iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VII **Month of publication:** July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73347

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

A Review on Heart Disease Prediction Using Exploratory Data Analysis

Rupali S. Awhad[1], Dinesh M. Barode[2], Amol P. Vakte[3], Seema S. Kawathekar[4]
[1,2,3] Research Student, [4]Assistant Professor, Department of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar, Maharashtra, India

Abstract: Healthcare companies produce vast amounts of raw information, commonly referred to as huge data, which can uncover invisible layouts and insightful perspectives to support informed decision-making. Data-driven decisions tend to be more reliable than those based on intuition, as they leverage large-scale datasets. Exploratory Data Analysis (EDA) plays a key role in this process by helping identify errors, recognize data characteristics, validate assumptions, and examine relationships between variables. In this context, EDA involves examining data without relying on statistical modeling or drawing formal conclusions. Analysts across various fields use EDA to uncover patterns and make informed forecasts. Recently, data analytics has become more accessible and increasingly important in healthcare, particularly for addressing disease outbreaks and emergencies. EDA serves as a foundational step in data analysis and supports the healthcare sector by enhancing treatments and promoting preventive care.

Keywords: Exploratory Data Analysis, Heart Disease, Decision Tree, Logistic Regression, SVM, Random Forest, KNN

I. INTRODUCTION

Heart disease is still one of the top causes of death in the world, which shows how important it is to find it early and make predictions to lower the death rate and improve healthcare outcomes. Predictive efforts for heart disease depend on a variety of datasets encompassing clinical information, lifestyle habits, and demographic factors. Exploratory Data Analysis (EDA) can be crucial as the initial phase in the data science workflow, allowing for the visualization and understanding of underlying patterns in the data before applying machine learning techniques.

A. Heart Disease

Cardiovascular disease has been identified and among the most dangerous medical disorders globally. It primarily affects the heart's capacity to circulate sufficient blood to the body's organs, thereby interfering with essential bodily functions [1]. Typical symptoms include tiredness, difficulty breathing, and swelling in the legs and feet. Accurate diagnosis of severe heart conditions requires advanced medical technologies, as these diseases present serious health risks [2]. However, diagnosing and treating heart-related issues remain difficult due to a lack of healthcare professionals and insufficient access to diagnostic tools, which negatively affects the quality of patient care [3]. Current as well as precise identification of heart disease is required for minimizing complications & avoiding serious health consequences [4]. Traditional diagnostic methods typically involve invasive procedures, such as reviewing the medical record of a patient, interpreting disease signs by medical professionals, and analyzing laboratory results. However, these approaches can be prone to delays and errors due to human judgment, and they often require significant time, resources, and financial investment throughout the diagnostic process [5].

Heart disease can be recognized by considering multiple factors, including age, gender, heart rate, and other related symptoms. In the healthcare industry, data analysis plays a vital role in disease prediction by improving diagnostic precision, assessing symptoms effectively, recommending appropriate treatments, enhancing patient care, lowering medical costs, increasing life expectancy, and reducing death rates among heart patients. Electrocardiograms (ECGs), which use sensors attached to the chest, are essential for detecting abnormal heart rhythms and potential stroke indicators by tracking the heart's electrical signals. Predictive analysis of heart disease typically utilizes detailed clinical data, supporting healthcare professionals in making accurate, data-driven decisions. Maintaining healthy blood circulation through the heart's vessels is critical for sustaining life, as inadequate blood flow can result in heart failure, kidney issues, neurological problems, or even sudden death. Numerous risk factors contribute to Cardiovascular illness, within obesity, tobacco use, a history of diabetes, hypertension, and high cholesterol levels, lack of physical activity, and unhealthy eating habits. One severe cardiovascular condition, Acute Myocardial Infarction (AMI) grows when the blood supply to the cardiac muscle is blocked, causing tissue damage or death [6].

Cardiovascular disease is mainly due to a decrease or blocking in blood circulation to the cardiac muscle. When this flow is limited, the oxygen supply delivered by red blood cells is also diminished an essential factor for maintaining consciousness and life. If the heart muscle is starved of oxygen for more than 6 to 8 minutes, it may lead to cardiac arrest and potentially death. A key factor in cardiovascular disease is the buildup of plaque a hardened substance mainly made of cholesterol (fat) within the coronary arteries, which severely narrows or obstructs blood flow. This condition, known as atherosclerosis, occurs when plaque accumulates in the arteries and is often associated with long-term inflammation. A high concentration of white blood cells contributes to this inflammatory process, increasing the risk of further complications like strokes or repeated heart attacks [7].

Healing generally involves two primary phases regulated by monocytes and macrophages: the inflammatory phase and the reparative phase. Both are crucial for proper recovery, but extended or excessive inflammation can be damaging and may ultimately lead to heart failure. A less common type of heart disease involves sudden spasms or contractions in the coronary arteries, which can happen unexpectedly and often without any signs of atherosclerosis [8]. These spasms restrict blood flow, reducing the oxygen supply to the heart and causing oxygen deficiency. Statistically, men are more likely to suffer from heart attacks than women. Additionally, chest pain in women often persists for more than an hour, whereas in men, it usually lasts for less than an hour. Cardiovascular disease impacts the entire body, not just the heart, causing widespread physiological changes that can affect other organs, including the bone marrow and spleen [9].

B. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is a method that is applied to investigate and analyze the essential attributes of a data set, primarily through visual methods. It serves as a preliminary step before engaging in more advanced modeling processes, offering early insights into the data. Exploratory Data Analysis can be classified into two primary categories: according to the approach employed graphical or non-graphical and based on the scope of analysis univariate (examining a single variable) or multivariate (analyzing multiple variables). The primary objectives of EDA include identifying patterns, spotting outliers, and detecting irregularities within the data. Furthermore, it plays a key role in generating hypotheses by enabling a clearer and more intuitive understanding of the dataset through visualization techniques [10][11].

After data collection and preprocessing, a crucial step known as Exploratory Data Analysis (EDA) is carried out. In this stage, data is visualized, plotted, and continuously refined without strict limitations to evaluate data quality and guide model development. Non-graphical EDA methods focus on calculating numerical values, while graphical methods display data through visual representations. Multivariate analysis explores the relationships among two or more variables at once, whereas univariate analysis focuses on examining a single variable independently. Although bivariate analysis—comparing two variables—is frequently used in multivariate EDA, there are cases where three or more variables are analyzed together. It is generally advised to first conduct univariate EDA on each variable involved in a multivariate analysis before proceeding with the full multivariate examination.

The main goals of performing Exploratory Data Analysis (EDA) include detecting errors in the dataset, confirming underlying assumptions, choosing initial model structures, discovering relationships among explanatory variables, and examining both the strength and direction of associations between predictor variables and the outcome variable [12].

II. LITERATURE SURVEY

A lot of studies have been carried out with disease prediction systems applying various machine learning algorithms in medical centres.

Karthick et al. [14] create an enhanced method for evaluating the probability of individuals under 50 years of age acquiring Cardiovascular Disease (CVD). This approach may facilitate the early detection of cardiovascular problems in adults over 50, thereby decreasing or averting mortality.

Thomas et al. [15] utilized the methods of data mining to predict an individual's risk of heart disease based on factors such as age, gender, blood pressure, cholesterol levels, and pulse rate. Techniques including Naive Bayes, K-Nearest Neighbours, Decision Tree Algorithm, & Neural Networks were utilised in their investigation. The decision tree methodology was employed for disease categorisation, whereas the Gaussian technique was utilised to estimate probability.

The study in paper [16] explores the prediction of heart disease using various machine learning algorithms such as KStar, J48, SMO, Bayes Net, and Multilayer Perceptron, implemented through the WEKA software. Among these techniques, SMO showed the best performance, followed by KStar, Multilayer Perceptron, and J48, based on results from k-fold cross-validation. However, the overall accuracy achieved by these models was still considered insufficient. Consequently, enhancing the precision of these models is crucial for delivering more dependable and effective selections in disease diagnosis.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

[17] A study applying the Cleveland dataset for heart disease, comprising 303 occurrences and 13 characteristics, employed 10-fold cross-validation with three distinct methods. The findings indicated that Gradient Boosting and Random Forest attained the maximum accuracy of 74.0 percent [18]. Experiments utilising the Framingham dataset from Massachusetts employed four models, with the K Neighbours Classifier and another classifier having maximum accuracies of 87 percent and 84 percent, respectively.

Chih-Wei Huang et al. [2015] performed a study on Chronic Kidney Disease utilising Exploratory Data Analysis (EDA) and visualisation methodologies [19]. They applied a divide-and-conquer technique to classify patients into homogeneous groups. Their suggested technique enables automatic correlation analysis and visual exploration of renal illness, while simultaneously monitoring individuals with different health issues throughout time. A Sankey diagram was used to visualise the extracted knowledge. The researchers discovered critical parameters affecting patient movements, thereafter preprocessing and segmenting the patient datasets. Ultimately, the knowledge was effectively visualised and filtered using the cohort trajectory network.

R. Indrakumari et al. [2020] conducted exploratory data analysis on heart disease using K-means clustering, with visualization carried out through Tableau software [20]. They worked with a publicly available dataset containing 209 records and eight features, such as age, blood pressure, blood glucose level, resting ECG, heart rate, and four types of chest pain. K-means clustering was applied to this dataset, and the results were visualized using Tableau.

El-Hossain A. Rady et al. [21] developed a heart disease prediction model that integrated data mining techniques with a variety of heterogeneous algorithms.

H.D. Marina et al. [22] employed a cardiovascular registry dataset to predict the occurrence of heart disease. Their approach incorporated various algorithms, including K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Logistic Regression, and Multilayer Perceptron (MLP). KNN achieved a sensitivity of 90.30%, while SVM reached an accuracy of 92.51%. Logistic Regression followed closely with an accuracy of 92.20%, and MLP outperformed the others with the highest accuracy of 95.59%.

W.L. Costa, L.S. Figueredo, et al. [18] identified a method applying Amazon EC2, Multilayer Perceptron (MLP), and cloud servers to facilitate cardiovascular rehabilitation. Cloud servers facilitate quick computations in scenarios when other platforms may collapse. This model got an efficiency of 97.79%. A logistic regression classifier was employed to analyse and categorise forms of chest discomfort [23].

Table 1: Comparative Analysis of Previous Studies

Author(s)	Year	Algorithms Used	Dataset	Result
Alkhamis, Moh A., et al. [24]	2024	Random Forest, Gradient Boost, XGBoost, SVM, Logistic Regression	1,976 people diagnosed with acute coronary syndromes in Kuwait	Random Forest achieved 80.92% accuracy
Peng, Mengxiao, et al. [25]	2023	XGBoost, Logistic Regression, Linear SVC, Random Forest , XGBH	Shanxi Baiqiuen Hospital dataset& Kaggle Competition Dataset	AUC of 0.8059 without BMI and 0.8069 with BMI
Srinivasan, Saravanan, et al. [26]	2023	Random Forest, Decision Tree, SVM, XGBoost, Radial basis functions, Knearest neighbour, Naïve Bayes , learning vector quantization	UCI repository	The proposed learning vector quantisation got an accuracy of 98.78%, precision of 98.07%, specificity of 97.1%, recall of 95.31%, F-measure of 97.89%, and sensitivity of 97.91%.
Cho, Sang-Yeong, et al. [27]	2021	AdaBoost, TreeBag, Neural Network (Using 8 and 16 variables), Logistic Regression	National Health Insurance Service-Health Screening (NHIS-HEALS) cohort from Korea	Pooled Cohort Equation (PCE) reported a C-statistic of 0.738
Schiborn, Catarina, et al. [28]	2021	Pooled Cohort Equation, Framingham CVD Risk Scores (FRS), PROCAM scores, Systematic Coronary Risk Evaluation (SCORE)	Potsdam and EPICHeidelberg (Not Available on Publicly)	C-indices indicated strong discrimination: 0.786 for EPIC-Potsdam and 0.762 for EPIC-Heidelberg
Ward, Andrew, et al. [29]	2020	Random forest, Gradient Boost, XGBoost, SVC, Decision Tree, Logistic Regression	Electronic Health Records from NorthernCalifornia	Gradient Boosting showed the highest accuracy
Grammer, Tanja B., et al. [30]	2019	ARRIBA, PROCAM I & II, FRS hard-CVE, ESC-HS, FRS-CHD1 & FRS-CHD2	Primary data from 4,044 participants in the DETECT study	Sensitivity to predict future cardiovascular disease was approximately 80%.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

III. METHODOLOGY

Each stage is crucial in the process of transforming raw data into valuable insights. The proposed research methodology outlines a systematic approach to extracting knowledge and developing predictive models from a given dataset.

A. Data Collection

The first phase entails gathering accurate information about patients afflicted with heart disease. We use the Heart Disease Prediction dataset from Kaggle, which has 303 patient records and 14 unique factors. The variables comprise: 41 distinct values for age, 2 for sex, 4 categories of chest pain, 49 for resting blood pressure, 2 for fasting blood sugar (fbs), 3 for resting electrocardiographic results (restecg), 91 for maximal heart rate reached (thalachh), and 2 for exercise-induced angina (exng). A detailed explanation of each feature is shown in Table 1, followed by the total of unique values for each.

Table 2: Dataset Features and Descriptions

Sr.No	Attribute	Description
1	Age	Patient's age in Year
2	Sex	Gender (1 = male, 0 = female)
3	Cp	Type of chest pain (values: 0–3, representing different pain categories)
4	Trest bps	Resting blood pressure in mm Hg (range: 50–150)
5	Chol	Serum cholesterol in mg/dl (range: 100–600)
6	Fbs	Fasting blood sugar > 120 mg/dl (1 = yes, 0 = no)
7	Restecg	Resting electrocardiographic (y=1, n=0)
8	Thalach	Maximum heart rate achieved (range: 71–200 bpm)
9	Exang	Exercise-induced angina (1 = yes, 0 = no)
10	Old Peak	St depression (y=1, n=0)
11	Slope	Slop peak exercise(y=1, n=0)
12	Ca	Number of major vessels (0–3) visualized by fluoroscopy
13	Thal	Thalassemia status (3 = normal, 6 = fixed defect, 7 = reversible defect)
14	Target	Heart disease (no=1, yes=2)

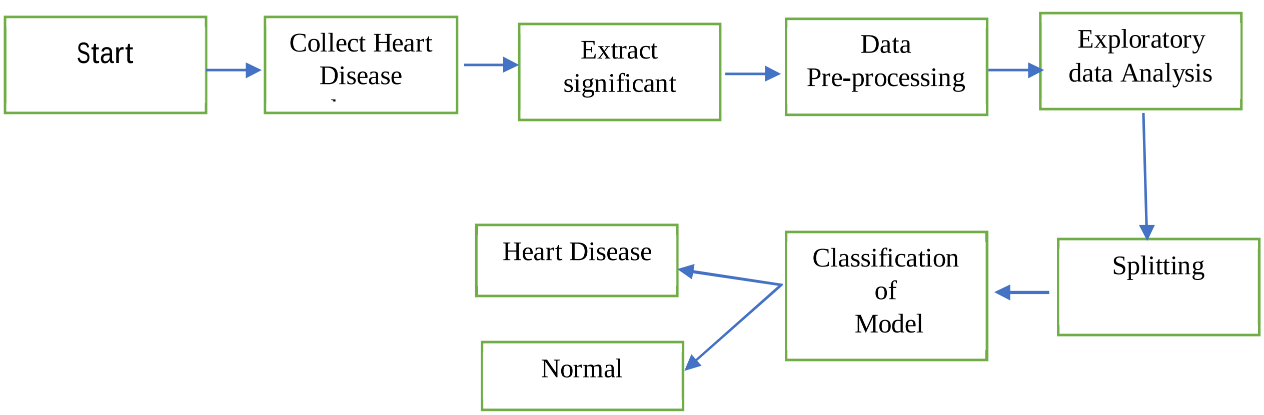


Fig.1: Workflow of Heart Disease Prediction System

B. Pre-processing of data

The next stage in the workflow is data pre-processing, which is essential for preparing raw data for machine learning applications. This stage entails cleaning the dataset, dealing with missing or extraneous values, and converting the data into formats appropriate for algorithmic analysis. It is universally acknowledged as one of the most challenging yet crucial elements of the data analytics pipeline, frequently reaching over fifty percent of the total time allocated to the entire process.

Efficient pre-processing is essential for making sure the precision and dependability of machine learning models, as raw input often keeps errors, noise, duplication, or incomplete entries. Essential tasks at this step consist of importing datasets, dividing them into training and testing sets, and implementing attribute scaling algorithms, all of which substantially enhance model performance.

C. Feature Selection

The precision of a predictive model is significantly influenced by the selection of the most pertinent characteristics from the pre-processed dataset. This procedure entails identifying key characteristics, including age, sex, blood pressure, cholesterol levels, smoking status, and family history, that deeply influence the risk of heart disease. Once the key variables are identified, feature extraction is performed. Not all features enhance model performance; some may be redundant or even detrimental to accuracy. Feature selection helps optimize the dataset by removing irrelevant variables and constructing new, informative features derived from existing ones. These newly created features are designed to retain the essential information of the original attributes. A correlation matrix is often employed to assess relationships between features and assist in selecting the most impactful ones for the final predictive model.

D. Model Selection

This phase deals with to determining which algorithm is most suited for the classification task. The selection process is predicated on assessing the way various machine learning algorithms perform on the dataset. The techniques logistic regression, decision trees, support vector machines (SVM), k-nearest neighbours (KNN), and random forests are frequently employed in the prediction of cardiac disease. The algorithm that is selected has a significant impact on the model's overall predicted accuracy. A comparison examination of each model's accuracy ratings is conducted in order to identify the best option, and the algorithm that predicts heart disease with the highest accuracy is ultimately chosen.

E. Training and Testing

After the selection of the model, it is next subjected to training and testing with the dataset that has been pre-processed. Both the training phase and the testing phase are used to evaluate the predicted accuracy of the model. The training phase gives the algorithm the opportunity to learn the underlying patterns and correlations that are present within the data. The effectiveness of the model is evaluated based on how effectively it generalises to test data that has not yet been discovered.

1) Algorithms Used

a) Decision Tree

A Decision Tree (DT) is a supervised learning technique primarily used for classification, while it can also be applied to regression tasks. The architecture of a decision tree resembles that of a tree, including root, branch, and leaf nodes. Each node represents a characteristic or attribute, branches identify decisions or rules, and leaves signify the final outcome or decision. Decision Trees operate by partitioning the data according to feature values. They work similarly to a flowchart and are regarded as highly specialised models. Decision Trees are preferred for their rapid and dependable outcomes relative to alternative machine learning methods, little data preprocessing requirements, and ease of comprehension. The Classification and Regression Trees (CART) technique allows Decision Trees to carry out both classification and regression tasks.

b) Random Forest

The Random Forest (RF) algorithm enhances performance by combining several Decision Trees (DTs). It employs a method called bootstrap aggregation, where multiple random samples are drawn from the original dataset. Each Decision Tree is built using different subsets of these samples through row sampling. These trees are trained independently, and the overall prediction is determined by taking a majority vote from all the individual trees.

c) K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a well-known and easy-to-use statistical learning algorithm. The parameter K specifies the number of nearest neighbors to consider. KNN can be applied to both classification and regression tasks by determining the majority class among the closest neighbors. As a non-parametric method, KNN does not assume any specific data distribution. It works well in various scenarios, especially when decision boundaries are irregular. Being an instance-based or “lazy” learning algorithm, KNN delays computation until a prediction is needed, approximating the function locally. It classifies new data points by measuring similarity, typically using distance metrics. Due to its simplicity and ability to handle non-linear data, KNN is commonly used in fields such as recommendation systems, anomaly detection, and pattern recognition.

d) Support Vector Machine

Support Vector Machine (SVM) is a powerful algorithm suitable for both classification and regression tasks. It employs statistical techniques to determine the optimal hyperplane that separates different data classes. With the help of the kernel trick, SVM can map data into higher-dimensional spaces, allowing it to identify the most suitable boundary for class separation. The goal of SVM is to maximize the margin between different classes, which leads to better classification performance. Thanks to the kernel method, SVM can manage nonlinear relationships between features, making it especially useful in high-dimensional spaces. It is known for its strong accuracy, good generalization across datasets, and resistance to overfitting. SVM is commonly employed in areas such as image recognition, text classification, and bioinformatics, where accuracy is vital.

e) Logistic Regression

Logistic Regression (LR) is a classification method used to predict categorical outcomes, most commonly binary results like 0 or 1, yes or no. It calculates the probability that a given instance belongs to a specific class, making it highly effective for tasks such as spam detection and medical diagnosis. The logistic function produces an output between 0 and 1, which suits classification problems well. Known for its simplicity, ease of interpretation, and efficiency, logistic regression is widely used across many domains. It performs best when there is a linear relationship between the features and the probability of the event. LR focuses on binary classification by estimating the chance of class membership, using the logistic function to model this probability, and then applying a threshold to classify the data into two categories.

Table-3: Comparison table

Algorithm	Accuracy
Decision tree	91%
Random forest	96%
KNN	73%
SVM	69%
Logistic regression	70%

Table 3 presents a comparison of various machine learning algorithms based on their accuracy performance. Among the listed algorithms, the Random Forest achieves the highest accuracy of 96%, indicating its superior ability to classify the data correctly. The Decision Tree achieves a strong accuracy of 91%, demonstrating its high effectiveness. On the other hand, K-Nearest Neighbors (KNN) attains an accuracy of 73%, while Logistic Regression and Support Vector Machine (SVM) have slightly lower accuracies of 70% and 69%, respectively. This comparison highlights that ensemble methods like Random Forest generally perform better than single-model algorithms for this dataset.

IV. CONCLUSION

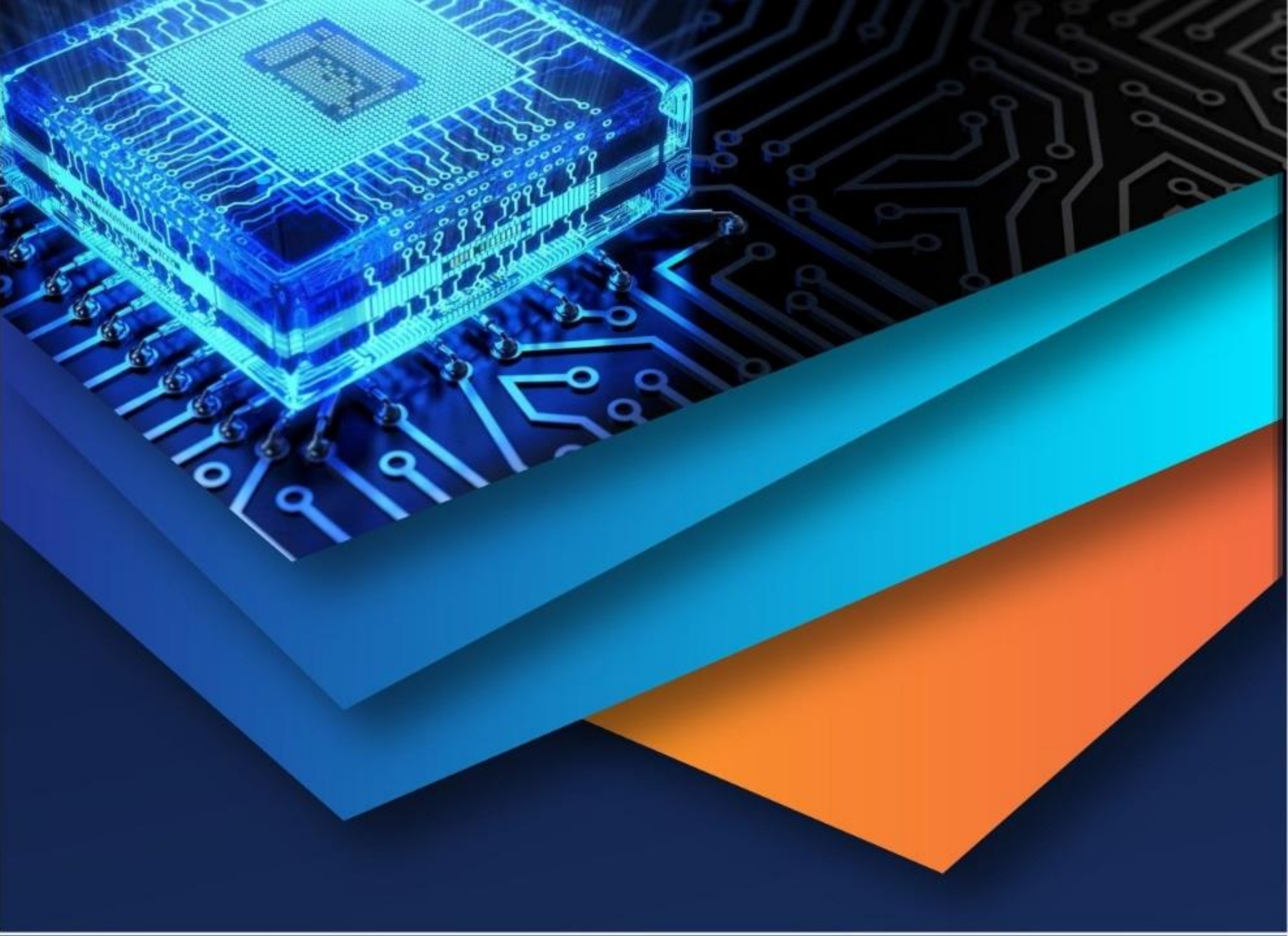
Based on the accuracy results of the different algorithms, Random Forest provides the highest accuracy at 96%, making it the most effective model for the given task. Decision Tree follows closely with an accuracy of 91%, showing strong performance as well. KNN, SVM, and Logistic Regression yield significantly lower accuracies, ranging from 69% to 73%. This indicates that these algorithms may not be as well-suited to the data for predicting heart disease in this case, compared to Random Forest and Decision Tree.

The high performance of Random Forest can be attributed to its ability to aggregate the results of multiple decision trees, which helps reduce overfitting and improve generalization. Therefore, Random Forest would be the recommended algorithm for heart disease prediction, as it provides the most reliable and accurate results.

REFERENCES

[1] L. Bui, T. B. Horwich, and G. C. Fonarow, “Epidemiology and risk profile of heart failure,” Nature Reviews Cardiology, vol. 8, no. 1, pp. 30–41, 2011.
[2] J.Mourão-Miranda, A.L.W.Bokde,C.Born, H.Hampel,and M. Stetter, “Classifying brain states and determining the discriminating activation patterns : support vector machine on functionalMRIdata,”NeuroImage,vol.28,no.4,pp.980–995, 2005.
[3] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, “Innovative artificial neural networks-based decision support system for heartdiseasesdiagnosis,”JournalofIntelligentLearningSystems and Applications, vol. 5, no. 3, pp. 176–183, 2013.
[4] Q. K. Al-Shayea, “Artificial neural networks in medical diagnosis,” International Journal of Computer Science Issues, vol. 8, no. 2, pp. 150– 154, 2011.

[5] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," International Journal of Computer Applications, vol. 19, no. 6, pp. 6–12, 2011.
[6] Al Mamoon I, Sani AS, Islam AM, Yee OC, Kobayashi F, Komaki S, "A proposal of body implementable early heart attack detection system", 1-4, 2013.
[7] Patterson K , Matthias Nahrendorf. Circ Res 119: 790-793, 2016.
[8] Soni, J., Ansari, U., Sharma, D., & Soni, S, "Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications", 17(8), 43-48, 2011.
[9] Masethe, H. D., & Masethe, M. A , "Prediction of heart disease using classification algorithms", In Proceedings of the world congress on engineering and computer science (Vol. 2, pp. 22-24), 2014-Oct.
[10] Komorowski M, Marshall D. C , J, Salciccioli J D and Crutain Y, Chapter 15- Exploratory Data Analysis - Secondary Analysis of Electronic Health Records. DOI: 10.1007/978-3-319-43742-2_15, 2016.
[11] Valdiviezo-Diaz, P., Reátegui, R., Barba-Guaman, L., Ortega, M, "Exploratory Data Analysis on Cervical Cancer Diseases. In: Botto-Tobar", M., Montes León, S., Torres-Carrión, P., Zambrano Vizuete, M., Durakovic, B. (eds) Applied Technologies. ICAT 2021. Communications in Computer and Information Science, vol 1535. Springer, Cham. https://doi.org/10.1007/978- 3-031-03884-6_32 ,2022.
[12] Huang, CW., Lu, R., Iqbal, U. et al., "A richly interactive exploratory data analysis and visualization tool using electronic medical records", BMC Med Inform Decis Mak 15, 92. https://doi.org/10.1186/s12911-015-0218-7, 2015.
[13] Rashik Rahmen, "Heart Attack Analysis Prediction Dataset", https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysisprediction-dataset, year = 2021-03-22.
[14] Alsmadi, Tibra, Nour Alqudah, and Hassan Najadat, "Prediction of Covid-19 patients states using Data mining techniques", 2021 International Conference on Information Technology (ICIT), IEEE, 2021.
[15] Khourdifi, Youness, and Mohamed Bahaj, "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization" International Journal of Intelligent Engineering and Systems 12.1: 242-252, 2019.
[16] A. H. M. S. U. Marjia Sultana , "Analysis of Data Mining Techniques for Heart Disease Prediction" , 2018.
[17] M. I. K. A. I.S. Musfiq Ali , "Heart Disease Prediction Using Machine Learning Algorithms".
[18] M. A. K. S. H. K. M. A. V. P. M Marimuthu, "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach".
[19] Huang, CW., Lu, R., Iqbal, U. et al., "A richly interactive exploratory data analysis and visualization tool using electronic medical records", BMC Med Inform Decis Mak 15, 92. https://doi.org/10.1186/s12911-015-0218-7. 2015.
[20] R. Indrakumaria, T Poongodi and Sowmya Rajnan Jena, (2020). Heart Disease Prediction using Exploratory Data Analysis, International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020, Procedia Computer Science 173 (2020) 130–139
[21] S. Nalluri, R. Vijaya Saraswathi, S. Ramasubbareddy, K. Govinda, and E. Swetha, "Chronic heart disease prediction using data mining techniques," in Data engineering and communication technology, Springer, 2020, pp. 903–912.
[22] Samuel Harford, Houshang Darabi, Marina [2019] Del Rios, Somshubra Majumdar, Fazle Karim, Terry Vanden Hoek, Kim Erwin, Dennis P. Watson, "A Machine Learning Based Model for Classification and Sensitivity Analysis of Out of Hospital Cardiac Arrest Outcomes" Elsevier, Resuscitation 138, pp. 134–140
[23] W. L. Costa, L. S. Figueredo, and E. T. A. Alves, "Application of an Artificial Neural Network for Heart Disease Diagnosis Brazilian Congress on Biomedical Engineering, Springer, 2019, pp. 753– 758.
[24] Alkhamis, Moh A., et al. "Interpretable machine learning models for predicting in-hospital and 30 days adverse events in acute coronary syndrome patients in Kuwait." Scientific Reports 14.1 (2024): 1243.
[25] Peng, Mengxiao, et al. "Prediction of cardiovascular disease risk based on major contributing features." Scientific Reports 13.1 (2023): 4778.
[26] Srinivasan, Saravanan, et al. "An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database." Scientific Reports 13.1 (2023): 13588.
[27] Cho, Sang-Yeong, et al. "Pre-existing and machine learning-based models for cardiovascular risk prediction." Scientific reports 11.1 (2021): 8886.
[28] Schiborn, Catarina, et al. "A newly developed and externally validated non-clinical score accurately predicts 10-year cardiovascular disease risk in the general adult population." Scientific Reports 11.1 (2021): 19609
[29] Ward, Andrew, et al. "Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population." NPJ digital medicine 3.1 (2020): 125.
[30] Grammer, Tanja B., et al. "Cardiovascular risk algorithms in primary care: Results from the DETECT study." Scientific reports 9.1 (2019): 1101



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429