A Comprehensive Analysis on Technique's Used to Predict Heart Disease

Anusha K S[1], Radhika A D[2]

[1]M.Tech Scholar, Department of Computer Science and Engineering, VVCE, Mysuru, Karnataka, India

[2]Assistant Professor, Department of Computer Science and Engineering, VVCE, Mysuru, Karnataka, India

ABSTRACT

As we are in modern and competitive world, we are tending to lead the sophisticated life where in we have forgotten to take care about our life. We are not concentrating on our food habit, mental and physical fitness. We are leading our life towards risk in fact to death. According to recent survey the major reason for death is due to heart attack, one in every four death is due to heart attack. Many technologies have come up effectively to discover the reason for heart attack and predict it before it occurs. In this paper we have made an attempt to analyze those various techniques used to predict the heart disease.

Keywords : Heart Disease, Naive Bayes, K-Means, Genetic Alogrithm, Newral Networks, Prediction.

I. INTRODUCTION

It is a world known fact that heart is the most essential organ in human body if that organ gets affected then it also affects the other vital parts of the body. Therefore it is very important for people to go for a heart disease diagnosis periodically. Poor clinical decisions can lead to tragic consequences which are therefore unacceptable. They can achieve the best results by employing appropriate computer-based information and/or decision support systems. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining and machine learning techniques can help remedy this situation. The basic flow for predicting heart disease is as follow [1] [2].

Patients database
The patient database will be collected, which will majorly contain 14 attributes/parameters. The 14 attributed are age, sex, Cp, trestbps, Chol, Fbs, restecg, thalach, exang, oldpeak, slope, Ca, thal, Num and this is listed with abbreviation in figure 1.

Data Analysis
Analyzing the data is the very important step, because they may contain some noisy data which is not required in our work, in order to remove that data we perform data cleaning and data integration to fill up the missing values.

Feature Selection
The data processing is done in this step, which helps to reduce the dimensionality; intern increases the accuracy and performance. The technique that is used in data processing is PDA, chi square test to name few.

Optimisation Algorithm
Selecting the algorithm to predict the heart disease is another important step, we should choose the algorithm which gives us better result with more accuracy. The algorithm which can be used is PSO, Genetic Algorithm, KNN, Artificial Neural Network, Naïve Baye's algorithm and many more.

Training and Clasification

The collected data will be trained and given as an input to the system and several classification techniques will be applied to extract hidden and useful information.

Prediction

The final step is to predict the heart disease using the information given by the patient, accordingly the system will predict whether the person will get heart disease or not.

Table 1 : list of major attributes to predict the heart disease

No.	Name	Description
1	Age	Age in Years
2	Sex	1=male, 0=female
3	Cp	Chest pain type(1=typical angina, 2=atypical angina, 3=non-anginal pain, 4=asymptomatic)
4	trestbps	Resting blood sugar(in mm Hg on admission to hospital)
5	Chol	Serum cholesterol in mg/dl
6	Fbs	Fasting blood sugar>120 mg/dl(1-true,0=false)
7	restecg	Resting electrocardiographic results(0=normal, 1=having ST-T wave abnormality, 2=left ventricularhypertrophy)
8	thalach	Maximum heart rate
9	exang	Exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	Slope of the peak exercise ST segment(1=upsloping, 2=flat, 3=downsloping)
12	Ca	Number of major vessels colored by fluoroscopy
13	thal	3=normal, 6=fixed defect, 7=reversible defect
14	Num	Class(0=healthy, 1=have heart disease)

The remainder of the paper is organized as follows: Section II presents the algorithms used to predict the heart disease. Section III presents the literature survey done on papers published in various journals. Section IV concludes the overall summary of the paper with the future work of the authors.

Algorithms to predict heart disease

There are many algorithms proposed by researches which effectively hepls to predict the heart disease, some of them are decision tree, K-means algorithms, Genetic algorithms, Naïve baye's algorithm and newral networks. Even now many works are being carried out to enhance the algorithm in order to get more accuracy in heart disease prediction.

Decision tree

The decision tree algorithm tries to solve the problem by using tree representation. The tree is built in a top-down format, which contains the root node and involves portioning the data into subsets that contain instances with similar values.

K-Means Algorithms

The K-Means algorithm is a clustering algorithm, which can be applied to the real world application. It groups the data in accordance with their characteristic value into k distinct cluster and the input data is taken for characterization into the same cluster having identical feature value. Let k be the positive integer denoting the number of clusters and it needs to be provided in advance. Then the pre processed heart disease data is clustered using K means algorithm with the K cluster of values [4].

Genetic Algorithms

This is the frequently used technique in data mining. It repeatedly modifies a population of individual solutions. The algorithm begins by creating a random initial population. The algorithm then creates a sequence of new populations. At each step, the algorithm uses the individuals in the current generation to create the next population. The algorithm stops when one of the stopping criteria is met.

Naïve Baye's Agorithm

It is a Classification technique based on Baye's theorem with an assumption of independence among predictors that is a Naive Baye's classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Baye's model is easy to build and particularly useful for very large data sets. It can be stated as the posterior probability is equal to likelihood and class prior probability to predictor prior probability.

Neural Networks

This is one of the most popular techniques in data mining. Neural network consists of 3 layer-input, hidden and output layers. There are set of connected input and output units; each connected unit has an associated weight present with respective units. During learning phase of neural network, networks learns from by adjusting weights so that it can be able to predict the correctly labeled class categories of the input tuples for the output prediction. The primary function of the neurons of input layer is to divide input into neurons in hidden layer. It maps a set of input data into a set of appropriate output data.

II. LITERATURE SURVEY

Many research works have been carried out to predict the heart disease using various techniques of data mining and machine learning, we have analyzed some of those algorithm and techniques and summarized them with the models accuracy.

Purushottam at el, [3] have proposed a system using data mining technique- decision tree, which effectively predict the heart disease. This system makes to take effective decision based on certain parameter. This system has given 86.3% accuracy in testing phase and 87.3% in training phase.

Sairabi H.Mujawar et al, [5] have used K-means and Naive Baye's to build a system which predicts heart disease using previous data set, they have considered 13 attributes. The model will predict whether the patient is having heart disease or not, they got 93% accuracy by considering total of 300 records.

S. seema et al, [6] aims to predict the chronic disease, using naïve Baye's, Decision tree, SVM, ANN. The comparative study is performed on classifiers. 95.56% accuracy was given by SVM in case of heart disease prediction.

Ashok Kumar Dwivedi et al, [7] suggested algorithms like naïve Baye's, KNN, Classification tree, Logistic regression, SVM and ANN. Among all Logistic regression gave the better result.

Chala Beyene et al, [8] the purpose is to predict the heart disease for early automatic diagnosis of the disease within result in short time. It uses the major attributes for prediction and analyses of data set are computed using WEKA software.

R. Sharmila et al, [9] have used non-linear classification algorithm to predict heart disease. Bigdata tools such as HADOOP, Mapreduce have been used and they suggest to use these tools to store large data set. The sequential and parallel SVM is used, accordingly 82.35% and 85% accuracy have obtained.

P. Sai Chandrashekar Reddy et al, [10] used ANN algorithm to predict heart disease. The model is used to predict the condition of the patient evaluating through various parameters. The accuracy of the system is provided in Java.

Table 2 : A Comprehensive analysis of various algorithms in literature Survey

YEAR	AUTHOR	PURPOSE	TECHNIQUES USED	ACCURACY
2015	Sharma Purushottam et al, [3]	Efficient Heart Disease Prediction System using Decision Tree.	Decision tree classifier	86.3% for testing phase. 87.3% for training phase.
2015	Sairabi H. Mujawar et al, [5]	Prediction of Heart Disease using Modified K-means and by using Naïve Bayes.	Modified k-means algorithm, naive Bayes algorithm.	Heart Disease detection=93%. Heart Disease undetection=89%.
2016	Dr. S. Seema et al, [6]	Chronic Disease Prediction by mining the data.	Naïve Bayes	Highest accuracy achieved by SVM, in case of heart disease 95.56%
			Decision Tree Support Vector Machine	Highest accuracy of 73.588% achieved by Naïve Bayes in case of diabetes.
2016	Ashok Kumar Dwivedi et al, [7]	Evaluate the performance of different machine learning techniques for heart disease prediction.	Naïve Baye's	83%
			KNN	80%
			Logistic Regression	85%
			Classification Tree	77%
2017	P. Sai Chandrasekhar Reddy et al, [10]	Heart disease prediction using ANN algorithm in data mining.	ANN	Accuracy proved in JAVA.
2018	Chala Bayen et al, [8]	Prediction and Analysis the occurrence of Heart Disease using data mining techniques.	J48, Naïve Bayes, Support Vector Machine.	It gives short time result which helps to give quality of services and reduce cost to individuals.
2018	R. Sharmila et al, [13]	A conceptual method to enhance the prediction of heart diseases using the data techniques.	SVM in parallel fashion	SVM provides better and efficient accuracy of 85% and 82.35%. SVM in parallel fashion gives better accuracy than sequential SVM.

III. CONCLUSION AND FUTURE WORK

In this paper we have discussed the basic flow for predicting the heart disease, the main attributes that are considered for evaluation. Various algorithms like Decision tree, K-Means algorithms, Genetic algorithms, naïve Baye's algorithms, Neural networks, used for the heart disease prediction was discussed. A comprehensive analysis was done on various techniques used by the researches with the model/system accuracy.

There are many possible ways to improve the scalability and accuracy of the prediction system. As noticed all authors have considered the standard 13-14 attributes in their work, in future work the attributes can be reduced aiming to get more accuracy by making use of feature selection-brute force algorithm with large data set and predict the percentage possibility of getting the heart disease including at what time/age that the patient may get heart attack using classification technique.

IV. REFERENCES

[1]. Bhuvaneswari Amma N.G "Cardiovascular Disease prediction system using Genetic Algorithm and Neural Network", IEEE.

[2]. Sana Bharti, Dr.Shaliendra Narayan Singh "Analytical study of heart Disearse prediction Comparing with Different Algorithms", International conference on computting, communication and automation(ICCCA2015).

[3]. Purushottam, Prof.(Dr.) Kanak saxena, Richa sharma, "Efficient heart disease prediction system", 2016, pp.962-969

[4]. C.Kalaiselvi, PhD, "Diagnosing of Heart Diseases using Average K-Nearest neighbor algorithm of Data Mining" 2016 International conference on computing for sustainable Global Development(INDIACom)

[5]. Sairabi H.Mujawar, P.R.Devale, "Prediction of Heart Disease using Modified K-meand and by using Naïve Baye's", International journal of Innovative research in computer and communication engineering, Vol.3, October 2015, pp.10265-10273.

[6]. Dr. S. Seema Shedole, Kuymari Deepika, "Predictive analytics to prevent and control chromic disease", https : // www. Researchgate.net/publication/316530782, January 2016

[7]. Ashok Kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation",springer, 17 september 2016.

[8]. Mr. Chala Beyene, Prof. Pooja kamat, "Survey on prediction and analysis the occurrence of heart disease using data mining techniques" Internaltional Journal of Pure and Applied Mathematics,2018.

[9]. R.Sharmila, S.Chellammal, "A Conceptual method to enhance the prediction of heart disease using the data mining techniques", International Journal of Computer Science and engineering, may 2018.

[10]. Mr.P.Sai Chandrasekhar Reddy, Mr.Puneet Palagi, S.Jaya, "Heart Disease prediction using ANN algorithm in Data mining, International journal of Compurt Science and Mobile Computing.April 2017,pp.168-172.