STM JOURNALS

International Journal of
Computer Science Languages

Volume 2, Issue 1, 2024
January–June
DOI (Journal): 10.37591/IJCSL

https://journals.stmjournals.com/ijcsl

Review IJCSL

Optimizing Heart Disease Prediction: Comparative Analysis of Machine Learning Algorithm for Early Detection

Purushotam Naidu K.[1,*], K. Roshini[2], J. Sravanthi[2], T. Kanchana Rekha[2], N. Sirisha[2], M. Thanushya[2]

Abstract

The expanding realm of data analysis holds considerable importance in healthcare, particularly in the medical sector where forecasting heart disease is considered a complex endeavor. Early prediction of serious health conditions can be the determining factor between survival and fatality, with heart disease being one such critical health issue. Over the past decade, the main reason for death has been heart disease. Heart disorders come in many different forms, and they are often referred to as cardiovascular diseases. These can range from heart rhythm issues to birth abnormalities to illnesses of the blood vessels. For several decades, it has continued to be the leading cause of death worldwide. It is imperative to find a precise and trustworthy method for automating the task in order to detect the sickness early and manage it effectively. Machine learning, a prominent application of artificial intelligence, is making significant strides in various research domains. This study examines supervised learning models, including logistic regression, naïve Bayes, support vector machine, K-nearest neighbors, decision tree, random forest, and the ensemble technique XGBoost, offering a comparative analysis to identify the most effective algorithm. Results indicate that random forest achieves the highest accuracy at 90.16% compared to other algorithms.

Keywords: Classification accuracy, logistic regression, naïve bayes, support vector machine, K-nearest neighbor, decision tree, random forest

INTRODUCTION

Since heart disease instances are increasing daily, it is both important and concerning to be aware of potential heart disease cases in advance. Making this diagnosis is a difficult undertaking that calls for efficiency and accuracy. The primary objective of the research article is to pinpoint individuals with an elevated risk of developing heart disease by considering various medical criteria. We created a heart disease prediction algorithm based on the patient's medical history to ascertain whether or not a heart disease diagnosis is likely. It can improve the decision-making abilities of healthcare professionals in terms of identifying an appropriate treatment plan and aid in the process of accurate diagnosis. These characteristics have a dataset that is gathered and examined. Gathering information from the right source is essential when building a prediction model since it affects the model's accuracy.

In this work, we aim to create a comprehensive predictive model for heart disease by utilizing machine learning (ML) algorithms such as naive

***Author for Correspondence**
Purushotam Naidu K.
E-mail: purushotam.k30@gmail.com

[1]Assistant Professor, Department of Computer Science and Engineering (AI & ML), GVP College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India
[2]Student, Department of Computer Science and Engineering (AI & ML), GVP College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

Received Date: March 28, 2024
Accepted Date: April 08, 2024
Published Date: May 9, 2024

Citation: Purushotam Naidu K., K. Roshini, J. Sravanthi, T. Kanchana Rekha, N. Sirisha, M. Thanushya. Optimizing Heart Disease Prediction: Comparative Analysis of Machine Learning Algorithm for Early Detection. International Journal of Computer Science Languages. 2024; 2(1): 1–10p.

Bayes, decision trees, K-nearest neighbors (KNN), support vector machine (SVM), and logistic regression. Although helpful, traditional risk variables frequently fail to capture the subtle correlations seen in complex datasets. We can model these complex patterns and improve the precision of predictive analytics by using machine learning techniques like logistic regression and naive Bayes. SVM also offers a strong framework for patient classification according to individual cardiovascular risk profiles.

We can now consider the closeness of comparable patient instances thanks to the addition of KNN, which gives the prediction model a more regional viewpoint. The commonly utilized methods are decision trees and random forest algorithms. By leveraging a comprehensive dataset encompassing clinical, lifestyle, and demographic information, we aim to create a holistic model that amalgamates the strengths of these diverse ML algorithms. The iterative refinement of our model through these algorithms ensures adaptability to varying patient profiles, fostering a nuanced understanding of cardiovascular risk factors. This project endeavors to not only enhance the accuracy of heart disease prediction but also to provide healthcare professionals with a flexible and interpretable tool for personalized risk assessment. The amalgamation of clinical expertise with the strengths of logistic regression, naive Bayes, SVM, KNN, decision trees, and random forests underscores our commitment to advancing proactive and personalized approaches to cardiovascular health.

LITERATURE SURVEY

The study by Hossen [1] compares classification algorithms for predicting heart disease using pre-processed data from the UCI dataset. Random forest (86.89%) and XGBoost (78.69%) achieved the highest accuracy. More complex models are suggested for improving early heart disease prediction.

The study by Yadav et al. [2] found that decision trees were less accurate for small datasets, whereas naïve Bayes was more precise when considering multiple possibilities. Despite ID3's adaptability, naïve Bayes consistently provided accurate results, indicating multiple algorithms. Combining naïve Bayes and K-means improves accuracy by analyzing individual variables. The use of machine learning to predict heart disease demonstrates its significant impact on healthcare, with future advancements likely to improve its efficiency. Bharti et al. [3] compared three different analysis methods, confirming machine learning's superiority over human performance. The recommendations for using ML with smaller datasets were validated. The assessment employed the following evaluation techniques: confusion matrix, precision, specificity, sensitivity, and F1 score. After data preprocessing, the K-neighbors classifier outperformed ML techniques on 13 dataset features, resulting in a significant reduction in deployment computing time. Normalization was identified as critical to avoiding overfitting and ensuring accurate model evaluation.

The paper by Malavika et al. [4] employed six approaches to enhance the precision of predicting heart disease. Random forest has the highest accuracy (91.8%), outperforming naive Bayes, logistic regression, SVM, decision tree, and KNN. The study by Biswas et al. [5] utilized six methodologies to enhance the accuracy of predicting heart disease. SVM and logistic regression performed well, but insufficient data hampered efficacy, emphasizing the need for more real-world patient data. Future research will focus on deep learning and better feature selection techniques. The study by Nashif et al. [6] looked at computer-based heart disease prediction methods and preferred SVM. An app was created to monitor patients' heart health. A sensor-computer system for continuous monitoring was developed, with recommendations for improvements. Shah et al. [7] assessed the effectiveness of the heart disease prediction model using criteria such as accuracy and precision. The random forest classification approach performed well in estimating heart conditions. The study suggested employing a support vector classifier to enhance predictions through machine learning methodologies. Machine learning algorithms, including naïve Bayes and random forest, accurately predict heart disease based on patient data analysis [8]. Random forest achieved an accuracy of 89.4%, demonstrating cost-saving and healthcare improvement potential. Future research may look into larger datasets and algorithm modifications to improve predictions. The experiment shows that KNN achieved 86.885% accuracy,

while random forest achieved 81.967% [7]. This demonstrates machine learning's potential for predicting heart disease, a major societal issue. Further advancements could improve healthcare solutions. The study by Jindal et al. [9] discovered that KNN, logistic regression, and random forest classifier performed well in detecting heart disease, with up to 88.5% accuracy using more medical data attributes. KNN had the highest accuracy, 88.52%. They tested various computer methods and chose logistic regression due to its 95% accuracy [7].

The confusion matrix and F1-score validated the model's accuracy and recall. They proposed using this approach to diagnose other diseases and combining different computer techniques for improved analysis, which could aid in the prediction of various health conditions. Mahmoud et al. [10] used the R program to analyze a heart disease dataset, ensuring data accuracy. They trained the computer to recognize patterns and found random forest to be the best predictor, with 85.05% accuracy. The study by Singh and Kumar [11] analyzed a heart disease dataset, finding more cases in men than women. Certain factors like 'cp', 'thalach', and 'slope' were linked to heart disease. Logistic regression accurately predicted 95% of cases.

The experiments tested different computer programs to predict heart disease [12]. Logistic regression worked well in most cases, identifying about 91.6% and 90.8% of cases. In one dataset, random forest did even better, identifying 98.6% of cases. The study by Gupta et al. [13] found logistic regression best at predicting heart problems, being right 92 times out of 100. They aim to explore random forest classifier and SVM for improved predictions.

METHODOLOGY

Logistic Regression

Logistic regression stands out as a widely used supervised machine learning technique, particularly adept at predicting categorical outcomes by assigning probabilities ranging from 0 to 1 as shown in the Figure 1. Its domain lies primarily in classification tasks, where it discerns between binary outcomes like yes/no or true/false. Unlike linear regression, logistic regression fits a sigmoid-shaped curve to model the likelihood of categorical outcomes. This curve, known as the logistic function, transforms real values into probabilities, aiding in classification tasks by assigning probabilities to new data points. The logistic regression equation, derived from the linear regression equation, provides the framework for this probabilistic modeling. There exist three primary categories of logistic regression based on the characteristics of the dependent variable: binomial, multinomial, and ordinal. Implementing logistic regression in Python for a binomial classification task involves several steps. Initially, data preprocessing is crucial, including loading the dataset, extracting independent and dependent variables, and splitting the data into training and testing sets. Feature scaling using Standard Scaler ensures uniformity in data representation. Subsequently, logistic regression is fitted to the training data using scikit-learn's logistic regression class. Predictions are then made on the test set, and the accuracy of these predictions is assessed using confusion matrix analysis.

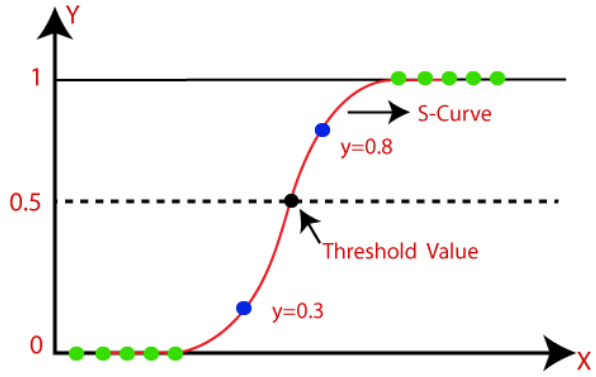

Figure 1. Logistic regression.

Naïve Bayes

The naïve Bayes classifier algorithm is a powerful tool in supervised learning, particularly suited for classification tasks. Leveraging Bayes' theorem, it finds widespread application in scenarios like text classification, spam filtration, sentiment analysis, and article categorization, especially in contexts with high-dimensional datasets. The term "naïve" in its name stems from the algorithm's assumption of feature independence, simplifying calculations but potentially oversimplifying real-world relationships. Meanwhile, "Bayes" highlights its reliance on Bayes' theorem for probability estimation. Characterized by its simplicity and efficiency, Naïve Bayes is renowned for quickly building effective machine learning models. It operates as a probabilistic classifier, determining class membership based on the probability of an object belonging to a specific category. Bayes' theorem, which forms the foundation of naïve Bayes, computes the probability of a hypothesis considering previous information and observed evidence. It factors in the likelihood of the evidence given the hypothesis, the prior probability of the hypothesis, and the marginal probability of the evidence. The advantages of the naïve Bayes classifier include its speed, simplicity, and suitability for both binary and multi-class classifications. It excels particularly in multi-class predictions and finds widespread use in text classification tasks. However, naïve Bayes comes with limitations. The graphical representation of naïve Bayes classifier is as shown in Figure 2.

Support Vector Machine

The SVM algorithm is a widely utilized supervised learning method, primarily employed for classification tasks in machine learning. Its versatility extends to regression problems as well, but its main application lies in classification. The primary objective of SVM is to establish an optimal line or decision boundary, known as a hyperplane, within an n-dimensional space, effectively segregating different classes as shown in Figure 3.

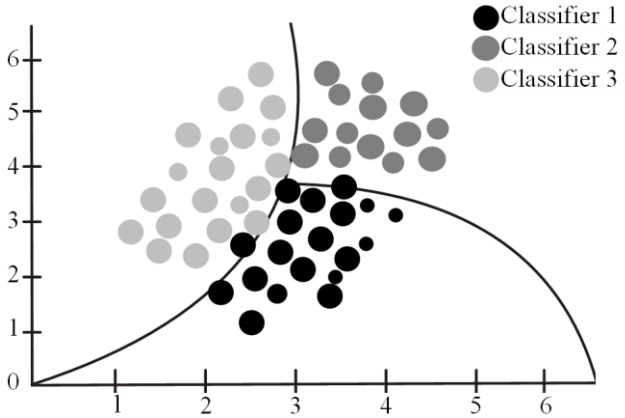

Figure 2. Naïve Bayes.

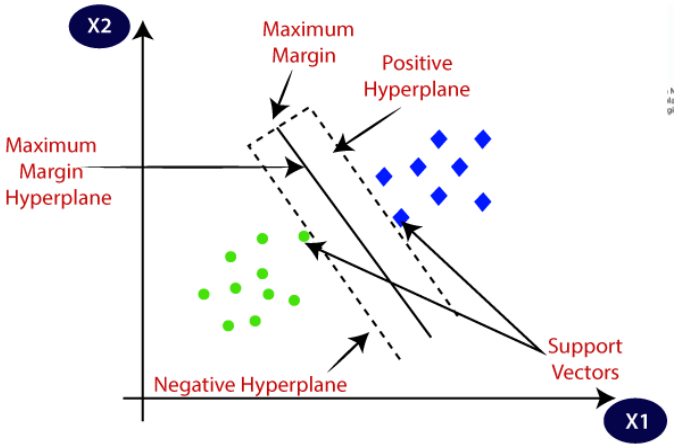

Figure 3. Support vector machine.

Types of SVM:

1. *Linear SVM:* Applied to datasets that are linearly separable, indicating they can be divided into two classes using a solitary straight line. In such cases, a linear SVM classifier is employed to create the decision boundary.
2. *Non-linear SVM:* Employed for datasets that cannot be separated by a straight line, known as non-linear data. In such scenarios, a non-linear SVM classifier is utilized to establish a decision boundary that effectively separates the classes despite their non-linear distribution. These two types of SVM classifiers cater to different data characteristics, ensuring SVM's applicability across diverse datasets and classification tasks.

K-Nearest Neighbors

K-nearest neighbors (KNN) stands as a fundamental yet pivotal classification algorithm within the realm of ML. Operating within the domain of supervised learning, it holds significant relevance in fields such as pattern recognition, data mining, and intrusion detection. Its applicability extends widely across real-life scenarios owing to its non-parametric nature, which implies that it refrains from making any underlying assumptions regarding the data distribution, unlike certain algorithms such as Gaussian mixture models (GMM). Representation of KNN is as shown in Figure 4.

Decision Tree

A decision tree is a hierarchical structure resembling a tree, utilized in supervised ML. Each internal node corresponds to a feature, branches depict decision rules, and leaf nodes signify the outcomes of the decision-making process. It serves as a versatile algorithm for both classification and regression tasks, being highly effective in various scenarios. Moreover, decision trees are fundamental components of ensemble methods like random forests, contributing to their robustness and efficacy in ML applications. Representation of decision tree is as shown in Figure 5.

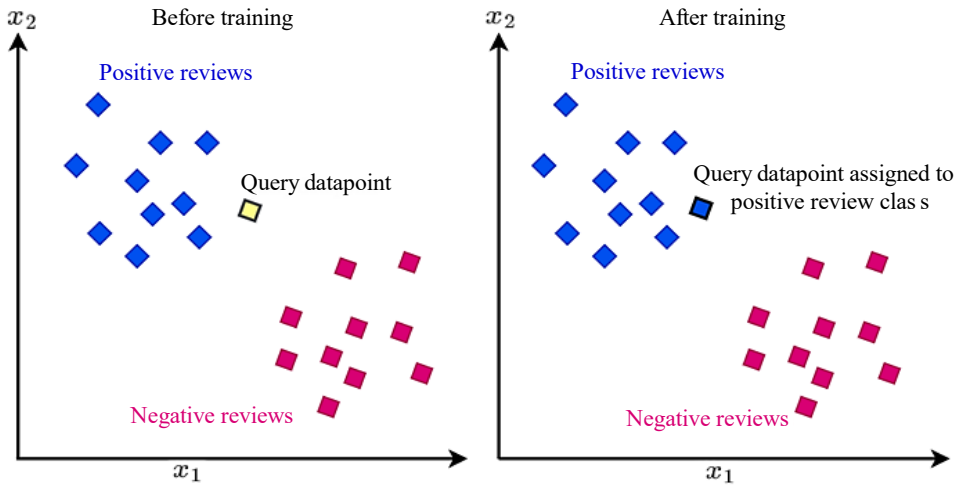

Figure 4. K-nearest neighbors.

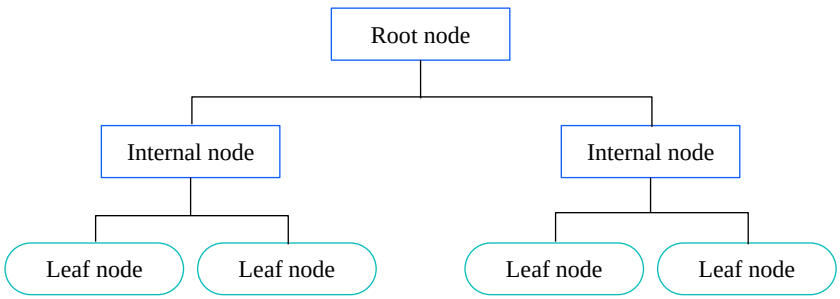

Figure 5. Decision tree.

Terminologies associated with decision trees:

1. *Root node:* The initial node at the top of the tree, representing the entire dataset, marking the starting point of the decision-making process.
2. *Decision/internal node:* A node representing a decision based on a specific input feature, leading to further branching or connecting to leaf nodes.
3. *Leaf/terminal node:* A node without child nodes, indicating the final class label or numerical value prediction.
4. *Splitting:* The process of dividing a node into multiple sub-nodes based on a chosen feature and split criterion.
5. *Branch/sub-tree:* A subsection of the decision tree starting from an internal node and extending to leaf nodes.
6. *Parent node:* The node that extends into one or multiple child nodes.
7. *Child node:* Nodes resulting from the splitting of a parent node.
8. *Impurity:* A measure of how similar the target variable is within a subset of data, indicating the level of randomness or uncertainty. Common impurity measures include the Gini index and entropy for classification tasks.
9. *Variance:* A measure of how much the predicted and target variables differ across different samples of a dataset, relevant for regression problems in decision trees. Several metrics like mean squared error, mean absolute error, among others, are employed to measure variance.
10. *Information gain:* A measure indicating the decrease in uncertainty obtained by dividing a dataset based on a particular feature.
11. *Pruning:* The process of removing unnecessary branches from the tree to prevent overfitting or improve efficiency by eliminating redundant information. Pruning ensures a simpler and more interpretable decision tree model.

Random Forest

The random forest method expands upon the bagging technique by merging bagging and feature randomness to produce an uncorrelated collection of decision trees known as a forest. Feature randomization, also referred to as feature bagging or "the random subspace method," involves selecting a random subset of features, ensuring minimal correlation among decision trees. This marks a significant difference between decision trees and random forests: while decision trees consider all available feature splits, random forests only choose a subset. Random forest algorithms rely on three primary hyperparameters, namely node size, the number of trees, and the number of features sampled, which must be set prior to training. Subsequently, the random forest classifier can be employed to address regression or classification tasks. The random forest algorithm consists of multiple decision trees, with each tree in the ensemble derived from a data sample drawn from a training set with replacement, termed the bootstrap sample. Within this training sample, one-third is allocated as test data, known as the out-of-bag (oob) sample, which will be revisited later. Additional randomness is introduced through feature bagging, augmenting dataset diversity and diminishing correlation.

System Architecture

The heart disease prediction system architecture involves several key stages. Initially, patient data is collected, encompassing various attributes such as age, gender, chest pain type, and physiological metrics like blood pressure and cholesterol levels. This data undergoes preprocessing, where cleaning, handling of missing values, and normalization are performed to ensure consistency. Feature engineering techniques may then be applied to extract relevant features or enhance existing ones. Following this, ML models are trained on the preprocessed data, including algorithms like naïve Bayes, KNN, and decision trees. Model evaluation is conducted using metrics like accuracy and precision to select the best-performing model for deployment. Once deployed, the model makes predictions on new patient data, with continuous monitoring for performance and feedback to refine the system over time. This architecture streamlines the process from data collection to real-time prediction, with a focus on accuracy and adaptability as shown in Figure 6.

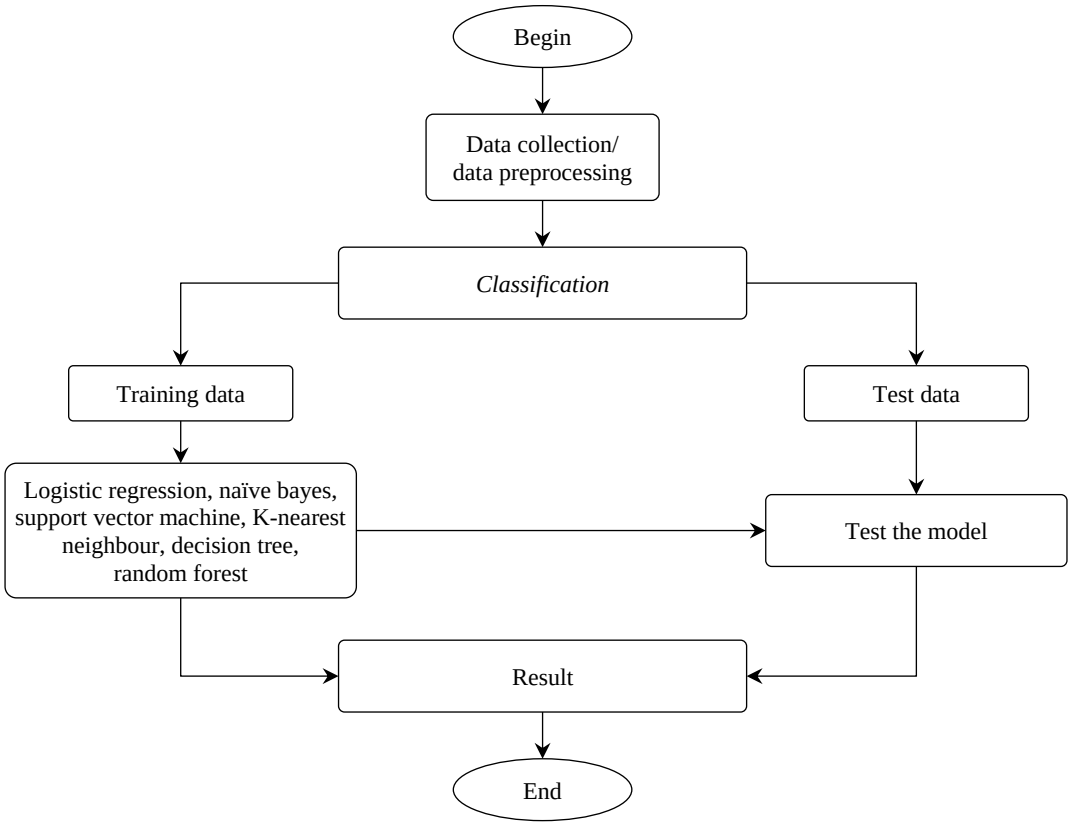

Figure 6. Archiecture diagram.

RESULTS AND DISCUSSION

With an accuracy score of 85.25%, logistic regression demonstrated a significantly higher level of performance, demonstrating its effectiveness in calculating the risk of heart disease incidence based on input features. This technique is highly suited for classification jobs since it can model binary outcomes. Utilizing the correlations between characteristics and the target variable, logistic regression offers a strong framework for heart disease prediction by evaluating the chance of a specific outcome using a logistic function.

Despite its naive assumptions, naive Bayes produced a respectable accuracy score of 85.25%, indicating its effectiveness in heart disease prediction. Naive Bayes did an amazing job of identifying patterns in the dataset, even with its "naive" assumption of feature independence. Naive Bayes offers a probabilistic framework for classification by utilizing the joint distribution of features and the target by computing the conditional probabilities of each feature given the target variable.

SVM nearly matched logistic regression with an accuracy score of 81.97%. Because SVM can find the optimal hyperplane for partitioning classes in a high-dimensional space, it can be used to distinguish between individuals with and without heart disease. SVMs are a dependable choice for classification problems such as heart disease prediction because they optimize the margin between classes, which attempts to generalize well to new data.

With a moderate accuracy score of 67.21%, KNN demonstrated its limits in terms of identifying the underlying patterns in the data. Although KNN is simple to use and intuitive, the intricate relationships between features may not be well captured by its reliance on local voting based on nearest neighbors. This implies that KNN may have trouble with high-dimensional datasets, such as those used to forecast heart disease, since the similarity between instances may not be well-represented by their distance from one another.

Decision tree performed reasonably well, with an accuracy rate of 81.97%. Decision trees can capture complicated interactions between information, making them ideal for classification problems such as predicting heart disease. However, their tendency to overfit the training data may have hindered their performance when compared to other methods. Pruning strategies and ensemble methods, such as random forest, can help reduce overfitting and improve decision tree performance.

Out[5]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Figure 7. Displaying first five rows in the dataset.

Out[6]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
251	43	1	0	132	247	1	0	143	1	0.1	1	4	3	0
268	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
56	48	1	0	122	222	0	0	186	0	0.0	2	0	2	1
199	65	1	0	110	248	0	0	158	0	0.6	2	2	1	0
185	44	1	0	112	290	0	0	153	0	0.0	2	1	2	0

Figure 8. Random samples from dataset.

Out[7]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.00
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.31
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.61
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.00
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.00
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.00
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.00
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.00

Figure 9. Summary statistics for dataset columns.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       303 non-null    int64
 1   sex       303 non-null    int64
 2   cp        303 non-null    int64
 3   trestbps  303 non-null    int64
 4   chol      303 non-null    int64
 5   fbs       303 non-null    int64
 6   restecg   303 non-null    int64
 7   thalach   303 non-null    int64
 8   exang     303 non-null    int64
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64
 11  ca        303 non-null    int64
 12  thal      303 non-null    int64
 13  target    303 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

Figure 10. Data frame summary: heart disease dataset.


```
age:                age
sex:                1: male, 0: female
cp:                 chest pain type, 1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic
trestbps:                  resting blood pressure
chol:                serum cholestoral in mg/dl
fbs:                fasting blood sugar > 120 mg/dl
restecg:                   resting electrocardiographic results (values 0,1,2)
thalach:                    maximum heart rate achieved
exang:              exercise induced angina
oldpeak:                   oldpeak = ST depression induced by exercise relative to rest
slope:              the slope of the peak exercise ST segment
ca:                 number of major vessels (0-3) colored by flourosopy
thal:               thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
```

Figure 11. Dataset attributes description.

```
[10]: count    303.000000
      mean       0.544554
      std        0.498835
      min        0.000000
      25%        0.000000
      50%        1.000000
      75%        1.000000
      max        1.000000
      Name: target, dtype: float64
```

Figure 12. Summary statistics for target variable in dataset.

```
target      1.000000
exang       0.436757
cp          0.433798
oldpeak     0.430696
thalach     0.421741
ca          0.391724
slope       0.345877
thal        0.344029
sex         0.280937
age         0.225439
trestbps    0.144931
restecg     0.137230
chol        0.085239
fbs         0.028046
Name: target, dtype: float64
```

Figure 13. Correlation of features with target variable in dataset.

```
The accuracy score achieved using Logistic Regression is: 85.25 %
The accuracy score achieved using Naive Bayes is: 85.25 %
The accuracy score achieved using Support Vector Machine is: 81.97 %
The accuracy score achieved using K-Nearest Neighbors is: 67.21 %
The accuracy score achieved using Decision Tree is: 81.97 %
The accuracy score achieved using Random Forest is: 90.16 %
```

Figure 14. Accuracy scores of machine learning algorithms for heart disease prediction.

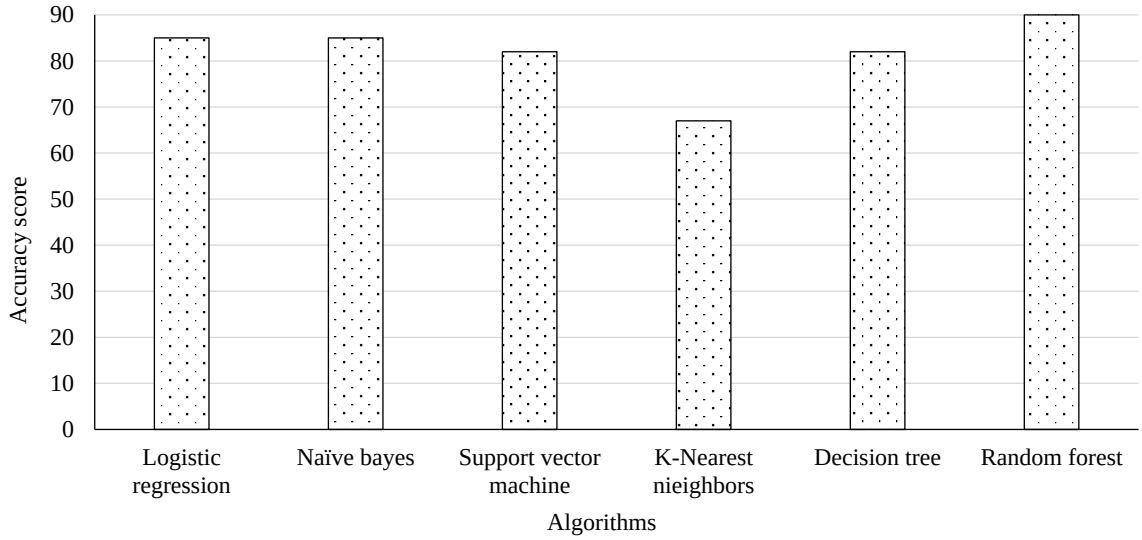

Figure 15. Bar chart showing accuracy comparison of machine learning algorithms.

Random forest has proven to be a valuable asset for predicting heart disease, achieving an accuracy rate of 90.16%. Because random forest combines many decision trees and aggregates their predictions, it reduces overfitting and improves generalization performance, making it a good choice for this classification task. Random forest can handle high-dimensional data and understand complex correlations between features, it is a reliable algorithm for heart disease prediction as shown in Figures 7–15.

CONCLUSION

In conclusion, the future of cardiac disease prediction with ML is a dynamic environment characterized by ongoing innovation. The integration of many data sources increases in interpretability, real-time monitoring, federated learning, improved imaging techniques, and individualized treatment plans all contribute to a more holistic and successful approach to cardiovascular health. As technology and healthcare mix, the prospect of employing machine intelligence to improve preventive cardiology remains appealing and transformative. Improvements in explainable artificial intelligence increase transparency, ultimately fostering trust between healthcare professionals and patients. The combination of ML and advanced imaging techniques leads to more accurate risk factor detection. Federated learning addresses privacy concerns by allowing for collaborative model training while safeguarding sensitive healthcare data. The seamless integration of ML with electronic health records enables a thorough examination.

REFERENCES

1. Hossen MK. Heart disease prediction using machine learning techniques. Am J Computer Sci Technol. 2022; 5 (3): 146–154.
2. Yadav A, Gediya L, Kazi A. Heart disease prediction using machine learning. Int Res J Eng Technol. 2021; 8 (9): 1325–1329.
3. 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. Comput Intell Neurosc. 2021; 2021: Article 8387680.
4. Malavika G, Rajathi N, Vanitha V, Parameswari P. Heart disease prediction using machine learning algorithms. Biosci Biotechnol Res Commun. 2020; 13 (11): 24–27.
5. Biswas N, Ali MM, Rahaman MA, Islam M, Mia MR, Azam S, Ahmed K, Bui FM, Al-Zahrani FA, Moni MA. Machine learning-based model to predict heart disease in early stage employing different feature selection techniques. BioMed Res Int. 2023; 2023: Article 6864343.
6. Nashif S, Raihan MR, Islam MR, Imam MH. Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. World J Eng Technol. 2018; 6 (4): 854–873.
7. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Computer Sci. 2020; 1 (6): Article 345.
8. Ahmed I. A Study of Heart Disease Diagnosis Using Machine Learning and Data Mining. MSc Thesis. San Bernardino, CA, USA: California State University San Bernardino; 2022.
9. Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using machine learning algorithms. IOP Conf Ser Mater Sci Eng. 2021; 1022 (1): 012072.
10. Mahmoud WA, Aborizka M, Amer FA. Heart disease prediction using machine learning and data mining techniques: Application of Framingham dataset. Turk J Computer Math Educ. 2021; 12 (14): 4864–4870.
11. Singh A, Kumar R. Heart disease prediction using machine learning algorithms. In: 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, February 14–15, 2020. pp. 452–457.
12. Radwan M, Mohamed Abdelrahman N, Wael Kamal H, Khaled Abdelmonem Elewa A, Moataz Mohamed A. MLHeartDisPrediction: heart disease prediction using machine learning. J Comput Commun. 2023; 2 (1): 50–65.
13. Gupta C, Saha A, Reddy NS, Acharya UD. Cardiac disease prediction using supervised machine learning techniques. J Phys Conf Ser. 2022; 2161 (1): 012013.

