



Original Article

Comparative Analysis of Logistic Regression and Decision Tree Models for Predicting Heart Disease Outcomes

Dr Syed Shafi Ahmed¹, Dr Arun Kumar Yadav², Prof. Arshiya Masood Siddiqui^{3*}, Dr Shivam Kamthan⁴

^{1,3,4}Government Medical College, Badaun, Uttar Pradesh, India

²Associate Professor, Hind Institute of Medical Sciences, Sitapur, Uttar Pradesh, India

 OPEN ACCESS

ABSTRACT

Cardiovascular disease has become a significant global health issue and remains one of the leading causes of mortality, requiring advanced and often costly detection methods. Heart failure, in particular, poses a severe threat to individuals, contributing to increased morbidity and mortality rates. Therefore, accurate prediction and diagnosis are essential to enable early intervention, timely detection, and effective treatment, reducing the life-threatening risks associated with heart disease—a challenge that persists in medical practice. Individuals diagnosed with or at high risk for cardiovascular disease, due to factors such as hypertension, diabetes, hyperlipidemia, or pre-existing conditions, need prompt identification and efficient management strategies. In this context, machine learning (ML) models play a pivotal role. Our study employed two ML techniques, including Logistic Regression (LR) and Decision Tree (DT), which yielded promising results. A comparative analysis of these algorithms was conducted to evaluate their predictive performance. The findings revealed that the Logistic regression achieved superior accuracy compared to the other model.

Corresponding Author:

Prof. Arshiya Masood Siddiqui
Government Medical College,
Badaun, Uttar Pradesh, India

Received: 16-03-2026

Accepted: 03-04-2026

Published: 17-04-2026

Copyright© International Journal of
Medical and Pharmaceutical Research

Keywords: Heart Disease Prediction, Logistic Regression, Decision Tree, Machine Learning, Risk Factors, Predictive Modeling.

INTRODUCTION

Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for approximately 20.5 million deaths in 2021, which represents nearly one-third of all global deaths. This high prevalence underscores the critical need for accurate and cost-effective diagnostic approaches to facilitate early detection and intervention. Traditional diagnostic methods, while effective, often involve significant costs and may not always provide the necessary predictive accuracy for timely intervention. In recent years, the integration of machine learning (ML) into healthcare has shown promise in enhancing predictive models for various diseases, including heart disease. ML algorithms can analyze complex datasets to identify patterns not readily apparent through conventional statistical methods, thereby improving the accuracy of disease prediction and patient outcomes.

The application of ML in predicting heart disease outcomes involves utilizing algorithms such as Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM) to analyze patient data and predict the likelihood of adverse cardiovascular events. These models can process vast amounts of information, including demographic data, medical history, and lifestyle factors, to provide a comprehensive risk assessment. The efficacy of these models is determined by their predictive accuracy, sensitivity, specificity, and overall reliability in diverse clinical settings.

Despite the potential benefits, the implementation of ML models in clinical practice presents several challenges. These include the need for large, high-quality datasets for training the algorithms, the complexity of integrating ML systems into existing healthcare infrastructures, and concerns regarding data privacy and security. Moreover, the interpretability of ML models is crucial for gaining the trust of healthcare professionals and ensuring that the predictions can be effectively translated into clinical decisions.

The aims and objective of this study are:

1. To evaluate the predictive performance of logistic regression and decision tree models in identifying heart disease outcomes based on clinical and demographic factors
2. To compare the accuracy and interpretability of logistic regression and decision tree models in predicting heart disease.
3. To seek insight into the most significant predictors of heart disease, such as age, chest pain type, maximum heart rate, ST depression, and ST slope, to better understand their contribution to heart disease risk.

By examining the performance metrics of the two described algorithms and identifying the factors influencing their effectiveness, this research seeks to provide insights into the practical applications of ML in cardiology and inform future developments in predictive healthcare technologies.

LITERATURE REVIEW

The integration of machine learning into cardiovascular disease prediction has been extensively explored over the past decade. Recent studies have demonstrated the potential of ML algorithms to enhance diagnostic accuracy and patient outcomes.

Ali et al. (2019) performed the study and developed a heart failure prediction system utilizing two support vector machine (SVM) models: one for feature selection and the other for prediction. In this approach, 70% of the data was allocated for training and 30% for testing. The feature selection model utilized L1-regularized linear SVM, while the prediction model implemented an L2-regularized SVM with a radial basis function (RBF) kernel. The hyper parameters of both SVM models were optimized to enhance performance [1].

Dwivedi (2018) performed a comparative study involving six machine learning methods—Artificial Neural Network (ANN), Support Vector Machine (SVM), Linear Regression (LR), K-Nearest Neighbor (KNN), Decision Tree (DT), and Naive Bayes (NB). The results revealed that LR outperformed the other techniques in predicting heart disease [2].

Khourdifi and Bahaj (2019) designed a system combining SVM, k-nearest neighbor (KNN), multilayer perceptron (MLP), random forest (RF), and Naive Bayes classifiers. The system's performance was enhanced using ant colony optimization and particle swarm optimization techniques, with KNN and RF achieving the highest accuracy. The results showed that the performance of the proposed system is superior to that of the classification technique presented in the study [3].

Latha and Jeeva (2019) performed the study proposing a diagnostic model combining results from Naive Bayes, multilayer perceptron, random forest, and Bayes network using majority voting. The results of the study indicated that ensemble techniques, such as bagging and boosting, are effective in improving the prediction accuracy of weak classifiers, and exhibit satisfactory performance in identifying risk of heart disease [4].

Mienye et al. (2020) performed the study and developed an artificial neural network (ANN) model optimized with a sparse autoencoder for heart disease diagnosis [5].

Mohan (2013) performed the study titled comparative analysis of classification function techniques for Heart Disease Prediction and found that SVM demonstrated strong performance in predicting cardiovascular disease when compared to DT and ANN [6].

Paragliola and Coronato (2020) introduced a model to assess cardiac event risks in hypertensive patients, employing a hybrid approach with long short-term memory (LSTM) networks and convolutional neural networks (CNNs). The model utilized ECG signals and time-series data for early predictions of hypertension-related complications [7].

Poornima and Gladis (2018) developed a hybrid heart disease prediction system, preprocessing data by removing missing values and reducing dimensionality with orthogonal local preserving projection (OLPP). The classification was performed using a neural network trained with Levenberg–Marquardt (LM) and group search optimization (GSO) for weight setting [8].

Pouriyeh et al. (2017) performed the study comparing DT, NB, ANN, KNN, and SVM found that SVM achieved superior accuracy of 84.15% for heart disease prediction [9].

Terrada et al. (2020) performed the study implementing a diagnostic system incorporating ANN, AdaBoost, and decision tree algorithms. Based on common performance indicators, this comparison shows that our proposed system has the highest accuracy of 94% in predicting and classifying atherosclerosis [10].

Thota et al. (2018) performed the study titled Heart Disease Prediction using random forest algorithm and obtained higher accuracy of 93.0% using RF to determine whether a patient suffers from heart failure [11].

Verma and Mathur (2020) performed the study and created a heart disease prediction system based on deep learning, selecting relevant features using correlation analysis and the cuckoo search algorithm [12].

Zheng et al. (2015) study assessed the performance of SVM, ANN, and the Hidden Markov Model (HMM) for diagnosing congestive heart failure, concluding that SVM outperformed the other two approaches. In addition to SVM, ensemble methods such as Random Forest (RF) have shown notable success [13].

In short, the literature reflects a growing body of evidence supporting the efficacy of machine learning models in predicting heart disease outcomes. While significant progress has been made, ongoing research is essential to address existing challenges and fully realize the potential of ML in improving cardiovascular health.

METHODOLOGY

The study utilizes the "Heart Failure Prediction" dataset downloaded from UCI Machine Learning Repository, comprising of 521 respondents on 11 clinical features relevant to heart disease prediction. These features include: 1. Age, 2. Sex, 3. Type of chest pain, 4. Resting Blood Pressure (Resting BP), 5. Serum cholesterol, 6. Blood sugar (BS), 7. Resting Electrocardiograph (Resting ECG), 8. Maximum Heart rate (Max HR), 9. Exercise Angina, 10. Old peak, 11. ST_Slope with Heart Disease as the dependent variable.

The methodology for this study follows a structured approach to compare logistic regression and decision tree models in predicting heart disease outcomes. The first step involves data preprocessing, where categorical variables such as Sex, Exercise Angina, and Resting ECG are converted into numerical values, while multi-category variables like Chest Pain Type and ST_Slope are encoded ordinal encoding. For model development, logistic regression is implemented as a probabilistic model to estimate the likelihood of heart disease presence based on predictor variables. The model is trained using the maximum likelihood estimation (MLE) method, and performance is assessed through accuracy, precision, recall, F1-score, and the AUC-ROC curve.

Concurrently, the decision tree model was developed using the CART (Classification and Regression Tree) algorithm in SPSS. The dependent variable was heart disease, and all predictors were included as independent variables. The CART algorithm was configured to use Gini impurity as the splitting criterion, and cross-validation was applied to prevent over fitting. The tree's configuration included a minimum of 10 cases per parent node and 5 cases per child node. The decision tree output included a tree diagram for visual interpretation, variable importance charts to identify influential predictors, and a classification table to evaluate accuracy, sensitivity, and specificity. Key predictors identified by the decision tree included ST_Slope, Exercise Angina, and Age, which provided clear decision rules for identifying heart disease cases. Model evaluation includes accuracy, confusion matrix analysis, and feature importance ranking to identify the most influential predictors in heart disease prediction.

Finally, a comparative analysis is conducted to assess the strengths and weaknesses of each model. Logistic regression is preferred for its statistical rigor and ability to quantify the relationship between predictors and the outcome, making it particularly useful for understanding risk factors. In contrast, the decision tree model provides a more interpretable, rule-based approach, making it an effective tool for clinical decision-making. While logistic regression offers better generalized ability, decision trees provide clearer decision paths but may be prone to over fitting. The study concludes by highlighting the complementary nature of both models, suggesting that integrating insights from logistic regression and decision trees could enhance the accuracy and interpretability of heart disease prediction models in clinical practice.

The ROC curve is a graphical representation of a classifier's performance across different threshold values. The x-axis represents 1 - Specificity (False Positive Rate), while the y-axis represents Sensitivity (True Positive Rate). The blue curve illustrates the classifier's ability to distinguish between classes, whereas the green diagonal line represents a random classifier with no discrimination ability. A strong classifier will have a curve that moves toward the top left corner, indicating high sensitivity while keeping the false positive rate low. The area under the curve (AUC) is a key metric for evaluating model performance, with values closer to 1 indicating a highly effective model, while an AUC of 0.5 suggests random guessing. Based on the shape of the ROC curve in the image, the classifier appears to perform well, demonstrating good predictive power. As the paper is derived from freely available secondary data, ethical clearance is not required.

ANALYSIS AND RESULTS

The descriptive statistics provide valuable insights into the characteristics of the dataset's continuous variables: Age, Cholesterol, HR (Maximum Heart Rate Achieved), Oldpeak (ST Depression Induced by Exercise), and Resting BP. The average age of participants is approximately 51.88 years, with a standard deviation of 9.38, indicating moderate variability. The distribution of age is nearly symmetric, with a slight negative skew (-0.161) and a flatter-than-normal shape, as reflected by the kurtosis value (-0.411). Cholesterol levels show high variability, with a mean of 165.42 mg/dL and a standard deviation of 127.16, likely capturing a wide range of participants, including those with normal and elevated cholesterol. The cholesterol distribution is also symmetric (Skewness = -0.082) and slightly flat (Kurtosis = -0.836) (Table 1).

Respondents achieved the maximum heart rate of 131.98 beats per minute, with a moderate spread around the mean (SD = 24.93) and a nearly symmetric distribution (Skewness = -0.076; Kurtosis = -0.271). In terms of old peak, the average value is 0.744, with a standard deviation of 0.9922, reflecting a narrow spread of ST depression values during exercise. However, the positive skew (0.816) indicates that a subset of participants exhibits elevated ST depression levels, which could be indicative of cardiac issues (Table 1). For Resting BP, the average value is 131.97 mmHg, with a standard deviation of 19.41, showing moderate variability across the dataset. The distribution of resting blood pressure is symmetric (Skewness = -0.040) but shows a high kurtosis (4.319), suggesting the presence of outliers, with some participants exhibiting exceptionally high or low values (Table 1).

Table 1: Descriptive Statistics

	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Age	51.88	9.377	87.930	-0.161	-0.411
Cholesterol	165.42	127.162	16170.217	-0.082	-0.836
Max HR	131.98	24.934	621.694	-0.076	-0.271
Old peak	0.744	0.9922	0.984	0.816	0.560
Resting BP	131.97	19.410	376.764	-0.040	4.319

The frequency distributions provide important insights into the characteristics of the dataset. The dataset is predominantly male, with 82.3% of participants being male and only 17.7% female, indicating a significant gender imbalance. Among chest pain types, Asymptomatic (ASY) chest pain is the most common, accounting for 55.3% of participants, followed by Atypical Angina (ATA) at 22.5% and Non-Anginal Pain (NAP) at 18.6%, while Typical Angina (TA) is the least frequent at 3.6%. This suggests that a substantial portion of participants might not exhibit typical angina symptoms, a characteristic often linked to heart disease. Additionally, most participants (74.7%) have normal fasting blood sugar levels, while 25.3% have elevated fasting blood sugar, which could be a contributing risk factor for heart disease (Table 2).

Regarding exercise-induced angina, 59.5% of participants report no angina during exercise, whereas 40.5% experience exercise angina, a potential indicator of cardiac stress. The ST slope distribution shows that a Flat ST slope (50.5%) is the most common, followed by an Upward slope (44.1%), while a Downward slope (5.4%) is rare. Flat or downward ST slopes are often associated with ischemia and heart disease, highlighting their importance as potential predictors. In terms of heart disease prevalence, 57.4% of participants have heart disease, while 42.6% do not, indicating a higher prevalence of heart disease cases in the dataset, which may influence the performance of predictive models (Table 2).

Table 2: Frequency Distribution for attributes

		Frequency(n)	Percentage (%)
Sex	Female	92	17.7
	Male	429	82.3
Chest Pain Type	ASY	288	55.3
	ATA	117	22.5
	NAP	97	18.6
	TA	19	3.6
Blood Sugar	Non Fasting	389	74.7
	Fasting	132	25.3
Exercise Angina	No	310	59.5
	Yes	211	40.5
ST_Slope	Down	28	5.4
	Flat	263	50.5
	Up	230	44.1
Heart Disease	No	222	42.6
	Yes	299	57.4

The cross tabulation of Sex and Heart Disease reveals gender-based trends. Among females, 3/4th i.e. 75% do not have heart disease, while only 1/4th i.e. 25% do have, whereas among males, 64.3% have heart disease and 35.7% do not. This indicates that males are excessively affected by heart disease compared to females in this dataset. These findings underscore the significance of variables such as Chest Pain Type, ST slope, and Exercise Angina as key factors influencing heart disease risk and suggest potential gender-based differences in heart disease prevalence (Table 3).

Table 3: Contingency table for Heart Disease and Sex

		Heart Disease		Total	
		No	Yes		
Sex	Female	Count	69	23	92
		%	75%	25%	

	Male	Count	153	276	429
		%	35.7%	64.3%	
		Total	222	299	521

The correlation matrix provides insights into the relationships between key variables in the dataset, including Age, Resting BP, Cholesterol, Max HR, Old peak, and Heart Disease. Age shows a significant positive correlation with Resting BP ($r = 0.230, p < 0.001$) and Old peak ($r = 0.255, p < 0.001$), indicating that older individuals tend to have higher resting blood pressure and greater ST depression during exercise. However, Age has a significant negative correlation with Max HR ($r = -0.456, p < 0.001$), suggesting that as age increases, the maximum heart rate achieved during exercise tends to decrease. Additionally, Age is positively correlated with heart disease ($r = 0.326, p < 0.001$), indicating that older individuals are more likely to develop heart disease (Table 4).

Resting BP has a weak but highly significant correlation with old peak ($r = 0.136, p = 0.002$) and a weak but significant positive correlation with Cholesterol ($r = 0.102, p = 0.020$), suggesting that higher resting blood pressure may be associated with higher cholesterol levels and ST depression. However, Resting BP does not show a significant correlation with heart disease ($r = 0.066, p = 0.131$), indicating that it may not be a strong direct predictor of heart disease in this dataset (Table 4). Cholesterol is weakly correlated but highly significant with Max HR ($r = 0.231, p < 0.001$), implying that individuals with higher cholesterol levels may achieve slightly higher maximum heart rates. However, Cholesterol shows no significant relationship with Old peak ($r = -0.018, p = 0.682$) and does not correlate strongly with heart disease ($r = -0.34, p < 0.001$) (Table 4). Max HR is negatively correlated with heart disease ($r = -0.365, p < 0.001$), suggesting that lower maximum heart rates are strongly associated with the presence of heart disease. Similarly, Old peak is positively correlated with heart disease ($r = 0.412, p < 0.001$), indicating that higher ST depression during exercise is linked to a higher likelihood of heart disease (Table 4).

Table 4: Correlation matrix

		Age	Resting BP	Cholesterol	Max HR	Old peak	Heart Disease
Age	R	1					
	p-value						
Resting BP	R	0.230**	1				
	p-value	<0.001					
Cholesterol	R	-0.287	0.102*	1			
	p-value	<0.001	0.020				
Max HR	R	-0.456	-0.149	0.231**	1		
	p-value	<0.001	0.001	<0.001			
Old peak	R	0.255**	0.136**	-0.018	-0.130**	1	
	p-value	<0.001	0.002	0.682	0.003		
Heart Disease	R	0.326**	0.066	-0.34	-0.365	0.412**	1
	p-value	<0.001	0.131	<0.001	<0.001	<0.001	

** . Correlation is significant at the 0.01 level (2-tailed); * . Correlation is significant at the 0.05 level (2-tailed)

Table 5: Classification table for Logistic Regression

Observed	Predicted: No Heart Disease	Predicted: Yes Heart Disease	Percentage Correct
No (0)	173	49	77.9%
Yes (1)	48	251	83.9%
Overall Accuracy			81.4%

*The cut value is .50

The classification table evaluates the predictive performance of the logistic regression model in distinguishing between individuals with and without heart disease. The model correctly predicted 77.9% of the cases for participants who do not have heart disease (True Negatives = 173 out of 222 actual "No" cases). However, it incorrectly classified 49 cases as having no heart disease when they actually do (False Negatives). For individuals with heart disease, the model performed slightly better, achieving a 83.9% correct prediction rate (True Positives = 251 out of 299 actual "Yes" cases). However, 48 cases were incorrectly predicted as having heart disease when they did not (False Positives). The model's overall accuracy was 81.4%, indicating that it correctly classified the majority of participants in the dataset. While the model demonstrated strong predictive performance, the slightly lower sensitivity for detecting cases without heart disease (77.9%) suggests room for improvement in reducing false negatives. This tradeoff between sensitivity and specificity may require further tuning of the model or threshold adjustments to optimize its clinical utility.

The default cut-off value for the logistic regression model is set at 0.50, meaning that the model classifies a case as "Yes" (Heart Disease) if the predicted probability is greater than or equal to 50%. Conversely, if the predicted probability is below 50%, the case is classified as "No" (No Heart Disease). This threshold represents a balance between sensitivity and

specificity. The logistic regression model summary provides insights into the fit and explanatory power of the model used to predict heart disease. The Cox & Snell R Square (0.392) and Nagelkerke R Square (0.527) measure the proportion of variation in the dependent variable explained by the predictors. While Cox & Snell R² indicates that approximately 39.2% of the variation is explained by the model, the Nagelkerke R²—a more robust measure—shows that the model explains about 52.7% of the variability. This suggests that the model has moderate to strong predictive power for heart disease.

The Hosmer-Lemeshow test evaluates the goodness-of-fit for a logistic regression model, determining how well the predicted probabilities align with the observed outcomes. In this case, the Chi-square value is 1.376 with 6 degrees of freedom, and the associated p-value is 0.967. A p-value greater than 0.05 indicates that the model's predictions are not significantly different from the observed data, meaning we fail to reject the null hypothesis that the model fits the data well. This result suggests that the logistic regression model provides an excellent fit for the dataset, as the predicted probabilities align closely with the actual outcomes of heart disease. The low Chi-square value further supports the model's suitability. Therefore, the Hosmer-Lemeshow test indicates that the logistic regression model is reliable for predicting heart disease in this context.

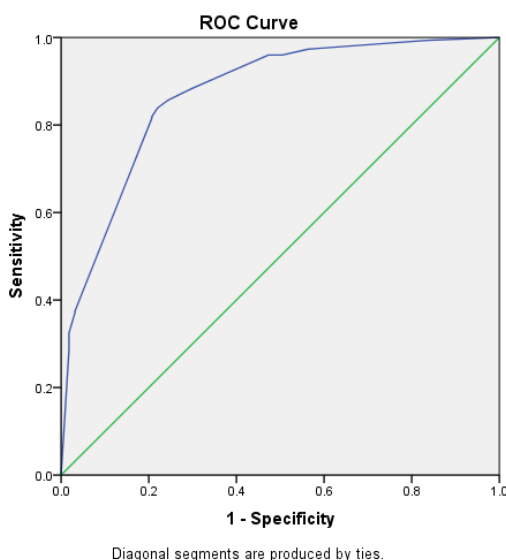


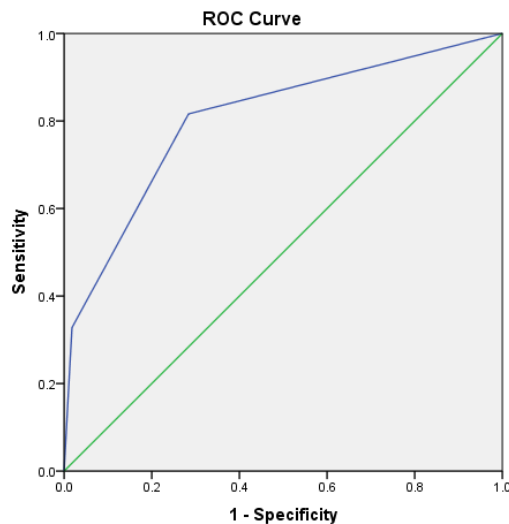
Figure 1: ROC curve for Logistic regression

The ROC (Receiver Operating Characteristic) Curve displayed in the figure evaluates the performance of the decision tree model for predicting heart disease. The Area under the Curve (AUC) value of 0.870 indicates that the classifier has good discriminatory ability in distinguishing between positive and negative classes. An AUC of 1.0 represents a perfect classifier, while an AUC of 0.5 suggests no better performance than random guessing. Since the AUC in this case is 0.870, the model demonstrates strong predictive performance, meaning it is effective at differentiating between positive and negative cases (Figure 1).

Table 6: Classification table for Decision Tree

Observed	Predicted		Percentage Correct
	No	Yes	
No	159	63	71.6%
Yes	55	244	81.6%
Overall Percentage	41.1%	58.9%	77.4%

The classification table results indicate that the model has an overall accuracy of 77.4%, meaning it correctly classifies 77.4% of all cases. Looking at the cases, the model correctly identifies 81.6% of patients with heart disease (True Positives) while misclassifying 55 cases as false negatives, meaning these individuals had heart disease but were predicted as not having it. On the other hand, the model successfully classifies 71.6% of individuals without heart disease (True Negatives) but mistakenly predicts 63 cases as false positives, where healthy individuals were incorrectly labeled as having heart disease. From a medical perspective, the high sensitivity (81.6%) is particularly important because it means the model is effective in identifying most patients who actually have heart disease. This is crucial since missing a heart disease case (false negative) could have serious health consequences. However, the lower specificity (71.6%) indicates that some individuals without heart disease are being misclassified as positive, which could lead to unnecessary medical tests or anxiety. While this trade-off is common in medical models, adjusting the decision threshold or incorporating additional risk factors could help improve specificity without significantly sacrificing sensitivity. The overall accuracy of 77.4% indicates that the model has decent predictive power, but improvements could be made, such as fine-tuning the decision tree parameters or incorporating additional features (Table 6).



Diagonal segments are produced by ties.

Figure 2: ROC curve for Decision Tree

The ROC Curve displayed in the figure evaluates the performance of the decision tree model for predicting heart disease. The Area under the Curve (AUC) for the Decision Tree model predicting heart disease is 0.805. This means that the model has a strong ability to distinguish between patients with and without heart disease. An AUC value of 0.805 indicates that, on average, the model will correctly rank a randomly chosen positive case (heart disease present) higher than a randomly chosen negative case (heart disease absent) 80.5% of the time. Since an AUC of 0.5 represents a model with no discrimination (random guessing), and an AUC of 1.0 represents a perfect classifier, a value of 0.805 suggests good predictive performance (Figure 2).

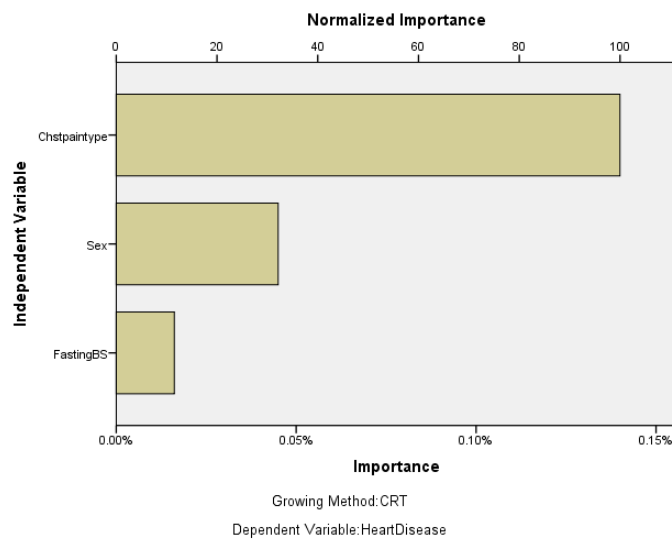


Figure 3: Variable Importance diagram

The decision tree model for predicting heart disease relies most heavily on chest pain type, which has the highest importance, normalized to 100%. This indicates that the presence and nature of chest pain play the most significant role in determining whether an individual is classified as having heart disease. The second most influential variable is sex, though its importance is significantly lower than that of chest pain type. This suggests that gender differences contribute to heart disease risk but are not as strong a predictor as chest pain characteristics. The third variable, Blood Sugar, has the lowest importance among the three but still plays a role in classification. While it is a contributing factor, it is less influential than chest pain type and sex. Overall, the model suggests that chest pain is the strongest predictor of heart disease, followed by sex and fasting blood sugar levels. These findings highlight the critical role of symptom-based factors in heart disease classification and suggest potential areas for model refinement by incorporating additional predictive variables.

CONCLUSION

Heart disease continues to be one of the leading causes of morbidity and mortality worldwide, necessitating the development of effective predictive models to aid early detection and intervention. Two machine learning approaches, Logistic Regression and Decision Tree, were employed to evaluate their effectiveness in predicting heart disease outcomes. The logistic regression model achieved an overall accuracy of 81.4%, correctly classifying most cases. It demonstrates

strong sensitivity in detecting heart disease (83.9%), though the slightly lower specificity (77.9%) suggests a need for further model refinement to reduce false negatives. Additionally, the ROC curve shows an AUC of 0.870, indicating strong discriminative ability. The decision tree model achieved an accuracy rate of 77.4%, effectively identifies patients with heart disease (81.6% sensitivity). However, its lower specificity (71.6%) results in more false positives, which could lead to unnecessary medical interventions. The model's AUC of 0.805 suggests a good but slightly lower predictive performance compared to logistic regression.

Overall, both models demonstrated strong predictive capabilities, with unique strengths. Logistic regression offered statistical interpretability, making it suitable for understanding the impact of individual risk factors on heart disease. On the other hand, the decision tree model's transparency and visual decision rules make it highly applicable in clinical decision support systems. Combining insights from both models could enhance the accuracy and practicality of heart disease prediction and diagnosis, contributing to better patient outcomes and more informed clinical decision-making.

Funding: Not applicable.

Conflict of interest: The authors declare no conflict of interests.

Data source: The data used in the study (Heart Failure Prediction) have been downloaded from UCI Machine Learning Repository.

REFERENCES

1. Ali, L., Niamat, A., Khan, J.A., Golilarz, N.A., Xingzhong, X., Noor, A., Nour, R., Bukhari, S.A.C. (2019). An optimized stacked support vector machines based expert system for the effective prediction of heart failure. *IEEE Access* 7,54007–54014. <https://doi.org/10.1109/ACCESS.2019.2909969>.
2. Dwivedi, A.K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing & Applications*, 29(10), 685–693.
3. Khourdifi, Y., Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by Particle Swarm Optimization and Ant Colony Optimization. *Int J Intell Eng Syst*, 12(1), 242–252. <https://doi.org/10.22266/ijies2019.0228.24>.
4. Latha, C.B.C., Jeeva, S.C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inf Med Unlocked* 16,100203. <https://doi.org/10.1016/j.imu.2019.100203>.
5. Mienye, I.D., Sun, Y., Wang, Z. (2020). Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. *Inf Med Unlocked* 18,1–5. <https://doi.org/10.1016/j.imu.2020.100307>.
6. Mohan, V. (2013). Comparative Analysis of Classification Function Techniques for Heart Disease Prediction. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(3), 735-741.
7. Paragliola, G., Coronato, A. (2020). An hybrid ECG-based deep network for the early identification of high-risk to major cardiovascular events for hypertension patients. *J Biomed Inform.* <https://doi.org/10.1016/j.jbi.2020.103648>.
8. Poornima, V., Gladis, D. (2018). A novel approach for diagnosing heart disease with hybrid classifier. *Biomed Res* 29, 2274–2280. <https://doi.org/10.4066/biomedicalresearch.38-18-434>.
9. Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H. & Gutierrez, J. (2017). A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease, 2017. *IEEE Symposium on Computers and Communications (ISCC)*, pp. 204-207.
10. Terrada, O., Hamida, S., Cherradi, B., Raihani, A., Bouattane, O. (2020). Supervised machine learning based medical diagnosis support system for prediction of patients with heart disease. *Adv Sci Technol Eng Syst J.*, 5(5), 269–277. <https://doi.org/10.25046/aj050533>
11. Thota, L., Nimmala, S., & Manasa, K. (2018). Heart Disease Prediction Using Random Forest Algorithm. *Global Journal of Engineering Science and Research*, 5(8), 248-252.
12. Verma, L., Mathur, M.K. (2020). Deep learning based model for decision support with case based reasoning. *Int J Innov Technol Explor Eng.*, 8(6C), 149–153
13. Zheng, Y., Guo, X., Qin, J. & Xiao, S. (2015). Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Computer Methods Programs Biomed*, 122, 372–83.