Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159

http://www.ripublication.com

Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review

[1]Animesh Hazra, [2]Subrata Kumar Mandal, [3]Amit Gupta, [4]Arkomita Mukherjee and [5]Asmita Mukherjee

[1,2,3] Assistant Professor, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India.

[4,5] Student, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India.

Abstract

A popular saying goes that we are living in an "information age". Terabytes of data are produced every day. Data mining is the process which turns a collection of data into knowledge. The health care industry generates a huge amount of data daily. However, most of it is not effectively used. Efficient tools to extract knowledge from these databases for clinical detection of diseases or other purposes are not much prevalent. The aim of this paper is to summarize some of the current research on predicting heart diseases using data mining techniques, analyse the various combinations of mining algorithms used and conclude which technique(s) are effective and efficient. Also, some future directions on prediction systems have been addressed.

Keywords: Heart Diseases; Machine Learning; Data Mining; Clustering; Classification.

INTRODUCTION

The heart is one of the main organs of the human body. It pumps blood trough the blood vessels of the circulatory system. The circulatory system is extremely important because it transports blood, oxygen and other materials to the different organs of the

body. Heart plays the most crucial role in circulatory system. If the heart does not function properly then it will lead to serious health conditions including death.

1 . T*ypes of Cardiovascular Diseases*

Heart diseases or cardiovascular diseases (CVD) are a class of diseases that involve the heart and blood vessels. Cardiovascular disease includes coronary artery diseases (CAD) like angina and myocardial infarction (commonly known as a heart attack).There is another heart disease, called coronary heart disease(CHD), in which a waxy substance called plaque develops inside the coronary arteries. These are the arteries which supply oxygen-rich blood to heart muscle. When plaque begins to build up in these arteries, the condition is called atherosclerosis. The development of plaque occurs over many years. With the passage of time, this plaque can harden or rupture (break open). Hardened plaque eventually narrows the coronary arteries which in turn reduces the flow of oxygen-rich blood to the heart. If this plaque ruptures, a blood clot can form on its surface. A large blood clot can most of the time completely block blood flow through a coronary artery. Over time, the ruptured plaque also hardens and narrows the coronary arteries. If the stopped blood flow isn't restored quickly, the section of heart muscle begins to die. Without quick treatment, a heart attack can lead to serious health problems and even death. Heart attack is a common cause of death worldwide. Some of the common symptoms of heart attack [2] are as follows.

1.1. *Chest pain*

It is the most common symptom of heart attack. If someone has a blocked artery or is having a heart attack, he may feel pain, tightness or pressure in the chest.

1.2. *Nausea,Indigestion, Heartburn and Stomach Pain*

These are some of the often overlooked symptoms of heart attack. Women tend to show these symptoms more than men.

1.3. *Pain in the Arms*

The pain often starts in the chest and then moves towards the arms, especially in the left side.

1.4. *Feeling Dizzy and Light Headed*

Things that lead to the loss of balance.

1.5. *Fatigue*

Simple chores which begin to set a feeling of tiredness should not be ignored.

1.6. *Sweating*

Some other cardiovascular diseases which are quite common are stroke, heart failure, hypertensive heart disease, rheumatic heart disease, Cardiomyopathy, Cardiacarrhythmia, Congenital heart disease, Valvular heart disease, Aortic aneurysms, Peripheral artery disease and Venous thrombosis. Heart diseases may

develop due to certain abnormalities in the functioning of the circulatory system or may be aggravated by certain lifestyle choices like smoking, certain eating habits, sedentary life and others. If the heart diseases are detected earlier then it can be treated properly and kept under control. Here, early detection is the main key. Being well informed about the whys and wherefores of heart disease will help in prevention summarily.

Prevalence of Cardiovascular Diseases

An estimated 17.5 million deaths occur due to cardiovascular diseases worldwide. More than 75% deaths due to cardiovascular diseases occur in the middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack [3]. India too has a growing number of CVD patients added every year. Currently, the number of heart disease patients in India is more than 30 million. Over two lakh open heart surgeries are performed in India each year. A matter of growing concern is that the number of patients requiring coronary interventions has been rising at 20% to 30% for the past few years [4].

The rest of the paper is organized as follows. Section 2 describes some of the well known data mining algorithms used for heart disease prediction. Section 3 describes some of the popular data mining tools used for the data analysis purpose. Section 4 summarizes the methodologies and results of previous research on heart disease diagnosis and prediction. Section 5 discusses the pros and cons on literature survey. Finally, Section 6 concludes the paper along with future scope.

DATA MINING ALGORITHMS

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Decision tree, Naïve Bayes, k-means, artificial neural network etc. They are discussed in the follows section.

1. *Decision Tree*

A Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences including chance event outcomes and utility. It is one of the ways to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis to help and identify a strategy that will most likely reach the goal. It is also a popular tool in machine learning. A Decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one. Finally by following these rules, appropriate conclusions can be reached

2. ***C4.5***

It is a classifier in the form of a Decision tree. It is a supervised learning method which uses information gain and pruning for improved results. It is quite fast, popular and the output is easily interpretable.

3. ***K-means Algorithm***

K-means creates k groups from a set of given objects so that the members of a group are more similar. Other than specifying the number of clusters, k-means also "learns" the clusters on its own without any information about which cluster a particular observation should belong to. That's why k-means can be called as semi-supervised learning method. K-means is specially effective over large datasets.

4. ***ID3 Algorithm***

The ID3 algorithm (Quinlan86) is a Decision tree building algorithm which determines the classification of objects by testing the values of the properties. It builds the tree in a top down fashion, starting from a set of objects and the specification of properties. At each node of the tree, a property is tested and the results used to partition the object at that point are set. This process is recursively continued till the set in a given sub tree is homogeneous with respect to the classification criteria. Then it becomes a leaf node. At each node, information gain is maximized and entropy is minimized. In simpler words, that property is tested which divides the candidate set in the most homogeneous subsets.

5. ***Support Vector Machine(SVM)***

It is a supervised learning method which classifies data into two classes over a hyper plane. Support vector machine performs a similar task like C4.5 except that it doesn't use Decision trees at all. Support vector machine attempts to maximize the margin (distance between the hyper plane and the two closest data points from each respective class) to decrease any chance of misclassification. Some popular implementations of support vector machine are scikit-learn, MATLAB and of LIBSVM.

6. ***Naive Bayes(NB)***

It is a simple technique for constructing classifiers. It is a probabilistic classifier based on Bayes' theorem. All Naive Bayes classifiers assume that the value of any particular feature is independent of the value of any other feature, given the class variable. Bayes theorem is given as follows: $P(C|X) = P(X|C) * P(C)/P(X)$, where X is the data tuple and C is the class such that P(X) is constant for all classes. Though it assumes an unrealistic condition that attribute values are conditionally independent, it performs surprisingly well on large datasets where this condition is assumed and holds.

7. *Artificial Neural Network (ANN)*

An artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. Information which flows through the network affects the structure of the artificial neural network because a neural network changes or learns in a sense-based on input and output, for that particular stage and consequently for each stage. ANN's are considered nonlinear statistical data modelling tools where the complex relationships between inputs and outputs are modelled or patterns are found .ANNs have layers that are interconnected. Artificial neural networks are fairly simple mathematical models to enhance existing data analysis technologies.

8. *CART*

CART stands for Classification and Regression Trees methodology. In classification trees the target variable is categorical and the tree is used to identify the "class" within which a target variable would likely fall into.In regression trees, the target variable is continuous and a tree is used to predict its value. The CART algorithm is structured as a sequence of questions, the answers to which determine what will be the next question if there should be any questions. The result of these questions look like a tree structure where the ends are terminal nodes which represent that there are no more queries.

9. *Random Forest*

Random Forests are an ensemble learning method (also thought of as a form of nearest neighbor predictor) for classification and regression techniques. It constructs a number of Decision trees at training time and outputs the class that is the mode of the classes output by individual trees. It also tries to minimize the problems of high variance and high bias by averaging to find a natural balance between the two extremes.. Both R and Python have robust packages to implement this algorithm.

10. *Regression*

Regression is a statistical concept which is used to determine the weight of relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables). Two basic types of regression are linear regression and multiple linear regression. Also, there are several non-linear regression methods that are used for more complicated data analysis.

11. *J48*

J48 is a Decision tree that is an implementation of ID3 (Iterative Dichtomiser 3) developed by the WEKA project team. R language also has a package to implement this. J48 does not require discretization of numeric attributes.

12. ***A-Priori Algorithms***

It is an algorithm for frequent item set mining and association rule learning. A-priori uses breadth-first search algorithm and a hash structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k-1 .Then it prunes the candidates which have an infrequent sub pattern.

13. ***Fuzzy Logic***

It is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1. Fuzzy logic is applicable in many fields from control theory to artificial intelligence. Fuzzy logic is mainly employed to handle the concept of partial truth where the truth value may range between completely true and completely false. Among various combinations of methodologies in soft computing, fuzzy logic and neuro computing are very practical and popular techniques leading to development of neuro-fuzzy systems.

14. ***Association Rules***

Association rules are basically if/then statements which help us to find out the relationships between apparently unrelated data in an information warehouse. It has two parts, an antecedent (if) and a consequent (then).Association rules are created by analyzing a data set for frequent if/then patterns. Using the criteria support and confidence, it identifies the most important relationships. Support indicates that how frequently the items appear in the database while confidence shows the number of times the if/then statements have been found to be true. In data mining, association rules are very useful for analyzing and predicting customer behaviour. Programmers use association rules to build programs capable of machine learning.

DATA MINING TOOLS

Data mining tools provide ready to use implementation of the mining algorithms. Most of them are free open source software's so that researchers can easily use them. They have an easy to use interface. Some of the popular data mining tools are WEKA, RapidMiner, TANAGRA, MATLAB etc. Some of them are discussed as follows.

1. ***WEKA***

It stands for Waikato Environment for Knowledge Learning. It is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data. WEKA supports different standard data mining tasks such as data pre-processing, classification, clustering, regression, visualization and feature selection. The basic premise of this application is to utilize a

computer application that can be trained to perform machine learning capabilities and derive useful information in the form of trends and patterns. Originally written in C, the WEKA application was then completely rewritten in Java and is now compatible with almost every computing platform. Its user friendly graphical interface allows for quick set up and operation.

2. *RapidMiner*

Formerly called as YALE (Yet Another Learning Environment), is an environment for providing data mining and machine learning procedures including data loading and transformation (ETL), data preprocessing and visualization, modeling, evaluation and deployment. Rapid Miner is written in the Java programming language. Also, it can be used for text mining, multimedia mining, feature engineering, data stream mining etc.

3. *TANAGRA*

It is a free data mining software designed for academic and research purposes. It proposes several data mining methods such as exploratory data analysis, statistical learning and machine learning. TANAGRA comprises some paradigms and algorithms such as clustering, association rule, parametric and nonparametric statistics, factorial analysis, feature selection and construction algorithms.

4. *Apache Mahout*

It is a project of the Apache Software Foundation designed for free implementations of distributed or otherwise scalable machine learning algorithms that focus primarily in the areas of collaborative filtering, clustering and classification. Apache Hadoop is another open source, Java-based programming framework which supports the processing and storage of extremely large data sets in a distributed computing environment. It is a part of the Apache project which is sponsored by the Apache Software Foundation.

5. *MATLAB*

It is the short form for matrix laboratory. It supports a multi-paradigm numerical computing environment. It is a fourth-generation programming language. MATLAB provides matrix manipulations, plotting of functions and data, algorithm implementations, creation of user interfaces and interfacing with programs written in other languages including C, C++,C#, Java, Fortran and Python [41].

6. ***Java***

Java is a high level programming language developed by Sun Microsystems and now owned by Oracle Inc. It is widely used for developing and delivering content on the web. Java has numerous object oriented programming features much like C++, but is simplified to eliminate language features that cause common programming errors. Java language is well suited for use on the World Wide Web. Java applets (small Java applications) can be downloaded from a web server and run on a computer by a Java-compatible web browser.

7. ***C***

C was developed by Dennis M. Ritchie at Bell Labs for the Unix Operating System in the early 1970s.It was originally intended for writing system softwares. C is a high-level, general-purpose programming language which is ideal for developing firmware and portable applications.

8. ***Orange***

It is a toolkit for data visualisation, machine learning and data mining. It is interactive and can be used as a Python library.

LITERATURE SURVEY

There are thirty five research papers that explore the computational methods to predict heart diseases. The summaries of them have been presented in a nutshell.

Shaikh Abdul Hannan et al. [5] used a Radial Basis Function(RBF) to predict the medical prescription for heart disease. About 300 patient's data were collected from the Sahara Hospital, Aurangabad. RBFNN (Radial Basis Function–Neural Network) can be described as a three-layer feed forward structure. The three layers are the input layer, hidden layer and output layer. The hidden layer consists of a number of RBF units (n_h) and bias (b_k). Each neuron on the hidden layer uses a radial basis function as a nonlinear transfer function to operate on the input data. The most often used RBF is usually a Gaussian function. Designing a RBFNN involves selecting centres, number of hidden layer units, width and weights. The various ways of selecting the centres are random subset selection, k-means clustering and others. The methodology was applied in MATLAB. Obtained results show that radial basis function can be successfully used (with an accuracy of 90 to 97%) for prescribing the medicines for heart disease**.**

AH Chen et al. [6] presented a heart disease prediction system that can aid doctors in predicting heart disease status based on the clinical data of patients. Thirteen important clinical features such as age, sex, chest pain type were selected. An artificial neural network algorithm was used for classifying heart disease based on

these clinical features. Data was collected from machine learning repository of UCI .The artificial neural network model contained three layers i.e. the input layer, the hidden layer and the output layer having 13 neurons, 6 neurons and 2 neurons respectively. Learning Vector Quantization (LVQ) was used in this study. LVQ is a special case of an artificial neural network that applies a prototype-based supervised classification algorithm. C programming language was used as a tool to implement heart disease classification and prediction trained via artificial neural network.The system was developed in C and C# environment.The accuracy of the proposed method for prediction is near to 80*%.*

Mrudula Gudadhe et al.[7] presented a decision support system for heart disease classification. Support vector machine (SVM) and artificial neural network (ANN) were the two main methods used in this system. A multilayer perceptron neural network (MLPNN) with three layers was employed to develop a decision support system for the diagnosis of heart disease. This multilayer perceptron neural network was trained by back-propagation algorithm which is computationally an efficient method. Results showed that a MLPNN with back-propagation technique can be successfully used for diagnosing heart disease.

Manpreet Singh et al. [8] proposed a heart disease prediction system based on Structural Equation Modelling (SEM) and Fuzzy Cognitive Map (FCM).They used Canadian Community Health Survey (CCHS) 2012 dataset. Here, twenty significant attributes were used. SEM is used to generate the weight matrix for the FCM model which then predicts a possibility of cardiovascular diseases. A SEM model is defined with correlation between CCC 121(a variable which defines whether the respondent has heart disease) along with 20 attributes. To construct FCM a weight matrix representing the strength of the causal relationship between concepts must be constructed first. The SEM defined in the previous section is now used as the FCM though they have achieved the required ingredients (i.e. weight matrix, concepts and causality).80% of the data set was used for training the SEM model and the remaining 20% for testing the FCM model. The accuracy obtained by using this model was 74%.

Carlos Ordonez [9] has studied association rule mining with the train and test concept on a dataset for heart disease prediction. Association rule mining has a disadvantage that it produces extremely large number of rules most of which are medically irrelevant. Also in general, association rules are mined on the entire data set without validation on an independent sample. In order to solve this, the author has devised an algorithm that uses search constraints to reduce the number of rules. The algorithm then searches for association rules on a training set and finally validates them on an independent test set. The medical significance of discovered rules is then evaluated with support, confidence and lift. Search constraints and test set validation significantly reduce the number of association rules and produce a set of rules with high predictive accuracy. These rules represent valuable medical knowledge.

Prajakta Ghadge et al. [10] have worked on an intelligent heart attack prediction system using big data. Heart attack needs to be diagnosed timely and effectively because of its high prevalence. The objective of this research article is to find a prototype intelligent heart attack prediction system that uses big data and data mining modeling techniques. This system can extract hidden knowledge (patterns and relationships) associated with heart disease from a given historical heart disease database. This approach uses Hadoop which is an open-source software framework written in Java for distributed processing and storage of huge datasets. Apache Mahout produced by Apache Software Foundation provides free implementation of distributed or scalable machine learning algorithms. Record set with 13 attributes (age, sex, serum cholesterol, fasting blood sugar etc.) was obtained from the Cleveland Heart Database which is available on the web. The patterns were extracted using three techniques i.e. neural network, Naïve Bayes and Decision tree. The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks.

Asha Rajkumar et al. [11] worked on diagnosis of heart disease using classification based on supervised machine learning. Tanagra tool is used to classify the data, 10 fold cross validation is used to evaluate the data and the results are compared. Tanagra is a free data mining software for academic and research purposes. It suggests several data mining methods from explanatory data analysis, statistical learning, machine learning and database area. The dataset is divided into two parts, 80% data is used for training and 20% for testing. Among the three techniques, Naïve Bayes shows lower error ratio and takes the least amount of time. It is shown in Table 1.

Table 1: Classification accuracy and time complexity of Naïve Bayes, Decision list and k-NN algorithms [11].

Algorithm	Accuracy	Time taken(ms)
Naïve Bayes	52.33%	609
Decision list	52%	719
k-NN	45.67%	1000

From the above results, Naïve Bayes algorithm plays a key role in shaping improved classification of a dataset.

K. S. Kavitha et al. [12] modelled and designed an evolutionary neural network for heart disease detection. This research describes a new system for detection of heart diseases using feed forward neural architecture and genetic algorithm. The proposed system aims at providing easier, cost effective and reliable diagnosis for heart disease. The dataset is obtained from UCI repository. The weights of the nodes for the

artificial neural network with 13 input nodes, 2 hidden nodes and 1 output node are once set with gradient descent algorithm and then with genetic algorithm. The performances of these methods are compared and it is concluded that genetic algorithm can efficiently select the optimal set of weights. In genetic algorithm tournament selection is a method of selecting an individual from a population of individuals. This work finds that more members are coming from the offspring population. It is an indication for generation of fitter offsprings which leads to greater diversity and exploration of search space. With the help of this work, expert disease prediction systems can be developed in the future.

K. Sudhakar et al. [13] studied heart disease prediction using data mining. The data generated by the healthcare industry is huge and "information rich". As such, it cannot be interpreted manually. Data mining can be effectively used to predict diseases from these datasets. In this paper, different data mining techniques are analyzed on heart disease database. Classification techniques such as Decision tree, Naïve Bayes and neural network are applied here .Associative classification is a new and efficient technique which integrates association rule mining and classification to a model for prediction and achieves maximum accuracy. In conclusion, this paper analyzes and compares how different classification algorithms work on a heart disease database.

Shantakumar B. Patil et al. [14] obtained important patterns from heart disease database for heart attack prediction. Enormous amount of data collected by the healthcare industry is unfortunately not 'mined' properly to find concealed information that can predict heart attack. Here, the authors have proposed MAFIA algorithm (Maximal Frequent Itemset Algorithm) to do so using Java. The data is preprocessed first, and then clustered using k-means algorithm into two clusters and the cluster significant to heart attack is obtained. Then frequent patterns are mined from the item set and significance weightages of the frequent data are calculated. Based on these weightages of the attributes(ex- age, blood pressure, cholesterol and many others), patterns significant to heart attack are chosen. This pattern can be further used to develop heart attack prediction systems.

Sairabi H. Mujawar et al. [15] predicted heart disease using modified k-means and Naïve Bayes. Diagnosis of heart disease is a complex task and requires great skills. The dataset is obtained from Cleveland Heart Disease Database. The attribute "Disease" with a value '1' indicates the presence of heart disease and a value '0' indicates the absence of heart disease. Modified k-means works on both categorical and combinational data which we encounter here. Using two initial centroids we obtain two farthest clusters. It finally gives a suitable number of clusters. Naive Bayes's creates a model with predictive capabilities. This predictor defines the class to which a particular tuple should belong to. This predictor has 93 % accuracy in predicting a heart disease and 89% accuracy in cases where it detected that a patient doesn't have a heart disease.

S. Suganya et al. [16] predicted heart disease using fuzzy cart algorithm. Fuzziness was introduced in the measured data to remove the uncertainty in data. A membership function was thus incorporated. Minimum distance CART classifier was used which proved efficient with respect to other classifiers of parametric techniques. The heart disease dataset is initially segregated into attributes that increase heart disease risk. Then fuzzy membership function is applied to remove uncertainty and finally ID-3 algorithm is run recursively through the non-leaf branches until all the data have been classified. The proposed method is implemented in Java.

Ashwini Shetty A et al. [17] proposed different data mining approaches for predicting heart disease. Their research work analyses the neural network and genetic algorithm to predict heart diseases. The initial weight of the neural network is found using genetic algorithm which is the main advantage of this method. Here, the neural network uses 13 input layers, 10 hidden layers and 2 output layers. The inputs are the attribute layers (here 13 attributes are used namely age, resting heart rate, blood pressure, blood sugar and others). Levenberg-Marquardt back propagation algorithm is used for training and testing. Optimization Toolbox is used to implement this system. 'configure' function is used with neural network where each weight lies between -2 to 2.Fitness function that is being used in the genetic algorithm is the Mean Square Error (MSE). Genetic algorithm is used for adjustment of weights. Based on MSE, fitness function will be calculated for each chromosome. Once selection is done, crossover and mutation in genetic algorithm replaces the chromosome having lower adaption with the better values. Fitter strings are obtained by optimizing the solution which corresponds to interconnecting weights and threshold of neural network. The resulting lower values those are close to zero, represent the generalized format of the network which is ready for classification problem. The system calculates accuracy using MATLAB. Preprocessing is done using WEKA. The results show that the hybrid system of genetic algorithm and neural network works much better than the performance of neural network alone.

K Cinetha et al. [18] proposed a decision support system for precluding coronary heart disease using fuzzy logic. This system predicts the possibility of heart disease in a patient for the next ten years. Data from normal and coronary heart disease patients were collected and it was observed whether a normal person developed coronary heart disease or what factors could have led to the onset of coronary heart disease. Prevention of risk factors are analyzed using fuzzy logic and Decision tree.The dataset contains 1230 instances. Decision tree is implemented for the establishment of fuzzy rules and the diagnosis of coronary heart disease. The method is used to produce the clustered data. Next, the fuzzy rule is obtained by extracting rules from the cluster using the Least Square Error (LSE). Determination of the best cluster is selected using fuzzy technique and variant analysis is performed during testing. Smaller values of variant boundaries are ideal for clustering. The best accuracy of the system for selected rules when applied to the TSK inference orde-1method is 97.67%.

Indira S. Fall Dessai [19] proposed an efficient approach for heart disease prediction based on Probabilistic Neural Network (PNN) technique. The data set containing 13 medical attributes was obtained from the Cleveland Heart Disease Database. It is clustered using k-means**.** Probabilistic Neural Network is a class of radial basis function (RBF) network which is useful for automatic pattern recognition, nonlinear mapping and estimating probabilities of class membership and likelihood ratios. An evaluation of the existing algorithms such as decision tree, Naïve Bayes, BNN for prediction is compared with PBN. This is done using Receiver Operating Characteristic Convex Hull (ROCCH) method. Results show that the proposed system gives 94.6% correct predictions.

Mai Shouman et al. [20] worked on the application of k-Nearest-Neighbors (k-NN) in diagnosis of heart disease. This paper shows that k-NN has higher accuracy compared to neural network ensemble. However, applying integrating voting could not enhance the k-NN accuracy in the diagnosis of heart disease patients, unlike Decision tree classifiers where voting increases accuracy. Voting is an aggregation technique which is used to combine decisions of multiple classifiers. K-NN without voting gave the highest accuracy of 97.4%. However the accuracy for k-NN with voting reduced to 92.7%.

Serdar AYDIN et al. [21] have studied and compared various methods of data mining for diagnosing heart disease. Techniques used are Bagging, AdaBoostM1, Random Forest, Naive Bayes, RBF Network, IBK and NN. The data has been collected from Long Beach VA Hospital. It includes 200 samples, each containing 14 features. The techniques are analyzed using WEKA software. Results show that RBF Network has the accuracy of 88.20%,making it the most accurate classification technique in the diagnosis of heart disease.

G Purusothaman et al. [22] have surveyed and compared different classification techniques for heart disease prediction. Instead of applying a single model such as Decision tree, artificial neural network and Naïve Bayes, the authors focus on the working of hybrid models i.e. models which combines more than one classification technique. They have surveyed the works of researchers who studied about the effectiveness of hybrid models. The performances of single models such as Decision tree, artificial neural network and Naïve Bayes are 76%, 85% and 69% respectively. However, hybrid approaches show an accuracy of 96%. Therefore, hybrid models lead to reliable and promising classifiers for predicting heart diseases with good accuracy.

Deepali Chandna [23] has incorporated a hybrid approach to merge a learning algorithm and a feature selection technique. The dataset is obtained from UCI. Among the 76 attributes in the set, only 14 attributes are selected using k-nearest neighbor's algorithms. This approach also uses information gain and Adaptive Neuro-Fuzzy Inference System (ANFIS). ANFIS is the combined effect of neural network and

fuzzy inference system. Information gain is used for selection of quality of attributes. The accuracy for the proposed approach is 98.24%.

S. Pravabathi et al. [24] presented an overview of research being carried out using DNFS (Decision tree based Neural Fuzzy System). The data mining techniques were used to enhance the heart disease diagnosis and prediction which include Decision trees, Naive Bayes classifiers, k-nearest neighbour classification (k-NN), support vector machine (SVM) and artificial neural networks techniques. Genetic algorithm was applied to improve the learning of neuro-fuzzy system which combined the adaptability of fuzzy inputs with neural network for accurate prediction. C4.5 Decision tree algorithm and RIPPER (Repeated Incremental Pruning to Produce Error Reduction) were used for classification.C4.5 classifier performed better than other data mining techniques for diagnosis like support vector machine and neural networks. Naïve Bayes classifier is also a better option. They concluded that Decision trees and Naïve Bayes classifiers are prominent for cardiovascular disease diagnosis with an accuracy reaching more than 95%.

Jaymin Patel et al. [25] compared different algorithms of Decision tree classification for better performance in heart disease diagnosis using WEKA. J48 algorithm, logistic model tree and random forest algorithms were compared. Datasets were taken from UCI repository consisting of 303 instances and 76 attributes, out of which 13 attributes were chosen to perform the tests.J48 is an open source, reliable Java implementation of the C4.5 algorithm in the WEKA.It uses divide and conquer approach to construct the tree, and attributes at each node are chosen such that it can further classify the part into samples. But here the greatest disadvantage is size, which increases linearly with the examples. Logistic model tree is a Decision tree structure with logistic regression function at the leaves. The algorithm has the choice of overseeing parallel and multi-class target variables, numeric and nominal attributes along with missing qualities as well. However, Logistic Model Tree(LMT) take longer time to be produced. Random forest is an ensemble classifier consisting of many Decision trees. Individual trees represent the output of the classes. It constructs Decision trees with controlled variations. The Decision tree classification was performed under the framework of WEKA 3.6.10 and the results are shown in Table 2.

Table 2: Demonstration of train error and test error for J48, Logistic Model Tree (LMT) and Random Forest classifiers [25].

Error Type	**Algorithms**		
	J48	Logistic Model Tree	Random Forest
Train Error	0.1423221	0.1656716	0
Test Error	0.1666667	0.237931	0.2

The best algorithm is J48 with highest accuracy of 56.76% and the total time to build the model is 0.04 seconds whereas LMT algorithm has the lowest accuracy of 55.77% and the total time to build the model is 0.39seconds.

Vikas Chaurasia et al. [26] presented a new model that enhanced the Decision tree accuracy for identifying heart disease in patients. Decision tree algorithms here include CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and C4.5 build model. CART model recursively separates observations in the branches to construct a tree for the purpose of improving prediction accuracy. It builds classification and regression trees to predict continuous dependent variables (regression) and categorical predictor variables (classification).ID3 (IterativeDichotomized3) uses iterative inductive approach to identify the root at first and then construct the binary tree. Decision tables (DTs) are tabular representations to describe and analyse decision situations. In this study data is used from the Cleveland Clinic Foundation. Only 11 attributes were chosen from the 76 raw attributes. It was analysed and implemented in WEKA tool. CART provided the highest accuracy of 83.49% followed by DT and ID3.

Gunsai Pooja Dineshgar et al. [27] surveyed on the current techniques of knowledge discovery in databases using data mining techniques and built a prototype of intelligent heart disease prediction system that gave diagnosis of heart disease using historical heart database. The data mining clustering techniques like k-means and k-medoid algorithms are analysed to achieve global optimality in partitioned-based clustering. PAM(Partitioning Around Medoids) uses iterative optimization that combines relocation of points between perspective clusters with re-nominating the points as potential medoids and CLARA(Clustering LARge Applications) which used random search to generate neighbours by starting with an arbitrary node and randomly checking max neighbour neighbours which are the versions of k-medoid algorithm. The k-means algorithm partitions a set of n objects into k desired cluster. After analysing the previous works did not used k-medoid algorithm, the authors have proposed to incorporate this method to classify data sets for predicting heart disease in an efficient and cost effective manner.

Jyoti Soni et al. [28] evaluated that the Weighted Associative Classifier (WAC) performed well as compared to other already existing Associative Classifiers. They designed a GUI to accept the patient's test results and predicted the presence of heart disease using CAR rules generated by WAC in Java platform**.** Weighted ARM uses weighted support and confidence framework to find out association rule from data repository. The WAC has been proposed as a new technique to get the exact significant rule instead of being flooded with insignificant relation. Experimental results show that WAC outperforms other associative classifiers such as CBA, CMAR and CPAR in terms of average accuracy. Maximum accuracy achieved is 81.51% with a support value 25% and confidence of 80%.

Kamal Kant et al. [29] proposed a prototype of heart disease prediction using data mining techniques, namely Naïve Bayes. Naïve Bayes is a statistical classifier which assigns no dependency between the attributes. The posterior probability needs to be maximized for determining the class. Here, Naïve Bayes classifier also performs well. In statistical probability and real time expert system, Naïve Bayes appears to be the most effective model for disease prediction followed by neural network and Decision trees.

Sharan Monica L et al. [30] surveyed current techniques of knowledge discovery in databases using data mining techniques such as J48, NB Tree and simple CART to predict heart disease more accurately with reduced number of attributes in the WEKA tool. J48, is an open source Java implementation of the C4.5 which uses information gain to take decisions. Naive Bayes classifier creates models with predictive capabilities, preferably for continuous dataset. Classification and Regression Trees (CART) is used to display important data relationships very quickly. These three Decision tree algorithms were applied in WEKA. J48 was the quickest to be built(0.08 sec) whereas CART gave the highest accuracy of 92.2%.

Nidhi Bhatla et al. [31] analysed various data mining techniques introduced in recent years to predict heart diseases. The observations revealed that neural networks with 15 attributes outperformed all other data mining techniques and the Decision tree also showed good accuracy with the help of genetic algorithm and feature subset selection using WEKA 3.6.6. This research work was incorporated by two more attributes i.e. obesity and smoking for efficient diagnosis apart from the other common attributes. Genetic algorithm was applied which uses natural evolution methodology. It continues generation until it evolves a population P where every rule in P satisfied the fitness threshold, starting from null. Decision tree gave 99.62% accuracy by using 15 attributes. Moreover, in combination with genetic algorithm with 6 attributes, Decision tree showed 99.2% efficiency.

Sumitra Sangwanet al. [32] developed a hybrid algorithm which uses k-means and A-priori algorithm for mining large volumes of data and extracting useful information. At first, clustering is done using k-means clustering algorithm. Then A-priori algorithm is used to find the frequent item sets. It is also used to mine the frequent term sets for Boolean association rule. It applies a "bottom up" approach i.e. frequent subsets are enlarged one item at a time and groups of candidates are tested altogether against the data. Results show that clustering followed by A-priori yielded better performance to predict heart disease.

Rishi Dubey et al. [33] have studied the different data mining techniques for prediction of heart disease. Most of the papers which they studied show that hybrid techniques outperform a single classification technique in terms of accuracy. They have concluded that neural network is an efficient technique for prediction. When the system is trained properly along with genetic algorithms, the system shows very promising results. This method can also be used to select the proper treatment

methods for a patient in future, instead of just predicting the chances of developing a heart disease among the patients.

Ashish Chhabbi et al. [34] have studied different data mining techniques for extracting hidden patterns from a dataset that can answer complex queries in prediction of heart disease. The dataset has been collected from UCI repository. They have applied Naive Bayes and modified k-means algorithm. Results show that modified k-means give better accuracy than simple k-means(where number of clusters were predefined).

Shadab Adam Pattekari et al. [35] developed a prototype of Heart Disease Prediction System using Naive Bayes, Decision trees and neural networks. It is implemented in a web application. In this system, user answers some predefined questions. Then it retrieves hidden data from the stored database and compares the user's values with trained dataset. The system discovers and extracts hidden knowledge associated with heart diseases from a historical heart disease database. It can answer the complex queries for diagnosing a disease. A set of 15 attributes was selected and then Naive Bayes classification method was applied to find out the chances of heart disease.

Boshra Baharami et al. [36] have evaluated different classification techniques such as J48 Decision tree ,k-Nearest Neighbors(k-NN), Naive Bayes(NB) and SMO(SMO is widely used for training SVM). On the dataset feature selection technique (gain ratio evaluation technique) is used to extract the important features. WEKA software is used for implementing the classification algorithms. 10 fold cross-validation technique is used to test the mining techniques. J48 shows the highest accuracy of 83.732%.

Dhanashree S. Medhekar et al. [37] presented a classifier technique for the heart sickness prediction and likewise they've confirmed how Naïve Bayes can be used for the classification purpose .They categorized clinical knowledge to five distinct classes namely no, low ,normal, excessive, very excessive. If any unknown sample is discovered, the method will classify it into respective class label. The dataset used here is the Cleveland medical institution ground work coronary heart disease set which contains 303observations and 14 parameters. The system works in two phases i.e. coaching phase and testing phase. In the coaching segment, the classification is supervised .The checking out segment involves the prediction of the unknown knowledge or the lacking values. The Naïve Bayes algorithm is used which is based on the Bayesian theorem. The outcome proves that the accuracy has been obtained by altering the number of occasions within the given dataset.

Noura Ajam [38] has studied that artificial neural networks(ANN) show significant results in heart disease diagnosis. The architecture of a neural network is formed by the number of processing units (neurons) and connections between them. A subgroup of processing elements is called layer. The number of neurons and the layers depends upon the complexity of the system. Artificial neural network is widely used in

medical diagnosis and health care applications because of it's high predictive power as classifier, fault tolerance and learning from environment. Artificial neural network is unsupervised learning type provided only with inputs associated with unknown targets. It is self organized. The dataset used here is obtained from Cleveland dataset which consists of 14attributes and 303 instances. Artificial neural network is trained using back propagation learning algorithm on the data. Input and target samples are divided as 60% training set, 20% validation set and 20% test set. The activation function used is tangent sigmoid for hidden layers and linear transfer function for output layer. Mean square error (MSE) is calculated which is equal to 0.1071 and the classification accuracy for heart disease is 88%.

S. Florence et al. [39] proposed a system which uses neural network and the Decision tree (ID3) for the prediction of heart attacks. The dataset used is provided by the UCI machine learning repository. CART, ID3, C4.5 Decision tree algorithms used Giniindex to measure the impurity of a partition or set of training attributes. The dataset contains six attributes like age, sex, cardiac duration, signal, possibility of attack etc. The final outcome is the class label. Depending upon the attribute values present in the dataset, the corresponding class label is predicted. 75% of the data is used for training and 25% is used for testing the system. The knowledge obtained from the classification is used to test the system. In the neural network, the input layer has 6 nodes, the hidden layer has 3 nodes and the output layer consists of 2 nodes. Finally it shows 2 outputs, that is the possibility of heart attacks. The prediction is done using the tool called RapidMiner Studio. Results are generated by using Decision tree as well as neural networks. They have used this method to predict whether there is an attack or not.

DISCUSSION

This paper summarises some of the recent works done in data mining related to cardiovascular diseases. Data mining algorithms can be effectively used to 'mine' relevant information from the huge amounts of data generated by the healthcare industry. These works show that rather than applying a single mining technique on a data set, results are far better if a collection of mining techniques are used. Java is chosen in most of the research work for practical execution of the project. WEKA, Tanagra, Matlab etc. are some of the other popular tools used for data analysis. Careful selection of the combination of mining techniques and accurate implementation of those techniques on the data set yields a fast and effective implementation of a system for heart disease management. The required dataset is divided into two parts, one is used for mining and the smaller partition is used for verifying. Most of the time, 10 fold cross validation technique is used. Some of the works are about the comparison of different classification techniques on a dataset to correctly classify if a given patient has any probability of a heart disease or not. Others have worked on 'mining' the causes that lead to heart diseases from a given dataset. Commonly used classification techniques are Decision tree, Naïve Bayes, artificial neural network, association rule mining and fuzzy logic. Apart from

analysing these commonly used techniques, some of the recent works have studied about "hybrid models". The idea of a hybrid model is to incorporate several known classification and selection techniques in a single model to give better results. It is observed that hybrid models give very high accuracy if proper combinations of different algorithms are chosen.

CONCLUSION

Heart diseases when aggravated spiral way beyond control. Heart diseases are complicated and take away lots of lives every year .When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time. Sedentary lifestyle and excessive stress in today's world have worsened the situation. If the disease is detected early then it can be kept under control. However, it is always advisable to exercise daily and discard unhealthy habits at the earliest. Tobacco consumption and unhealthy diets increase the chances of stroke and heart diseases. Eating at least 5 helpings of fruits and vegetables a day is a good practice. For heart disease patients, it is advisable to restrict the intake of salt to one teaspoon per day.

One of the major drawbacks of these works is that the main focus has been on the application of classification techniques for heart disease prediction, rather than studying various data cleaning and pruning techniques that prepare and make a dataset suitable for mining. It has been observed that a properly cleaned and pruned dataset provides much better accuracy than an unclean one with missing values. Selection of suitable techniques for data cleaning along with proper classification algorithms will lead to the development of prediction systems that give enhanced accuracy.

In future an intelligent system may be developed that can lead to selection of proper treatment methods for a patient diagnosed with heart disease. A lot of work has been done already in making models that can predict whether a patient is likely to develop heart disease or not. There are several treatment methods for a patient once diagnosed with a particular form of heart disease. Data mining can be of very good help in deciding the line of treatment to be followed by extracting knowledge from such suitable databases.

REFERENCES

[1] https://en.wikipedia.org/wiki/Cardiovascular_disease.

[2] http://www.heart.org/HEARTORG/Conditions/HeartAttack/WarningSignsofaHeartAttack/Warning-Signs-of-a-HeartAttack_UCM_002039_Article.jsp#.WNpKgPl97IU.

[3] www.who.int/cardiovascular_diseases/en/.

[4] http://food.ndtv.com/health/world-heart-day-2015-heart-disease-in-india-is-a-growing-concern-ansari-1224160.

[5] Shaikh Abdul Hannan, A.V. Mane, R. R. Manza,and R. J. Ramteke, Dec 2010,"Prediction of Heart Disease Medical Prescription using Radial Basis Function",IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), DOI: 10.1109/ICCIC.2010.5705900 ,28-29 .

[6] AH Chen, SY Huang, PS Hong, CH Cheng, and EJ Lin,2011, "HDPS: Heart Disease Prediction System",Computing in Cardiology, ISSN: 0276-6574, pp.557- 560.

[7] Mrudula Gudadhe, Kapil Wankhade, and Snehlata Dongre, Sept 2010,"Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network",International Conference on Computer and Communication Technology (ICCCT),DOI:10.1109/ICCCT.2010.5640377, 17-19.

[8] Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K. Mago,2016, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map",IEEE International Conference on Fuzzy Systems (FUZZ),pp. 1377-1382.

[9] Carlos Ordonez,2006, "Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction", IEEE Transactions on Information Technology in Biomedicine (TITB), pp. 334-343, vol. 10, no. 2.

[10] Prajakta Ghadge,Vrushali Girme, Kajal Kokane, and Prajakta Deshmukh, 2016,"Intelligent Heart Attack Prediction System Using Big Data", International Journal of Recent Research in Mathematics Computer Science and Information Technology,Vol. 2, Issue 2, pp.73-77, October 2015–March.

[11] Asha Rajkumar, and Mrs G. Sophia Reena, 2010, "Diagnosis of Heart Disease using Data Mining Algorithms",Global Journal of Computer Science and Technology,Vol. 10,Issue 10, pp.38-43, September.

[12] K. S. Kavitha, K. V. Ramakrishnan, and Manoj Kumar Singh, September 2010,"Modelling and Design of Evolutionary Neural Network for Heart Disease Detection" ,International Journal of Computer Science Issues (IJCSI), Vol. 7, Issue 5, pp. 272-283.

[13] K.Sudhakar, and Dr. M. Manimekalai, January 2014, "Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 1157-1160.

[14] Shantakumar B. Patil,and Dr. Y. S. Kumaraswamy, February 2009, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction", IJCSNS International Journal of Computer Science and Network Security, Vol. 9 , No. 2,pp. 228-235.

[15] Sairabi H. Mujawar, and P. R. Devale, October 2015,"Prediction of Heart Disease using Modified k-means and by using Naive Bayes", International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.

[16] S. Suganya, and P.Tamije Selvy, January 2016, "A Proficient Heart Disease Prediction Method using Fuzzy-Cart Algorithm", International Journal of Scientific Engineering and Applied Science (IJSEAS), Vol. 2, Issue1,ISSN: 2395-3470.

[17] Ashwini Shetty A, and Chandra Naik, May 2016,"Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative Research in Science,Engineering and Technology(An ISO 3297: 2007 Certified Organization), Vol. 5, Special Issue 9, pp. 277-281.

[18] K Cinetha, and Dr. P. Uma Maheswari, Mar.-Apr. 2014,"Decision Support System for Precluding Coronary Heart Disease using Fuzzy Logic.", International Journal of Computer Science Trends and Technology (IJCST), Vol. 2, Issue 2, pp. 102-107.

[19] Indira S. Fal Dessai,2013,"Intelligent Heart Disease Prediction System Using Probabilistic Neural Network", International Journal on Advanced Computer Theory and Engineering (IJACTE),Vol. 2, Issue 3, pp. 38-44.

[20] Mai Shouman, Tim Turner, and Rob Stocker, June 2012,"Applying k-Nearest Neighbors in Diagnosing HeartDiseasePatients",International Journal of Information and Education Technology, Vol. 2, No. 3,pp. 220-223.

[21] Serdar AYDIN, Meysam Ahanpanjeh,and Sogol Mohabbatiyan,February 2016, "Comparison And Evaluation of Data Mining Techniques in the Diagnosis of Heart Disease",International Journal on Computational Science & Applications (IJCSA), Vol. 6,No.1, pp. 1-15.

[22] G. Purusothaman, and P. Krishnakumari, June 2015,"A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", Indian Journal of Science and Technology**,** Vol. 8(12), DOI:10.17485/ijst/2015/v8i12/58385, pp. 1-5.

[23] Deepali Chandna, 2014,"Diagnosis of Heart Disease Using Data Mining Algorithm", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (2), pp. 1678-1680.

[24] S. Prabhavathi, and D. M. Chitra, Jan. 2016, "Analysis and Prediction of Various Heart Diseases using DNFS Techniques**",** International Journal of

Innovations in Scientific and Engineering Research(IJISER), Vol.2, Issue 1, pp. 1-7.

[25] Jaymin Patel, Prof. Tejal Upadhyay, and Dr. Samir Patel, Sep 2015-Mar 2016, "Heart Disease Prediction using Machine Learning and Data Mining Technique", Vol. 7, No.1, pp. 129-137.

[26] Vikas Chaurasia, and Saurabh Pal, 2013, "Early Prediction of Heart Diseases Using DataMiningTechniques",Caribbean Journal of Science and Technology, ISSN: 0799-3757, Vol.1, pp. 208-218.

[27] Gunsai Pooja Dineshgar, and Mrs. Lolita Singh, February 2016, "A Review on Data Mining for Heart Disease Prediction",International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Vol. 5, Issue 2, pp. 462-466.

[28] Jyoti Soni, Uzma Ansari, Dipesh Sharma, and SunitaSoni,June 2011, "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers", International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 6, pp. 2385-2392.

[29] Kamal Kant, and Dr. Kanwal Garg,2014, "Review of Heart Disease Prediction using Data Mining Classifications",International Journal for Scientific Research & Development(IJSRD), Vol. 2, Issue 04, ISSN (online): 2321-0613, pp. 109-111.

[30] Sharan Monica L, and Sathees Kumar B,February 2016, "Analysis of Cardiovascular Heart Disease Prediction Using Data Mining Techniques", International Journal of Modern Computer Science (IJMCS), ISSN: 2320-7868(Online), Vol. 4, Issue 1, pp. 55-58.

[31] Nidhi Bhatla, and Kiran Jyoti, Oct. 2012,"An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), Vol. 1, Issue 8, ISSN: 2278-0181, pp. 1-4.

[32] Sumitra Sangwan, and Tazeem Ahmad Khan, Mar. 2015,"Review Paper Automatic Console for Disease Prediction using Integrated Module of A-priori and k-mean through ECG Signal",International Journal For Technological Research In Engineering, Vol. 2, Issue 7, ISSN(Online): 2347-4718, pp. 1368-1372.

[33] Rishi Dubey, and Santosh Chandrakar, Aug. 2015,"Review on Hybrid Data Mining Techniques for The Diagnosis of Heart Diseases in Medical Ground" ,Vol. 5, Issue 8, ISSN: 2249-555X, pp. 715-718.

[34] Ashish Chhabbi,Lakhan Ahuja,Sahil Ahir, and Y. K. Sharma,19 March 2016,"Heart Disease Prediction Using Data Mining Techniques", International Journal of Research in Advent Technology,E-ISSN:2321-9637,Special Issue National Conference "NCPC-2016", pp. 104-106.

[35] Shadab Adam Pattekari,and Asma Parveen, 2012,"Prediction System for Heart Disease using Naive Bayes", International Journal of Advanced Computer and Mathematical Sciences, ISSN: 2230-9624,Vol. 3, Issue 3, pp. 290-294.

[36] Boshra Bahrami, and Mirsaeid Hosseini Shirvani,February 2015,"Prediction and Diagnosis of Heart Disease by Data Mining Techniques",Journal of Multidisciplinary Engineering Science and Technology(JMEST), ISSN:3159-0040, Vol. 2, Issue 2, pp. 164-168.

[37] Dhanashree S. Medhekar, Mayur P. Bote, and Shruti D. Deshmukh, March 2013,"Heart Disease Prediction System Using Naive Bayes", International Journal of Enhanced Research in Science Technology & Engineering, ISSN No: 2319-7463, Vol. 2, Issue 3, pp. 1-5.

[38] Noura Ajam, 2015, "Heart Diseases Diagnoses Using Artificial Neural Network", Network And Complex Systems, ISSN: 2224-610X (Paper), ISSN: 2225-0603(Online), Vol.5, No.4,pp. 7-11.

[39] S.Florence, N. G. Bhuvaneswari Amma, G. Annapoorani, and K. Malathi, November 2014, "Predicting The Risk of Heart Attacks using Neural Network and Decision Tree", International Journal Of Innovative Research In Computer And Communication Engineering, ISSN (Online): 2320-9801, Vol. 2, Issue 11, pp. 7025-7028.

[40] https://www.saylor.org/site/wpcontent/uploads/2012/06/Wikipedia-Decision-Tree.pdf.

[41] https://en.wikipedia.org/wiki/MATLAB.