International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 03, ISSN No. 2455-2143, Pages 121-125
Published Online July 2023 in IJEAST (http://www.ijeast.com)

COMPARATIVE ASSESSMENT OF MACHINE LEARNING ALGORITHMS FOR HEART DISEASE PREDICTION

Vikash Sawan
Assistant Professor
Department of Computer Engineering & Applications
GLA University, Mathura, Uttar Pradesh, India

Kumari Jugnu
Research Scholar
Department of Computer Science & Engineering
National Institute of Technology, Patna, Bihar, India

Durga Prasad Roy, Rakesh Pandey
Assistant Professor
Department of Computer Science & Engineering
G. H. Raisoni College of Engineering, Nagpur .

Abstract— **Heart disease prediction is a complex and critical task in the field of medicine, considering the alarming rate at which people succumb to heart-related issues. The rapid advancements in data science offer a promising avenue for processing vast healthcare datasets. Automating the prediction process can help mitigate associated risks and enable timely patient alerts. This study leverages the heart disease dataset from the UCI machine learning repository to develop an automated prediction system. By employing various data mining techniques including Naive Bayes, Decision Tree, Logistic Regression, and Random Forest, the system aims to predict the likelihood of heart disease and categorize the patient's risk level. Through a comprehensive comparative analysis, this paper evaluates the performance of these machine learning algorithms. The experimental results underscore the superiority of the Random Forest algorithm, achieving an impressive accuracy rate of 90.16% in contrast to the other algorithms studied.**

Keywords— **Heart disease prediction, data science, machine learning, Naive Bayes, Decision Tree, Logistic Regression, Random Forest, comparative analysis, accuracy.**

I. INTRODUCTION

Heart disease remains a leading cause of mortality worldwide, necessitating accurate and efficient predictive models. The advent of data science has ushered in new opportunities to leverage extensive healthcare data for improved diagnostics and patient care. Automated prediction systems offer a proactive approach to mitigate risks associated with heart disease, enabling timely interventions. In this paper, we explore the implementation of various machine learning algorithms for heart disease prediction and risk classification.

II. RELEATED WORK

Numerous research efforts have been dedicated to the prediction of heart disease utilizing the UCI Machine Learning dataset. These studies have explored a range of data mining techniques, yielding varying degrees of accuracy. Notably, Avinash Golande and colleagues have conducted research to investigate the classification of heart disease using different machine learning algorithms [1]. Their study specifically examined the efficacy of Decision Tree, K-Nearest Neighbors (KNN), and K-Means algorithms for this purpose.

Avinash Golande et al.'s research involved a comprehensive evaluation of the aforementioned machine learning algorithms. The primary focus was on classification tasks related to heart disease prediction. The study sought to assess the accuracy of each algorithm in this context and subsequently compare their performance.

The research found that among the algorithms tested, the Decision Tree algorithm exhibited the highest accuracy in predicting heart disease. This outcome highlights the potential of Decision Tree as a powerful tool for medical classification tasks. Importantly, the study emphasized the possibility of further enhancing the efficiency of the Decision Tree algorithm through a combination of different techniques and parameter tuning. In summary, Avinash Golande and co-

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 03, ISSN No. 2455-2143, Pages 121-125
Published Online July 2023 in IJEAST (http://www.ijeast.com)

authors' work demonstrates the application of machine learning algorithms, including Decision Tree, KNN, and K-Means, for heart disease classification. The study's conclusion, which suggests the superiority of the Decision Tree algorithm, underscores the significance of algorithm selection in achieving accurate predictions. Furthermore, the research points towards the optimization potential of combining techniques and fine-tuning parameters to enhance the efficiency of the chosen algorithm.

T. Nagamani et al. [2] proposed a system that integrated data mining techniques with the MapReduce algorithm. The accuracy achieved by this system surpassed that of conventional fuzzy artificial neural networks. The use of dynamic schema and linear scaling contributed to the improved accuracy.

Fahd Saleh Alotaibi [3] designed a machine learning model that compared five algorithms: Decision Tree, Logistic Regression, Random Forest, Naive Bayes, and Support Vector Machine (SVM). Using the Rapid Miner tool, higher accuracy was achieved compared to Matlab and Weka tools. The Decision Tree algorithm exhibited the highest accuracy in this study.

Anjan Nikhil Repaka et al. [4] proposed a system that employed Naive Bayesian techniques for classification and the Advanced Encryption Standard (AES) algorithm for secure data transfer in disease prediction. The focus was on ensuring secure data transmission alongside accurate disease prediction.

Theresa Princy R et al. [5] conducted a survey that explored various classification algorithms for heart disease prediction, including Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, and Neural Network. The study analyzed classifier accuracy across different numbers of attributes.

Nagaraj M Lutimath et al. [6] performed heart disease prediction using Naive Bayes classification and Support Vector Machine (SVM). Performance measures such as Mean Absolute Error, Sum of Squared Error, and Root Mean Squared Error were used, and SVM emerged as the superior algorithm in terms of accuracy over Naive Bayes.

III. PROPOSED MODEL

The focus of the proposed work is on predicting heart disease through the utilization of four specific classification algorithms: Decision Tree, Random Forest, Logistic Regression, and Naive Bayes. The primary goal of this study is to achieve an accurate and efficient prediction of whether a patient is likely to suffer from heart disease. In this process, a health professional plays a pivotal role by inputting the relevant health report data of the patient. These input values encompass a range of clinical and demographic attributes that are indicative of heart health. The collected data is then fed into the predictive model that has been developed, which employs the mentioned classification algorithms to process and analyze the information. The predictive model's core function is to assess the input data and calculate the probability of the patient having heart disease based on the selected algorithms' analyses. Each classification algorithm contributes its unique approach to the prediction, and the ensemble of these algorithms aids in generating a more robust and accurate prediction. To illustrate the entire process, Figure 1 visually depicts the sequence of steps involved in the heart disease prediction framework. This figure provides a clear overview of how health report data is collected, processed, and utilized by the model to make a probability-based prediction regarding the presence of heart disease. In summary, the proposed work focuses on predicting heart disease by utilizing a combination of classification algorithms. The objective is to provide an effective and efficient tool for health professionals to assess the likelihood of heart disease in patients. The process involves inputting patient health report data, which is then analyzed by the predictive model employing the selected algorithms, ultimately culminating in a probability-based prediction. Figure 1 visually outlines this entire process.

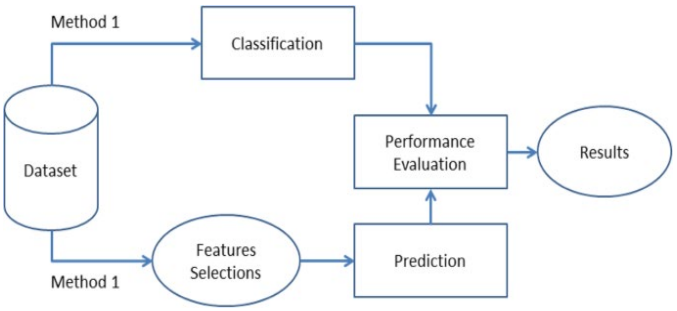

Fig. 1: Generic Model Predicting Heart Disease

Data Collection: Health report data, including clinical and demographic attributes, is collected from the patient. Data Preprocessing: The collected data undergoes preprocessing steps, which may include normalization, feature scaling, handling missing values, and data transformation.

Feature Selection: Relevant features are selected from the preprocessed data to optimize the prediction model's performance.

Classification Algorithms:

a. Decision Tree: The Decision Tree algorithm analyzes the data's attributes and splits them based on specific conditions, forming a tree-like structure to make predictions. b. Random Forest: The Random Forest algorithm constructs an ensemble of Decision Trees, where each tree contributes to the final prediction.

c. Logistic Regression: Logistic Regression calculates the probability of a binary outcome (heart disease or no heart disease) based on the input features.

d. Naive Bayes: Naive Bayes calculates the probability of a particular outcome given the input data, making predictions using Bayes' theorem.

Ensemble Prediction: The individual predictions from each algorithm are combined using an ensemble approach to improve prediction accuracy.

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 03, ISSN No. 2455-2143, Pages 121-125
Published Online July 2023 in IJEAST (http://www.ijeast.com)

Probability Calculation: The ensemble prediction yields a probability score indicating the likelihood of the patient having heart disease.
Threshold Application: A predefined threshold is applied to the probability score to classify the patient into either the "Heart Disease" or "No Heart Disease" category.

Prediction Outcome: The model presents the final prediction outcome to the healthcare professional, indicating the likelihood of heart disease based on the input data.
This generic model showcases the key steps involved in predicting heart disease using a combination of classification algorithms. The process starts with data collection and preprocessing, followed by the application of Decision Tree, Random Forest, Logistic Regression, and Naive Bayes algorithms. The ensemble prediction and probability calculation lead to a final prediction outcome, which helps guide medical professionals in assessing a patient's risk of heart disease.

Data Collection and Preprocessing: For our study, we utilized the Heart Disease Dataset, which is a compilation of data from four different databases. However, our analysis focused exclusively on the UCI Cleveland dataset. This specific dataset originally contains a total of 76 attributes, but for the purpose of published experiments, a subset of only 14 features has been consistently used [9]. In line with this, we opted to work with the UCI Cleveland dataset, which had already undergone processing, and was accessible on the Kaggle website.

The decision to use the preprocessed UCI Cleveland dataset was driven by its established relevance and compatibility with prior research efforts. The dataset's 14 selected attributes have been widely recognized as pertinent to heart disease prediction, contributing to the dataset's accuracy and reliability.

Table 1: Attributes Used in the Proposed Work

Attribute	**Description**
Age	Age of the patient
Sex	Gender (0 = Female, 1 = Male)
CP (Chest Pain)	Type of chest pain
Trestbps	Resting blood pressure
Chol	Serum cholesterol
FBS (Fasting Blood Sugar)	Fasting blood sugar > 120 mg/dl (1 = true, 0 = false)
Restecg	Resting electrocardiographic results
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina (1 = yes, 0 = no)
Oldpeak	ST depression induced by exercise
Slope	Slope of the peak exercise ST segment
CA	Number of major vessels colored by fluoroscopy
Thal	Thalassemia type
Target	Presence of heart disease (0 = No, 1 = Yes)
Attribute	**Description**
Age	Age of the patient
Sex	Gender (0 = Female, 1 = Male)
CP (Chest Pain)	Type of chest pain
Trestbps	Resting blood pressure
Chol	Serum cholesterol
FBS (Fasting Blood Sugar)	Fasting blood sugar > 120 mg/dl (1 = true, 0 = false)
Restecg	Resting electrocardiographic results
Thalach	Maximum heart rate achieved

This table outlines the 14 attributes used in our study. These attributes encompass various patient characteristics and medical indicators that have been shown to be relevant for heart disease prediction. Leveraging this preprocessed dataset allows us to focus our analysis on the specific attributes that have demonstrated significant predictive value in prior research.

IV. CONCLUSION

With the increasing number of deaths due to heart diseases, it has become mandatory to develop a system to predict heart diseases effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of heart diseases. This study compares the accuracy score of Decision Tree, Logistic Regression, Random Forest and Naive Bayes algorithms for predicting heart disease using UCI machine learning repository dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 90.16% for prediction of heart disease. In future the work can be enhanced by developing a web application based on the Random Forest algorithm as well as using a larger dataset as compared to the one used in this analysis which will help to provide better results and help health professionals in predicting the heart disease effectively and efficiently.

V. REFERENCE

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

[5]. 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.

[6]. M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp. 85–94, 2013.

[7]. C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, ―Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field,‖ in Machine Learning Paradigms, 2019, pp. 71–99.

[8]. Fajr Ibrahem Alarsan., and Mamoon Younes 'Analysis and classification of heart diseases using heartbeat features and machine learning algorithms',Journal Of Big Data,2019;6:81.

[9]. Kondababu A, Siddhartha V, Kumar BB, Penumutchi B. A comparative study on machine learning based heart disease prediction. Materials Today: Proceedings. 2021 Feb 19.

[10]. Riyaz L, Butt MA, Zaman M, Ayob O. Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review. InInternational Conference on Innovative Computing and Communications 2022 (pp. 81-94). Springer, Singapore.

[11]. Gao XY, Amin Ali A, Shaban Hassan H, Anwar EM. Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. Complexity. 2021 Feb 10;2021.

[12]. Abdullah AS, Rajalaxmi R. A data mining model for predicting the coronary heart disease using random forest classifier. InInternational Conference in Recent Trends in Computational Methods, Communication and Controls 2012 Apr (pp. 22-25).

[13]. Hasan R. Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction. InITM Web of Conferences 2021 (Vol. 40, p. 03007). EDP Sciences.

[14]. Singh B, Prabhakar Tiwari SN, Singh RP, Vishwakarma M, Patel DK, Kumar A, Pratap A, Singh SP, Mishra S, Raj R, Lohia P. SN Paper ID. InInternational Conference on Electrical and Electronics Engineering (ICE3) 2020 Feb (Vol. 14, p. 15).

[15]. Mythili T, Mukherji D, Padalia N, Naidu A. A heart disease prediction model using SVM-decision trees-logistic regression (SDL). International Journal of Computer Applications. 2013 Jan 1;68(16).

[16]. Baban.U. Rindhe, Nikita Ahire, Rupali Patil, Shweta Gagare, Manisha Darade "Heart Disease Prediction Using Machine Learning", Volume 5, Issue 1, May 20.

[17]. Hasan R. Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction. InITM Web of Conferences 2021 (Vol. 40, p. 03007). EDP Sciences.

[18]. Shah D. Heart Disease Prediction using Machine Learning Techniques Springer Nature Singapore Pte Ltd, 2020.

[19]. Kondababu A, Siddhartha V, Kumar BB, Penumutchi B. A comparative study on machine learning based heart disease prediction. Materials Today: Proceedings. 2021 Feb 19.

[20]. Gao XY, Amin Ali A, Shaban Hassan H, Anwar EM. Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. Complexity. 2021 Feb 10;2021.

[21]. Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience. 2021 Jul 1;2021.