e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

COMPARATIVE ANALYSIS USING MACHINE LEARNING ALGORITHMS FOR PREDICTING HEART DISEASE

M R Rahul[*1], Anitha Krishnan G[*2]

[*1]Department Of MCA, SCMS School Of Technology And Management, Ernakulam, Kerala, India.

[*2]Associate Professor, Department Of MCA, SCMS School Of Technology And Management, Ernakulam, Kerala, India.

ABSTRACT

Today, the death rate has increased worldwide due to heart diseases. The better way to prevent this disease is by having a system that can foresee the symptoms or irregularities in the beginning which can save many lives. Recently research in machine learning had gained lots of attention and had been utilized in different sorts of applications including within the medical field. The researchers are able to predict the probability of getting heart diseases among susceptible patients with the help of various machine learning models. The main aim of this study is to draw a comparison among different classification algorithms that predict irregularities in the heart. In this paper, the results state that recent methodologies develop an understanding for predicting heart disease. This research paper provides the result analysis of significant machine learning models that are been used in building a highly efficient and accurate prediction model that will help doctors treat heart disease at the earliest and reduce deaths caused. For finding the efficient model this paper uses various factors like confusion matrix, recall, precision, f1-score, and accuracy.

Keywords: Machine Learning, Classification Algorithms, Accuracy, Precision, Recall, F1-Score Confusion Matrix.

I. INTRODUCTION

The work proposed in this paper is mainly a comparative study of different machine learning algorithms that are used for predicting heart disease. The heart is the center of the circulation system that regulates blood flow throughout our body. So, it is the principal part of the human body. So, if any irregularity is seen in the heart can cause distress in other parts of the body. Any kind of disturbance in the normal functionality of the heart can lead to heart problems. In today's world, the primary reason for the occurrence of many deaths is due to heart disease. An unhealthy lifestyle, alcohol, smoking, and a high intake of fat can lead to hypertension can cause heart disease. According to the WHO, more than 10 million deaths occur every year around the world because of heart disease. Keeping a healthy lifestyle and the earliest detection of irregularities are the only ways heart-related diseases can be prevented.

In the medical field, various diseases can be detected, predicted, and diagnosed with the help of machine learning. The main objective of this paper is that provides doctors a tool that could detect heart disease at an early phase. This will provide effective treatment for patients with severe consequences. Machine learning plays an important role in detecting the discretely hidden patterns and thereby analyzing the given data. After analyzing the data, machine learning algorithms help in early diagnosis and heart disease prediction. This paperwork presents the performance analysis of different classification algorithms such as Logistic Regression, K-Nearest Neighbors, Decision Tree, and Random Forest using various factors such as confusion matrix, recall, precision, f1-score, and accuracy for predicting heart disease at an early stage.

Machine Learning is the method of automatically learning and improve from the experience with the existing data and, building usable models for its prediction. The field of machine learning has not only been helpful in medical sciences but also this trend has a fast-track path of change. Nowadays, the healthcare industry delivers complex information that is broad in measure with respect to diagnosis, disease prognosis, patients, and medical health care equipment. This huge amount of data that are needed to be filtered enable us to extract useful and helpful information that can be useful to detect these heart diseases at an early stage to avoid disastrous consequences.

II. METHODOLOGY

The project is based on heart disease prediction based on machine learning. So, different classification algorithms are compared with various factors for finding the accuracy to find the effective algorithm among the four algorithms.

Heart disease prediction undergoes various stages: -

1) Classification Algorithms

a. Logistic Regression
b. K-Nearest Neighbors (KNN)
c. Decision Tree
d. Random Forest

2) Factors used for calculation

a. Confusion Matrix
b. Precision
c. Recall
d. F1-Score
e. Accuracy

III. MODELING AND ANALYSIS

To initiate this work, we have started collecting data in each and every aspect towards the goal of the system. In the first place, the research was in the direction of the main causes or the factors which have a strong influence on heart health. Some factors are unmodifiable like age, sex, and family background but there are some parameters like blood pressure, heart rate, etc. which can be kept in control by following certain measures. Many doctors suggest keeping a healthy diet and regular exercise to have a healthy heart. The parameters which are considered for the study in designing the system which has major risk percentages with relevance CAD.

The next step was to collect a dataset. For this, we have taken the dataset from the UCI Machine Learning Repository. The dataset contains as many as 76 parameters describing the complete health status of the heart. These parameters are obtained by expensive clinical tests like ECG, CT scan, etc. Out of these, the traditional heart disease prediction system uses 14 major parameters. Since these parameters require expensive lab tests to find ECG, chest pain type, ST depression, etc. To avoid these and to make the system less complex we selected the above-mentioned parameters can easily be measured using different sensors available in the market. Then these features are applied to the proposed models: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest. Then accuracy has been calculated to find the efficiency for different classification algorithms. The result is been evaluated on the basis of precision, recall, and f1-score. At last, we retrieve the best algorithm based on efficiency for the given datasets.

IV. RESULTS AND DISCUSSION

The classification algorithms are compared with each other to find the best among the five algorithms. The efficiency of the proposed models is calculated in terms of confusion matrix, recall, precision, f1-score, and accuracy.

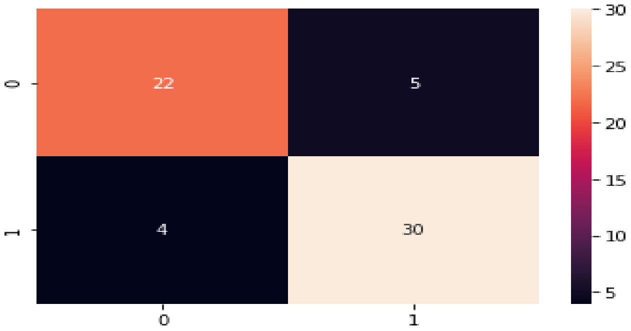

Figure 1: Confusion Matrix of Logistic Regression

Table 1. Accuracy of Logistic Regression

Heart Disease Prediction	Precision	Recall	F-Score
	0.86	0.88	0.87
Accuracy			0.85

The above Figure 1 shows the confusion matrix of Logistic Regression for Heart Disease Prediction. The result has been obtained on the basis of precision, recall, and f1-score. Table 1 shows the accuracy of Logistic Regression that has been calculated from the above-mentioned parameters.

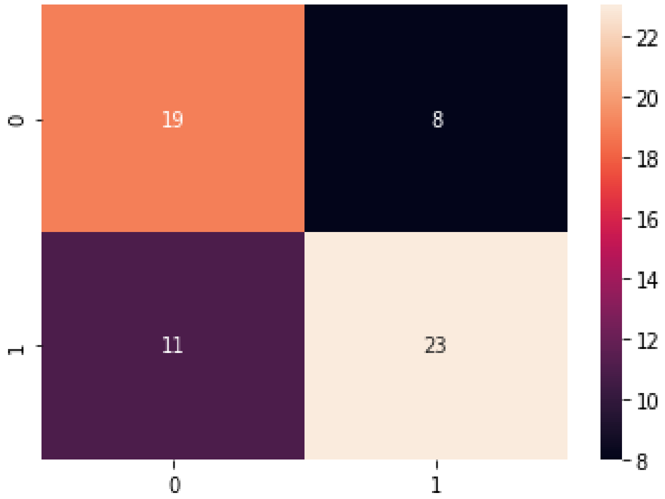

Figure 2: Confusion Matrix of K-Nearest Neighbors

Table 2. Accuracy of K-Nearest Neighbors

Heart Disease Prediction	Precision	Recall	F-Score
	0.74	0.67	0.70
Accuracy			0.63

The above Figure 2 shows the confusion matrix of K-Nearest Neighbors for Heart Disease Prediction. The result has been obtained on the basis of precision, recall, and f1-score. Table 2 shows the accuracy of K-Nearest Neighbors that has been calculated from the above-mentioned parameters.

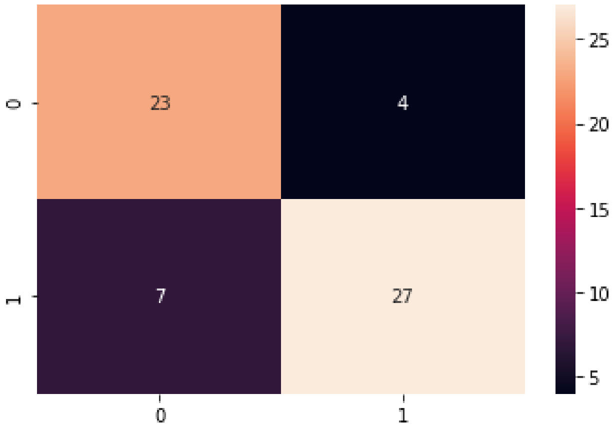

Figure 3: Confusion Matrix of Decision Tree

Table 3. Accuracy of Decision Tree

Heart Disease Prediction	Precision	Recall	F-Score
	0.87	0.79	0.83
Accuracy			0.78

The above Figure 3 shows the confusion matrix of Decision Tree for Heart Disease Prediction. The result has been obtained on the basis of precision, recall, and f1-score. Table 3 shows the accuracy of Decision Tree that has been calculated from the above-mentioned parameters.

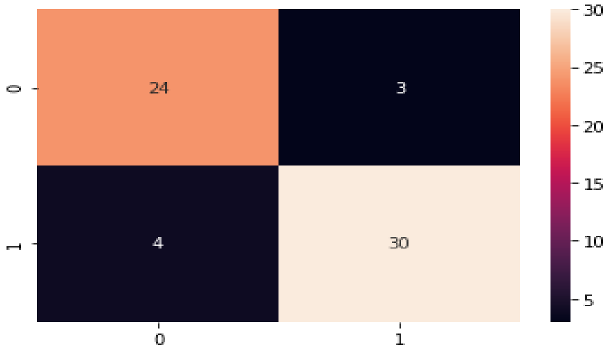

Figure 4: Confusion Matrix of Random Forest

Table 4. Accuracy of Random Forest

Heart Disease Prediction	Precision	Recall	F-Score
	0.90	0.88	0.89
Accuracy			0.88

The above Figure 4 shows the confusion matrix of Random Forest for Heart Disease Prediction. The result has been obtained on the basis of precision, recall, and f1-score. Table 4 shows the accuracy of Random Forest that has been calculated from the above-mentioned parameters.

Table 5. Summarized result of the four Machine Learning models

Models	**Precision**	**Recall**	**F-Score**	**Accuracy**
Logistic Regression	0.86	0.88	0.87	0.85
KNN	0.74	0.67	0.70	0.63
Decision Tree	0.87	0.79	0.83	0.78
Random Forest	0.90	0.88	0.89	0.88

The main aim of this project was to compare and analyze the various machine learning algorithm in terms of their recall, precision, f1-score, and accuracy. The expected results of this project are to find the accurate classification model for predicting the heart disease at earliest.

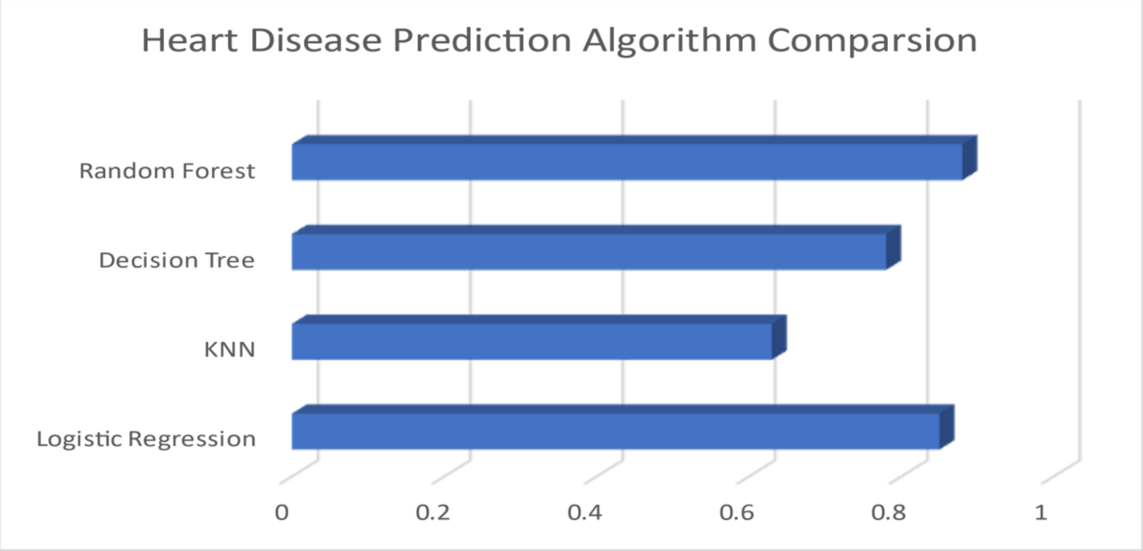

Figure 5: Comparison of all the four Machine Learning models

From these, we can conclude that the accurate model for predicting the heart disease is Random Forest because by comparing the algorithm based on the precision, recall and f1-score its accuracy is 88 % and the second best is Logistic Regression which is followed by Decision Tree. The least accurate is KNN.

V. CONCLUSION

In this project, I have predicted heart disease using four Classification Algorithms: Logistic regression, K-Nearest Neighbors (KNN), Random Forest and Decision Tree. I have pre-processed the data with respect to the given requirement. Then applied the above-mentioned models for finding the effective algorithm among the four. The results obtain concludes that Random Forest is the best model to predict heart disease with an accuracy of 0.88 which is followed by Logistic Regression with an accuracy of 0.85. KNN is 0.63 which is the least accurate algorithm among the four.

VI. REFERENCES

[1] Senthil Kumar Mohan and Chandrasekar Tirumala “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” IEEE, Issue 19, June 2019

[2] Sheik A Abdullah and R Raja Laxmi.” Data mining Model for predicting the coronary heart disease using Random Forest Classifier”. IJCA Proceedings on International Conference in Recent trends in Computational Methods, Issue April 2012

[3] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease", Proc. Int. Conf. Comput. Appl. (ICCA), pp. 306-311, Sep. 2017

[4] SP Rajamhoana and C Akalya Devi “Analysis of Neural Networks Based Heart Disease Prediction System” Department of Information. Technology. PSG College of Technology Issue in June 2018

[5] Cheng A and W. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database", Proc. 39th Annual IEEE Eng. Med. Biol. Soc. (EMBC), pp. 2566-2569, Jul. 2017

[6] F. Dammak and A. M. Alimi, "The impact of criterion weights techniques in multi-criteria decision making in crisp and intuitionistic fuzzy domains”, Fuzzy Syst. (FUZZ-IEEE), Aug. 2015

[7] V. Revathi, "Prediction of heart disease using MLP algorithm", *Int. J. Sci. Technol. Res.*, vol. 5, no. 8, pp. 235-238, 2015

[8] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining”, (IJAEMS) June- 20117 vol-10

[9] Krishnaiah G "Heart disease prediction system using data mining techniques and intelligent fuzzy approach: A review", *Int. J. Computer. Appl.*, vol. 136,2016

[10] Sona wane J S and Patil D, "Prediction of heart disease using multilayer perceptron neural network", *Proc. Int. Conf. Inf. Common. Embedded System.,* Feb. 2014

[11] R Singh” Detection of coronary artery disease by reduced features and extreme learning machine” Medicine and Pharmacy Reports, vol. 91 2018

[12] Y. Nataliani, “A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy,” IEEE Transactions on Fuzzy Systems, vol. 26, no. 2, pp. 817–835, 2018