IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

# Heart Disease Prediction using Machine Learning and Comparative Classification

**Sanika Subhash Patil[1] and Mrs. Snigdha Shukla[2]**
[1]Student, Master of Computer Applications,[2]
Assistant Professor, Master of Computer Applications,
Sadhu Vaswani Institute of Management Studies for Girls, Pune,
Savitribai Phule Pune University, Pune, Maharashtra, India.

**Abstract :** *Heart disease is a major health concern, and recognizing patterns in routine clinical measurements may help identify patients who require further medical assessment. This study examines the use of supervised machine-learning methods to classify the presence or absence of heart disease using the Cleveland Heart Disease dataset. The dataset contains 303 patient records described by thirteen clinical attributes. Seven classification algorithms—Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, Gaussian Naive Bayes, and Gradient Boosting—were trained and compared under a common experimental procedure. The records were divided through a stratified 80:20 training–testing split. Numerical features were standardized using statistics calculated from the training data only. The models were assessed using Accuracy, Precision, Recall, F1-score, ROC-AUC, and stratified ten-fold cross-validation. Random Forest achieved the strongest test performance, with 83.6% accuracy, an F1-score of 0.865, and a ROC-AUC of 0.916. The leading features in its importance ranking were chest-pain type, major-vessel count, thalassemia status, and maximum heart rate. The results indicate that machine learning can help identify patterns in clinical data; however, the model should be viewed as an academic decision-support experiment and not as a diagnostic tool. Validation on larger and independent patient groups is necessary before any clinical application can be considered.*

**Keywords:** Heart Disease Prediction, Machine Learning, Classification, Random Forest, Cleveland Dataset, Clinical Features, Model Evaluation.

## I. INTRODUCTION

Heart disease affects people across different age groups and places a continuing burden on individuals, families, and healthcare services. Early recognition of possible heart-related problems is important because timely medical assessment can help healthcare professionals decide whether additional tests or treatment are required. Clinical evaluation may involve several kinds of information, including a patient's symptoms, blood pressure, cholesterol level, electrocardiogram findings, and response to exercise. Looking at these measurements together can be challenging, particularly when many attributes need to be considered at the same time.

Machine learning offers a way to examine relationships within existing clinical records. A model can learn patterns from patient data for which the outcome is already known and then use those patterns to classify another record. In a heart-disease classification task, the input may include age, chest-pain category, resting blood pressure, cholesterol, exercise-induced angina, and maximum heart rate. The model produces a predicted class based on the information supplied. Such an output can support further investigation, but it cannot replace a medical examination or the judgement of a qualified healthcare professional.

This research compares seven established machine-learning algorithms using the Cleveland Heart Disease dataset. Each model is evaluated through the same data split and a common set of performance measures. The purpose is to identify

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

which of the selected classifiers performs most effectively on this dataset and to examine the clinical attributes that contribute to the leading model's predictions.

## II. OBJECTIVE OF STUDY

The objective of this study is to examine the use of machine-learning techniques for predicting the presence or absence of heart disease from clinical attributes. The study aims to prepare the dataset for classification, develop and compare seven supervised learning models, evaluate their performance using multiple metrics, identify influential clinical features, and discuss the practical value and limitations of the resulting predictions.

## III. LITERATURE REVIEW

Research into computer-assisted heart-disease prediction has used a range of statistical and machine-learning methods. Classical approaches such as Logistic Regression have been applied because they provide a familiar method for estimating the relationship between clinical variables and a binary outcome. Decision Trees have also been used because their rule-based structure can make some classification decisions easier to follow. These methods established a useful foundation for later work involving more flexible algorithms.

Ensemble methods, particularly Random Forest and Gradient Boosting, have attracted attention because they combine multiple decision trees and can represent nonlinear relationships among input variables. Support Vector Machines have also been used for classification tasks, although their performance can be affected by feature scaling and model settings. K-Nearest Neighbors and Naive Bayes provide additional approaches based on distance and probability, respectively. Comparing several algorithms on the same data can help show whether one method offers a more suitable balance of predictive performance.

The Cleveland Heart Disease dataset is commonly used in academic experiments because it contains clinical measurements and a labelled outcome. However, results from a small dataset need to be interpreted carefully. Accuracy alone does not describe the types of errors a model makes. Precision, recall, F1-score, and ROC-AUC provide additional information about how the classifier distinguishes between the two classes. Careful separation of training and testing data is also important so that information from the test records does not influence model development.

## IV. RESEARCH GAP

A comparison of heart-disease classifiers can be difficult when studies use different preprocessing methods, data partitions, and evaluation measures. A high accuracy score by itself may hide the number of positive cases missed by a model or the number of negative cases incorrectly flagged. In addition, a prediction is more useful to examine when the features influencing the model are reported. Performance on a single dataset also does not establish that a model will generalize to patients from other hospitals or populations.

To address these concerns within the selected dataset, this study evaluates seven classifiers using a shared stratified train–test split and a consistent set of performance metrics. It also examines the confusion matrix and feature-importance ranking of the best-performing classifier. The study is limited to a comparative experiment on the Cleveland dataset; it does not claim to provide external clinical validation.

## V. RESEARCH METHODOLOGY

### Research Design

The research follows an experimental comparative design. The process begins with checking the clinical dataset, followed by stratified partitioning and feature standardization. Seven classification algorithms are trained on the prepared training data. Their predictions are evaluated using the held-out test data, while stratified ten-fold cross-validation is carried out on the training subset to provide an additional view of model performance. The strongest model is then examined through its confusion matrix and feature-importance ranking.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

**System / Model Architecture**

The proposed workflow begins with the Cleveland dataset containing 303 patient records and thirteen input features. The records are checked for missing values and duplicates before being divided into training and testing subsets. Standardization is performed without using information from the test subset. The prepared data are passed to the seven classifiers, and their outputs are compared using the selected evaluation measures. Finally, the leading classifier is interpreted through its classification errors and feature-importance ranking.

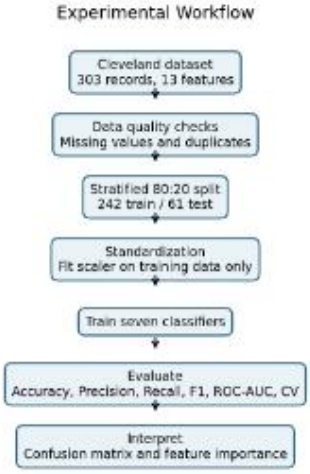



**Figure 1. Research workflow for heart-disease prediction.**

## VI. DATASET AND FEATURE SELECTION

**Data Collection and Feature Selection**

The study uses the Cleveland Heart Disease dataset, originally assembled at the Cleveland Clinic Foundation and made available through the UCI Machine Learning Repository. The collection contains 303 patient records, thirteen clinical input attributes, and a target variable representing the presence or absence of heart disease. The features include demographic information, symptom categories, blood measurements, electrocardiographic observations, and exercise-related indicators.

The thirteen input attributes are age, sex, chest-pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), fasting blood sugar (fbs), resting electrocardiographic result (restecg), maximum heart rate (thalach), exercise-induced angina (exang), ST depression (oldpeak), slope of the peak exercise ST segment (slope), number of major vessels (ca), and thalassemia status (thal). The target is coded as disease present (1) or absent (0).

**Table 1. Clinical features used in the classification task.**

| Feature | Description |
|---|---|
| age | Patient age in years |
| sex | Sex (1 = male, 0 = female) |
| cp | Chest-pain type (0–3) |
| trestbps | Resting blood pressure (mm Hg) |
| chol | Serum cholesterol (mg/dl) |
| fbs | Fasting blood sugar above 120 mg/dl |

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

| Feature | Description |
|---|---|
| restecg | Resting electrocardiographic result (0–2) |
| thalach | Maximum heart rate achieved |
| exang | Exercise-induced angina (1 = yes, 0 = no) |
| oldpeak | ST depression induced by exercise relative to rest |
| slope | Slope of peak exercise ST segment (0–2) |
| ca | Number of major vessels coloured by fluoroscopy |
| thal | Thalassemia status (0–3) |
| target | Heart disease present (1) or absent (0) |

**Data Quality and Class Distribution**

The supplied study reports that the 303 records contain no missing values after inspection. The class distribution consists of 165 positive records (54.5%) and 138 negative records (45.5%). Since the two groups are relatively close in size, the experiment did not introduce aggressive resampling or explicit class weighting.

**Data Preprocessing**

Before model training, the records were checked for missing entries and duplicate rows. The dataset was divided using an 80:20 stratified split, resulting in 242 training records and 61 testing records. Stratification was used to preserve the class proportions as closely as possible in both subsets. Continuous features were standardized to a zero mean and unit variance. The StandardScaler was fitted only on the training subset and then applied to the test subset. This procedure prevents test-set information from influencing the scaling step.

Standardization is particularly relevant for distance-based and gradient-based methods, including K-Nearest Neighbors, Support Vector Machine, and Logistic Regression. Applying the same preprocessing procedure across the classifiers supports a more consistent comparison.

## VII. RISK / TARGET CLASSIFICATION

The prediction task is framed as binary classification. A target value of 1 represents a record labelled as having heart disease, while a target value of 0 represents a record labelled as not having heart disease. The model learns the relationship between the thirteen clinical input features and this target label. The predicted class is an estimate based on patterns in the training records and should not be interpreted as a confirmed medical diagnosis.

## VIII. MODEL TRAINING AND MACHINE LEARNING MODELS

The prepared dataset was divided into 80% training data and 20% testing data using a stratified split. The training subset was used to fit the classifiers, while the test subset was kept separate for final performance evaluation. A fixed random state of 42 was used to support reproducibility. The study compared seven supervised machine-learning algorithms:

Logistic Regression: A statistical classification method that estimates the probability of a binary outcome. It provides a useful baseline for comparing the other models.

K-Nearest Neighbors: A distance-based method that assigns a class using the labels of nearby training examples. Feature scaling is important because distances are calculated from the input measurements.

Decision Tree: A tree-based classifier that divides records through a sequence of feature-based rules. It can represent nonlinear decision boundaries, although an individual tree may be sensitive to the training sample.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Random Forest: An ensemble method that combines predictions from multiple decision trees trained using random samples and feature subsets. It can capture interactions among clinical variables and was the best-performing classifier in this experiment.

Support Vector Machine: A classifier that identifies a decision boundary separating the classes. Its results can depend on feature scaling and the selected model parameters.

Gaussian Naive Bayes: A probabilistic classifier that applies Bayes' theorem and assumes conditional independence between input features given the class.

Gradient Boosting: An ensemble approach that builds decision trees sequentially, with later trees learning to reduce errors made by earlier stages.

## IX. MODEL EVALUATION

The classifiers were compared using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Stratified ten-fold cross-validation was also performed on the training data. Accuracy indicates the proportion of correct predictions among all test records. Precision measures how many records predicted as positive were actually positive. Recall measures the proportion of actual positive records identified by the model. The F1-score combines precision and recall into a single measure. ROC-AUC indicates how well the classifier separates the two classes across different decision thresholds.

Cross-validation provides an additional estimate of model performance across different portions of the training data. Reporting multiple measures is important because false-positive and false-negative predictions may have different practical consequences in a healthcare context.

## X. EXPERIMENTAL RESULTS

The seven classifiers produced the test-set results presented in Table 2. Random Forest achieved the highest accuracy, F1-score, and ROC-AUC among the evaluated models.

**Table 2. Performance of the seven classifiers on the held-out test set.**

| Model | Accuracy | Precision | Recall | F1-score | ROC-AUC | CV Mean |
|---|---|---|---|---|---|---|
| Random Forest | 0.836 | 0.781 | 0.970 | 0.865 | 0.916 | 0.844 ± 0.090 |
| Support Vector Machine | 0.820 | 0.775 | 0.939 | 0.849 | 0.883 | 0.802 ± 0.084 |
| K-Nearest Neighbors | 0.820 | 0.790 | 0.909 | 0.845 | 0.884 | 0.826 ± 0.071 |
| Naive Bayes | 0.820 | 0.790 | 0.909 | 0.845 | 0.876 | 0.818 ± 0.106 |
| Logistic Regression | 0.803 | 0.769 | 0.909 | 0.833 | 0.869 | 0.827 ± 0.105 |
| Gradient Boosting | 0.803 | 0.784 | 0.879 | 0.829 | 0.875 | 0.794 ± 0.106 |
| Decision Tree | 0.787 | 0.750 | 0.909 | 0.822 | 0.818 | 0.748 ± 0.082 |

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

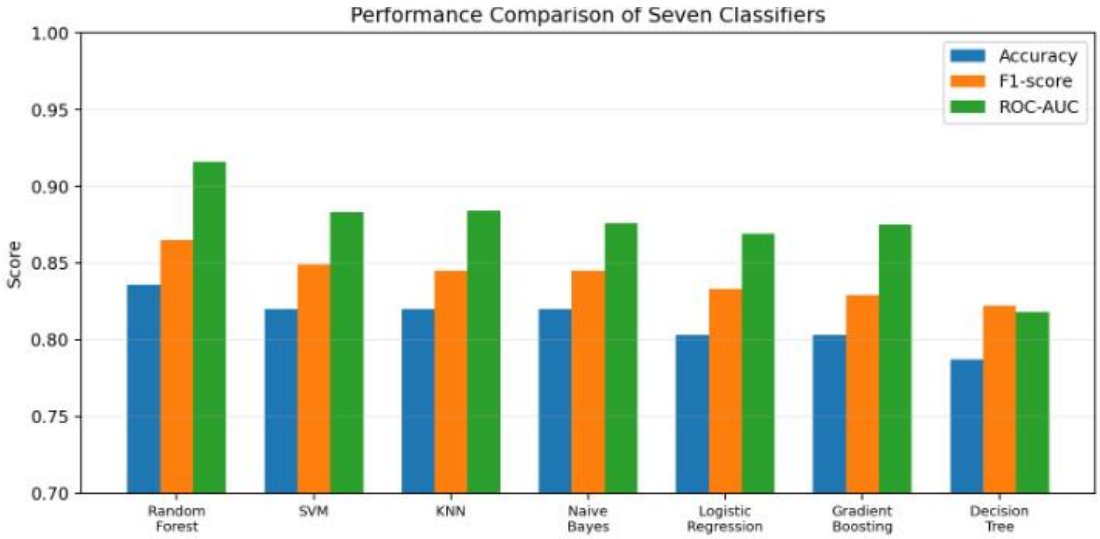



**Figure 2. Comparison of accuracy, precision, recall, and F1-score.**

The Random Forest classifier recorded 83.6% accuracy, 0.781 precision, 0.970 recall, an F1-score of 0.865, and a ROC-AUC of 0.916. Support Vector Machine, K-Nearest Neighbors, and Naive Bayes each achieved 82.0% accuracy. Decision Tree recorded the lowest accuracy at 78.7%. Under the conditions of this experiment, Random Forest provided the strongest overall results. These values describe performance on the selected test subset and should not be treated as a guarantee of performance on other patient populations.

## XI. CONFUSION MATRIX ANALYSIS

The confusion matrix provides a more detailed view of the Random Forest model's test predictions. Out of 61 test records, the classifier correctly identified 19 records as not having heart disease and 32 records as having heart disease. It incorrectly labelled 9 records as positive and missed 1 positive record by predicting the negative class. In other words, the matrix contains 19 true negatives, 32 true positives, 9 false positives, and 1 false negative.

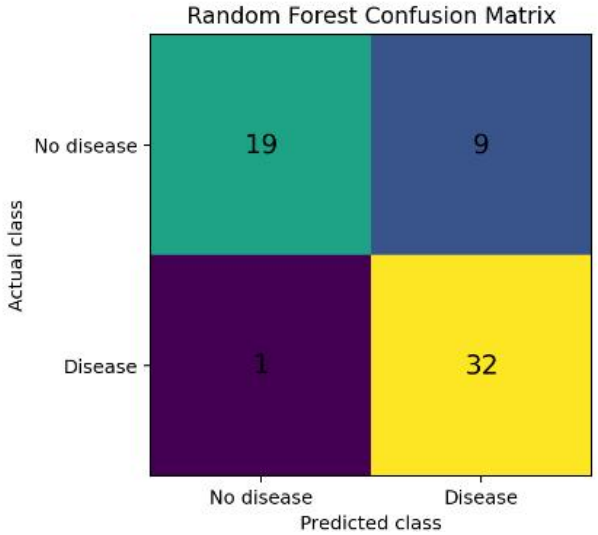



**Figure 3. Confusion matrix of the Random Forest classifier (61 test records).**

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

The low number of false negatives is relevant when considering a model for preliminary screening support. However, false-positive predictions can lead to additional examinations, cost, and concern for patients. The model's output must therefore be interpreted with care and alongside clinical evidence.

## XII. FEATURE IMPORTANCE AND MODEL INTERPRETATION

The feature-importance analysis reported for the Random Forest model identified chest-pain type (cp), the number of major vessels coloured by fluoroscopy (ca), thalassemia status (thal), and maximum heart rate (thalach) among the most influential predictors. These features represent different types of clinical information, including reported symptoms, examination findings, and the body's response to exercise.

The feature ranking helps describe which inputs the fitted model relied on most strongly. It should not be interpreted as evidence that any individual feature causes heart disease. Similarly, a feature's importance in this model does not mean that the same ranking will apply to every dataset or patient group.

## XIII. DISCUSSION

Random Forest achieved the strongest test results among the seven classifiers. Its ensemble structure can represent nonlinear relationships and combinations of clinical attributes, which may help explain its performance on this dataset. The results also show that recall was generally higher than precision across the evaluated models. This means that many positive records were identified, but some records without the disease label were also classified as positive. Both types of error are important in healthcare: missed positive cases may delay further assessment, while false alarms can lead to unnecessary follow-up.

The comparison supports the use of classical machine-learning methods as a research approach for examining clinical data. However, the dataset is relatively small and comes from a single source. The results should therefore be considered an experimental baseline rather than evidence that the selected model is ready for clinical use.

## LIMITATIONS

This study has several limitations. First, the Cleveland dataset contains only 303 patient records, so performance estimates may change with another sample or data partition. Second, the records originate from a single clinical source and may not represent the full range of patients seen in other hospitals or regions. Third, the hyperparameters were not subjected to an exhaustive optimization process. Fourth, no independent external dataset was available to test how well the model generalizes beyond the original collection. Finally, feature importance describes how the model uses input variables; it does not establish causal relationships. For these reasons, the model is intended for academic research and should not be used as an autonomous diagnostic or emergency decision system.

## XIV. CONCLUSION

This study compared seven supervised machine-learning classifiers for predicting the presence or absence of heart disease using the Cleveland Heart Disease dataset. The models were developed under a common experimental procedure that included a stratified 80:20 train–test split and feature standardization fitted only on the training data. Random Forest produced the strongest overall test performance, achieving 83.6% accuracy, an F1-score of 0.865, and a ROC-AUC of 0.916. Chest-pain type, major-vessel count, thalassemia status, and maximum heart rate were identified among the leading predictors in the fitted model.

The results show that machine-learning methods can identify patterns in routinely recorded clinical attributes. At the same time, the confusion matrix and the difference between precision and recall demonstrate why a model should be assessed using more than one performance measure. The findings are best understood as an academic comparison on a small dataset. They do not establish clinical readiness, and any real medical decision must remain with qualified healthcare professionals.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

## XV. FUTURE SCOPE

Future research can extend this work by evaluating the classifiers on larger datasets collected from multiple hospitals and patient populations. Systematic hyperparameter tuning may provide a more complete comparison of the algorithms. Additional explainability techniques could also be explored to understand individual predictions, provided that they are implemented and validated as part of the model pipeline. Independent external validation and prospective clinical studies would be required before considering the use of such a system in healthcare practice.

## REFERENCES

[1] Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., and Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64(5), 304–310.

[2] World Health Organization. (2021). Cardiovascular diseases (CVDs) fact sheet. World Health Organization.

[3] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[4] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[5] Cortes, C., and Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[6] UCI Machine Learning Repository. Heart Disease dataset. Center for Machine Learning and Intelligent Systems, University of California, Irvine.

Publication note: The experimental dataset details, metrics, and confusion-matrix counts in this manuscript have been retained from the supplied heart-disease manuscript. Before journal or conference submission, verify the numerical results against the final code and dataset, complete the institution and contact details, and check the references against the exact sources consulted.