JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# HEART DISEASE DETECTION USING MACHINE LEARNING

Swarali Dabake
Student
Computer Engineering
Pillai HOC College of Engineering & Technology
Rasayani, India

Shraddha Patil
Student
Computer Engineering
Pillai HOC College of Engineering & Technology
Rasayani, India

Smriti Kadu
Student
Computer Engineering
Pillai HOC College of Engineering & Technology
Rasayani, India

Snehal Shinde
Faculty
Computer Engineering
Pillai HOC College of Engineering & Technology
Rasayani, India

***Abstract:*** Recently, the diagnosis of heart disease has been one of the most complex tasks in the medical field. With the growing population, it becomes more and more difficult to diagnose and begin treatment early. Machine learning strategies have accelerated the health sector with more research. The machine learning model for heart rate predictor supports related parameters. The UCI heart prediction benchmark database will be used in this research project, which contains different parameters on heart condition. The machine learning algorithm which is a random forest will be used for the model event. The Random Forest algorithm will achieve much higher accuracy compared to other machine learning algorithms. This model will be useful to medical staff at their clinic as a decision support system.

**Keywords: Random Forest, Heart Disease Prediction.**

## I. INTRODUCTION

The work proposed in this paper focuses on the various machine learning techniques used to predict heart disease. According to the World Health Organization (WHO), 12 million people worldwide die each year from heart disease. The burden of cardiovascular disease has increased rapidly worldwide over the past few years. Numerous studies have been conducted to identify the most important signs of cardiovascular disease and accurately predict any risk. Cardiovascular disease, also called the silent killer, leads to death in a person without obvious symptoms. Early diagnosis of cardiovascular disease plays an important role in lifestyle decisions in high-risk patients and reduces the likelihood of complications. The project aims to predict future heart disease by using machine learning algorithms to analyze patient data that discriminates whether or not they have heart disease. Predictive heart disease is a web-based machine-learning program. The user enters his or her

specific medical information in order to obtain a predictor of that user's heart disease. The algorithm will calculate the risk of heart disease. The result will be displayed on the web page itself. The algorithm will be trained using a database. Further steps will be taken to improve the algorithms thereby improving accuracy. These steps include cleaning the database and pre-processing the data. The main web-based application that accepts various parameters from the user as input and calculation.

## II. RELATED WORK

[1] T. Nagamani et al. developed a device concept using statistical data collection strategies and algorithmic loading maps. For 45 cases in the test set, the accuracies reported in this article exceeded those obtained using standard custom neural networks. The accuracy of the algorithm used here was improved by using flexible circuits and line sizes.

[2] Avinash Goland et al. etc; Read on for a variety of machine learning algorithms that can be used to differentiate between heart disease. We looked at the selection trees, KNN, and OK algorithms available for differentiation and compared their accuracy. Studies have shown that the accuracy obtained from the selection tree is very high and suggested that similar strategies and parameter tuning could be combined to make them more efficient.

[3] Teresa Prince. R. et al. conducted a study involving a single-category algorithm used to predict coronary heart disease. Classification methods using naive bays, k closed communities (KNNs), selection tree neural communities, and divisor accuracy are analyzed using a propriety range of properties.

[4] Nagraj M. Lutimat et al., Heart disease was predicted using naive Bayes classification and a support vector machine (SVM). The overall performance size used in the analysis represents the subtext.

## III. PROPOSED MODEL

The purpose of this study was to successfully predict whether a patient had a heart condition. The health professional incorporates incoming values from the patient's health report. Data are entered into a model that predicts the risk of heart disease. Figure 1 shows the whole process involved.


Fig: Heart disease prediction model.

**A. Data collection and preprocessing.**

The database used was the heart database downloaded from the kaggle website. A full description of the 14 symbols used in the proposed event is provided in Table 1 below.

| Sl No. | Attribute Description | Distinct Values of Attribute |
|---|---|---|
| 1. | Age | Multiple values between 29 & 71 |
| 2. | Sex | 0,1 |
| 3. | CP | 0,1,2,3 |
| 4. | Blood Pressure | Multiple values between 94& 200 |
| 5. | Cholesterol | Multiple values between 126 & 564 |
| 6. | FBS-It represent the fasting blood sugar in the patient. | 0,1 |
| 7. | Resting ECG-It shows the result of ECG | 0,1,2 |
| 8. | Heartbeat | Multiple values from 71 to 202 |
| 9. | Exang- used to identify exercise induced angina. If yes=1 or else no=0 | 0,1 |
| 10. | OldPeak- It describes the patient's depression level. | Multiple values between 0 to 6.2. |
| 11. | Slope- describes patient condition uring peak exercise. It is divided into three segments(Unsloping,Flat, Down sloping) | 1,2,3. |
| 12. | CA- Result of fluoroscopy. | 0,1,2,3 |
| 13. | Thal- test is required for a patient suffering from chest pain or difficulty breathing. There are 4 types of values representing Thallium testing. | 0,1,2,3 |
| 14. | Target - This is the last column of the database. It's a class or label Column. Represents the number of classes in the database. This database has binary categories i.e. two stages (0,1). In the "0" class it means less risk of heart disease while "1" represents a higher risk of heart disease. The value of "0" or "1" depends on another 13 attribute. | 0,1 |

## B. Classification

The attributes mentioned in Table 1 are given as inputs to various ML algorithms such as Random Forest. The input database is divided into 80% of the training database and the remaining 20% to the test database. Database training is a database used to train a model. The test database is used to test the performance of a trained model. In each algorithms the performance is calculated and analyzed based on the various metrics used such as accuracy, precision, recall and F scoring scores as described further. The algorithms tested in this paper are listed below.

**Random Forest Algorithm:**

Any forest algorithm is used for editing and searching. This will cause the data tree to be created and the prediction is based on this. Random forest algorithms can be used in large databases and can generate the same results even if many recording values are not available. Samples derived from the solution can be saved for use in other data. In any forest, there are

two steps and first create an arbitrary forest and predict using any scheduling of the forest generated in the first step.

Accuracy and important flexible information can also be provided with results. A random forest is a subdivision that includes a group of formal tree dividers k, where k is independent trees, which are distributed evenly and

each random tree includes a unit of input partition vote. The random forest uses the Gini index to separate and determine the final section of each tree. The last section of each tree is grouped together and voted with the weighted values to form the final separator. Random forest operation, Random seed selection is selected, a collection of samples from the training database while class distribution is maintained.

## IV. RESULT AND ANALYSIS

Results obtained with a random forest are presented in this section. Metrics are used to analyze the performance of algorithm points for Precision, Precision(P), Recall(R), and F measure. The accuracy metric (see equation (1)) measures the good accuracy of an analysis. Remembering [mentioned in math (2)] defines the average real-time probability. The F rate [mentioned in measure (3)] for accuracy testing.

Accuracy = (TP) / (TP + FP) (1)

Remember = (TP) / (TP + FN) (2)

F– Measure = (2*Accuracy* Remember) / (Accuracy + Remember) (3) TP True positive: the patient has the disease and the test is positive.

FP False positive: the patient does not have the disease but tests do. TN True negative: the patient is free of the disease and the test is positive.

FN False negative: the patient has the disease but the test says you do not have it. In the test the previously analyzed data is used to perform the tests and the abovementioned algorithm is tested and used. The accuracy school found in Random Forest, is shown below in the Table.

| Algorithm | Precision | Recall | F- measure | Accuracy |
|---|---|---|---|---|
| Random Forest Algorithm | 1.00 | 1.00 | 1.00 | 94% |

## V. CONCLUSION

As deaths from cardiovascular disease increase, it is important to develop efficient and accurate systems for predicting cardiovascular disease. The purpose of this study was to find the most efficient machine learning algorithm for diagnosing heart disease. This study compares the accuracy of random forest heart attack prediction algorithms using the UCI machine learning database set. The results of this study show that the Random Forest algorithm is the most efficient algorithm with 100% accuracy for predicting heart disease.

# VI. REFERENCES

[1] Avinash Golande, Pavan Kumar T,"Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering, Vol 8, pp.944-950 2019.

[2] T.Nagamani, S.Logeswari, B.Gomathy," Heart Disease Prediction using Data Mining with Mapreduce Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Fahd Saleh Alotaibi," Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin"Design And Implementation Heart Disease Prediction Using Naives Bayesian", International Conference on Trends in Electronics and Information (ICOEI 2019).

[5] Theresa Princy R,J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[6] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,'Prediction Of Heart Disease using Machine Learning', International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

[7] Sayali Ambekar, Rashmi Phalnikar,"Disease Risk Prediction by Using Convolutional Neural Network",2018 Fourth International Conference on Computing Communication Control and Automation.