator and RST as an attribute reducer of the input. The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision feature as the original.

As an illustration, consider a data set in the form of a decision table that consists of m condition attributes: $\{c_1, c_2, \ldots, c_m\}$ and single decision attribute: $\{d\}$. Let $\{c_2\}$ be the attribute that has missing values. Using ANN, a missing value estimator can be built to estimate the missing values of attribute $\{c_2\}$ by using $\{c_1, c_3, \ldots, c_m\}$ and $\{d\}$ as the input attributes of input layer and $\{c_2\}$ as the output attribute of output layer. Thus, there will be m input attributes and one output attribute. Using RST, the decision table can be reduced using the reduct concept. Then, the condition attributes can be reduced to: $\{c_1, c_2, \ldots, c_n\}$ where $n < m$. For missing values of $\{c_2\}$, an ANN imputer can be built with smaller size of input attributes: $\{c_1, c_3, \ldots, c_n\}$ and $\{d\}$. The diagram of ANNRST is shown in Figure 3.2, where c_k denotes the kth attribute that has missing values. As an illustration, consider Table 3.6.

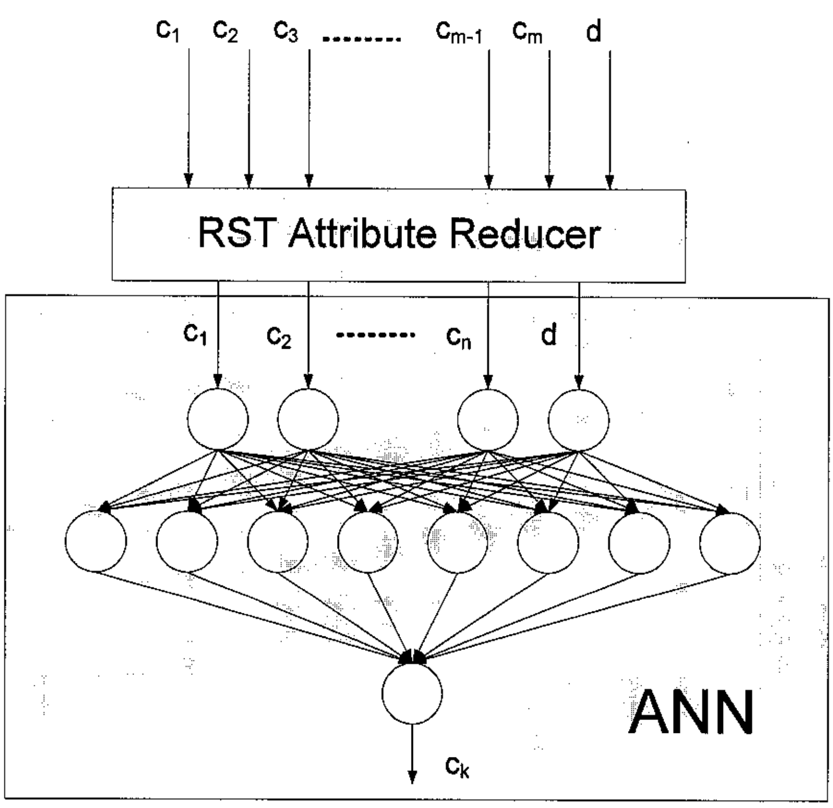


Figure 3.2 ANNRST topology

Table 3.6 Example of data set with three condition attributes and missing values

$x \in U$	c_1	c_2	c_3	d
x_1	0	3	1	1
x_2	0	0	0	0
x_3	1	1	?	0
x_4	1	1	1	1
x_5	2	2	0	0
x_6	?	3	1	1
x_7	2	2	0	0
x_8	2	?	0	1
x_9	2	3	0	1
x_{10}	2	3	1	1

To compute the reduct, the instances that have missing attribute values are removed and thus Table 3.6 becomes Table 3.7.

Table 3.7 Example of data set with missing values is removed

$x \in U$	c_1	c_2	c_3	d
x_1	0	3	1	1
x_2	0	0	0	0
x_4	1	1	1	1
x_5	2	2	0	0
x_7	2	2	0	0
x_9	2	3	0	1
x_{10}	2	3	1	1

RST reduct computation using Johnson's algorithm yields the reduct of Table 3.7 which is Table 3.8 [102].

Discretization must be conducted for the case of Table 3.2 which has continuous values of its conditional attributes before applying reduct computation.

The algorithm of discretization operates by first creating a Boolean function f_K from the set of candidate cuts, and then computing a prime implicant of this function.

For each condition attribute c, its value set V_c can be sorted to obtain the following ordering:

$$v_c^1 < \ldots < v_c^i < \ldots < v_c^{|V_c|} \tag{3.14}$$

Let K_c represents the set of all generated cuts for attribute c with decision d.

$$X_c^i = \{x \in U \mid c(x) = v_c^i\} \tag{3.15}$$

$$\Delta_c^i = \{v \in V_d \mid \exists x \in X_c^i \text{ such that } d(x) = v\} \tag{3.16}$$

$$K_c = \left\{ \frac{v_c^i + v_c^{i+1}}{2} \mid \left|\Delta_c^i\right| > 1 \text{ or } \left|\Delta_c^{i+1}\right| > 1 \text{ or } \Delta_c^i \neq \Delta_c^{i+1} \right\} \tag{3.17}$$

The function f_K is defined as:

$$f_K = \prod_{(x,y)} \sum_{c} \left\{ \sum k^* \mid k \in K_c \text{ and} c(x) < k < c(y) \text{ and} \partial_C(x) \neq \partial_C(y) \right\} \tag{3.18}$$

where $\partial_C(x) = \{v \in V_c \mid \exists y \in R_C(x) \text{ such that } d(y) = v\}$ and $R_C(x) = \{y \in U \mid xR_C y\}$

The processing steps of computation based on Johnson's algorithm are illustrated as follows:

Let B represent the reduct or the discretization cuts. U denotes the set of sets corresponding to the function f_C (3.13) or f_K (3.23), and $w(S)$ denotes a weight for set S in U that automatically gets computed from the data [55].

i. Let $B = \varnothing$

ii. Let c or k denotes the variable that maximizes $\sum w(S)$, where the sum is taken over all sets S in U that contain c or k.

iii. Add c or k to B.

iv. Remove all sets S from U that contain c or k.

v. If $S = \varnothing$ return B. Otherwise, go to step ii.

Table 3.8 Example of reduct of data set with three condition attributes and missing values

$x \in U$	c_2	d
x_1	3	1
x_2	0	0
x_4	1	1
x_5	2	0
x_7	2	0
x_9	3	1
x_{10}	3	1

Based on Table 3.8, a simple MLP-ANN can be built to estimate the missing data of attribute c_3. Although a single hidden layer is enough to make ANN become universal approximator, there are no strict formulas to determine the number of hidden neurons. For this case, log_2 of the number of training samples is chosen to determine the number of hidden neurons [103]. The topology is shown in Figure 3.3.

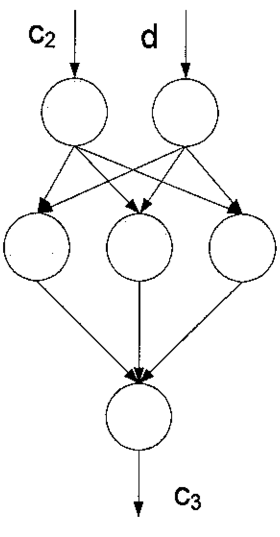


Figure 3.3 ANN topology to impute the missing values in attribute c_3

The same method can be applied to construct the ANN for missing attribute values imputation in attribute c_1. For attribute c_2, complete attributes must be used since the reduct has the c_2 attribute, i.e. c_1, c_3 and d are the input and c_2 is the output.

To train the network, the data set shown in Table 3.8 is used. The data set is split into training data, validation data and testing data with some split ratio. For example, the split ratio is usually 60:40, 70:30 or 80:20 between training and validation. It depends on the amount of the data. Validation is included in the training process for early stopping method. After training – testing process, the well trained ANN is ready to impute the missing values which are shown in Table 3.9. Usually at the start of training process the validation error is decreasing. In the early stopping method, the stopping condition is when the validation error starts to increase.

Table 3.9 Example of data set with missing values which will be imputed

$x \in U$	c_1	c_2	c_3	d
x_3	1	1	?	0
x_6	?	3	1	1
x_8	2	?	0	1

Using the ANN topology shown on Figure 3.4, the missing value in attribute c_3 can be imputed with the input of c_2 and d.

A study on imputation technique using simulation of missing data is carried out to obtain the effectiveness of ANNRST. All UCI data sets except Switzerland are used in this study. Referring to the Table 3.1 and Table 3.2 in section 3.3, the attributes that have real missing values which are the slope of the peak of ST segment (*slope*), number of major vessels coloured by fluoroscopy (*ca*) and exercise thallium scintigraphic defects (*thal*) are removed (not used). The objects with most of their attributes have missing values are also removed. For example, if any object x_i has missing value on most of its attributes, then x_i is removed from the decision table. After the above preprocessing, the decision table consists of 597 objects with 10 conditions and single decision. ROSETTA is used for RST data analysis. Boolean reasoning algorithm shown in equations (3.19 – 3.23) is applied to discretize the numerical attributes. The decision table of data set then is reduced by RST using Johnson's reduct computation [102]. Fasting blood sugar over 120 mg/dl (*fbs*), resting electrocardiographic results (*restecg*) and exercise inducing angina (*exang*) are arbitrarily chosen as the attribute that has missing values. Ten to 300 missing values are simulated on *fbs* attribute.

The ANN topology that is used is an MLP with six inputs (all reduct conditions and decision), single hidden layer and single output (*fbs*). The number of hidden layer is chosen to be $\log_2(T)$, where T is the number of training samples. In this case the maximum missing data is 300. It means 297 samples are used as training process. The early stopping method is used with the ratio of training and validation is 80:20, because the more the number of training samples the better the results of training process with acceptable number of validation samples. Hence the actual training samples are 238 sets with 59 sets of validation samples, which are unseen data. 59 sets are more than 30 sets thus are acceptable in statistical point of view. Then the hidden layer is $\log_2(238) = 7.89 \approx 8$ neurons [103]. Rprop learning method and normalization of inputs value by scaling the inputs between values of -1 to 1 are applied. The normalized value of x which is y can be calculated using the formula:

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \tag{3.19}$$

where $y_{max} = 1$, $y_{min} = -1$, x_{max} is the maximum value of vector X and x_{min} is the minimum value of input vector X.

Fifty training simulations with different weight initialization are carried on the data set using Nguyen-Widrow weight initialization [104].

Initialization algorithm based on Nguyen-Widrow is as below:

i. Let $w_{ij}(\text{old})$ = random number between -0.5 and 0.5
ii. Compute $\|w_j(\text{old})\| = \sqrt{(w_{1j}(\text{old})^2 + w_{2j}(\text{old})^2 + \ldots + w_{nj}(\text{old})^2)}$
iii. Reinitialize weights :

$$w_{ij}(\text{new}) = \frac{\delta w_{ij}(\text{old})}{\|w_j(\text{old})\|}.$$

iv. Set bias: w_{0j} = random number between $-\delta$ and δ.

where $\delta = 0.7\sqrt[n]{p}$ with n = number of input neurons and p = number of hidden neurons.

The typical values of parameters are $\eta^- = 0.5$, $\eta^+ = 1.2$ and $\Delta_{ij}^{(0)} = 0.07$ [62].

For comparison, ANN without RST attribute reduction, k-Nearest Neighbour (k-NN) and Concept Most Common value Filling (CMCF) are applied to the data set for data imputation.

The algorithm of k-NN is as per flow chart shown in Figure 3.4: take the row that contains missing value as the target, determine its k-nearest neighbours by computing the Euclidean distance between incomplete row and complete rows, identify the k closest distance and impute the missing value in the row that contains missing value by averaging the corresponding complete rows of the k closest [28]. In this case k = 1 because Table 3.2 consists of non-numerical values. It is not possible to take the average of non-numerical values.

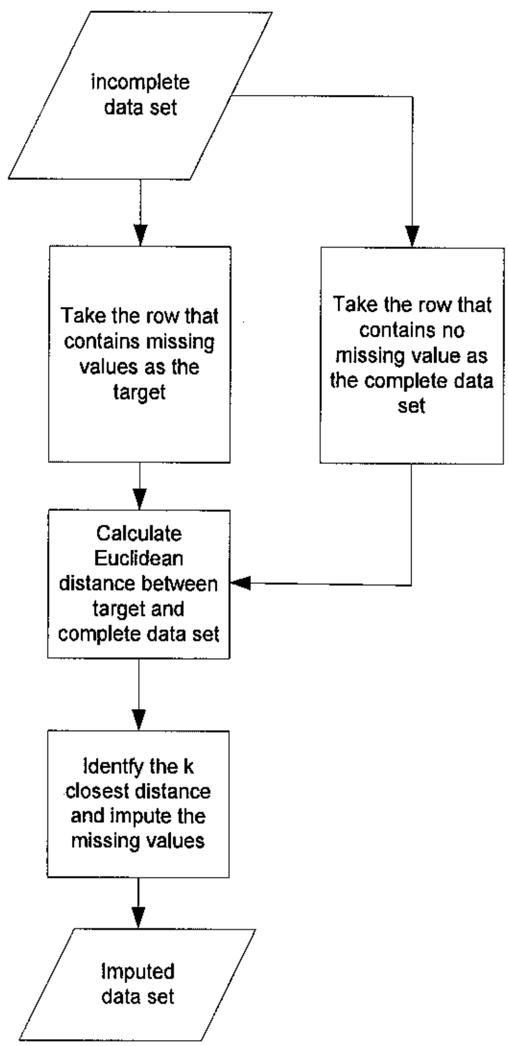


Figure 3.4 k-NN imputation

Most common attribute value filling is the simplest methods to deal with missing attribute values. The value that occurs most frequent is selected to be the value of the unknown missing attribute value. CMCF is a restriction of most common attribute value filling by its concept or decision. In CMCF, the value that occurs most frequent within concept is selected to be the imputed value. CMCF is also called maximum relative frequency method or maximum conditional probability method [22].

Referring to Table 3.2, ANN imputation uses ten input attributes (nine from conditions and one from decision) to impute missing values for example in three attributes, namely *fbs*, *restecg* and *exang* attributes (Section 3.1).

The accuracy among different methods is calculated. Maximum accuracy among 50 training-testing simulations of ANN and ANNRST is considered as the accuracy for these two methods. Average accuracy for 50 training-testing simulations with different weight initializations using Nguyen-Widrow for ANN and ANNRST is calculated to find the stability of these two methods. The higher average accuracy means the higher probability to get the better accuracy.

After applying ANNRST to the simulated missing data, ANNRST is applied to the actual missing values on *slope, ca* and *thal* attributes (Table 3.2) using the same process. At the first step, ANNRST is used to predict the ten simulated missing values on *slope, ca* and *thal* attributes. Fifty simulations with random initial weight are conducted on each simulated missing data with Nguyen-Widrow weight initialization, scaling to -1 to 1 as in (3.24), resilient backpropagation training, and tansig activation function at hidden and output layer. Secondly, the best accuracy of 50 simulations is chosen to predict the real missing values on *slope, ca* and *thal* attributes. k-NN imputation method and CMCF to estimate the missing values are also implemented. There are three data sets after missing data estimation, which are ANNRST, k-NN and CMCF imputed data sets. The process of imputation using ANNRST is shown in Figure 3.5.

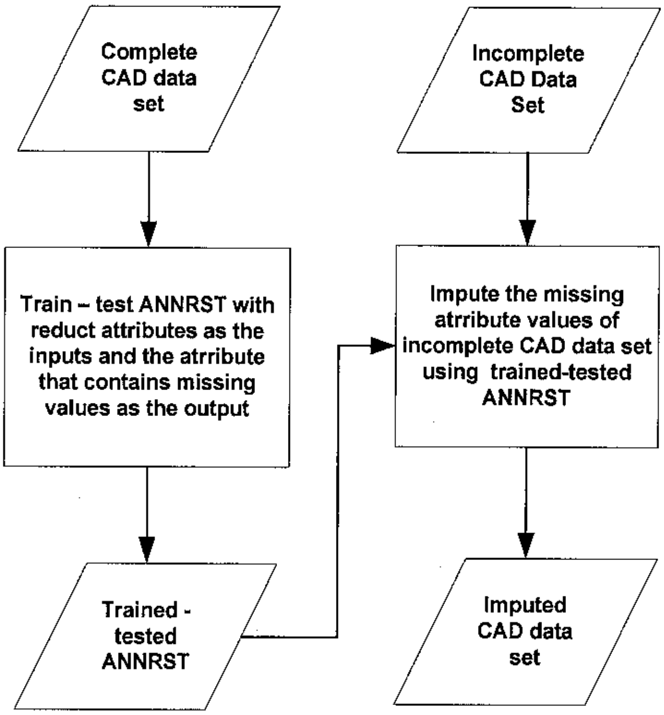


Figure 3.5 ANNRST imputation process

3.5.2 The evaluation of ANNRST imputation

Accuracy of imputations on simulated missing data in *fbs, restecg* and *exang* attributes (Table 3.2) is used to evaluate ANNRST, ANN, k-NN and CMCF imputations. However, accuracy can not be used to evaluate the quality of the imputation for the case of real missing attribute values case because the actual of the missing data is unknown [23]. The result of the imputation can not be guaranteed as the true values that were missing. In this work the imputed data set evaluation to measure the performance of the imputation methods is proposed. The evaluation is conducted by discovering the knowledge from the imputed data set and then building the classifier based in the discovered knowledge to classify the complete data set. The accuracy and coverage of the classifier that is based on the knowledge of certain imputed data set can be considered as the performance of those certain imputation method. For example, the accuracy and coverage of classifier based on imputed data set by ANNRST can be considered as the performance index of ANNRST imputation method.

There are two methods to find the classifier accuracy and coverage. First, using 10-fold cross validation, the accuracy and coverage of all data sets are calculated using different classifiers, which are RST, decomposition tree, LTF-C (Local Transfer Function Classifier), and k-NN. Cross validation is a method for evaluation of classifiers. k-fold cross validation is performed by splitting the data into k sub-samples. One sub-sample is used as training set and the remaining (k-1) subsamples are treated as testing sets. The training-testing process is repeated k times with each of k sub-samples used only once as the testing set. The accuracy is the average of k individual accuracy. 10-fold cross validation is commonly used and gives good performance [105]. Decomposition is partitioning a large table into smaller parts. The decomposition scheme of large table into smaller ones is based on tree template which represents common features. The results will be binary tree of subtables. This binary tree can be used for classification [57]. LTF-C is an ANN based classifier. Its topology or architecture is similar to Radial Basis Function (RBF) or Support Vector Machine (SVM). It has one hidden layer with Gaussian transfer function that is connected to output layer with linear transfer function. It differs from RBF and SVM in the training algorithm. More detailed explanation can seen in [56][106]. k-NN classifier is an instance based method classifier. The algorithm is by inducing a distance measure from the training set. Then for each instance, it assigns a decision based on the k nearest neighbours of this instance according to distance measurement. Further explanation can be found in [56].

Second, all the data sets of UCI, except Switzerland, are split into two sets. The first set is Cleveland data set as the testing data set that has only six missing values and the second set is imputed data sets (Hungarian and Long Beach) as the training data set. Then, accuracy and coverage of all three data sets are calculated with the same classifier as the first method. This second method is to find the quality of imputed data sets according to original complete data set, which is Cleveland data set. All the accuracy and coverage calculations are done using the software RSES.

RST based imputation evaluation is also proposed in this work. RST is used to extract the rules from Hungarian and Long Beach data sets as the training data. The resulting three training data sets, namely ANNRST, k-NN and CMCF data sets are used for rule extraction. Using ROSETTA, there will be a lot of extracted rules. Simple filtering method based on rule support is conducted. This rule filtering method is proposed as an evaluator of the imputation performance. The imputation method effects on the classifier performance on the testing data while rule filtering is applied on these rules are demonstrated.

Accuracy is a statistical measure of how well a binary classifier correctly classifies the objects, i.e., the accuracy is the proportion of true positive and true negative results in the population. In the case of disease, an accuracy of 100% means that the system identifies all sick and well people correctly. Coverage is a measure of how well a binary classifier recognizes the objects. Coverage also represents the generality of a classifier. Both high accuracy and coverage are requirements of binary classifiers. Michalski introduced a quality measure of classifier that combines accuracy and coverage. Sensitivity measures the proportion of positives which are correctly identified. Specificity measures the proportion of negatives which are correctly identified. If sick is considered as positive, a sensitivity of 100% means that the system identifies all sick people as sick and a specificity of 100% means that the system identifies all healthy people as healthy.

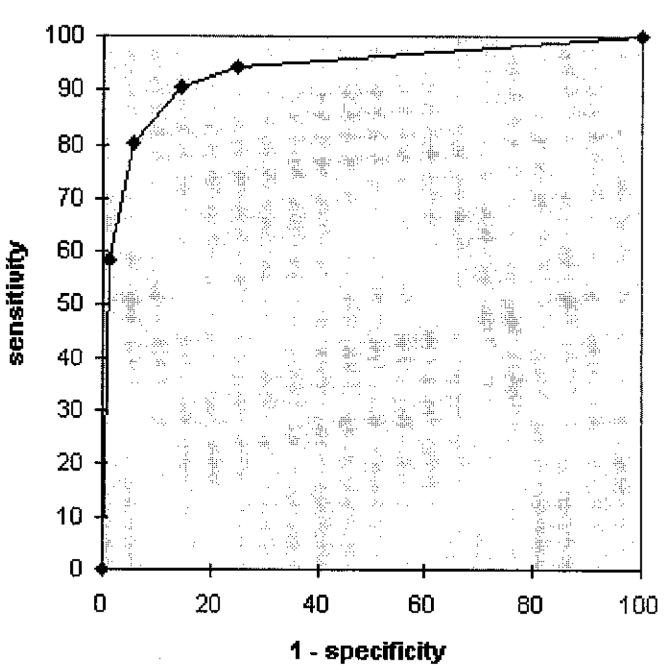


Figure 3.6 An Example of ROC curve

A Receiver Operating Characteristic (ROC) is a graphical representation of discrimination. ROC is a graphical plot of the sensitivity versus (1 - specificity) for a binary classifier system with its discrimination threshold is varied. The Area Under Curve (AUC) of ROC represents the discrimination capability of the classifiers. To measure the RST classifier performance during rule filtering, the quality metrics of classifier, which are accuracy, coverage, sensitivity and specificity, are calculated. AUC of ROC is also calculated [96]. Michalski rule quality is adopted and modified as a metrics to measure the quality of classifier based on the testing data [33]. Example of ROC curve is shown in Figure 3.7. All the quality calculations are based on confusion matrix as shown in Table 3.10.

Table 3.10 Confusion Matrix [55]

		Predicted		
		Negative	Positive	Undefined
Actual	Negative	TN	FP	UN
	Positive	FN	TP	UP

TP (true positive): correct prediction on presence of CAD.

TN (true negative): correct prediction on no-presence of CAD.

FP (false positive): wrong prediction on presence of CAD.

FN (false negative): wrong prediction on no-presence of CAD.

UP (undefined positive): unknown presence of CAD case.

UN (undefined negative): unknown no-presence of CAD case.

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Coverage = (TP+TN+FP+FN)/(TP+TN+FP+FN+UN+UP)

Sensitivity = TP/(TP+FN)

Specificity = TN/(TN+FP)

Michalski quality = μ.Accuracy + (1-μ).Coverage, where $0 \le \mu \le 1$ [33].

For example, consider Table 3.11 as the rules produced by RST.

Table 3.11 Example of rules with their support

Rules	Support
if A = A1 and B = B2 then Y = Y1	80
if A = A2 and C = C1 then Y = Y2	70
if B = B1 and C = C2 and D = D1 then Y = Y1	60
if C = C1 and D = D2 and E = E1 then Y = Y2	50
if A = A2 and E = E2 then Y = Y1	40
if E = E2 and F = F1 then Y =Y2	30
if B = B2 and F = F2 then Y = Y1	20
if D = D1 and F = F1 then Y = Y1	10

These rules are assumed to be the result of RST rule generation from imputed data set by ANNRST. To evaluate the imputation method, rule filtering based on support is applied. As an example, for support = 10, no rule is removed because there is no rule that has support less than 10. When support is 60, the RST classifier will only have three rules. In this work, rule filtering is applied with support start from 1 to *n*. Every time the filtering is applied, quality metrics of classifier, which are accuracy, coverage, sensitivity, specificity, are calculated. The filtering process is stopped when there is no significant change on the metrics of the classifier. For example, the coverage still high enough with acceptable accuracy. By plotting the data into graph, the performance of imputation method can be determined.

3.6 RULE DISCOVERY AND SELECTION

Rules are generated on the basis of computed relative reducts, and constitute one of the most important results of RST data analysis. These rules will be the "core" of fuzzy decision support system to diagnose CAD and will give the explanation of the decision support inference and reasoning.

One of the advantages of RST method for building the classifier is that the system has transparent knowledge instead of being "black-box", such as ANN classifier. RST classifier becomes transparent through the generated rules. However, in practical application, the generated rules by RST are often too large especially with large amount of attributes and data. Very large number of rules, hence, is difficult to understand. With this conditions, RST becomes "black-box". To face this problem, rule filtering is feasible to select the high quality rules and removing redundancy as well. A combination of rule quality filtering and rule importance measure is discussed in this thesis. A rule selection scheme based on rule quality based on support with RST rule importance measure is proposed to make the RST classifier interpretable while maintaining the performance.

3.6.1 Rule discovery by RST

A set of decision rules can be generated from all computed relative reducts. Consider $DS = (U, C \cup D)$ as decision table. Referring to section 3.4.3 and using information from Table 3.3, its relative reducts are $\{c_1\}$, $\{c_2\}$, $\{c_3\}$ and $\{c_1, c_3\}$. By considering one of its relative reducts which is $\{c_2\}$, the rules can be generated as antecedent consequent form with their supports as shown in equation (3.10 – 3.13). The other rules can be generated from the other relative reducts with the same method.

3.6.2 Rule selection based on RST

In this section, RST based rule importance measure [64] is modified to select the rules by converting the rules to decision tables. Support filtering as one of rule quality based filtering is applied at the first stage of the selection. It is time consuming and high computation cost to apply only RST based rule selection because the number of attributes is very high. Filtering method based on rule support is applied to reduce the number of rules before applying rule importance measure to select the most importance rules. The modification is proposed by applying this method to decision system and converting rules to form decision tables based on testing data instead of training data for rule importance measurement. This idea is proposed to increase the generalization of the RST classifier. The classifier learns mainly from training data but also let the testing data contribute to the learning process through the rule selection. This method originally inspired by ANN learning scheme that splits the data to three sets namely training, testing and validation. The testing data involved in learning process to determine the stopping condition of training process. The ANN stops the iteration when the Mean Squared Error (MSE) of testing data starts to increase although the MSE of training data still decreases. Finally the ANN will be validated using validation data which is not involved in training process to test the generality. Consider $R = \{Rule_1, Rule_2, ..., Rule_j\}$ as a set of rules generated from training decision tables. If there are i objects on testing decision table, a new decision table $DS_{i \times (j+1)}$ can be formed. The value of $Rule_a$ attribute of object x_b is 1 if $Rule_a$ both its antecedent and consequent can be applied to x_b. The value is 0 if the rule cannot be applied. The value for column $j+1$ equals decision value. With $a = 1,\ldots,j$ and $b = 1,\ldots,i$. The new decision table then can be reduced using RST reduct concept. The attribute of the shortest reduct are chosen as the selected rules based on their importance. Selected rule based classifier performance is calculated using classifier quality metrics such as accuracy and coverage. Consider Table 3.12 as the table of rules and their support.

Table 3.12 Example of k rules with their support

Rules	**Support**
$Rule_1$	80
$Rule_2$	70
.	.
.	.
.	.
$Rule_{k-2}$	3
$Rule_{k-1}$	2
$Rule_k$	1

Using support filtering with stopping criteria based on accuracy and coverage, the number of rule will be reduced. If there are l number of removed rules, the number of rules become $j = k - l$. The j number of rules then applied to the testing data set which is the complete CAD data set to create new decision tables with rules as the attributes as shown in Table 3.13.

Table 3.13 Decision table with rules as attributes

$x \in U$	$Rule_1$	$Rule_2$	...	$Rule_{j-1}$	$Rule_j$	d
x_1	0	1	...	1	1	1
x_2	0	0	...	0	1	0
x_3	1	1	...	1	0	0
x_4	1	1	...	1	1	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x_{i-2}	0	0	...	0	1	1
x_{i-1}	0	0	...	1	1	1
x_i	0	0	...	1	1	1

Reduct computation is applied to the Table 3.13 using ROSETTA based on Johnson's algorithm. The attributes of the reduct is then the selected rules. The process of rule selection is shown in Figure 3.7.

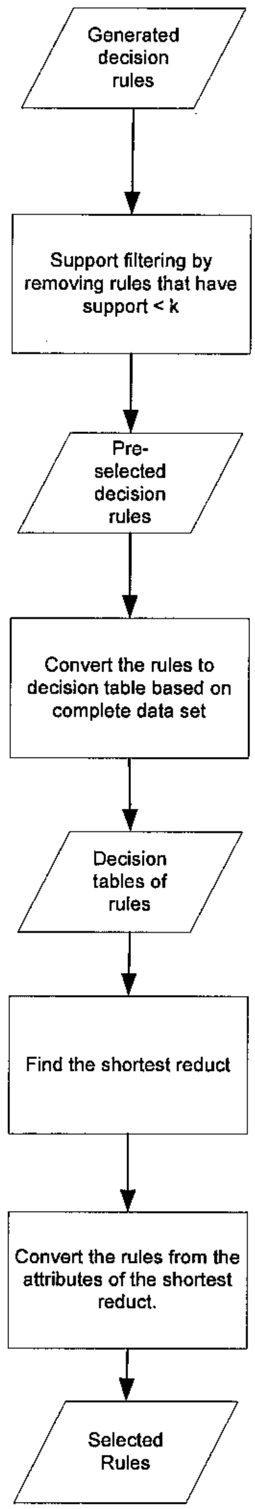


Figure 3.7 Rule selection process

Various rule filtering schemes are implemented for comparison. The formulae used for filtering criteria of rules can be seen in Table 3.14.

Table 3.14 Various rule filtering criteria [33]

Filtering Criteria	Formula
Michalski	μ x *accuracy(r)* + (1 - μ) x *coverage(r)*
Torgo	Michalski with μ = 1/2 + (1/4)(*accuracy(r)*)
Bradzil	$accuracy(r).e^{coverage(r)-1}$
Pearson	$(n_{\alpha,\beta} . n_{\neg\alpha,\neg\beta} - n_{\alpha,\neg\beta} . n_{\neg\alpha,\beta}) / (n_{\beta} . n_{\neg\beta} . n_{\alpha} . n_{\neg\alpha})$
Cohen	$\lvert U\rvert . n_{\alpha,\beta} + \lvert U\rvert n_{\neg\alpha,\neg\beta} - n_{\alpha} . n_{\beta}) / (\lvert U\rvert^2 - n_{\alpha} . n_{\beta} - n_{\neg\beta} . n_{\neg\alpha})$

where $0 \leq \mu \leq 1$, r is individual rule, α is rule antecedent, β is rule consequent, $n_{\alpha\beta}$ is the number of objects that satisfy both α and β, $n_{\alpha,\neg\beta}$ is the number of objects that satisfy α but not β and $|U|$ is the cardinality of the universe. Accuracy and coverage of the RST classifier applied to testing and training data sets is calculated and compared with different filtering criteria. The Michalski, Torgo and Bradzil formulae are empirical, ad-hoc formulae, but are intuitive and interpretable. Pearson and Cohen are based on statistics and the theory of contingency table [33].

3.7 FUZZY MODELLING

The types of applications for which fuzzy systems are particularly useful are mainly difficult cases where traditional techniques do not work well. Medical problem is one of such difficult problem. Some people consider field of medicine is more an "art" than a "science" [39]. Fuzzy system approaches also allow us to represent descriptive or qualitative expression such as "high", "medium" or "low" instead of quantitative expression. These qualitative expressions are more natural than mathematical equations for many decisions making.

Although this RST system with fuzzy inference is a data-driven or evident-based system with no or minimum involvement of domain expert, verification with domain expert is necessary. Evaluation and verification from medical expert (cardiologist) are conducted to validate the results in this work.

3.7.1 Fuzzy logic

Fuzzy logic is the extension of classical logic with the new concept of "degree of membership". The degree of membership represents the possibility of the elements of the set belong to the set. This membership concept differs from crisp set concept which allows only the elements as member or not member of the set.

Let X be the universe of discourse. The elements contained in X is defined by x. Consider A and B to be the sets which contains x. The classical set theory defines:

$$x \in X \rightarrow x \text{ belongs to } X$$
$$x \in B \rightarrow x \text{ belongs to } B$$
$$x \notin A \rightarrow x \text{ does not belong to } A$$

The element in the universe of discourse X that belongs to the set A or does not belong to the set A can be represented by the function

$$\mu_A(x) = \begin{cases} 1 & iff \;\; x \in A \\ 0 & iff \;\; x \notin A \end{cases} \tag{3.20}$$

Different from classical set theory that classifies the elements of the set into crisp set, fuzzy set has an ability to classify elements into a continuous set using the concept of degree of membership. The function that maps the element into the degree of membership is called membership function. The fuzzy membership function is similar to equation (3.20) but it gives the continuous values between 0 and 1 other than discrete values 0 or 1. According to equation (3.20), in the fuzzy set, it cannot be said that x belongs to A or does not belong to A, but x belong to A with certain degree of membership between 0 and 1.

For example, consider resting heart rate. Classical set theory can only classify the heart rate as low and high (i.e., either 0 or 1). It cannot interpret the heart rate between 60 and 80 beats per minute. The function that represents this classical set theory is:

$$\mu_{high}(x) = \begin{cases} 1 & iff \;\; x \geq 70 \; beats \; per \; min \\ 0 & iff \;\; x < 70 \; beats \; per \; min \end{cases} \tag{3.21}$$

The boundary 70 beats per minute is taken because classical logic cannot interpret intermediate values. On the other hand, fuzzy logic can solve this problem using membership function concept as given by:

$$\mu_{high}(x) = \begin{cases} 1 & \text{if } x \geq 80 \text{ beats per min} \\ \dfrac{x-60}{20} & \text{if } 60 \leq x < 80 \text{ beats per min} \\ 0 & \text{if } x < 60 \text{ beats per min} \end{cases} \tag{3.22}$$

This membership function shown in equation (3.22) can be represented as a graph as shown in Figure 3.8. The degree of lowness is the complement of the degree of highness of the heart rate.

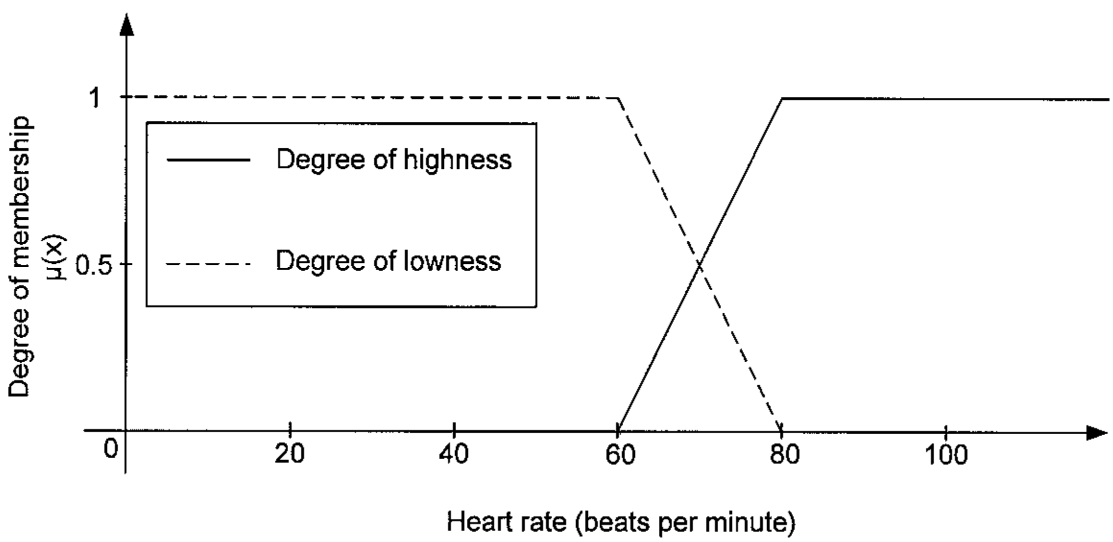


Figure 3.8 Membership function for the degree of highness and degree of lowness of the heart rate

The common method for fuzzy representation is

$$A = \{x, \mu_A(x)\} \quad x \in X \tag{3.23}$$

where x is an element in X and $\mu_A(x)$ is the membership function of set A which defines the membership of fuzzy set in the universe of discourse X. The term $\{x, \mu_A(x)\}$ is the singleton pair. For the case of heart beat, the fuzzy set can be represented as

high = {(60, 0), (65, 0.25), (70, 0.5), (75, 0.75), (80, 1)}.

In the above fuzzy set second element of the set high denotes that the heart rate 65 beats per minute belongs to the set high by 0.25. On the other hand using the graph of Figure 3.8, the second element of the set high is also belongs to the set low by 0.75.

There is an alternative method to represent the singleton function in general as in (3.24)

$$A = \sum_{x_i \in X} \mu_A(x)/x_i \tag{3.24}$$

The representation of (3.24) is for the discrete universe of discourse. For the continuous membership functions, the fuzzy set representation is given by:

$$A = \int_x \mu_A(x)/x_i \tag{3.25}$$

Like classical logic, fuzzy logic also has logic operators. Some important operators are union (disjunction), intersection (conjunction) and complement.

The union is the maximum degree of membership of sets A and B.

$$\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x)) \tag{3.26}$$

The intersection is the minimum degree of membership of sets A and B.

$$\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) = \min(\mu_A(x), \mu_B(x)) \tag{3.27}$$

The complement of the membership of set A is

$$\overline{\mu_A}(x) = 1 - \mu_A(x) \tag{3.28}$$

3.7.2 Fuzzification techniques

Fuzzification is the process of changing the crisp value into fuzzy value by creating fuzzy membership function. This process can be achieved with different types of fuzzifiers. There are many types of fuzzifiers, but in this research only triangular and trapezoidal fuzzifier will be used. Triangular and trapezoidal fuzzifiers that make triangular and trapezoidal membership function are the simplest and the most common possible models of value of membership as they are fully defined by only three and four parameter respectively. The linear change in the membership value is the simplest possible model people can think of.

Fuzzification of a real-valued variable is done by using combination of intuition, experience and analysis of the set of rules and conditions associated with the input data variables. There are no fixed procedures for the fuzzification process.

Triangular membership function is described in the form:

$$\mu_A(x,a,m,b)=\begin{cases} 0, & if\ x \le a \\ \dfrac{x-a}{m-a}, & if\ x \in [a,m) \\ \dfrac{b-x}{b-m}, & if\ x \in [m,b] \\ 0, & if\ x \ge b \end{cases} \tag{3.29}$$

The graphical representation of equation (3.29) is shown in Figure 3.9.

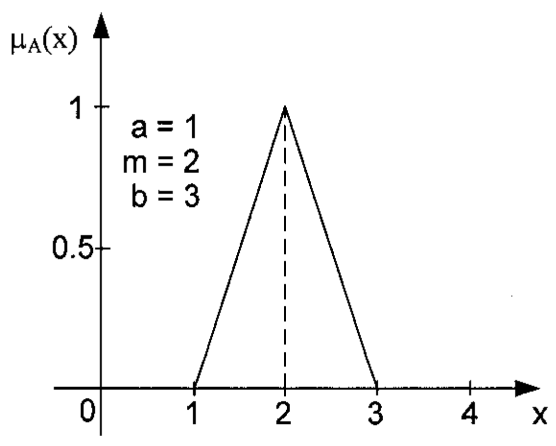


Figure 3.9 Triangular membership function

This membership function can be represented in more concise form as:

$$\mu_A(x,a,m,b)=\max\{\min[(x-a)/(m-a),(b-x)/(b-m)],0\} \tag{3.30}$$

Trapezoidal membership function is described in the form:

$$\mu_A(x,a,m,n,b)=\begin{cases} 0, \; if \; x<a \\ \dfrac{x-a}{m-a}, \; if \; x\in[a,m) \\ 1, \; if \; x\in[m,n) \\ \dfrac{b-x}{b-n}, \; if \; x\in[n,b] \\ 0, \; if \; x>b \end{cases} \tag{3.31}$$

This membership function can be represented in more concise form as:

$$\mu_A(x,a,m,n,b)=\max\{\min[(x-a)/(m-a),1,(b-x)/(b-n)],0\} \tag{3.32}$$

The graphical representation of equation (3.31) is shown in Figure 3.10.

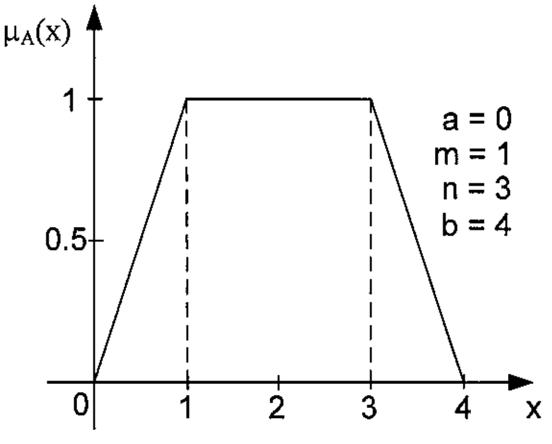


Figure 3.10 Trapezoidal membership function

3.7.3 Fuzzy rules and inference engine

A fuzzy proposition can be an atomic or compound sentence. For example

- "Heart rate is high" is an atomic fuzzy proposition.

- "Heart rate is high and blood pressure is low" is a compound fuzzy proposition.

- Compound fuzzy relations are expressed with fuzzy connectives or operators such as AND (conjunction), OR (disjunction) and NOT (complement). These fuzzy logical connectives are similar to the set operators in the fuzzy set theory.

The fuzzy rules are written as:

- IF <fuzzy proposition> THEN <fuzzy proposition>

- The fuzzy proposition can be atomic or compound sentence.

The IF-THEN rules can be interpreted in classical logic theory by the implication operators. Suppose there is a statement such as "IF a THEN b", this statement can be represented in classical logic as $a \Rightarrow b$.

The implication operator can also be written as

$$\bar{a} \vee b \quad \text{or} \quad (a \wedge b) \vee \bar{a} \tag{3.33}$$

The above equivalence can easily be shown with the truth table of Table 3.15.

Table 3.15 Truth table of implication

a	b	$a \Rightarrow b$ or $\bar{a} \vee b$
F	F	T
F	T	T
T	F	F
T	T	T

Fuzzy logic rule which is IF <fuzzy proposition> THEN <fuzzy proposition> can be represented as $a \Rightarrow b$, where a and b are fuzzy variables. The first fuzzy proposition is rule antecedent and the second is rule consequent.

Fuzzy implication can be replaced by using fuzzy operator such as union (OR), intersection (AND) and fuzzy complement (NOT). This fuzzy implication is called fuzzy inference method. There are many fuzzy inference methods. Fuzzy logic version of proposition (3.33) is given by:

$$\mu_{A \Rightarrow B}(x, y) = \max[1 - \mu_A(x), \mu_B(y)] \ or$$
$$\mu_{A \Rightarrow B}(x, y) = \max[\min\{\mu_A(x), \mu_B(y)\}, 1 - \mu_A(x)] \quad \forall x \in X, \forall y \in Y \tag{3.34}$$

The most common and widely used fuzzy implication is given in (3.35) [107]:

$$\mu_{A \Rightarrow B}(x, y) = \min[\mu_A(x), \mu_B(y)], \ \forall x \in X, \forall y \in Y \tag{3.35}$$

Fuzzy rules and implication make a fuzzy inference system.

The system can be explained by example below:

Let heart beat and blood sugar be fuzzy variables. Let the rules be:

"IF the heart rate is high or heart rate is low THEN heart problem is yes"

The first fuzzy proposition or antecedent is "the heart rate is high or heart rate is low". The second proposition or consequent is "heart problem is yes". The linguistic values of heart rate are high and low. The membership function for heart rate is given by Figure 3.8. Equation (3.28) represents highness membership function for heart rate. The lowness membership function of heart rate is:

$$\mu_{low}(x) = \begin{cases} 1 & if \ \ x \le 60 \ \ beats\ per\ min \\ \dfrac{80 - x}{20} & if \ \ 60 < x \le 80 \ \ beats\ per\ min \\ 0 & if \ \ x > 80 \ \ beats\ per\ min \end{cases} \tag{3.36}$$

The heart problem variable can have the value of yes or no. It is normalized in the range of 0 and 1, which indicates the severity of heart problem. The 0 condition represent the healthiest people and the 1 condition represents the most suffered people of heart disease.

The membership function is given as:

$$\mu_{yes}(x)=\begin{cases}0 & if\ \ x<0.25\\ \dfrac{x-0.75}{0.5} & if\ \ 0.25\leq x<0.75\\ 1 & if\ \ x\geq 0.75\end{cases} \tag{3.37}$$

and

$$\mu_{no}(x)=\begin{cases}1 & if\ \ x\leq 0.25\\ \dfrac{0.75-x}{0.5} & if\ \ 0.25<x\leq 0.75\\ 0 & if\ \ x>0.75\end{cases} \tag{3.38}$$

These functions can be plotted as shown in Figure 3.11.

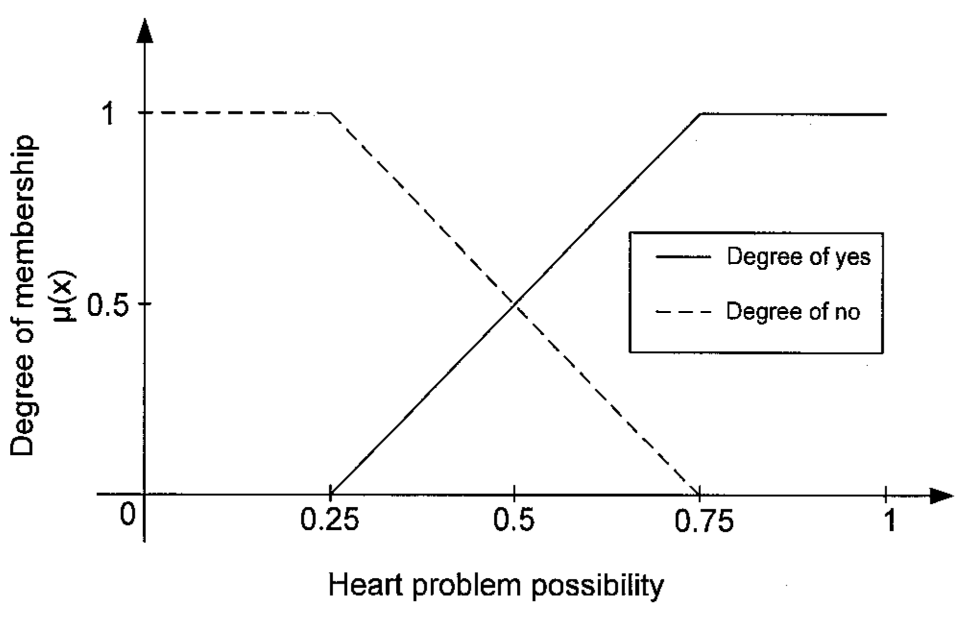
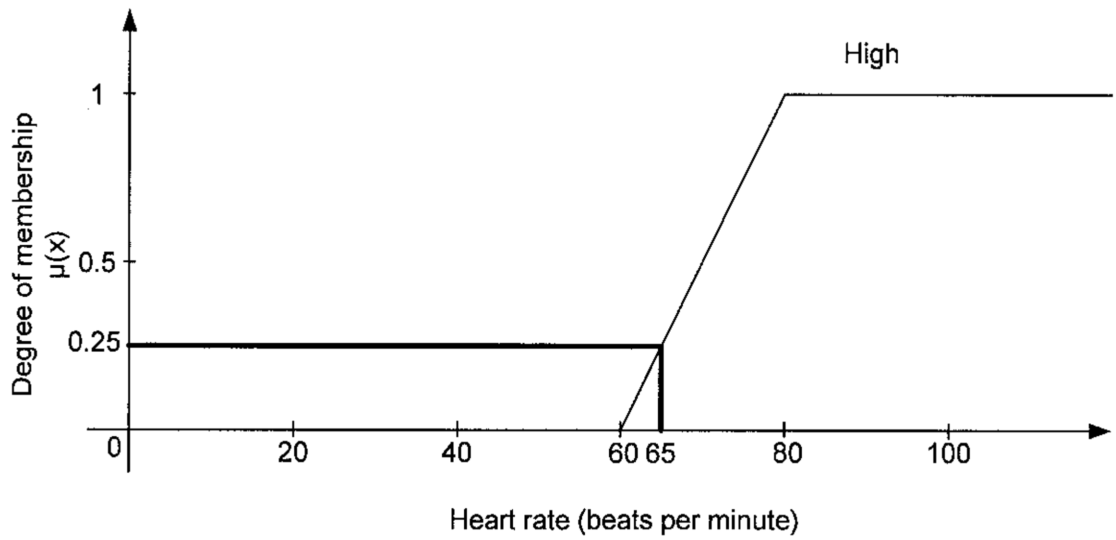


Figure 3.11 Membership function for the degree of yes and no of the heart problem

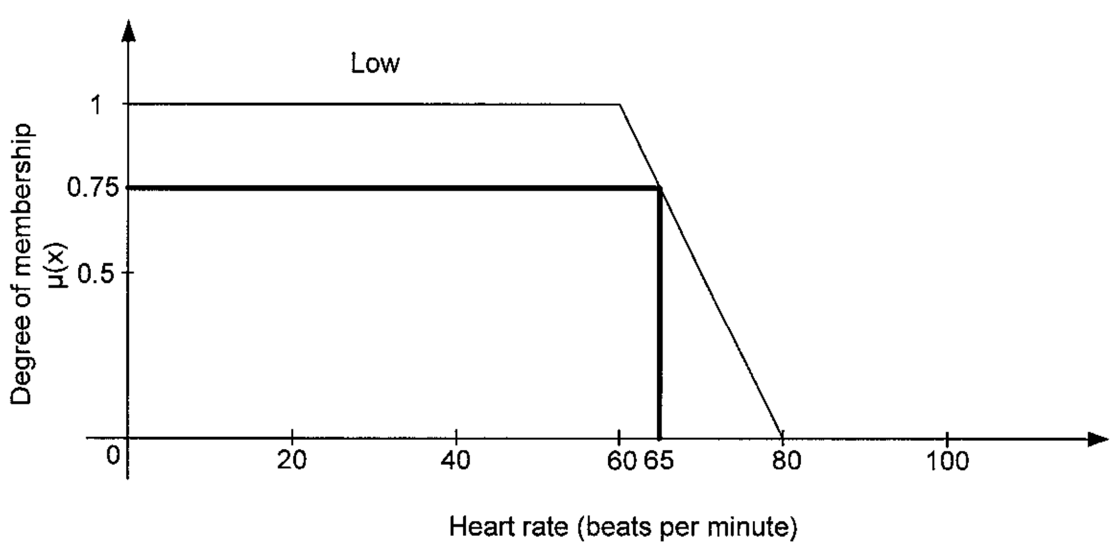
The Mamdani implication can be applied using the procedures as follows [107] :

1. The OR connective is replaced with the max operator.
2. The maximum of two membership functions in the antecedent part of fuzzy rules is evaluated.

3. The Mamdani implication which is min operator is applied between the result of antecedent membership function and consequent membership function.



(a)



(b)

Figure 3.12 Application of or (max) operator for the given rule of (a) high membership function (b) low membership function

Figure 3.12 is an example when the heart rate is 65 beats per minute.

The membership function is:

$$\mu_{heart-rate}(65) = \mu_{low}(65) \vee \mu_{high}(65) = \max[\mu_{low}(65), \mu_{high}(65)]$$
$$= \max\left[\left(\frac{65-60}{20}\right), \left(\frac{80-65}{20}\right)\right] = \max[0.25, 0.75] = 0.75$$

The Mamdani implication operator (*min*) can be applied to the rule antecedent and the rule consequent which is heart problem possibility. The membership function using equation (3.41) of the implication *heart-rate* $\Rightarrow$ *heart problem* is:

$$\mu_{heart-rate \Rightarrow heart-problem}(65, y) = \mu_{heart-rate}(65) \wedge \mu_{heart-problem}(y)$$
$$= \min[\mu_{heart-rate}(65), \mu_{heart-problem}(y)] = \min[0.75, \mu_{heart-problem}(y)]$$

This result can be plotted as follows in Figure 3.13.

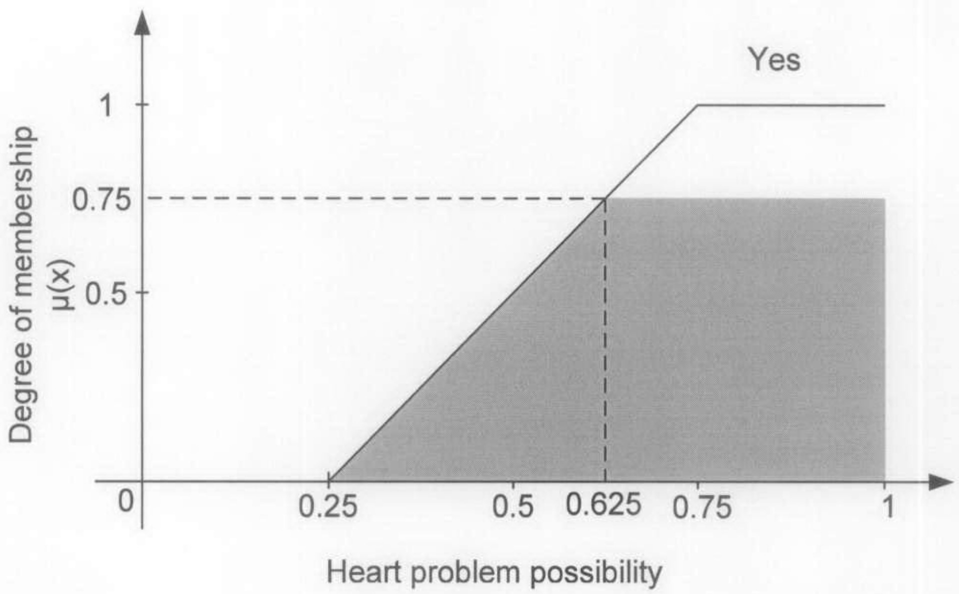


Figure 3.13 Membership function of heart problem based in Mamdani implication.

The grey area is the result of Mamdani implication.

In many cases, it is necessary to convert the fuzzy value to crisp value. The process of converting the fuzzy value into crisp value is called defuzzification. One of the most common defuzzification methods is centroid or centre of gravity defuzzification. If the fuzzy rules are more than one then aggregation of each fuzzy rules antecedent membership function is necessary. The aggregation method usually uses fuzzy union or max operator.

$$\mu_{total} = \cup_i (\mu_i(x)) = \mu_1(x) \vee \mu_2(x) \vee ... \vee \mu_i(x) \qquad (3.39)$$

where i is the number of fuzzy rules.

The centroid defuzzification technique is expressed as:

$$x^* = \frac{\int \mu(x) x dx}{\int \mu(x) dx} \qquad (3.40)$$

The discrete version of centroid defuzzification is:

$$x^* = \frac{\sum_{j=1}^{n} x_j \mu(x_j)}{\sum_{j=1}^{n} \mu(x_j)} \qquad (3.41)$$

The centroid fuzzification is illustrated as follows in Figure 3.14.

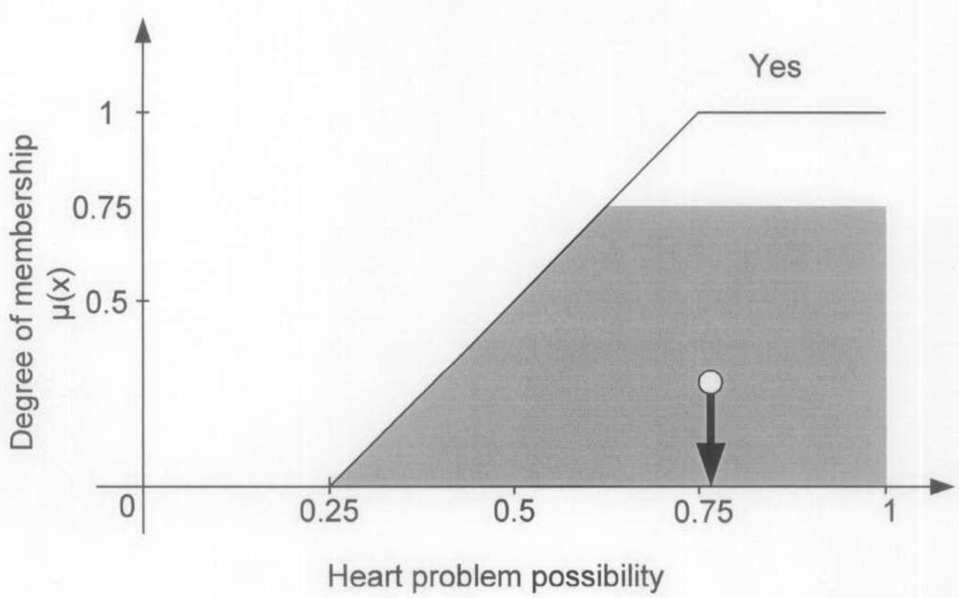


Figure 3.14 Centroid defuzzification

3.7.4 Fuzzy decision support system development

Fuzzy system has been proven to be powerful to many application areas. Fuzzy system has the ability to handle imprecise knowledge by means of fuzzy linguistic terms. The main disadvantage of fuzzy system is the difficulty in preparing the knowledge. Fuzzy system can not learn from the data directly. The need of human experts is obvious in the development of fuzzy system. The involving of human expert makes the fuzzy system subjective. For example, determining the membership functions will need a domain expert.

RST can learn directly from the data. This is the advantage of RST system. The system will be objective because it does not need a domain expert during the development. One of the drawbacks of RST is that the generated rules are crisp instead of fuzzy. Combining RST and fuzzy system will make the system powerful. The system will be objective and be able to handle the imprecise knowledge and fuzziness. The selected generated RST rules are crisp. Fuzzification must be applied to these crisp rules. The fuzzy membership functions are in the form of triangular and trapezoid shapes as shown in Figure 3.15.

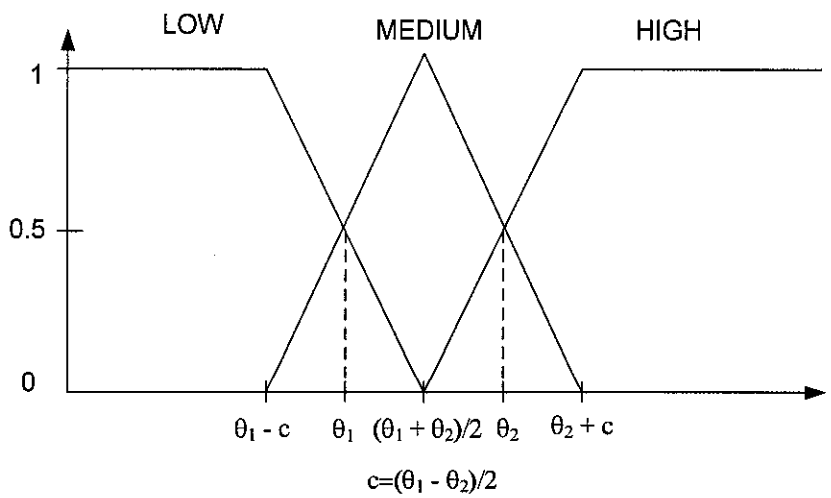


Figure 3.15 Membership function of attribute age

All numerical conditions are fuzzified based on the value of discretization "cuts". As an example, the numerical attribute *age* is discretized using Boolean reasoning into three discrete values which are [*, θ_1), [θ_1, θ_2) and [θ_2, *) meaning "less than θ_1", "greater than or equal θ_1 and less than θ_2 " and " equal or more than θ_2 " respectively. Therefore the "cuts" are θ_1 and θ_2. Two trapezoidal and single triangular membership functions of attribute *age* which are LOW, MEDIUM and HIGH can be generated.

For the attribute that has only two values, the membership function can be made as shown in Figure 3.16.

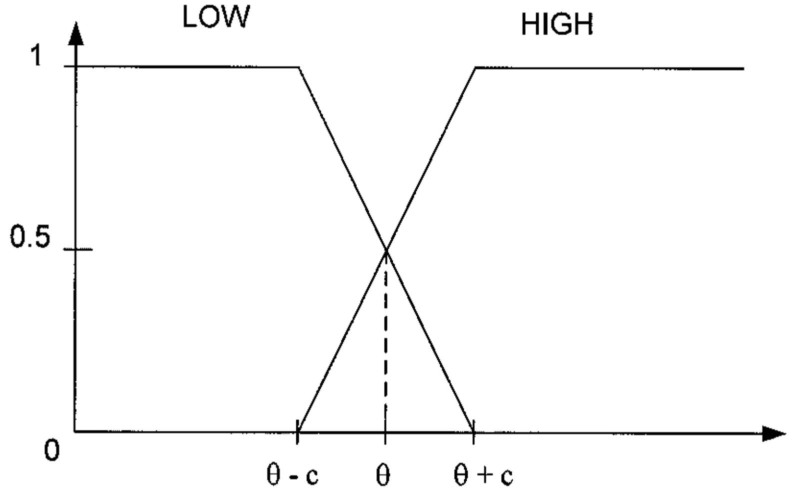


Figure 3.16 Membership function of attribute thalach

In this case, c is determined intuitively. As an example, the numerical attribute *thalach* is discretized using Boolean reasoning into three discrete values which are [*, θ) and [θ, *) meaning "less than θ" and " equal or more than θ " respectively. Therefore the "cuts" are θ. Two trapezoidal and single triangular membership functions of attribute *age* which are LOW and HIGH can be generated. The nominal attributes can have the fuzzy membership function. Crisp value is the fuzzy value with the degree of membership function is 1 with no overlapping between the membership functions. For simplification, no optimization is applied to the fuzzy membership functions. Fuzzy weighing method is introduced here based on selected RST rules support of training data. If the *n*th crisp rule has support $sp(n)$ with i number of rules, then the corresponding fuzzy rule weight is:

$$w(n) = \frac{sp(n)}{\max(sp(1),...,sp(i))} \tag{3.42}$$

and i is the total number of rules.

Consider Table 3.16 for an illustration of fuzzification of crisp rules for medical data in Table 3.2. The rules are crisp.

Table 3.16 Example of crisp rules

No.	Rules
1	oldpeak([0.3, *)) AND slope(2) AND thal(7) => num(1)
2	fbs(0) AND thalach([133, *)) AND slope(1) AND ca ([*, 1)) AND thal(3) => num(0)
3	fbs(0) AND ca ([1, *)) AND thal(7) => num(1)
4	sex(1) AND fbs(0) AND thalach([133, *)) AND exang(0) AND ca ([*, 1)) AND thal(3) => num(0)
5	sex(1) AND fbs(0) AND restecg(0) AND oldpeak([0.3, *)) AND thal(7) => num(1)

After fuzzification, the fuzzy rules generated are shown in Table 3.17.

Table 3.17 Example of fuzzy rules

No.	Rules
1	IF oldpeak is high AND slope is flat AND thal is reversable defect THEN CAD is yes
2	IF fbs is low AND thalach is high AND slope is upsloping AND ca is low AND thal is normal THEN CAD is no
3	IF fbs is low AND ca is high AND thal is reversable defect THEN CAD is yes
4	IF sex is male AND fbs is low AND thalach is high AND exang is no AND ca is low AND thal is normal THEN CAD is no
5	IF sex is male AND fbs is low AND restecg is normal AND oldpeak is high AND thal is reversable defect THEN CAD is yes

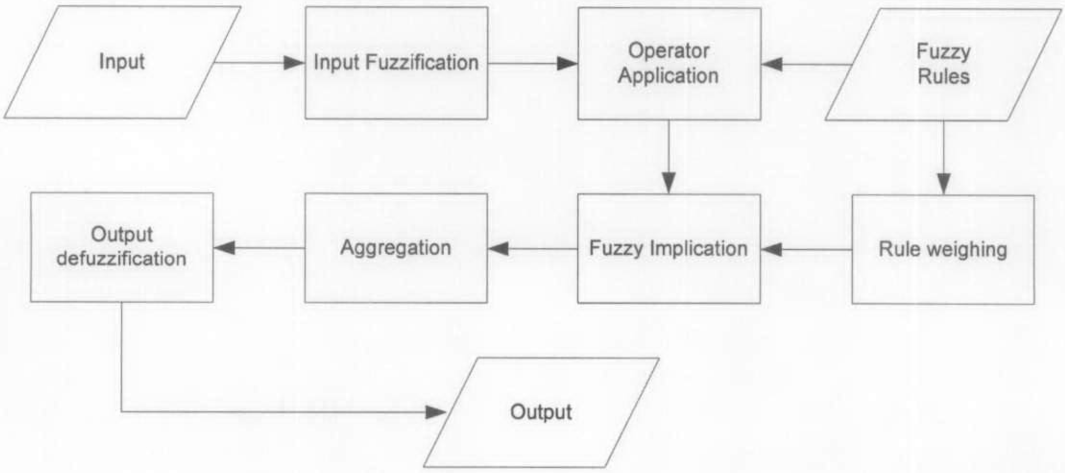


Figure 3.17 Fuzzy inference flowchart

The Mamdani inference engine is used for the inference process. Centroid defuzzification is chosen to get the numerical output of CAD diagnosis. For AND operator and implication, *min* function is used. The function of *sum* is used for aggregation. MATLAB Fuzzy Logic Toolbox is used in the development. The process of fuzzy inferencing steps shown in Figure 3.17 can be explained as follows:

Step 1

Input fuzzification. The first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. The input is always a crisp numerical value limited to the universe of discourse of the input variable and the output is a fuzzy degree of membership in the qualifying linguistic set (always the interval between 0 and 1). Nominal values which are crisp also can be mapped to their respective membership function with membership degree is 1. When there are missing attribute values, all the respective membership function will take the maximum values.

Step 2

Fuzzy operator application. Once the inputs have been fuzzified, the degree to which each part of the antecedent has been satisfied for each rule is known. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule. This number will then be applied to the output function. The input to the fuzzy operator is two or more membership values from fuzzified input variables. In this work, the rules only have AND operator which is *min* operator because there is no OR operator in the extracted rules.

Step 3

Implication method. Before applying the implication method, rule's weight must be considered. Every rule has a weight defined in Equation (3.47), which is applied to the number given by the antecedent. Once proper weighting has been assigned to each rule, the implication method is implemented. A consequent is a fuzzy set represented by a membership function, which weights appropriately the linguistic characteristics that are attributed to it. The consequence is reshaped using a function associated with the antecedent (a single number). The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Implication is implemented for each rule. The function of *min* is applied in this work.

Step 4

Output aggregation. Since decisions are based on the testing of all of the rules in a fuzzy inference system, the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable, just prior to the fifth and final step, defuzzification. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable. It may be noticed that as long as the aggregation method is commutative (which it always should be), then the order in which the rules are executed is unimportant. The function of max or *sum* (simply the sum of each rule's output set) is used in this work.

Step 5

Output defuzzification. The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set) and the output is a single number. As much as fuzziness helps the rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. However, the aggregate of a fuzzy set encompasses a range of output values, and so must be defuzzified in order to resolve a single output value from the set. The most popular defuzzification method which is the centroid calculation is used in this work. This method returns the centre of area under the curve. The diagnosis result of CAD is taken from the output. If the output is equal or more than 0.5, then it is considered as CAD. If the output is less than 0.5, then it is not CAD. The numerical value can be considered as the percentage of blocking or possibility of CAD.

The example of inference process of Figure 3.17 for the medical data set is shown in Figure 3.18.

3.7.5 Evaluation of fuzzy decision support system

Fuzzy decision support system (FDSS) undergoes two phases of evaluation. The first phase is evaluation using data mining and knowledge discovery. In this evaluation FDSS is compared to other data mining and knowledge discovery model. Four models of the classifiers are MLP-ANN, k-NN, C4.5 and RIPPER. The WEKA software is used to model these classifiers [108].

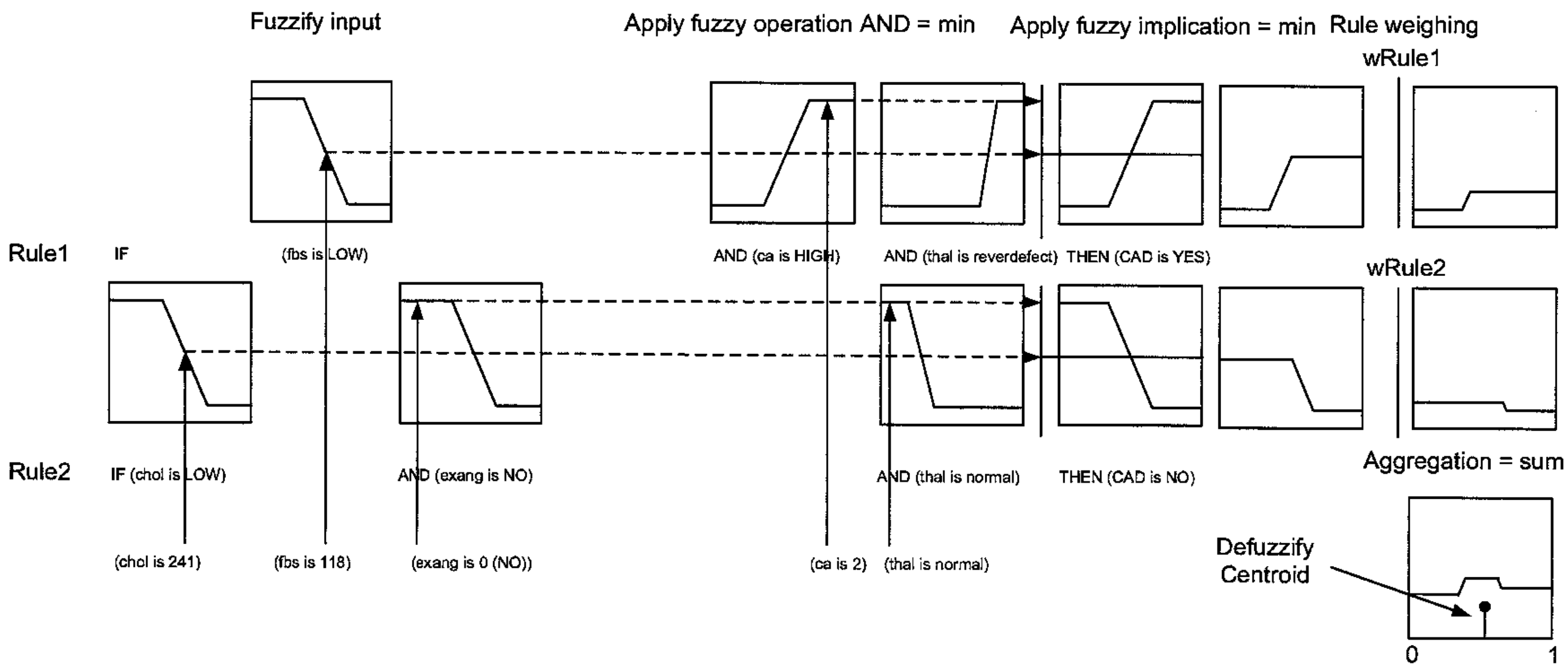


Figure 3.18 Example of fuzzy inference process for medical data set

MLP-ANN is the widely used classifier. The topology is 24 input nodes, 13 hidden nodes and 2 output nodes (Section 3.4). Basic back-propagation learning algorithm (4.6) with learning rate 0.3 and momentum 0.2 is used which is the default value of WEKA. This method usually gives high accuracy but it has lack of knowledge transparency and interpretability. The evaluation process using MLP-ANN is shown in Figure 3.19.

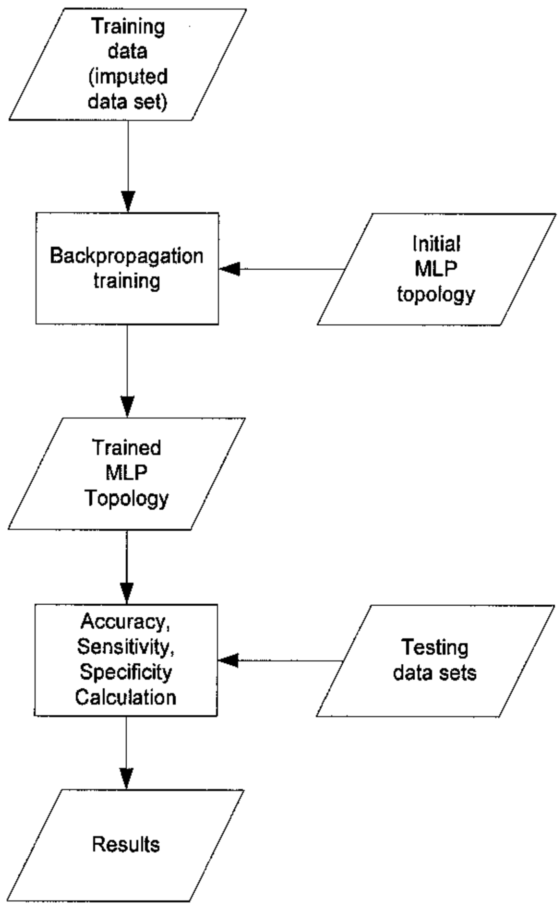


Figure 3.19 Evaluation process of MLP-ANN using four UCI data sets

k-NN is based on the similarity with the training data and it is instance-based learning that generates classification predictions using only specific instances or objects. It uses normalized Euclidean distance to find the training object which is closest to the applied test object, and estimates the same decision as this training object. If multiple objects have the same Euclidean (smallest) distance to the test object, only the first one found is used. The model uses only one neighbour (k=1) [108][109]. The distance between an instance x_1 with the attribute values of $a_1(x_1), a_2(x_1), ..., a_k(x_1)$ (where k is the number of attributes) and an instance x_2 with the attributes values of $a_1(x_2), a_2(x_2),..., a_k(x_2)$ is defined as:

$$\sqrt{(a_1(x_1)-a_1(x_2))^2+(a_2(x_1)-a_2(x_2))^2+...+(a_k(x_1)-a_k(x_2)} \qquad (3.43)$$

The process of classification and its performance of four UCI data sets are illustrated in Figure 3.20.

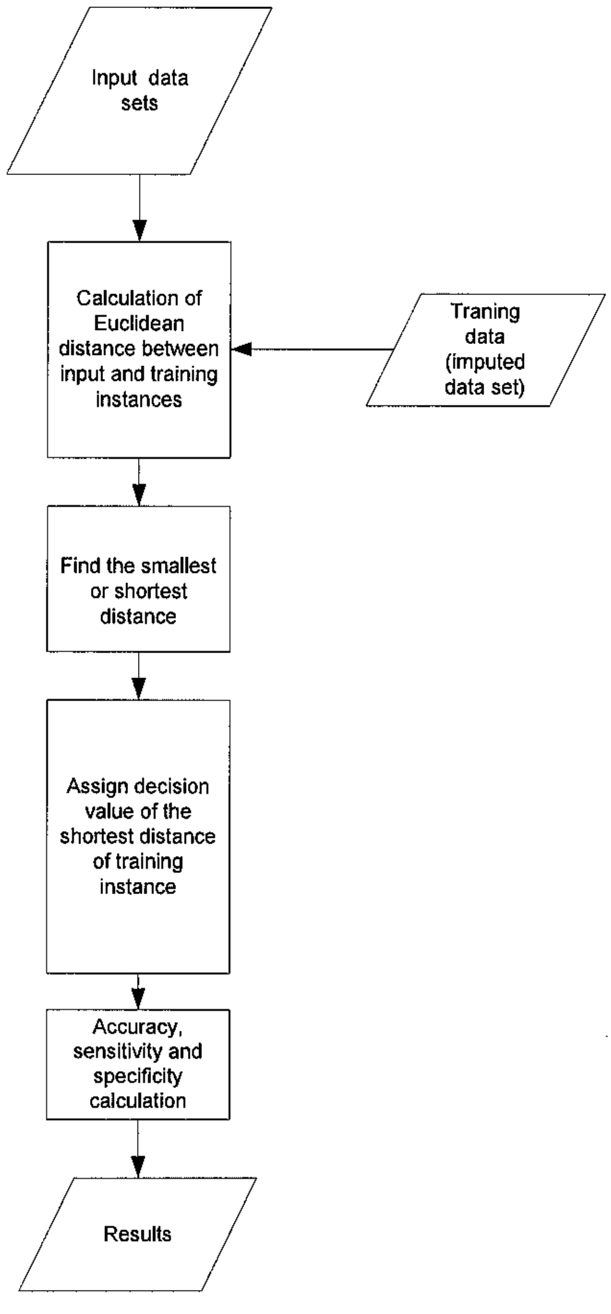


Figure 3.20 The evaluation process of k-NN using four UCI data sets

Usually, normalization of input values to the range between 0 and 1 is carried out before performing Euclidean distance calculation. This method is lack of interpretability and it needs high computation resources when the training data set is large because every single classification will perform the k-NN training from the training set.

C4.5 is a classifier based on Induction Decision Tree (ID3) with some extension [110]. C4.5 builds decision trees from a set of training data based on information entropy concept. The training data is a set of already classified objects. The training data is augmented with the decision attribute and becomes decision table that has condition attributes and decision. The aim is to choose the attributes that contribute best for classifying the objects into certain categories of decision. The information entropy is used to measure the contribution of these attributes. Every attribute's entropy is measured then the best one, which has highest information gain, is chosen. This first chosen attribute is the root of the tree. For the levels following the root of the tree, the same procedure is repeated recursively for the further classification of the data. If there are N decision classes, proportion of instances with classification i is denoted by p_i for i= 1 ... N and p_i is the value of the number of occurrence of class i divided by the total number of instances. The information entropy of variable X is defined as

$$E(X) = -\sum_{i=1}^{N} p_i \log_2(p_i) \tag{3.44}$$

Suppose S is a candidate split that partitions training data T into several subsets $T_1, T_2, \ldots, T_K$. The new entropy can be calculated using

$$E_S(T) = \sum_{j=1}^{K} P_j H_S(T_j) \tag{3.45}$$

where P_j represents the proportion of object record in subset j. Then information gain can be defined as

$$Gain(S) = E(T) - E_S(T) \tag{3.46}$$

C4.5 uses the biggest information gain to choose the optimal split in each decision node. The basic algorithm of C4.5 can be illustrated in Figure 3.21. As an extension of ID3 which can not handle continuous value, C4.5 can handle continuous attribute by using threshold that categorize into two classes which are less than threshold and more than threshold. C4.5 is able to operate with the data which has missing value by marking it as "?" and then ignored during the calculation of gain and entropy [43]. C4.5 is decision tree based classifier which is less interpretable than rule based classifier.

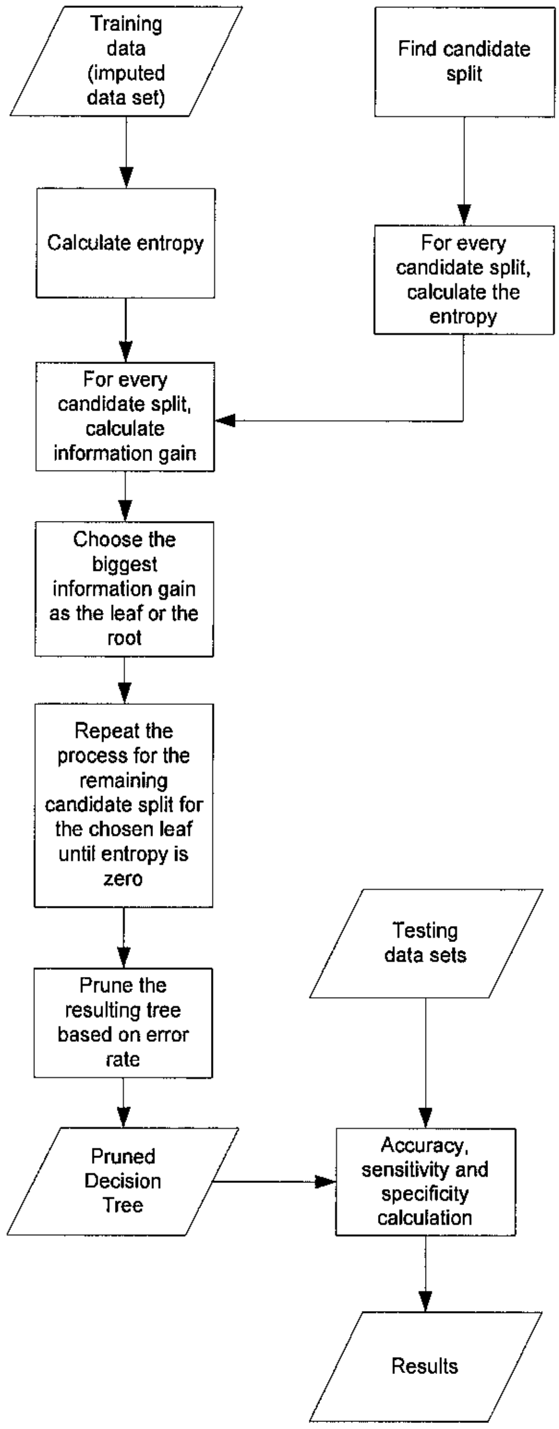


Figure 3.21 The evaluation process of C4.5 using four UCI data sets

RIPPER (Repeated Incremental Pruning to Produce Error Reduction) is classifier based on rule generation. It starts from empty rule set. For every class of the decision, rule generation is applied by starting from the less prevalent one to the more frequent one. The rule generation algorithm is by growing one rule through the adding of antecedents to the rule until the rule is perfect (i.e. 100% accurate). The procedure tries every possible value of each attribute and selects the condition with highest information gain. After this rule generation, the algorithm incrementally prunes each rule and allows the pruning of any final sequences of the antecedents. This process of rule growing and pruning is controlled by the description length and error rate. RIPPER is greedy algorithm which needs high computing resources. Figure 3.22 show the evaluation process using RIPPER algorithm. The stopping condition of the growing-pruning process are (a) there are no more uncovered examples or (b) the description length (DL) of rule set and examples is 64 bits greater than the smallest DL found so far, or (c) the error rate exceeds 50%. At the optimization stage, for every rule, again the training data set is split into growing set and pruning set, but all the instances of pruning set that are covered by another rules are removed. Growing-pruning process is applied to the new generated rule set and the previous rule set that added greedily with antecedents [108][111].

C4.5 and RIPPER can be considered similar method to the RST, but C4.5 and RIPPER are based on information entropy and RST is based on discernibility relation that is based on classification of the data. MLP-ANN and k-NN are the effective classifier but they do not produce rules.

FDSS developed here and the other classifiers are applied on all four sources of UCI CAD data sets and 22 samples from Ipoh Specialist Hospital. Accuracy, sensitivity and specificity are calculated and compared.

The second phase of FDSS evaluation is the verification and the validation of the results obtained in the first phase by cardiologists. Thirty samples taken randomly from each source of UCI CAD data sets and 22 samples taken from Ipoh Specialist Hospital are diagnosed by FDSS and verified by three cardiologists. The accuracy, sensitivity and specificity are also calculated and compared.

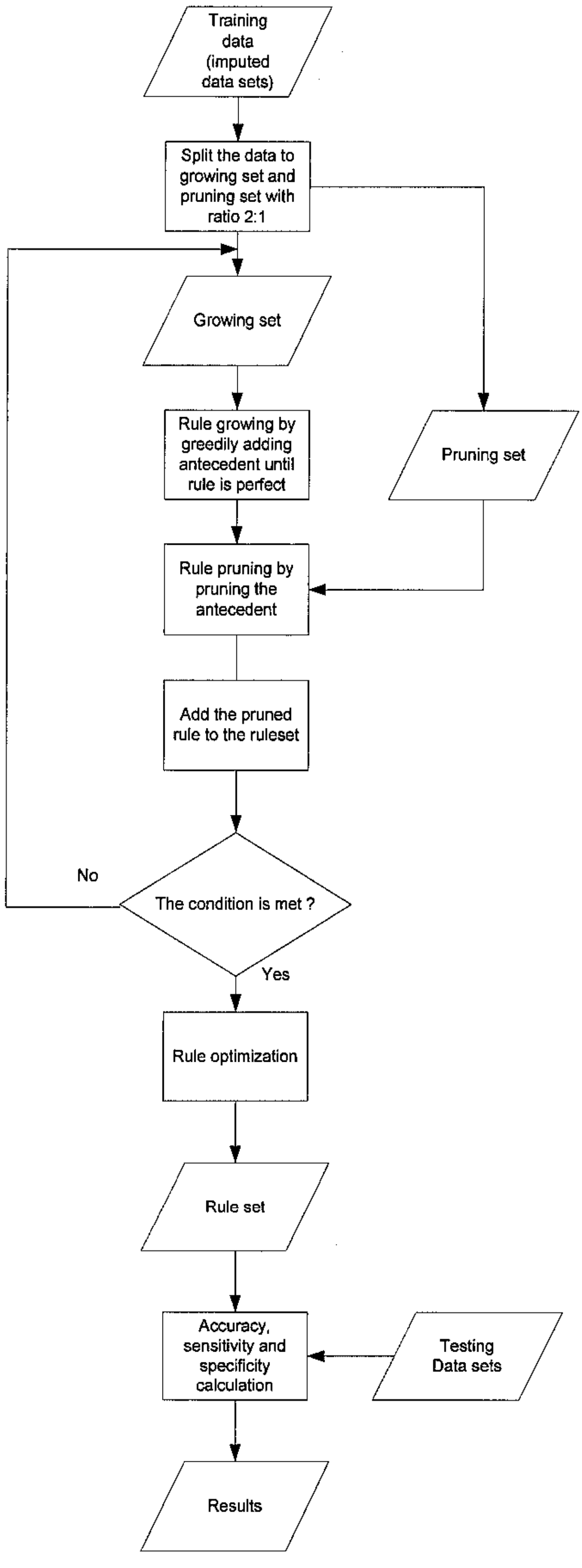


Figure 3.22 The evaluation process of RIPPER using four UCI data sets

3.8 SUMMARY

A new method for development of FDSS for the diagnosis of coronary heart disease has been presented. The model developed uses coronary artery disease data sets from UCI database repository and data collected in local hospital (Ipoh Specialist Hospital). All of four sources have been used in the development of FDSS. The development of FDSS started from new missing data imputation by using the hybrid MLP-ANN and RST namely ANNRST. RST is used as a feature selection of the ANN input. ANNRST is applied to impute simulated missing values of UCI data sets. The accuracy among different methods which are ANN, CMCF, k-NN and ANNRST is calculated. Maximum accuracy among 50 training-testing simulations of ANN and ANNRST is considered as the accuracy for these two methods. Average accuracy for 50 training-testing simulations with different weight initializations using Nguyen-Widrow for ANN and ANNRST is calculated to find the stability of these two methods. The higher average accuracy means the higher probability to get the better accuracy. The new evaluation method by applying ANNRST to impute real missing values is introduced. The evaluation is conducted by discovering the knowledge from the imputed data set and then building the classifier based in the discovered knowledge to classify the complete data set. The accuracy and coverage of the classifier that is based on the knowledge of certain imputed data set can be considered as the performance of those certain imputation method. The performance of discovering rules by RST during rule filtering is also introduced as the new evaluation of the imputation methods. The next stage is rule generation based on RST. The new rule filtering scheme which combines rule quality and rule importance is proposed. The rule importance selection is based on RST. The filtered or selected rules are used as fuzzy rules to develop the decision support system. The simple fuzzification scheme and the new proposed rule weighing based on the support of rules are introduced to build the fuzzy inference engine of FDSS. Using the fuzzy inference engine, the results of diagnosis will indicate the percentage of coronary artery blocking and possibility level of CAD. Finally the FDSS is evaluated using data mining techniques and verified by three cardiologists.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 INTRODUCTION

As mentioned in the methodology, the aim of this research is to develop an evidence based decision support system for the diagnosis of coronary artery disease that learns directly from raw data. The system must be able to handle incomplete and continuous data with high accuracy at acceptable coverage and interpretable or transparent knowledge. The first problem that must be solved is incompleteness of the training data which imputation is needed. The second problem is managing too many generated rules. The last problem is building the fuzzy decision support system.

4.2 MISSING DATA IMPUTATION

One of the important steps of knowledge discovery task is data preprocessing. Missing data imputation is part of data preprocessing step. The quality of knowledge discovery, which is rule extraction, can be degraded when the data consists of many missing values. Therefore, imputation method has significant impact to the quality of knowledge discovery. A large number of high quality complete data sets are not always available. It is very difficult to get both large number and complete data sets. The knowledge discovery must be able to handle incomplete and very limited number of data sets using imputation of missing attribute values. The data sets explained in section 3.3 are used.

4.2.1 ANNRST data imputation for simulated missing data

Most of the researchers used Cleveland data set that consists of 303 patients because it contains only six patients that have incomplete attribute values. To justify how strong the proposed imputation method, more data is needed. All the UCI-CAD data sets, except Cleveland, have severe incompleteness. These data sets still can be used by selecting only the complete attributes to verify the quality of proposed imputation methods. Most of the values of the attributes namely the slope of the peak of ST segment (*slope*), number of major vessels coloured by fluoroscopy (*ca*), and exercise thallium scintigraphic defects (*thal*) in the Hungarian and Long Beach data sets are missing. Hence they are not used in this ANNRST model verification. All the objects of Switzerland data set are not used due to severe incompleteness.

After preprocessing by removing the objects having too many missing values on their attributes and removing the slope of the peak of ST segment (*slope*), number of major vessels coloured by fluoroscopy (*ca*), and exercise thallium scintigraphic defects (*thal*), the CAD decision table consists of 597 objects with 10 attributes and single decision as shown in Table 4.1. ROSETTA is used for RST data analysis. Boolean reasoning based discretization algorithm as shown in equations (3.14 – 3.18) is applied to discretize the numerical attributes.

Table 4.1 Data set in the form of decision table

No	*age*	*sex*	*cp*	*trestbps*	*chol*	*fbs*	*restecg*	*thalach*	*exang*	*oldpeak*	*num*
1	63	1	1	145	233	1	2	150	0	2.3	0
2	37	1	3	130	250	0	0	187	0	3.5	0
3	41	0	2	130	204	0	2	172	0	1.4	0
4	56	1	2	120	236	0	0	178	0	0.8	0
5	57	0	4	120	354	0	0	163	1	0.6	0
6	57	1	4	140	192	0	0	148	0	0.4	0
7	56	0	2	140	294	0	2	153	0	1.3	0
8	44	1	2	120	263	0	0	173	0	0	0
9	52	1	3	172	199	1	0	162	0	0.5	0
10	57	1	3	150	168	0	0	174	0	1.6	0
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
588	61	1	3	120	337	0	0	98	1	0	1
589	58	1	3	150	219	0	1	118	1	0	1
590	74	1	4	155	310	0	0	112	1	1.5	1
591	62	1	4	160	254	1	1	108	1	3	1
592	53	1	4	144	300	1	1	128	1	1.5	1
593	62	1	4	158	170	0	1	138	1	0	1
594	46	1	4	134	310	0	0	126	0	0	1
595	55	1	4	122	223	1	1	100	0	0	1
596	62	1	2	120	254	0	2	93	1	0	1
597	63	1	4	140	260	0	1	112	1	3	1

The decision table of data set is then reduced by RST using Johnson's computation of equation (3.8) with the processing steps as follows:

Let B represent the reduct. U denotes the set of all sets corresponding to the function f_C as shown in equation (3.8) and $w(S)$ denotes a weight for set S in U that is automatically computed from the data [55].

i. Let $B = \varnothing$
ii. Let c denotes the variable that maximizes $\sum w(S)$, where the sum is taken over all sets S in U that contain c.
iii. Add c to B.
iv. Remove all sets S from U that contain c.
v. If $S = \varnothing$ return B. Otherwise, go to step ii.

Because the above algorithm finds the reduct that starts from empty set, then the result will be the shortest reduct without calculating all possible reducts. Johnson's computation of reduct using ROSETTA results in reduct that has five attributes, which are {*age, trestbps, chol, thalach, oldpeak*}.

Fasting blood sugar over 120 mg/dl (*fbs*), resting electrocardiographic results (*restecg*), and exercise inducing angina (*exang*) are arbitrarily chosen as the attributes that have missing values. Ten to 300 missing values are simulated in the table on *fbs*, *restecg*, and *exang* attributes. For comparison, ANN without RST attribute reduction, k-Nearest Neighbour (k-NN), and Concept Most Common value Filling (CMCF) are applied to the data set for data imputation.

ANN and ANNRST do not always give the best accuracy for every training process. Many training-testing simulations need to be conducted to find the best accuracy. 50 training-testing simulations with Nguyen-Widrow weight initialization are chosen (Section 3.6.1), because more than 30 samples are enough to give normal distribution in statistical point of view. The best accuracy of 50 training-testing simulations is called maximum accuracy. The average of 50 training-testing simulations accuracy is called average accuracy. Hence the term of average accuracy is found only in ANN and

ANNRST. CMCF and k-NN do not have average accuracy. They only have the accuracy which is called maximum accuracy in this work.

The accuracy among different methods is calculated. Average accuracy for 50 training-testing simulations with different weight initializations using Nguyen-Widrow for ANN and ANNRST is calculated to find the stability of these two methods. The higher average accuracy means the model (ANN or ANNRST) has the higher probability to have better accuracy.

Performance of ANNRST, ANN, k-NN and CMCF on simulated data sets for attributes *fbs*, *exang* and *restecg* are shown in Figure 4.1 to Figure 4.6. It can be seen from Table 4.3 to Table 4.5 and Figure 4.1 to Figure 4.3 that the accuracy of ANNRST is comparable with that of the ANN and CMCF and better than that of the k-NN.

Figure 4.1 Comparison of maximum accuracy among imputation methods for attribute *fbs*.

Table 4.2 Comparison of maximum accuracy among imputation methods for attribute *fbs*

	Number of Missing Data	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	Average
Maximum Accuracy	ANN	0.90	0.80	0.80	0.75	0.74	0.75	0.76	0.75	0.74	0.77	0.77	0.78	0.77	0.77	0.78	0.79	0.79	0.79	0.79	0.80	0.80	0.81	0.82	0.83	0.83	0.83	0.83	0.84	0.84	0.85	0.80
	ANNRST	0.90	0.80	0.73	0.75	0.76	0.72	0.73	0.73	0.73	0.75	0.77	0.78	0.77	0.76	0.77	0.79	0.79	0.79	0.79	0.79	0.80	0.81	0.81	0.82	0.83	0.83	0.83	0.84	0.84	0.85	0.79
	k-NN	0.70	0.65	0.67	0.68	0.66	0.62	0.64	0.63	0.62	0.64	0.64	0.66	0.68	0.66	0.67	0.69	0.69	0.69	0.73	0.73	0.73	0.75	0.73	0.75	0.76	0.77	0.77	0.78	0.78	0.78	0.70
	CMCF	0.80	0.75	0.70	0.70	0.72	0.72	0.73	0.73	0.73	0.75	0.76	0.77	0.76	0.76	0.77	0.78	0.78	0.78	0.78	0.79	0.80	0.80	0.81	0.82	0.82	0.83	0.83	0.84	0.84	0.85	0.78

From Figure 4.1 and Table 4.2, it can be seen that the average accuracy of ANNRST is somewhat less than the average accuracy of ANN. The maximum accuracy of ANNRST is 0.9 for 10 missing data, which is the same as ANN.

CMCF has average accuracy close to ANNRST but the individual accuracy for each number of missing data is always less than or equal that of the ANNRST. The imputation ability of CMCF depends on the nature of data. Attributes *fbs* can only have binary values. In this case CMCF performs well. CMCF has good accuracy because the common value of complete data set is the same as the common value of incomplete data set. The accuracy of k-NN is less than the others.

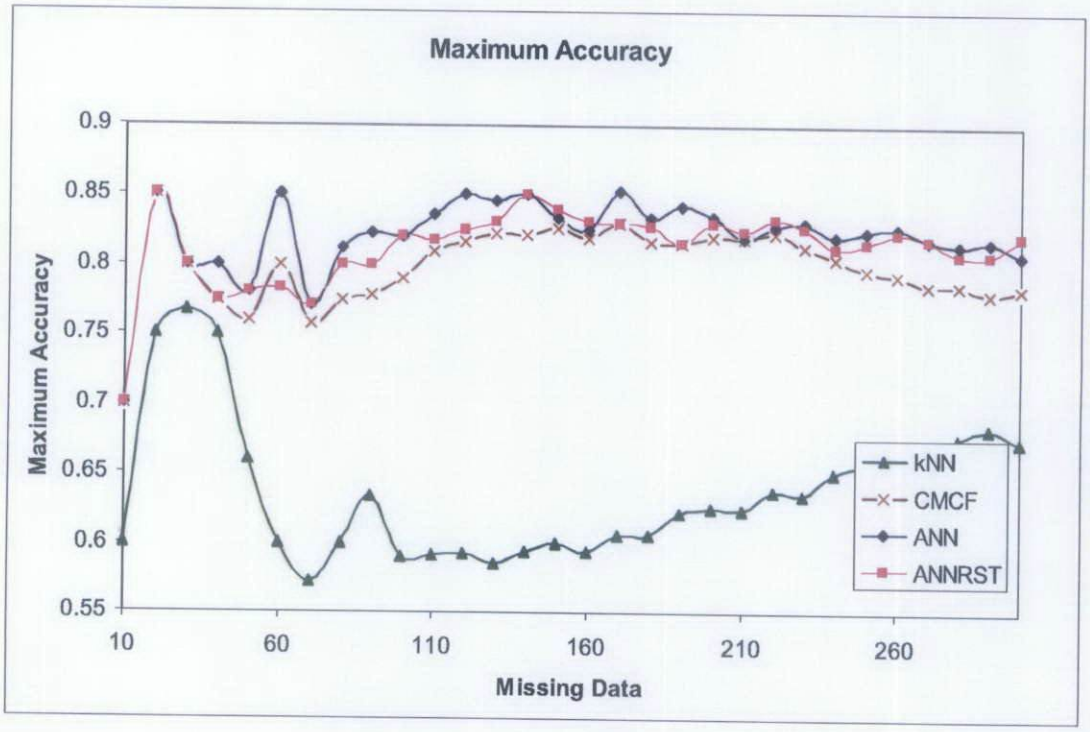


Figure 4.2 Comparison of maximum accuracy among imputation methods for attribute *exang*.

Figure 4.2 shows the similar result as the result in Figure 4.1. ANNRST accuracy is also somewhat less than ANN. CMCF has similar performance compared to ANN and ANNRST but it has less accuracy. k-NN accuracy is also less than the others.

Table 4.3 Comparison of maximum accuracy among imputation methods for attribute *exang*

	Number of Missing Data	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	Average
Maximum Accuracy	ANN	0.70	0.85	0.80	0.80	0.78	0.85	0.77	0.81	0.82	0.82	0.84	0.85	0.85	0.85	0.83	0.83	0.85	0.83	0.84	0.84	0.82	0.83	0.83	0.82	0.82	0.83	0.82	0.81	0.82	0.81	0.82
	ANNRST	0.70	0.85	0.80	0.78	0.78	0.78	0.77	0.80	0.80	0.82	0.82	0.83	0.83	0.85	0.84	0.83	0.83	0.83	0.82	0.83	0.82	0.83	0.83	0.81	0.82	0.82	0.82	0.81	0.81	0.82	0.81
	k-NN	0.60	0.75	0.77	0.75	0.66	0.60	0.57	0.60	0.63	0.59	0.59	0.59	0.58	0.59	0.60	0.59	0.61	0.61	0.62	0.63	0.62	0.64	0.63	0.65	0.66	0.67	0.67	0.68	0.68	0.67	0.64
	CMCF	0.70	0.85	0.80	0.78	0.76	0.80	0.76	0.78	0.78	0.79	0.81	0.82	0.82	0.82	0.83	0.82	0.83	0.82	0.82	0.82	0.82	0.82	0.81	0.80	0.80	0.79	0.79	0.79	0.78	0.78	0.80

Table 4.3 shows that ANNRST has comparable performance compared to ANN. This table also shows that CMCF has average accuracy somewhat less than ANNRST. This result is similar to the imputation of the attribute *fbs*. k-NN average accuracy is the worst among the others.

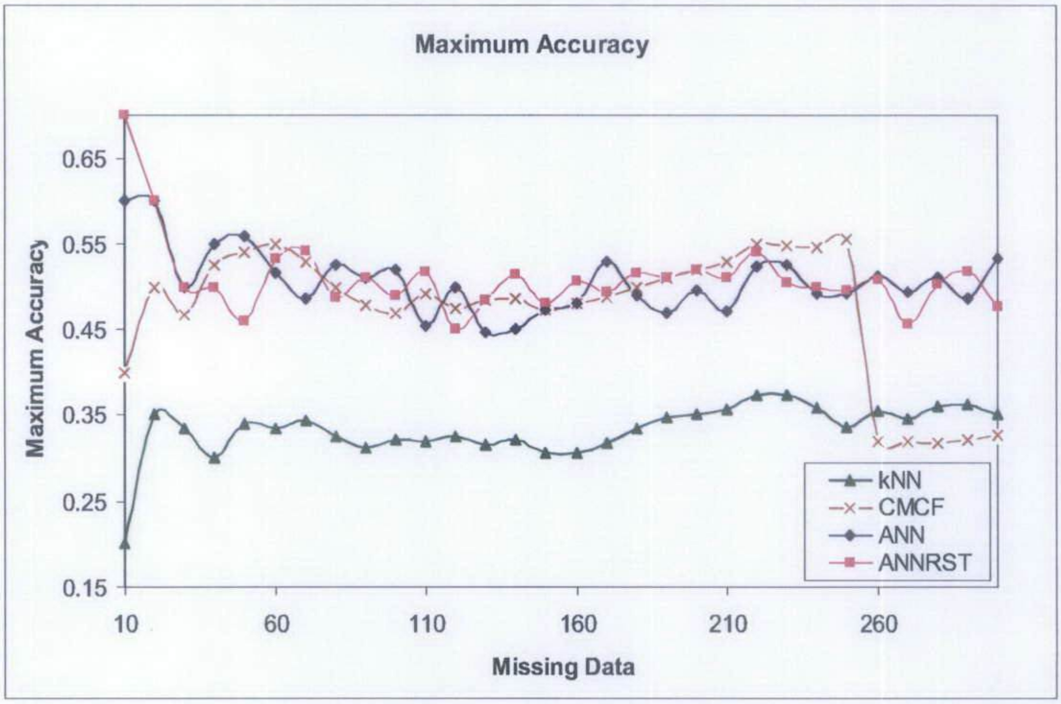


Figure 4.3 Comparison of imputation accuracy among methods for attribute *restecg*.

Figure 4.3 shows the similar result to Figure 4.2 except for CMCF, it decreases extremely in the case of 260 missing data. It may happen because the common value in the complete data set is different from the common value in the incomplete data set when the missing data is 260. This is the drawback of CMCF.

The accuracy of all imputation methods in the attribute *restecg* is worse than that in the previous attributes (*fbs* and *exang*). This is because of the attribute *restecg* can have three nominal values, where *fbs* and *exang* only have binary value.

Table 4.4 Comparison of maximum accuracy among imputation methods for attribute *restecg*

	Number of Missing Data	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	Average
Maximum Accuracy	ANN	0.60	0.60	0.50	0.55	0.56	0.52	0.49	0.53	0.51	0.52	0.45	0.50	0.45	0.45	0.47	0.48	0.53	0.49	0.47	0.50	0.47	0.52	0.53	0.49	0.49	0.51	0.49	0.51	0.49	0.53	0.51
	ANNRST	0.70	0.60	0.50	0.50	0.46	0.53	0.54	0.49	0.51	0.49	0.52	0.45	0.48	0.51	0.48	0.51	0.49	0.52	0.51	0.52	0.51	0.54	0.50	0.50	0.50	0.51	0.46	0.50	0.52	0.48	0.51
	k-NN	0.20	0.35	0.33	0.30	0.34	0.33	0.34	0.33	0.31	0.32	0.32	0.33	0.32	0.32	0.31	0.31	0.32	0.33	0.35	0.35	0.36	0.37	0.37	0.36	0.34	0.35	0.34	0.36	0.36	0.35	0.33
	CMCF	0.40	0.50	0.47	0.53	0.54	0.55	0.53	0.50	0.48	0.47	0.49	0.48	0.48	0.49	0.47	0.48	0.49	0.50	0.51	0.52	0.53	0.55	0.55	0.55	0.56	0.32	0.32	0.32	0.32	0.33	0.47

The performance of ANN and ANNRST during the learning process with different weight initialization is shown in Table 4.5 to Table 4.7 and Figure 4.4 to Figure 4.6. These tables and figures show only ANN and ANNRST and do not show k-NN and CMCF, which are non iterative method means k-NN and CMCF, always give the constant results for the same number of missing values. The average accuracy of ANNRST on 50 simulations running with different weight initialization as shown in Table 4.5 to Table 4.7 and Figure 4.4 to Figure 4.6 is slightly better than that of the ANN. It means that ANNRST is more stable than ANN to achieve better imputation accuracy.

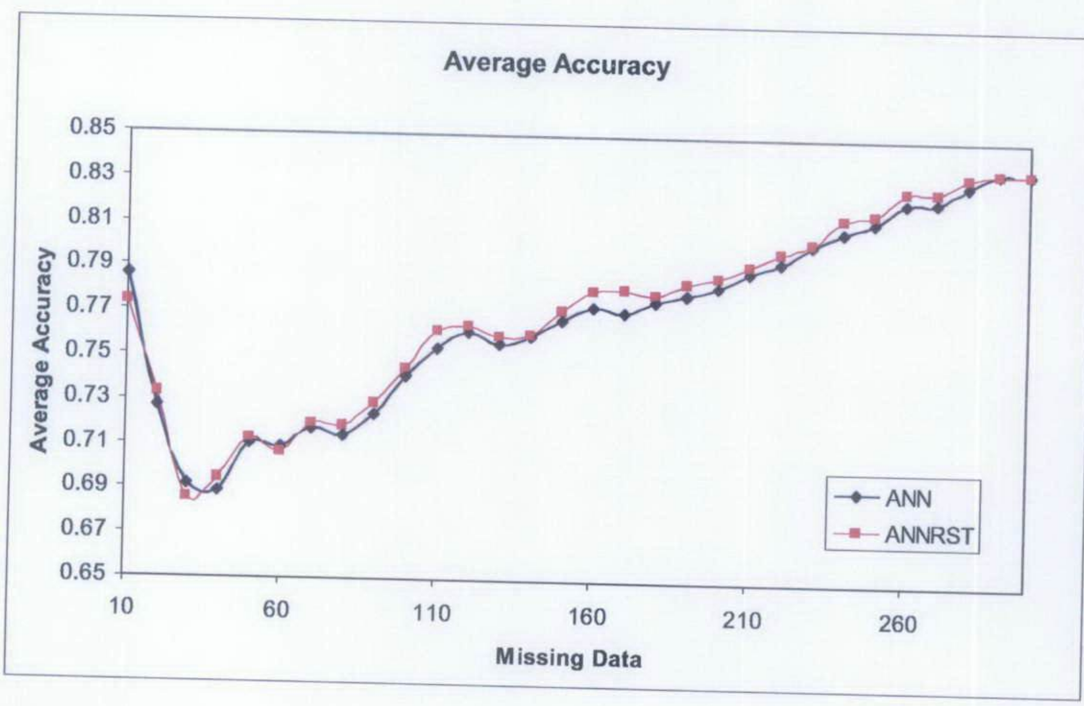


Figure 4.4 Comparison of average accuracy among ANN and ANNRST methods for attribute *fbs*.

Figure 4.4 shows that ANNRST has slightly better average accuracy than ANN. The resulting graph is quite similar between ANN and ANNRST.

Table 4.5 Comparison of average accuracy among ANN and ANNRST methods for attribute *fbs*

Number of Missing Data		10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	Average
Average Accuracy	ANN	0.786	0.727	0.692	0.689	0.710	0.709	0.717	0.714	0.724	0.741	0.754	0.761	0.756	0.760	0.767	0.773	0.771	0.776	0.779	0.783	0.790	0.794	0.802	0.808	0.813	0.822	0.823	0.830	0.835	0.836	0.768
	ANNRST	0.774	0.733	0.686	0.695	0.712	0.707	0.719	0.719	0.729	0.745	0.762	0.764	0.760	0.760	0.772	0.781	0.781	0.779	0.784	0.787	0.793	0.798	0.803	0.814	0.817	0.827	0.827	0.834	0.836	0.836	0.771

In Table 4.5, the average of average accuracy of ANNRST is better than ANN. ANNRST average accuracy outperforms ANN for almost of the number of missing data.

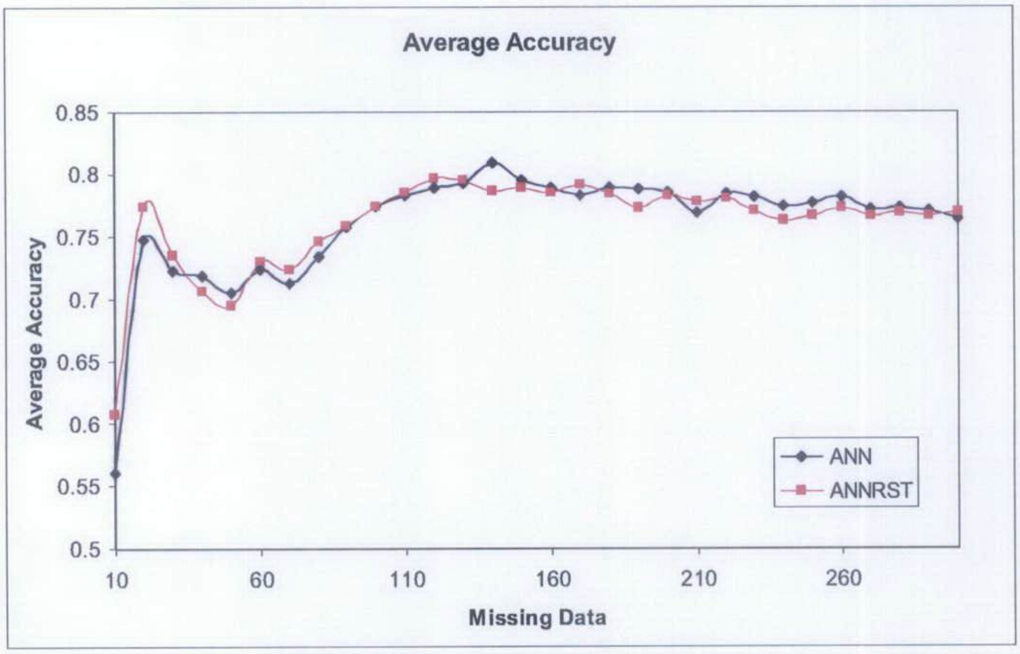


Figure 4.5 Comparison of average accuracy among ANN and ANNRST methods for attribute *exang*

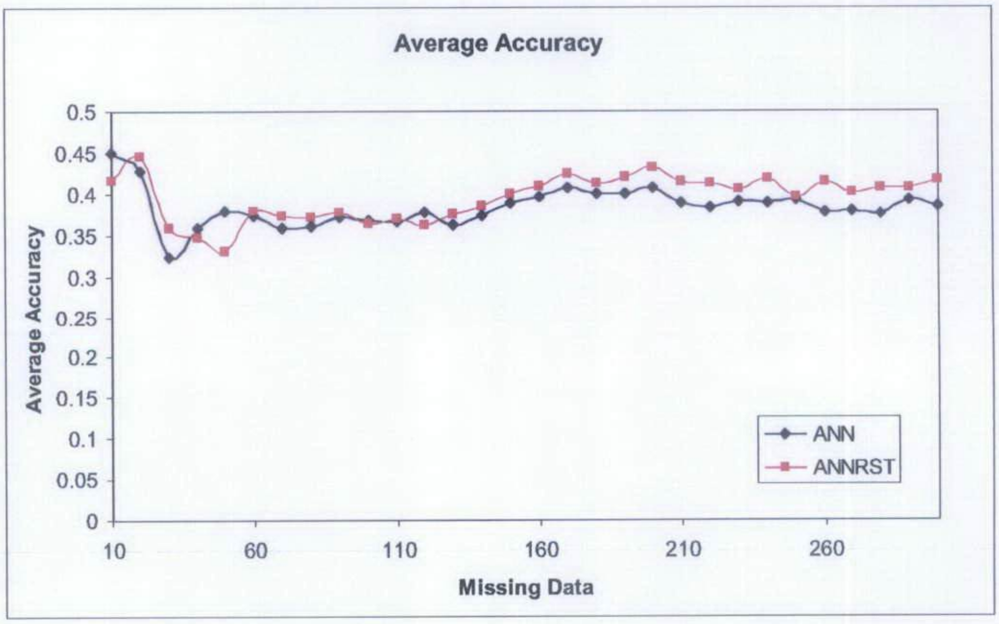


Figure 4.6 Comparison of imputation average accuracy among methods for attribute *restecg*

Table 4.6 Comparison of average accuracy among ANN and ANNRST methods for attribute *exang*

Number of Missing Data		10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	Average
Average Accuracy	ANN	0.560	0.747	0.723	0.719	0.705	0.723	0.712	0.733	0.757	0.774	0.783	0.788	0.792	0.809	0.795	0.789	0.783	0.789	0.788	0.785	0.768	0.784	0.781	0.774	0.776	0.781	0.771	0.772	0.770	0.764	0.760
	ANNRST	0.608	0.774	0.735	0.706	0.695	0.730	0.723	0.746	0.759	0.774	0.785	0.796	0.795	0.787	0.789	0.786	0.792	0.784	0.772	0.782	0.778	0.780	0.770	0.763	0.766	0.772	0.766	0.769	0.766	0.768	0.761

Table 4.7 Comparison of average accuracy among ANN and ANNRST methods for attribute *restecg*

Number of Missing Data		10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
Average Accuracy	ANN	0.450	0.428	0.323	0.359	0.380	0.373	0.359	0.361	0.373	0.367	0.366	0.377	0.362	0.374	0.388
	ANNRST	0.416	0.447	0.359	0.348	0.332	0.379	0.373	0.373	0.378	0.364	0.369	0.362	0.375	0.384	0.399

Number of Missing Data		160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	Average
Average Accuracy	ANN	0.397	0.407	0.400	0.399	0.407	0.389	0.383	0.390	0.388	0.392	0.377	0.379	0.376	0.392	0.385	0.383
	ANNRST	0.409	0.424	0.412	0.421	0.432	0.414	0.413	0.406	0.417	0.397	0.414	0.402	0.407	0.407	0.416	0.395

The average accuracy of imputation of 10 to 300 missing data for attributes *fbs*, *exang* and *restecg* are 70.73 % and 70.39% for ANN and ANNRST respectively as shown in Table 4.8. For average accuracy, the averages are 63.70% and 64.22% for ANN and ANNRST respectively as shown in Table 4.9.

Table 4.8 Average of maximum accuracy of ANNRST and ANN

Attributes	Average of maximum accuracy	
	ANNRST	ANN
fbs	78.86%	79.51%
exang	81.21%	82.05%
restecg	51.10%	50.64%
Average=	70.39%	70.73%

Table 4.9 Average of average accuracy of ANNRST and ANN

Attributes	Average of average accuracy	
	ANNRST	ANN
fbs	77.11%	76.80%
exang	76.05%	75.98%
restecg	39.49%	38.32%
Average=	64.22%	63.70%

Average of average accuracy is counted by calculating the average of the mean of 50 simulation accuracies on 10-300 missing values. From the Table 4.8 and Table 4.9, ANNRST can be considered similar to ANN in performance but it has simpler topology and smaller number of input. Smaller input attribute will make ANNRST more flexible for imputation of missing attribute values.

4.2.2 ANNRST data imputation for real missing data

After applying ANNRST to the simulated missing data, ANNRST is applied to the actual missing values on *slope*, *ca* and *thal* attributes using the same process. Using complete decision table, which has 13 attributes as shown in Table 4.1, reduct computation with Boolean reasoning discretization and Johnson's algorithm results in a reduct with the attributes of {*age, trestbps, chol, thalach, oldpeak, ca*}. Using this reduct, then ANNRST for *slope*, *ca* and *thal* imputation is developed.

k-NN imputation method and CMCF are also implemented to estimate the missing values. After missing data estimation, there are three imputed data sets based on ANNRST, k-NN and CMCF. The number of objects is out of 661, 303 objects belong to Cleveland data set. The cross validation accuracy and coverage of all data sets are calculated using different classifiers, which are RST, Decomposition Tree, LTF-C (Local Transfer Function Classifier), and k-NN.

The result of accuracy, coverage, sensitivity and specificity using 10-fold cross validation of randomized and mixed Cleveland, Hungarian and Long Beach data set are shown in Table 4.10 to Table 4.13. The classifiers used are RST, decision tree (D-Tree), k-NN and Local Transfer Function Classifier (LTF-C).

Table 4.10

Accuracy of ten folds cross validation on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Accuracy of Imputation	ANNRST	0.865	0.894	0.861	0.862
	k-NN	0.823	0.847	0.808	0.811
	CMCF	0.9	0.909	0.889	0.906

Table 4.11

Coverage of ten folds cross validation on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Coverage of Imputation	ANNRST	1	0.677	1	1
	k-NN	1	0.639	1	1
	CMCF	1	0.739	1	1

It can be seen in Table 4.10 and Table 4.11 that CMCF data imputation is considered as the best method as its classifier performance of accuracy outperforms the others.

Table 4.12

Sensitivity of ten folds cross validation on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Sensitivity of Imputation	ANNRST	0.842	0.901	0.874	0.832
	k-NN	0.806	0.845	0.808	0.8
	CMCF	0.899	0.911	0.906	0.9

Table 4.13

Specificity of ten folds cross validation on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Specificity of Imputation	ANNRST	0.887	0.889	0.85	0.897
	k-NN	0.839	0.854	0.811	0.82
	CMCF	0.902	0.903	0.876	0.911

It can also be seen in Table 4.12 and Table 4.13 that CMCF data imputation outperforms ANNRST and k-NN as its classifier performance of sensitivity and specificity is better than the others.

However, the objective of ANNRST is to impute missing values that have strong relationship with complete data sets. Ten fold cross validation method can not indicate the strength of the relationship between the imputed data set and the complete data set. This strength of the relationship can be recognized by considering Hungarian and Long Beach data sets with 358 objects, which are almost incomplete as the training data, and Cleveland with 303 objects, which are almost complete as the testing data, that will be classified. This method is called "imputed classify complete" because the imputed data set is used to classify the complete data sets to know the correctness of imputation in term of extracted knowledge from the imputed data set.

Table 4.14

Accuracy of "Imputed classify Complete" on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Accuracy of Imputation	ANNRST	0.815	0.856	0.815	0.812
	k-NN	0.759	0.779	0.729	0.776
	CMCF	0.792	0.77	0.779	0.789

Table 4.15

Coverage of "Imputed classify Complete" on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Coverage of Imputation	ANNRST	1	0.482	1	1
	k-NN	1	0.432	1	1
	CMCF	1	0.617	1	1

Table 4.16

Sensitivity of "Imputed classify Complete" on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Sensitivity of Imputation	ANNRST	0.842	0.877	0.806	0.777
	k-NN	0.712	0.774	0.676	0.741
	CMCF	0.698	0.542	0.691	0.719

Table 4.17

Specificity of "Imputed classify Complete" on three data sets with four classifiers

		Classifiers			
		RST	D-Tree	k-NN	LTF-C
Specificity of Imputation	ANNRST	0.793	0.836	0.823	0.841
	k-NN	0.799	0.783	0.774	0.805
	CMCF	0.872	0.875	0.854	0.848

The results are shown in Table 4.14 to Table 4.17. It is shown that ANNRST is better than CMCF and k-NN on overall accuracy and sensitivity. CMCF has better specificity than ANNRST, but for the accuracy ANNRST is better than CMCF. The gap between sensitivity and specificity of CMCF is 0.872 - 0.698 = 0.174, where the gap between sensitivity and specificity of ANNRST is 0.842 - 0.793 = 0.049. It means that ANNRST is generally better than CMCF. The results show that ANNRST imputation results in data set that is more similar to the complete data set than that of the CMCF. However, CMCF is better than ANNRST while using 10-fold cross validation method.

The example of imputation is illustrated in Table 4.18 which shows 30 samples taken from Cleveland data set. In the table, "?" indicates simulated missing values. The imputed data set using ANNRST is shown in Table 4.19. The darker cells indicate wrong imputed values. The original data set before the missing values are simulated is shown in Table 4.20. There are 3 wrong imputations out of 38 missing values. Hence, the accuracy is 0.92. The blackened cell is the wrong imputation values.

4.2.3 Evaluation of ANNRST based on RST rule filtering

RST classifier is used in this study for extracting rules from data sets. With three sets of imputed data produced by ANNRST, k-NN and CMCF, then RST could extract 3881, 5566 and 1537 rules, respectively. Investigation of the effects of imputation on the quality of RST classifier is carried on while rule filtering based on support is applied. "Support = *m*" means that rules having support below the value of *m* are filtered. During the filtering method the number of rules will be reduced.

Table 4.21 and Figure 4.23 show rule filtering effects on the number of rules. For example when support = 19, then the rules that have support below 19 are filtered so that the number of rules remains 206 as shown in the bold values in Table 4.21.

Table 4.18 Example of data set that contains missing data

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	?	0
67	1	4	160	286	0	2	108	1	1.5	2	?	3	1
67	1	4	120	229	0	2	129	1	2.6	2	?	7	1
37	1	3	130	250	0	0	187	0	3.5	?	0	?	0
41	0	2	130	204	0	2	172	0	1.4	1	?	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	?	0
62	0	4	140	268	0	2	160	0	3.6	?	?	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	?	0
63	1	4	130	254	0	2	147	0	1.4	?	?	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	?	1
57	1	4	140	192	0	0	148	0	0.4	?	?	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	?	0
56	1	3	130	256	1	2	142	1	0.6	2	1	?	1
44	1	2	120	263	0	0	173	0	0	?	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	?	?	0
56	1	2	130	221	0	2	163	0	0	?	0	7	0
56	1	2	120	240	0	0	169	0	0	3	?	?	0
67	1	3	152	212	0	2	150	0	0.8	2	0	7	1
55	0	2	132	342	0	0	166	0	1.2	?	0	?	0
44	1	4	120	169	0	0	144	1	2.8	3	?	?	1
63	1	4	140	187	0	2	144	1	4	1	2	7	1
63	0	4	124	197	0	0	136	1	0	2	0	3	1
41	1	2	120	157	0	0	182	0	0	?	?	?	0
59	1	4	164	176	1	2	90	0	1	?	?	?	1
57	0	4	140	241	0	0	123	1	0.2	2	0	7	1
45	1	1	110	264	0	0	132	0	1.2	2	?	7	1
68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
57	1	4	130	131	0	0	115	1	1.2	?	?	7	1
57	0	2	130	236	0	2	174	0	0	2	1	?	1
59	1	1	134	204	0	0	162	0	0.8	?	2	3	1

Table 4.19 Example of data set that contains imputed values for missing data using ANNRST

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	[illegible]	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	1	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	0	6	0
56	1	2	130	221	0	2	163	0	0	1	0	7	0
56	1	2	120	240	0	0	169	0	0	3	0	3	0
67	1	3	152	212	0	2	150	0	0.8	2	0	7	1
55	0	2	132	342	0	0	166	0	1.2	1	0	3	0
44	1	4	120	169	0	0	144	1	2.8	3	0	6	1
63	1	4	140	187	0	2	144	1	4	1	2	7	1
63	0	4	124	197	0	0	136	1	0	2	0	3	1
41	1	2	120	157	0	0	182	0	0	1	0	3	0
59	1	4	164	176	1	2	90	0	1	2	2	6	1
57	0	4	140	241	0	0	123	1	0.2	2	0	7	1
45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
57	0	2	130	236	0	2	174	0	0	2	1	3	1
59	1	1	134	204	0	0	162	0	0.8	1	2	3	1

Table 4.20 Example of data set before simulating the missing data (complete)

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	3	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
56	1	2	130	221	0	2	163	0	0	1	0	7	0
56	1	2	120	240	0	0	169	0	0	3	0	3	0
67	1	3	152	212	0	2	150	0	0.8	2	0	7	1
55	0	2	132	342	0	0	166	0	1.2	1	0	3	0
44	1	4	120	169	0	0	144	1	2.8	3	0	6	1
63	1	4	140	187	0	2	144	1	4	1	2	7	1
63	0	4	124	197	0	0	136	1	0	2	0	3	1
41	1	2	120	157	0	0	182	0	0	1	0	3	0
59	1	4	164	176	1	2	90	0	1	2	2	6	1
57	0	4	140	241	0	0	123	1	0.2	2	0	7	1
45	1	1	110	264	0	0	132	0	1.2	2	0	7	1
68	1	4	144	193	1	0	141	0	3.4	2	2	7	1
57	1	4	130	131	0	0	115	1	1.2	2	1	7	1
57	0	2	130	236	0	2	174	0	0	2	1	3	1
59	1	1	134	204	0	0	162	0	0.8	1	2	3	1

$\mu = 0.5$ are shown in Table 4.22. Figure 4.8 – 4.10 show the graphical results of Table 4.22. $\mu = 0.5$ is chosen by considering that accuracy is the same important as coverage.

Table 4.22 Filtering effects on accuracy, coverage and Michalski quality measure with $\mu = 0.5$

Support	Accuracy			Coverage			Michalski Quality $\mu = 0.5$		
	ANNRST	k-NN	CMCF	ANNRST	k-NN	CMCF	ANNRST	k-NN	CMCF
0	0.8218	0.7591	0.7987	1.0000	1.0000	1.0000	0.9109	0.8795	0.8993
1	0.8185	0.7591	0.8020	1.0000	1.0000	1.0000	0.9092	0.8795	0.9010
2	0.8185	0.7558	0.7987	1.0000	1.0000	1.0000	0.9092	0.8779	0.8993
3	0.8185	0.7657	0.7987	1.0000	1.0000	1.0000	0.9092	0.8828	0.8993
4	0.8185	0.7591	0.7921	1.0000	1.0000	1.0000	0.9092	0.8795	0.8960
5	0.8152	0.7558	0.7888	1.0000	1.0000	1.0000	0.9076	0.8779	0.8944
6	0.8152	0.7550	0.7855	1.0000	0.9967	1.0000	0.9076	0.8758	0.8927
7	0.8086	0.7508	0.7822	1.0000	0.9934	1.0000	0.9043	0.8721	0.8911
8	0.8119	0.7542	0.7781	1.0000	0.9802	0.9967	0.9059	0.8672	0.8874
9	0.8119	0.7525	0.7781	1.0000	0.9736	0.9967	0.9059	0.8631	0.8874
10	0.8119	0.7372	0.7718	1.0000	0.9670	0.9835	0.9059	0.8521	0.8777
11	0.8119	0.7578	0.7744	1.0000	0.9538	0.9802	0.9059	0.8558	0.8773
12	0.8119	0.7563	0.7744	1.0000	0.9208	0.9802	0.9059	0.8385	0.8773
13	0.8106	0.7491	0.7770	0.9934	0.8944	0.9769	0.9020	0.8217	0.8770
14	0.8067	0.7568	0.7763	0.9901	0.8548	0.9736	0.8984	0.8058	0.8749
15	0.8121	0.7796	0.7755	0.9835	0.8086	0.9703	0.8978	0.7941	0.8729
16	0.8191	0.8198	0.7755	0.9670	0.7327	0.9703	0.8931	0.7762	0.8729
17	0.8219	0.8257	0.7755	0.9637	0.7195	0.9703	0.8928	0.7726	0.8729
18	0.8185	0.8295	0.7789	0.9637	0.7162	0.9703	0.8911	0.7728	0.8746
19	**0.8219**	**0.8520**	**0.7789**	**0.9637**	**0.6469**	**0.9703**	**0.8928**	**0.7495**	**0.8746**
20	0.8144	0.8383	0.7789	0.9604	0.5512	0.9703	0.8874	0.6947	0.8746
21	0.8179	0.8667	0.7789	0.9604	0.4950	0.9703	0.8891	0.6809	0.8746
22	0.8179	0.8978	0.7789	0.9604	0.4521	0.9703	0.8891	0.6750	0.8746
23	0.8213	0.8963	0.7789	0.9604	0.4455	0.9703	0.8909	0.6709	0.8746
24	0.8229	0.8955	0.7816	0.9505	0.4422	0.9670	0.8867	0.6689	0.8743
25	0.8229	0.8871	0.7842	0.9505	0.4092	0.9637	0.8867	0.6482	0.8740

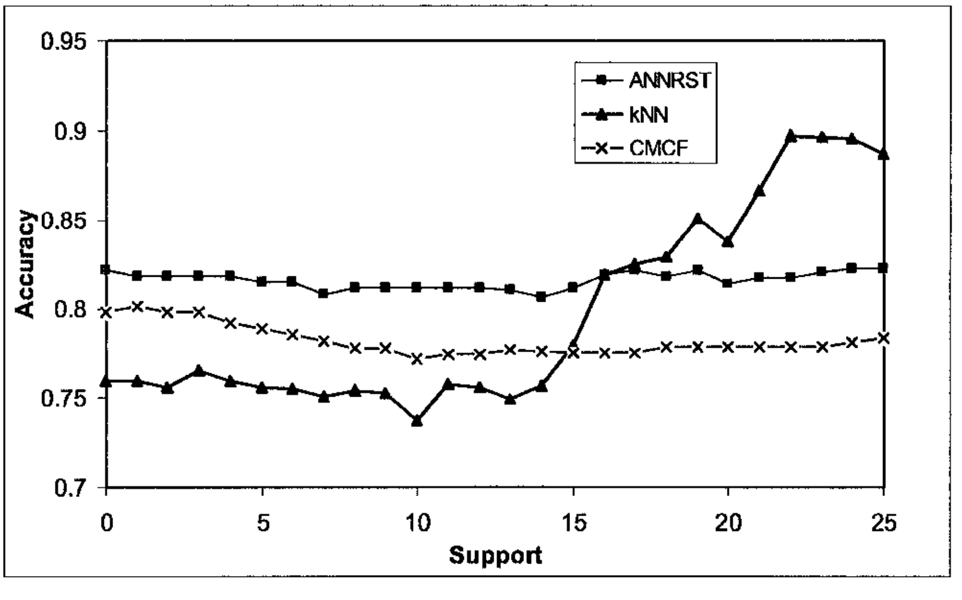


Figure 4.8 Filtering effect on accuracy of RST classifier on three different data sets.

It is shown in Figure 4.8 that ANNRST is better than the others except when support is 16 of which k-NN has better accuracy. If we consider the coverage, it can be seen that the coverage of k-NN is badly decreasing. Michalsky quality measure with $\mu = 0.5$ in Figure 4.10 shows clearly that ANNRST outperforms the others.

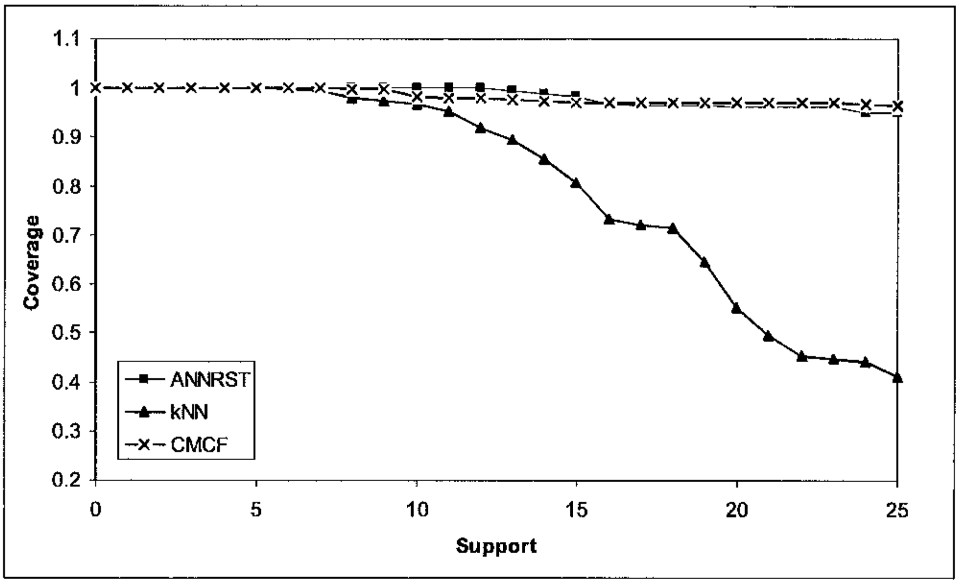


Figure 4.9 Filtering effect on coverage of RST classifier on three different data sets.

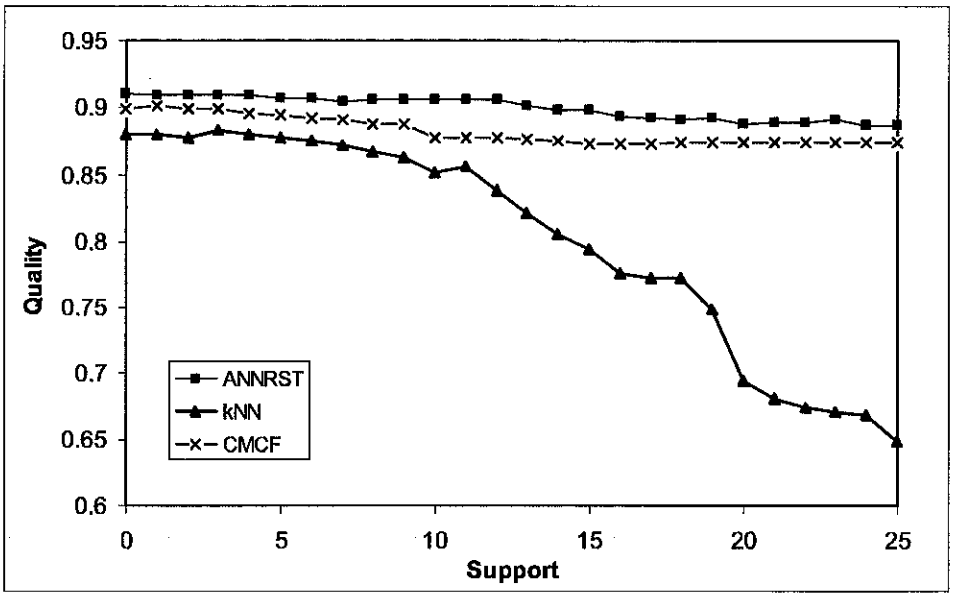


Figure 4.10 Filtering effect on Michalski quality measure (μ=0.5) of RST classifier on three different data sets.

The Michalski quality measure as a trade off between accuracy and coverage shows the performance of ANNRST. The increasing of accuracy of classifier will usually be followed by decreasing of coverage.

The number of rules must be considered when the rule based classifiers are compared. Fewer numbers of rules give better interpretability of the classifier. Hence, it is necessary to see the accuracy, coverage and Michalsky quality measure on the same number of rules while filtering the rules. Thus, the effects of rule filtering with the same number of rules can be seen in Figure 4.11-4.13. ANNRST is still considered better than CMCF and k-NN. For small number of rules, k-NN has the highest accuracy but lowest coverage. Michalsky quality measure shows that k-NN is the worst and ANNRST is the best as shown in Figure 4.13.

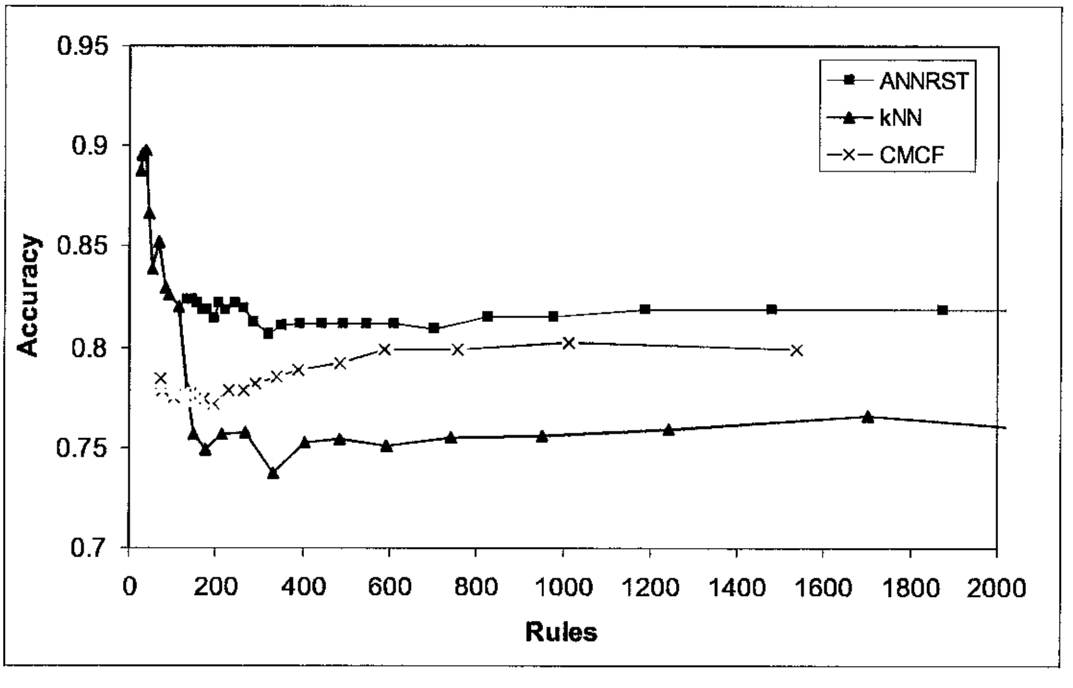


Figure 4.11 Filtering effect on accuracy of RST classifier on three different data sets on the basis of rule number.

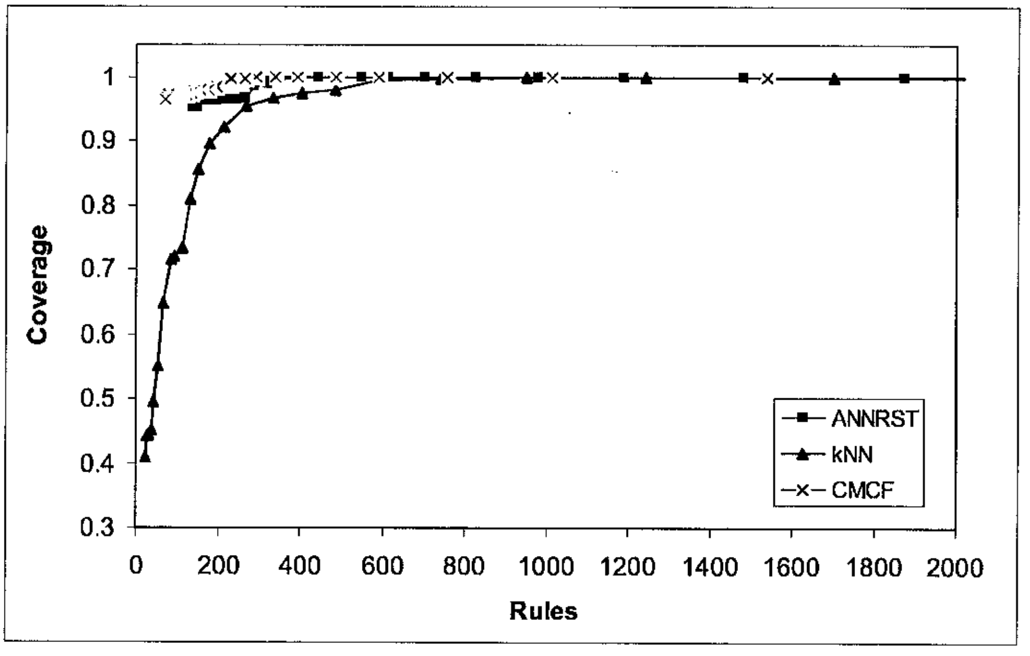


Figure 4.12 Filtering effect on coverage of RST classifier on three different data sets on the basis of rule number.

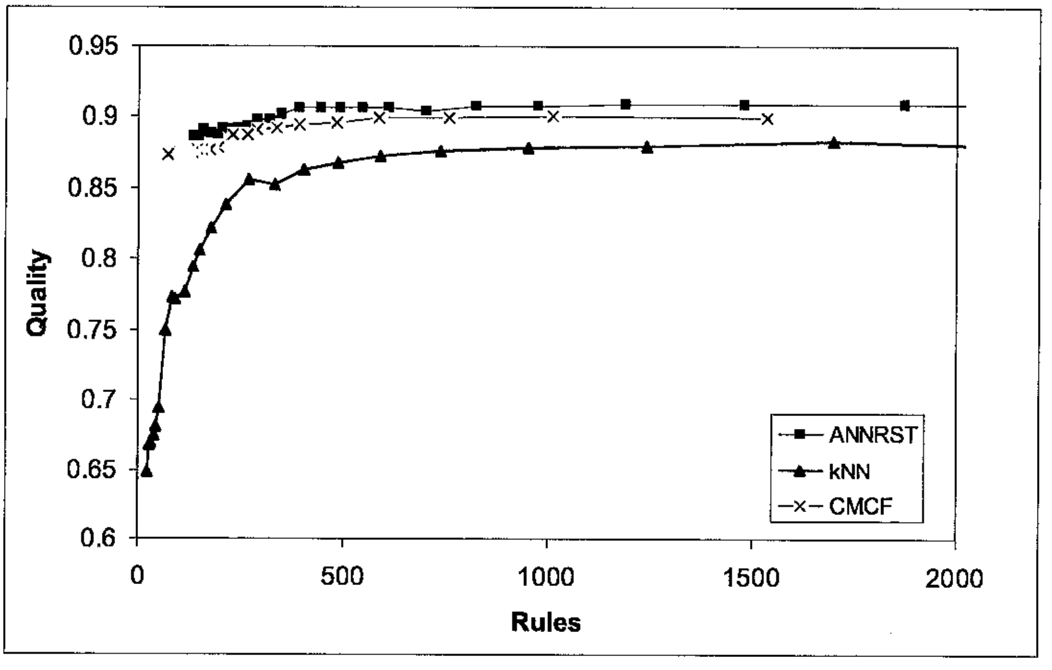


Figure 4.13 Filtering effect on Michalski quality measure (μ=0.5) of RST classifier on three different data sets on the basis of rule number.

Sometimes sensitivity and specificity are more interesting than accuracy and coverage. AUC (Area Under Curve) of ROC (Receiver Operating Characteristics) is also important parameter of binary classifiers. Table 4.23 shows the sensitivity, specificity and AUC of ROC during rule filtering based on support. k-NN can not be considered good performance due to very bad coverage as shown in Table 4.23. ANNRST looks better than CMCF in sensitivity but worse in specificity, but AUC of ROC ANNRST is better than CMCF. The gap between sensitivity and specificity of ANNRST is also smaller than CMCF. Sensitivity means how accurate the classifier recognizes the CAD suffered patients. While Specificity means how accurate the classifier recognizes non-suffered patients. As a trade off between sensitivity and specificity, then AUC of ROC has been introduced. It can be seen that ANNRST is still better than the others in AUC-ROC except after removing rules having support below 20 where k-NN is better. In this case the coverage of k-NN must be considered since k-NN has the worst coverage. ANNRST is better than k-NN when the coverage is considered.

Table 4.23 Filtering effects on sensitivity, specificity and AUC of ROC

Support	Sensitivity			Specificity			AUC of ROC		
	ANNRST	k-NN	CMCF	ANNRST	k-NN	CMCF	ANNRST	k-NN	CMCF
0	0.8345	0.7122	0.7266	0.8110	0.7988	0.8598	0.8926	0.8499	0.8568
1	0.8345	0.7122	0.7338	0.8049	0.7988	0.8598	0.8943	0.8509	0.8562
2	0.8345	0.7194	0.7194	0.8049	0.7866	0.8659	0.8953	0.8487	0.8571
3	0.8345	0.7122	0.7122	0.8049	0.8110	0.8720	0.8925	0.8496	0.8568
4	0.8417	0.6978	0.6978	0.7988	0.8110	0.8720	0.8892	0.8441	0.8561
5	0.8417	0.6835	0.6978	0.7927	0.8171	0.8659	0.8889	0.8409	0.8527
6	0.8417	0.6957	0.6906	0.7927	0.8049	0.8659	0.8874	0.8234	0.8388
7	0.8345	0.7101	0.6835	0.7866	0.7853	0.8659	0.8821	0.8135	0.8333
8	0.8345	0.7090	0.6739	0.7927	0.7914	0.8659	0.8762	0.8043	0.8225
9	0.8345	0.7164	0.6667	0.7927	0.7826	0.8720	0.8697	0.7957	0.8220
10	0.8273	0.7068	0.6493	0.7988	0.7625	0.8720	0.8683	0.7893	0.8124
11	0.8273	0.7615	0.6541	0.7988	0.7547	0.8720	0.8688	0.7961	0.8100
12	0.8273	0.7742	0.6541	0.7988	0.7419	0.8720	0.8637	0.7924	0.8133
13	0.8130	0.7731	0.6541	0.7988	0.7303	0.8773	0.8524	0.7844	0.8185
14	0.8130	0.7895	0.6515	0.7866	0.7310	0.8773	0.8498	0.7885	0.8147
15	0.8201	0.8302	0.6489	0.7805	0.7410	0.8773	0.8453	0.8072	0.8156
16	0.8130	0.8163	0.6489	0.7744	0.8226	0.8773	0.8504	0.8366	0.8075
17	0.8130	0.8105	0.6489	0.7744	0.8374	0.8773	0.8507	0.8304	0.8096
18	0.8130	0.8105	0.6565	0.7683	0.8443	0.8773	0.8467	0.8339	0.8101
19	**0.8130**	**0.8353**	**0.6565**	**0.7744**	**0.8649**	**0.8773**	**0.8464**	**0.8580**	**0.8096**
20	0.8058	0.7969	0.6565	0.7622	0.8641	0.8773	0.8441	0.8431	0.8096
21	0.8058	0.7593	0.6565	0.7683	0.9271	0.8773	0.8441	0.8506	0.8104
22	0.8058	0.7234	0.6565	0.7683	0.9889	0.8773	0.8472	0.8561	0.8131
23	0.8058	0.7111	0.6565	0.7744	0.9889	0.8773	0.8399	0.8500	0.8132
24	0.7986	0.7045	0.6565	0.7683	0.9889	0.8827	0.8398	0.8467	0.8181
25	0.7986	0.7045	0.6565	0.7683	0.9875	0.8882	0.8397	0.8460	0.8157

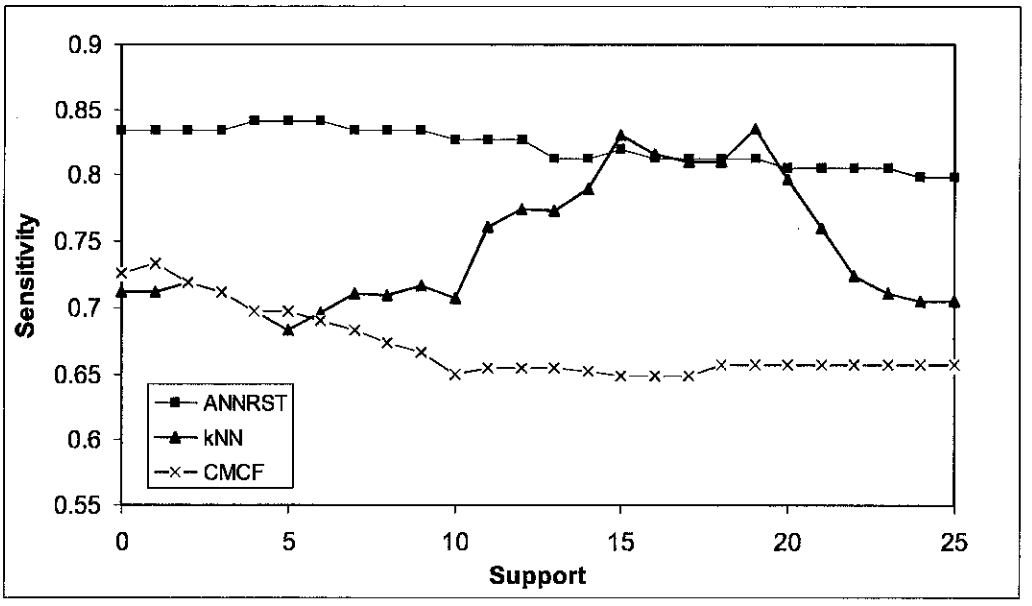


Figure 4.14 Filtering effect on sensitivity of RST classifier on three different data sets.

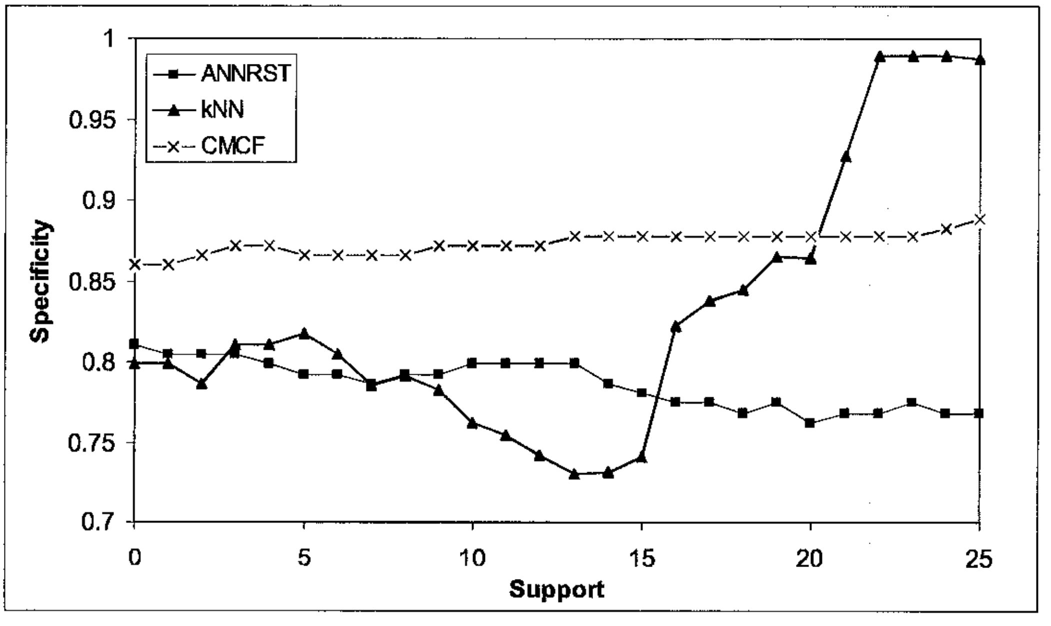


Figure 4.15 Filtering effect on specificity of RST classifier on three different data sets.

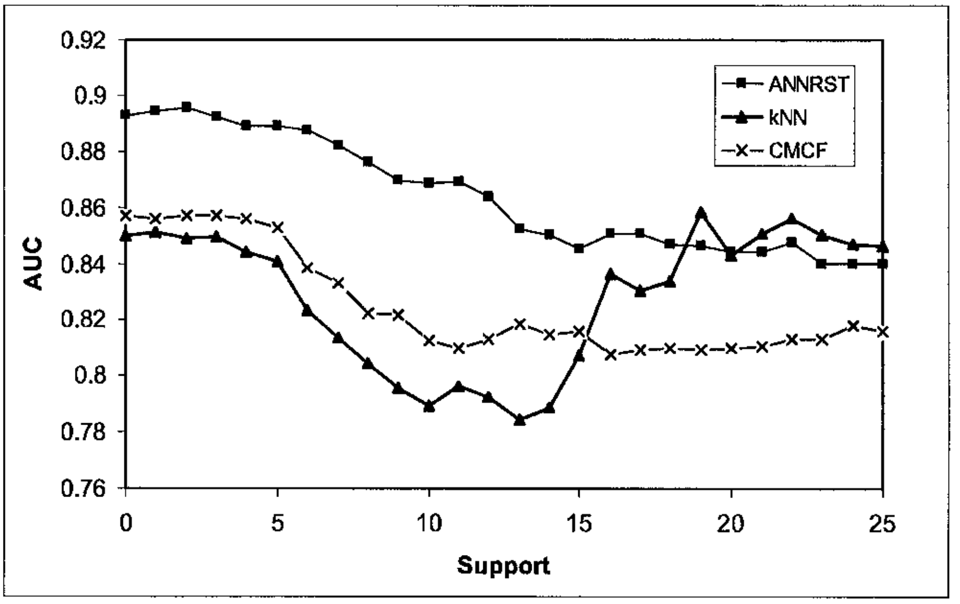


Figure 4.16 Filtering effect on AUC of ROC of RST classifier on three different data sets.

The number of rules among classifiers is not the same. For example, after filtering rules that have support below 20, the number of rules is 206, 92 and 67 for ANNRST, CMCF and k-NN as shown in Table 4.21. Figure 4.17 – 4.19 shows the sensitivity, specificity and AUC of ROC on the basis of number of rules.

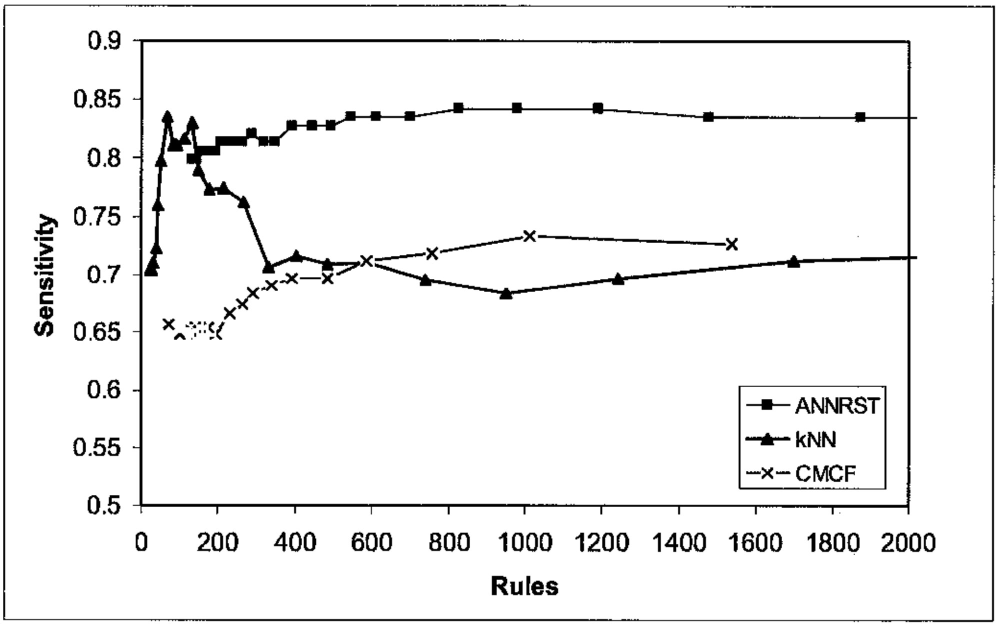


Figure 4.17 Filtering effect on sensitivity of RST classifier on three different data sets on the basis of rule number.

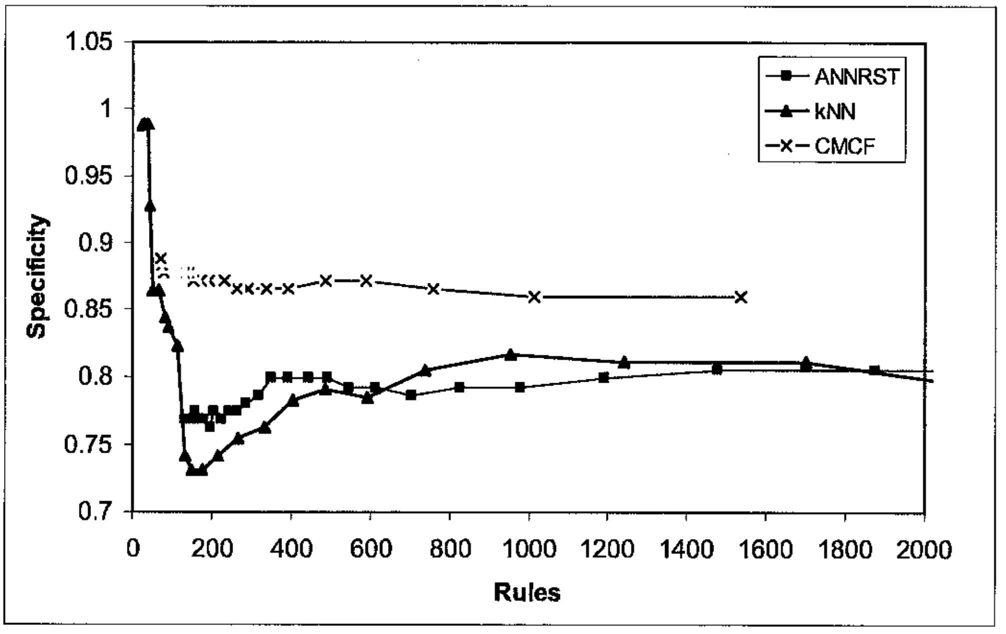


Figure 4.18 Filtering effect on specificity of RST classifier on three different data sets on the basis of rule number.

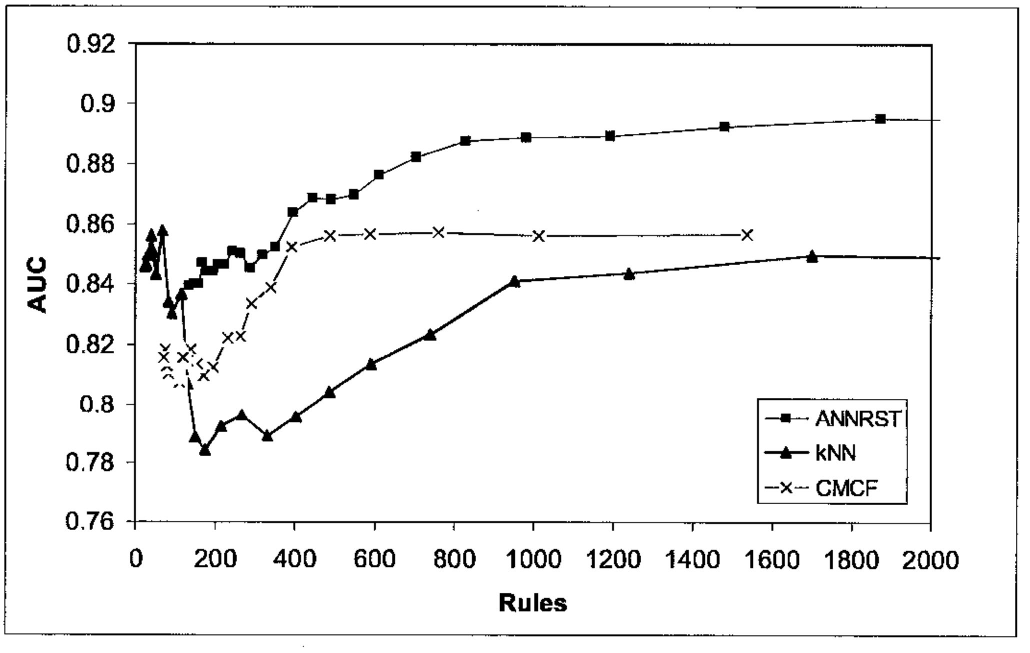


Figure 4.19 Filtering effect on AUC of ROC of RST classifier on three different data sets on the basis of rule number.

ANNRST generally is better than the others except for specificity where CMCF performs better, but ANNRST is better than CMCF on AUC of ROC. The above graphs can be used to select the appropriate number of rules with acceptable accuracy and coverage.

4.3 RULE DISCOVERY AND SELECTION

Based on the results reported in previous sub section, ANNRST imputed Hungarian and Long Beach data sets with 358 objects are chosen to discover the knowledge. Discretization of numerical attributes based on Boolean reasoning using equations (3.14) - (3.18) results in values as in Table 4.24.

Table 4.24 Discretization results

Numerical Attributes	Discrete Value		
age	[*, 41)	[41, 53)	[53, *)
trestbps	[*, 129)		[129, *)
chol	[*, 241)		[241, *)
thalach	[*, 133)		[133, *)
oldpeak	[*, 0.3)		[0.3, *)
ca	[*, 1)		[1, *)

[*, 41) means age < 41, [53,*) means age ≥ 53 and [41,53) means 41≤ age <53. The other attributes can be seen in Table 3.1.

Rule filtering by removing the rules having support below 20 results in 206 rules with their accuracy and coverage are 0.822 and 0.964, respectively, as shown in Table 4.22 with bold font.

The rules are shown in Table 4.25. These 206 rules will be selected using RST rule selection based on testing data. The rules in this table are still in crisp form. Fuzzification will be made after RST rule selection process.

Table 4.25 Extracted rules using RST

No	Decision Rules	Support
1	oldpeak([0.3, *)) AND slope(2) AND thal(7) => num(1)	81
2	cp(4) AND slope(2) AND thal(7) => num(1)	77
3	fbs(0) AND exang(1) AND oldpeak([0.3, *)) AND thal(7) => num(1)	73
4	cp(4) AND fbs(0) AND exang(1) AND thal(7) => num(1)	71
5	fbs(0) AND thalach([133, *)) AND oldpeak([*, 0.3)) AND ca ([*, 1)) AND thal(3) => num(0)	70
6	sex(1) AND cp(4) AND fbs(0) AND oldpeak([0.3, *)) AND thal(7) => num(1)	69
7	sex(1) AND exang(1) AND slope(2) AND thal(7) => num(1)	65
8	fbs(0) AND exang(1) AND slope(2) AND thal(7) => num(1)	64
.	.	.
.	.	.
.	.	.
201	trestbps([129, *)) AND chol([241, *)) AND oldpeak([*, 0.3)) AND slope(1) AND ca ([*, 1)) => num(0)	20
202	cp(4) AND trestbps([129, *)) AND chol([241, *)) AND restecg(0) AND thalach([*, 133)) AND slope(2) => num(1)	20
203	age([53, *)) AND trestbps([129, *)) AND restecg(0) AND exang(0) AND ca ([*, 1)) => num(0)	20
204	age([53, *)) AND restecg(1) AND slope(2) AND thal(7) => num(1)	20
205	age([41, 53)) AND trestbps([*, 129)) AND chol([*, 241)) AND thalach([133, *)) AND ca ([*, 1)) => num(0)	20
206	cp(4) AND trestbps([129, *)) AND chol([241, *)) AND fbs(0) AND restecg(0) AND thalach([*, 133)) => num(1)	20

This set of rules from support based filtering is chosen and will be selected using RST rule importance measure as explained in section 3.4.2. The results of new decision table converted from extracted rules are shown in Table 4.26.

The new constructed decision table consists of 303 objects. There are 206 rules. Hence the decision table has 206 conditional attributes with single decision, which is attribute *num*. RST reduct computation using Johnson's algorithm results in reduct with 27 conditions, which represent 27 rules (Section 3.4.2).

Table 4.26 New decision table from extracted rules using RST

No	Rule1	Rule2	Rule3			Rule204	Rule205	Rule206	Decision
1	0	0	0	.	.	0	0	0	0
2	0	0	0	.	.	0	0	0	1
3	1	1	1	.	.	0	0	0	1
4	0	0	0	.	.	0	0	0	0
5	0	0	0	.	.	0	0	0	0
6	0	0	0	.	.	0	0	0	0
7	0	0	0	.	.	0	0	0	1
8	0	0	0	.	.	0	0	0	0
9	1	1	0	.	.	0	0	0	1
10	0	0	0	.	.	0	0	0	1
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
294	0	0	1	.	.	0	0	0	1
295	0	0	0	.	.	0	0	0	1
296	0	0	0	.	.	0	1	0	0
297	0	0	0	.	.	0	0	0	1
298	0	1	0	.	.	0	0	1	1
299	1	0	0	.	.	0	0	0	1
300	1	1	0	.	.	0	0	0	1
301	1	1	1	.	.	0	0	0	1
302	0	0	0	.	.	0	0	0	1
303	0	0	0	.	.	0	0	0	0

The reduct attributes of Table 4.26 are {*Rule1, Rule10, Rule17, Rule19, Rule20, Rule23, Rule26, Rule30, Rule33, Rule49, Rule51, Rule66, Rule70, Rule73, Rule77, Rule81, Rule101, Rule114, Rule117, Rule118, Rule150, Rule154, Rule158, Rule160, Rule161, Rule180, Rule203*}.

These selected 27 rules have accuracy and coverage of 0.852 and 0.937 respectively on testing data set. Table 4.27 shows all the 27 selected rules based on RST reduct.

Table 4.27 Selected rules using RST

No.	Rule	Rule No.
1	oldpeak([0.3, *)) AND slope(2) AND thal(7) => num(1)	Rule1
2	fbs(0) AND thalach([133, *)) AND slope(1) AND ca ([*, 1)) AND thal(3) => num(0)	Rule10
3	fbs(0) AND ca ([1, *)) AND thal(7) => num(1)	Rule17
4	sex(1) AND fbs(0) AND thalach([133, *)) AND exang(0) AND ca ([*, 1)) AND thal(3) => num(0)	Rule19
5	sex(1) AND fbs(0) AND restecg(0) AND oldpeak([0.3, *)) AND thal(7) => num(1)	Rule20
6	chol([*, 241)) AND exang(0) AND thal(3) => num(0)	Rule23
7	trestbps([129, *)) AND restecg(0) AND ca ([*, 1)) AND thal(3) => num(0)	Rule26
8	cp(4) AND exang(1) AND slope(2) AND ca ([1, *)) => num(1)	Rule30
9	cp(2) AND chol([*, 241)) AND ca ([*, 1)) => num(0)	Rule33
10	sex(0) AND fbs(0) AND exang(0) AND oldpeak([*, 0.3)) => num(0)	Rule49
11	chol([241, *)) AND fbs(0) AND slope(2) AND thal(7) => num(1)	Rule51
12	trestbps([129, *)) AND slope(1) AND ca ([*, 1)) AND thal(3) => num(0)	Rule66
13	chol([241, *)) AND slope(2) AND ca ([1, *)) => num(1)	Rule70
14	exang(1) AND oldpeak([0.3, *)) AND ca ([*, 1)) AND thal(7) => num(1)	Rule73
15	cp(4) AND chol([241, *)) AND fbs(0) AND ca ([1, *)) => num(1)	Rule77
16	age([41, 53)) AND chol([*, 241)) AND thal(3) => num(0)	Rule81
17	sex(1) AND cp(4) AND trestbps([129, *)) AND chol([241, *)) AND thalach([*, 133)) AND slope(2) => num(1)	Rule101
18	age([41, 53)) AND sex(1) AND thalach([133, *)) AND exang(0) AND ca ([*, 1)) => num(0)	Rule114
19	thalach([133, *)) AND exang(1) AND thal(7) => num(1)	Rule117
20	trestbps([129, *)) AND chol([241, *)) AND fbs(0) AND exang(0) AND ca ([*, 1)) => num(0)	Rule118
21	age([53, *)) AND sex(1) AND trestbps([129, *)) AND fbs(1) AND slope(2) => num(1)	Rule150
22	sex(1) AND cp(4) AND thalach([133, *)) AND thal(7) => num(1)	Rule154
23	age([53, *)) AND cp(2) AND slope(1) => num(0)	Rule158
24	age([41, 53)) AND sex(0) AND thal(3) => num(0)	Rule160
25	trestbps([*, 129)) AND fbs(0) AND ca ([1, *)) => num(1)	Rule161
26	chol([*, 241)) AND restecg(0) AND oldpeak([0.3, *)) AND ca ([1, *)) => num(1)	Rule180
27	age([53, *)) AND trestbps([129, *)) AND restecg(0) AND exang(0) AND ca ([*, 1)) => num(0)	Rule203

The comparison between rule selection methods applied on training and testing data sets can be seen on Table 4.28 and Table 4.29. Two support based filtering are implemented. This selection method is based on support of rules on training data set and testing data set.

The formula of selection methods, which are Michalski, Torgo, Brazdil, Pearson and Cohen, is explained in Table 3.14.

Table 4.28 Comparison of rule selection method performance applied on training data set

Selection Methods	Accuracy	Coverage	Number of rules
Proposed Method	1	0.816	27
Support Based (Training Data)	1	0.665	29
Support Based (Testing Data)	1	0.749	27
Michalski μ=0.5	1	0.682	27
Torgo	1	0.682	27
Brazdil	1	0.682	27
Pearson	1	0.682	27
Cohen	1	0.598	29

Table 4.29 Comparison of rule selection method performance applied on testing data set

Selection Methods	Accuracy	Coverage	Number of rules
Proposed Method	0.852	0.937	27
Support Based (Training Data)	0.847	0.799	29
Support Based (Testing Data)	0.844	0.868	27
Michalski μ=0.5	0.845	0.785	27
Torgo	0.845	0.785	27
Brazdil	0.845	0.785	27
Pearson	0.845	0.785	27
Cohen	0.863	0.65	29

Table 4.28 and Table 4.29 show that the proposed method in general performs better than the others on both testing and training data sets. The proposed method is also compared with other four classifiers such as decision tree, k-NN, LTF-C and unfiltered RST rules on testing data. The results are shown in Table 4.30. LTF-C and k-NN do not produce rules.

Table 4.30 Comparison of proposed method and other four classifiers applied on testing data set

Methods	Accuracy	Coverage	Number of rules
Proposed Method	0.852	0.937	27
Unfiltered RST rules	0.815	1	4095
Decision Tree	0.856	0.482	83
LTF-C	0.812	1	-
k-NN	0.815	1	-

4.4 FUZZY MODELLING

4.4.1 Fuzzification

The first step to build fuzzy decision support system is rule fuzzification. The results of RST rule extraction is in crisp mode. In order to apply these rules into fuzzy inference engine, fuzzification must be carried on using the step explained in section 3.8.2 and 3.8.4 and Table 4.26. Triangular and trapezoidal fuzzifiers that make triangular and trapezoidal membership function are the simplest and the most common possible models of value of membership as they are fully defined by only three and four parameter respectively. The linear change in the membership value is the simplest possible model people can think of.

There are six discretized numerical attributes. Based on discretization of Table 4.26, the fuzzy membership functions are listed below.

Membership function of *age*

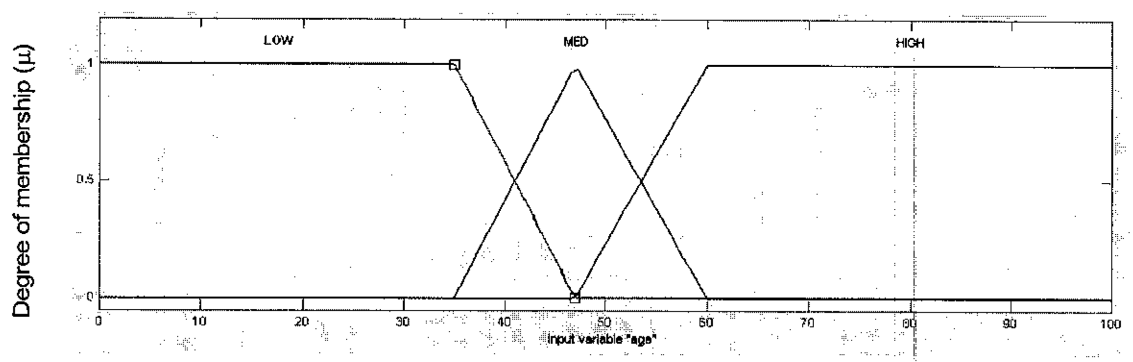


Figure 4.20 Membership function of attribute *age*

Membership function of *trestbps*

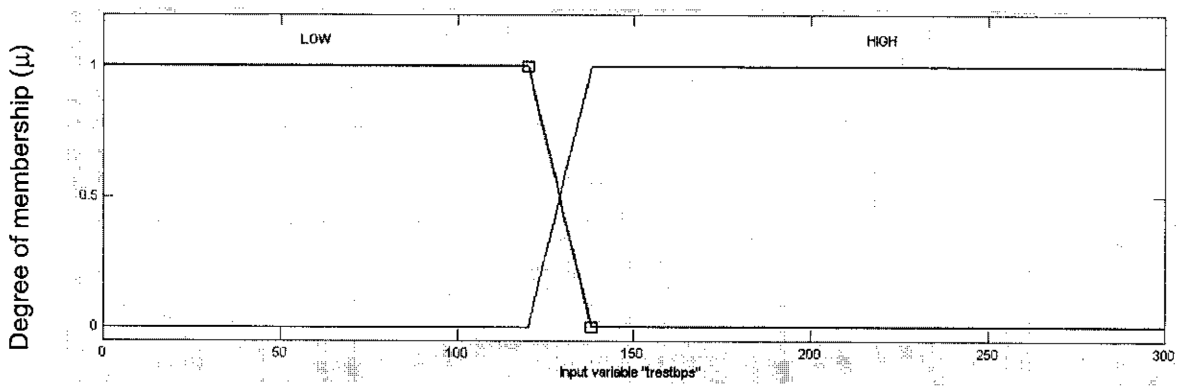


Figure 4.21 Membership function of attribute *trestbps*

Membership function of *chol*

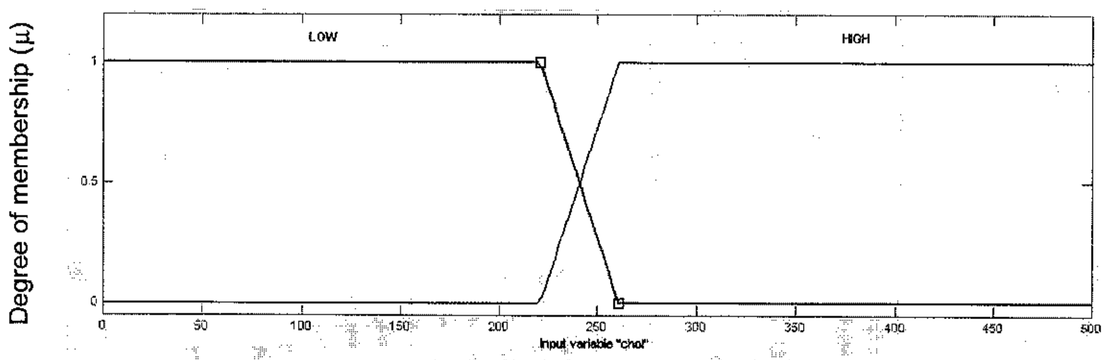


Figure 4.22 Membership function of attribute *chol*

Membership function of *thalach*

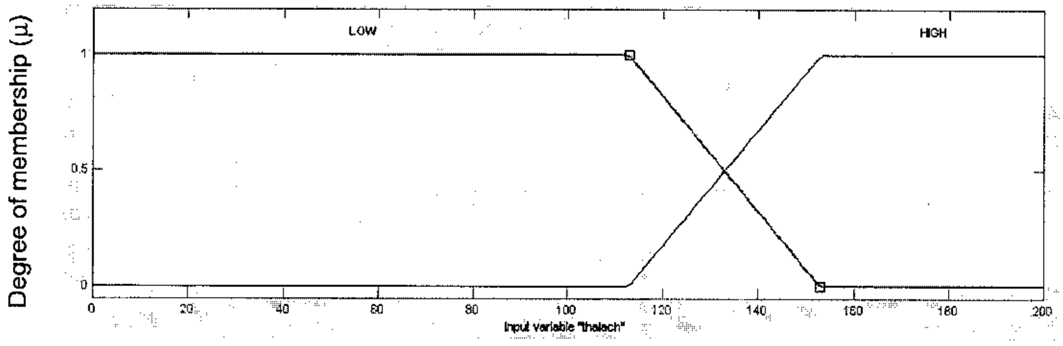


Figure 4.23 Membership function of attribute *thalach*

Membership function of *oldpeak*

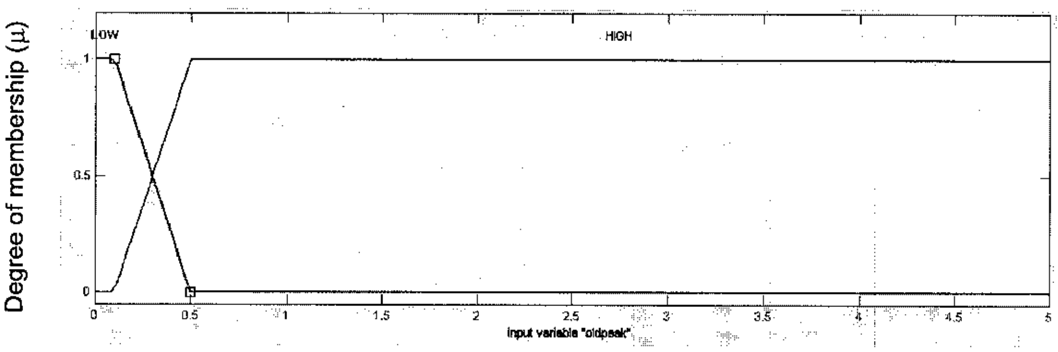


Figure 4.24 Membership function of attribute *oldpeak*

Membership function of *ca*

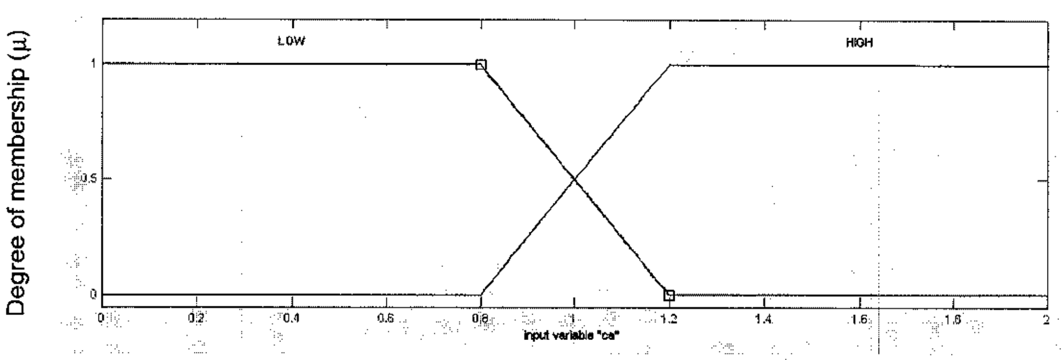


Figure 4.25 Membership function of attribute *ca*

The output which is *num* is CAD possibility. Referring to Table 3.1, in a crisp manner, if more than 50% narrowing, the CAD is yes otherwise CAD is no. This CAD output can be fuzzified to be fuzzy output with the membership function as shown in Figure 4.26.

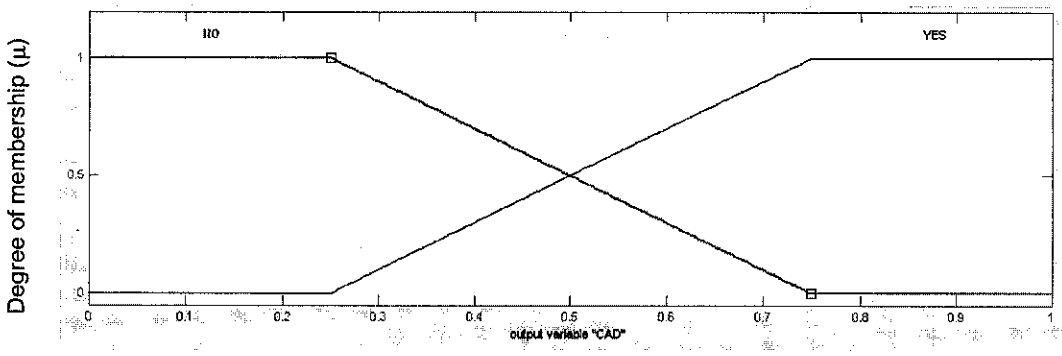


Figure 4.26 Membership function of attribute *num* that represents the CAD possibility

The remaining attributes which are non-numerical can be converted to crisp membership. The value of their membership functions can only have 1 or 0. Table 4.27 can be fuzzified similar to Table 3.16 which becomes Table 3.17 after fuzzification.

4.4.2 Fuzzy inference engine

The inference engine uses *min* operator for AND and Mamdani implication which is min operator with weighing factor for each rule. The step of inferring is explained in section 3.7.4.

The rules can then be included in Mamdani fuzzy inference engine with the same way as in Figure 3.17. The weights of each rule can be calculated using equation (3.42) results in Table 4.31. The centroid method, as shown in equation (3.40), is used to defuzzify the output into numerical values. The numerical values represent the possibility of CAD or the approximate percentage of blocking of one of the coronary arteries.

Table 4.31 Weight of rules

Rule No	Weight
1	1
2	0.765432099
3	0.679012346
4	0.666666667
5	0.654320988
6	0.62962963
7	0.604938272
8	0.580246914
9	0.555555556
10	0.518518519
11	0.518518519
12	0.469135802
13	0.456790123
14	0.456790123
15	0.444444444
16	0.432098765
17	0.37037037
18	0.345679012
19	0.345679012
20	0.333333333
21	0.296296296
22	0.296296296
23	0.296296296
24	0.283950617
25	0.283950617
26	0.259259259
27	0.24691358

4.5 EVALUATION OF FUZZY DECISION SUPPORT SYSTEM

4.5.1 Evaluation using data mining methods

Evaluation of FDSS with imputation is done by comparing to other data mining and knowledge discovery models using similar training and testing data sets. The training data sets are imputed data set. Four models of classifiers which are MLP-ANN, k-NN, C4.5 and RIPPER are used to classify all the UCI data sets including Hungarian and Long Beach before their missing values are imputed and Ipoh Specialist Hospital data set (22 patients) called Ipoh data set. The results of diagnosis of CAD for different classifiers are shown in Table 4.32 – Table 4.34.

Table 4.32 Accuracy of different diagnosis methods on UCI-CAD data sets and Ipoh data set

Diagnosis Method	UCI Data Sets and Ipoh Data Set					
	Cleveland	Hungarian	Long Beach	Switzerland	Ipoh	Average
FDSS	0.83	0.84	0.75	0.70	0.82	0.79
MLP-ANN	0.81	0.54	0.76	0.81	0.77	0.74
k-NN	0.81	0.92	0.85	0.62	0.73	0.78
C4.5	0.82	0.66	0.77	0.53	0.55	0.66
RIPPER	0.83	0.68	0.78	0.41	0.68	0.67

Table 4.33 shows that FDSS has the best accuracy in Ipoh data set. FDSS and RIPPER have the best accuracy in Cleveland data set. MLP-ANN has the best accuracy in Switzerland data set. k-NN has the best accuracy in Hungarian and Long Beach data sets. The average shows that FDSS has better accuracy than the others. k-NN gets its high accuracy on Hungarian and Long Beach in which the training data are mostly taken from these data sets. k-NN is distance based classifier which will have very high accuracy if it is applied on the training data set. FDSS has better accuracy than k-NN in Cleveland and Switzerland data sets. The knowledge of FDSS is transparent where k-NN knowledge is less transparent. The result of FDSS is also numerical values which represent the percentage of coronary artery blocking where k-NN is categorical having only yes or no value.

Table 4.33 Sensitivity of different diagnosis methods on UCI-CAD data sets and Ipoh data set

Diagnosis Method	UCI Data Sets and Ipoh Data Set					
	Cleveland	Hungarian	Long Beach	Switzerland	Ipoh	Average
FDSS	0.81	0.70	0.83	0.71	1.00	0.81
MLP-ANN	0.77	0.44	0.75	0.93	0.92	0.76
k-NN	0.84	0.87	0.85	0.96	0.69	0.84
C4.5	0.79	0.52	0.78	0.97	0.85	0.78
RIPPER	0.82	0.54	0.81	0.96	0.69	0.76

Table 4.34 shows that k-NN has better average of sensitivity than FDSS. FDSS has the best sensitivity on Ipoh data set but has the worse sensitivity on Switzerland data set. If specificity is considered as in Table 4.35, then FDSS has the best of specificity on Switzerland data set.

Table 4.34 Specificity of different diagnosis methods on UCI-CAD data sets and Ipoh data set

Diagnosis Method	UCI Data Sets and Ipoh Data Set					
	Cleveland	Hungarian	Long Beach	Switzerland	Ipoh	Average
FDSS	0.85	0.91	0.49	0.50	0.56	0.66
MLP-ANN	0.85	1.00	1.00	0.06	0.56	0.69
k-NN	0.79	0.95	0.81	0.10	0.78	0.69
C4.5	0.85	0.89	0.60	0.10	0.11	0.51
RIPPER	0.84	0.87	0.59	0.08	0.67	0.61

The average specificity of FDSS is below MLP-ANN and k-NN classifiers. From the results, it can be considered that FDSS is comparable with the other classifiers. However, FDSS has advantage of transparent knowledge, handling continuous values directly and numerical output that represent the approximate blockage of coronary arteries. FDSS has good diagnosis accuracies for all data sets which are between 0.7 and 0.84 and the average is 0.79. The sensitivities of FDSS are between 0.7 and 1 and the average is 0.81. The specificities of FDSS are between 0.49 and 0.91 and the average is 0.66. FDSS can be considered robust and has the best diagnosis performance compared to the others.

4.5.2 Evaluation by cardiologists

The performance of FDSS is evaluated and compared using 30 samples taken from each four data sets from UCI. The samples are diagnosed by FDSS and three local cardiologists. The actual results are taken from coronary angiogram. Table 4.36 and Figure 4.27 show the results of FDSS compared to two cardiologists using Cleveland data set. The 30 samples of Cleveland data set with the results of FDSS and three cardiologists are shown in Table 4.35.

Figure 4.27 shows that the performance of FDSS is better than the decisions by cardiologists especially on accuracy and specificity. The first cardiologist has indicated the same sensitivity as FDSS. FDSS has high specificity compared to the three cardiologists.

Table 4.35 Cleveland data set with the diagnosis of FDSS, cardiologists and angiography

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Cardio#1	Cardio#2	Cardio#3	FDSS	Angio
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0.95	1	1	0.50	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	<0.5	1	0	0.84	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	0.95	1	1	0.88	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	<0.5	0	0	0.15	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	<0.5	0	0	0.11	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	<0.5	0	0	0.13	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	0.6	0	1	0.86	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	<0.5	0	1	0.11	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	0.9	1	1	0.82	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	0.95	1	1	0.84	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0.8	1	0	0.21	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0.6	0	0	0.19	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	0.6	1	1	0.80	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0.6	1	1	0.18	0
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0.6	1	0	0.20	0
56	1	2	130	221	0	2	163	0	0	1	0	7	0.55	1	1	0.15	0
56	1	2	120	240	0	0	169	0	0	3	0	3	<0.5	1	0	0.15	0
67	1	3	152	212	0	2	150	0	0.8	2	0	7	0.8	1	1	0.96	1
55	0	2	132	342	0	0	166	0	1.2	1	0	3	0.55	0	0	0.15	0
44	1	4	120	169	0	0	144	1	2.8	3	0	6	0.8	1	1	0.50	1
63	1	4	140	187	0	2	144	1	4	1	2	7	0.95	1	1	0.87	1
63	0	4	124	197	0	0	136	1	0	2	0	3	<0.5	0	1	0.21	1
41	1	2	120	157	0	0	182	0	0	1	0	3	<0.5	0	0	0.10	0
59	1	4	164	176	1	2	90	0	1	2	2	6	0.8	1	0	0.82	1
57	0	4	140	241	0	0	123	1	0.2	2	0	7	0.8	1	0	0.81	1
45	1	1	110	264	0	0	132	0	1.2	2	0	7	0.8	1	1	0.86	1
68	1	4	144	193	1	0	141	0	3.4	2	2	7	0.95	1	1	0.89	1
57	1	4	130	131	0	0	115	1	1.2	2	1	7	0.9	1	1	0.88	1
57	0	2	130	236	0	2	174	0	0	2	1	3	<0.5	0	0	0.27	1
59	1	1	134	204	0	0	162	0	0.8	1	2	3	0.55	0	1	0.37	1

Table 4.36 Diagnosis performances of FDSS on Cleveland data set

Metrics	Diagnosis Results			
	FDSS	Cardiologist#1	Cardiologist#2	Cardiologist#3
Accuracy	0.87	0.67	0.67	0.73
Sensitivity	0.82	0.82	0.76	0.76
Specificity	0.92	0.46	0.54	0.69

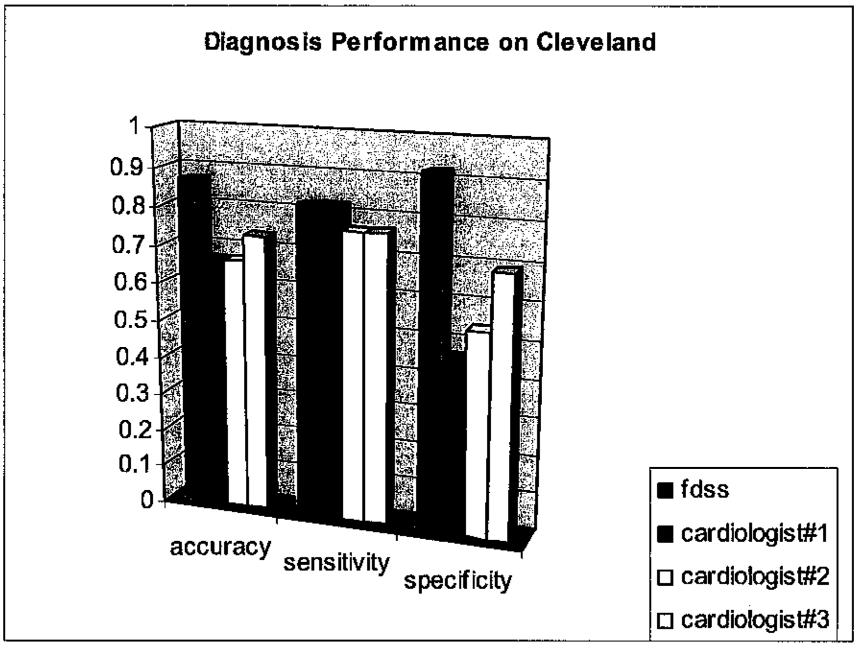


Figure 4.27 Diagnosis performances of FDSS on Cleveland data set

Table 4.37 show the 30 samples of Hungarian data set with the results of FDSS and two cardiologists. The actual results are taken from coronary angiogram. This data set consists of missing attribute values on the attributes of *slope*, *ca* and *thal*. Table 4.38 and Figure 4.28 show the results of FDSS compared to three cardiologists using Hungarian data set.

Table 4.37 Hungarian data set with the diagnosis of FDSS, cardiologists and angiography

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Cardio#1	Cardio#2	Cardio#3	FDSS	Angio
28	1	2	130	132	0	2	185	0	0	?	?	?	<0.5	0	1	0.10	0
29	1	2	120	243	0	0	160	0	0	?	?	?	<0.5	0	1	0.11	0
29	1	2	140	156	0	0	170	0	0	?	?	?	<0.5	0	0	0.10	0
30	0	1	170	237	0	1	170	0	0	?	?	6	0.85	1	0	0.14	0
31	0	2	100	219	0	1	150	0	0	?	?	?	<0.5	0	0	0.11	0
35	1	2	120	308	0	2	180	0	0	?	?	?	<0.5	0	0	0.08	0
35	1	2	150	264	0	0	168	0	0	?	?	?	<0.5	0	0	0.11	0
36	1	2	120	166	0	0	180	0	0	?	?	?	<0.5	0	0	0.10	0
36	1	3	112	340	0	0	184	0	1	2	?	3	<0.5	0	0	0.11	0
36	1	3	130	209	0	0	178	0	0	?	?	?	<0.5	0	1	0.11	0
36	1	3	150	160	0	0	172	0	0	?	?	?	<0.5	0	0	0.10	0
60	1	3	120	246	0	2	135	0	0	?	?	?	<0.5	0	0	0.80	0
61	0	4	130	294	0	1	120	1	1	2	?	?	0.75	1	1	0.87	0
61	1	4	125	292	0	1	115	1	0	?	?	?	0.65	1	1	0.81	0
62	0	1	160	193	0	0	116	0	0	?	?	?	0.6	1	1	0.13	0
62	1	2	140	271	0	0	152	0	1	1	?	?	<0.5	0	0	0.19	0
31	1	4	120	270	0	0	153	1	1.5	2	?	?	0.8	0	0	0.50	1
33	0	4	100	246	0	0	150	1	1	2	?	?	0.8	0	0	0.50	1
34	1	1	140	156	0	0	180	0	0	?	?	?	0.65	0	0	0.10	1
35	1	2	110	257	0	0	140	0	0	?	?	?	<0.5	0	0	0.14	1
36	1	2	120	267	0	0	160	0	3	2	?	?	<0.5	0	0	0.76	1
49	0	3	160	180	0	0	156	0	1	2	?	?	<0.5	0	0	0.12	1
49	1	3	115	265	0	0	175	0	0	?	?	?	<0.5	0	0	0.11	1
49	1	4	130	206	0	0	170	0	0	?	?	?	<0.5	0	0	0.18	1
50	0	3	140	288	0	0	140	1	0	?	?	7	0.8	1	1	0.81	1
50	1	4	145	264	0	0	150	0	0	?	?	?	<0.5	0	0	0.18	1
55	1	1	140	295	0	?	136	0	0	?	?	?	0.6	1	1	0.15	1
55	1	2	160	292	1	0	143	1	2	2	?	?	0.65	1	1	0.81	1
55	1	4	145	248	0	0	96	1	2	2	?	?	0.65	1	1	0.87	1
56	0	2	120	279	0	0	150	0	1	2	?	?	<0.5	0	0	0.22	1

Table 4.38 Diagnosis performances of FDSS on Hungarian data set

Metrics	Diagnosis Results			
	FDSS	Cardiologist#1	Cardiologist#2	Cardiologist#3
Accuracy	0.63	0.63	0.53	0.47
Sensitivity	0.43	0.50	0.29	0.29
Specificity	0.81	0.75	0.75	0.63

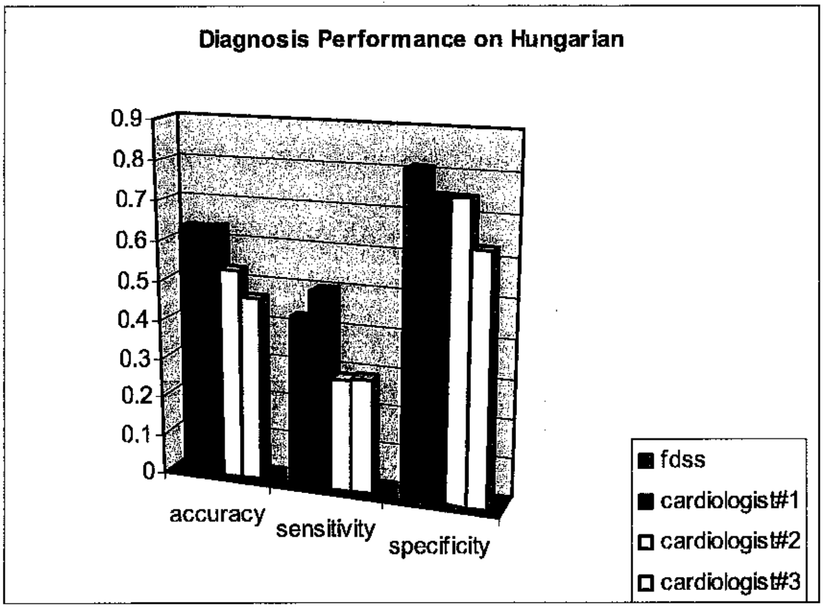


Figure 4.28 Diagnosis performance of FDSS on Hungarian data set

Table 4.38 and Figure 4.28 show that FDSS has better specificity. FDSS has the same accuracy as the first cardiologist and better than the others. The sensitivity of FDSS is less than the first cardiologist but better than the others.

Table 4.39 show the 30 samples of Long Beach data set with the results of FDSS and two cardiologists. The actual results are taken from coronary angiogram. This data set consists of missing attribute values on the attributes of *tresbps, chol, fbs, thalach, exang, slope, ca* and *thal.* Table 4.40 and Figure 4.29 show the results of FDSS compared to three cardiologists using Long Beach data set.

Table 4.39 Long Beach data set with the diagnosis of FDSS, cardiologists and angiography

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Cardio#1	Cardio#2	Cardio#3	FDSS	Angio
60	1	2	160	267	1	1	157	0	0.5	2	?	?	<0.5	1	0	0.83	1
56	1	2	126	166	0	1	140	0	0	?	?	?	<0.5	1	0	0.14	0
59	1	4	140	326	0	1	117	1	1	2	?	?	0.65	0	1	0.82	1
62	1	4	110	?	0	0	120	1	0.5	2	?	3	0.55	1	1	0.81	1
63	1	3	?	?	0	2	?	?	?	?	?	?	<0.5	0	1	0.61	1
57	1	4	128	?	1	1	148	1	1	2	?	?	0.65	1	0	0.81	1
62	1	4	120	220	0	1	86	0	0	?	?	?	<0.5	0	1	0.79	0
63	1	4	170	177	0	0	84	1	2.5	3	?	?	0.85	1	1	0.89	1
46	1	4	110	236	0	0	125	1	2	2	?	?	0.7	1	0	0.86	1
63	1	4	126	?	0	1	120	0	1.5	3	?	?	0.7	0	1	0.79	0
60	1	4	152	?	0	1	118	1	0	?	?	7	0.85	1	0	0.81	0
58	1	4	116	?	0	0	124	0	1	1	?	?	<0.5	0	0	0.82	1
64	1	4	120	?	1	1	106	0	2	2	?	?	<0.5	0	1	0.93	1
52	1	3	128	?	0	1	180	0	3	1	?	?	<0.5	1	1	0.79	1
59	1	4	154	?	0	1	131	1	1.5	?	0	?	0.55	1	0	0.78	0
61	1	3	120	?	0	0	80	1	0	2	?	?	0.55	1	1	0.80	1
40	1	4	125	?	1	0	165	0	0	?	?	7	0.8	1	1	0.62	1
61	1	4	?	?	1	1	86	0	1.5	2	?	7	0.85	1	0	0.92	1
41	1	4	104	?	0	1	111	0	0	?	?	?	<0.5	0	0	0.50	0
57	1	4	?	277	1	1	?	?	?	?	?	?	<0.5	0	0	0.81	1
63	1	4	136	?	0	0	84	1	0	?	?	7	0.85	1	1	0.83	1
59	1	4	122	233	0	0	117	1	1.3	3	?	?	0.85	0	1	0.84	1
51	1	4	128	?	0	0	107	0	0	?	?	?	<0.5	0	1	0.23	0
59	1	3	?	?	0	0	128	1	2	3	?	?	0.85	0	0	0.79	1
42	1	3	134	240	?	0	160	0	0	?	?	?	<0.5	1	1	0.16	0
55	1	3	120	?	0	1	125	1	2.5	2	?	7	0.9	1	1	0.89	1
63	0	2	?	?	0	0	?	?	?	?	?	?	<0.5	0	1	0.12	0
62	1	4	152	153	0	1	97	1	1.6	1	?	7	0.9	1	0	0.83	1
56	1	2	124	224	1	0	161	0	2	2	?	?	0.6	1	1	0.18	0
53	1	4	126	?	0	0	106	0	0	?	?	?	<0.5	0	1	0.44	1

Table 4.40 Diagnosis performances of FDSS on Long Beach data set

Metrics	Diagnosis Results			
	FDSS	Cardiologist#1	Cardiologist#2	Cardiologist#3
Accuracy	0.80	0.63	0.57	0.53
Sensitivity	0.95	0.65	0.60	0.60
Specificity	0.50	0.60	0.50	0.40

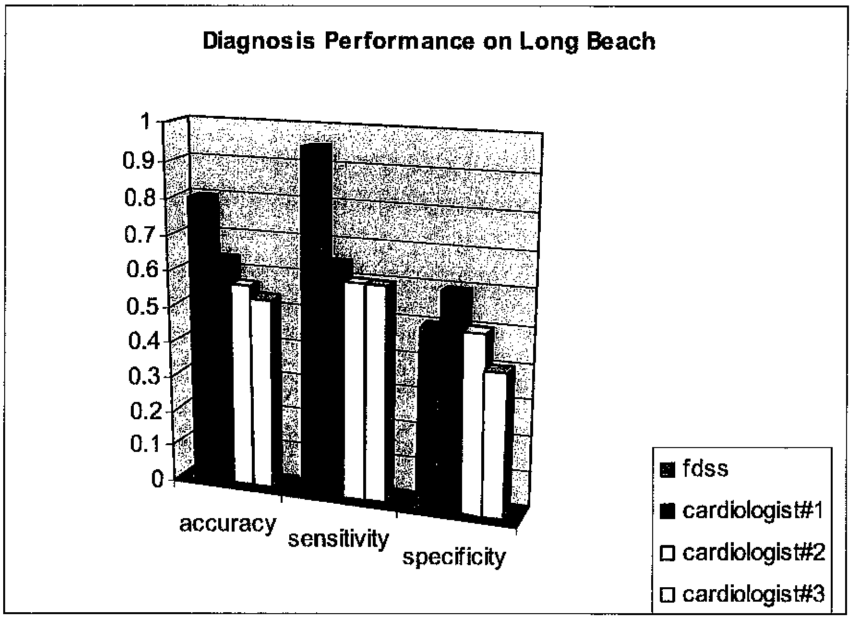


Figure 4.29 Diagnosis performances of FDSS on Long Beach data set

Table 4.40 and Figure 4.29 show that FDSS has better accuracy and sensitivity but worse specificity compared to the decision results of the first cardiologist. The case is similar to Table 4.34 where FDSS also has worse specificity.

Table 4.41 shows the 30 samples of Switzerland data set with the results of FDSS and three cardiologists. The actual results are taken from coronary angiogram. This data set consists of missing attribute values on the attributes of *chol, fbs, slope, ca* and *thal.* All *chol* attribute values are missing. This data set only has one patient that has negative result of CAD based on coronary angiogram. Table 4.42 and Figure 4.30 show the results of FDSS compared to two cardiologists using Switzerland data set. The Switzerland is the most independent data set that has no relationship to the training data of FDSS.

Table 4.41 Switzerland data set with the diagnosis of FDSS, cardiologists and angiography

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Cardio#1	Cardio#2	Cardio#3	FDSS	Angio
43	1	4	100	?	?	0	122	0	1.5	3	?	?	0.8	0	1	0.81	1
43	1	4	115	?	0	0	145	1	2	2	?	7	0.85	1	1	0.87	1
43	1	4	140	?	0	1	140	1	0.5	1	?	7	0.8	1	0	0.83	1
45	1	3	110	?	?	0	138	0	0.1	1	?	?	<0.5	0	1	0.13	0
46	1	4	100	?	?	1	133	0	2.6	2	?	?	0.55	0	1	0.83	1
46	1	4	115	?	0	0	113	1	1.5	2	?	7	0.8	1	1	0.91	1
47	1	3	110	?	?	0	120	1	0	?	?	3	0.55	1	1	0.16	1
47	1	3	155	?	0	0	118	1	1	2	?	3	0.55	1	1	0.13	1
47	1	4	110	?	?	1	149	0	2.1	1	?	?	<0.5	0	1	0.61	1
47	1	4	160	?	0	0	124	1	0	2	?	7	0.8	1	1	0.84	1
48	1	4	115	?	?	0	128	0	0	2	?	6	0.65	1	1	0.21	1
50	0	4	160	?	?	0	110	0	0	?	?	3	<0.5	0	1	0.14	1
50	1	4	115	?	0	0	120	1	0.5	2	?	6	0.7	1	1	0.50	1
50	1	4	120	?	0	1	156	1	0	1	?	6	0.6	1	1	0.50	1
51	0	4	120	?	?	0	127	1	1.5	1	?	?	0.6	0	0	0.81	1
57	1	4	110	?	?	1	131	1	1.4	1	1	?	0.65	0	1	0.79	1
57	1	4	140	?	0	0	120	1	2	2	?	6	0.75	1	1	0.81	1
57	1	4	140	?	?	0	100	1	0	?	?	6	0.7	1	1	0.83	1
57	1	4	160	?	?	0	98	1	2	2	?	7	0.9	1	1	0.90	1
57	1	4	95	?	?	0	182	0	0.7	3	?	?	0.75	0	0	0.11	1
58	1	4	115	?	?	0	138	0	0.5	1	?	?	<0.5	0	1	0.73	1
58	1	4	130	?	0	1	100	1	1	2	?	6	0.6	1	1	0.81	1
58	1	4	170	?	?	1	105	1	0	?	?	3	<0.5	0	0	0.74	1
59	1	3	125	?	?	0	175	0	2.6	2	?	?	0.65	0	0	0.40	1
59	1	4	110	?	?	0	94	0	0	?	?	6	<0.5	1	1	0.50	1
59	1	4	120	?	0	0	115	0	0	2	?	3	<0.5	0	1	0.25	1
59	1	4	125	?	?	0	119	1	0.9	1	?	?	0.55	1	1	0.84	1
59	1	4	135	?	0	0	115	1	1	2	?	7	0.85	1	1	0.88	1
60	1	3	115	?	?	0	143	0	2.4	1	?	?	<0.5	0	0	0.82	1
60	1	4	125	?	?	0	110	0	0.1	1	2	?	<0.5	0	1	0.87	1

Table 4.42 Diagnosis performances of FDSS on Switzerland data set

Metrics	Diagnosis Results			
	FDSS	Cardiologist#1	Cardiologist#2	Cardiologist#3
Accuracy	0.77	0.73	0.57	0.77
Sensitivity	0.76	0.72	0.55	0.79
Specificity	1.00	1.00	1.00	0.00

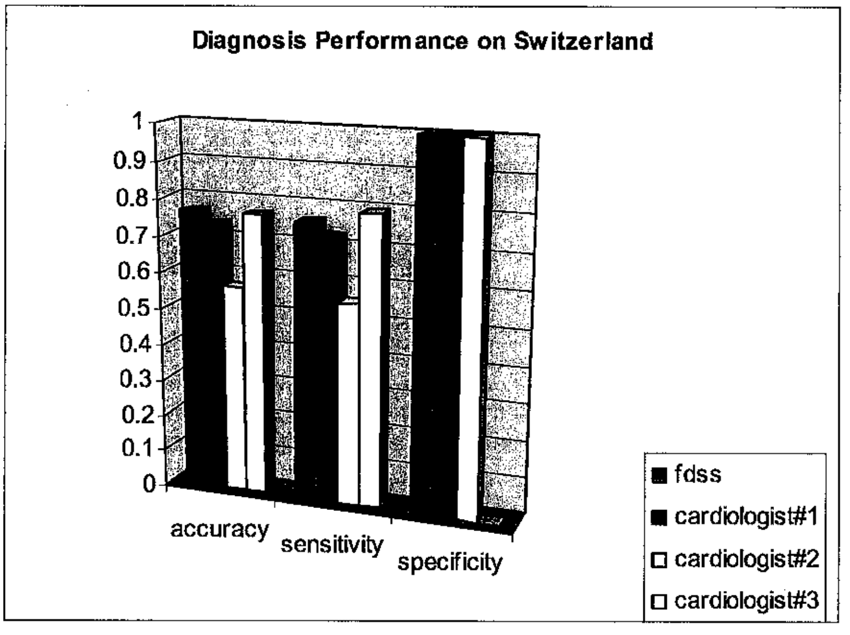


Figure 4.30 Diagnosis performances of FDSS on Switzerland data set

Figure 4.30 shows that FDSS and the third cardiologist have better accuracy and sensitivity. FDSS sensitivity is less than third cardiologist. The specificity is the same between FDSS and the first and second cardiologists. They can have high specificity of 100% because the patient having no CAD is only one and all of them have correct diagnosis.

The ability of FDSS to estimate the possibility of CAD or percentage of blockage of coronary artery based on Table 4.35, Table 4.37, Table 4.39 and Table 4.41 are demonstrated in Figure 4.31 – Figure 4.34. There is only one cardiologist that gives the results as percentage of blockage of the coronary artery. The other cardiologists only give the results of "yes" or "no" which are similar to coronary angiogram results.

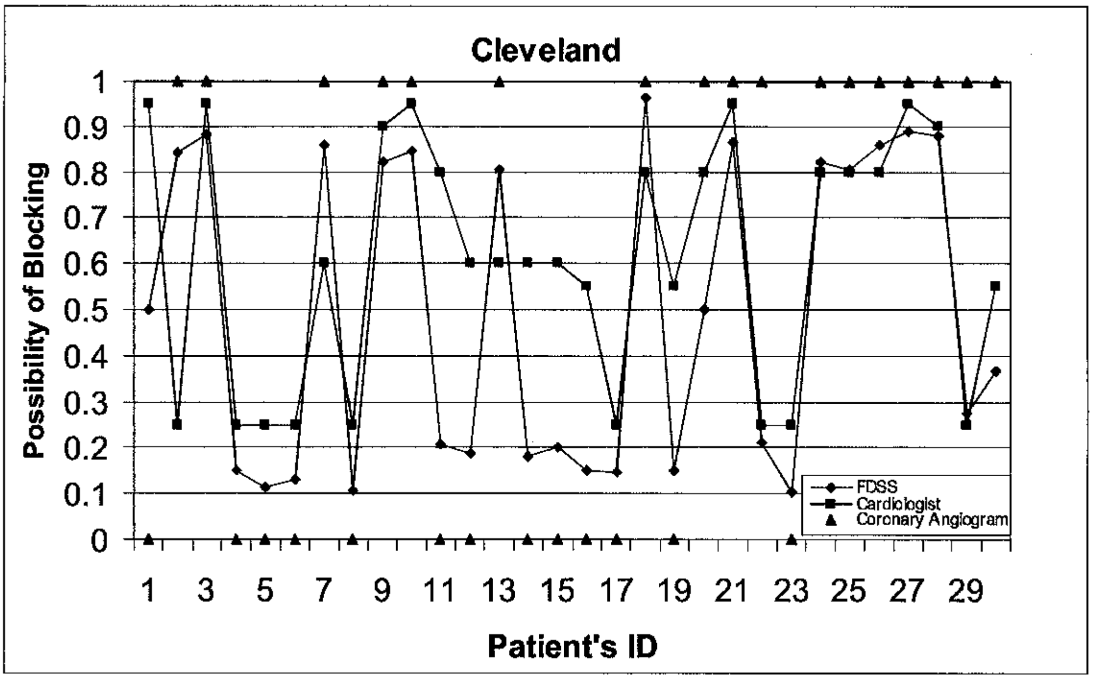


Figure 4.31 Diagnosis Results of FDSS and Cardiologist on Cleveland data set

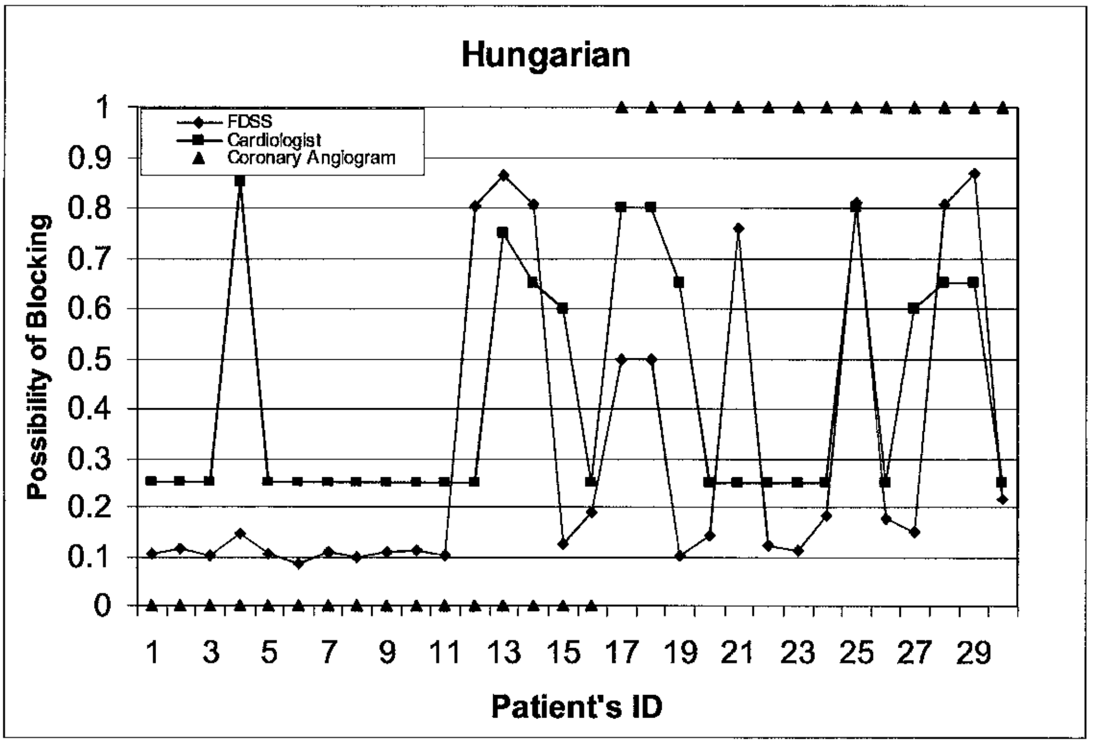


Figure 4.32 Diagnosis Results of FDSS and Cardiologist on Hungarian data set

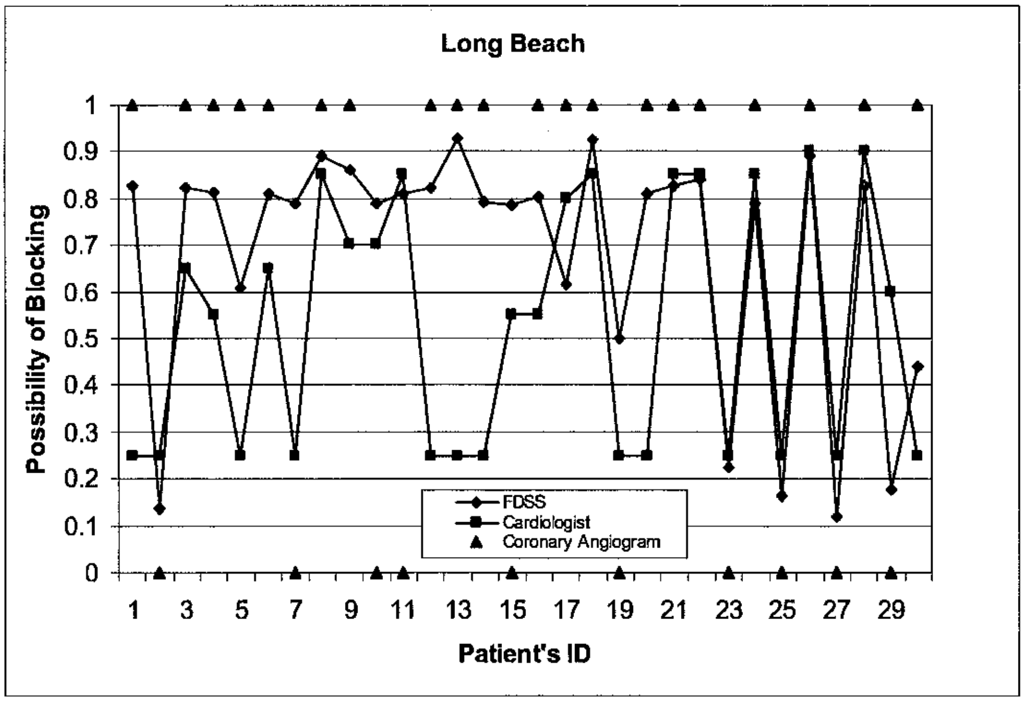


Figure 4.33 Diagnosis Results of FDSS and Cardiologist on Long Beach data set

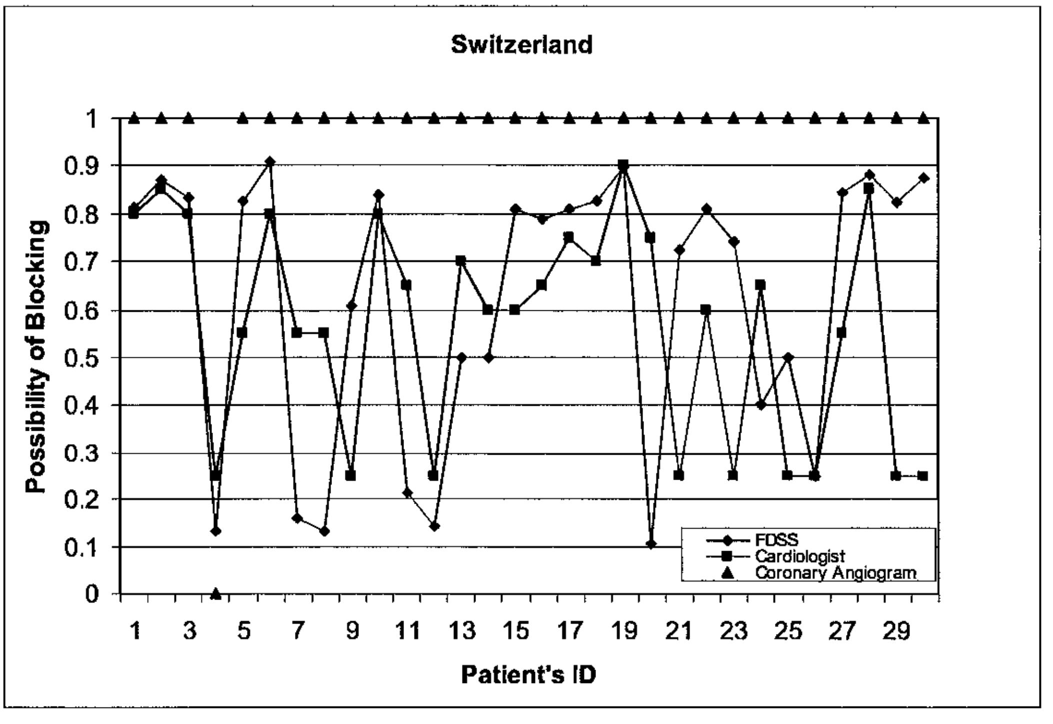


Figure 4.34 Diagnosis Results of FDSS and Cardiologist on Switzerland data set

All the results show that the results of FDSS are generally inline with the cardiologist.

4.6 SUMMARY

In this chapter, the proposed method has been demonstrated in the case of the diagnosis of coronary artery disease. The problem of missing attribute values of the UCI coronary artery disease data sets has been solved using ANNRST data imputation. The ability of RST to reduce the input attributes of ANN has also been demonstrated. It can be seen that the results of ANNRST to impute the simulated missing data are similar to ANN without the input attribute reduction. Hence, ANNRST can be used to replace ANN with simpler topology. The ANNRST performance is verified and evaluated on real missing data. The performance of ANNRST is assessed using four classifiers that built based on imputed data. The quality of classifiers in classifying the testing data can be considered as the performance of the imputation methods which are ANNRST, k-NN and CMCF. In this evaluation method, ANNRST outperforms other methods. RST classifier uses extracted rules to classify the unseen data sets. The extracted rule set of RST consist of many rules. Filtering method that is based on rule support can select small number good quality rules among all rules. During the filtering process, the performance of the RST classifier is decreasing. These filtering effects can be used to measure the quality of imputation methods. ANNRST outperforms the other methods, which are k-NN and CMCF, in this filtering process.

The results of rule filtering is based on support and then refiltered by using the proposed RST rule selection. This second filtering or selection results in only 27 decision rules with acceptable accuracy and coverage. This proposed RST rule filtering also outperforms other rule filtering methods. These selected rules are then fuzzified and included in fuzzy inference engine to build fuzzy decision support system. The membership function development is based on the results of discretization of numerical attributes. Mamdani implication is then used as inference engine. The proposed rule weighing based on supports is also implemented. The results show that fuzzy decision support system has stable and good performance among other classifiers such as MLP-ANN, k-NN, C.45 and RIPPER. This proposed fuzzy decision support system is verified by cardiologists and gives very good results in accuracy, sensitivity and specificity. It also has similar estimation of the possibility of CAD or approximation of coronary artery blockage.

CHAPTER 5: CONCLUSION

5.1 INTRODUCTION

The aim of this research is to develop a decision support system for the diagnosis of CAD based on facts or evidences. Literature survey shows that existing methods for conventional CAD diagnosis still have many limitations which are that they rely too much on human experts, invasive and expensive. The existing computer based and AI based diagnosis methods also still have limitations such as non- interpretable, can not handle heterogeneous and incompleteness of data. They can not learn from evidences directly. Based on the survey, the proposed system must be able to handle incomplete and heterogeneous data with high accuracy at acceptable coverage and has interpretable knowledge. The first problem that must be solved is missing attribute values of the training data which imputation is needed. The second problem is selection of large number and redundant generated rules. The last problem is building the fuzzy decision support system and its evaluations. This chapter is the conclusions that answer the aforementioned problems and explanation of limitation of the proposed methods.

5.2 MISSING DATA IMPUTATION

The problem of missing attribute values of the University California Irvine (UCI) Coronary Artery Disease (CAD) data sets has been solved using proposed ANNRST data imputation. ANNRST imputation accuracy for simulated missing data is similar to pure ANN without RST input attribute reduction. Hence, ANNRST can be used to replace ANN with simpler topology. Both ANNRST and ANN have better accuracy CMCF and k-NN imputations.

The result of the imputation can not be guaranteed as the true values that were missing. In this work the imputed data set evaluation to measure the performance of the imputation methods is proposed. The evaluation is conducted by discovering the knowledge from the imputed data set and then building the classifier based in the discovered knowledge to classify the complete data set. The accuracy and coverage of the classifier that is based on the knowledge of certain imputed data set can be considered as the performance of those certain imputation method. In case of real missing values, ANNRST outperforms other methods in term of accuracy, coverage, sensitivity and specificity of classifiers that uses imputed data set by ANNRST.

RST classifier uses extracted rules to classify the unseen data sets. The extracted rule set of RST consist of many rules. Filtering method that based on rule support can selected the small number good quality rules from the other rules. During the filtering process, the performance of the RST classifier is reduced. These filtering effects can be used to measure the quality of imputation methods. ANNRST outperforms the other methods, which are k-NN and CMCF in this filtering process.

5.3 RULE DISCOVERY AND SELECTION

Support filtering as one of rule quality based filtering is applied at the first stage of the selection. It is time consuming and high computation cost to apply only RST based rule selection because the number of attributes is very high. Filtering method based on rule support is applied to reduce the number of rules before applying rule importance measure to select the most importance rules. The modification is proposed by applying this method to decision system and converting rules to form decision tables based on testing data instead of training data for rule importance measurement. This idea is proposed to increase the generalization of the RST classifier. The classifier learns mainly from training data but also let the testing data contribute to the learning process through the rule selection. The results of rule filtering based on support then refiltered by using the proposed RST rule selection. This second filtering or selection results in only 27 decision rules with acceptable accuracy and coverage. This proposed RST rule filtering also outperformed other rule filtering methods which are Michalski, Torgo, Bradzil, Pearson and Cohen.

5.4 FUZZY DECISION SUPPORT SYSTEM AND EVALUATIONS

Fuzzy system has been proven powerful to many application areas. Fuzzy system has the ability to handle imprecise knowledge by means of fuzzy linguistic terms. The main disadvantage of fuzzy system is the difficulty in preparing the knowledge. Fuzzy system can not learn from the data directly. The need of human expert is obvious in the development of fuzzy system. The involving of human expert sometimes makes the fuzzy system subjective. For example, determining the membership functions will need domain expert.

RST can learn directly from the data. This is the advantage of RST system. The system will be objective. One of the drawbacks of RST is that the generated rules are crisp instead of fuzzy. Combining RST and fuzzy system will make the system powerful. The system will be objective and be able to handle the imprecise knowledge and fuzziness.

The selected rules base on RST are fuzzified and included in fuzzy inference engine to build fuzzy decision support system. The membership function development is based on the results of discretization of numerical attributes. Mamdani implication is then used as inference engine. The proposed rule weighing based on supports is also implemented. The results show that fuzzy decision support system has stable and good performance among other classifiers which are MLP-ANN, k-NN, C.45 and RIPPER. This proposed fuzzy decision support system is verified by three local cardiologists and in general gives very good results in accuracy, sensitivity and specificity. It also has similar estimation of the possibility of CAD or approximation of coronary artery blockage.

5.5 SUMMARY

The related works on coronary artery disease diagnosis using artificial intelligent and soft computing methods are reviewed. This can be the starting point for future researchers.

Diagnosis of coronary artery disease needs much information in order to draw the correct decision. Even though the complete historical data, symptoms and medical test from patient are available, cardiologist still may get confused to draw conclusion of the presence of coronary artery disease. Coronary angiography is preferred by cardiologists to diagnose CAD with high accuracy even it is invasive, risky and expensive.

Advanced computer methods such as artificial intelligence, soft computing and expert systems is applied as alternative methods to diagnose coronary artery disease. The new emerging data analysis and rule induction method, rough set theory, are applied on attributes reduction as well as rule induction.

Artificial neural network based missing data imputations have been proven to be success for estimating missing values in input attributes. The rough set concept of attribute reduction is combined with artificial neural network to impute missing attribute values in data sets.

Rough set theory is used to measure the importance of rules then is applied in rule selection problem. Rule quality and importance measure is used in rule selection problem to give the high quality and important selected rules.

The use of fuzzy theory is very useful for decision making problems. The combination of fuzzy theory and rough set theory to build decision support system for the diagnosis of coronary artery disease is proposed in this thesis.

To overcome incompleteness of data, a missing data imputation is developed based on artificial neural network with rough set theory attribute reduction, namely ANNRST. ANNRST is applied on simulated missing values on selected UCI data sets. The results show that ANNRST is better than other methods such as CMCF and k-NN imputation methods.

A new evaluation method for real missing attribute values imputation is proposed. The imputation method can be considered better when the filtered rules are still has high performance during filtering process.

Rule selection based on combined rule quality and importance is proposed. Rule quality based on support is fast and simple to filter very large number of rules. This proposed rule selection can select 27 from 3881 rules. The selected rule set is compared to Michalski, Torgo, Brazdil, Pearson and Cohen rule quality filtering. The proposed rule selection perform better that the others.

Fuzzy decision support system for the diagnosis of coronary artery disease is proposed. The membership function creation is based on Boolean reasoning discretization of numerical attributes which is considered indirect learning from training decision table. The rule weighing based on rule support is proposed to assign the strength of fuzzy rules during inference process. The defuzzification of output provides numerical values than can be considered as percentage of blocking possibility of coronary artery. The results show that fuzzy decision support system has stable and good performance among other classifiers such as MLP-ANN, k-NN, C.45 and RIPPER. This proposed fuzzy decision support system is verified by cardiologists and gives very good results in accuracy, sensitivity and specificity. It also has similar estimation of the possibility of CAD or approximation of coronary artery blockage.

The result of FDSS provides the percentage of blocking of coronary artery. Cardiologist can not provide the percentage of blocking by using historical, simple blood test data and ECG, even if thallium scintigraphic is performed. Coronary angiography which is invasive diagnosis also can not give the percentage of blocking. Cardiologist gives decision by observing the angiogram results and indicates CAD if the blocking is more than 50%. By FDSS, cardiologist can approximate the blocking of coronary artery and draw a decision whether coronary angiography should be performed or not.

5.6 RECOMMENDATIONS

The following recommendations are suggested in order to improve the proposed method in this thesis.

i. Optimization of artificial neural network parameters may improve the accuracy of ANNRST.

ii. Continuous or numerical value handling for RST data analysis may be developed. Tolerance based or variable precision RST can be considered.

iii. A method of choosing correct parameters and performance measures of rule selection using rule quality and importance measure may be developed to improve the rule selection become more objective and less subjective.

iv. Membership function optimization and adjustment may be applied to improve the diagnosis performance of fuzzy decision support system. Some optimization algorithms can be used such as genetic algorithm, particle swarm optimization and similar soft computing techniques. Adaptive Neuro Fuzzy Inference System (ANFIS) can also be considered.

REFERENCES

[1] B. L. Zaret, M. Moser, and L. S. Cohen, *Yale University School of Medicine Heart Book.* New York: Hearst Books, 1992.

[2] B. Phibbs, *The Human Heart: A Basic Guide to Heart Disease.* Philadelphia: Lippincott Williams & Wilkins, 2007.

[3] A. Selzer, *Understanding Heart Disease.* Berkeley: University of California Press, 1992.

[4] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, M. May, and P. Brindle, "Derivation and Validation of Qrisk, a New Cardiovascular Disease Risk Score for the United Kingdom: Prospective Open Cohort Study," *British Medical Journal*, pp. 136, 2007.

[5] J. Mackay and G. A. Mensah, "The Atlas of Heart Disease and Stroke," World Health Organization 2009.

[6] O. S. Randall and D. S. Romaine, *The Encyclopedia of the Heart and Heart Disease.* New York, NY: Facts on File, 2005.

[7] "Coronary Heart Disease Statistics Fact Sheets 2008/2009," British Heart Foundation, London 2008.

[8] National Heart Association of Malaysia, "Heart Diseases on the Rise, Second Leading Killer", http://www.malaysianheart.org/article.php?aid=35, Accessed: 21-07-2009.

[9] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A Multilayer Perceptron-Based Medical Decision Support System for Heart Disease Diagnosis," *Expert Systems with Applications*, vol. 30, pp. 272, 2006.

[10] K. G. Jayanta and V. Marco, "Building a Bayesian Network Model of Heart Disease," presented at Proceedings of the 38th annual on Southeast regional conference, Clemson, South Carolina, 2000. pp. 239-240

[11] D. Gamberger, G. Krstačić, and T. Šmuc, "Medical Expert Evaluation of Machine Learning Results for a Coronary Heart Disease Database," in *Medical Data Analysis*, 2000, pp. 119.

[12] J. Komorowski and A. Ohrn, "Modelling Prognostic Power of Cardiac Tests Using Rough Sets," *Artificial Intelligence in Medicine*, vol. 15, pp. 167, 1999.

[13] M. Kukar, C. Groselj, I. Kononenko, and J. J. Fettich, "An Application of Machine Learning in the Diagnosis of Ischaemic Heart Disease," presented at

Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems (CBMS '97), 1997. pp. 70

[14] G. Bologna, A. Rida, and C. Pellegrini, "Intelligent Assistance for Coronary Heart Disease Diagnosis: A Comparison Study," in *Artificial Intelligence in Medicine*, 1997, pp. 199.

[15] M. G. Tsipouras, T. P. Exarchos, D. I. Fotiadis, A. Kotsia, A. Naka, and L. K. Michalis, "A Decision Support System for the Diagnosis of Coronary Artery Disease," presented at 19th IEEE Symposium on Computer-Based Medical Systems, 2006. pp. 279-284

[16] M. G. Tsipouras, T. P. Exarchos, D. I. Fotiadis, A. P. Kotsia, K. V. Vakalis, K. K. Naka, and L. K. Michalis, "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 12, pp. 447, 2008.

[17] G. T. Anderson, J. Zheng, R. Wyeth, A. Johnson, J. Bissett, and T. P. Group, "A Rough Set/Fuzzy Logic Based Decision Making System for Medical Applications," *International Journal of General Systems*, vol. 29, pp. 879 - 896, 2000.

[18] S. Mitra and Y. Hayashi, "Bioinformatics with Soft Computing," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 36, pp. 616-634, 2006.

[19] Z. Pawlak, "Rough Sets," *International Journal of Computer and Information Sciences*, vol. 11, pp. 341-355, 1982.

[20] Z. Pawlak, "Rough Set Approach to Knowledge-Based Decision Support," *European Journal of Operational Research*, vol. 99, pp. 48-57, 1997.

[21] C. Wu, Y. Yue, M. Li, and O. Adjei, "The Rough Set Theory and Applications," *Engineering Computations: Int J for Computer-Aided Engineering*, vol. 21, pp. 488, 2004.

[22] J. Grzymala-Busse and M. Hu, "A Comparison of Several Approaches to Missing Attribute Values in Data Mining," in *Rough Sets and Current Trends in Computing*, 2001, pp. 378.

[23] L. Al Shalabi, M. Najjar, and A. Al Kayed, "A Framework to Deal with Missing Data in Data Sets," *Journal of Computer Science*, vol. 2, pp. 740-745, 2006.

[24] J. Li and N. Cercone, "Assigning Missing Attribute Values Based on Rough Sets Theory," presented at IEEE International Conference on Granular Computing, 2006. pp. 607

[25] I. Wasito and B. Mirkin, "Nearest Neighbour Approach in the Least-Squares Data Imputation Algorithms," *Information Sciences*, vol. 169, pp. 1, 2005.

[26] I. Wasito and B. Mirkin, "Nearest Neighbours in Least-Squares Data Imputation Algorithms with Different Missing Patterns," *Computational Statistics & Data Analysis*, vol. 50, pp. 926, 2006.

[27] A. Ragel and B. Cremilleux, "Mvc--a Preprocessing Method to Deal with Missing Values," *Knowledge-Based Systems*, vol. 12, pp. 285, 1999.

[28] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. Altman, "Missing Value Estimation Methods for DNA Microarrays," *Bioinformatics*, vol. 17, pp. 520-525, 2001.

[29] X. Wang, A. Li, Z. Jiang, and H. Feng, "Missing Value Estimation for DNA Microarray Gene Expression Data by Support Vector Regression Imputation and Orthogonal Coding Scheme," *BMC Bioinformatics*, vol. 7, pp. 32, 2006.

[30] H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, "Methods for Imputation of Missing Values in Air Quality Data Sets," *Atmospheric Environment*, vol. 38, pp. 2895, 2004.

[31] B. Bhattacharya, D. L. Shresta, and D. P. Solomatine, "Neural Networks in Constructing Missing Wave Data in Sedimentation Modelling," presented at XXXth IAHR Congress, Thessaloniki, Greece, 2003. pp. -

[32] P. Siripitayananon, C. Hui-Chuan, and J. Kang-Ren, "Estimating Missing Data of Wind Speeds Using Neural Network," presented at IEEE SoutheastCon., 2002. pp. 343

[33] T. Agotnes, "Filtering Large Propositional Rule Sets While Retaining Classifier Performance," in *Department of Computer and Information Science*: Norwegian Universtiy of Science and Technology, 1999, pp. 143.

[34] M. Maddouri and J. Gammoudi, "On Semantic Properties of Interestingness Measures for Extracting Rules from Data," *Adaptive and Natural Computing Algorithms*, pp. 148, 2007.

[35] J. Li and N. Cercone, "Introducing a Rule Importance Measure," *Transactions on Rough Sets V*, pp. 167, 2006.

[36] N. H. L. a. B. Institute, "Atherosclerosis", "http://www.nhlbi.nih.gov/health/dci/Diseases/Cad/CAD_WhatIs.html", Accessed: 22-05-09.

[37] N. H. L. a. B. Institute, "Heart with Muscle Damage and a Blocked Artery", "http://www.nhlbi.nih.gov/health/dci/Diseases/HeartAttack/HeartAttack_WhatIs.html", Accessed: 22-05-09.

[38] H. Chen, *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. New York, NY: Springer, 2005.

[39] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease," *The American Journal of Cardiology*, vol. 64, pp. 304-310, 1989.

[40] K. G. Jayanta and V. Marco, "Probabilistic Model Building of Heart Disease," Department of Computer Science, University of South Carolina 1999.

[41] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-Based Learning Algorithms," *Machine Learning*, vol. 6, pp. 37, 1991.

[42] D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz, "Uci Repository of Machine Learning Databases," University California Irvine, Department of Information and Computer Science, 1998.

[43] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, Calif.: Morgan Kaufmann Publishers, 1993.

[44] A. Liping and T. Lingyun, "A Rough Neural Expert System for Medical Diagnosis," presented at International Conference on Services Systems and Services Management, 2005. pp. 1130

[45] D. Gamberger, "A Minimization Approach to Propositional Inductive Learning," presented at Proceedings of the 8th European Conference on Machine Learning, 1995. pp. 151

[46] K. Polat, S. Sahan, and S. Gunes, "Automatic Detection of Heart Disease Using an Artificial Immune Recognition System (Airs) with Fuzzy Resource Allocation Mechanism and K-Nn (Nearest Neighbor) Based Weighting Preprocessing," *Expert Systems with Applications*, vol. 32, pp. 625-631, 2007.

[47] K. Polat, S. Gunes, and S. Tosun, "Diagnosis of Heart Disease Using Artificial Immune Recognition System and Fuzzy Weighted Pre-Processing," *Pattern Recognition*, vol. 39, pp. 2186-2193, 2006.

[48] C. Ordonez, "Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction," *IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society*, vol. 10, pp. 334, 2006.

[49] J. Grzymala-Busse, "Lers—a Data Mining System," in *Data Mining and Knowledge Discovery Handbook*, 2005, pp. 1347.

[50] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*: Addison Wesley, 2006.

[51] Z. Suraj, "Special Session on Rough Set Theory and Its Applications," presented at 1st International Computer Engineering Conference on New Technologies for the information society (ICENCO'2004), Egypt, 2004 of Conference. pp. Pages

[52] T. Munakata, *Fundamentals of the New Artificial Intelligence: Beyond Traditional Paradigms*. New York: Springer, 1998.

[53] Z. Pawlak, aw, and A. Skowron, "Rough Membership Functions," in *Advances in the Dempster-Shafer Theory of Evidence*: John Wiley \& Sons, Inc., 1994, pp. 251.

[54] W. Ziarko, "Rough Sets: Trends, Challenges, and Prospects," in *Rough Sets and Current Trends in Computing*, 2001, pp. 1.

[55] A. Ohrn, "Discernibility and Rough Sets in Medicine: Tools and Applications," in *Department of computer and information science*. Trondheim: Norwegian University of Science and Technology, 1999, pp. 223.

[56] J. G. Bazan, M. S. Szczuka, A. Wojna, and M. Wojnarski, "On the Evolution of Rough Set Exploration System," in *Rough Sets and Current Trends in Computing*, 2004, pp. 592.

[57] J. Bazan and M. Szczuka, "Rses and Rseslib - a Collection of Tools for Rough Set Computations," in *Rough Sets and Current Trends in Computing*, 2001, pp. 106.

[58] B. Predki, R. Stowinski, J. Stefanowski, and R. Susmaga, "Rose - Software Implementation of the Rough Set Theory," *Lecture notes in computer science.*, pp. 605, 1998.

[59] B. Predki and S. Wilk, "Rough Set Based Data Exploration Using Rose System," *Lecture notes in computer science.*, pp. 172, 1999.

[60] I. Duntsch and G. Gediga, *Rough Set Data Analysis: A Road to Non-Invasive Knowledge Discovery*. Bangor, N.I.: Methodos Publishers (UK), 2000.

[61] S. Haykin, *Neural Network a Comprehensive Foundation*: Prentice Hall International Inc., 1999.

[62] M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The Rprop Algorithm," presented at IEEE International Conference on Neural Networks, 1993. pp. 586

[63] I. Wasito, "Least Squares Algorithms with Nearest Neighbour Techniques for Imputing Missing Data Values," in *School of Computer Science and Information Systems*: University of London, 2003.

[64] J. Li, "Rough Set Based Rule Evaluations and Their Applications." Waterloo, Ontario: University of Waterloo, 2007, pp. 191.

[65] F. V. Nelwamondo and T. Marwala, "Rough Set Theory for the Treatment of Incomplete Data," presented at Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International, 2007. pp. 1

[66] F. Fessant and S. Midenet, "Self-Organising Map for Data Imputation and Correction in Surveys," *Neural Computing and Applications*, vol. 10, pp. 300-310, 2002.

[67] W. Wei and Y. Tang, "Ageneric Neural Network Approach for Filling Missing Data in Data Mining," presented at IEEE International Conference on Systems, Man and Cybernetics, Washington, DC, 2003. pp. 862-867

[68] B. Bhattacharya and D. P. Solomatine, "Machine Learning in Sedimentation Modelling," *Neural Networks*, vol. 19, pp. 208, 2006.

[69] C.-H. Wu, C.-H. Wun, and H.-J. Chou, "Using Association Rules for Completing Missing Data," presented at 4th International Conference on Hybrid Intelligent Systems, 2004. pp. 236-241

[70] A. Ragel and a. Crémilleux, "Treatment of Missing Values for Association Rules," in *Research and Development in Knowledge Discovery and Data Mining*, 1998, pp. 258.

[71] H. Kiers, "Weighted Least Squares Fitting Using Ordinary Least Squares Algorithms," *Psychometrika*, vol. 62, pp. 251, 1997.

[72] M. Giardina, Y. Huo, F. Azuaje, P. McCullagh, and R. Harper, "A Missing Data Estimation Analysis in Type Ii Diabetes Databases," presented at Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on, 2005. pp. 347

[73] J. W. Grzymala-Busse, "Rough Set Strategies to Data with Missing Attribute Values," presented at Workshop on Foundations and New Directions in Data

Mining, associated with the third IEEE International Conference on Data Mining, Melbourne, FL, USA, 2003. pp. 56-63

[74] J. W. Grzymala-Busse and S. Siddhaye, "Rough Set Approaches to Rule Induction from Incomplete Data," presented at 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia, Italy, 2004. pp. 923-930

[75] Q. Shen and A. Choucholas, "Rough Set-Based Dimensionality Reduction for Supervised and Unsupervised Learning," *International Journal of Applied Mathematics and Computer Science*, vol. 11, pp. 583-601, 2001.

[76] N. O. Attoh-Okine, "Combining Use of Rough Set and Artificial Neural Networks in Doweled-Pavement-Performance Modeling-a Hybrid Approach," *Journal of Transportation Engineering*, vol. 128, pp. 270-275, 2002.

[77] J. Peters and M. Szczuka, "Rough Neurocomputing: A Survey of Basic Models of Neurocomputation," in *Rough Sets and Current Trends in Computing*, 2002, pp. 951.

[78] B. S. Ahn, S. S. Cho, and C. Y. Kim, "The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction," *Expert Systems with Applications*, vol. 18, pp. 65, 2000.

[79] S. Tsumoto, "Accuracy and Coverage in Rough Set Rule Induction," in *Rough Sets and Current Trends in Computing*, J. J. Alpigini, et. al., Ed. Berlin, Heidelberg: Springer-Verlag, 2002, pp. 373-380.

[80] J. Li and N. Cercone, "Discovering and Ranking Important Rules," presented at Granular Computing, 2005 IEEE International Conference on, 2005. pp. 506

[81] J. Li, P. Pattaraintakorn, and N. Cercone, "Rule Evaluations, Attributes, and Rough Sets: Extension and a Case Study," *Transactions on Rough Sets VI*, pp. 152, 2007.

[82] J. Li and N. Cercone, "A Rough Set Based Model to Rank the Importance of Association Rules," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 2005, pp. 109.

[83] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353, 1965.

[84] C. C. Siu, Q. Shen, and R. Milne, "A Fuzzy Expert System for Turbomachinery Diagnosis," presented at IEEE Int. Conf. on Fuzzy Systems, 1997. pp. 555-560

[85] C. Koutsojannis and I. Hatzilygeroudis, "Fesmi: A Fuzzy Expert System for Diagnosis and Treatment of Male Impotence," in *Knowledge-Based Intelligent Information and Engineering Systems*, 2004, pp. 1106.

[86] L. A. Carreno and Y. Jani, "A Fuzzy Expert System Approach to Insurance Risk Assessment Using Fuzzyclips," presented at WESCON/'93. Conference Record, 1993. pp. 536

[87] Y. Cho, K. Lee, J. Yoo, and M. Park, "Autogeneration of Fuzzy Rules and Membership Functions for Fuzzy Modelling Using Rough Set Theory," *IEE Proc.-Control Theory Appl.*, vol. 145, pp. 437-442, 1998.

[88] L. Yinghua and G. A. Cunningham, "A New Approach to Fuzzy-Neural System Modeling," *Fuzzy Systems, IEEE Transactions on*, vol. 3, pp. 190, 1995.

[89] G. Drwal and M. Sikora, "Fuzzy Decision Support System with Rough Set Based Rules Generation Method," in *Rough Sets and Current Trends in Computing*, 2004, pp. 727.

[90] M. Sikora, "Fuzzy Rules Generation Method for Classification Problems Using Rough Sets and Genetic Algorithms," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 2005, pp. 383.

[91] M. Tsipouras, T. Exarchos, C. Papaloukas, A. Bechlioulis, A. Kotsia, T. Nanou, C. Bazios, Y. Antoniou, D. Fotiadis, A. Naka, and L. Michalis, "Automatic Creation of Decision Support Systems: Application and Results in the Cardiovascular Diseases Domain," *The Journal on Information Technology in Healthcare*, vol. 4, pp. 222-230, 2006.

[92] M. G. Tsipouras, T. P. Exarchos, C. Papaloukas, A. Bechlioulis, A. Kotsia, T. Nanou, C. Bazios, Y. Antoniou, D. I. Fotiadis, A. Naka, and L. K. Michalis, "Automatic Creation of Decision Support Systems: Application and Results in the Cardiovascular Diseases Domain," presented at 4th International Conference on Information and Communication Technologies in Health, Greece, 2006. pp.

[93] H. Watanabe, W. J. Yakowenko, K. Yong-Mi, J. A. A. J. Anbe, and T. A. T. T. Tobi, "Application of a Fuzzy Discrimination Analysis for Diagnosis of Valvular Heart Disease," *IEEE Transactions on Fuzzy Systems*, vol. 2, pp. 267, 1994.

[94] R. Jensen, "Combining Rough and Fuzzy Sets for Feature Selection," in *School of Informatics*: University of Ediburgh, 2005, pp. 221.

[95] K. L. Du and M. N. S. Swamy, *Neural Network in a Softcomputing Framework*: Springer, 2006.

[96] T. Fawcett, "An Introduction to Roc Analysis," *Pattern Recognition Letters*, vol. 27, pp. 861, 2006.

[97] J. H. Gennari, P. Langley, and D. Fisher, "Models of Incremental Concept Formation," *Artificial Intelligence*, vol. 40, pp. 11-61, 1989.

[98] W. Duch, R. Adamczak, and K. Grabczewski, "A New Methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logical Rules," *IEEE Transactions on Neural Networks*, vol. 11, pp. 1-31, 2000.

[99] N. Jankowski and N. Kadirkamanathan, "Statistical Control of Rbf-Like Networks for Classification," presented at 7th International Conference on Artificial Neural Networks, Lausanne, Switzerland, 1997. pp. 385-290

[100] C. E. Pedreira, L. Macrini, and E. S. Costa, "Input and Data Selection Applied to Heart Disease Diagnosis," presented at International Joint Conference on Neural Networks, Montreal, Canada, 2005. pp. 2389-2393

[101] A. Skowron and C. Rauszer, "The Discernibility Matrices and Functions in Information Systems," in *Intelligent Decision Support: Handbook of Applications and Advances of Rough Sets Theory*, R. Slowinski, Ed. Dordrecht: Kluwer Academic Publisher, 1992, pp. 331-362.

[102] D. S. Johnson, "Approximation Algorithms for Combinatorial Problems," *Journal of Computer and System Sciences*, vol. 9, pp. 256-278, 1974.

[103] N. Wanas, G. Auda, M. S. Kamel, and F. A. K. F. Karray, "On the Optimal Number of Hidden Nodes in a Neural Network," presented at IEEE Canadian Conference on Electrical and Computer Engineering., 1998. pp. 918

[104] D. Nguyen and B. Widrow, "Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights," presented at 1990 IJCNN International Joint Conference on Neural Networks, 1990. pp. 21

[105] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," presented at IJCAI, 1995. pp. 1145

[106] M. Wojnarski, "Ltf-C: Architecture, Training Algorithm and Applications of New Neural Classifier," *Fundamenta Informaticae*, vol. 54, pp. 89-105, 2003.

[107] E. H. Mamdani, "Applications of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis," *IEEE Transactions on Computers*, vol. 26, pp. 1182-1191, 1977.

[108] I. H. Witten, *Data Mining: Practical Machine Learning Tools and Techniques*. s.l.: Morgan Kaufmann Publications, 2005.

[109] W. A. David, K. Dennis, and K. A. Marc, "Instance-Based Learning Algorithms," *Mach. Learn.*, vol. 6, pp. 37-66, 1991.

[110] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 01, pp. 81, 1986.

[111] W. W. Cohen, "Fast Effective Rule Induction," *Machine Learning -International Workshop then Conference-*, pp. 115, 1995.

List of Publications and Awards

Publications

1. Setiawan, N.A., P.A. Venkatachalam, and A.F.M. Hani. *Missing Data Estimation on Heart Disease Using Artificial Neural Network and Rough Set Theory*. in *International Conference on Intelligent and Advanced Systems*. 2007. Kuala Lumpur, Malaysia: IEEE.
2. Setiawan, N.A., P.A. Venkatachalam, and A.F.M. Hani. *Missing Attribute Value Prediction Based on Artificial Neural Network and Rough Set Theory*. in *International Conference on Biomedical Engineering and Informatics*. 2008. Sanya, Hainan, China: IEEE.
3. Setiawan, N.A., P.A. Venkatachalam, and A.F.M. Hani. *A Comparative Study of Imputation Methods to Predict Missing Attribute Values in Coronary Heart Disease Data Set*. in *4th Kuala Lumpur International Conference on Biomedical Engineering*. 2008. Kuala Lumpur: Springer.
4. Setiawan, N.A., P.A. Venkatachalam, and A.F.M. Hani. *A Rough-Set-Based Knowledge Discovery from Incomplete Coronary Artery Disease Data Sets*. Accepted in IEEE Transactions on Man, System and Cybernetics, Part C: Application and Review.
5. Setiawan, N.A., P.A. Venkatachalam, and A.F.M. Hani. *Fuzzy Decision Support System for Coronary Artery Disease Diagnosis Based on Rough Set Theory*. Accepted in Informatics in Primary Care, Radcliffe Publication.

Awards

1. P.A. Venkatachalam, N.A. Setiawan and Ahmad Fadzil M Hani, *Fuzzy Decision Support System to Diagnose Heart Problem by RST (CAD-RFDSS)*, Silver Medal in the Category of Biotechnology, Health and Fitness, International Invention, Innovation and Technology Exhibition (ITEX) 2008, Kuala Lumpur, Malaysia.
2. P.A. Venkatachalam, N.A. Setiawan and Ahmad Fadzil M Hani, *Fuzzy DSS to Diagnose Heart Problem by RST* , Gold Medal in the Category of Health/Fitness, Invention & New Exposition (INPEX) 2008, Pittsburgh, U.S.A.

3. P.A. Venkatachalam, N.A. Setiawan and Ahmad Fadzil M Hani, *Fuzzy DSS to Diagnose Heart Problem by RST*, Silver Medal in the Category of Specialized Technology, Invention & New Product Exposition (INPEX) 2008, Pittsburgh, U.S.A.
4. P.A. Venkatachalam, N.A. Setiawan and Ahmad Fadzil M Hani, *Coronary Artery Disease Blocking Estimator (CADBE)*, Gold Medal in the Category of Biotechnology, Health and Fitness, International Invention, Innovation and Technology Exhibition (ITEX) 2009, Kuala Lumpur, Malaysia.

Appendix

Short information of the cardiologists:

Dr. Choong Choon Hooi
Consultant Cardiologist and Physician
Qualifications:
Doctor of Medicine (MD) from Universiti Kebangsaan Malaysia
Membership of The Royal Colleges of Physicians of The United Kingdom (MRCP(UK))
Office:
Hospital Fatimah Ipoh, No 1, Leboh Chew Peng Loon, Ipoh Garden, 31400 Ipoh, Perak.
Tel/Fax: 05-547 8281 (D/L)

Dr Tan Huat Chai
Consultant Cardiologist and Physician
Qualifications:
Bachelor of Medicine and Bachelor of Surgery (MBBS) from Universiti Malaya
Membership of The Royal Colleges of Physicians of The United Kingdom (MRCP(UK))
Office:
Hospital Fatimah Ipoh, No 1, Leboh Chew Peng Loon, Ipoh Garden, 31400 Ipoh, Perak.
Tel/Fax: 05-547 9663 (D/L)

Dr Hasral Noor Hasni
Consultant Cardiologist and Physician
Qualifications:
Bachelor of Medicine and Bachelor of Surgery (MBBS) from University of Queensland
Membership of The Royal Colleges of Physicians of The United Kingdom (MRCP(UK))
Office:
Ipoh Specialist Hospital 26 Jalan Raja Dihilir 30350 Ipoh
Tel: 05-240 8777

Experiences:

POST	HOSPITAL	PERIOD
House Officer	Hospital Teluk Intan	1997-1998
Medical Officer	Hospital Gerik	1998-1999
Medical Officer in Paediatric Institute	Hospital Kuala Lumpur	1999-2000
Medical Officer in Department of Medicine	Hospital Selayang	2000-2002
Physician in Department of Medicine	Hospital Kuala Lumpur	2002-2003
Clinical attachment in Depatment of Medicine	Ninewells Hospital Dundee,UK	June 2002-July 2002
Senior Register Cardiologist in Department of Cardiology	Institut Jantung Negara (National Heart Institute)	July 2003-may 2007