Analysis of myocardial infarction risk factors in heart disease data set.

Chitra R1*, Chenthil Jegan TM2, Ezhilarasu R3

1Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kanyakumari, Tamil Nadu, India
2Department of Mechanical Engineering, St. Xaviers Catholic College of Engineering, Kanyakumari, Tamil Nadu, India
3Department of Electrical and Electronics Engineering, Vel-Tech High Tech Engineering College, Thiruvallur, Kerala, India

Abstract

A major type of heart disease identified all over the world is Myocardial Infarction (MI), commonly called as heart attack. Sudden blockage in a coronary artery by a blood clot due to damage or death of heart muscle is called Myocardial Infarction (MI). Many researches have attempted to detect MI from identified risk factors using intelligent and data mining algorithms. The commonly used data set for MI detection is Cleveland data set. The purpose of the present study is to investigate the risk factors of Cleveland data set and their significant role in MI detection.

Keywords: Malfunctioning, Dysfunction of heart, Atherosclerosis, Cardiologist.

Accepted on August 03, 2017

Introduction

Heart Disease (HD) is a commonly prevailing killer disease in modern world. Diagnosis and treatment of heart disease are very important to reduce mortality rate [1,2]. Cardiovascular disease (CVD) is an abnormality such as malfunctioning and dysfunction of heart. It reduces the supply of oxygen to all major parts of the body and causes critical problems in the body. Most of the other cardiovascular diseases and coronary heart diseases are caused by the progression of atherosclerosis [3].

Myocardial infarction is the progressions of atherosclerosis, and it is triggered by the shortage of oxygen and nutrients to the contractile cells [3,4]. MI is the main health problem throughout the world. Atherosclerosis is caused because of the cholesterol and fatty substance deposits in arteries walls. As time passes, these fatty plaques grow in size and harden which may lead to a narrowing of the arteries like the scale formation on the inside of a metal water pipe.

Myocardial Infarction results in the death of the heart muscle because of blockage in the coronary artery. MI leads to heart attack and is a potentially lethal disease. MI is the result of lack of oxygen supply than demand. MI can be normally identified from patient’s history and ECG.

Nowadays, computer-aided detection systems are used to detect MI. It is the secondary tool to assist the cardiologist to predict MI. To develop computer-aided detection system using data mining and intelligent techniques, the proper risk factors should be identified. Most of the researchers use Cleveland data set for detection of MI using data mining and soft computing techniques. The aim of the study is to analyze the various risk factors of Cleveland data set and to identify their significance.

Related Works

During 2008, 48% of non-controllable CVD was caused [5]. Various surveys indicated that ischemic heart disease is the major health problem in India. Hence intensive research is required to prevent MI.

The role of ECG is also important to predict MI. If there is a considerable change in ECG, special attention is needed to treat the patients. [6]. Cardiologists generally diagnose heart disease based on three types of tests ECG information, patient symptoms and enzymatic test. Since enzymatic tests are expensive and time consuming, ECG and symptoms of the patient are used to decide on the disorder in cases of emergency.

According to the WHO reports, aging population and continuous prevalence of heart disease are the number one reason for 30% of the global total of all deaths. Existing healthcare systems are universally poor in terms of cost and quality [7]. Low and middle-income countries are mostly affected as 82% of CVD deaths take place in these countries occurring almost equally in men and women. CVD is presently the important root of death universally, and is estimated to reach 48.2 percent by 2030 [4].

Currently, the best practice for reducing mortality rates caused by complex diseases is to detect the symptoms at the early stages. Through this 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 MI is angina, but in many people MI occurs without pain. Computer aided system can overcome the problems and detect abnormalities in the early stage. It is clear that assessment of data taken from patients and decisions of medical experts are the most vital factors in diagnosis.

Framingham validated the MI risk factors [8] that have consequently been exposed to be universal through the world [9]. MI risk factors never occur in isolation, but they are correlated consequently been exposed to be universal through the world [9]. MI risk factors never occur in isolation, but they are correlated
Cholesterol, diet, and hypertension, left ventricular hypertrophy and smoking [12]. Angina pectoris is the common symptom of myocardial infarction [13].

Angina is the common symptom of heart disease in women than in men. In women after five years post onset, 4.8% died from heart disease [14,15]. The possibility of CVD is more with diabetes than without it. More significantly, sex and age add to the risk of MI in patients without diabetes [16-18].

HD is identified as the higher percentage (39%) non-controllable disease in people below 70. The level of HD risk for Asian Indian is more than 35 years for male and 45 for females. The score of risk can be calculated in a different manner for men and women. Angina is the chest pain due to impaired coronary flow [19].

In normal ECG results heart beats are regular. In abnormal ECG T waves are termed tall and if the height exceeds 1mV it is a sign of HD. Left ventricular hypertrophy is calculated by Romhiltestes points scoring system and it is the sign of HD. Left ventricular hypertrophy on the ECG is used in many CHD prediction algorithms and it is highly associated with hypertension. Hypertension associated with diabetes increases the mortality of MI patients [20]. Similarly, high cholesterol also increases the cause of MI and stroke [21].

Blood sugar also plays a significant role in MI diagnosis [22]. It is a common testing used to elicit cardiac abnormalities and the dynamic exercise includes treadmill test. Thallium test is used to produce nuclear cardiology imaging or otherwise called myocardial perfusion images. An imperfection existing on the initial stress images is termed as reversible defect and it is the severe sign of MI [23].

ECG signal is the key factor identified by the cardiac specialist for the diagnosis of MI [24]. The large QRS complex indicates the abnormality and the possibility of HD [25-27]. The P-wave may even be absent in some ECG recordings. In the ECG, the ST elevation also plays a vital role to detect the presence of ischaemia [28]. An approach based on ECG ST-T segment analysis has been proposed for intelligent MI detection [29]. ECG waves can be analyzed automatically without the influence of cardiologists [30]. Since MI is strongly reflected in the ST segments of ECGs [31], experiments on MI detection in this paper have been performed through segmentation and analysis of ST segments. Simple amplitude measurements are still the basis for the interpretation of the exercise ECG, e.g., the diagnostic criterion of 1mm flat ST depression was established in 1957 although more advanced processing of the exercise electrocardiogram has been attempted [32-34].

Cleveland Data Set Description

Heart attack data set is acquired from UCI (University of California, Irvine C.A). The Cleveland heart disease data was obtained from V.A. Medical Center, Long Beach and Cleveland Clinic Foundation from Dr. Robert Detrano. Totally, Cleveland dataset contains 17 attributes and 270 patients’ data. Among the 270 data, 150 refer to normal cases, i.e., without the risk of MI and the remaining 120 data with high risk, i.e., with the possibility of MI. In the published experiments, only 13 attributes are used for calculating the risk of MI. The data set contains the data in the age range between 25 and 75 and it also contains the data of women as well as men.

A study is made to analyze the abnormality of ischaemic heart disease patient. The data is observed from the patient who had heart attack and normal patient. The data obtained is distributed among various age groups and both genders. Ischemic heart disease has several symptoms. The 13 input attributes commonly considered in MI prediction system are discussed below.

Age

Age plays a major role in the prediction of MI. If age advances, the risk of damaged and narrowed arteries also increases. It weakens or thickens heart muscle that contributes to ischaemic heart disease and thus lead to MI. 44% of all deaths is caused only because of non-controllable diseases taking place before the age of 70. In this experimentation the age selected is in the range of 25-75.

Gender

Men are generally at greater risk of heart disease. However, the risk for a woman increases after menopause. The cause of disability and death in women is high after menopause. At age 75 the risk of MI is the same for both men and women. According to a survey after 50 years, 46% of women have high risk of MI. In digitized Cleveland dataset, 0 indicates male and 1 indicates female.

Chest pain

The main symptom of MI is angina which is commonly called chest pain. If the 9blood flows to the heart decreases, then the delivery of oxygen to heart muscle also decreases. Thus, there is a cause of discomfort, squeezing or painful feeling which is known as angina.

The byproduct lactic acid builds up in the heart muscle because of less efficiency of heart when chest pain occurs. Stable angina is felt as predictable sensation of chest pain. It is usually due to stenosis coronary artery producing transient MI. Angina is said to be unstable when the symptom pattern worsens abruptly in terms of frequency and duration without an obvious cause of increased oxygen consumption. The typical clinical representation of angina refers to be poor.

There are two types of angina. In type 1 angina, the site of pain is retrosternal and its duration is 1-5 minutes. Type 2 angina is the emotion induced angina which may be relieved with rest. Non-angina pain also causes sudden MI and its possibility is high in diabetic patients. Asymptomatic chest pain occurs at any time even at rest. In actual MI, the characteristics of the pain are in all categories, and if it lasts longer, it is more severe. In the Cleveland data set typical type 1 angina is encoded as ‘1’, typical type 2 angina as ‘2’, non-angina pain as ‘3’ and asymptomatic chest pain is encoded as ‘4’.

Resting blood pressure

The marginal blood pressure recorded is 120/80 mmHg. Here, 120 is the systolic blood pressure, which is the measure of
blood pressure when the heart beats. 120 is the diastolic blood pressure which is measured in the arteries between the heart beats. The range of pressure 140/90 mmHg is marked as high blood pressure (HBP). Below the age of 45, men have more percentage of HBP compared to women whereas the percentage of women and men affected by HBP is same in the age range 45-64 years. After 65 years, the HBP risk is more for women. The increased value of systolic blood pressure plays a major role in cardiovascular disease risk factor compared to diastolic blood pressure. People with systolic BP of 120–139 mmHg or diastolic BP of 80–89 mmHg are considered to be high risk persons of MI, who need regular health monitoring to avoid MI. Table 1 shows different levels of systolic and diastolic blood pressure. Hypertension refers to high BP. In the selected data set the BP is recorded from 90 to 190.

**Cholesterol**

The cholesterol is classified as Low density lipoprotein (LDL) cholesterol, High density lipoprotein (HDL) cholesterol, triglycerides and total cholesterol. LDL plays a key role in artery blockage. General level of LDL cholesterol is 100 mg/dL. High risk MI persons have the LDL level below 70 mg/dL. HDL is a reverse transport protein.

High planes of HDL cholesterol have less risk of MI. An HDL level of 60 mg/dL is maintained to achieve good protection against MI. The combined values of LDL, HDL and other lipids are called total cholesterol. The appropriate value of total cholesterol is less than 200 mg/dL. In Cleveland data set, serum cholesterol ranges from 160-410 mg/dL.

**Blood sugar**

Diabetes mellitus (increased blood sugar) is well-defined as fasting blood glucose of greater than 125 mg/dL. Diabetes increases the risk of MI. The role of increased diabetes in MI is the same as that of other risk factors high cholesterol level and high blood pressure. If diabetes is not controlled initially, it will cause damage to blood vessel and causes severe risk of MI. Hypertension increases twice in diabetes persons than the people with desirable glucose value. People with diabetes are severely affected by MI and their condition is worse. In people affected with diabetes, MI occurs with fewer symptoms or sometimes without symptoms. It is otherwise known as silent attack. Further, diabetes stops out the defensive effects estrogen, and thus premenopausal women with diabetes are identified as the high MI risk persons. In the selected data set, 120 mg/dL is considered as threshold value, fasting blood sugar >120 mg/dL is encoded as ‘1’ and fasting blood sugar <120 mg/dL is encoded as ‘0’.

**Resting electrocardiographic results**

The contraction and relaxation of heart is represented in the form of bioelectrical activity which is termed as ECG.

The ECG signal is a representation of the bioelectrical activity of the heart representing the cyclical contractions and relaxations of the human heart muscles. In the Figure 1 the P wave in ECG is obtained because of the depolarization of the atria. The large QRS complex is due to the depolarization of the ventricles. This is the complex with largest amplitude and is easy to detect. In an ECG, the time between P wave (atrial depolarization) and the QRS complex (ventricular depolarization) has normal value as 0.12-0.20 seconds. A distinct variation in PR interval indicates the heart disease. The time interval between QRS complex and the T wave is known as QT interval. QT interval prolongation is the severe risk factor of MI. Typical value for QRS complex duration will be less than 0.10 seconds. The part of the ECG from the end of the QRS wave to the next ECG is termed as the ST segment.

For people without heart disease, the ST segment is neither elevated nor depressed. Any change in ST from baseline may indicate cardiac disease such as ischaemia. The normal ECG is encoded as ‘0’, ECG with abnormal ST-T wave is encoded as ‘1’ and probable or definite left ventricular hypertrophy is encoded as ‘2’. The ECG for level ‘0’, level ‘1’ and level ‘2’ are shown in Figures 2-4 respectively.

<table>
<thead>
<tr>
<th>BP Classification</th>
<th>Systolic BP mmHg</th>
<th>Diastolic BP mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>More than 120</td>
<td>and</td>
</tr>
<tr>
<td></td>
<td>and</td>
<td>Less than 80</td>
</tr>
<tr>
<td>Prehypertension</td>
<td>120–139</td>
<td>or</td>
</tr>
<tr>
<td>Stage 1 HBP</td>
<td>140–159</td>
<td>or</td>
</tr>
<tr>
<td>Stage 2 HBP</td>
<td>More than 160</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>More than 100</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Levels of blood pressure

![Figure 1. ECG wave with its intervals and segments](image-url)
Figure 2. Normal ECG waveform

Figure 3. ECG with ST-T wave abnormality

Figure 4. ECG with left ventricular hypertrophy
**Maximum heart rate**

An arrhythmia is an irregular heart rhythm. Arrhythmia can occur with a normal heart rate or with fast or slow heart rates that may include coronary artery disease, heart attack, blood imbalances, and more. There are many types of arrhythmias including atrial fibrillation and atrial flutter. A normal heart rate is 50 to 100 beats per minute. Arrhythmias can occur with a normal heart rate or with heart rates that are slow or rapid. The sinus rate increases progressively with exercise. Abnormal heart rate reflects abnormal tone and it is associated with higher risk. Heart rate is recorded between 71-202 in the selected Cleveland data set.

**Exercise induced angina**

The exercise stress test called thread mill is the most common method used to diagnose the patients with suspected MI. Compared to ECG this method is not affected by low sensitivity. Stress test includes ECG measurements and non-ECG measurements. The factors considered in the proposed heart data set are Maximum heart rate, Exercise induced angina, ST depression and slope of ST segment. In Cleveland data set exercise induced angina is represented as ‘1’ and ‘0’ represents the absence of angina during thread mill test.

**ST depression induced by exercise relative to rest**

The three type of modalities are dynamic, static and resistive. Exercise induced ST segment depression does not localize site of MI. The magnitude of ST depression at peak exercise also indicates the existence of severe disease. Major ST depression at low workload indicates severe disease. ST segment depression during recovery (10%) is a reliable indicator in predicting coronary artery disease. The depth of ST depression due to ischemia is influenced by R wave height.

**Slope of the ST segment**

A new paradigm emerged with the result of thrombolysis is in MI benefitted patients with ST elevation. Slowly up sloping ST segment depression (>1.5 mm, 80msec from the J point) usually indicates MI. Horizontal ST segment depression is considerable abnormal response. Down sloping ST segment depression represents severe MI. In the selected data sets value 1 represents up sloping, value 2 represents horizontal ST wave and value 3 represents down sloping ST. But ST segment elevation only is not a perfectly sensitive marker for MI. The various slopes of ST segment are shown in Figure 5.

**Angiogram results**

Angiography is a medical imaging method used to visualize the organs of the body like arteries, veins and the heart chambers. In this method, a radio opaque contrast agent is injected into the blood vessel and then imaging using fluoroscopy.

Angiography helps to explore usual and pathological conditions of the vessel structure mostly luminal tapering and obstruction or aneurysmal broadening. From the obtained coronary angiogram, the number of major vessels (0-3) colored by fluoroscopy can be measured. It is used to examine the entire coronary anatomy including bypass graft increased in number that indicates the severe MI. Angiography is made under local anesthesia condition.

**Thallium test**

A thallium stress test is a nuclear imaging test that shows how well blood flows into the heart during exercise and at rest. A radioisotope is administered intravenously. It settles into the heart muscle and pinpoints spots that are abnormal. A heart muscle or coronary arteries damaged by previous MI can be identified by thallium heart scan. In this test, thallium is carried to the bloodstream of the entire body and the perfusion defects are identified. People having perfusion defects during exercise have the high risk of MI. In the HD data set, the normal value obtained in thallium stress test is coded as ‘3’, fixed defect is encoded as ‘6’ and reversible defect is encoded as ‘7’. Sample set of data attributes from Cleveland data set used for MI prediction is shown in Table 2.
Table 2. Cleveland data set sample attributes.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Gender</th>
<th>Chest pain</th>
<th>Resting BP (mmHg/and/or)</th>
<th>Cholesterol (mg/dL)</th>
<th>Blood sugar (mg/dL)</th>
<th>ECG</th>
<th>Maximum heart rate (beats per minute)</th>
<th>Exercise induced angina</th>
<th>ST segment depression (mV)</th>
<th>Slope of ST segment</th>
<th>Angiogram results</th>
<th>Thallium Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>60.0</td>
<td>1.0</td>
<td>4.0</td>
<td>130.0</td>
<td>206.0</td>
<td>0.0</td>
<td>2.0</td>
<td>132.0</td>
<td>1.0</td>
<td>2.4</td>
<td>2.0</td>
<td>2.0</td>
<td>7.0</td>
</tr>
<tr>
<td>56.0</td>
<td>1.0</td>
<td>1.0</td>
<td>120.0</td>
<td>193.0</td>
<td>0.0</td>
<td>2.0</td>
<td>162.0</td>
<td>0.0</td>
<td>1.9</td>
<td>2.0</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>71.0</td>
<td>0.0</td>
<td>4.0</td>
<td>112.0</td>
<td>149.0</td>
<td>0.0</td>
<td>0.0</td>
<td>125.0</td>
<td>0.0</td>
<td>1.6</td>
<td>2.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>58.0</td>
<td>0.0</td>
<td>1.0</td>
<td>150.0</td>
<td>283.0</td>
<td>1.0</td>
<td>2.0</td>
<td>162.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>35.0</td>
<td>1.0</td>
<td>4.0</td>
<td>126.0</td>
<td>282.0</td>
<td>0.0</td>
<td>2.0</td>
<td>156.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>55.0</td>
<td>0.0</td>
<td>4.0</td>
<td>180.0</td>
<td>327.0</td>
<td>0.0</td>
<td>1.0</td>
<td>117.0</td>
<td>1.0</td>
<td>3.4</td>
<td>2.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>48.0</td>
<td>1.0</td>
<td>4.0</td>
<td>130.0</td>
<td>256.0</td>
<td>1.0</td>
<td>2.0</td>
<td>150.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
<td>7.0</td>
</tr>
<tr>
<td>44.0</td>
<td>1.0</td>
<td>4.0</td>
<td>110.0</td>
<td>197.0</td>
<td>0.0</td>
<td>2.0</td>
<td>177.0</td>
<td>0.0</td>
<td>2.5</td>
<td>1.0</td>
<td>1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>63.0</td>
<td>1.0</td>
<td>1.0</td>
<td>145.0</td>
<td>233.0</td>
<td>1.0</td>
<td>2.0</td>
<td>150.0</td>
<td>0.0</td>
<td>3.0</td>
<td>1.0</td>
<td>0.0</td>
<td>6.0</td>
</tr>
<tr>
<td>54.0</td>
<td>0.0</td>
<td>3.0</td>
<td>350.0</td>
<td>304.0</td>
<td>1.0</td>
<td>0.0</td>
<td>170.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>41.0</td>
<td>0.0</td>
<td>2.0</td>
<td>126.0</td>
<td>306.0</td>
<td>0.0</td>
<td>0.0</td>
<td>163.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Conclusion

The various MI risk factors in Cleveland data set and their role in MI prediction are discussed in this paper. In data mining-based heart disease prediction algorithm, the patient data set is used to identify the risk. From the analysis, it is concluded that the major risk factors of MI are presented in the Cleveland data set, and hence it can be used as a standard database for MI prediction. Further, feature extraction techniques can be used to predict more relevant risk factors.

References


*Correspondence to:
Chitra R
Associate Professor, Department of Computer Science and Engineering
Noorul Islam Centre for Higher Education
Kanyakumari
Tamil Nadu
India
Tel no: + (91) 4651-250462
E-mail: jesi_chit@yahoo.co.in