International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 13 Issue 4, April-2024, Impact Factor: 8.375

Heart Disease Prediction Using Machine Learning

Riya Gujar[1], Sakshi Jadhav[2], Shriparna Jadhav[3], Pranali Kapse[4], Prof. Dinesh Singh Dhakar[5]

[1,2,3,4,5] CSE in AIML, Saraswati College of Engineering, Navi Mumbai, India

ABSTRACT

Nowadays, people are getting caught in their day-to-day lives doing their work and other things and ignoring their health. Due to this hectic life and ignorance towards their health, the number of people getting sick increases every day. Moreover, most of the people are suffering from diseases like heart disease. Global deaths of almost 31% of the population are due to heart-related disease as data contributed by the World Health Organization (WHO). So, the prediction of heart disease happening or not becomes important for the medical field. However, data received by the medical sector or hospitals is so huge that sometimes it becomes difficult to analyze. Using machine learning techniques for this prediction and handling of data can become very efficient for medical people. Hence in this study, we have discussed heart disease and its risk factors and explained machine learning techniques. Using that machine learning techniques, we have predicted the rate of heart disease and provided a comparative analysis of the algorithms for machine learning used for the experiment of the prediction. The goal or objective of this research is completely related to the prediction of heart disease via a machine learning technique and analysis of them.

Keywords— machine learning, random forest, logistic regression algorithm, heart disease prediction

INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to heart disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Much research has been conducted in an attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future heart disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analyzing them to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

PROBLEM STATEMENT AND OBJECTIVE

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it is expensive, or it is not efficient to calculate the chance of heart disease in humans. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not feasible to precisely determine the possibility of heart disease occurrence for each patient everyday in all cases accurately, and consultation of a patient for 24 hours by a doctor is not available as it requires more sapience, time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

The main *objective* of doing this research is to present a heart disease prediction model for the prediction of rate of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of rate of heart disease in a patient. Hence, the algorithms are evaluated at numerous levels and types of evaluation strategies. This will provideresearchers and medical practitioners to establish a better perspective.

LITERATURE SURVEY

For a model to work effectively, appropriate algorithms must be used. The research paper by Najmu Nissa, Sanjay Jamwal and Mehdi Neshat,"A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques" published in year 2023, describes various models and their analysis in a detailed manner.[1]

An efficient cardiovascular disease monitoring system using IoT and Random Forest algorithm was proposed by Kellen Sumwiza , Celestin Twizere, Gerard Rushingabigwi ,Pierre Bakunzibake, Peace Bamurigire,"Enhanced cardiovascular disease prediction model using random forest algorithm, 2023 for cardiac disease detection and an accuracy of 98%.[2]

In B. Naga Vardhana, B. Rohitha, G. Anusha, B. Sneha Latha, Ch. Aiswarya, R. Sudha Kishore, "CARDIAC DISEASE PREDICTION USING RANDOM FOREST WITH LINEAR MODEL", 2023, a combined a linear model with the Random Forest method was presented with a fresh strategy. By combining the best features of both techniques, this hybrid model seeked to improve prediction accuracy without sacrificing interpretability. [3]

Among the various algorithms, the research paper by Bhagyesh Randhawan, Ritesh Jagtap, Amruta Bhilawade , Durgesh Chaure, "Heart Disease Prediction Using Logistic Regression Algorithm"(2022)determined how well the logistic regression algorithm performs in predicting cardiovascular disease.[4]

The Debabrata Swain, Badal Parmar, Hansal Shah, Aditya Gandhi, Manas Ranjan Pradhan, Harprith Kaur and Biswaranjan Acharya Cardiovascular Disease Prediction using Various Machine Learning Algorithms"(2022) paper shows analysis of five different modelLogistic regression, Support Vector Machine (SVM), MLP Classifier with Principal Component Analysis, Deep Neural Network, and finally Bootstrap aggregation using Random Forest. [5]

In Kompella Sri Charan1, Kolluru S S N S Mahendranath2, "Heart Disease Prediction Using Random Forest Algorithm", 2022,the most efficient ML algorithm for the detection of heart diseaseswas discussed. A brief explanation of each machine learning algorithm is done. [6]

PROPOSED SYSTEM

The working of the system starts with the collection of data and selecting the important attributes. Then, the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied, and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. The following modules are used in the implementation of this system:

A. Collection of Dataset

Initially, we collected a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model.

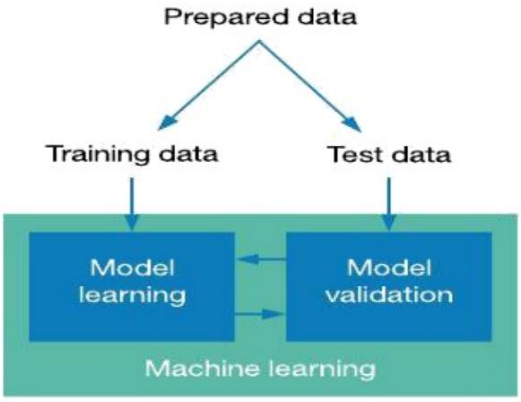

Fig. IV.A.1 Collection of Data

B. Selection of attributes

Attribute or Feature selection includes the selection of appropriate attributes for the prediction system. This is used to increase the system's efficiency. Various attributes of the patient like gender, chest pain type, fasting blood pressure, serum cholesterol, etc. are selected for the prediction.

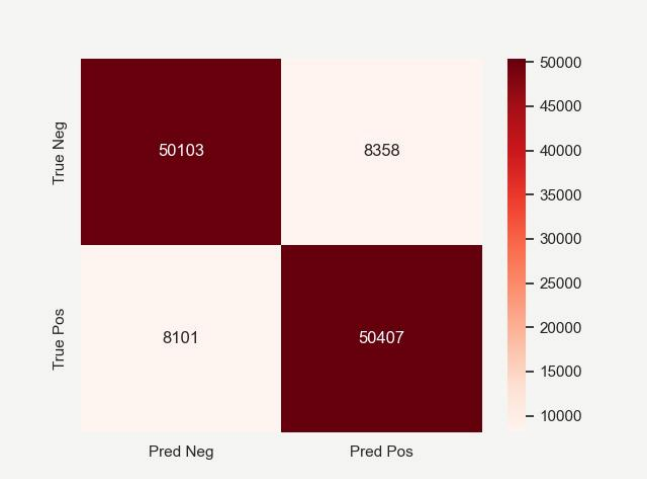

Fig. IV.B.1 Confusion Matrix

C. Pre-processing of Data

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.

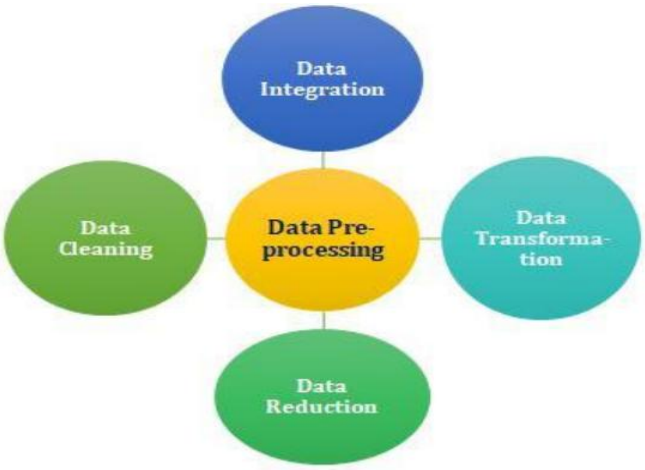

Fig.IV.C.1 Data Pre-processing

D. Balancing of Data

Imbalanced datasets can be balanced in two ways. They are Under Sampling and Over Sampling-

- Under Sampling: In Under Sampling, dataset balance is done by the reduction of the size of the ample class. This process is considered when the amount of data is adequate.

- Over Sampling: In Over Sampling, dataset balance is done by increasing the size of the scarce samples. This process is considered when the amount of data is inadequate.

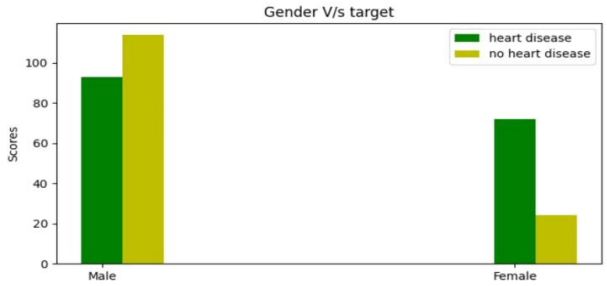

Fig. IV.D.1 Data Balancing

E. Prediction of Disease

Various machine learning algorithms like SVM, Naive Bayes, Decision Tree, Random Tree, Logistic Regression are used for classification. In our model, we decided to use Logistic Regression and Random Tree. Comparative analysis is performed among these two and the algorithm that gives the highest accuracy is used for heart disease prediction.

RESULT AND ANALYSIS

The results of our project displayed that the performance of both the algorithms revealed significant variations in predicting the rate of heart disease. Logistic Regression demonstrated moderate accuracy, with an accuracy score of approximately 90%. However, its precision and recall rates were relatively lower compared to Random Forest. Random Forest exhibits higher accuracy and outperforms Logistic Regression in terms of precision and recall. The ensemble nature of Random Forest enables it to capture complex relationships within the data, resulting in improved predictive performance and anticipating better outcomes. Also, the model, using the patient's details, gives an almost accurate prediction whether the patient has chances of heart disease or not.

ACKNOWLEDGMENT

A project is something that could not have been materialized without co-operation of many people. This project shall be incomplete if I do not convey my heartfelt gratitude to those people from whom I have got considerable support and encouragement.

It is a matter of great pleasure for us to have a respected **Prof. Dinesh Singh Dhakar** as my project guide. We are thankful to her for being a constant source of inspiration.

We would also like to give our sincere thanks to **Prof. Shraddha Subhedar, H.O.D, Computer Science & Engineering in Artificial Intelligence and Machine Learning** Department, **Prof. Dinesh Singh Dhakar**, **Project co-ordinator** for their kind support.

We would like to express our deepest gratitude to **Dr. Manjusha Deshmukh,** our principal of Saraswati College of Engineering, Kharghar, Navi Mumbai.

Last but not the least I would also like to thank all the staffs of Saraswati college of Engineering (Computer Science & Engineering in Artificial Intelligence and Machine Learning) for their valuable guidance with their interest and valuable suggestions brightened us.

CONCLUSION

In this study, the application of promising technology like machine learning for the initial prediction of heart diseases has the potential to make a profound impact on society, given the alarming prevalence of heart diseases in India and worldwide. Early prognosis of heart disease can play a pivotal role in facilitating lifestyle changes for high-risk patients and subsequently reducing complications, marking a significant milestone in the field of medicine. As the number of individuals afflicted with heart diseases continues to rise annually, the imperative for early diagnosis and treatment becomes increasingly apparent. The utilization of suitable technological support in this context can prove highly beneficial to both the medical community and patients. Attribute selection was crucial to enhance efficiency, as the inclusion of all features led to decreased system performance. The correlation among certain features in the dataset resulted in their removal. By comparing the accuracies of both the machine learning methods, an Extreme Gradient Boosting classifier emerged as the most accurate, achieving an accuracy of 81%.

REFERENCES

[1] Najmu Nissa, Sanjay Jamwal and Mehdi Neshat,"A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques", 2023.

[2] Kellen Sumwiza, Celestin Twizere, Gerard Rushingabigwi, Pierre Bakunzibake, Peace Bamurigire,"Enhanced cardiovascular disease prediction model using random forest algorithm", 2023
[3] B. Naga Vardhana, B. Rohitha, G. Anusha, B. Sneha Latha, Ch. Aiswarya, R. Sudha Kishore, "CARDIAC DISEASE PREDICTION USING RANDOM FOREST WITH LINEAR MODEL", 2023.
[4] Yu Qiu, "Logistic Regression Model based on heart disease and Its Potential Influencing Factors", 2023.
[5] Aminu Bashir Suleiman, Stephen Luka and Muhammad Ibrahim, "CARDIOVASCULAR DISEASE PREDICTION USING RANDOM FOREST MACHINE LEARNING ALGORITHM", 2023.
[6] Nadikatla Chandrasekhar and Samineni Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization", 2023.
[7] Dr. Vadhri Suryanarayana, Dr. Satyabrata Dash, Dr. Shameena Begum, Sujata Chakarvarty, Y. Nagendra Kumar, K. Venkatesh, "Heart Disease Prediction Using Machine Learning Techniques.", 2022.
[8] C.B.M Karthi, A. Kalaivani, "Heart Disease Prediction Based on Age Detection Using Logistic Regression over Random Forest", 2022
[9] Bhagyesh Randhawan, Ritesh Jagtap, Amruta Bhilawade , Durgesh Chaure, "Heart Disease Prediction Using Logistic Regression Algorithm", 2022
[10] Debabrata Swain, Badal Parmar, Hansal Shah, Aditya Gandhi, Manas Ranjan Pradhan, Harprith Kaur and Biswaranjan Acharya Cardiovascular Disease Prediction using Various Machine Learning Algorithms",2022.
[11] Kompella Sri Charan, Kolluru S S N S Mahendranath, "Heart Disease Prediction Using Random Forest Algorithm", 2022.
[12] B. Manoj Kumar1, Uma Priyadarsini, "Accuracy Analysis of Heart Disease Prediction using Logistic Regression in Comparison with the Linear Regression Algorithm",2022.
[13] C.B.M.Karthi, A. Kalaivani, " Heart Disease Prediction Based On Age Detection Using Novel Logistic Regression Over K-Nearest Neighbor ", 2022.
[14] C.B.M Karthi, A. Kalaivani, " Heart Disease Prediction Based on Age Detection using Novel Logistic Regression over Decision Tree ", 2022.
[15] G. Pavithraa, Sivaprasad, " Analysis and Comparison of Prediction of Heart Disease Using Novel Support Vector Machine and Logistic Regression Algorithm ", 2022.
[16] R.Vasanthi, S. Nikkath Bushra, K.Manojkumar, N.Suguna, "Heart Disease Prediction Using Random Forest Algorithm", 2022.

[17] P. Prasanna Sai Teja, Veeramani T, "Improving the Efficiency of Heart Disease Prediction Using Novel Random Forest Classifier Over Support Vector Machine Algorithm", 2022.
[18] P. Prasanna Sai Teja,,Veeramani T, "Comparing the Efficiency of Heart Disease Prediction using Novel Random Forest, Logistic Regression and Decision Tree And SVM Algorithms",2022.
[19] Garg, Apurv & Sharma, Bhartendu & Khan, Rizwan, "Heart disease prediction using machine learningtechniques", 2021.
[20] S. Zaman and R. Toufiq,"Codon based back propagation neural network approach to classify hypertension gene sequences," in Proc. Int. Conf. Elect., Comput. Commun. Eng. (ECCE), Feb. 2017.
[21] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease" in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017.
[22] Pahwa K, Kumar R.,"Prediction of heart disease using hybrid technique for selecting features".4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON). IEEE, 2017.
[23] Pouriyeh S, Vahid S, Sannino G, De Pietro G, Arabnia H, Gutierrez J. A "Comprehensive investigation and comparison of machine learning techniques in the domain of heart disease"IEEE symposium on computers and communications (ISCC). IEEE, 2017.
[24] S. Radhimeenakshi, ''Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural network,'' in Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom), New Delhi, India, Mar. 2016.
[25] Deepika K, Seema S. "Predictive analytics to prevent and control chronic disease"2nd international conference on applied and theoretical computing and communication technology (iCATccT). IEEE, 2016.
[26] Otoom AF, Abdallah EE, Kilani Y, Kefaye A, Ashour M. "Effective diagnosis and monitoring of heart disease", 2015.
[27] A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012.
[28] Anamta Siddiqui, Syed Wajahat Abbas Rizvi, "Heart Disease Prediction by using Random Forest Classifier",2023.