



Survey Paper On Different Types Of Prediction Algorithm For Heart Disease prediction And Comparative Study On Types Of algorithms Used

¹ Anju Elsa Koshy; ²Preetha S.L.; ³ Sam .G.Benjamin ;⁴ Dr.B.Radhakrishnan

¹ CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

² CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

³ CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

⁴ CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India.

ABSTRACT

Heart is a one of the most important organ of our body which is essential for sustaining our life. The main cause for huge number of death in the world over the last few decades is critical cardio vascular disease and now it rated as the most life threatening disease of the day. So, there is a need of reliable, accurate and feasible system for diagnose such heart diseases early as soon as possible proper medication and treatment. Machine learning shows effective results in making decisions and early predictions from the broad set of data produced by the health care industry. This survey paper present a various research studies, which includes machine learning algorithms it analysis of most relevant machine learning algorithms available on the literature survey. The machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), NaïveBayes, Decision Trees (DT),

Random Forest (RF). This paper is working on the algorithm with best accuracy. *Keywords-neural network,random forest, Machine learning,accuracy,decision tree.*

I. INTRODUCTION

Heart disease or coronary artery disease (CAD) is one of the severe life threatening diseases due to the complicated working processes, different symptoms and tests.The correct prediction of heart diseases is a difficult task and it led to delay in the proper treatment. Machine learning (ML) has been proven to be successful in making decisions and predictions from the large quantity of data produced by the healthcare industry. This study strongly opinioned that the main reason behind the large number of deaths is due to ill identification mechanisms at an early stage. If it predicted at an early stage would help prevent the death at great extent. There are several factors which raise the risk of heart diseases .Medical scientists have classified the factors into two categories; one

of them is risk factors that can be changed, and another one is risk factors that cannot be changed. Smoking habits, family history of heart disease, high cholesterol level, obesity, high blood pressure all these come under risk factors, and lack of physical exercise. Family history, sex, age comes under risk factors that cannot be changed. Hence, we need a reliable and efficient model to predict the risk factor very early. Now a days there are many automated techniques to detect the heart disease such as machine learning, deep learning, data mining etc. Different person body can show different symptoms of heart disease. There are a variety of different heart disease which includes,

- Coronary artery: Coronary artery happens because of the blockage of arteries in the heart.
- Arrhythmia: An arrhythmia is a heart rhythm abnormality.
- Heart infection: Heart infections may be caused by bacteria, viruses.
- Heart failure: A chronic condition in which the heart does not pump blood as well as it should.
- Vascular disease: Vascular disease happens when bloodflow reduced to the heart.
- Stroke: Interruption of blood supply occur damage to the brain

II. LITERATURE SURVEY

Senthil kumar Mohan et al [1] explains the objective to find critical heart disease by applying machine learning. That bringing about improving the exactness in the prediction of cardiovascular malady. In the study for the purpose we used the Cleveland dataset of UCL repository and all about 14 attributes of the dataset were considered. The process starts from preprocessing data phase followed by feature selection based on DT entropy, classification and model performance evaluation, and obtained the results with improved accuracy. The dataset contains an attribute named “num” for indicating the diagnosed heart disease in patients ranging from different scales ie, 0 to 4. In the target, 0 represents the absence of heart

disease and all the values from 1 to 4 represent patients with heart disease and 4 being the dangerous. The result of data pre-processing show the presence of heart disease and also absence state. Many machine learning techniques such as used for the prediction of heart disease, NB, GLM, LR, DL, GBT, DT, RF, SVM using for the prediction of heart disease.

Devansh Shah et al [2] proposed a method to predict whether a patient will develop heart disease or not. In the paper, various data mining techniques are used for effective heart disease prediction. Efficient and accurate prediction with a lesser number of attributes and tests. This paper aims to foresee scope of having heart disease as probable cause by computerized prediction of heart disease. This will be helpful in the medical field especially for medical practitioners and patients. This paper additionally depicts which attributes contribute more than the others to anticipation of higher precision. This may spare the expense of different trials of a patient, as all the attributes may not contribute such a substantial amount to expect the outcome. In this paper, they used the Cleveland dataset of UCL repository. They are using different classifier algorithm were conducted test through the WEKA data mining tool, which is in ARFF format. The author has used four data mining classification techniques, k- nearest neighbor, naive bayes, decision tree, and random forest.

Noor Basha et al [3] proposed an early detection of heart syndrome using machine learning techniques, to predict and analyze the heart related syndrome in patients. The model for heart disease patients with various machine learning algorithms such as, NB, SVM, DT, RF, and KNN algorithms and techniques to achieve accuracy and efficiency on analysis and prediction of heart syndromes. An exploratory data analysis (EDA) heart disease dataset to summarize their main characteristics with visual methods. Out of which KNN algorithm found to be that very effective and efficient performance on accuracy score on heart disease prediction.

Chaitrali S. Dangare et al [4] propose a method to predict heart disease more accurately. This paper has analyzed prediction systems for heart disease using more number of input attributes. This research paper added two more input attributes i.e. obesity and smoking are used to get more accurate result. After pre-processing, data mining classification techniques such as, Decision Tree, Neural Network, and Naïve Bayes are used. The result analysis shows that the accuracy of Neural Network is more accurate than Decision Trees, and Naïve Bayes.

A. Lakshmanarao et al [5] suggest some machine learning techniques are used for the detection of heart disease. He applied sampling techniques for handling unbalanced datasets. For the purpose he selected kaggle dataset which contains n number of patients records with about 15 features like demographic, behavioral factors, medical risk factors. Based on the attributes, machine predict whether the patient has a 10 year risk of future coronary heart disease (CHD). Various data mining methods are used to predict the overall risk. When dealing with unbalanced datasets, oversampling and under sampling are helpful to balance the samples of two different classes.

Mr. Santhana Krishnan.J et al [6] this paper mainly propounded two supervised data mining algorithm for predict the possibilities of having heart disease of a patient and analyzed it with classification model namely and Naive bayes classifier and Decision tree classification. The aim of this paper is to know whether the patient has heart disease or not. The decision tree model has predicted the heart disease patient highly accurate and naïve Bayes classifier has predicted heart disease patients with an accuracy level lower than that of the decision tree. The reason for the higher accuracy other this algorithm is this model analyses the dataset in the tree shape format.

Aditi Gavhane et al [7] proposed a method to develop an application which can predict the vulnerability of a heart disease based on basic symptoms like age, sex, pulse rate etc. The proposed system used here is multi-

layer perceptron (MLP) neural network to train and test the dataset. MLP provides its users with an uncompromising accurate prediction .The output of the system will give a prediction result if the person has a heart disease easy to understandable terms of YES or NO. The system gives an idea about the heart status leading to CAD beforehand. Due to recent advancements in technology, the machine learning algorithms are evolved a lot and hence they use multi layered perceptron (MLP) in the proposed system because of its efficiency and accuracy.

Priyanka N et al [8] this paper aims to predict the heart disease based on the dataset values stored by MYSQL database as backend server. Training will made the dataset values to predict the presence or absence of heart disease in the patient record. This paper use classification techniques like Naive Bayes and Decision Tree algorithm for the effective prediction of heart disease. It is a web based application, and found Decision tree was the best. It gave most accurate result whether the patient had the possibility of the heart disease. This system can also be used in future projects to detects the specify type of heart disease in particular. There by the diagnosis and management of heart disease can be made simpler.

Aditya Methaila et al [9] in his research work focused on using different algorithms and combinations of several target attributes for effective heart attack prediction using data mining. The different supervised machine learning algorithm such as Naive Bayes, Neural Network, Decision Tree, along with it weighted association of Apriori algorithm and MEFIA algorithm ie, Maximal Frequent Itemset Algorithm for heart disease is considered. In this study, they have used the Cleveland dataset from among the 15 attributes of the dataset. The accuracy rate of the Decision Tree and Bayesian classification further improves by applying genetic algorithm which reduce the actual data size to get the optimal subset of attribute sufficient for the best heart disease prediction.

Nidhi Bhatla et al [10] this paper aims at analyzing

various data mining techniques introduced in recent years for heart disease prediction. In this paper heart disease prediction system has been developed using about 15 attribute, 13 attributes are used for prediction, 2 more attributes obesity and smoking efficient diagnosis of heart disease. The proposed method in this research work is an extended version of the model that combines the genetic algorithms for feature

selection and fuzzy expert system for effective classification. This research work has presented the data classification based on various supervised machine learning algorithm such as, naive bayes, decision list and KNN. The result analysis shows that neural network has shown the highest accuracy among other working algorithms.

ALGORITHMS IN LITERATURE SURVEY

TABLE 1. A COMPARATIVE STUDY OF VARIOUS

YEAR	AUTHOR	PURPOSE	TECHNIQUE USED	REASON AND ACCURACY
2019	Senthil kumar Mohan [1]	Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques	a) Decision Tree b) Random Forest c) SVM	Random forest is best among others because of: Excellent prediction ability and best optimization options.
2020	Devansh Shah [2]	Heart Disease Prediction Using Machine Learning Techniques.	a) KNN b) Random Forest c) Naïve Bayes	KNN is best among other ML algorithms because of Can learn non linear decision boundaries and enormous choice of decision boundaries.
2019	Noor Basha [3]	Early Detection Of Heart Syndrome Using Machine Learning Technique	a) Random Forest b) SVM c) Naïve Bayes d) KNN	RF performed with high accuracy because there is no need of normalization and its training is very faster.
2012	Chaitrali S. Dangare [4]	Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques	a) Neural Network b) Decision Tree c) Naïve Bayes	Neural Network is produce best result here because Neural Networks have the ability to learn by themselves and produce the output that is not limited to the input provided to them. The input is stored in its own networks instead of a database, hence the loss of data does not affect its working.
2019	A. Lakshmanar	Machine Learning Techniques For Heart Disease Prediction	a) Random Over Sampling b) Synthetic Minority Over Sampling c) Adaptive synthetic sampling approach	For random oversampling, SVM given the best accuracy (82.30%) For synthetic Minority over sampling Random Forest Accuracy (91.3%) For Adaptive synthetic sampling, Random forest given the Accuracy (90.3%)
2019	Mr.Santhana	Prediction of Heart	a) Naïve Bayes	Decision Tree (91%) produce best

	Krishnan.J [6]	Disease Using Machine Learning Algorithms	b)Decision Tree	result as compared to naive bayes because of its some disadvantages: NB algorithm faces the ‘zero-frequency problem’ where it assigns zero probability to a categorical variable whose category in the test data set wasn’t available in the training dataset.
2018	Aditi Gavhane [7]	Prediction of Heart Disease Using Machine Learning	a)Decision Tree b) KNN c)K-mean clustering d) Adaboost	Decision Tree (86.60%) is best because In decision tree data pre-processing step requires very less time. The working concept behind decision tree is more familer to programmers and comparatively easier to understand and implement than other similar algorithms
2017	Priyanka N [8]	Usage Of Data Mining Techniques In Predicting The Heart Disease Naïve Bayes And Decision Tree	a) Naïve Bayes b)Decision Tree	Decision Tree (92.21%) produce marginal efficiency over naive bayes because : Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.
2014	Aditya Methaila [9]	Early heart disease prediction using data mining techniques	a)Decision Tree b) Naïve Bayes c)Neural Network	Decision Tree is best over naive bayes because: Naive Bayes implicitly assumes that all the attributes are mutually independent. There is no specific rule for determining the structure of a neural network. The appropriate network structure is achieved through experience and trial and error.
2012	Nidhi Bhatla [10]	An Analysis of Heart Disease Prediction using Different Data Mining Techniques	a) Decision Tree b) Neural Network c) Naïve Bayes	Decision Tree predict accurate result (93.62%) over other algorithms because: It is easy to grasp because it follows a constant method that somebody follows whereas creating any call-in real-life.

III. 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. This paper studied different types of machine learning algorithms for prediction of heart disease. This paper elaborated various machine learning algorithms and worked towards finding the best algorithm by analyzing their features. Every algorithm has given different result in different situations. Further it is analyzed a marginal accuracy is achieved by Random Forest among others. Hence more complex models are needed to increase the accuracy of predicting heart disease at early stage. In future this paper propose methodology for early prediction of heart disease with high accuracy and minimum cost and complexity.

REFERENCES

- [1.] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava —Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, —On the Compatibility of Adaptive Controllers, Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)
- [2.] Devansh Shah, Samir Patel, Santosh Kumar Bharti "Heart Disease Prediction Using Machine Learning Techniques" Published online: 16 October 2020 © Springer Nature Singapore Pte Ltd 2020.
- [3.] Noor Basha, Ashok Kumar P S, Gopal Krishna C, Venkatesh P "Early Detection Of Heart Syndrome Using Machine Learning Technique" 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT).
- [4.] Chaitrali S. Dangare, Sulabha S. Apte, —Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques; International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.
- [5.] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar," Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.
- [6.] Mr.Santhana Krishnan.J, Dr.Geetha.S," Prediction of Heart Disease Using Machine Learning Algorithms",2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT),doi:10.1109/ICIICT1.2019.8741465.
- [7.] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD)," Prediction of Heart Disease Using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).IEEE Conference Record # 42487; IEEE Xplore

ISBN:978-1-5386-0965-1.

[8.] Priyanka N, Dr.Pushpa RaviKumar "Usage Of Data Mining Techniques In Predicting The Heart Disease Naïve Bayes And Decision Tree" 2017 International Conference on circuits Power and Computing Technologies [ICCPCT].

[9.] Aditya Methaila, Early Heart Disease Prediction Using Data Mining Techniques; CCSEIT, DMDB, ICBB, MoWiN, AIAP pp. 53–59, 2014.

[10.] Nidhi Bhatla, Kiran Jyoti, " An Analysis of Heart Disease Prediction using Different Data Mining Techniques" International Journal of Engineering and Technology Vol.1 issue 8 2012.

[11.] Kaan Uyar, Ahmet İlhan –" Diagnosis Of Heart Disease Using Genetic Algorithm Based Trained Recurrent Fuzzy Neural Networks" 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception

