IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

Effective Heart Disease Diagnosis Using Machine Learning Techniques

Sumanth M [1] and Adarsh M J[2]
Student, Department of Computer of Application[1].
Assistant professor, Department of Computer of Application[2].
Jawaharlal Nehru New College of Engineering, Shivamogga, Karnataka, India
sumanthsumanth908@gmail.com and adarshmj@jnnce.ac.in

Abstract: *Heart disorders, also known as Cardiovascular Diseases (CVDs), are one of the world's worst problems. They are invisible and strike without warning when the body's limits are reached. Every day, the health-care business generates a vast amount of data about patients and conditions. Researchers and workers, on the other hand, do not make proper use of this information. The healthcare sector today is rich in data but deficient in competence. There are a variety of data processing and machine learning approaches and tools available for extracting useful information from databases and using that information to make more accurate predictions and decisions. The main idea of this research paper is to summarize recent research on heart disease prediction with comparative results, as well as to draw analytical conclusions and reduce the amount of attributes utilized in heart disease diagnosis, which may result in fewer tests being required. that a patient must undergo. Our effort also aims to improve the proposed system's efficiency. Electrocardiograms are a non-invasive and low-cost approach to diagnose cardiac problems (ECG).*

Keywords: Cardiovascular disease, ECG, Data mining, Electrocardiogram, Support Vector Machines, K-Nearest Neighbor; Naïve Bayes, Decision Trees.

I. INTRODUCTION

The term "heart disease" refers to a variety of illnesses that damage your heart. The word "heart disease" refers to a variety of disorders, including blood vessel disorders like coronary artery disease, heart rhythm abnormalities (arrhythmias), and birth defects (congenital heart defects), among others. Cardiovascular disease is another term for heart disease. Cardiovascular disease (CVD) is a term used to describe conditions in which blood vessels are restricted or obstructed, which can result in a heart attack (myocardial infarction), chest discomfort, or stroke. Various heart disorders, such as those affecting your cardiac muscle, valves, or rhythm, are also classified as other types of heart disease's claim the lives of 17.9 million people each year, accounting for almost 31% of all deaths worldwide. Nowadays, the healthcare industry generates a huge amount of data about patients, disease diagnoses, and other topics; yet, researchers and practitioners do not appropriately utilise this data. Today, one of the most important challenges facing the healthcare business is service quality (QoS). QoS entails accurately forecasting disease and providing patients with effective remedies. Poor forecasting can result in unacceptably severe consequences. There are numerous risk factors for heart disease. Some risk factors are uncontrollable, such as family history, ethnicity, advancing age, and being male. However, smoking, high BP, high cholesterol, diabetes, inactivity and obesity are all preventable problems.

II. LITERATURE SURVEY

The survey is one such method of scientific investigation that enables researchers to gather information from a large population. (Generalization). During the survey process, the students faced a number of difficulties.

In [1], This study focuses on a number of data mining methods that are essential for identifying locally prevalent diseases like heart disease, lung cancer, and breast disease, among others, using pharmaceutical data mining. Information mining is a technique for extracting data that seeks out inactive examples and is used to assess and pinpoint cardiac disease. In this case, the Naive Bayes approach was used. In the Nave Bayes algorithm, the Bayes theorem was used. Naive Bayes has a lot of power when it comes to making assumptions on its own as a result. The study's data set was gathered from

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

a renowned diabetes research facility in Chennai, Tamil Nadu. There are more than 500 patients in the dataset. The tool used is Weka

[2]The dataset's data has undergone preliminary processing in order to support categorization. It is recommended to use a convolution neural network technique to create effective classification rules. In order to classify medical data, the neural network is trained using the convolutions method. A multilayer perceptron with a specific architecture is used in the convolution neural network technique to recognize two-dimensional picture content. Input, convolution, sample, and output should always be added as additional layers. The convolution layer and sample layer in a deep network design may also have numerous iterations. The proposed model has an accuracy of 87.26% and a 78.70% prediction rate.

In [3], Utilizing information from two ECG leads, the procedure is carried out independently, and the final classification is determined by averaging the two classification findings. For the purpose of providing dynamic features, RR interval statistics is also computed. A SVM classifier is used to classify heartbeats into one of 16 categories using these two different types of attributes together. Distinct Lead Signals have different test beat classifications based on accuracy.

The research of [4] A CNN network that had been properly trained received ECG signals directly. From outpatient ECG scans, the repository received more than 4000 ECG signals. The testing set's confusion matrix reveals that the "regular" class's average accuracy was 99 percent. 98.34% of the sample was sensitive, and 98.36% of the sample was specific.

Similarly, [5] One of the leading causes of death worldwide is heart illness or anomalies. A sound signal called a phonocardiogram is produced by the mechanical beating of the heart (PCG). The use of an ECG is often expensive and time-consuming. In this study, a non-segmentation method for information extraction from a PCG signal is presented. To remove DC and limit the frequency to the required range, the signal must first go through pre-processing. Then, four characteristics and four classifiers are used to identify heart murmur sound from PCG data.

III. PROPOSED METHODOLOGY

The proposed approach was created with the purpose of distinguishing between persons who have cardiac disease and those who are healthy. A variety of machine learning predictive models for heart disease were assessed to see how well the selected features performed. Significant features are chosen using feature selection algorithms, The efficiency of the classifiers was evaluated based on these features extracted. The Detroit heart disease dataset and the UCI heart disease dataset have both been used in a number of studies, including ours. The method used the common machine learning models logistic regression, K-NN, SVM, decision tree, and random forest. The verification and performance assessment metrics for the model are calculated. The suggested system's methodology is divided into many steps, involving patient data gathering, attribute selection, dataset pre-processing, machine learning model use, and classifier measuring performance methods.

We suggested a method for detecting heart disease using machine learning approaches in this research, and the findings demonstrated a significant accuracy standard for delivering a considerably good approximation outcome. A accurate determination of healthcare datasets was and continues to be a major machine learning problem the figure 2 shows the accuracy chart and figure 3 shows the accuracy graph of the various models.

3.1 System Design

A. ECG Signal Classification

With a frequency spectrum of 0.01-150Hz, the ECG is a symmetrical low frequency weak signal. For many decades, data collecting techniques have been introduced to reliably sense, gather, record, and assess the physical status of all the patients [20]. Any small alterations in any part of the ECG waveform might cause a variety of disorders. Because of its inexpensive cost, the electrocardiogram (ECG) is the most used test for screening cardiac disorders. Furthermore, ECG pattern identification is frequently used as an early cardiac disease alert system

B. Support Vector Machine (SVM)

It is a method for supervised learning that divides data into two categories on a hyperplane. Support vector machine do a similar duty as C4.5, but without the need of Decision trees. To limit the possibility of misclassification, the support vector machine aims to increase the margin (the distance between the hyper plane and the two nearest data points from each respective class). Scikit-learn, MATLAB, and LIBSVM are some well-known support vector machine implementations.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

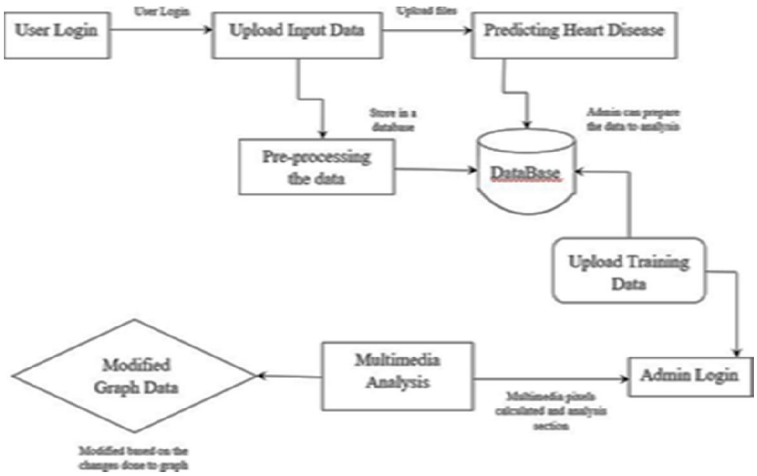

Fig 1. Flowchart representation

C. Naive Bayes

It is a straightforward method for creating classifiers. It is a Bayes' theorem-based Classifier that is probabilistic. All Naive Bayes classifiers assume that the value of any given feature is independent of the value of any other feature given the class variable. The Bayes theorem is as follows:

$$P(C|X) = P(X|C) * P(C)/P(X)$$

P(X), where X is a data tuple and C is the class, is constant for all classes. Despite the fact that it assumes an unrealistic requirement that attribute values are conditionally independent, it performs quite well on large datasets where this circumstance is expected and holds.

.

D. Decision Tree

A decision tree is a decision-making supporting tool that employs a decision-making model or graph that looks like a tree. It requires a record or object that contains a set of attributes as input and returns a "decision with anticipated output value for the input." The input attributes can be continuous or discrete. After running a number of experiments, the decision tree comes to a conclusion. A decision tree's non leaf nodes correspond to tests for the appropriate attribute value, and the node's branches are marked with the test's possible results. Within the tree, every leaf node specifies the value (decision) that will be returned if that leaf is reached. Decision tree implementation algorithms include J48, Random Forest, and Logistic Tree Model.

E. Random Forest

Random Forests are a regularization technique classifier that can be used for classification and regression (also known as closest neighbor predictor). It builds a variety of Decision trees during training and outputs the category that is the mode of the classes produced by each tree. By averaging to find a natural compromise between high variance and key tenets, it also seeks to mitigate the concerns of high variance and significant tenets. This approach is well-supported in both R and Python

Predictive model	Accuracy (%)	Sensitivity (%)	Processing time (s)
Logistic Regression	84	83	19.213
K-nearest Neighbour	76	73	29.400
Support Vector Machine	75	75	18.239
Naïve Bayes	83	78	34.101

Fig 2. Accuracy chart

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

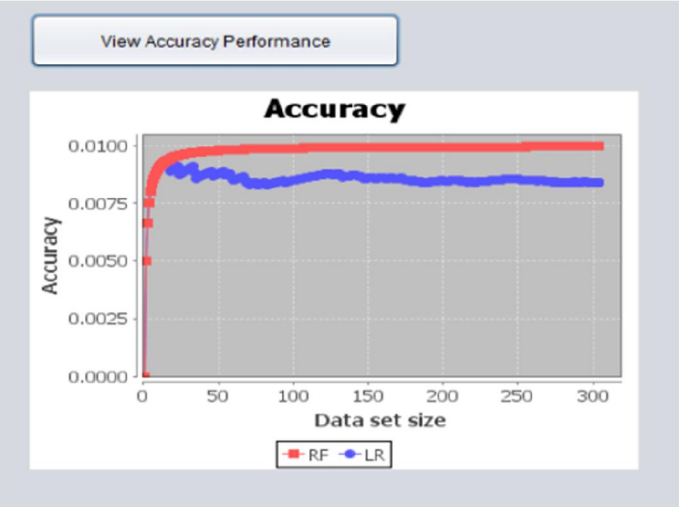

Fig 3. Accuracy graph

F. Training

It contains a total of 303 patient records. The data set is then separated into training and testing sets in a 70:30 ratio, and the machine learning models are given each set. To screen for cardiac disease, the user is permitted to enter the clinical parameters given above. After the data has been entered, it will validate it. The user can enter the correct values by clicking the corresponding buttons when the appropriate messages are displayed.

G. Prediction

The proposed method was developed to make a distinction between people who have cardiac illness and healthy people. We performed heart disease prediction in our system using the Cleveland UCI library. It provides a straightforward visual representation of the dataset, working environment, and predictive analytics architecture.

If a person has heart disease or not, machine learning algorithms can predict this based on the data that the user has provided. The 4 machine learning models' individual forecasts and all of their results are displayed.

IV. RESULTS AND DISCUSSION

Here it will predict the risk of heart disease using the model's user will login and enters the details of Blood and ECG it will shows the risk of heart and if user found risk he can also book an appointment with doctors at the required time and date here admin in the admin login admin can add doctors and he can see the feedback of the user's and admin can also add the new doctors.

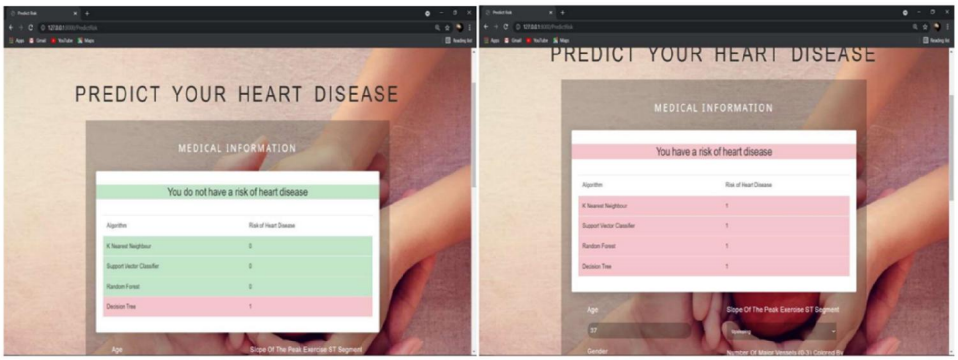

Fig 4.Result of Prediction

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, June 2022
Impact Factor: 6.252

V. CONCLUSION

By suggesting a strategy for heart disorder/ailment diagnosis through machine learning approach, fully utilized diagnosis can be obtained; the findings revealed a high precision standard for providing a superior assessment result. Well organized classification of healthcare dataset is a major machine learning problem then and now. Data mining techniques with the patient dataset and ECG signal classification helps in the detecting heart related abnormalities and diseases precisely (Above 80%). Depending on the type of input, we can also use sound signals for classification, which returns different types of heartbeats.

REFERENCES

[1]. ECG signal analysis and arrhythmia detection on IoT wearable medical devices," D. Azariadi, V. Tsoutsouras, S. Xydis and D. Soudris, 2016 5th International Conference on Modern Circuits and Systems Technologies, Thessaloniki, Greece.SSS

[2]. Automatic ECG Diagnosis Using Convolutional Neural Network" Roberta Avanzato and Francesco Beritelli. Department of Electrical, Electronic and Computer Engineering, University of Catania, Italy.Published: 8 June 2020.

[3]. ECG signal analysis and arrhythmia detection on IoT wearable medical devices," D. Azariadi, V. Tsoutsouras, S. Xydis and D. Soudris, 2016 5th International Conference on Modern Circuits and Systems Technologies, Thessaloniki, Greece.

[4]. Automatic ECG Diagnosis Using Convolutional Neural Network" Roberta Avanzato and Francesco Beritelli. Department of Electrical, Electronic and Computer Engineering, University of Catania, Italy.Published: 8 June 2020.

[5]. A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases" Mohammed Abdul Khaleel (Research Scholar Sambalpur University, India), Sateesh Kumar Pradham (P.G.Department of Computer Science ,Utkal University, India), G.N. Dash(P.G.Department of Physics,Sambalpur University, India)

[6]. S P Raghavendra, M.J.Adarsh, Shoieb Ahamed and J. Shree Hari, "Estimation of Human Age and Gender Based on LBP Features Using Two Level Decision by SVM" _ K. C. Santosh and B.Gawali (Eds.): RTIP2R 2020, CCIS 1380, pp. 82– 94, 2021. https://doi.org/10.1007/978-981-16-0507-9_8, Springer Nature Singapore.

[7]. Adarsh M J, Md. Irshad Hussian B," A Survey on Digitization of Historical Document with Image Enhancement Techniques" "- Jan. 18 | Volume XII | Issue I | Article Number: 14 | ISSN 2321-3469. In International Journal of Computer Engineering and Applications.

[8]. M J Adarsh, Steven Lawrence Fernandes, S N Bharath Bhushan, Vinlal Vinod, "A Novel Representation for Classification of User Sentiments" Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, Pages 379-386 Springer, Singapore, DOI: 10.1007/978-981-10-3156-4, FICTA-2016, Vol -2.

[9]. Adarsh. M.J., Praveen Kumar Katigar, "Neural Network Based Approach for Classification of Text Documents"- Apr 2014 | Volume number: 1 | Issue: 2 | Article Number: 9 | ISSN: 2348 –5477. Kaav International Journal of Science, Engineering & Technology.