Volume-17,Issue-6 (June-22)
ISSN (O) :- 2349-3585

A Survey on Predicting Heart Disease at Early Stages using Machine learning techniques

Divya K [1] , Saravanan G [2]

[2] ASSISTANT PROFESSOR

Department of Computer Science and Engineering, Erode Sengunthar College of Engineering(Autonomous), Perundurai, erode-638057

Abstract— Predicting and detection of heart disease has always been a critical and challenging task for healthcare practitioners. Hospitals and other clinics are offering expensive therapies and operations to treat heart diseases. S o, predicting heart disease at the early stages will be useful to the people around the world so that they will take necessary actions before getting severe. Heart disease is a significant problem in recent times; the main reason for this disease is the intake of alcohol, tobacco, and lack of physical exercise. Over the years, machine learning shows effective results in making decisions and predictions from the broad set of data produced by the health care industry. S ome of the supervised machine learning techniques used in this prediction of heart disease are artificial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (SVM), native Bayes) (NB) and k- nearest neighbour algorithm. Furthermore, the performances of these algorithms are summarized.

Keywords— Machine learning, supervised learning, health care services, heart disease.

I. INTRODUCTION

The heart is one of the main parts of the human body after the brain. The primary function of the heart is to pumping blood to the whole body parts. Any disorder that can lead to disturbing the functionality of the heart is called heart disease. Several types of heart disease are there in the world; Coronary artery disease (CAD), and heart failure (HF) are the most common heart diseases that are present. The main reason behind the coronary heart disease (CAD) is blockage or narrowing down of the coronary arteries [1] Coronary arteries also responsible for supplying blood to the heart. CAD is the leading cause of death over 26 million people are suffering from coronary heart disease(CAD) around the world, and it is increasing 2% annually due to CAD 17.5 million deaths happened globally in 2005[2]. In the growing world, 2% of the population around the world is suffering from CAD, and 10% of the people are older than 65 years. Approximately 2% of the annual healthcare budget spent only to treat CAD disease. USA government spent 35 billion dollars for CAD in 2018[3].

Different factors can raise the risk of heart failure. Medical scientists have classified those factors into two different categories; one of them is risk factors that cannot be changed, and another one is risk factors that can be changed. Family history, sex, age comes under risk factors that cannot be changed. High cholesterol, smoking, physical inactivity, high blood pleasure all these come under risk factors [4].

Heart disease is a significant issue, so there is a need for diagnosis or prediction of heart disease there areseveral methods to diagnose heart disease among them Angiography is the trending method which is used by most of the physicians across the world. However, there are some drawbacks associated with angiography technique. It is an expensive procedure and physicians have to analyze so many factors to diagnose a patient hence this process makes physician job very difficult, so these limitations motivate to

develop a non-invasive method for prediction of heart disease. These conventional methods deal with medical reports of the patients moreover these conventional methods are time-consuming, and it may give erroneous results because these conventional methods are performed by humans[2][4]. To avoid these errors and to achieve better and faster results, we need an automated system. Over the past years, researchers find out that machine learning algorithms perform very well in analyzing medical data sets. These data sets will be directly given to machine learning algorithms, and machine learning algorithms will perform according to their nature, and those algorithms will give some outputs. There are some common attributes which are used to predict the heart diseases are:

- Gender (it is a binary attribute 1 for female, 0 for male).
- Age.
- Resting blood pressure.
- Types of chest pain.
- Serum cholesterol in mg/dl.
- Fasting blood sugar.
- ECG results.
- Heart rate.
- Thalassemia.
- Old peak[5].

TABLE I. VARIOUS TYPESOF HEART DISEASES[6]:

HEART DISEASE TYPE:	Description:
Coronary artery	Coronary artery happens because of the blockage of arteries in the heart
Vascular disease	Vascular disease happens when blood flow reduced to the heart.
Heart rhythm disorder	It is another type of heart disease; it is nothing but a disturbance in heart rhythm; it could be heartbeats too faster or too slowly or abnormal way.
Structural heart disease	It means muscles or blood vessels, valves; walls near the heart are present in a disorganized way. This disorganization of heart structure will cause heart failure.
Heart failure	Heart failure will happen when the heart is completely damaged. Heart attacks and high blood pressure these two will lead to heart failure.

II. MACHINE LEARNING ALGORITHMS

A. Artificial neural network (ANN):

Artificial neural networks mainly designed for computational purposes; the main theme of this model is to do a job faster than the traditional model. This model is similar to the biological structure of neurons in the human brain. How the neurons connect in the brain the same way here, also neurons (nodes) will connect. This model consists of a vast number of interconnected elements (neurons) working collectively to perform a task. A single layer neural network is called a perceptron it gives us a single output[7].

B. Support vector machine (SVM):

A support vector machine is a supervised learning technique in machine learning algorithms. If you give any labeled training data to support vector machine algorithms , it will produce a classifier that will divide the labeled data into different classes.

In the one-dimensional (1D) space, this classifier is called a point.

In two-dimensional (2D) space, this classifier is called a

line.

In three-dimensional (3D) space, this classifier is called a plane.

ISSN:-2349-3585 |www.ijrdt.org

In four-dimensional(4D) or more space, this classifier is

called a hyperplane.[8][9].

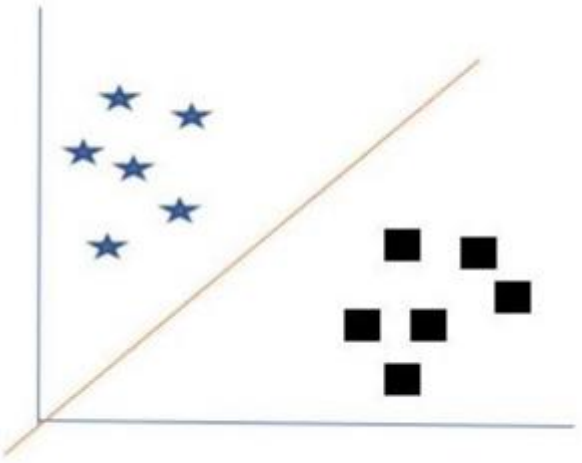

Fig. 1. Example of SVM.

In the above example, classifier is a line.

C. Decision tree (DT):

A decision tree is one of the supervised learning techniques in machine learning algorithms. It is used for both classification and regression. In this algorithm, data will be split according to the parameters. A decision tree is a tree that will contain nodes and leaves. At leaves, we will get outcomes or decision, and at the nodes, data will be split

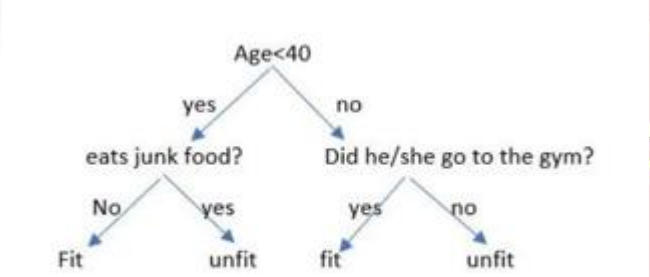

Fig. 2. Example of DT.

The decision tree is two types:

1. Classification tree
2. Regression tree

Classification tree: Here, we will get a decision (outcome) variable as categorical like the above example.

Regression tree: Here will get a decision (outcome) variable as continuous.

D. Random forest (RF):

It is one of the supervised machine learning algorithms which is used for both classification and regression also. However, it is mainly used for classification purposes. The name itself is suggested that it is a forest, a forest is a group of trees similarly in a random forest algorithm we will have trees these trees are the decision t rees. If we have a higher number of decision trees prediction results will be more accurate. Random forest algorithm works this way at; first it will collect random samples from the dataset and then it will create decision trees for each sample from those available

trees we will select the tree which will produce the best prediction results [10].

E. Naïve Bayes (NB):

Naïve Bayes is one of the supervised machine learning classification algorithms. Earlier it is used for te xt classification. It deals with the datasets which have the highest dimensionality. Some examples are s entimental analysis, spam filt ration, etc. This naïve Bayes algorithm is based on Bayes theorem with the assumption that attributes are independent of each other. It is nothing but attributes in one class that is independent of any other attributes that are present in the same class[11].

III. LITERATURE SURVEY

LIAQAT ALI et al. [4] recommended a model which is consists of two methods one is X^2 statistical and deep neural network(DNN). Feature refinement is done by X^2 statistical model and classification is done by a deep neural network(DNN). In their study, they have used the Cleveland dataset. There are 303 instances in that dataset, among them, 297 have no missing data, and the remaining 6 have missing data. Among 297, 207 instances are used for training data, and the remaining 90 are used as testing data. This model gives better results compared to conventional ANN models which are present earlier. As a result of this using this proposed model, they have got 93.33% classification accuracy using DNN. It is 3.33% more than that of the conventional ANN model.

ISSN:-2349-3585 |www.ijrdt.org

Dr.kanak Saxena et al. [12] developed a data mining model to predict heart disease efficiently. It mainly helps the medical practitioners to make efficient decisions way based on the given parameters. The author has used Cleveland dataset from UCI, and they have used age, sex, resting blood pressure, chest pain, serum cholesterol, fasting blood sugar, etc. as attributes. Furthermore, they have divided the datasets into two parts one is for testing, and the other one is for training. They have used a 10-fold method to find accuracy.

AWAIS NIMAT et al. [1] proposed an expert system based on two support vector machines(SVM) to predict heart disease efficiently. These tow SVM's have their purpose; first, one is used to remove the unnecessary features, and the second one is used for prediction. Moreover, they have used the HGSA (hybrid gird search algorithm) to optimize the two methods. By using this model, they have achieved 3.3% better accuracy than the conventional SVM models that are present earlier.

Deepika et al. [13] proposed predictive analytics to prevent and control the chronic disease with the help of machine learning techniques such as naive Bayes, support vector machine, decision tree, and artificial neural network and they have used UCI machine learning repository datasets to calculate the accuracy. Among them, Support vector machine gives the best accuracy of 95.55%.

Ashir Javeed et al. [2] developed a model to improve the prediction of heart disease by overcoming the problem of over fitting; over fitting means the proposed model performs and gives better accuracy on testing data and gives unfortunate accuracy result for training data while predicting the heart disease. To solve this problem, they have developed a model that will give the best accuracy on both training and testing data. That model consists of two algorithms one is RAS(Random search algorithm) other one is a random forest algorithm that is used to predict the model. This proposed model gave them better results in training data as well as testing data.

D.M Chitra et al. [14] suggested a system to predict various heart diseases using the DNFS technique. DNFS means a decision tree based fuzzy neural system. Here authors proposed a system that will predict heart disease with the help of data mining techniques, as well as machine learning techniques such as decision tree, naïve Bayes, k-nearest neighbours, support vector machines, and artificial neural network for the prediction. The authors have used the Cleveland database for the prediction with the 13 features. Moreover, they have conducted a comparative study on various algorithms also and finally, they found out that naïve Bayes and decision tree gives the best accuracy.

IV. COMPARATIVE STUDY OF LITERATURE SURVEY

TABLE II. COMPARISON TABLE OF LITERATURE SURVEY

Authors	Techniques used	Accuracy
Liaqat Ali et al. [4]	X^2 statistical model, deep neural network	93.33%(holdout) 91.57%(k-fold)
Dr.Kanak Saxena et.al[12]	Decision tree	86.3% (testing phase) 87.3% (training phase)
Awais Nimat et al. [1]	Support vector machine, Hybrid grid search algorithm (HGSA)	92.22% (L1 linear SVM+L2 linear & RBF SVM)
Deepika et.al [13]	Naïve Bayes, Decision tree, Support vector machine	SVM gives the best accuracy with 95.55%
Ashir Javeed et al. [2]	Random search algorithm (RSA), Random forest.	93.33% (RSA+RF)
D.M Chitra et.al[14]	Decision tree, Support vector machine,	Chronic disease diagnosis between 82% and 92%

ISSN:-2349-3585 |www.ijrdt.org

	Naïve Bayes.	

V. CONCLUSION

Heart disease is a very critical issue in the present growing world. So, there is a need for an automated system to predict heart disease at earlier stages. So that it will be useful for the physician to diagnose the patients efficiently, and it will be useful to the people also because they can track their health issues by using this automated system. Some of the expert automated systems were summarized in this paper. Feature selection and prediction, these two are essential for every automated system. By choosing features efficiently, we can achieve better results in predicting heart disease. We have summarized some algorithms which are useful while selecting the features, like hybrid grid search algorithm and random search algorithm, etc. So, in the future, it is better to use search algorithms for selecting the features and then applying machine learning techniques for prediction will give us better results in the prediction of heart disease.

REFERENCE

[1] L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," IEEE Access, vol. 7, pp. 54007–54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

[2] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," IEEE Access, vol. 7, pp. 180235– 180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[3] M. Gjoreski, A. Gradisek, B. Budna, M. Gams, and G. Poglajen, "Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure from Heart Sounds," IEEE Access, vol. 8, pp. 20313–20324, 2020, doi: 10.1109/ACCESS.2020.2968900.

[4] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on $\chi 2$ Statistical Model and Optimally Configured Deep Neural Network," IEEE Access, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[5] M. R. Ahmed, S. M. Hasan Mahmud, M. A. Hossin, H. Jahan, and S. R. Haider Noori, " A cloud based four-tier architecture for early detection of heart disease with machine learning algorithms," 2018 IEEE 4th Int. Conf. Comput. Commun. ICCC 2018, pp. 1951–1955, 2018, doi: 10.1109/CompComm.2018.8781022.

[6] "types of heart disease."[Online]. Available: https://www.heartandstroke.ca/heart/what-is-heart - disease/types- of-heart-disease.

[7] J. Schmidhuber, "Deep Learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015

[8] N. H. Farhat, "Photonit neural networks and learning mathines the role of electron-trapping materials," IEEE Expert. Syst. their Appl., vol. 7, no. 5, pp. 63–72, 1992, doi: 10.1109/64.163674

[9] M. Asaduzzaman, and S. Akhter Hossain, "An analysis of computational intelligence techniques for diabetes prediction Machine Learning View project An analysis of computational intelligence techniques for diabetes prediction," Int. J. Eng. &Technology, vol. 7, no. 4, pp. 6229–6232, 2018, doi: 10.14419/ijet.v7i4.28245.

[10] G. H. Tang, A. B. M. Rabie, and U. Hägg, "Indian hedgehog: A mechanotransduction mediator in condylar cartilage," J. Dent. Res., vol. 83, no. 5, pp. 434–438, 2004, doi: 10.1177/154405910408300516.

[11] Y. Karaca and C. Cattani, "7. Naive Bayesian classifier," Comput. Methods Data Anal., pp. 229–250, 2018, doi: 10.1515/9783110496369-007.

[12] Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," Procedia

ISSN:-2349-3585 |www.ijrdt.org

Comput. Sci., vol. 85, pp. 962–969, 2016, doi: 10.1016/j.procs.2016.05.288.

[13] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016, no. January 2016, pp. 381–386, 2017, doi: 10.1109/ICATCCT.2016.7912028.

[14] "Analysis and Prediction of Various Heart Diseases Using Dnfs Techniques," vol. 2, no. 1, pp. 1–7, 2015.

ISSN:-2349-3585 |www.ijrdt.org