PREDICTION OF HEART DISEASE USING HYBRID METHOD

[1]Dr. M. RAGHAVA NAIDU, [2]Dr. M. RATNABABU
[1]Assistant Professor, Department of CSE, Krishna University College of Engineering and Technology, Machilipatnam, Andhra Pradesh, India
[2]Professor, Department of ECE, St. Ann's College of Engineering and Technology, Vetapalem Mandal, Chirala, Andhra Pradesh, India

ABSTRACT:
Heart disease is one of the main causes of death in the modern world. Cardiovascular disease prediction is a major issue in clinical data analysis. While accurately forecasting heart illness is a challenging task, contemporary Machine Learning (ML) techniques make it feasible. By putting in place a strong machine learning system, cardiovascular illnesses can be accurately predicted, human intervention can be reduced, and supplementary medical tests may be avoided. The severity and death rate of the condition can be decreased with this kind of evaluation. Few research demonstrate how machine learning methods might be used to predict heart disease. This research presents a hybrid method (Support Vector Machine and Decision Tree) for utilizing Machine Learning (ML) approaches to increase the accuracy of cardiovascular disease prediction. The prediction model is introduced using famous categorization algorithms and a multitude of feature combinations. A mixed machine learning (ML) prediction model for heart disease offers higher accuracy and better performance.

KEY WORDS: Heart Disease prediction; Cardiovascular Disease Prediction; Naïve Bayes (NB); Machine Learning (ML); Support Vector Machine (SVM).
DOINumber:10.48047/nq.2022.20.2.NQ22360 NeuroQuantology2022;20(2):652-658

I.INTRODUCTION

The heart is one of the most important organs in the body susceptible to Cardiovascular Diseases (CVD). Heart and blood diseases generally refer to cardiovascular diseases that occur as a result of coronary heart disease and are considered heart attack, stroke, heart failure and other pathological conditions. The cause of millions of deaths is CVDs. Since the age of 50, the United Kingdom has experienced the greatest increase in death rates [1].

Other medical conditions like diabetes, high blood pressure, etc. are discovered when the heart is unable to circulate enough blood throughout the body.

An Electronic Health Record (EHR), or EHR, is a device that can identify a patient's record. The medical staff may easily use this information because it is safe, current, and patient-centered. This facilitates the discovery of hidden information and establishes a connection between patient data that may be applied to clinical and research procedures. This procedure assists in the elimination of traditions.

With an estimated 17.9 million fatalities per year, cardiovascular diseases (CVDs) are the leading cause of death worldwide. Coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other conditions are included in the group of heart and blood vessel disorders known as CVDs. Heart attacks and strokes account for four out of every five CVD deaths, with premature deaths accounting for one-third of these deaths in people under the age of 70.

Age, sex, smoking, family ancestry, cholesterol, less than stellar eating routine, hypertension, corpulence, physical inertia, and liquor admission are viewed as hazard factors for heart disease , and inherited

hazard factors, for example, hypertension and diabetes likewise lead to heart disease [2]. Some hazard factors are controllable. Aside from the above components, way of life propensities, for example, dietary patterns, physical latency, and heftiness are additionally viewed as significant hazard factors [3]. Among them, the tobacco biting, undesirable eating regimen, physical dormancy and liquor are the essential driver of heart ailments. Analysts are utilizing an assortment of classes of numerical information mining instruments that are existing in the investigation of heart illnesses [4].As well as being risk factors for CVD, obesity and being overweight also raise blood pressure, glucose levels, and cholesterol. In settings for primary care, these can all be easily measured. Premature deaths can be prevented by identifying those who are most at risk for CVDs and ensuring they receive the proper treatment. To ensure treatment and assistance for those in need, all primary care settings must have access to essential medications and fundamental health knowledge for non-communicable diseases.

Many peoples do not recognize that they are at risks, and millions of people around the world experience difficulties controlling risk factors that contribute to stroke and heart attack. Through lifestyle modifications and the use of medication when needed, controlling main risk factors can lower the incidence of heart attacks and strokes [5]. CVD risk factors include hypertension, high blood cholesterol, and high blood sugar or glucose levels. Those physical traits are related to fundamental social and behavioural issues like elderly, wealth, and urbanization. Using tobacco, poor diet, excessive alcohol use, and insufficient exercise are all risk factors.

The symptoms of heart diseases can currently be predicted using a range of Machine Learning techniques. Among these algorithms are Decision Tree, Logistic Regression, KNN, Navy Base, and SVM [6]. These algorithms divide into several classes based on the models that are created, using restricted data such as age, sex, etc. This technique main contribution is the ability to predict heart failure. To data below shows, a variety of Machine Learning approaches are utilised, including decision trees, naïve base logistic regression, and K-Nearest Neighbour (KNN). The primary objective of this analysis is to enhance the accuracy of heart disease prediction. The results of numerous investigations have led to feature selection constraints that can be applied in algorithm development. The suggested hybrid strategy has a better capacity to predict cardiac disease than existing approaches, according to the findings of the experiments.

II. LITERATURE SURVEY

Guardian et al. [7] Multiple physiological signal sensors, an Arduino microcontroller, and the Internet of Things (IoT) idea are used to develop a cardiac patient monitoring system. Methods used are SVM, random forest, simple logistic models platform (WEKA) and Arduino based microcontroller system. Whether compared to different approaches and technologies for continuously sensor monitoring, support vector machines have the highest accuracy, as determined by this methodology. The design verifies an integrative cardiac patient monitoring system that could be utilized in the patient's home environment while also providing the patients with an online centralised monitoring system based on the findings. Real-time data cannot be transmitted to the server through a blood pressure sensor module based on PhotoPlethysmoGraphy (PPG) or an electronic sphygmomanometer.

S. Mohan et. al [8] Prediction algorithm was created to give doctors and other healthcare professionals individualised information. The methodologies utilized are K-Nearest Neighbors, support vector machine, logistic regression and random forest. Their research revealed that the SVM model's accuracy is 0.8947 its sensitivity is 0.9434 its specificity is 0.7826, and is 0.8868. The RF model is the

most sensitive, whereas the KNN model is the least sensitive. CAD has higher accuracy, sensitivity and specificity, and than other models, making it more effective and accurate. The method may need more research to enhance the effectiveness of the Machine Learning algorithm before it can be applied as a clinical solution, according to the constraints discovered.

Maruf Ahmed Tamal et al. [9] A Machine Learning approach has been used to develop a model for heart disease prediction that has a 95% accuracy rate. Methods used decision tree, svm, naive bayes, random forest, logistic regression, QDA. The techniques involved resulted in the second highest classification accuracy in SVM research, with LR and QDA having the highest accuracy. The results support the conclusion that SVM performs more accurately than any other approaches. The system's shortcomings include the exclusion of crucial elements like menopause, which have a significant impact on heart disease.

Based on the findings, SVM outperformed all other approaches in terms of accuracy. The system's shortcomings include the exclusion of crucial elements like menopause, which have a significant impact on heart disease.

S. M. M. Hasan, M. A. Mamun, M. P. Uddin, and M. A. Hossain et.al, [10] worked with the Cleveland Heart Disease dataset of 14 attributes and 303 records downloaded from the UCI machine learning repository. They utilized information gain as a feature selection technique to omit unnecessary features and different classification algorithms like KNN, Decision Tree (ID3), Gaussian Naive Bayes, Logistic Regression, and Random Forest to predict coronary illness. After analyzing the results, they found that Logistic Regression yielded maximum accuracy with than all the other algorithms in the research outcomes.

K. C. Howlader, and A. Mazumder M. S. Satu, et al. [11] utilizing 116 records and 34 variables gathered from three Dhaka, Bangladesh, hospitals. Seven different methods were used to categorize the diseases.J48, Heoffding Tree, SMO, IBk, KStar, and Naive Bayes are among them. With a low accuracy of 70.83 percent, naive Bayes performed the best in the developed framework.

Deeanna Kelley et.al [12] performed a work, to analyze and detect heart disease. In this the algorithm used was Naive Bayes algorithm. In Naïve Bayes algorithm they used Bayes theorem. Hence Naive Bayes has a very power to make assumption independently. The used data-set is obtained from a diabetic research institutes of Chennai, Tamilnadu which is leading institute. There are more than 500 patients in the dataset. The tool used is Weka and classification is executed by using 70% of Percentage Split. The accuracy offered by Naive Bayes is 86.419%.

Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador, et.al [13] described the Prediction for similarities of disease by using ID3 algorithm in television and mobile phone. This paper gives a programmed and concealed way to deal with recognize designs that are covered up of coronary illness. The given framework utilize information min-ing methods, for example, ID3 algorithm. This proposed method helps the people not only to know about the diseases but it can also help's to reduce the death rate and count of disease affected people.

DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem, et.al, [14] described the disease Predicting system using data mining techniques. In this paper they talk about MAFIA (Maximal Frequent Item set algorithm) and K-Means clustering. As classification is important for prediction of a disease. The classification based on MAFIA and K-Means results in accuracy.

M. Sultana, A. Haider, and M. S. Uddin et.al, [15] worked with two data sets. One is the

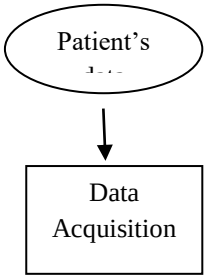
standard data set of 13 attributes with 270 records downloaded from the UCI Machine Learning Repository. Other is the dataset of 6 attributes with mere 100 records collected. After collecting the data sets, the researchers examined their heart disease prediction procedure through Weka software using KStar, J48, SMO, Bayes Net, and Multilayer Perceptron algorithms. After analyzing the results, they found that SMO and Bayes Net showed optimum performances than the performances of KStar, Multilayer Perceptron, and J48 algorithms based on performance factors in their proposed system.

III. METHODOLOGY

Hybrid machine learning for heart disease prediction is discussed in this section. Fig. 1 shows the presented system's workflow.

The severe cardiac syndrome in critically ill patients is examined and predicted using the system framework of the learning algorithms.

Heart disease data from the website should be used as input data and processed in this model. It can be anticipated that a person's risk of developing severe heart syndrome increases if they have diabetes and smoke or drink regularly. For instance, someone who smokes and drinks alcohol, and has severe diabetes is almost certain to experience acute cardiac syndrome. The dataset of 6 attributes with mere 100 records collected from Enam Medical Diagnosis Centre, Savar, Dhaka, Bangladesh. They build the site to store manual data into MySQL format and to use MySQL data to convert into CSV data format so that we can use the CSV data file in our python programming.



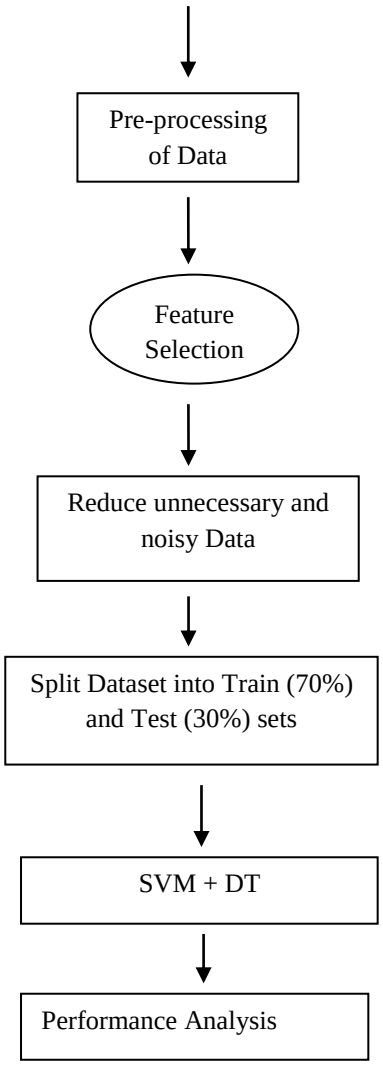


Fig.1. The Workflow Of The Hybrid System

The Cleveland heart dataset from the UCI machine learning. Analyze the dataset taken from the Cleveland heart disease database. After storing all the manual data into our website database, we checked the noisy data and eliminated them from our dataset.This analysis dataset was taken from the web, which basically supports many datasets that are open source and accessible in a variety of data formats. The data source is achieved with various attributes, where the data set is used to create an efficient model to classify and predict the severity of heart disease in critically ill patients.

As all know, high blood pressure is a major warning sign for such people, as it causes the heart to work too hard to pump blood throughout the body, causing stress and strain on the heart and damaging the blood veins and blood vessels. Age, gender, a poor or unhealthy diet, stress at work, and inherited conditions are some of the characteristics of healthy people that make them more susceptible to chronic heart syndrome. Working people or corporate employees are now the prospective victim community of major heart disease sufferers in today's society.

After gathering multiple records, pre-processing of heart disease data occurs. Six patient records out of a total of 303 in the sample had missing data. Following the removal of those six records from of the dataset, pre-processing was conducted on the remained 297 patient information. For only the dataset's properties, a multiclass variables and a binary classification are implemented. To determine whether heart disease is present or not, the multi-class variable is used. The value is set to 1 when the patient has heart disease; otherwise, it is set to 0 when the patient doesn't have heart disease. By converting diagnosis values from medical records, data is pre-processed. The data pre-processing outcomes for the 297 patient records show that 137 patient information with a value of 1 have cardiac problems, while 160 patient information with a value of 0 have no heart disease.

The feature selection and modeling keep on repeating for various combinations of attributes. From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease.As previously mentioned, this experiment makes use of a variety of (ML) techniques, including NB, LR, DT, KNN, and SVM. The experiment was conducted again using all 13 attributes and ML techniques.

The dataset contains 100 records and it is divided into training and test sets once it has been finalised. 70% of the data was left over for developing the machine learning model. The final model is tested using the remaining 30% of the data once the model has been trained. Later, Machine Learning (ML) was applied using these split datasets. The following are the algorithms: 1. logistic regression, 2. decision tree, 3. naive bayes, 4. K Nearest Neighbor (KNN), and 5. support vector machine (SVM). The Integrated Development Environment that we used was Anaconda Jupyter Notebook (IDE). We utilised the following Python libraries: Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn, and Scikit-Learn. To divide the datasets between testing and train sets, the pandas package was used to extract our data file, NumPy was used to perform mathematical calculations, and Matplotlib and Seaborn were used to show the model output. The five machine learning algorithms listed below are now being utilized to forecast heart disease:

Decision Tree: Trees are generated utilizing high entropy inputs for data D training samples. These trees are built rapidly and easily using the top-down recursive Divide-And-Conquer (DAC) approach. On D, tree trimming was done to get rid of unnecessary samples.

Logistic Regression: The probability idea serves as the foundation for the predictive analysis technique known as Logistic Regression. A Logistic Regression model is comparable to a Linear Regression model, but it makes use of a more complex cost function. The range of the cost function in the logistic regression hypothesis is between 0 and 1, and it is sometimes referred to as the "Sigmoid function" or the "Logistic Function." Implementing linear functions correctly is impossible since they can have a value lower than 0 or higher than 1.

Support Vector Machine (SVM): Both classification and regression tasks use the Support Vector Machine (SVM), a supervised ML algorithm. But classification issues frequently make use of it. They represent each unique data item as a position in an n-dimensional space (where n is the number of features you have) in this method to apply SVM, in which the estimate of each components is the estimation of a certain coordinates. At that moment, classification is accomplished by locating the hyperplane that divides the 2 classes as a whole.

The performance analysis of this hybrid system is evaluated in terms of Accuracy, and Sensitivity.

IV. RESULTS ANALYSIS

This section discusses the hybrid Machine Learning for analysing the results and predicting cardiac disease.

Table 1: Performance Analysis

ML classifiers	Accuracy	Sensitivity
Logistic Regression	82.9	86.1
Hybrid SVM + DT	97.2	94.3

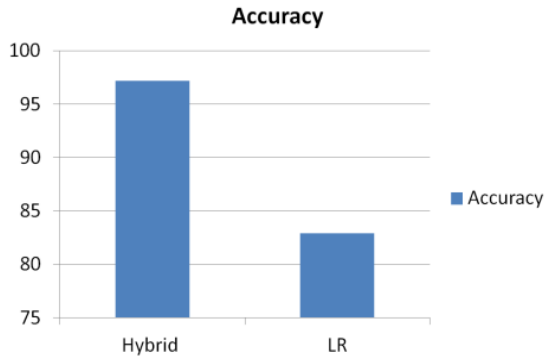


Fig. 2: Accuracy Comparison Graph

Fig.2 shows comparative graph of accuracy for LR and hybrid model for prediction of heart disease.

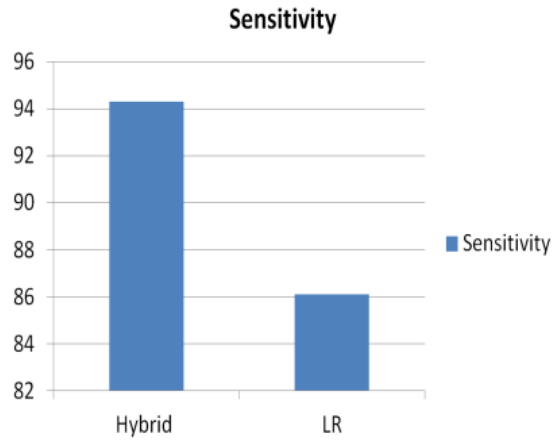


Fig.3: Sensitivity Comparison Graph

The comparative graph of the sensitivity for the hybrid model for heart disease prediction and the LR is shown in Fig. 3.

V. CONCLUSION

Gaining insight into the analysis of unprocessed cardiac healthcare data can help ensure long-term preservation and early identification of issues with heart illnesses. The method provided a novel viewpoint on cardiac disease by applying machine learning techniques to interpret raw data. Heart disease evaluation is a crucial and challenging task in medicine. On the other hand, early illness identification and preventative actions can significantly lower the death rate. This investigation has to be extended further in order to concentrate on real-world datasets instead of only theoretical techniques and simulations. Combining features from Support Vector Machines (SVM) and Decision Trees (DT) is the recommended Hybrid Machine Learning technique. It has been demonstrated that hybrid machine learning is incredibly accurate in predicting cardiac disease. In this research, they suggest a hybrid method that uses machine learning techniques to identify important highlights, hence increasing the accuracy of the cardiovascular disease prediction. A few well-established grouping techniques and several highlight combinations are provided to the forecast model. Hence, this model achieves

better results interms of accuracy and sensitivity.

VI. REFERENCES

[1] Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dha-vachelvan, P., Venkatachalapathy, V.S.K., "Recursive ant colony optimization routing in wireless mesh network", (2016) Advances in Intelligent Systems and Computing, 381, pp. 341-351.

[2] Subhashini, R., KeerthiSamhitha, B., Mana, S.C., Jose, J,"Data Analytics To Study The Impact Of Firework Emission On Air Quality: A Case Study", AIP Conference Proceedings. 2019

[3] Mana, S.C., Samhitha, B.K., Jose, J., Swaroop, M.V., Reddy, P.C.K., "Traffic violation detection using principal component analysis and viola jones algorithms", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019

[4] Amin Mohammed Shafennor, et al. Identification of Significant features and datamining techniques in predicting heart disease. Telematics Inf 2019:82–93.

[5] Ramandeep Kaur, Er. Prabhsharn Kaur "Heart Disease Forecasting Pattern using Various Data Mining Techniques" (June 2016)

[6] M. A. Jabbar, P. Chandra, and B. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 628–634, Nov. 27-29, 2012, doi: 10.1109/ISDA.2012.6416610.

[7] The Guardian." UK heart disease fatalities on the rise for first time in 50 years".https://www.theguardian.com/society /2019/may/13/heart_x0 002_circulatory-disease-fatalities-on-rise-in-uk. Accessed 25 Oct 2019.

[8] S. Mohan: "Effective Heart Disease Prediction Using Hybrid ML Techniques", VOLUME 7, 2019,DOI 10.1109/ACCESS.2019.2923707

[9] Maruf Ahmed Tamal 2019, Heart Disease Prediction based on External Factors: A Machine Learning Approach, International Journal of Advanced Computer Science and Applications(IJACSA), Volume 10 Issue 12, 2019.

[10] S. M. M. Hasan, M. A. Mamun, M. P. Uddin, and M. A. Hossain, "Comparative Analysis of Classification Approaches for Heart Disease Prediction," *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Feb. 8-9, 2018, doi: 10.1109/IC4ME2.2018.8465594.

[11] K. C. Howlader, M. S. Satu, and A. Mazumder, "Performance Analysis of Different Classification Algorithms that Predict Heart Disease Severity in Bangladesh," 2017 International Journal of Computer Science and Information Security (IJCSIS), vol. 15, no. 5, pp. 332-340, May, 2017, doi: 10.1109/CEEICT.2016.7873142.

[12] Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal

[13] Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial" in JMIR Cardio 2017, vol. 1, issue 2, pg no:1

[14] DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "The Utilisation of Machine Learning Approaches for Med-ical Data Classification" in Annual Conference on New Trends in Information & Communications Technology Applications - march-2017

[15] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Sept. 22-24, 2016, doi: 10.1109/CEEICT.2016.7873142.