A Comparative Study for Predicting Heart Diseases using Data Mining Techniques

Ravi Ranjan

UG Student, Department of Computer Science

Sambhram Institute of Technology, Bengaluru

Rahul Kumar Yadav

UG Student, Department of Computer Science

Sambhram Institute of Technology, Bengaluru

T. Srinivasa Rao

Professor, Department of Computer Science

Sambhram Institute of Technology, Bengaluru

Abstract

Heart disease diagnosis is a complex task that requires much experience and knowledge. The traditional way of predicting Heart disease is a very complex task to predict the disease like ECG Heart MRI and Stress Test etc. In today's era deaths due to heart disease increase According to world health organization heart disease are the number 1 cause of death globally. Computer-based information along with advanced Data Mining classification techniques are used for appropriate results.in this paper has analyzed prediction systems of Heart disease using more number of input. the system uses medical parameters such as sex, age, blood pressure, cholesterol etc. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database, from the results, it has been seen that neural network and Naive Bayes predict Heart Disease with the highest accuracy.

Keywords- *Data Mining, Decision Trees, Heart Disease, Naive Bayes, Neural Networks.*

1.INTRODUCTION

The heart is an important organ of the human body. It pumps blood to every part of our anatomy. life is itself dependent on the efficient working of the heart. if the operation of the heart is not proper, it will affect the other body parts of humans such as the brain, kidney, etc. According to WHO, 17 million total global deaths are due to heart attacks and strokes. WHO, estimated by 2030, almost 23.6 million people will die to heart disease. Medical organizations, all around the world, collect data on Heart disease. These data can be exploited using various data mining techniques to gain useful insights. But the data collected is very large and, many times, this data can be very noisy. These data sets, which are too overwhelming for human minds to comprehend, can be easily explored by using the Data Mining technique to predict heart disease with the highest accuracy.

2.ABOUT HEART DISEASE

Heart means "cardio" hence all heart diseases concern to a category of cardiovascular diseases. Life is completely dependent on the efficient working of the heart.

Some of the major symptoms of heart diseases [4]:

- Hypertension

Hypertensive heart disease refers to heart conditions caused by high blood pressure.

- High blood pressure

High blood pressure (hypertension) can quietly damage your body for years before symptoms develop. Uncontrolled high blood pressure can lead to disability, a poor quality of life, or even a fatal heart attack or stroke. High blood pressure (BP), or hypertension, is defined by two levels by 2017 American College of Cardiology/American Heart Association (ACC/AHA) guidelines.

- High blood cholesterol

High blood cholesterol is a condition that causes the levels of certain bad fats, or lipids, to be too high in the blood.

- Smoking

Smoking harms nearly every organ in the body, including the heart, blood vessels etc.

- Family history

A family history of heart diseases, meaning one or more of your blood relatives has or has had heart disease, is one of the primary risk factors in experiencing cardiovascular issues, according to the National Institutes of Health.

- Obesity

Obesity is a common problem in the United States. Current research suggests that one in three Americans is obese. Obesity and overweight are linked to several factors that increase ones risk for cardiovascular disease.

The different kinds of heart diseases are as follows [5]:

- Coronary heart diseases

The coronary arteries supply the heart muscle with nutrients and oxygen by circulating blood.

- Angina pectoris

Angina pectoris is the medical term for chest pain or discomfort due to coronary heart disease.

- Congestive heart failure

Congestive heart failure (CHF) is a chronic progressive condition that affects the pumping power of your heart muscles. While often referred to simply as "heart failure," CHF specifically refers to the stage in which fluid builds up around the heart and causes it to pump inefficiently.

- Cardiomyopathy

Cardiomyopathy is a disease in which your heart muscle weakens and has difficulty pumping blood.

- Congenital heart diseases.

A congenital heart defect is a problem with the structure of the heart. It is present at birth. Congenital heart defects are the most common type of birth defect. The defects can involve the walls of the heart, the valves of the heart, and the arteries and veins near the heart.

3.LITERATURE SURVEY

Monika Gandhi et.al, [1] made use of Naive Bayes, Decision tree and neural network algorithms. There is a huge number of features involved. So, there is a need to reduce the number of features. This can be done by feature selection. On doing this, they say that time is reduced. They made use Decision tree and neural network

V.V.Ramalingam, Ayantan Dandapath, M Karthik Raja [2] they say that prediction of Heart Disease, Random Forest and Ensemble models have performed very well because they solve the problem of overfitting by employing multiple algorithms.

Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte [3] made use of Data mining techniques for appropriate results. They use Neural network to predict the Heart Diseases. For prediction, the system uses sex, blood pressure like 13 medical parameters.

Chaitrali S. Dangare, Sulabha S. Apte[4] Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques.Made use of Naive Bayes, Decision tree and Neural Network they say that Neural Network provides accurate results as compare to Decision tree & Naive Bayes.

M.Marimuthu, S. Deivarani, Gayathri.R[5] Analysis of Heart Disease Prediction Using Various Machine Learning Techniques like the k-nearest neighbor, Naive Bayes, Support vector machine and Decision tree are used. they calculated the accuracy of each algorithm. Then the one with good is Support Vector Machine.

Rajesh N, T Maneesha, Shaika Hafeez, Hari Krishna[6] they say that with Naive Bayes algorithm getting more accurate prediction compare to the Decision Tree Algorithm

N.Raihan, Parichay Kumar Mandal[7] made use of Artificial Neural Network presented a Risk Prediction Of   Ischemic Heart   Disease.

Avinash Golande, Pavan Kumar T[8] made use of various   Machine Learning Techniques to predict the Heart Disease.

Hariharan K, Vigneshwar W.S, Subramaniyaswamy V[9]   discussed in detail The Heart Disease and different types of Heart Disease. And also the popular techniques available in data mining.

Sheena Angra, Sachin Ahuja[10] discussed the various machine learning techniques and approaches in various fields and application areas.

Isra'a Ahmed Zriqat, Ahmad Altamimi, Mohammad Azzeh[11] A Comparative study for predicting Heart Disease Using Data Mining Classification Methods. They are sad that the Decision Tree gives the highest accuracy compared to the Random Forest on the same Data set.

4. DECISION TREE

The decision tree is a classification technique used in data mining, statistics and machine learning. It is a supervised learning algorithm. They can be used to solve both regression and classification problems. First, calculate the entropy of every attribute. Then the data set to split with the help of the variables or predictor. In Decision Tree, the major challenge is to the identification of the root node in each level. This process is known as attribute selection. We have three popular attribute selection measures.The leaf nodes of the decision tree represent the class label and branches represent the affiliation of the feature that leads to the class labels [9].

1)   Information Gain

The information gain is based on the decrease in entropy after a data-set is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain.

2)   Gain Ratio

In decision tree learning, Information gain ratio is a ratio of information gain to the intrinsic information.

3)   Gini Index

The Gini Index is calculated by subtracting the sum of the squared probabilities of each class from one. It favors larger partitions. Information Gain multiplies the probability of the class times the log (base=2) of that class probability.

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

5. NAIVE BAYES

It simple and effective classification techniques which are based on the Bayes Theorem with strong independence assumptions.it considers that the presence/absence of a particular attribute of a class is unrelated to the presence/absence of any other attribute when the class variable is given [11]. The Naïve Bayes Classifier technique is mainly applicable when the dimensionality of the inputs is high [2]. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. Naïve Bayes model recognizes the characteristics of patients with heart disease. It shows the probability of each input attribute for the predictable state.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Likelihood

Class Prior Probability

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

6. Neural Network

An artificial neural network(ANN). one of the most important models in Artificial Neural Network is the Multilayer perceptron(MLP). the MLP consists of one input layer, one output layer, and one or more hidden layers. each layer consists of one or more nodes. the output of the input layer is given to the hidden layer, through weighted connection links. the output of the hidden layer is forwarded to the output layer, it performs computations and produces final results [3]. The ANN can be classified in two main groups according to their working.

    1) Supervised learning: It is a simple model, in which the networks compute a response to each input and then compare it with target value. If the computed response differs from target value, the weights of the network are adapted according to a learning rule. e.g. Single-layer perceptron, Multi-layer perceptron.

    2) Unsupervised learning: These networks learn by identifying special features in the problems they are exposed to. e.g.: Self-organizing feature maps.

7. CONCLUSIONS

In this research, different data Mining techniques were studied to enhance the early prediction of heart disease. With two or more data set that gives different accuracy, However, it confirms that the decision tree was ranked first across all evaluation Measures on data sets, followed by ANN in the second position. This can give important conclusions about all the techniques. This investigation tells us about dissimilar technologies that are used in dissimilar papers with dissimilar count of attributes with different accuracies depending on the tools designed for execution. The accurateness of the structure can be further upgraded by creating various combinations of data mining techniques and by parameter tuning also.

8. REFERENCES

[1] Monika Gandhi, Shailendra Narayanan Singh ,"Predictions in heart disease using techniques of data mining" 2015.

[2] V. V. Ramalingam, Ayantan Dandapath, M Karthik Raja ,"Heart Disease prediction using machine learning techniques :a survey" IJET 684-687, 2018

[3] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A Data Mining Approach For Prediction Of Heart Disease Using Neural Networks". IJCET 2012.

[4] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques",IJET 0975-888, June 2012

[5] M. Marimuthu, S.Deivarani, Gayathri.R,"Analysis of Heart Disease prediction using various Machine Learning Techniques"

[6] Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna,"Prediction of heart Disease using machine learning Algorithms", IJET 363-366, 2018.

[7] M. Raihan, Parichay Kumar Mandal, Muhammad Muinul Islam et.al ," Risk Prediction Of Ischemic Heart Disease Using Artificial Neural Network", ECCE 2019.

[8] Avinash Golanda, Pavan Kumar T, "Heart Disease Prediction Using Effective Machine Learning   Techniques", ISSN:2277-3878, June 2019

[9] Hariharan K, Vigneshwar W.S, Sivaramakrishan   N, Subramaniyaswamy V," A Comparative Study On Heart Disease Analysis Using Classification Techniques",IJPAM 13357-13366, 2018.

[10] Sheena Angra, Sachin Ahuja," Machine Learning and its Applications: A Review", IEEE 2017

[11] Isra'a Ahmed Zriqat, Ahmad Monsa Altamimi, Mohammad Azzeh,"A Comparative Study for Prediction Heart Disease Using Data Mining Classification Methods", IJCSIS, December 2016