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 5, June 2026
Impact Factor: 8.2

Heart Disease Prediction Using Machine Learning Techniques

Arjun Patil[1] and Anjali Roy[2]
[1] HOD and Assistant Professor, Department of M. Sc.IT
[2]Student, M. Sc.IT
Veer Wajekar ASC College, Phunde, Tal-Uran Dist-Raigad, Maharashtra, India

Abstract: *Heart disease is one of the leading causes of death worldwide. Early prediction and diagnosis of heart disease can help reduce mortality rates and improve healthcare systems. Machine learning techniques provide efficient methods for analysing medical data and predicting diseases with high accuracy. This research paper presents a comparative study of various machine learning algorithms used for heart disease prediction. The study uses the UCI Heart Disease dataset and applies preprocessing, feature selection, and classification techniques to evaluate model performance. Algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbours (KNN) are analysed using evaluation metrics including accuracy, precision, recall, and F1-score. Experimental results indicate that ensemble-based models provide better prediction accuracy compared to traditional classifiers.*

Keywords: Heart Disease; Machine Learning; K Nearest Neighbour (K-NN); Random Fores

I. INTRODUCTION

Heart disease, also known as cardiovascular disease, affects millions of people every year and has become a major public health concern. Early diagnosis plays a significant role in preventing severe complications and reducing healthcare costs. Traditional diagnostic methods depend heavily on medical experts and laboratory tests, which may be time-consuming and expensive.

Machine learning (ML) has emerged as a powerful technology in healthcare for disease prediction and medical decision-making. ML algorithms can identify hidden patterns in patient data and provide accurate predictions. By using patient health parameters such as age, blood pressure, cholesterol level, chest pain type, and heart rate, machine learning models can assist doctors in diagnosing heart disease efficiently.

The objective of this research is to analyse and compare different machine learning techniques for heart disease prediction and determine the most effective model.

II. KEYWORDS

Heart Disease; Machine Learning; K Nearest Neighbour (K-NN); Random Fores

III. LITERATURE REVIEW

Several researchers have applied machine learning techniques for heart disease prediction.

- Logistic Regression is widely used for binary classification problems and provides good interpretability.
- Decision Tree classifiers are useful for generating understandable decision rules.
- Random Forest improves prediction accuracy by combining multiple decision trees.
- Support Vector Machine (SVM) performs well on high-dimensional datasets.

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 5, June 2026
Impact Factor: 8.2

- Deep learning models such as Artificial Neural Networks (ANN) have recently shown promising results in medical diagnosis.

Previous studies reported prediction accuracies ranging from 80% to 95% depending on the dataset and preprocessing techniques used.

IV. DATASET DESCRIPTION

The study uses the UCI Heart Disease dataset, which contains patient medical records. The dataset includes several attributes related to cardiovascular health.

Dataset Attributes

Attribute	Description
Age	Age of the patient
Sex	Gender
Chest Pain Type	Type of chest pain
Resting Blood Pressure	Blood pressure level
Cholesterol	Serum cholesterol
Fasting Blood Sugar	Blood sugar level
ECG Results	Electrocardiographic results
Maximum Heart Rate	Maximum heart rate achieved
Exercise-Induced Angina	Presence of angina
Oldpeak	ST depression value
Target	Presence or absence of heart disease

V. METHODOLOGY

The proposed methodology consists of several stages:

Data Collection
↓
Data Preprocessing
↓
Feature Selection
↓
Model Training
↓
Model Testing
↓
Performance Evaluation

5.1 Data Preprocessing

Data preprocessing improves data quality and model performance. The following steps are performed:

- Handling missing values
- Removing duplicate 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 5, June 2026
Impact Factor: 8.2

- Data normalization
- Encoding categorical variables

5.2 Feature Selection

Feature selection helps identify the most relevant attributes affecting heart disease prediction. Correlation analysis and statistical methods are used to reduce unnecessary features.

VI. MACHINE LEARNING ALGORITHMS

6.1 Logistic Regression

Logistic Regression is a statistical classification algorithm used for binary outcomes.

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}$$

Advantages:

- Simple and interpretable
- Efficient for small datasets

6.2 Decision Tree

Decision Trees classify data using hierarchical decision rules.

Advantages:

- Easy visualization
- Handles categorical data effectively

Disadvantages:

- Risk of overfitting

6.3 Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy.

Advantages:

- High accuracy
- Reduces overfitting

6.4 Support Vector Machine (SVM)

SVM separates classes using an optimal hyperplane.

Advantages:

- Effective in high-dimensional spaces
- Good generalization capability

6.5 K-Nearest Neighbours (KNN)

KNN classifies data points based on nearest neighbouring samples.

Advantages:

- Simple implementation
- No training phase required

Disadvantages:

- Computationally expensive for large datasets

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 5, June 2026
Impact Factor: 8.2

VII. PERFORMANCE EVALUATION METRICS

The performance of machine learning models is evaluated using the following metrics:

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

VIII. EXPERIMENTAL RESULTS

The machine learning models are trained and tested using the dataset. The following table shows comparative results.

Algorithm	Accuracy
Logistic Regression	85%
Decision Tree	82%
KNN	84%
SVM	88%
Random Forest	91%

The Random Forest algorithm achieved the highest accuracy due to its ensemble learning capability.

IX. DISCUSSION

The experimental analysis shows that machine learning techniques can effectively predict heart disease. Ensemble methods such as Random Forest outperform traditional models because they reduce variance and improve stability.

SVM also demonstrates strong performance due to its capability of handling complex feature spaces. Logistic Regression provides interpretable results, making it useful in healthcare applications where transparency is important.

Despite high prediction accuracy, several challenges remain:

- Limited dataset size
- Data imbalance
- Missing medical records
- Privacy concerns in healthcare data

Future systems can integrate deep learning and real-time patient monitoring for improved prediction accuracy.

DOI: 10.48175/IJARSCT-35634

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 5, June 2026
Impact Factor: 8.2

X. CONCLUSION

This research paper presented a comparative analysis of machine learning techniques for heart disease prediction. Different classification algorithms including Logistic Regression, Decision Tree, KNN, SVM, and Random Forest were analyzed using medical datasets.

Among all models, Random Forest achieved the highest prediction accuracy. The study demonstrates that machine learning techniques can assist healthcare professionals in early diagnosis and treatment planning.

Future work may include:

- Deep learning integration
- Real-time IoT healthcare systems
- Explainable AI models
- Larger clinical datasets

REFERENCES

1. UCI Machine Learning Repository – Heart Disease Dataset
2. Chaurasia, V., & Pal, S. (2018). Early Prediction of Heart Diseases Using Data Mining Techniques.
3. Detrano, R. et al. International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease.
4. Krittanawong, C. et al. Machine Learning Prediction in Cardiovascular Diseases.
5. World Health Organization (WHO) Reports on Cardiovascular Diseases.