

Research Article

Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction

Gufran Ahmad Ansari,¹ Salliah Shafi Bhat,² Mohd Dilshad Ansari ,³ Sultan Ahmad ,^{4,5} Jabeen Nazeer ,⁴ and A. E. M. Eljialy ,⁶

¹Department of Computer Science, Dr. Vishwanath Karad MIT World Peace University, Pune 411038, India

²Department of Computer Applications, B.S Abdur Rahman Institute of Science & Technology, Chennai 600048, India

³Guru Nanak University, Hyderabad 501506, India

⁴Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, P. O. Box. 151, Alkharj 11942, Saudi Arabia

⁵University Center for Research and Development (UCRD), Department of Computer Science and Engineering, Chandigarh University, Gharuan, Mohali, 140413 Punjab, India

⁶Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, P.O.Box. 151, Alkharj 11942, Saudi Arabia

Correspondence should be addressed to Sultan Ahmad; s.alisher@psau.edu.sa

Received 23 January 2023; Revised 9 March 2023; Accepted 10 March 2023; Published 29 May 2023

Academic Editor: Yaser Ahangari N.

Copyright © 2023 Gufran Ahmad Ansari et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The leading cause of death worldwide today is heart disease (HD). The heart is recognised as the second-most significant organ behind the brain. A successful outcome of treatment can be improved by an early diagnosis which can significantly reduce the chance of death in health care. In this paper, we proposed a method to predict heart disease. We used various machine learning algorithms (MLA), namely, logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), Naive Bayes (NB), random forest (RF), and decision tree (DT). With the testing data set, we evaluated the model's accuracy in heart disease prediction. When compared to the other five models, the random forest and k-nearest neighbor approaches perform better. With a 99.04% accuracy rate, the k-nearest neighbor algorithm and random forest provide the best match to the data as compared to other algorithms. Six feature selection algorithms were used for the performance evaluation matrix. MCC parameters for accuracy, precision, recall, and F measure are used to evaluate models.

1. Introduction

One of the most difficult and severe illnesses affecting individuals worldwide is heart disease. The heart which regulates blood flow throughout the body is a crucial component of the human body. The human lifespan will be shortened by heart disease. HD affects around 15 million people each year [1]. Heart disease is one of the top causes of death in the contemporary world. Heart illnesses are caused by many risk factors, such as high blood pressure and excessive cholesterol, high cholesterol, diabetes, and irregular heartbeats. Doctors, researchers, and scientists are working to identify

the causes of heart disease in its early stages to make human life better [2]. Due to the limited accessibility of diagnostic tools, the lack of specialists, and other resources that affect the accurate diagnosis and treatment of heart patients, heart disease diagnosis and therapy are particularly difficult in developing countries [3]. Since cardiac illness has a complex character, it requires cautious management. Regression, KNN, SVM, NB, and DT are used to categorise the severity of the condition. In order to help with decision-making and prediction from the vast quantity of data generated by the healthcare business, machine learning (ML) has been proven to be useful [4]. Around 17.9 million people died in 2016

which is 31% of all deaths worldwide. Among them, heart attack and stroke account for 85% of the deaths. Patients are facing more cardiac problems due to a variety of factors, including lifestyle choices like smoking, eating poorly, and having high blood pressure [5]. The RF and KNN approaches outperform the other five methods. The k-nearest neighbor method and RF when compared to other algorithms, offer the best match to the data with a 99.04% accuracy rate. Based on symptoms such as pulse rate, age, gender, asthma, smoking, and blood pressure, heart disease is predicted with accuracy [6]. Additionally, recently many researchers create machine learning-based methods for forecasting the prevalence of heart illnesses [7]. The categorization and prediction for the diagnosis of cardiac disease have been the subject of numerous studies, and a variety of machine learning models are being applied. Using a simulated classifier, the patients with high and low risks of congestive heart failure are displayed [8]. Shortness of breath, muscular weakness, swollen feet, and exhaustion are among the indications and symptoms of heart disease [9]. Heart illness can be fatal and should not be ignored. Males are more likely than females to suffer heart disease, according to Harvard Health Publishing [10]. We gathered a dataset for the research of heart disease from different sources, namely, the University of California (UCI). Using machine learning techniques, the UCI database is used to identify heart disease. Using NB, DT, LR, and the random forest algorithm, they demonstrated the accuracy of the random forest algorithm at 90.16 percent. As a result, the accuracy achieved with logistic regression is 89.06 percent, whereas the accuracy achieved without using logistic regression is 87.77 percent [11, 12]. Researchers applied the random forest and nearest neighbor algorithms for improving accuracy. A detailed analysis of heart disease prediction using machine learning was published in 2020. As a result, the annual decline in heart disease deaths has been significant. However, it is really helpful to utilize machine learning techniques to forecast results from existing data. This research employs a classification-based machine learning technique to anticipate the risk of heart disease from the risk factors. It also aims to improve the accuracy of heart disease risk predictions.

1.1. Motivation of Study. There are several diseases that affect people everywhere in the world. Today, HD is a serious problem that has a big impact on mortality in both men and women. 17.9 million deaths from heart disease are reported annually by the WHO, which accounts for 31% of all deaths from heart diseases. Although there are machine learning tools and approaches available, there are no models that are now suitable for quickly and accurately predicting the disease. There is currently no reliable automated system that can improve heart disease prognosis or reduce its consequences. Because of this, using machine learning algorithms to lessen the effects of the disease would be a significant accomplishment. It might improve the quality of life for heart patients while also significantly delaying the onset of the condition. The major goals of this research are to build a model to predict the presence of heart diseases. Additionally, the goal of this research is to determine the

classification algorithm that will predict the above sickness with the highest level of accuracy. This research will be supported by a comparative analysis such as logistic regression, KNN, support vector machine, Naive Bayes, decision tree, and random forest for prediction of heart disease, and the most accurate algorithm would be considered to be the better one. The segmentation of the paper is organized as follows: Section 1 is the Introduction. Section 2 discusses related work with existing methods. Section 3 discusses the flow chart of the proposed framework. Section 4 describes data collection and methodology. Section 5 is about results and analysis. Finally, Section 6 ends with a conclusion as well as a future enhancement.

2. Related Work

HD is a common disease that affects many people during middle age or old age. A wide variety of issues can solve related to heart diseases using a machine learning approach. Marimuthu et al. conducted a review for the prediction of heart disease using a data analytical technique. For predicting cardiac disease, machine learning techniques (MLT) has included DT, NB, KNN, and SVM [13]. A comprehensive review of heart disease prediction using machine learning was written by Battula et al. They have created a table that contrasts every MLT used to predict heart disease since 2012 [14]. Comparative analysis of cardiac disorders using MLA has been done in numerous research articles. The literature evaluation has shown the classification effectiveness of various machine learning algorithms on the dataset for heart disease [15]. A suggestion of a decision support system based on a logistic regression classifier for categorising heart disease attained a classification accuracy of 77%. Machine learning is useful for a variety of problems. One use for this method is to a dependent variable can be predicted using the values of the independent variables. Due to its extensive data resources, which are difficult to manage manually, the health sector has advanced analytics. Even in developed economies, heart disease has been found to be one of the leading causes of death. Heart disease deaths are caused in part because the risks are not identified or are detected much later than they ought to be. However, using machine learning techniques can help resolve this problem and provide early risk predictions. Support vector machines (SVM), DT, regression, and NB classifiers are a few of the methods utilised for these prediction issues. With 92.1% accuracy, SVM was found to be the strongest predictor followed by neural networks (91%) and decision trees (89.6%) diabetes, hypertension [16]. It was believed that gender and smoking were risk factors for heart disease. [17]. Machine learning techniques such as DT, NB, and associative classification are effective at predicting cardiac disease according to analytical research. Comparing associative classification to standard classifiers, especially when dealing with unstructured data, it produces higher accuracy and flexibility. Decision tree classifiers are easy to use and precise, according to a comparison of classification methods. The best algorithm was discovered to be Naive Bayes which was then followed by neural networks and decision trees [18]. Additionally used for disease prediction

are artificial neural networks. Supervised networks have been utilised for diagnosis as well as the back propagation algorithm can be used to train them. The test results have demonstrated satisfactory accuracy. It introduced the Intelligent Heart Disease Prediction System (IHDPs) and techniques like DT, NB, and neural networks (NN) [19]. The authors' experiments showed that the NB model had the highest prediction accuracy (86.1%). DT came in third with a score of 80.4%, and NN came in second with a score of 86.12% for right prediction. The majority of high-accuracy reduction research employs a mixed method that involves categorization algorithms. Our research, which is summarized here is aimed at improving the classification of algorithms by using machine learning techniques. Both the effectiveness of these classification algorithms and the accuracy of heart disease prediction is enhanced. Research on LR, KNN, SVM, NB, DT, and RF is performed out, and the outcomes are evaluated. Applying feature selection improves the outcomes much more. The results are used to evaluate how effectively these classifiers may be used in the healthcare sector.

3. Flow Chart of Proposed Framework

The proposed flow chart for the entire experiment from data collection to result development is shown in Figure 1. Data is first preprocessed after being collected from sources (as described earlier).

Preprocessing data is used to reduce bias, noise, and inaccuracy. Following the data preprocessing stage, there are training and testing sets for the database.

In addition, many machine learning technologies are utilised to train and test the data. The technique is finished with the generation of accurate results that are compared across various machine learning techniques.

4. Data Collection and Methodology

The purpose of the research paper is to explore, and the creative process is briefly covered in the following subsections.

4.1. Data Set. The researchers analyze the use of Dataset for Cleveland Heart from UCI's machine learning. The dataset has 12 attributes and 520 occurrences. The dataset's description can be found in Table 1 This proposed research used the dataset to create a machine-learning-based method for diagnosing heart problems. The features are age, gender, Trestbps, Chol, fbs, Thalch, smoker, CP, skin cancer, BMI, blood pressure, and outcome. The main class has two values, "False" and "True," which represent the absence or presence of any heart disease, respectively.

4.2. Data Preprocessing. When using machine learning algorithms cleaning, the data is crucial for maximizing precision and effectiveness. Data preparation is required for accurate data representation and machine learning classifiers which must be trained and tested properly. In order for MLT to effectively represent data and be trained and validated data must first be preprocessed. The standard scalar guarantees that each feature has a mean of 0 and a variance of 1 resulting in an equal coefficient for all features. The data is modified similarly in MinMax Scaler so that all features fall

between 0 and 1. The dataset basically contains a deletion of the missing values feature row. This research implemented each of these data preparation methods.

4.2.1. Data Cleaning. Data were used to acquire unprocessed information. As a result, a variety of methods has been used to clean the data including eliminating duplicates and irrelevant information.

4.3. Feature Selection. The most pertinent information is chosen by feature selection, a type of dimensionality reduction in order to categorise and predict the disease. In many well-known classification applications, the feature selection process is one of the fundamental elements [20]. Before classifying the data, more relevant features must be chosen in order to produce a better result in accuracy, and unnecessary features must be eliminated [21]. In order to classify the input data, the most relevant feature is selected. This feature selection approach is frequently used in all application domains because it removes duplicate data without sacrificing any information. As a result, this technique is used with a variety of algorithms. The following reasons support the implementation of the feature selection technique:

- (i) Reduced training program
- (ii) It facilitates the identification of the data by the algorithm
- (iii) The removal of unnecessary data from high-dimensional space
- (iv) By lowering the variables, the output data can be enhanced

4.3.1. Correlation Matrix. When creating a useful dataset analysis, it is frequently simpler to take the relationship between variables into consideration. A statistic known as correlation determines how closely two variables move in relationship to one another. Two variables are considered to be positively linked when they move in the same direction and negatively correlated when they move in the opposite direction. The correlation map based on the diabetes dataset is shown in Figure 2. The dataset is evaluated, and a heat map is created to show the correlation between the values. From this, it can be seen that age, gender, and Thalch characteristics that most strongly match the target variable. The correlation between age and outcome is 1:0.11 in Figure 2 which is greater than other attributes.

4.4. K-Fold and Data Splitting. Researchers and practitioners frequently utilize the K-fold cross-validation method to build models and get rid of information bias. With a k value of 10, the K-fold cross-validation method has been applied. Ten equal-sized partitions of the full dataset were created at random. Ten partitions were created; however, only one was utilised to validate (test) the model. The remaining ten partitions are used as training data. Each of the 10 partitions was used as the validation data exactly once during the course of the entire process' ten iterations. The accumulation function was used to combine the results

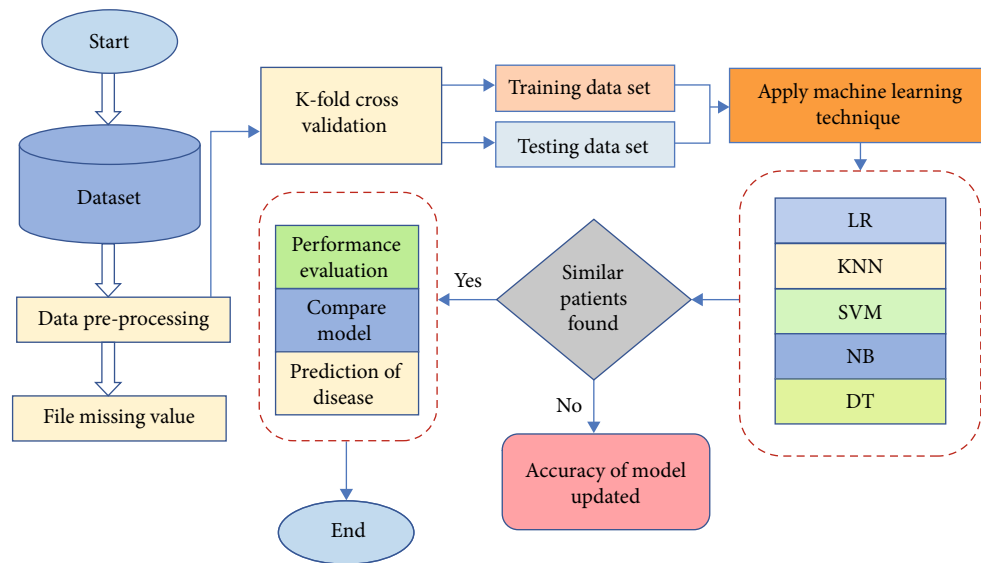


FIGURE 1: Proposed flow chart for predicting heart disease.

TABLE 1: Features of the dataset.

S.no	Attributes	Description
1	Age	Age in years
2	Gender	If a person is male or female
3	Trestbs	Hospital admission mm/hg for the resting blood pressure
4	Chol	Mg/dl of serum cholesterol
5	Fbs	Blood sugar levels at breakfast (more than 120 mg/dl). 1 indicates true. 0 indicates false.
6	Thalch	It achieved maximum heart rate
7	Smoker	Whether the person is a smoker or not
8	CP	Whether the person is having chest pain or not
9	Skin cancer	Whether the person is suffering from skin cancer or not.
10	BMI	Body mass index of a person
11	Blood pressure	Whether the person is having blood pressure or not.
12	Outcome	0 denotes the absence of heart disease. 1 denotes cardiovascular disease

of all iterations. To match the performance of both training and testing datasets, the issue of overfitting and underfitting has been reduced in the dataset. The advantage of this strategy was that it eliminated bias from the data when creating ML models to produce accurate results. In order to validate the results, each bin of testing data has been used exactly once. All data samples are used for both training and testing. The dataset is split into 70% for testing and 30% for training, and the analysis is carried out using the method identified below.

4.5. Apply Machine Learning Technique. Using machine learning classification, groups of patients with heart disease and healthy people are segregated. Using open-source Anaconda 2020, the entire experimental work was performed uses of data science and machine learning in scientific computing. The preprocessing of large amounts of data, predictive analysis, and other applications using the free and unrestricted open-source Python distribution known as Anaconda. It was developed to simplify package manage-

ment and distribution. Together with Python, Spyder is used as an integrated development environment for programming tasks and calculations (3.7.6). A machine is trained using machine learning to take information from the data and predict the results of new sets of information. As a result, we now have training and test sets of data. After the machine has been trained using the training data set, the results are verified using the test data set. A software will be created as part of the machine learning model that we will create. Supervised learning and unsupervised learning are the two subcategories of machine learning. the supervised education, in supervised learning, the computer receives instruction (mentoring), but in unsupervised learning, the machine picks up skills on its own (self-study). The examples which follow will help us understand how the two vary. Supervised learning (SL) algorithms:

- (i) The machine must determine if an incoming mail is spam or not given the data of emails designated by users as trash or not

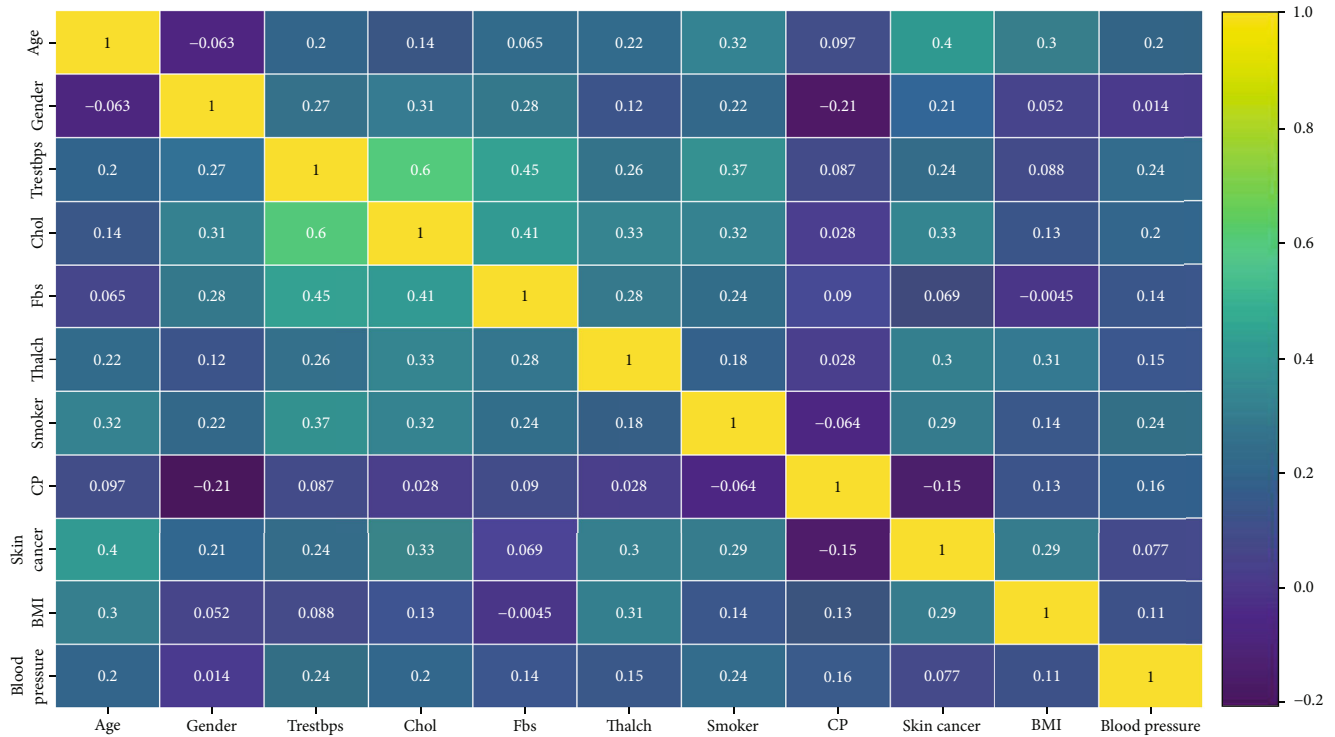


FIGURE 2: Heat Map of a data set.

- (ii) The machine should be able to determine whether a new patient has cancer based on the data of individuals who have been diagnosed with the disease
- (iii) The machine must predict the cost of the property with the specific size given data on the costs of homes in a certain area of varying sizes

And the following in unsupervised learning algorithms

- (i) Finding patterns in the data using the scientific data
- (ii) Noise reduction in the audio input
- (iii) Obtaining song background music for the chorus

In short, SL uses labeled data, whereas unsupervised learning uses unlabelled data. A list of various machine learning algorithms is provided below. decision tree, Naive Bayes, support vector machine, logistic regression, k-nearest neighbor, random forest.

4.5.1. Logistic Regression. Supervised learning which includes classification and regression problems can be resolved using the technique of logistic regression. The range of logistic regression's result is between 0 and 1. The maximum likelihood estimate is the foundation of this technique. In logistic regression, the Sigmoid function whose probability is presented as a binary one is used as an activation function [22]. In equation (1), it is shown as

$$q = (1 + e)^{(c+dx)-1}, \quad (1)$$

where q is the probability c , d will be the parameter of the model, and x is a factor.

4.5.2. K-Nearest Neighbor. On the basis of the samples Euclidean distance, it extracts knowledge $e(y_1, y_j)$ and the vast majority of k nearest neighbor.

$$E(y_1, y_j) = \sqrt{(y_{i1} - y_{j1})^2 + (y_{im} - y_{jm})^2}. \quad (2)$$

4.5.3. Support Vector Machine. Models are described as finite-dimensional vector spaces, where each dimension denotes a "feature" of a particular object. It has been demonstrated to be a successful strategy in high-dimensional space issues. Due to its computational effectiveness on huge datasets, this technique is typically utilised in sentiment analysis and the classification of data.

4.5.4. Naïve Bayes. The Naïve Bayes algorithm classifies the dataset using the Bayes rule. Based on the probability observed in the training data, the classification is made using all the features. It is a supervised learning algorithm. The classification is made based on the probability

$$P(A) = \frac{P(A|B) \cdot P(B)}{P(B)}, \quad (3)$$

where $P(A|B)$ is the conditional probability of A given B , $P(B|A)$ is the conditional probability of B given A , $P(A)$ is the probability of event A , and $P(B)$ is the probability of event B .

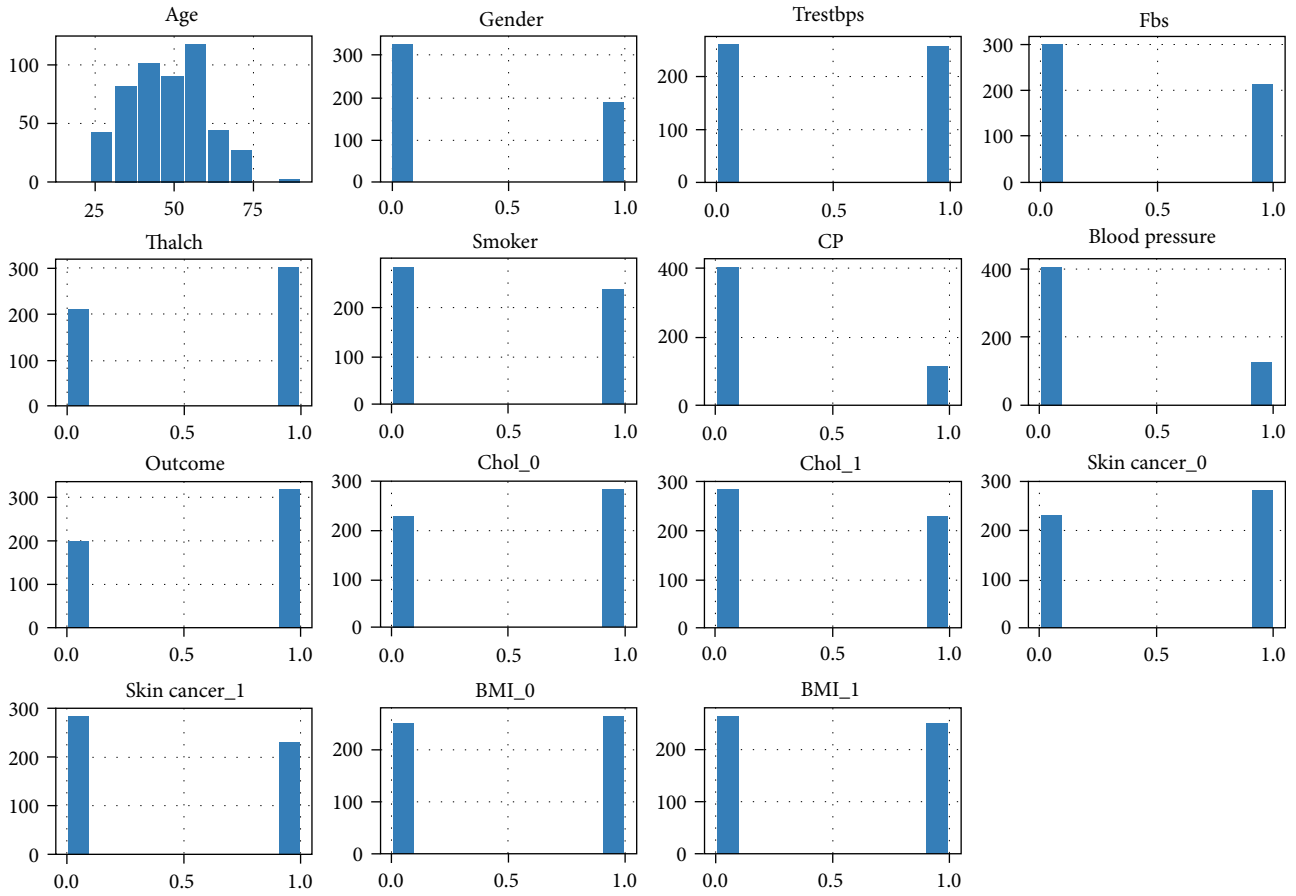


FIGURE 3: Histogram of attributes.

4.5.5. Decision Tree. Each leaf node has a class label, and each branch shows the outcome of a test on a specific variable in these supervised machine algorithms. At the top of the tree is the parent node, also referred to as the root node. To identify a different separate category based on the most data collected, decision-makers can choose the best option and make their way up a decision tree from root to leaf [23]. DT can handle constant and continuous parameters. The major benefit of the decision tree is that it can overfit.

4.5.6. Random Forest. One algorithm for classification is the random forest approach. Based on the bagging process, instruction is given. In the algorithm for supervised learning, the classification of the algorithm is given in

$$\text{MSE} = \frac{1}{N} \sum (N_i = 1f_i - y_i)^2, \quad (4)$$

where 'N' is the occurrence count, f_i is the model's output, and y_i represents the instances' true values.

4.6. Performance Evaluations. A comparison of several categorization techniques has been done using the Cleveland dataset. The performance matrices Accuracy, Precision, Recall, F-Measure, and MCC are all explained by Equations

(5–9). These evaluation measures are utilised to contrast the efficiency of our suggested strategy and possible alternatives.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100, \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} * 100, \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100, \quad (7)$$

$$\text{F Measure} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} * 100, \quad (8)$$

$$\text{MCC} = \frac{\text{TP} * \text{TN} - \text{FP} * \text{FN}}{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP}) + (\text{TN} + \text{FN})} * 100. \quad (9)$$

5. Result and Analysis

Numerous classification models and their statistical analyses are provided in this section of the research. On the Cleveland heart disease data, we assess the effectiveness of LR, KNN, SVM, NB, RF, and DT in the first stage. In this research, we investigated different machine learning algorithms for the prediction of cardiac disease using an

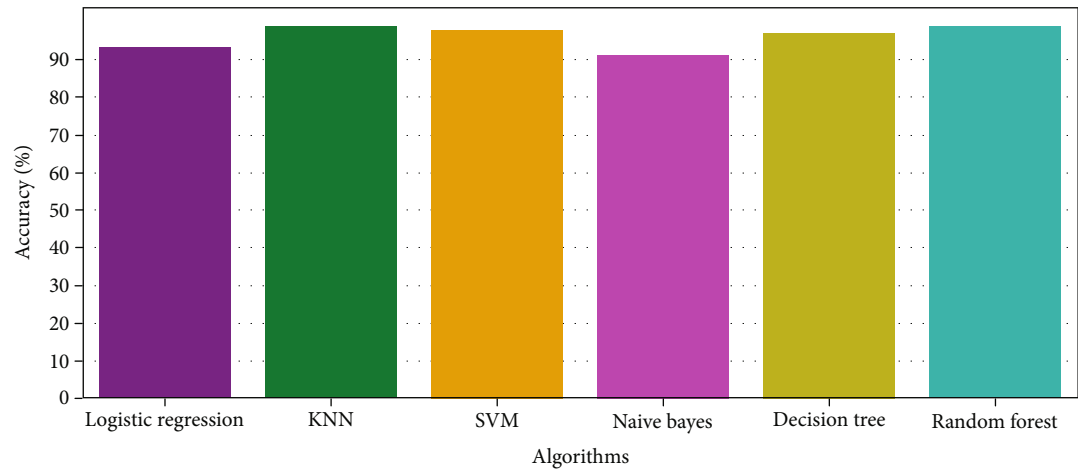


FIGURE 4: Accuracy of models for different algorithms.

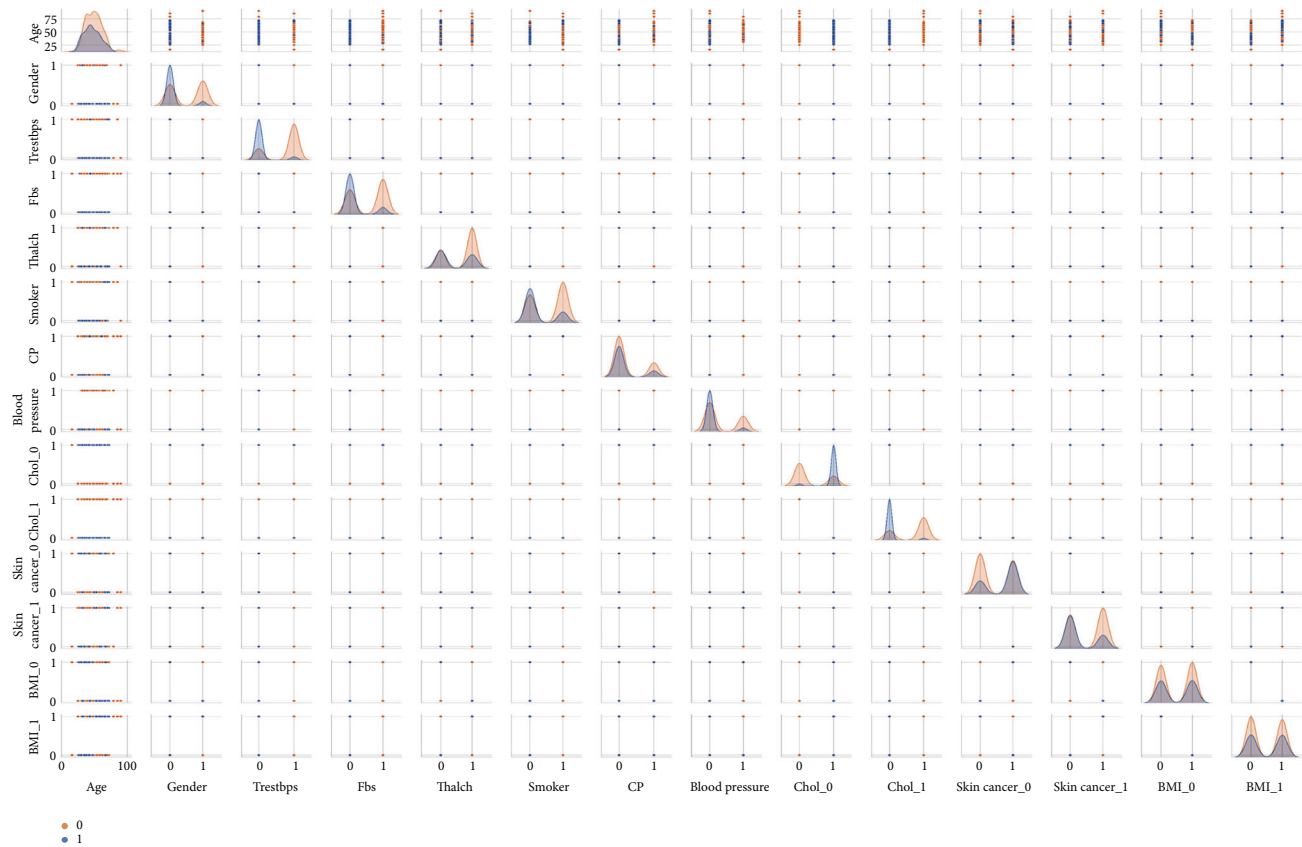


FIGURE 5: Scatterplot in a data set representation.

experimental and analytical techniques. Figure 3 displays the histogram that was created in addition to the plots that depict the distribution of each dataset attribute.

5.1. Model Accuracy. Twelve features are used in the development of the prediction models, and the modelling techniques' accuracy is evaluated. Figure 4 compares multiple algorithms and shows the accuracy numbers so that we

can better understand the variations. MLA style depends on their consistency. The comparison shows that RF and KNN are more accurate than the other models. The following bar graph illustrates how accurate various algorithms are depicted.

Six machine learning algorithms were used in this paper for predicting heart disease. The relationship between the features used in the dataset is depicted in the scatterplot in

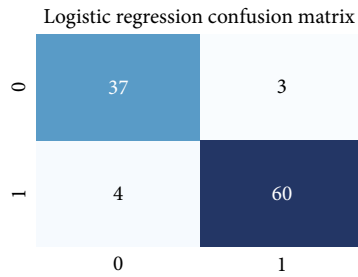


FIGURE 6: Logistic algorithms.

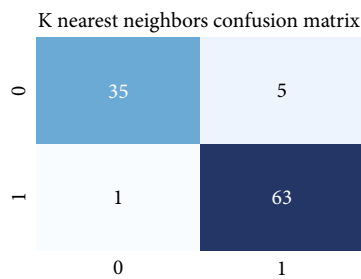


FIGURE 7: K-nearest neighbor.

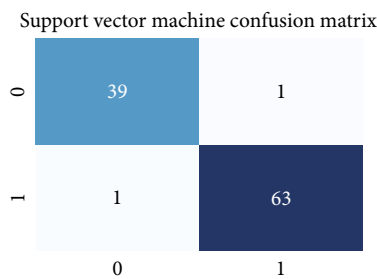


FIGURE 8: Support vector machine.

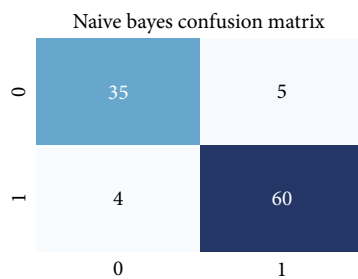


FIGURE 9: Naïve Bayes algorithm.

Figure 5. For each dot's location along the X and Y axes, the values that are utilised to quantify a specific data point are displayed.

In machine learning algorithms, the performance of the algorithms is evaluated using a confusion matrix. In a tabular arrangement, the rows reflect the actual values, the columns the expected values, and the rows display the actual

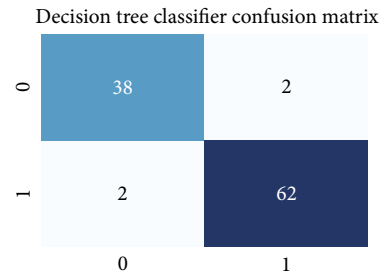


FIGURE 10: Decision tree algorithm.

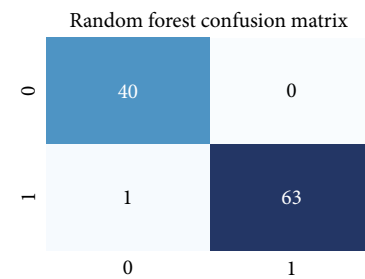


FIGURE 11: Random forest algorithm.

values. In Figures 6–11, these classifier confusion matrices are displayed. The performance assessment of ML models is checked using the confusion matrix to look for mistakes or miscalculations while predicting heart disease. Based on four factors, including true positive (TP), true negative (TN), false positive (FP), and false negative, it compares the actual results with the predicted ones (FN). The different ML classifiers have been analyzed using statistical metrics including accuracy, precision, recall, F measure, and MCC using the confusion matrices.

Additionally as shown in Figure 12, certain other statistical measures are also calculated. The machine learning classifiers are evaluated using these parameters. Accuracy, precision, recall, F measure, and MCC are some of the different parameters.

5.2. Comparative Analysis. Table 2 compares the effectiveness of our proposed framework with a variety of relevant types of literature in terms of the methodologies employed, the dataset, and the analysis. Most cardiac markers are consistent throughout all studies conducted for comparison with the suggested study. It was discovered that our well-planned approach produced positive outcomes for several evaluation measures, especially accuracy for the prediction of heart disease. The employment of techniques like data imputation for handling missing values, scatterplot method for identifying and replacing outliers, and transformation method for standardizing and normalizing data has led to superior outcomes than those of other relevant research. When creating the proposed framework, the K-fold cross-validation technique was used to get results that were more reliable than those from similar research.

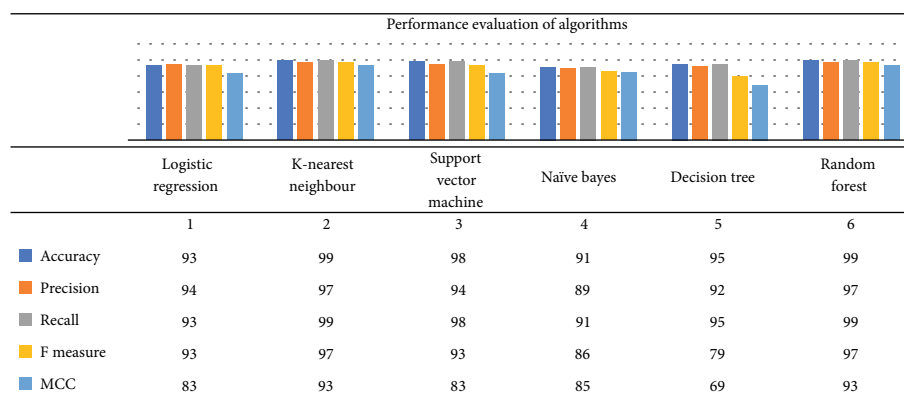


FIGURE 12: Performance evaluation of algorithms.

TABLE 2: Comparison with the existing system.

Authors	Techniques used	Dataset	Accuracy
Chauhan et al. [24]	RF, LR, SVM	PIDD data set	78%
Pouriyeh et al. [25]	KNN, SVM, RF, NB	Cleveland data	89%
Kedia et al. [26]	LR, KNN, SVM, DT	UCI data	90%
Atallah et al. [27]	SVM, DT KNN, RF	Cleveland data	87%
Proposed method	LR, KNN, SVM, NB, DT, RF	Cleveland Heart from UCI's machine learning	99.04%

6. Conclusions

Comparing various ML for the early detection of heart disease is the main contribution, preprocessing techniques were used to enhance the dataset's quality. With the primary objectives being the handling of corrupted and missing values as well as the removal of outliers in order to predict the illness. Additionally, we used a variety of machine learning techniques, and the outcomes were compared using various statistical metrics. The experimental finding indicates a 70:30 ratio between testing and training the data. In this study, we perform 10-fold cross-validation to a number of machine learning methods, and we find that random forest and k-nearest neighbor are 99.04% accurate compared to other algorithms. Future work can be carried out using various combinations of machine learning methodologies to enhance prediction techniques. For the purpose of better comprehension of the critical features and increasing the precision of heart disease prediction, new feature selection approaches can also be developed.

Data Availability

The data used to support the findings of this study are available from the first author upon request (gufran.ansari@mitwpu.edu.in).

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

All authors have contributed equally to this work and have also read and agreed to submit the current version of the manuscript to this journal.

Acknowledgments

This study is supported via funding from the Prince Sattam Bin Abdulaziz University, project number (PSAU/2023/R/1444).

References

- [1] K. Battula, R. Durgadinesh, K. Suryapratap, and G. Vinaykumar, "Use of machine learning techniques in the prediction of heart disease," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pp. 1–5, Mauritius, Mauritius, 2021.
- [2] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.
- [3] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," *Journal of Intelligent Learning Systems and Applications*, vol. 5, no. 3, pp. 176–183, 2013.
- [4] V. Vijayaganth and M. Naveenkumar, "Smart sensor based prognostication of cardiac disease prediction using machine learning techniques," in *Applications of Machine Learning in Big-Data Analytics and Cloud Computing*, pp. 63–80, River Publishers, 2022.
- [5] H. B. Kibria and A. Matin, "The severity prediction of the binary and multi-class cardiovascular disease – a machine

- learning-based fusion approach," *Computational Biology and Chemistry*, vol. 98, article 107672, 2022.
- [6] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," in *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, pp. 520–525, Greater Noida, India, 2015.
 - [7] S. S. Bhat, V. Selvam, G. A. Ansari, M. D. Ansari, and M. H. Rahman, "Prevalence and early prediction of diabetes using machine learning in North Kashmir: a case study of district Bandipora," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2789760, 12 pages, 2022.
 - [8] P. Melillo, N. de Luca, M. Bracale, and L. Pecchia, "Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 727–733, 2013.
 - [9] S. Ahmad, S. Khan, M. Fahad AlAjmi et al., "Deep learning enabled disease diagnosis for secure internet of medical things," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 965–979, 2022.
 - [10] A. A. Ahdal, M. Rakhra, S. Badotra, and T. Fadhaeel, "An integrated machine learning techniques for accurate heart disease prediction," in *2022 International Mobile and Embedded Technology Conference (MECON)*, pp. 594–598, Noida, India, 2022.
 - [11] A. Noor, L. Ali, H. T. Rauf, U. Tariq, and S. Aslam, "An integrated decision support system for heart failure prediction based on feature transformation using grid of stacked autoencoders," *Measurement*, vol. 205, article 112166, 2022.
 - [12] Y. A. Nanekaran, Z. Licai, J. Chen et al., "Anomaly detection in heart disease using a density-based unsupervised approach," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6913043, 14 pages, 2022.
 - [13] M. Marimuthu, M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra, "A review on heart disease prediction using machine learning and data analytics approach," *International Journal of Computer Applications*, vol. 181, no. 18, pp. 20–25, 2018.
 - [14] B. L. Y. Agbley, J. P. Li, A. U. Haq et al., "Federated Fusion of Magnified Histopathological Images for Breast Tumor Classification in the Internet of Medical Things," in *IEEE Journal of Biomedical and Health Informatics*, 2023.
 - [15] I. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, "Multi-label active learning-based machine learning model for heart disease prediction," *Sensors*, vol. 22, no. 3, p. 1184, 2022.
 - [16] S. Goel, A. Deep, S. Srivastava, and A. Tripathi, "Comparative analysis of various techniques for heart disease prediction," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, pp. 88–94, Mathura, India, 2019.
 - [17] R. Poonguzhali, S. Ahmad, P. T. Sivasankar et al., "Automated brain tumor diagnosis using deep residual u-net segmentation model," *Computers, Materials & Continua*, vol. 74, no. 1, pp. 2179–2194, 2023.
 - [18] F. Ma, T. Sun, L. Liu, and H. Jing, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network," *Future Generation Computer Systems*, vol. 111, pp. 17–26, 2020.
 - [19] Y. A. Nanekaran, Z. Licai, J. Chen et al., "Diagnosis of chronic diseases based on patients' health records in iot healthcare using the recommender system," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 5663001, 14 pages, 2022.
 - [20] A. Tsanas, "Relevance, redundancy, and complementarity trade-off (RRCT): a principled, generic, robust feature-selection tool," *Patterns*, vol. 3, no. 5, article 100471, 2022.
 - [21] A. Alharbi, K. Equbal, S. Ahmad, H. U. Rahman, and H. Alyami, "Human gait analysis and prediction using the Levenberg-Marquardt method," *Journal of Healthcare Engineering*, vol. 2021, Article ID 5541255, 11 pages, 2021.
 - [22] S. Ahmad, H. A. Abdeljaber, J. Nazeer, M. Y. Uddin, V. Lingamuthu, and A. Kaur, "Issues of clinical identity verification for healthcare applications over mobile terminal platform," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6245397, 10 pages, 2022.
 - [23] G. A. Ansari and S. S. Bhat, "Exploring a link between fasting perspective and different patterns of diabetes using a machine learning approach," *Educational Research*, vol. 12, no. 2, pp. 500–517, 2022.
 - [24] T. Chauhan, S. Rawat, S. Malik, and P. Singh, "Supervised and unsupervised machine learning based review on diabetes care," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 581–585, Coimbatore, India, 2021.
 - [25] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, "A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 204–207, Heraklion, Greece, 2017.
 - [26] S. Kedia and M. Bhushan, "Prediction of mortality from heart failure using machine learning," in *2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, pp. 1–6, Patna, India, 2022.
 - [27] R. Atallah and A. Al-Mousa, "Heart disease detection using machine learning majority voting ensemble method," in *2019 2nd International Conference on New Trends in Computing Sciences (Ictcs)*, pp. 1–6, Amman, Jordan, 2019.