

Heart Disease Prediction with Machine Learning Models: Superior Performance of Support Vector Machines

Nariman A. Khalil ¹ , Mohamed Elkholy ² , and Mohamed Eassa ^{3,4,*} 

¹ Assistant professor in Egyptian Chinese College for Applied Technology, Suez University, Egypt;
Narimankhalil.eccat@suez.edu.eg.

² Faculty of Computer Science & Engineering, Alamein International University, Egypt; melkholy@aiu.edu.eg.

³ College of Informatics, Midocean University, 98123, Moroni, Comoros.

⁴ Computer Science Department, Faculty of Information System and Computer Science, October 6 University, Giza, 12585, Egypt; Mohamed.eassa.cs@o6u.edu.eg.

* Correspondence: Mohamed.eassa.cs@o6u.edu.eg.

Abstract: Heart disease results from long-term damage or inflammation affecting the heart and blood vessels, and it can impact individuals regardless of age, gender, or social background. Recent studies highlight approaches in identifying and predicting the risk of this condition. Such models estimate the probability of heart disease in percentage terms, enabling early detection and intervention. In this work, the dataset was processed using Python, and two primary algorithms Random Forest and Support Vector Machine (SVM) were applied for classification. While both methods produced promising outcomes, the SVM model achieved the highest performance, recording an accuracy of 95%. These results demonstrate the capability of machine learning techniques in enhancing the reliability of heart disease prediction.

Keywords: Cardiovascular Disease; Machine Learning Models; Support Vector Machine (SVM); Logistic Regression; Decision Trees; Random Forest; Data Preprocessing; Exploratory Data Analysis (EDA); Performance Metrics.

1. Introduction

Cardiovascular diseases (CVDs) remain the primary cause of global mortality, accounting for 17.9 million deaths annually as reported by the World Health Organization. These conditions encompass a wide spectrum of disorders, including coronary artery disease, myocardial infarction, arrhythmias, and heart failure. Although medical research and treatment methods have progressed significantly, the early detection and effective management of heart disease continue to be difficult due to the intricate interplay of genetic, environmental, and lifestyle influences.

In recent years, advances in data-driven technologies have introduced new opportunities to enhance healthcare outcomes [1]. Among these, machine learning (ML), a branch of artificial intelligence, has emerged as a powerful tool with the potential to revolutionize cardiovascular medicine [2]. By analyzing large and heterogeneous medical datasets, ML algorithms can detect hidden trends, recognize risk factors, and support both diagnostic and prognostic decision-making with elevated levels of accuracy [3].

2. Related Work

In the significance of the heart as a living human organ cannot be overstated. Its role in blood circulation, as essential as oxygen's role in man's survival, underscores the significance of preserving the heart. Scientists across various disciplines, such as artificial intelligence, machine learning, and data mining, have devoted their efforts to this life-preserving field. The field of prediction of heart disease is not new; others have ventured into the field and developed satisfactory results. We shall mention some of the papers we have read and studied to direct us in our research, and all papers mentioned shall be found in the references section (see Table 1).

Table 1. Comparison of various existing ML methods.

Papers/year	Methods/Classifiers	Datasets	Highest Accuracy	Best model
2025 [4]	KNN/Decision Tree/Random Forest	UCI Dataset	81	KNN
2022 [5]	SVM/KNN/Decision Tree/Logistic regression/Naïve Bayes/neural networks	UCI Dataset	85	SVM
2023 [6]	Gradient Boosting/ KNN/Decision Tree/Logistic regression/Naïve Bayes/neural networks	UCI Dataset	90.8	Gradient Boosting/ Logistic regression
2023 [7]	SVM/KNN/Decision Tree/Logistic regression/Random Forest/XG boost	Cleveland	87.91	SVM
2024 [8]	SVM, Logistic Regression, Random Forest	Pro-AI-Robikul dataset	72	SVM
2024 [9]	Random Forest	Framingham	86.52	Random Forest
2023 [10]	SVM/KNN/Decision Tree/Logistic regression/Naïve Bayes/Random Forest	UCI dataset	94.79	Random Forest

Heart disease includes conditions like coronary artery disease that affect blood flow and heart function. Common symptoms are chest pain, shortness of breath, fatigue, and dizziness. The study uses machine learning classifiers to predict heart disease and improve accuracy.

Heart disease refers to a group of disorders that impact the heart's normal function. While the term is often used to describe coronary artery disease (CAD) and the heart attacks associated with it, problems can also occur in other parts of the heart, such as the muscle, the valves, or its electrical conduction system. When the heart is not working efficiently, it fails to supply the body with adequate blood, oxygen, and essential nutrients. Since the heart acts as the engine that powers the body, any disruption in its ability to deliver this "fuel" affects the performance of all bodily systems [11].

Heart disease may present itself in a variety of ways, and symptoms do differ from one individual to another. Symptoms that many individuals have in common are: Chest pressure or pain (angina), Shortness of breath, Pain or discomfort in the neck, jaw, throat, upper abdomen, or back, Fatigue, Irregular heartbeat, Swelling in the legs, ankles, or feet (edema), Dizziness or lightheadedness.

Prompt identification of these symptoms is crucial as early identification, and treatment can significantly improve the patient's outcome. Machine learning algorithms trained to work with indicators of patient health attempt to recognize individuals with patterns in them suggestive of these symptoms [11].

3. Methodology

This section discusses the machine learning classifiers applied for heart disease prediction, along with the proposed strategy to enhance accuracy. Four different classification techniques are employed and compared in this study. Their performance is evaluated by using accuracy as the key metric, after which the most effective model is selected for making predictions.

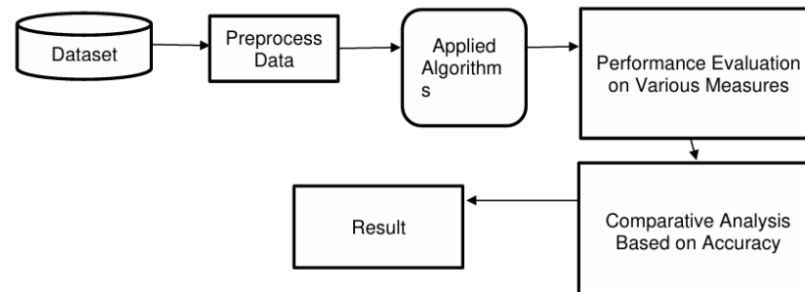


Figure 1. Methodology of proposed work.

4.1 Dataset

The Cleveland Heart Disease dataset is widely utilized in supervised machine learning studies for predicting cardiovascular conditions. This dataset, available through the Kaggle repository, consists of 1,025 patient records with fourteen features, including the target variable. The attributes cover demographic, clinical, and diagnostic information such as age, sex, blood pressure, cholesterol, blood sugar, and other medical indicators. Originally, the dataset defined five outcome classes ranging from 0 (no disease) to 4 (severe presence). However, due to the uneven distribution of cases across these categories (with class counts of 164, 55, 36, 35, and 13), most research simplifies the problem into a binary classification task: zero representing the absence of disease and one indicating its presence. In this work, the target variable is used to classify whether a patient is diagnosed with heart disease provides a detailed overview of the dataset's attributes.

4.1.1 Dataset Overview

The dataset consists of 1,025 fully complete entries without any missing information. It covers a variety of demographic, clinical, and diagnostic attributes, including age, gender, chest pain category, blood pressure, cholesterol level, fasting blood sugar, electrocardiogram outcomes, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the ST segment, number of major vessels, and thallium stress test results.

4.1.2 Statistics for numerical variables

Table 2. Descriptive statistics.

Variable	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Age	1025.0	54.43	9.07	29.0	48.0	56.0	61.0	77.0
Trestbps	1025.0	131.61	17.52	94.0	120.0	130.0	140.0	200.0
Chol	1025.0	246.0	51.59	126.0	211.0	240.0	275.0	564.0
Thalach	1025.0	149.11	23.01	71.0	132.0	152.0	166.0	202.0
Oldpeak	1025.0	1.07	1.18	0.0	0.0	0.8	1.8	6.2

4.1.3 Statistics for categorical variables

Table 3 summarizes the distribution of categorical variables in the dataset. For the sex attribute, males (coded as 0) are the most common group, appearing in most cases. Chest pain (cp) is

represented by four categories, with type 0 being the most frequent. In fasting blood sugar (fbs), most patients fall below the 120 mg/dl threshold, making category zero the dominant value. Resting electrocardiographic results (restecg) show three outcomes, with type 1 occurring most often. Regarding exercise-induced angina (exang), most patients did not experience angina during exercise, corresponding to category zero. The slope of the peak exercise ST segment (slope) is divided into three classes, where type 2 is the most prevalent. For the number of major vessels observed by fluoroscopy (ca), five values are possible, with zero being the most common. The thallium stress test (thal) demonstrates four outcomes, with the reversible defect (type 2) recorded most frequently. Finally, the target variable indicates whether heart disease is present or not, and most cases are classified as having the disease (value 1).

Table 3. Descriptive stats for object columns.

Variable	Count	Unique	Top	Freq
sex	1025	2	1	713
cp	1025	4	0	497
fbs	1025	2	0	872
restecg	1025	3	1	513
exang	1025	2	0	680
slope	1025	3	2	482
ca	1025	5	0	578
thal	1025	4	2	544
target	1025	2	1	526

4.2 Data Exploration and Analysis (EDA)

The EDA process involved the following steps to get accustomed to the data set and identify patterns that would be helpful for heart disease prediction:

4.2.1 Target Variable Distribution

A count plot was used to represent the target variable distribution with a balanced data set between being influenced by heart disease (1) and not being influenced (0).

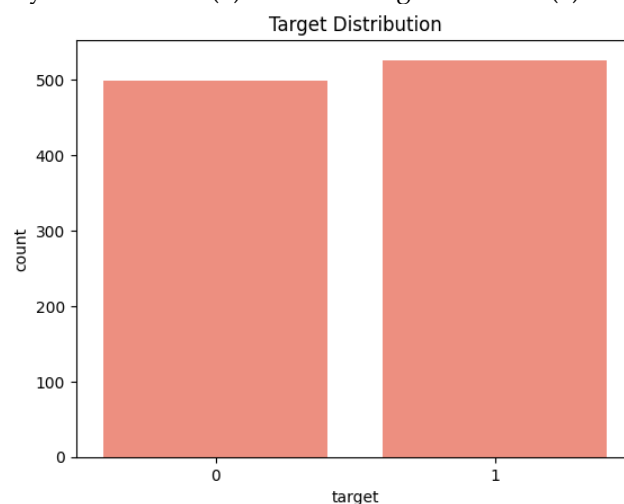


Figure 2. Target distribution.

4.2.2 U-univariate Analysis:

- **Numerical Features:** Histograms and box plots were used to track feature distributions like age, cholesterol, resting blood pressure, and maximal heart rate. Figure 3 illustrates the distribution of continuous numerical features in the dataset. The age variable is evenly spread, with a noticeable concentration among patients in their late fifties, and an overall mean of 54.43 years (SD = 9.07). Resting blood pressure (trestbps) is clustered between 120- and 140-mm Hg, averaging 131.61 mm Hg with a standard deviation of 17.52. Serum cholesterol (chol) levels are mostly within the 200–300 mg/dl range, with a mean of 246 mg/dl (SD = 51.59). For maximum heart rate achieved (thalach), most individuals fall between 140 and 170 bpm, with the average at 149.11 bpm (SD = 23.01). Finally, exercise-induced ST depression (oldpeak) is concentrated near zero, indicating that many participants did not experience significant depression during exercise, with a mean of 1.07 and a standard deviation of 1.18.

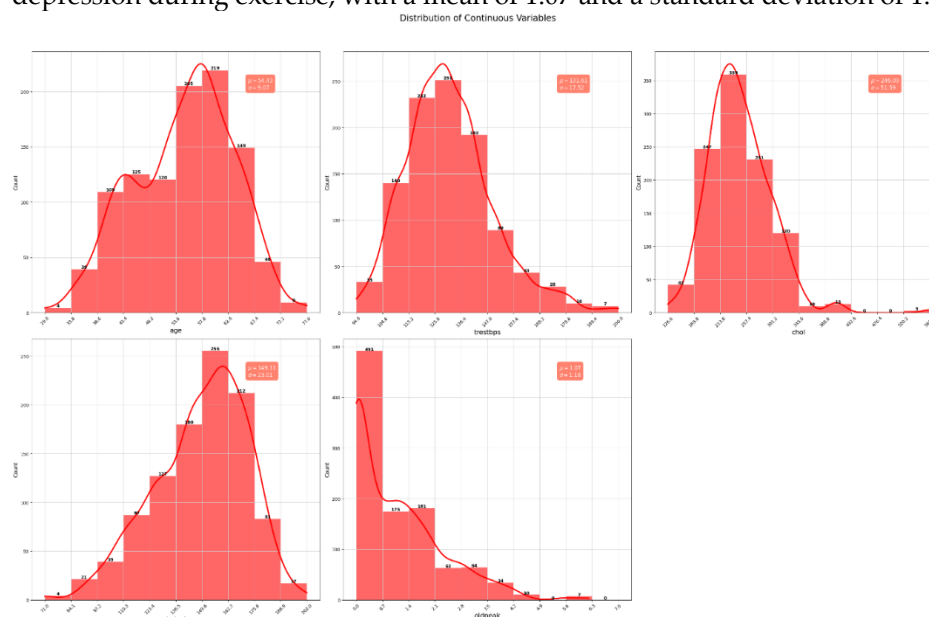


Figure 3. Distribution of continuous features.

- **Categorical Features:** Count Plots were used for features like chest pain type, fasting blood sugar, and electrocardiographic findings to determine class frequencies. Figure 4 presents the distribution of categorical variables within the dataset. The gender variable (sex) indicates that females constitute most of the sample. For chest pain type (cp), type 0, which represents typical angina, is the most frequently observed, while the other types occur at lower but varying rates. The fasting blood sugar attribute (fbs) shows that most patients have levels below 120 mg/dl, suggesting that elevated blood sugar is uncommon in this population. Resting electrocardiographic results (restecg) display diverse outcomes, with certain categories appearing more often than others. In terms of exercise-induced angina (exang), most patients do not report this condition, implying that it is not a dominant symptom in the dataset. The slope of the peak exercise ST segment (slope) also varies, with one type being more prevalent than the rest. For the number of major vessels colored by fluoroscopy (ca), the value of zero is the most common, meaning many patients had no vessels highlighted. The thallium stress test results (thal) reveal a range of categories, with one defect type appearing most frequently. Finally, the target variable shows that the dataset is balanced, with approximately 54.5% of patients having heart disease and 45.5% without it.

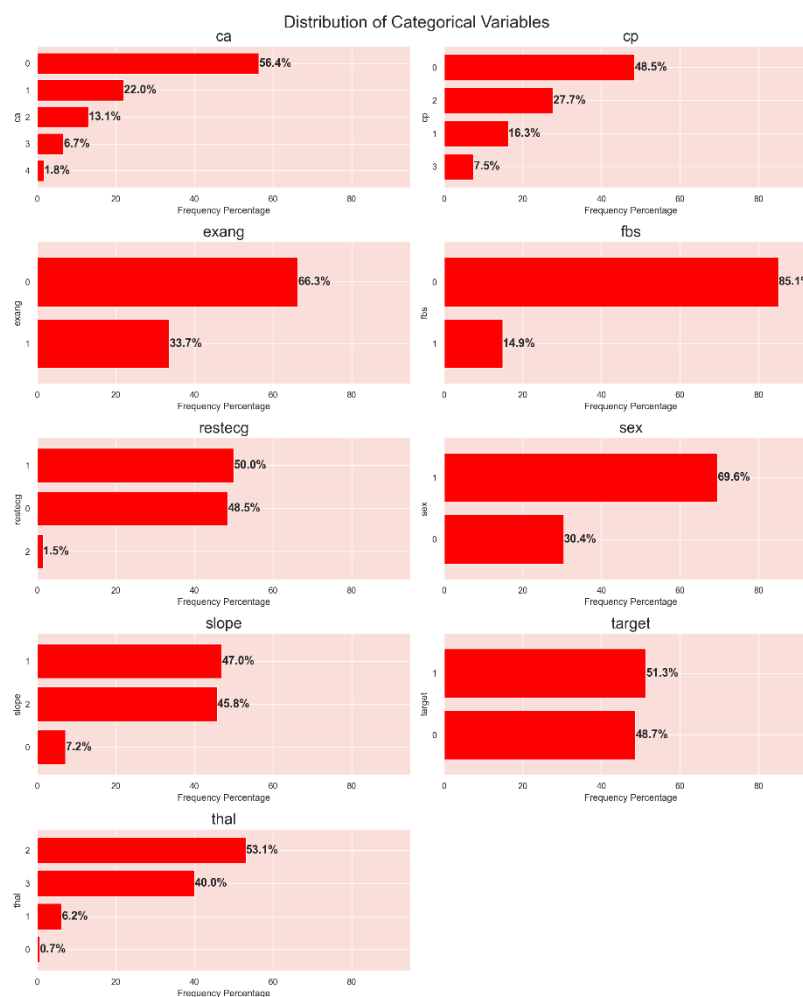


Figure 4. Distribution of categorical features.

4.2.3 Bi-variate Analysis

Association of each feature with the target variable was illustrated using boxplots and violin plots, i.e., for:

Continuous features vs. heart disease

Figure 5 highlights the distribution of continuous features in relation to heart disease. The analysis shows that patients diagnosed with heart disease tend to be slightly younger on average compared to those without, as the non-disease group exhibits a higher mean age. For resting blood pressure (trestbps), both categories display identical distributions, indicating that this attribute has limited discriminative power. In terms of serum cholesterol (chol), the values for both groups appear quite close, though the mean cholesterol level for patients with heart disease is marginally lower. A more noticeable distinction is observed in maximum heart rate achieved (thalach), where patients with heart disease reach higher rates during stress tests than those without the condition. Finally, exercise-induced ST depression (oldpeak) is found to be lower among patients with heart disease, with values clustering near zero, whereas the non-disease group shows a broader range of distribution.

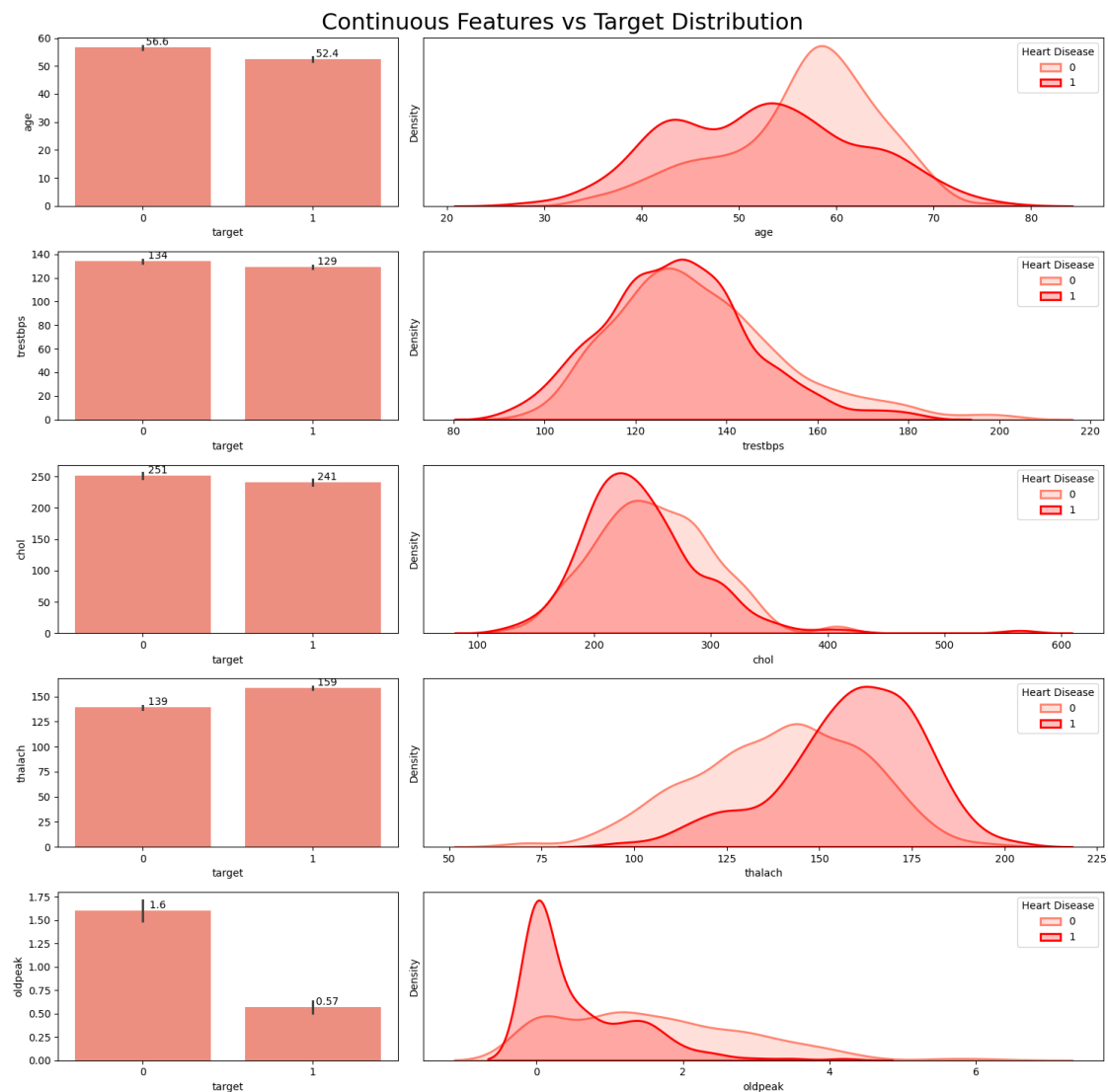


Figure 5. Continuous features vs. target distribution.

Categorical features vs. heart disease

Key predictors are ca, cp, exacting, sex, slope, thal. Fewer vessels, chest pain types (1–3), no exercise angina, male-gender, slope type 2, and reversible defect show higher disease presence, while) fb(s has weak impact and restate moderate effect.

Figure 6 illustrates the relationship between categorical features and the presence of heart disease. The variable Number of Major Vessels (ca) shows that patients with fewer vessels highlighted through fluoroscopy are more likely to have heart disease, while the probability decreases as the number of colored vessels increases. Similarly, Chest Pain Type (cp) demonstrates that individuals with pain types 1, 2, or 3 have a greater likelihood of heart disease compared to those with type 0, making chest pain a useful predictive feature. The feature Exercise-Induced Angina (exang) indicates that patients without angina during exercise exhibit a higher rate of disease than those who experience it, emphasizing its importance as a predictor. On the other hand, Fasting Blood Sugar (fbs) has little influence, as the distribution of cases is the same for values above and below the 120 mg/dl threshold. In contrast, Resting Electrocardiographic Results (restecg) suggest that category one is more strongly associated with the presence of heart disease, which may contribute to prediction. The Sex (sex) attribute highlights that males have a higher rate of heart disease than females, pointing to gender differences as a relevant factor. Additionally, it shows that type 2 has a

stronger association with heart disease, confirming its predictive value. Finally, the Thallium Stress Test (thal) reveals that patients with a reversible defect (category 2) are more likely to suffer from heart disease, marking this variable as one of the most significant indicators.

Categorical Features vs Target Stacked Barplots

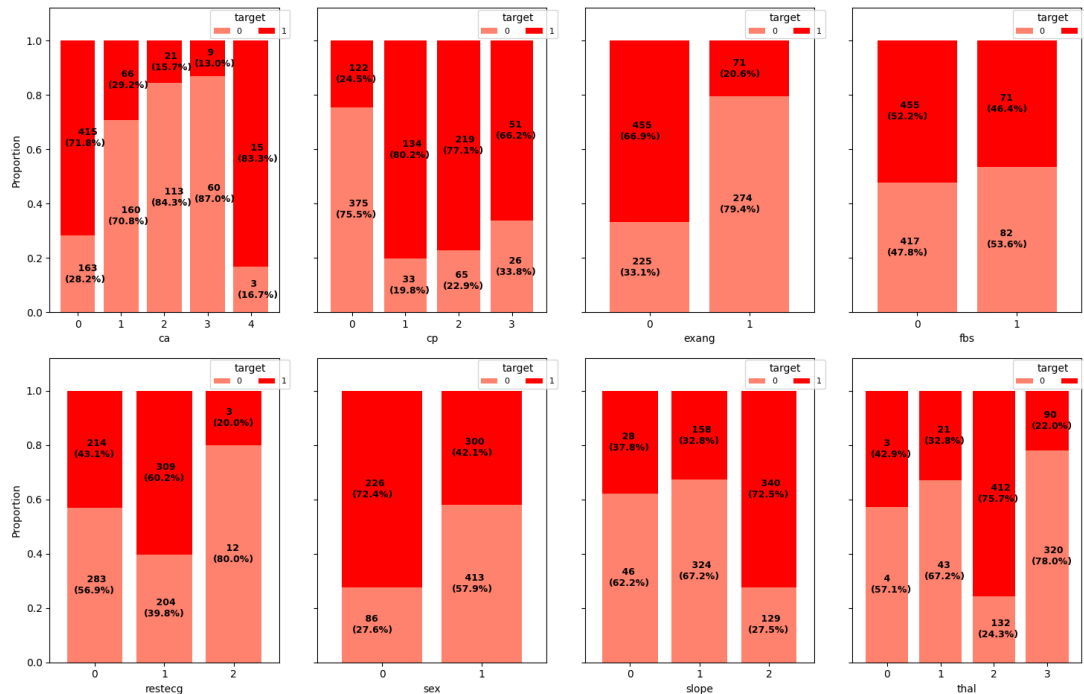


Figure 6. Categorical features vs. target.

Correlation Map

As shown in Figure 7 cp (Chest Pain Type): +0.43 → Positive. Higher chest pain score is linked with higher probability of heart disease. thalach (Max Heart Rate): +0.42 → Increased. Increased heart rate is linked with occurrence of heart disease. exang (Exercise-induced angina): -0.44 → Negative. Exercise-induced angina is likely to indicate lower chance of having a healthy heart. oldpeak (ST depression): -0.45 → Negative correlation. ST depression at exercise more → likely to have heart issues. ca (Number of large vessels filled in by fluoroscopic): -0.38 → More vessels uncovered = poorer condition = less likely to have healthy heart. thal: -0.34 → Certain thalassotherapy levels more connected with disease.

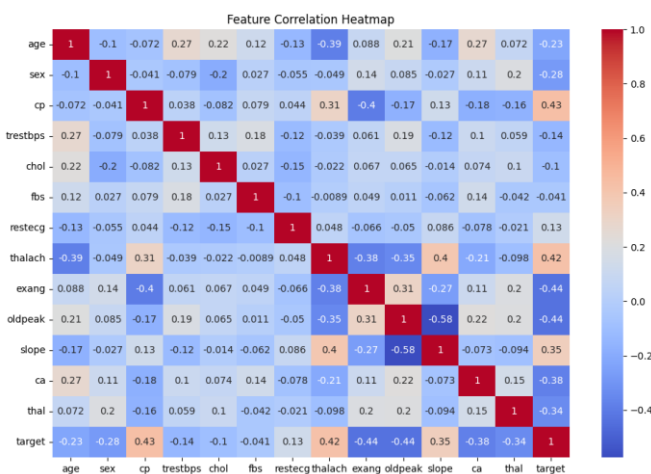


Figure 7. Correlation heatmap.

4.3 Data Preprocessing

Below are the steps used for preprocessing to clean data for modeling:

4.3.1 Handling Missing Values

We identified any missing values in the data. We then either replaced them or removed them to complete the dataset. Although the data was mostly clean, we did some preprocessing to ensure consistency throughout.

4.3.2 Treatment of Outliers

Outliers identified during EDA (like extremely high cholesterol levels) were capped or dropped to prevent the learning process of the model.

- **Outliers check**

Identifying outliers is a key step in ensuring reliable machine learning results. Properly handling these abnormal values prevents them from distorting predictions, enhances the model's accuracy, and makes its behavior more understandable.

- **Treating outliers**

This involves identifying and removing outliers from the data set before training. Outlier detection is a crucial step in data preprocessing, as unusual values can negatively influence the performance and accuracy of machine learning models. If not treated properly, these points may bias the results and reduce the interpretability of the model. To address this issue, several strategies are commonly applied. One approach is thresholding, where data instances that exceed a certain statistical boundary, such as a Z-score greater than three, are considered outliers. Another method relies on distance measures, where observations that are located far from their nearest neighbors are identified as abnormal. A third technique is clustering, in which outliers are recognized as data points that do not fit into any cluster or appear only in exceedingly small groups. By applying such methods during preprocessing, the dataset becomes more consistent, which in turn improves the reliability and predictive power of the trained model.

4.3.3 Encoding Categorical Variables

Categorical attributes such as chest pain, thalassemia, and rest Electrocardiogram (ECG) result were converted into numerical form using one-hot encoding to be prepared for use by machine learning algorithms.

4.3.4 Feature Scaling

The numerical features were adjusted using normalization methods like Min-Max scaling and standardization (z-score). These techniques make sure that all attributes fall within a similar range. This is especially important for distance-based algorithms such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM).

4.3.5 Feature Selection (Optional)

Features of low correlation or redundant features were monitored and removed to increase the efficiency of models and generalization.

5. Results and Discussion

This section presents the results of four classifiers: Support Vector Machine (SVM), Logistic Regression, Decision Tree, and Random Forest. Their performance was evaluated using standard metrics: Accuracy (overall correctness), Precision (ratio of correctly predicted positives), Recall (ability to detect actual positives), and F1-score (harmonic mean of precision and recall).

5.1 Model evaluation Metrics

Evaluation metrics are used to measure how well machine learning models perform and to compare different algorithms. In this study, the classifiers tested include Support Vector Machine (SVM), Logistic Regression, Decision Tree, and Random Forest. Their performance was assessed using four common metrics: Accuracy, which shows the overall correctness of predictions; Precision, which indicates how many predicted positives were correct; Recall (Sensitivity), which measures how well the model identifies actual positives; and F1-score, which combines precision and recall into a single balanced value.

5.1.1 A Confusion/classification Matrix

A widely used method for assessing the effectiveness of a classification model is by comparing its predicted outputs with the actual labels. From this comparison, several evaluation measures can be derived, including accuracy, precision, recall, and the F1-score. Among these, precision and recall are particularly important and are frequently applied when analyzing the performance of binary classification models.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Figure 8. Confusion matrix.

Table 4. Comparison between models' accuracy.

Decision Tree	91.1%
SVM	95.09%
Random forest	94.82%
Logistic regression	84.9%

Table 5. Comparison between evaluation metrics for class 0.

Model	Precision	Recall	F1-score
Decision tree	0.89	0.94	0.92
Random forest	0.96	0.94	0.95
Logistic Regression	0.87	0.82	0.84
SVM	0.97	0.93	0.95

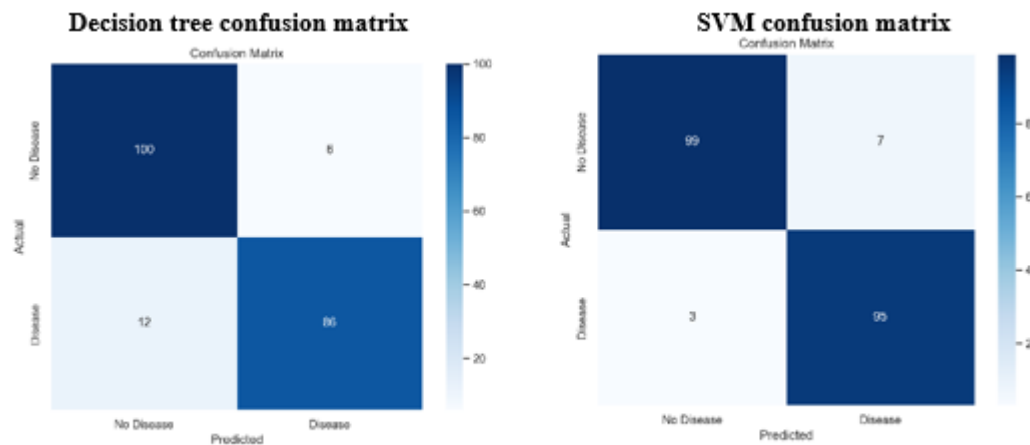


Figure 9-A. Decision tree Confusion matrix

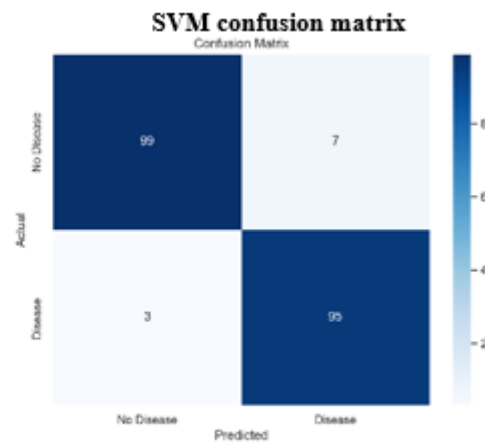


Figure 9-B. SVM Confusion matrix

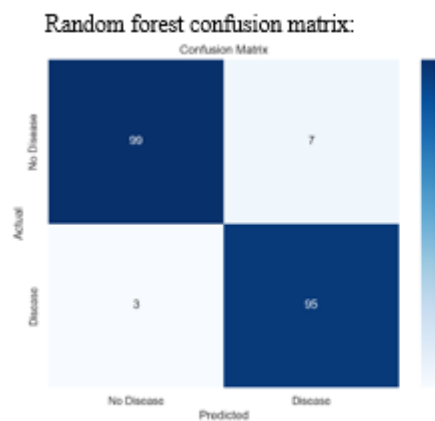


Figure 9-C. Random Forest Confusion Matrix

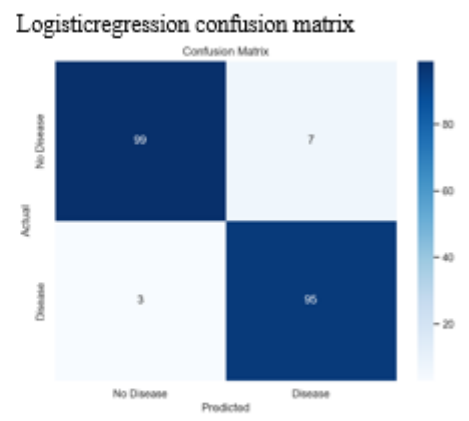


Figure 9-D. LogisticRegression Confusion

Table 6. Comparison between proposed work and past work.

Papers/year	Methods/Classifiers	datasets	Highest Accuracy	Best model
Proposed work	SVM/ Decision Tree/Logistic regression/Random Forest/	Cleveland	95.09%	SVM
[6]	KNN/Decision Tree/Random Forest	UCI Dataset	81%	KNN
[4]	SVM/KNN/Decision Tree/Logistic regression/Naïve Bayes/neural networks	UCI Dataset	85%	SVM
[9]	Gradient Boosting/ KNN/Decision Tree/Logistic regression/Naïve Bayes/neural networks	UCI Dataset	90.8%	Gradient Boosting/ Logistic regression
[10]	SVM/KNN/Decision Tree/Logistic regression/Random Forest/XG boost	Cleveland	87.91%	SVM
[11]	SVM, Logistic Regression, Random Forest	UCI Kaggle	72%	SVM
[12]	Random Forest	Framingham	86.52%	Random Forest
[13]	SVM/KNN/Decision Tree/Logistic regression/Naïve Bayes/Random Forest	UCI dataset	94.79%	Random Forest

6. Optimizations

In machine learning, different models and datasets require specific hyperparameters, which act as adjustable variables that influence training. Since the best values are not known in advance, they are determined through repeated experimentation, a process called hyperparameter tuning. This involves testing the model with various parameter combinations, either manually or through automated optimization methods.

Hyperparameters define aspects of the model's structure and behavior, directly affecting its accuracy and efficiency. Finely tuning those enables researchers to enhance performance and achieve more reliable results, making this step a critical component of the machine learning workflow.

7. Conclusions

We identify the application of which machine learning algorithms are effective in predicting heart disease from duly formulated health data. With a sufficient variety of supervised machine learning methods implemented on Cleveland Heart Disease, we verified the predictability of the model by using Decision Tree, Random Forest, Logistic Regression and Support Vector Machine (SVM). Out of all of them, Random Forest and Support Vector Machine (SVM) consistently outperform both classes, on all the metrics of measurement precision, recall, and F1-score—and hence are the most trustworthy models to recognize individuals at risk. Use of exploratory data analysis and sufficient preprocessing of data, i.e., removal of outliers, one-hot encoding of categorical features, and attribute normalization, played a significant role in enhancing model precision and generalizability. Recall reduction also enabled the models to be highly effective at identifying true positive cases, a criterion which is vitally important in medical diagnosis since failure to identify a positive case would be catastrophic. In conclusion, this study illustrates the potential for machine learning to emerge as a valuable tool in early diagnosis of heart disease. With ongoing development, clinically significant application, and improved access to large and diverse data, the models will enable clinical experts to make better and earlier-informed decisions, therefore leading to better patient outcomes and less cardiovascular mortality.

Declarations

Ethics Approval and Consent to Participate

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration by another publisher. All the material is owned by the authors, and/no permissions are required.

Consent for Publication

This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of Data and Materials

The dataset is publicly available at Kaggle (<https://www.kaggle.com/paultimothymooney/breast-histopathology-images>).

Competing Interests

The authors declare no competing interests in the research.

Funding

This research was not supported by any funding agency or institute.

References

- [1] Mohamed Elkholy and Marzouk MA, "Deep learning-based classification of eye diseases using Convolutional Neural Network for OCT images," *Frontiers in Computer Science* 5:1252295. 2024, doi: 10.3389/fcomp.2023.1252295.

- [2] N. Khalil, M. Elkholy, and M. Eassa, "A Comparative Analysis of Machine Learning Models for Prediction of Chronic Kidney Disease", Sustainable Machine Intelligence Journal (SMIJ), vol. 5, no. 5, Oct. 2023.
- [3] Marzouk, Marwa A. and Elkholy, Mohamed. "Combining bag of visual words-based features with CNN in image classification" Journal of Intelligent Systems, vol. 33, no. 1, 2024, pp. 20230054. <https://doi.org/10.1515/jisys-2023-0054>
- [4] Muhammad, Bakhtawar & Umar, Hooria & Fatima Yousaf, Hoor & Nasir, Usama & Hussain, Muhammad Zunnurain & Hasan, Muhammad Zulkifl & Mustafa, Muzzamil & Yaqub, Muhammad. (2025). Heart Disease Prediction Using Machine Learning. 10.1109/IDICAIEI61867.2024.10842908.
- [5] Anderies, Anderies & Tchin, Jalaludin & Putro, Prambudi & Darmawan, Yudha & Gunawan, Alexander. (2022). Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms. Engineering, Mathematics, and Computer Science (EMACS) Journal. 4. 87-93. 10.21512/emacsjournal.v4i3.8683.
- [6] AbdElminaam, D. S., Mohamed, N., Wael, H., Khaled, A., & Moataz, A. (2023). MLHeartDisPrediction: Heart Disease Prediction using Machine Learning. Journal of Computing and Communication, 2(1), 50–65.
- [7] Ahamad, G.N.; Shafiullah; Fatima, H.; Imdadullah.; Zakariya, S.; Abbas, M.; Alqahtani, M.S.; Usman, M. Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease. Processes 2023, 11, 734.
- [8] Bhowmik, P. K., Miah, M. N. I., Uddin, M. K., Sizan, M. M. H., Pant, L. P., Islam, R., & Gurung, N. (2024). Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health. British Journal of Nursing Studies, 4(2), 35–50.
- [9] Mir, A., Ur Rehman, A., Ali, T. M., Javaid, S., Almufareh, M. F., Humayun, M., & Shaheen, M. (2024). A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques. ESC Heart Failure.
- [10] Torthi, R., Marapatla, A. D. K., Mande, S., Gadiraju, H. K. V., & Kanumuri, C. (2023). Heart disease prediction using random forest-based hybrid optimization algorithms.
- [11] Cleveland Clinic. Coronary Artery Disease. Cleveland Clinic – Health Library, 2023. [online]. Available.
- [12] Luísa Soares, Tatiana Leal, Ana Lúcia Faria, Ana Aguiar and Cátia Carvalho. Cardiovascular Disease: A Review. Biomed J Sci & Tech Res 51(3)- 2023. BJSTR. MS.ID.008101
- [13] S. Belgic and L. Dragun. Random forest in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 114:24–31, 2016.

Received: 29 Jan 2025, **Revised:** 30 Apr 2025,

Accepted: 15 Sep 2025, **Available online:** 17 Sep 2025.



© 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Science Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.