WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Predicting Early-Stage Heart Disease using Machine Learning: A Survey

Shital Bhor[1], Dr. Prof.Rokade.M.D[2]

[1]Department of Computer Engineering, Otur Savitribai Phule University, Pune, India
[2]Department of ComputerEngineering,OturSavitribai Phule University, Pune, India.

ABSTRACT

The prediction and detection of heart diseases has always been a critical and challenging task for healthcare professionals. Hospitals and other clinics offer expensive therapies and surgeries to treat heart disease. So, predicting heart diseases in early stages will be helpful for people all over the world to take necessary measures before it gets worse. 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 has shown effective results in making decisions and making predictions from a wide range of data produced by the healthcare industry. Some of the supervised machine learning techniques used in this heart disease prediction are Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes) (NB) and k-Nearest Neighbor Algorithm. Next, the performances of these algorithms are summarized.

Keywords— Machine Learning, Supervised Learning, Healthcare Services, Heart Disease.

INTRODUCTION

After the brain, the heart is one of the main parts of the human body. The primary function of the heart is to pump blood throughout the body. Any disorder that can lead to impaired heart function is called heart disease. There are several types of heart disease in the world; coronary artery disease (CAD) and heart failure (HF) are the most common heart diseases that occur. The main cause of coronary heart disease (CAD) is the blockage or narrowing of the coronary arteries The coronary arteries are also responsible for supplying blood to the heart. CAD is the leading cause of death with more than 26 million people suffering from coronary heart disease (CAD) worldwide and is increasing by 2% annually due to CAD with 17.5 million deaths worldwide in 2005. In the growing world, CAD affects 2% of the population worldwide and 10% of people over 65 years of age. Approximately 2% of the annual health care budget is spent on treating CAD alone. The US government spent $35 billion on CAD in 2018.

Various factors can increase the risk of heart failure. Doctors have classified these factors into two different categories; one is risk factors that cannot be changed and the other is risk factors that can be changed. Family history, gender, age are risk factors that cannot be changed. High cholesterol, smoking, physical inactivity, high blood pressure are all risk factors.

Heart disease is a significant problem, so there is a need to diagnose or predict heart disease, there are several methods to diagnose heart disease, among them Angiography is a trending method used by most of the doctors around the world. However, there are some disadvantages associated with the angiography technique. It is an expensive procedure and doctors have to analyze so many factors to diagnose a patient, therefore this

process makes the doctor's job verydifficult, so these limitations motivate the development of a non-invasive method to predict heart disease. These conventional methods deal with patients' medical reports, moreover, these conventional methods are time-consuming and may give erroneous results because these conventional methods are performed by humans. To avoid these mistakes and achieve better and faster results, we need an automated system. In recent years, researchers have found that machine learning algorithms work very well in analyzing medical data sets. These data sets will be directly fed to the machine learning algorithms and the machine learning algorithms will work according to their nature and these algorithms will provide certain outputs. There are some common attributes that are used to predict heart disease:

Gender (this is a binary attribute of 1 for female, 0 for male).

- Age.
- Resting blood pressure.
- Types of chestpain.
- Serum cholesterol inmg/dl.
- Fasting bloodsugar.
- ECG results.
- Heartrate.
- Thalassemia.
- Oldpeak.

TABLE I. DIFFERENT TYPES OF HEART DISEASE [6]:

TYPE OF HEART DISEASE:	**Description:**
Coronary artery	A coronary artery is caused by a blockage of the arteries in the heart
Vascular disease	Vascular disease occurs when blood flow to the heart is reduced.
Heart rhythm disorder	It is a different type of heart disease; it is nothing more than a heart disorder rhythm; it may be a heartbeat that is too fast or too slow or an abnormal heartbeat.
Structural heart disease	It means muscles or vessels, valves; the walls near the heart are present in a disordered manner. This disorganization of heart structure causes heart failure.
Heart failure	Heart failure occurs when the heart is completely damaged. Heart attacks and high blood pressure these two will lead to heart failure.

MACHINE LEARNING ALGORITHMS

A. Artificial neural network (ANN):

Artificial neural networks intended primarily for computing purposes; The main theme of this model is to get the job done faster than the traditional model. This model is similar to the biological structure of neurons in the human brain. As neurons connect in the brain, the neurons (nodes) also connect here. This model consists of a large number of interconnected elements (neurons) that work together to accomplish a task. A single-layer neural network is called a perceptron and 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 some labeled training data to support vector machine algorithms, a classifier will be created to divide the labeled data into different classes.

In 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.

In four-dimensional (4D) or more space is this classifier the so-called hyperplane.[8][9]

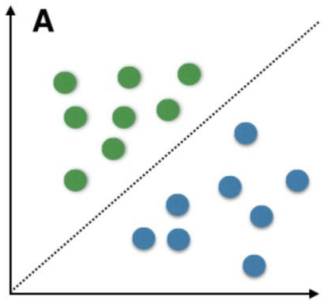

Fig. 1. Example of SVM.

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, the data will be split according to the parameters. A decision tree is a tree that will contain nodes and leaves. On the leaves we get the results or decisionsand on the nodes the data is divide

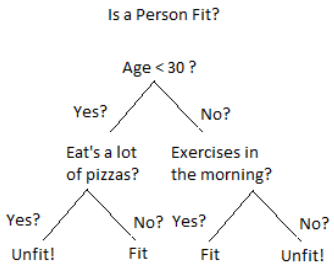

Fig. 2. Example of DT.

The decision tree is two types:

1. Classification tree
2. Regression tree

Classification Tree: Here we get the decision (outcome) variable as categorical as the above example.

Regression tree: Here we get the decision (outcome) variable as continuous.

Random forest(RF):
It is one of the supervised machine learning algorithms used for both classification and regression. However, it is primarily used for classification purposes. The name itself suggests 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 decision trees. If we have a higher number of decision trees, the prediction results will be more accurate. The random forest algorithm works this way on; first collects random samples from the data set and then creates decision trees for each sample from these available trees we select the tree that gives the best prediction result.[10]

D. Naïve Bayes(NB):
Naïve Bayes is one of the supervised machine learning classification algorithms. Previously used for text classification. It deals with datasets that have the highest dimensionality. Some examples are sentiment analysis, spam filtering ratio, etc. This naive Bayes algorithm is based on Bayes theorem with the assumption that the attributes are independent of each other. It is nothing but attributes in one class that are independent of any other attributes that are present in the same class [11].

LITERATURESURVEY

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

Dr. Kanak Saxena et al. [12] developed a data mining model for effective heart disease prediction. It primarily helps doctors to make effective decisions based on the given parameters. The author used the Cleveland dataset from UCI and used age, gender, resting blood pressure, chest pain, serum cholesterol, fasting blood sugar, etc. as attributes. Furthermore, they divided the datasets into two parts, one is for testing and the other is for training. They used the 10-fold method to find the accuracy.

AWAIS NIMAT et al. [1] proposed an expert system based on two support vector machines (SVM) to effectively predict heart disease. These trailing SVMs have a purpose; the former is used to remove unnecessary features and the latter is used for prediction. In addition, they used HGSA (hybrid gird search algorithm) to optimize these two methods. Using this model, they achieved 3.3% better accuracy than the conventional SVM models that were present before.

Deepika et al. [13] proposed predictive analytics for chronic disease prevention and control using machine learning techniques such as naive Bayes, support vector machine, decision tree, and artificial neural network and used UCI machine learning datasets to calculate accuracy. Among them, Support vector machine provides the best accuracy of 95.55%.

Ashir Javeed et al. [2] developed a model to improve heart disease prediction by overcoming the overfitting problem; overfitting means that the proposed model works and gives better accuracy to the test data and gives unfortunate accurate results for the training data in predicting heart disease. To solve this problem, they developed a model that would provide the best accuracy for both training and test data. This model consists of two algorithms, one is RAS (Random search algorithm), the other is random forest algorithm, which is used to predict the model. This proposed model gave them better results in both training data and testing data.

D.M. Chitra et al. [14] proposed a system to predict various heart diseases using the DNFS technique. DNFS stands for Decision Tree Based Fuzzy Neural System. Here, the authors proposed a system to predict heart disease using data mining techniques as well as machine learning techniques such as decision tree, naive Bayes, k-nearest neighbors, support vector machines, and artificial neural network for prediction. The authors used the Cleveland database for the 13-element prediction. In addition, they also conducted a comparative study of different algorithms and finally found that Naive Bayes and Decision Tree provide the best accuracy.

COMPARATIVE STUDY OF LITERATURE SURVEY

TABLE II.COMPARISON TABLE OF LITERATURE SURVEY

Authors	**Techniques used**	**Accuracy**
Liaqat Ali et al. [4]	X^2statistical model, deepneural 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, Naïve Bayes.	Chronic disease diagnosis between 82% and 92%

CONCLUSION

Heart disease is a very critical problem in today's growing world. Thus, there is a need for an automated system to predict heart disease at earlier stages. So, it will be useful for doctors to diagnose patients effectively and it will also be useful for people because they can track their health problems using this automated system. Some of the expert automated systems have been summarized in this post. Feature selection and prediction, these two features are essential for any automated system. By effectively selecting features, we can achieve better results in the prediction of heart diseases. We have summarized some algorithms that are useful in feature selection, such as hybrid grid search algorithm and random search algorithm, etc. So, in the future, it is better to use search algorithms for feature selection and then apply machine learning techniques for prediction will give us better results in cardiac prediction disease.

REFERENCES

[1] Rahul Katarya. and Polipireddy Srinivas."Predicting Heart Disease at Early Stages using Machine Learning: A Survey" 978-1-7281-4108-4/20/$31.00 ©2020 IEEE

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] J. Schmidhuber, "Deep Learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[9] 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.

[10] A. K. M Sazzadur Rahman, M. Mehedi Hasan, S. Asaduzzaman, 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.

[11] 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.

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

[13] Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," Procedia Comput. Sci., vol. 85, pp. 962–969, 2016, doi: 10.1016/j.procs.2016.05.288.

[14] K 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.

[15].Monika D.Rokade ,Dr.Yogesh kumar Sharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic."IOSR Journal of Engineering (IOSR JEN),ISSN (e): 2250-3021, ISSN (p): 2278-8719

[16] Monika D.Rokade ,Dr.Yogesh kumar Sharma"MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE

[17]Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2324 - 2331.

[18] Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Health Care Patient Monitoring using IoT and Machine Learning.", **IOSR Journal of Engineering (IOSR JEN),** ISSN (e): 2250-3021, ISSN (p): 2278-8719

[19]Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Data Hiding In Audio-Video Using Anti Forensics Technique ForAuthentication ", IJSRDV4I50349, Volume : 4, Issue : 5

[20] SunilS.Khatal Dr. Yogesh Kumar Sharma. (2020). Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2340 - 2346.