Prediction of Cardiac Disease Using Machine Learning Techniques

N Sarada Kumari [1], Mr.K.Subhash Chandra [2], P.Punnarao [3]
Student[1, 3], Assistant.Professor[2]
Amritha Sai Institute of Science and Technology Paritala-521180
Autonomous NAAC with A Grade,Andhra Pradesh,India

ABSTRACT

In the medical realm, diagnosing heart disease presents a formidable challenge due to the reliance on extensive clinical and pathological data for decision-making. This complexity has spurred heightened interest among researchers and clinical professionals in achieving efficient and accurate heart disease prediction. Timely diagnosis of heart disease is paramount, given its status as a leading cause of widespread mortality. Thus, early detection holds significant importance. Machine learning has emerged as a reliable and evolving tool in the medical domain, offering invaluable support in disease prediction through meticulous training and testing processes. This study aims to explore various prediction models for heart disease, focusing on selecting crucial features using the Random Forest algorithm. Known for its high accuracy, Random Forests is a supervised machine learning algorithm that outperforms other counterparts such as logistic regression. Leveraging the Random Forests algorithm, we endeavor to predict whether an individual is likely to have heart disease or not.

Keywords: Heart Disease Diagnosis, Machine Learning in Medicine, Random Forest Algorithm, Early Disease Prediction, Clinical Data Analysis

I. INTRODUCTION

The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood vessels, for example, veins, arteries, and capillaries. These blood vessels deliver blood all over the body. Abnormalities in normal blood flow from the heart cause several types of heart diseases which are commonly known as cardiovascular diseases (CVD). Heart diseases are the main reasons for death worldwide. According to the survey of the World Health Organization (WHO), 17.5 million total global deaths occur because of heart attacks and strokes. More than 75% of deaths from cardiovascular diseases occur mostly in middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are

because of stroke and heart attack . Therefore, prediction of cardiac abnormalities at the early stage and tools for the prediction of heart diseases can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases.

Due to the development of advance healthcare systems, lots of patient data are nowadays available (i.e. Big Data in Electronic Health Record System) which can be used for designing predictive models for Cardiovascular diseases. Data mining or machine learning is a discovery method for analyzing big data from an assorted perspective and encapsulating it into useful information. “Data Mining is a non-trivial extraction of implicit, previously unknown and potentially useful information about data”. Nowadays, a huge amount of data pertaining to disease diagnosis, patients etc. are generated by healthcare industries. Data mining provides a number of techniques which discover hidden patterns or similarities from data.

Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart disease prediction system which was validated on two open access heart disease prediction datasets. Data mining is the computer based process of extracting useful information from enormous sets of databases. Data mining is most helpful in an explorative analysis because of nontrivial information from large volumes of evidence

.Medical data mining has great potential for exploring the cryptic patterns in the data sets of the clinical domain.

These patterns can be utilized for healthcare diagnosis. However, the available raw medical data are widely distributed, voluminous and heterogeneous in nature .This data needs to be collected in an organized form. This collected data can be then integrated to form a medical information system. Data mining provides a user-oriented approach to novel and hidden patterns in the Data The data mining tools are useful for answering business questions and techniques for predicting the various diseases in the healthcare field. Disease prediction plays a significant role in data mining. This paper analyzes the heart disease predictions using classification algorithms. These invisible patterns can be utilized for health diagnosis in healthcare data.

Data mining technology affords an efficient approach to the latest and indefinite patterns in the data. The information which is identified can be used

by the healthcare administrators to get better services. Heart disease was the most crucial reason for victims in the countries like India, United States. In this project we are predicting the heart disease using classification algorithms. Machine learning techniques like Classification algorithms such as Random forest, Logistic Regression are used to explore different kinds of heart based problems.

II. LITERARURE SURVY

Heart disease prediction has garnered considerable attention in medical research, with studies such as [1] focusing on clinical decision support systems employing weighted fuzzy rules for risk level prediction. Meanwhile, [2] conducted an analysis comparing various data mining techniques for heart disease prediction, underscoring the importance of algorithm selection. Additionally, [3] provided an overview of predictive data mining techniques in medical diagnosis, emphasizing their relevance to heart disease prediction. Further enhancing predictive models, [6] utilized genetic algorithms for feature subset selection, while [7] proposed a hybrid technique in data mining classification. Moreover, [8] contributed insights into the diagnosis of coronary artery stenosis using data mining, while [9] explored heart disease classification with the nearest neighbor classifier. [10] Presented a prediction system based on the Naive Bayes algorithm, while [11] investigated heart disease diagnosis using the K-Nearest Neighbor algorithm. [12] Proposed a heart disease prediction system leveraging data mining methods, complementing [13]'s comprehensive study on prediction using machine learning and data mining.

III. PROBLEM STATEMENT

EXISTING SYSTEM:

Clinical decisions often rely on doctors' intuition and experience rather than leveraging the wealth of data available in medical databases. This practice can introduce biases, errors, and unnecessary costs, ultimately impacting the quality of patient care. Medical misdiagnoses can manifest in various ways, leading to severe consequences for patients. According to the National Patient Safety Foundation, 42% of medical patients believe they have experienced a missed diagnosis or medical error. Patient safety is sometimes overshadowed by concerns such as the cost of medical procedures and medications. However, medical misdiagnoses pose a significant risk to the healthcare profession and can instill fear in patients seeking treatment. Addressing this issue requires raising awareness among the public and holding accountable any medical

practitioners responsible for misdiagnoses through legal action and claims.

Early-stage prediction of heart disease remains a challenge, with current systems unable to reliably forecast the condition. Moreover, the practical utilization of collected data within existing systems proves to be a time-consuming endeavor, hindering swift decision-making processes in clinical settings. Furthermore, any inaccuracies or faults in the prediction process, whether attributed to medical professionals or hospital staff, carry significant risks, potentially leading to fatal outcomes for patients. Additionally, the extensive and costly procedures required to ascertain a patient's susceptibility to heart disease place a considerable burden on healthcare resources and personnel, highlighting the urgent need for more efficient and accessible predictive models in healthcare.

PROPOSED SYSTEM:

Proposed system, detailing the components, techniques, and tools utilized in its development. To create an intelligent and user-friendly heart disease prediction system, an efficient software tool capable of handling large datasets and facilitating the comparison of multiple machine learning algorithms is essential. Once the most robust algorithm with the highest accuracy and performance metrics is identified, it will be integrated into the development of a smartphone-based application designed to detect and predict heart disease risk levels. Additionally, hardware components such as Arduino or Raspberry Pi, various biomedical sensors, a display monitor, and a buzzer are required to construct a continuous patient monitoring system.

Methodology:

Logistic Regression is a widely used statistical technique for predicting binary outcomes, where the predictions are expressed as probabilities of an event occurring. It models data points using the standard logistic function, which yields results ranging between 0 and 1. The assumptions of logistic regression include a binary dependent variable, with one factor level representing the desired outcome, and the inclusion of only meaningful variables. Additionally, the independent variables should be independent of each other, and logistic regression typically requires large sample sizes. While commonly used for binary variables, logistic regression can also be applied to categorical dependent variables with more than two classes, known as Multinomial Logistic Regression.

On the other hand, Random Forest is a supervised learning algorithm used for classification and regression tasks, predominantly for classification problems. It operates by creating decision trees on data samples and obtaining predictions from each tree, eventually selecting the best solution through voting. This ensemble method is advantageous over single decision trees as it mitigates overfitting by averaging the results. The working of Random Forest involves several steps, starting with the selection of random samples from the dataset, followed by the construction of decision trees for each sample and obtaining prediction results from each tree. Subsequently, voting is performed for each predicted result, and the most voted prediction results are selected as the final prediction outcome.

IV. RESULTS & DISCUSSION

Using Random forest algorithm we got an accuracy of 75.40.

Fig-1. Accuracy of Logistic Regression

Using SVC algorithm we got an accuracy of 80.32 and using K-Neighbours an accuracy of 78.68 is achieved.

Fig-2. Accuracy of SVC & K-Neighbour

Using Decision Tree algorithm we got an accuracy of 75.40 and using Random Forest an accuracy of 885.24 is achieved and an accuracy of 80.32 with Gradient Boost Classifier.

Fig-3. Accuracy of DT, RF, GBC

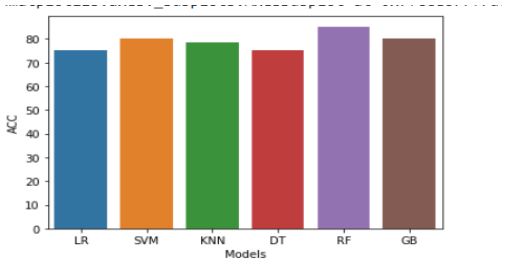

Fig-4. Accuracy Comparison of Algorithms

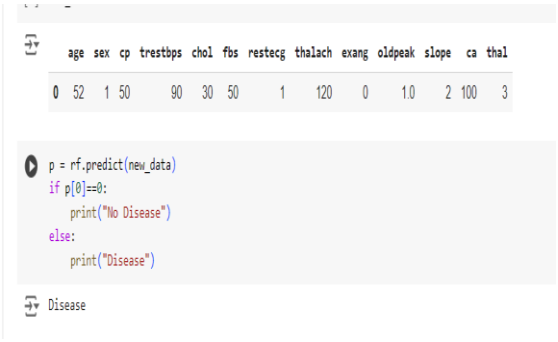

Fig-5. Prediction using RF algorithms

VI. CONCLUSION

In this paper, we introduce a heart disease prediction system utilizing different classifier techniques, specifically Random Forest and Logistic Regression. Through our analysis, we have determined that Random Forest provides better accuracy compared to Logistic Regression. Our primary objective is to enhance the performance of the Random Forest classifier by eliminating unnecessary and irrelevant attributes from the dataset, focusing solely on the most informative features for the classification task.

VII. REFERENCE

[1] P .K. Anooj, ―Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules‖; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40. Computer Science & Information Technology (CS & IT) 59

[2] Nidhi Bhatla, Kiran Jyoti"An Analysis of Heart Disease Prediction using Different Data Mining Techniques".International Journal of Engineering Research & Technology

[3] Jyoti Soni Ujma Ansari Dipesh Sharma, Sunita Soni. "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction".

[4] Chaitrali S. Dangare Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888)

[5] Dane Bertram, Amy Voida, Saul Greenberg, Robert Walker, "Communication, Collaboration, and Bugs: The Social Nature of Issue Tracking in Small, Collocated Teams".

[6] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, ―Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm‖; International Journal of Engineering Science and Technology, Vol. 2(10), 2010.

[7] Ankita Dewan, Meghna Sharma," Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", 2nd International Conference on Computing for Sustainable Global Development IEEE 2015 pp 704-706. [2].

[8] R. Alizadehsani, J. Habibi, B. Bahadorian, H. Mashayekhi, A. Ghandeharioun, R. Boghrati, et al., "Diagnosis of coronary arteries stenosis using data mining," J Med Signals Sens, vol. 2, pp. 153-9, Jul 2012.

[9] M Akhil Jabbar, BL Deekshatulu, Priti Chandra," Heart disease classification using nearest neighbor classifier with feature subset selection", Anale. Seria Informatica, 11, 2013

[10] Shadab Adam Pattekari and Asma Parveen," PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624, Vol 3, Issue 3, 2012, pp 290-294.

[11]C. Kalaiselvi, PhD, "Diagnosis of Heart Disease Using K-Nearest Neighbor Algorithm of Data Mining", IEEE, 2016

[12] Keerthana T. K., " Heart Disease Prediction System using Data Mining Method", International Journal of Engineering Trends and Technology", May 2017.

[13] Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, ELSEVIER.Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Prediction Using Machine Learning and Data Mining July 2017, pp.2137-2159.