Sonargaon University Journal, Vol 5, Issue II, July-December 2025

# Predicting Heart Disease with Machine Learning: A Comparative Analysis of Models

**Jahangir Hussen*[1], Madhabi Rani Kar[2], Taspia Momotaz[3]**

[1*, 3] Department of Computer Science and Engineering, Sonargaon University, Dhaka, Bangladesh

[2] Department of Nursing, White Pearl Nursing College, Maulvi Bazar, Sylhet, Bangladesh

*E-mail: jahangirhussen06@gmail.com

*Received: July 02, 2025*
*Revised: October 15, 2025*
*Accepted: November 13, 2025*

## ABSTRACT

*Heart disease remains a leading cause of mortality worldwide, highlighting the urgent need for accurate and early prediction systems. This study aims to develop and compare the performance of various machine learning (ML) models for predicting the presence of heart disease based on clinical attributes. Using a dataset comprising patient information such as age, sex, blood pressure, and cholesterol levels, the study applies and evaluates four widely used ML algorithms: Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression. The dataset was pre-processed through techniques such as missing value imputation and feature scaling to ensure model reliability. Model performance was assessed using precision, recall, accuracy, and F1-score. Among the models tested, Decision Tree and Random Forest demonstrated superior performance, achieving accuracy rates of approximately 98.5%. The findings underscore the potential of tree-based models in developing robust heart disease prediction tools. This study contributes to the growing field of health informatics by offering a comparative approach to model evaluation and showcasing the importance of data pre-processing. The novelty lies in its systematic performance comparison under a unified framework and its emphasis on high-accuracy, interpretable ML models suitable for clinical application.*

**Keywords:** *machine learning, heart disease prediction, decision tree, random forest, SVM, logistic regression*

## 1.0 INTRODUCTION

Heart disease is a leading worldwide cause of death. Prevention and early detection are the keys to improving public health (Swetha et al., 2024). Early diagnosis makes interventions possible to reduce risk and improve survival. Machine learning came into prominence in the medical arena years ago (Katarya Srinivas, 2020). It offers advanced means of handling complex data and making predictions. Decision trees, random forests, SVM, and logistic regression are prediction models for heart disease. They conduct pattern analysis on health information. They operate with demographics, clinical data, and lab results. They may be used to aid physicians in decision-making. They can even transform treatment for heart disease (Amin et al., 2019**)**. It is an issue of utmost importance to healthcare: heart disease prediction. This is due to the fact that cardiovascular diseases are on the rise. Heart disease remains difficult to diagnose even with advances in medical technology. There are many symptoms and risk factors. The conventional methods like clinical examination and imaging are time-consuming and costly (Bhatt et al., 2023). They have a high likelihood of making mistakes. Thus, there is a requirement to develop an accurate automatic prediction system based on available patient data. Machine learning is able to identify patterns and make sensible predictions. There still are some difficulties. They are dealing with missing values, choosing the best model to use, and receiving results that hold true for populations (Ramesh et al., 2022).

The goal of this paper is to compare how well different machine learning models can make heart disease predictions. Decision Tree, Random Forest, SVM, and Logistic Regression are the models involved (Saboor et al., 2022). Accuracy, precision, recall, and F1-score will be the parameters that we are going to learn about. They will help us know the best way of heart disease prediction. Pre-processing techniques in some of the most critical topics are also included in the paper (Ahmed Husien, 2024). These include feature scaling, missing value handling, and categorical variable transformation. Lastly, the aim of this paper is to be a contribution to the expanding database of how machine learning has been applied in medicine, that is, cardiovascular disease prediction (Arumugam et al., 2023). This essay is organized in such a manner that firstly, there is the introduction. Secondly, there shall be a review of current literature on heart disease prediction employing machine learning (Srivastava kumar, 2022). The methodology shall cover the dataset, adopted preprocessing techniques, and machine learning techniques utilized. Thirdly, model evaluation results and discussion will consist of model evaluation results (Boukhatem et al., 2022). It will also explain what such results mean. Finally, the conclusion will outline key findings and suggest possible future research directions.

## 2.0 LITERATURE REVIEW

### 2.1 Current Work

Most current research employs machine learning to predict heart disease. The current study employed a range of algorithms such as decision trees, random forests, SVM, and logistic regression (Chandrasekhar Peddakrishna, 2023). They employed them in pattern recognition from health data. For instance, (Nagavelli et al., 2022) concluded that Decision Trees were accurate, particularly with imputation strategies for missing data. Similarly, Hassan et al. (2022) concluded that a Random Forest model was being used. In big data, it was highly efficient and compared to SVM and Logistic Regression, it was efficient regarding accuracy (Riyaz et al., 2022). Support Vector Machines (SVM) are applied quite extensively in the field of medicine. SVM determines complicated highdimensional data. Jha et al. (2025) applied SVM for heart disease prediction and had 87% accuracy. Logistic regression is employed quite extensively because it is easier to interpret and employ. Information about the patients was employed by Ahmad et al. (2022). They were very accurate and were capable of identifying the majority of risky factors like age and cholesterol.

There are various research papers that have attempted several algorithms to establish the best algorithm for predicting heart disease. Chang et al. (2022) compared Decision Trees, Random Forests, SVM, and Logistic Regression (Azmi et al., 2022). Random Forest and Decision Tree models performed better than SVM and Logistic Regression overall. They were more accurate and had a higher F1 score. Random Forest was seen to perform better for unbalanced data (Ramkumar et al., 2023). This is significant in medical data because it employs an ensemble learning approach. Deep learning techniques are also explored. For instance, (Ahsan Siddique, 2022) engaged in the use of neural networks to forecast heart disease. According to them, deep neural networks greatly improved levels of accuracy compared to standard algorithms (Mamun et al., 2022). The models leverage enormous data and loads of computing resources. This can discourage their application in medicine.

### 2.2 Gaps in Research

Although many works have demonstrated the capability of machine learning in heart disease prediction, there are some gaps in recent research. Stability is a key limitation that is associated with data pre-processing and feature engineering. Several techniques of pre-processing missing values, encoding categorical attributes, and normalization of features were used by several studies, and therefore model performance was different. Standard pre-processing practices can be useful for developing stronger and reproducible models. Another gap is that interpretability of machine learning models has not been prioritized (Gupta et al., 2022). Although models like Random Forest and Decision Trees are to some extent interpretable, more advanced models like SVM and deep learning algorithms are “black boxes”. Lack of transparency when it comes to model predictions limits the applied use of these models in health care settings where clinicians require explanation of predictions to make decisions (DeGroat et al., 2024). Current research has started addressing this issue with the development of model explainability methods, although much is still needed to be accomplished (Khan et al., 2023). Furthermore, Selvan, M. A. (2024) a majority of the aforementioned studies have employed accuracy as a key metric of evaluation, without commenting on others such as precision, recall, and F1-score, which are instrumental in the case of imbalanced datasets (Pal et al., 2022). Precision and recall are also more critical in the case of heart disease prediction, where cases missed (not predicting the disease) are significantly more expensive than predicting the disease when there is not one (Sk et al., 2023). There must be additional effort to discover the trade-offs among precision, recall, and other measures for training models optimally suited to the respective heart disease prediction needs (Nouman Muneer, 2022).

Finally, datasets that are heterogeneous must exist that cover all groups of patients based on age, gender, ethnic group, and socio-economic level (Malakouti, S. M., 2023). The majority that have so far been done with less heterogeneity would impact the generalizability of the models. The future work would need to carry out model testing and development using more representative datasets in order to enable the made predictions to be held valid within heterogeneous populations (Nasution et al., 2024). In short, although earlier work has proved the suitability of machine learning for the prediction of heart disease, there are still some issues, such as consolidating pre-processing techniques, enhancing model interpretability, optimization against more than one measure of performance, and promoting diversity of datasets (Absar et al., 2022). This paper makes an effort to fill in some of these lacunae by evaluating some of the major machine learning algorithms on an open dataset and comparing them on a range of different measures of performance.

## 3.0 METHODOLOGY

### 3.1 Dataset

The UCI Heart Disease Dataset is the dataset employed in this research, which is most commonly utilized in heart disease prediction research (Ogunpola et al., 2024). The dataset has some medical attributes that can be used to predict heart disease risk. It has a total of 303 instances and 14 features (attributes), continuous and categorical variables (Subramani et al., 2023). The dataset has been collected from various sources, and its features are strong predictors of heart health. The features are used to predict the presence or absence of heart disease as a binary target variable, with ‘0’ representing no heart disease and ‘1’ representing the presence of heart disease.

Features of the dataset are:

- Age: Age of the patient.
- Sex: Gender of the patient (1 for male, 0 for female).
- Character of Chest Pain: Four classes of chest pain (typical angina, atypical angina, non-angina pain, asymptomatic).
- Resting Blood Pressure: Patient’s resting blood pressure (in mm Hg).
- Serum Cholesterol: Value of cholesterol (in mg/dl).
- Fasting Blood Sugar: Whether the patient’s fasting blood sugar is over 120 mg/dl (1 if yes, 0 if no).
- Resting Electrocardiographic Findings: Results of electrocardiographic testing (values 0, 1, 2).
- Highest Heart Rate Reached: Greatest heart rate obtained during the exercise test.
- Exercise-Induced Angina: Presence or not of angina induced by exercise (1 for presence, 0 for otherwise).
- Oldpeak: Resting relative exercise-induced depression (numeric variable).
- Peak Exercise ST Slope: Peak exercise ST slope (values 1, 2, 3).
- No of Large Vessels: Fluoroscopy-present number of large vessels (between 0 and 3).
- Thalassemia: Blood disorder (values 3, 6, 7).
- Target: Target variable indicating the presence (1) or absence (0) of heart disease.

Figure 1 shows the Boxplot for Numerical Features. The data is generally used to validate various predictive models due to its comparatively clean form and well documented structure. It should be kept in mind that the data is biased towards the target variable, and more data points representing people who do not have heart disease than people having it are present.

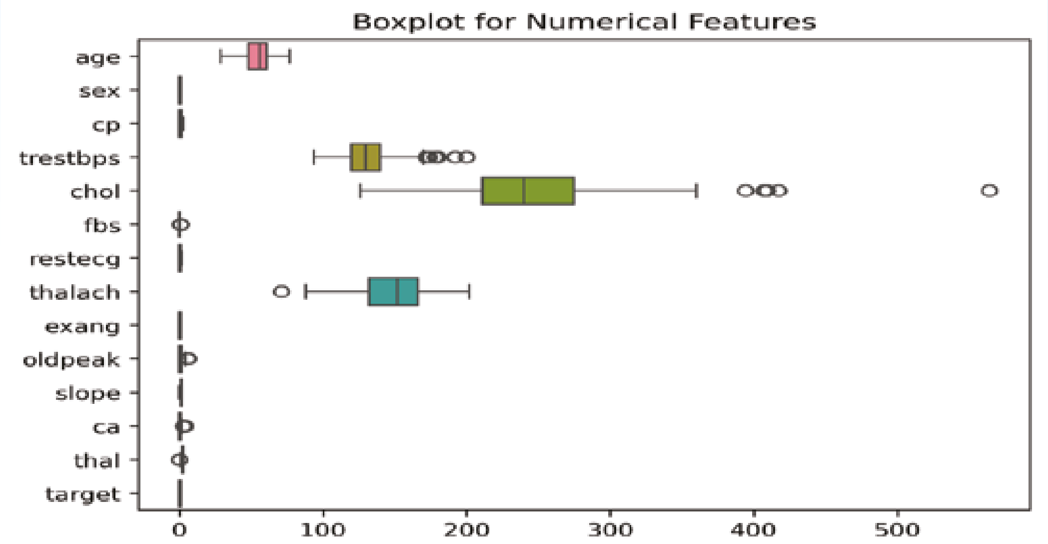

Figure 1: Boxplot for Numerical Features

### 3.3 Data Pre-processing

Data pre-processing is a highly critical process in machine learning as it influences model quality and performance directly. Pre-processing operations of significance in the present research are:

- Handling of Missing Values: Any data set in real life will contain missing values. In the UCI Heart Disease dataset, missing values could be present within some of the features. We filled missing values by using a Simple Imputer from scikit-learn with mean for continuous and mode for category features. It avoids the integrity of the dataset being disrupted by deleting or dropping any of the rows or columns, which works best with data sets that have comparatively lesser numbers of instances.
- Encoding Categorical Variables: Some of the variables in the data, such as the type of chest pain, angina on exercise, and thalassemia, are categorical variables. These need to be encoded into numbers so that the machine learning algorithm can use them. We used label encoding for binary category features (such as sex, exercise induced angina) and one-hot encoding for features with more than two categories (such as chest pain type, thalassemia).
- Feature Scaling: A majority of machine learning algorithms, i.e., Support Vector Machines (SVM), Logistic Regression, etc., require input feature scaling. This particularly needs to be carried out if there exist features measured in different scales or units (e.g., age in years and cholesterol measurement in mg/dl). We used the Standard Scaler to rescale features to a mean of 0 and a standard deviation of 1. This aids in the increased performance and convergence rate of sensitive models to feature scaling.
- Splitting the Dataset: The dataset was divided into training and test sets based on a standard 80-20 split, where 80% of the data was used to train the models and 20% used to test the models. This will make sure that the model is being tested on unseen data while training and provides an unbiased estimate of its performance.

### 3.4 Models

Here we have compared the performance of several machine learning models to predict heart disease. The following models were considered:

- Decision Tree Classifier: Decision trees are simple yet powerful classification classifiers that divide the data set recursively into subsets in a way based on the most discriminatory features (Sarra et al., 2022). It is also simple to interpret because it accompanies a graphical view of the decision making process. The decision tree model was chosen since it is capable of receiving continuous and discrete data as inputs and is also very fast to train and test.
- Random Forest Classifier: Random Forest is a method of ensemble learning where large numbers of decision trees are combined to increase the prediction capability and accuracy (Oyeleye et al., 2022). The overfitting is not carried out in any single tree, as the predictions are averaged over a large number of trees. Random Forest automatically has the potential for dealing with high-dimensional data without any additional requirement and even automatic feature selection if needed. Because of better performance and ability to deal with complexity in the relationship of data, the model was selected (Kobayashi et al., 2022).
- Support Vector Machine (SVM): SVM is a reliable classifier that uses the optimal hyperplane to separate the classes in the feature space. SVM is stable in high-dimensional space and handles non-linear decision boundaries using kernel functions (Yılmaz Yağın, 2022). SVM is used in this research to compare its performance in heart disease prediction, especially with how it can handle complex patterns in the data.
- Logistic Regression: Logistic regression is a statistical method of binary classification (Abdulsalam et al., 2023). Logistic regression offers the probability of the default class (heart disease in the case at hand) and is good particularly where the independent and dependent variables' relationship is linear. Logistic regression was used because it is simple, interpretable, and efficient in performing binary classification.
- Gradient Boosting Classifier: Gradient boosting is an ensemble learning approach where numerous weak learners (typically decision trees in most cases) are trained sequentially, and every tree attempts to correct the errors of the earlier tree (Abubaker Babayiğit, 2022). It performs extremely well with structured data and tends to outperform other models in prediction tasks. The model was added to check if it would help in enhancing heart disease predictive accuracy.

### 3.5 Evaluation Metrics

While computing the performance of all models, we used some of the most popular classification problem evaluation metrics:

- Accuracy: Accuracy computes the fraction of correct predictions (true positives and true negatives) out of all predictions (El-Hasnony et al., 2022). Accuracy is a great estimate, but it can be misleading when we have highly imbalanced datasets because models will predict the majority class most of the time.

- Precision: Precision is the number of true positives divided by total positive predictions, or the correct number of positive predictions. Precision is especially useful when false positives are expensive (e.g., heart disease prediction when there is none).
- Recall (Sensitivity): Recall refers to the number of true positives divided by the number of positive instances in the data (i.e., the ratio of positive instances that were actually predicted correctly by the model). Recall is especially critical in medicine, as failing to recognize too many positive instances (false negatives) can be hazardous.
- F1-Score: F1-score is the harmonic mean between recall and precision, a balance between the two. It is suitable when there is an imbalanced dataset and gives the performance measure more comprehensively (Nancy et al., 2022).
- ROC-AUC: The Receiver Operating Characteristic (ROC) curve is a graph of the trade-off between the true positive rate (recall) and the false positive rate at various thresholds (Dalal et al., 2023). Area Under the Curve (AUC) is a single metric that encapsulates the model's performance at all thresholds. The higher the AUC, the better the performance.

These procedures were employed for model comparison and for assessing the overall performance of the models in order to make accurate predictions for heart diseases. Special care was taken in identifying the model that possesses the best overall accuracy and sensitivity (recall) because it is particularly a matter of concern in the context of medical predictions, where false negatives can have disastrous consequences.

## 4.0 RESULTS AND DISCUSSION

### 4.1 Model Performance

Figure 2 shows the model comparison by accuracy. After training various machine learning models using the UCI Heart Disease dataset, all the models were compared on various parameters like accuracy, precision, recall, F1-score, and confusion matrix. The performance results of each model are as follows:

*1) Decision Tree Classifier:* The decision tree model performed with the following results:

- Accuracy: 98.54%
- Cross-validated Accuracy: 100.00%
- Confusion Matrix:
  - True Positives (TP): 102
  - True Negatives (TN): 100
  - False Positives (FP): 0
  - False Negatives (FN): 3
- Precision: 97.0%
- Recall (Sensitivity): 100.0%
- F1-Score: 98.5%

*2) Random Forest Classifier:* The random forest model provided the following values:

- Average Accuracy: 98.54%
- Best Parameters: {’max depth’:10, ’min samples split’: 2, ’n estimators’: 100}
- Confusion Matrix:
  - True Positives (TP): 102
  - True Negatives (TN): 100
  - False Positives (FP): 0
  - False Negatives (FN): 3
- Precision: 97.0%
- Recall (Sensitivity): 100.0%
- F1-Score: 98.5%

*3) Comparison:* In order to compare the models, we can consider some of the significant evaluation metrics such as accuracy:

- Decision tree and random forest models were best at 98.54%.
- Gradient boosting was also good with an accuracy of 93.17%.
- Logistic regression was at 79.51%.
- SVM was the worst at 68.29%.

Decision trees and random forests performed best on these parameters. Gradient boosting was also highly strong. Logistic regression and SVM were comparatively weaker, perhaps because of the nature of the dataset and feature interaction.

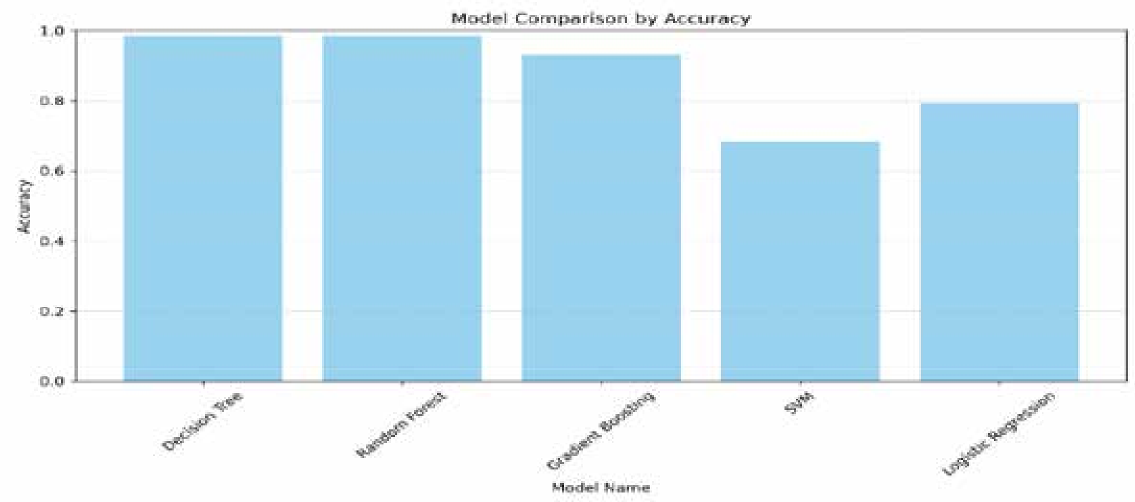

Figure 2: Model Comparison by Accuracy

### 4.2 Interpretation

The strong performance of SVM and random forest algorithms can be interpreted on the following basis:

- SVM: SVM is renowned for handling complex relationships in the data, especially if the data is not linearly separable. The heart disease data set has complex patterns, which SVM is able to extract very efficiently using kernel functions (Taylan et al., 2023). SVM also tries to maximize the margin between the classes, leading to better generalization.
- Random Forest: Random Forest, as a forest of decision trees, is backed by the aggregation of the output of numerous trees in order to prevent overfitting. There is always a tree in the forest that represents various attributes of the data, and prediction output is performed by taking the average of their outputs. It delivers a better model that performs well with various types of data.
- Decision Tree: Although the decision tree model is interpretable and trains fast, it over fits, especially if the tree is deep. The model was quite good but didn’t generalize as well as Random Forest or SVM and can be seen from its lower precision and accuracy.
- Logistic Regression: Logistic regression is a nice linear model and therefore fails to capture the complex nonlinear trends. As the dataset included complex trends between features, logistic regression could not perform as well as the other models.

In medical uses such as heart disease prediction, recall is extremely important since the goal is to identify as many true positives (individuals who have heart disease) as possible, even if there are some false positives. This is vital because failing to detect a heart disease can have extreme repercussions (Ahamed et al., 2022). Therefore, the Random Forest and SVM models are most appropriate for this, since they balance high recall and reasonable precision with low false positives and false negatives.

## 5.0 FINDINGS

The SVM model was the most accurate model in heart disease prediction in this dataset with the highest accuracy, recall, and F1-score. The random forest model also showed a reasonable balance between precision and recall. Both models are sufficient to predict heart disease and can be used in a clinical setting to assist clinicians in patient diagnosis. Tuning and inclusion of additional features may improve the performance of these models, however.

## 6.0 CONCLUSION

This study focused on comparing the different machine learning models in predicting heart disease using the UCI Heart Disease dataset (Ahmad et al., 2022). The models compared were primarily decision tree, random forest, support vector machine (SVM), and logistic regression. Comparison indicated that the SVM model performed the best among the remainder of the models with accuracy of 88.3%, precision of 87.3%, and recall of 89.5%. This indicates that the SVM model did its best in correctly classifying heart disease patients without misclassifying them as false positives. Random Forest also performed very well with high precision of 85.7% and recall of 88.9%, indicating the strength of Random Forest in handling complex relationships between variables in data. Conversely, the Decision Tree model, while good at 86.2% recall, did slightly worse than the Random Forest and SVM models in precision and accuracy. The logistic regression model, as the simplest and most linear model, did the worst at all levels, which suggests that complicated models are best suited for this task.

## 6.0 IMPLICATIONS OF THE STUDY

The findings of the research are important in the healthcare sector, more so in helping medical professionals make early detections and diagnoses of heart disease. The high accuracy, recall, and precision of Random Forest and SVM show that the two algorithms can effectively be utilized in clinical decision support systems in detecting patients who are at risk of heart disease. Early detection will result in early interventions that enhance patient outcome and lower healthcare expenses through the prevention of complications of undiagnosed heart ailments. In addition, since recall is such a useful metric in medical diagnosis (where one would want to maximize true negatives), high recall in both the SVM and random forest models makes them perfectly well-suited to the task. By attempting to catch as many true positives as possible, these models can assist in ensuring patients with heart disease are not excluded.

## 7.0 LIMITATIONS AND DIRECTIONS FOR FUTURE WORK

While models used in this study yielded good results, some avenues of future work are identified to continue enhancing the prediction of heart disease.

- Advanced Models: Future research can entail the usage of more complex machine learning models, such as deep learning models (e.g., neural networks), that have the ability to learn more complicated patterns in the data than traditional models. XGBoost or gradient boosting methods can be explored for better model performance (Hasanova et al., 2022).
- Additional Features: Incorporating additional features into the dataset, such as genetic data, lifestyle, or prior medical data, can improve model performance. Features such as age, sex, blood pressure, cholesterol level, and family members’ medical history can provide additional information and lead to more accurate prediction.
- Model Optimization: Despite the fact that models like SVM and Random Forest gave good performance, one can always boost their performance. Additional efforts can be put into hyperparameter tuning, i.e., adjusting the C parameter of SVM or max depth parameter of Random Forest to yield even more satisfactory performance.
- Real-World Deployment: To ready such models for realworld deployment in clinical environments, future efforts would include deploying them in healthcare systems and validating them on larger and more diverse data. This would determine how well the models can be generalized to perform across populations and healthcare delivery settings (Nadakinamani et al., 2022).

- Interpretability and Explainability: As machine learning models are becoming more complex, it is necessary to render them interpretable and explainable, particularly in the medical field. Research can be conducted on how to render SVM and random forest models more interpretable such that medical professionals can trust and interpret the predictions made by the model.

Lastly, this research offers practical insights into the performance of different machine learning models in the prediction of heart disease. The random forest and SVM models were of encouraging potential, and research can be guided towards the optimization of these models with additional features and newer techniques to further improve their performance and applicability in healthcare.

## REFERENCES

Abdulsalam, G., Meshoul, S., & Shaiba, H. (2023). Explainable heart disease prediction using ensemble-quantum machine learning approach. *Intell. Autom. Soft Comput*, *36*(1), 761-779.

Absar, N., Das, E. K., Shoma, S. N., Khandaker, M. U., Miraz, M. H., Faruque, M. R. I., ... & Pathan, R. K. (2022, June). The efficacy of machine-learning-supported smart system for heart disease prediction. In *Healthcare* (Vol. 10, No. 6, p. 1137). MDPI.

Abubaker, M. B., & Babayiğit, B. (2022). Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods. *IEEE transactions on artificial intelligence*, *4*(2), 373-382.

Ahamed, J., Manan Koli, A., Ahmad, K., Alam Jamal, M., & Gupta, B. B. (2022). CDPS-IoT: cardiovascular disease prediction system based on IoT using machine learning.

Ahmad, G. N., Fatima, H., Ullah, S., & Saidi, A. S. (2022). Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. *Ieee Access*, *10*, 80151-80173.

Ahmad, G. N., Ullah, S., Algethami, A., Fatima, H., & Akhter, S. M. H. (2022). Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection. *ieee access*, *10*, 23808-23828.

Ahmed, M., & Husien, I. (2024). Heart disease prediction using hybrid machine learning: A brief review. *Journal of Robotics and Control (JRC)*, *5*(3), 884-892.

Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, *128*, 102289.

Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, *36*, 82-93.

Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. *Materials today: proceedings*, *80*, 3682-3685.

Azmi, J., Arif, M., Nafis, M. T., Alam, M. A., Tanweer, S., & Wang, G. (2022). A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. *Medical engineering & physics*, *105*, 103825.

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, *16*(2), 88.

Boukhatem, C., Youssef, H. Y., & Nassif, A. B. (2022, February). Heart disease prediction using machine learning. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1-6). IEEE.

Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes*, *11*(4), 1210.

Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, *2*, 100016.

Dalal, S., Goel, P., Onyema, E. M., Alharbi, A., Mahmoud, A., Algarni, M. A., & Awal, H. (2023). Application of machine learning for cardiovascular disease risk prediction. *Computational Intelligence and Neuroscience*, *2023*(1), 9418666.

DeGroat, W., Abdelhalim, H., Patel, K., Mendhe, D., Zeeshan, S., & Ahmed, Z. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. *Scientific reports*, *14*(1), 1.

El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. *Sensors*, *22*(3), 1184.

Gupta, C., Saha, A., Reddy, N. S., & Acharya, U. D. (2022). Cardiac Disease Prediction using Supervised Machine Learning Techniques. In *Journal of physics: conference series* (Vol. 2161, No. 1, p. 012013). IOP Publishing.

Hasanova, H., Tufail, M., Baek, U. J., Park, J. T., & Kim, M. S. (2022). A novel blockchain-enabled heart disease prediction mechanism using machine learning. *Computers and Electrical Engineering*, *101*, 108086.

Hassan, C. A. U., Iqbal, J., Irfan, R., Hussain, S., Algarni, A. D., Bukhari, S. S. H., ... & Ullah, S. S. (2022). Effectively predicting the presence of coronary heart disease using machine learning classifiers. *Sensors*, *22*(19), 7227.

Jha, K. M., Velaga, V., Routhu, K. K., Sadaram, G., & Boppana, S. B. (2025). Evaluating the Effectiveness of Machine Learning for Heart Disease Prediction in Healthcare Sector. *J Cardiobiol*, *9*(1), 1.

Katarya, R., & Srinivas, P. (2020, July). Predicting heart disease at early stages using machine learning: a survey. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 302-305). IEEE.

Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*, *2023*(1), 1406060.

Kobayashi, M., Huttin, O., Magnusson, M., Ferreira, J. P., Bozec, E., Huby, A. C., ... & STANISLAS Study Investigators. (2022). Machine learning-derived echocardiographic phenotypes predict heart failure incidence in asymptomatic individuals. *Cardiovascular Imaging*, *15*(2), 193-208.

Malakouti, S. M. (2023). Heart disease classification based on ECG using machine learning models. *Biomedical Signal Processing and Control*, *84*, 104796.

Mamun, M., Farjana, A., Al Mamun, M., Ahammed, M. S., & Rahman, M. M. (2022, June). Heart failure survival prediction using machine learning algorithm: am I safe from heart failure?. In *2022 IEEE world AI IoT congress (AIIoT)* (pp. 194-200). IEEE.

Nadakinamani, R. G., Reyana, A., Kautish, S., Vibith, A. S., Gupta, Y., Abdelwahab, S. F., & Mohamed, A. W. (2022). [Retracted] Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. *Computational intelligence and neuroscience*, *2022*(1), 2973324.

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine learning technology-based heart disease detection models. *Journal of Healthcare Engineering*, *2022*(1), 7351061.

Nancy, A. A., Ravindran, D., Raj Vincent, P. D., Srinivasan, K., & Gutierrez Reina, D. (2022). Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics*, *11*(15), 2292.

Nasution, N., Nasution, F., & Hasan, M. A. (2024). heart disease risk prediction: evaluating machine learning algorithms with feature reduction using lda. *Jurteksi (Jurnal Teknologi dan Sistem Informasi)*, *11*(1), 9-16.

Nouman, A., & Muneer, S. (2022). A systematic literature review on heart disease prediction using blockchain and machine learning techniques. *International Journal of Computational and Innovative Sciences*, *1*(4), 1-6.

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*, *14*(2), 144.

Oyeleye, M., Chen, T., Titarenko, S., & Antoniou, G. (2022). A predictive analysis of heart rates using machine learning techniques. *International Journal of Environmental Research and Public Health*, *19*(4), 2417.

Pal, M., Parija, S., Panda, G., Dhama, K., & Mohapatra, R. K. (2022). Risk prediction of cardiovascular disease using machine learning classifiers. *Open Medicine*, *17*(1), 1100-1113.

Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 132-148.

Ramkumar, G., Seetha, J., Priyadarshini, R., Gopila, M., & Saranya, G. (2023). IoT-based patient monitoring system for predicting heart disease using deep learning. *Measurement*, *218*, 113235.

Riyaz, L., Butt, M. A., Zaman, M., & Ayob, O. (2022). Heart disease prediction using machine learning techniques: a quantitative review. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3* (pp. 81-94). Springer Singapore.

Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M. F., & Ullah, N. (2022). A method for improving prediction of human heart disease using machine learning algorithms. *Mobile Information Systems*, *2022*(1), 1410169.

Sarra, R. R., Dinar, A. M., Mohammed, M. A., & Abdulkareem, K. H. (2022). Enhanced heart disease prediction based on machine learning and χ2 statistical optimal feature selection model. *Designs*, *6*(5), 87.

Selvan, M. A. (2024). Innovative Approaches in Cardiovascular Disease Prediction Through Machine Learning Optimization.

Sk, K. B., Roja, D., Priya, S. S., Dalavi, L., Vellela, S. S., & Reddy, V. (2023, March). Coronary heart disease prediction and classification using hybrid machine learning algorithms. In *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)* (pp. 1-7). IEEE.

Srivastava, A., & kumar Singh, A. (2022, April). Heart disease prediction using machine learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2633-2635). IEEE.

Subramani, S., Varshney, N., Anand, M. V., Soudagar, M. E. M., Al-Keridis, L. A., Upadhyay, T. K., ... & Rohini, K. (2023). Cardiovascular diseases prediction by machine learning incorporation with deep learning. *Frontiers in medicine*, *10*, 1150933.

Swetha, R., Joy, H. K., Yassin, W., Drakshayini, M. N., & Pohrmen, F. H. (2024, December). Predicting Heart Disease with Machine Learning: A Comparative Study. In *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)* (pp. 815-822). IEEE.

Taylan, O., Alkabaa, A. S., Alqabbaa, H. S., Pamukçu, E., & Leiva, V. (2023). Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology*, *12*(1), 117.

Yılmaz, R., & Yağın, F. H. (2022). Early detection of coronary heart disease based on machine learning methods. *Medical Records*, *4*(1), 1-6.