CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Data mining is a useful tool to extract information and pattern from data for use in decision support and estimation. In medical field, data mining is used to recognize and classify patterns in multivariate patient attributes. Future outcomes can also be predicted based on previous experience and present conditions. Classification is a two-step process, in which a model can be constructed from a set of predetermined classes and it is used for classifying future or unknown objects. Cardiovascular Disease (CVD) is a disease that affects the functioning of the heart. The common type of cardiovascular disease is known as myocardial infarction. The data mining concepts, role of data mining in medicine, the types of cardiovascular diseases, the motivation and objective of the developed classifiers for the common type of cardiovascular disease known as myocardial infarction are discussed in this chapter. This chapter also describes how the entire research is organized in this thesis.

1.2 DATA MINING

Data mining is the analysis step of the knowledge discovery in database’s process. It is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics and systems. Data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The
commonly used data mining techniques are association, classification, prediction, clustering and sequential pattern analysis.

Machine learning is the study of computer algorithms that improve automatically through experience. Applications of machine learning range from data mining programs that discover general rules in large data sets to information filtering systems that automatically learn users' interests. Machine learning can be used to develop systems resulting in increased efficiency and effectiveness.

One of the commonly used machine learning techniques is classification. Classification is probably the oldest and most widely-used of all the knowledge discovery approaches. Classification is learning a function that maps a data item on to one of several redefined classes. Patterns that are extracted using machine intelligence can be used to predict which class the data falls under. Classification is the process in which one uses some of the data to categorize the present data set into different classes for future use. Classification is done in such a way that the properties of each data sample in a particular class are similar to each other.

Medical data mining is becoming increasingly important in healthcare. Medical databases have accumulated large quantities of information about patients and their clinical histories. It may contain data with characteristics such as incompleteness, incorrectness, and inexactness. Classification analysis is the widely adopted data mining techniques for healthcare applications to support and improve the quality of medical diagnosis. In medical classification, all approaches known use the knowledge of the already present data sets. Usually, a subset of the present data set is used as the training sample for the classification technique. Different data mining approaches can be used for classification.
1.3 CARDIOVASCULAR DISEASES

Any kind of disease that affects the functioning of cardiovascular system is known as CVD. Ischaemic heart disease is the most common type of cardiovascular disease around the world. It refers to problems with the circulation of blood to the heart muscle.

An ischaemic stroke is the most common type of stroke, and it happens when a blood vessel that feeds the brain gets blocked, usually from a blood clot. It is refers to a problem with the circulation of blood in the blood vessels of the brain. When the blood supply to a part of the brain is shut off, the brain cells will die. A complete blockage with long-term effects is referred to as a cerebrovascular thrombosis or a cerebrovascular stroke.

Heart failure means the heart is not pumping blood as well as it should. This does not mean that the heart stops beating. The heart keeps working, but the body’s need for blood and oxygen is not being met. It occurs when the pumping action of the heart cannot provide enough blood to the rest of the body as it is needed.

Arrhythmia is an abnormal rhythm of the heart. There are various types of arrhythmia. The heart can beat too slowly, too fast or irregularly. Bradycardia is when the heart rate is less than 60 beats per minute. Tachycardia is when the heart rate is more than 100 beats per minute. An arrhythmia can affect the heart functions. The heart may not be able to pump enough blood to meet the body's needs.

Heart valve problem is the problem that occurs in one or more number of the heart valves. It occurs if heart valves do not open enough to allow the blood to flow through or do not close properly and allow blood to leak through.

Rheumatic heart disease is a major problem in many poor countries. This disease begins with a bacterial infection in childhood, affecting joints and heart
valves. Congenital heart disease is a problem with the structure of the heart arising because of a birth defect.

1.4 MYOCARDIAL INFARCTION

A Myocardial Infarction (MI), also known as heart attack, is the damage and death of heart muscle from the sudden blockage of a coronary artery by a blood clot. Coronary arteries are blood vessels that supply the heart muscle with blood and oxygen. The dead heart muscle is eventually replaced by scar tissue. The coronary arteries provide the heart with this critical blood supply. If there exists coronary artery disease, those arteries become narrow and blood cannot flow as well as they should.

Fatty matter, calcium, proteins, and inflammatory cells in the arteries build up plaques of different sizes within the artery. The plaque deposits are hard on the outside and soft and mushy on the inside. When the plaque is hard, the outer shell cracks, platelets come to the area and blood clots form around the plaque. This slow process is known as atherosclerosis.

If a blood clot totally blocks the artery the heart muscle becomes starved for oxygen. When the heart muscle is starved for oxygen and nutrients, it is called ischaemia. When damage or death of any part of the heart muscle occurs as a result of ischaemia, it is called a heart attack or myocardial infarction. Figure 1.1 shows the schematic view of Myocardial Infarction. Ischaemia often causes chest pain or discomfort, known as angina pectoris.
Each coronary artery supplies blood to a region of heart muscle. The amount of damage to the heart muscle depends on the size of the area supplied by the blocked artery and the time between injury and treatment. Healing of the heart muscle begins soon after a heart attack and takes about eight weeks. The most common symptoms of a heart attack are chest pressure or pain, shortness of breath, pain or discomfort in the arms or shoulder, pain or discomfort in the jaw, neck or back, feeling weak, lightheaded or nausea.

There is a weak relationship between the severity of pain and the degree of oxygen deprivation in the heart muscle. The heart attack can be occurring with severe pain or without pain. In some cases, angina can be extremely serious, and has been known to cause death.

Knowing the early warning signs of heart attack is critical for prompt recognition and treatment. A person experiencing a heart attack may not even be sure of what is happening. Heart attack symptoms vary among individuals and even a person who has had a previous heart attack may have different symptoms in a...
subsequent heart attack. Although chest pain or pressure is the most common symptom of a heart attack, heart attack victims may experience a diversity of symptoms. Hence early prediction of MI plays a major role in the diagnoses of the disease.

1.5 PROBLEM DESCRIPTION

CVD being the number one killer of the modern era, a faster solution is of crucial importance for diagnosis and treatment of CVD, as delay of every second counts toward patient’s mortality (Blount et al. 2007, Luca et al. 2004). CVD is the heart abnormalities which mean malfunctioning and dysfunction of heart in which it eventually leads to less oxygen supply to all vital parts of the body. It affects oxygen transportation to brain, lungs, and heart itself, and supply of less oxygen leads to more problems in the body (WHO 2008).

Most of the other cardiovascular diseases and coronary heart diseases are caused by the progression of atherosclerosis. One of the progressions of atherosclerosis is myocardial, and this condition is caused by the lack of oxygen and nutrients to the contractile cells (Pham, 2008).

Today, medical services have come a long way to treat patients with various diseases. Among the most lethal one is the heart disease problem which cannot be seen with a naked eye and comes instantly when its limitations are reached. Today, diagnosing patients correctly and administering effective treatments have become quite a challenge. Poor clinical decisions may lead to patients’ death which cannot be afforded by the hospital as it loses its reputation.

The cost to treat a patient with a heart problem is quite high and not affordable by every patient. To achieve a correct and cost-effective treatment
computer-aided decision support systems can be developed to do the task. Most hospitals today use some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images.

The diagnosis of diseases is a vital and intricate job in medicine. The recognition of MI from diverse features or signs is a multilayered problem that is not free from false assumptions and is frequently accompanied by impulsive effects. Thus, the attempt to exploit knowledge and experience of several specialists and clinical screening data of patients composed in databases to assist the diagnosis procedure is regarded as a valuable option. Hence it is necessary to develop CAD system for diagnosis of MI using intelligent classifiers.

1.6 MOTIVATION AND OBJECTIVES

A total of 57 million deaths occurred in the world during 2008. 36 million (63%) were due to non-controllable disease principally cardiovascular diseases, diabetes, cancer and chronic respiratory diseases. The leading causes of deaths in 2008 were cardiovascular diseases 17 million deaths or 48% of non-controllable disease(Alwan et al. 2010). Different statistical surveys indicate that heart disease, especially ischaemic heart disease, has become a major health burden also in India (Ashley et al. 2001).

Therefore, advanced research is needed as a preventive measure against this silent killer. Cardiovascular diseases are currently the leading cause of death globally accounting for 21.9 percent of total deaths and are projected to increase to 26.3 percent by 2030 (WHO2008). Currently, the best practice for reducing human mortality rates caused by complex diseases is to detect their
symptoms at early stages. Through the early recognition of symptoms, one can get the most effective clinical treatment for the best outcome.

The cardiologist can diagnose the disease after the investigation of the MI test. The main symptom of the MI is angina, but in many people MI occurs without pain. An automated system can overcome these problems by reducing the number of false positive and false negative readings and increase the chance of detecting abnormalities early. This is a favorable prognosis for patients as incorrect or late detections often result in mortality. As with many labor-intensive occupations, cardiologists use CAD systems that can identify potential MI on heart disease data set.

These software systems are entering clinical practice as a way to improve cardiologist’s ability to detect few a cases of MI without chest pain. The studies indicate the importance of analyzing the problem and efforts made to improve the performance of MI detection in digital data. Researchers are responsible to conceive new and improved analytical tools to solve a problem. When a new tool is available, the problem should be reexamined to find better and more accurate solutions.

Classification systems can help in minimizing possible errors and also can provide instant examination of medical data in shorter time and in a more detailed manner. This research work concentrates on developing a CAD system as an artificial second cardiologist using MATLAB software. New artificial intelligent techniques such as neural network have been used in medical applications for detecting the normal and abnormal patient from data set. The patient with high risk of MI is classified as abnormal, and the others are classified as normal.

The thriving of artificial intelligence which utilizes the human experience in a more relaxed form than the conventional mathematical approach has recently
attracted more attention. The classification is done using supervised classifiers. Three most popular supervised classifiers used for classification are Feed Forward Neural Network (FFNN), Cascaded correlation Neural Network (CNN) and Support Vector Machine (SVM). Designing optimal neural network architecture is made by a human expert, and it requires a tedious trial and error process.

Automatic determination of artificial neural network parameters is the most critical task. Soft computing evolutionary approaches like genetic algorithms and particle swarm intelligence can be effectively used for tuning the intelligent networks (Park et al. 1994, Zhenya et al. 1998, Sing and Postlethwaite 1996). These techniques were used for optimizing the initial network parameters like hidden layer neurons, learning rate and momentum constant.

This research work focuses mainly on designing a CAD system based on the optimized neural network parameters evaluated using evolutionary approaches to improve the classification accuracy in MI detection thereby reducing the misclassification rate.

1.7 ORGANIZATION OF THE THESIS

In addition to the first introductory chapter, this thesis is composed of seven chapters. The data mining techniques in medicine, cause of myocardial infarction, problem description and the motivation of the research are explained in the first introductory chapter. Chapter 2 presents a detailed literature survey and the past works involved in the design, application and analysis of heart disease detection system using data mining and intelligent algorithm. It also describes the commonly used risk factors for heart MI detection.
Chapter 3 explores the proposed system model and the detailed model used for MI detection. The application of supervised classifiers for detecting the MI is analyzed for UCI machine learning, and real clinical database is presented in Chapter 4 which includes the intelligent algorithms like FFNN, CNN and SVM.

The artificial neural network needs tuning of its parameters for optimal performance. To achieve the optimality in artificial neural network, various evolutionary approaches like Genetic Algorithm and Particle Swarm Optimization are used and their performance is analyzed. The various parameters considered for optimization are learning rate, momentum factor and number of neurons in the hidden layer. These paradigms are presented in Chapter 5.

In Chapter 6, a concise review of work reported, contributions made and comparative performance results for MI detection is described. The major conclusions reached are presented with the recommendation for future enhancement in Chapter 7.

1.8 SUMMARY

The role of data mining in medicine, cardiovascular diseases, MI and the need for computer-aided MI detection system are briefly discussed in this chapter. The problem description and the motivation of the research are also presented in this chapter. The existing methods and techniques for MI detection are presented in the next chapter. Chapter 2 also describes the intelligent machine learning techniques for classification of cardiovascular diseases and the influencing risk factors considered in MI.