

MACHINE LEARNING APPLICATION IN CARDIOLOGY.PREDICTION OF HEART DISEASE

Sukanya Kokani*¹

*¹UG Student, Department Of Information Technology, B.K Birla College Kalyan
Empowered Autonomous Status, Kalyan, Maharashtra, India.

ABSTRACT

Heart disease is a general but dangerous disease that term to a type of disease that affects the heart or blood vessels. Due to this millions of people die every year. The annual number of deaths from cardiovascular diseases in India is projected to rise by 620 million people living with cardiovascular diseases. Heart disease has been the leading cause of death worldwide over the past 10 years. This research paper presents a comparative study of two popular Machine learning algorithms, the Decision Tree algorithm and the Random Forest algorithm. The study aims to compare the results of both algorithms and determine which one of them predicts the diseases early and accurately. It will be an important consideration for the medical professional when making a clinical decision. Furthermore, the results of this research contribute to the ongoing efforts aimed at developing accurate and predictable models for predicting heart disease. These findings will provide valuable assistance to the healthcare department by leveraging technology

Keywords: Machine Learning, Heart Disease Prediction, Decision Tree Algorithm, Random Forest Algorithm, Machine Learning Algorithm.

I. INTRODUCTION

Health is the basic human right, and care that improves the lives of the citizens will improve a count. Research concludes that the burden due to Heart diseases in India is remarkable that what is being experienced at a global level. Heart disease along with other illnesses, still causes a lot of deaths and health problems around the world. Heart disease prediction is a complex task that involves the analysis of numerous patient attributes such as age, gender, cholesterol level, blood pressure, and various medical tests. Machine learning techniques outperform traditional statistical modes when it comes to capturing complex patterns within the above data. Decision Tree algorithm a simple yet powerful algorithm, offers transparency and insights into the decision-making process, making it a valuable asset in healthcare applications. Random Forest an ensemble learning method, leverages the strength of multiple decision trees to improve predictive performance while maintaining interpretability. Early and accurate prediction of heart diseases is crucial for effective preventive measures and timely medical precautions. In recent years, Machine Learning algorithms have emerged as valuable tools for predictive modeling in healthcare. This research paper explores the application of two algorithms, the Decision Tree algorithm and Random Forest, to predict heart disease with a focus on enhancing accuracy and interpretability. The improvement of health is the prevention, diagnosis, treatment, and cure of diseases. The research is divided into 5 Sections; Section 1 primary objectives of this study are to employ the Decision Tree and Random Forest algorithm to develop a prediction model for heart disease identification, and secondly, to compare the performance of these models against established clinical risk assessment tools. By doing so, this research seeks to evaluate the feasibility of machine learning in augmenting traditional diagnostic approaches, ultimately enhancing the accuracy and reliability of heart disease prediction. This paper is organized as follows: Section 2 provides a review of related work in the field of heart disease prediction, and highlights the significance of machine learning. Section 3 presents the methodology, including data collection, preprocessing, and the implementation of the Decision Tree algorithm and Random Forest algorithm. Section 4 reports the experimental results and discusses the findings, including model accuracy and interpretability. Finally, Section 5 summarizes the conclusions drawn from this research and suggests future directions for improving heart disease prediction using machine learning techniques. The study utilizes a detailed dataset that includes various clinical attributes.

II. LITERATURE REVIEW

The topic we have chosen for the research has its primary objective is to quickly predict heart disease that will a beneficial for the patient. It will lead to fast cures of diseases after accurate prediction of the diseases. The

large cause of death of people due to this we have chosen the prediction research. The following experiment will take place in the Google platform it is cloud-grounded and provided by Google as a short-term Google collaboration it allows one to write Python code in a jupyter notebook environment. Googlecolab is particularly popular among data scientists and machine learning practitioners, it is free to access, has collaborative editing, participatory context, and most importantly no need to install anything like libraries on our local machine, as everything runs in the cloud. This model will use the patient data to make an accurate prediction of the diseases. It will be real-time monitoring it continues the use of the models has enabled real-time heart prediction and early intervention and, enhanced patient care. In this following research, we have referred to several research papers [1] this paper is a review paper they have concluded with machine learning heart diagnosis using non-invasive they observed that three different workflows for machine learning-based heart disease diagnosis are possible.[2] This paper is based on a diabetic patient's diagnosis of heart disease using ML but they have used algorithms data mining technology, naive Bayes method, and SVM using three Machine Learning methods, they concluded that the SVM model can be recommended for the classification of the diabetic dataset.

III. METHODOLOGY

The Methodology outlined in this part forms the basis for the comparative analysis of the Decision Tree and Random Forest algorithm. The following will present discuss and analyze the result of this evaluation. We have a dataset of 304 patient records to perform the experimentation. In this, dataset we have the clinical attributes like chest pain, resting electrocardiography test results, cholesterol, Fbs, etc.

The procedure of this experiment follows the following concept

Sklearn:

Scikit-learn, often abbreviated as "sklearn", is a popular Python library for machine learning. It provides a wide range of tools and algorithms for tasks such as classification, regression, clustering, dimensionality reduction, and more.

Numpy:

Numpy is the Python library that focuses on numerical and scientific computing, offering support the large arrays and matrices, as well as high-level mathematical functions for working with them.

Matplotlib:

Matplotlib is a Python library commonly used for creating static, animated, and interactive visualization and scientific computing to display data and results in a clear and informative way.

Pandas:

Pandas is a Python library designed for data manipulation and analysis. It provides data structures and functions for efficiently working with structured data, such as tables and time series. Pandas are particularly valuable for tasks like data cleaning, exploration, transformation, and analysis. It is a fundamental tool in the data science and data analysis ecosystem in Python.

Seaborn:

Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for creating informative and attractive statistical graphics. Seaborn simplifies many of the complexities of creating visually appealing plots and charts, making it a popular choice for data visualization in data analysis and data science tasks.

The following figure displays the correlation between the features positively or negatively correlated using seaborn for example if the chest pain correlates with the target, it results that the person is infected with a heart disease.

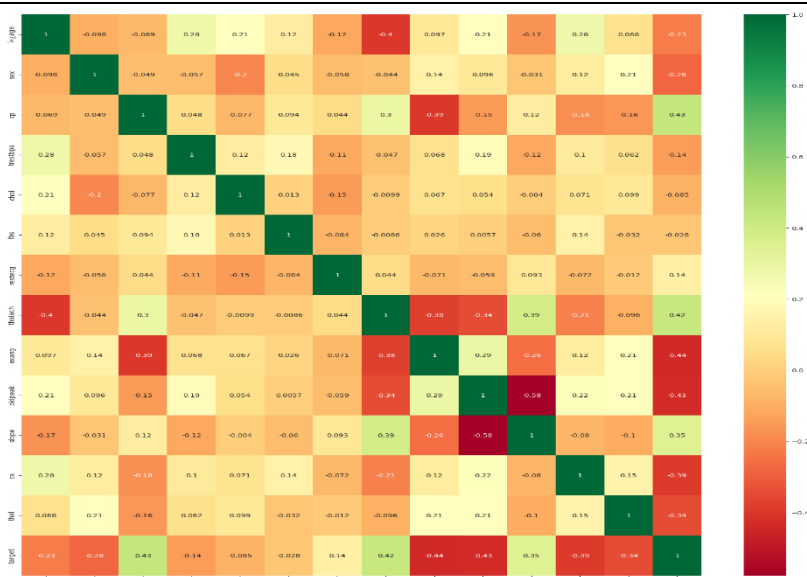


Figure 1: - correlation between features

The main practice of good research is the dataset should be balanced so here is the graph showing the balanced dataset. Here we have given the variables targets, targets include whether the person is infected with heart disease or not infected with heart disease. In the following graph target is in 1 and 0, where 1 is for the person having heart disease 0 is for the person not having heart disease.

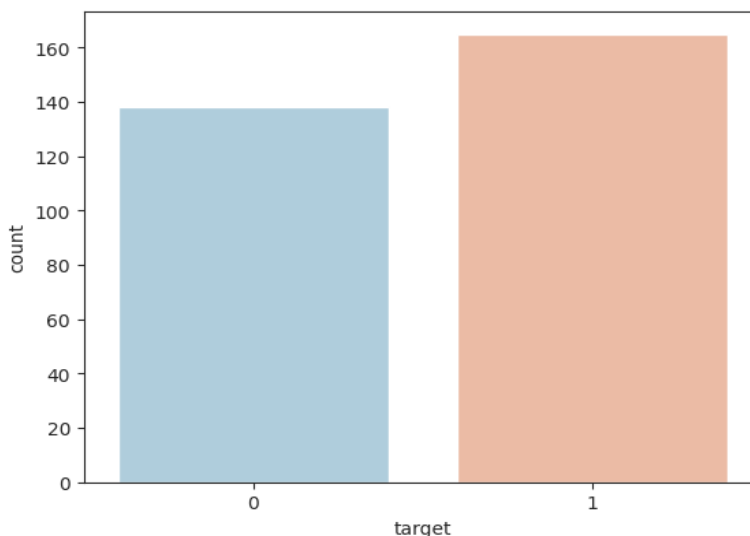


Figure 2:- balanced dataset

IV. MODELING AND ANALYSIS

Represents a decision based on a feature, each leaf node represents a decision based on a feature, and each node presents a class label. Decision trees are simple to understand and interpret, making them valuable for research.

Random Forest algorithm:

A Random Forest algorithm is an ensemble Machine learning model that is used for both classification and regression tasks. It's based on the decision tree algorithm and combines the prediction of multiple decision trees to improve accuracy and reduce overfitting. Each tree is trained on a random subset of the data and complex relationships.

Analysis

Heart disease is indeed a leading cause of death worldwide. It encompasses a range of conditions that affect the heart. Preventative measures, a healthy lifestyle, and regular medical check-ups can help reduce the risk of heart disease.

For the accurate prediction of diseases, we have decided to do a research-based experiment we have chosen a huge powerful platform for machine learning and used the machine learning algorithm Decision Tree and Random Forest

V. RESULTS AND DISCUSSION

The results we have gotten are very impressive both algorithm Decision Tree and Random Forest given nearly identical levels of accuracy, which underscores their robust performance and highlights their effectiveness in our task. Through the comparison of both the algorithms, we come to the decision the decision tree algorithm gives the most effective and highest accuracy as compared to random forest. The following table shows the accuracy rate.

Table 1: comparison of the algorithm

ML algorithm	Prediction rate
Random Forest algorithm	0.85%
Decision Tree algorithm	0.95%

VI. CONCLUSION

The research we have conducted regarding using Machine learning algorithms Random Forest and Decision Tree algorithm for the prediction of heart diseases. We performed the experiments using the Jupyter platform the results we received were that comparatively the decision tree algorithm gives 0.95% the effective and highest prediction accuracy rate than the Random Forest algorithm. To keep in consideration the lead cause of death is because of heart disease worldwide. Experiment leads to prediction decision tree algorithm gives the accuracy means it can predict the disease more accurately than the Random Forest algorithm.

In our work, we have tried to predict the chances of heart disease using medical attributes from the patient's data.

The results are highlighted in the decision tree classifier as it gives an effective accuracy rate, in the future, it can be used by the medical department to develop a heart disease predictive model.

VII. REFERENCES

- [1] N. Kumar and D. Kumar, 'Machine Learning based Heart Disease Diagnosis using Non-Invasive Methods: A Review', J. Phys. Conf. Ser., vol. 1950, no. 1, p. 012081, Aug. 2021, doi: 10.1088/1742-6596/1950/1/012081.
- [2] G. Parthiban and S. K. Srivatsa, 'Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients', Int. J. Appl. Inf. Syst., vol. 3, no. 7, pp. 25–30, Aug. 2012, doi: 10.5120/ijais12-450593.
- [3] J. Patel and D. S. Patel, 'Heart Disease Prediction Using Machine Learning and Data Mining Technique'.
- [4] S. Jiang, 'Master of Applied Statistics'.
- [5] V. Sharma, S. Yadav, and M. Gupta, 'Heart Disease Prediction using Machine Learning Techniques', in 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICANN), Greater Noida, India: IEEE, Dec. 2020, pp. 177–181. doi: 10.1109/ICACCCN51052.2020.9362842.