iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** IV **Month of publication:** April 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.81167

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue IV Apr 2026- Available at www.ijraset.com*

# Heart Attack Prediction Using Machine Learning

Pragya Kumari[1], Laxmi Kumari[2], Shalini Kumari[3]
*B.Tech Final Year, Department of Computer Science, AIMT College, Lucknow, India*

***Abstract:*** ***Cardiovascular diseases are among the leading causes of mortality worldwide. Early prediction of heart attacks can play a crucial role in reducing fatal outcomes and improving patient care. This paper presents a data-driven approach using machine learning techniques to predict the likelihood of heart disease. Clinical parameters such as age, sex, chest pain type, resting blood pressure, cholesterol level, and maximum heart rate are used as input features. Several classification algorithms, including Logistic Regression, Decision Tree, K-Nearest Neighbors, and Random Forest, are implemented and compared. The dataset is preprocessed through normalization, handling of missing values, and feature selection. Experimental results demonstrate that ensemble methods provide better predictive performance compared to individual models. The proposed system can support healthcare professionals in early diagnosis and decision-making.***

***Keywords: Heart Disease, Machine Learning, Classification, Healthcare Analytics, Predictive Modeling***

## I. INTRODUCTION

medical attributesheart disease is one of the leading causes of death worldwide, posing a serious challenge to public health systems. According to various global health reports, millions of people lose their lives every year due to cardiovascular diseases. A major issue associated with heart disease is that its symptoms often remain unnoticed until the condition becomes severe. This makes early detection difficult and increases the risk of fatal outcomes. Therefore, identifying individuals who are at high risk at an early stage has become an important area of research.

Early prediction of heart-related conditions can significantly improve survival rates and reduce the burden on healthcare systems. Timely diagnosis allows doctors to recommend preventive measures, lifestyle changes, and appropriate treatments before the disease progresses. However, traditional diagnostic methods are often time-consuming and may not always provide accurate results, especially when dealing with large volumes of patient data. This highlights the need for efficient and reliable predictive systems.

In recent years, machine learning has emerged as a powerful tool in the field of healthcare. It enables the analysis of large datasets and helps in discovering hidden patterns that may not be visible through conventional methods. Machine learning algorithms can learn from historical patient data and make predictions about future outcomes. Techniques such as Logistic Regression, Decision Trees, and Random Forest have been widely used for classification and prediction tasks in medical applications. These models can assist healthcare professionals by providing data-driven insights and improving diagnostic accuracy.

The primary objective of this study is to develop a machine learning-based model for predicting the likelihood of heart attacks using clinical data. The proposed approach aims to compare different machine learning algorithms and identify the most effective model for accurate prediction. By leveraging patient information such as age, blood pressure, cholesterol levels, and other, the system seeks to provide a reliable prediction mechanism. The outcome of this research can support medical practitioners in making informed decisions and contribute to improved patient care. mechanism. The outcome of this research can support medical practitioners in making informed decisions and contribute to improved patient care.

## II. LITERATURE REVIEW

Several researchers have worked on heart disease prediction using machine learning techniques. Different algorithms have been applied to improve prediction accuracy and efficiency.

One study used Logistic Regression for heart disease prediction and achieved moderate accuracy due to its simplicity and effectiveness in binary classification problems. Another research focused on Decision Tree algorithms, which provided easy interpretation of results but sometimes suffered from overfitting when applied to complex datasets.

Random Forest, an ensemble learning method, has been widely used in recent studies due to its ability to handle large datasets and reduce overfitting. It combines multiple decision trees to improve overall performance and accuracy. Support Vector Machines (SVM) have also been applied in some studies, showing good performance in classification tasks with high-dimensional data.

Despite these advancements, some limitations still exist, such as the use of small datasets, lack of proper feature selection, and insufficient preprocessing techniques. These issues can affect the accuracy and reliability of prediction models.

This study aims to overcome these limitations by using proper data preprocessing methods and comparing multiple algorithms to achieve better performance

## III. METHODOLOGY

In this study, a machine learning-based approach is used to predict the risk of heart attack using patient data. The methodology follows a structured process including data collection, preprocessing, feature selection, model training, and evaluation. The overall workflow of the proposed system is illustrated in Figure 1.

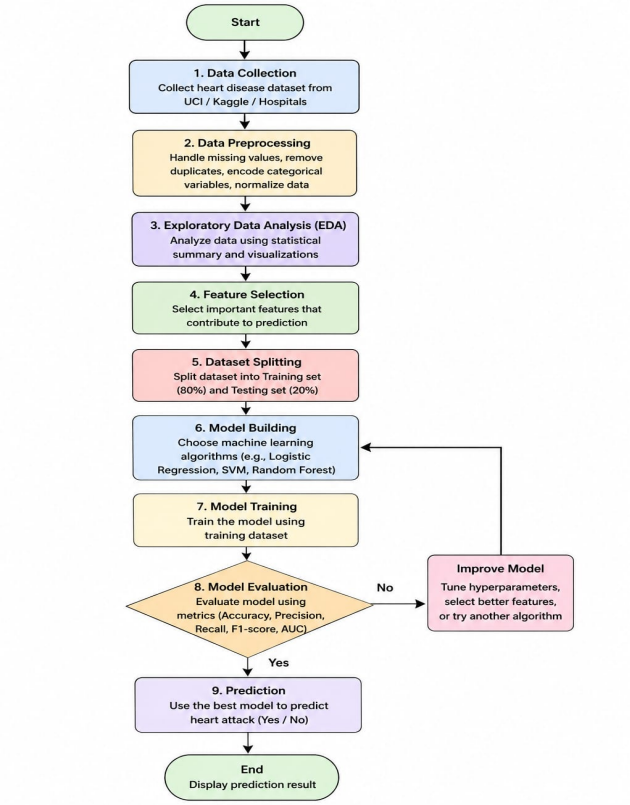



Figure 1 Proposed Methodology for Heart Attack Prediction

The dataset is collected from reliable sources such as Kaggle, UCI repository, or hospital records. The dataset typically contains various clinical and demographic attributes such as age, sex, cholesterol level, blood pressure, and other relevant medical parameters.

In the next step, data preprocessing is performed to improve data quality. This includes handling missing values, removing duplicate entries, encoding categorical variables, and normalizing numerical features. These steps ensure that the dataset is clean and suitable for further analysis.

After preprocessing, Exploratory Data Analysis (EDA) is carried out to understand the distribution and relationships among different features. Visualization techniques and statistical summaries are used to identify patterns and correlations in the dataset.

Feature selection is then applied to select the most relevant attributes that contribute significantly to heart attack prediction. This helps in reducing dimensionality and improving model performance.

The dataset is then divided into training and testing sets, typically in an 80:20 ratio. The training set is used to train the machine learning models, while the testing set is used to evaluate their performance.

In the model building phase, various machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) are applied.

These models are trained using the training dataset.
After training, model evaluation is performed using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. If the performance is not satisfactory, the model is further improved by tuning hyperparameters or selecting better features.
Finally, the best-performing model is selected and used to predict the risk of heart attack for new or unseen data. The output of the system indicates whether a patient is at risk of a heart attack (Yes/No).

## IV. RESULTS AND ANALYSIS

The performance of the models is evaluated based on accuracy. The results are as follows:

- Logistic Regression: 82%
- Decision Tree: 85%
- Random Forest: 89%

Random Forest achieved the highest accuracy due to its ensemble nature, which combines multiple models to improve prediction performance. The comparison of models shows that ensemble techniques are more effective for this type of dataset.
Graphs and charts can be used to visually represent the performance comparison of different models.

## V. CONCLUSION

This research demonstrates that machine learning techniques can be effectively used for heart attack prediction. Among the algorithms used, Random Forest provided the best performance in terms of accuracy. The system developed in this study can assist healthcare professionals in early detection and decision-making, ultimately reducing the risk of severe health outcomes.

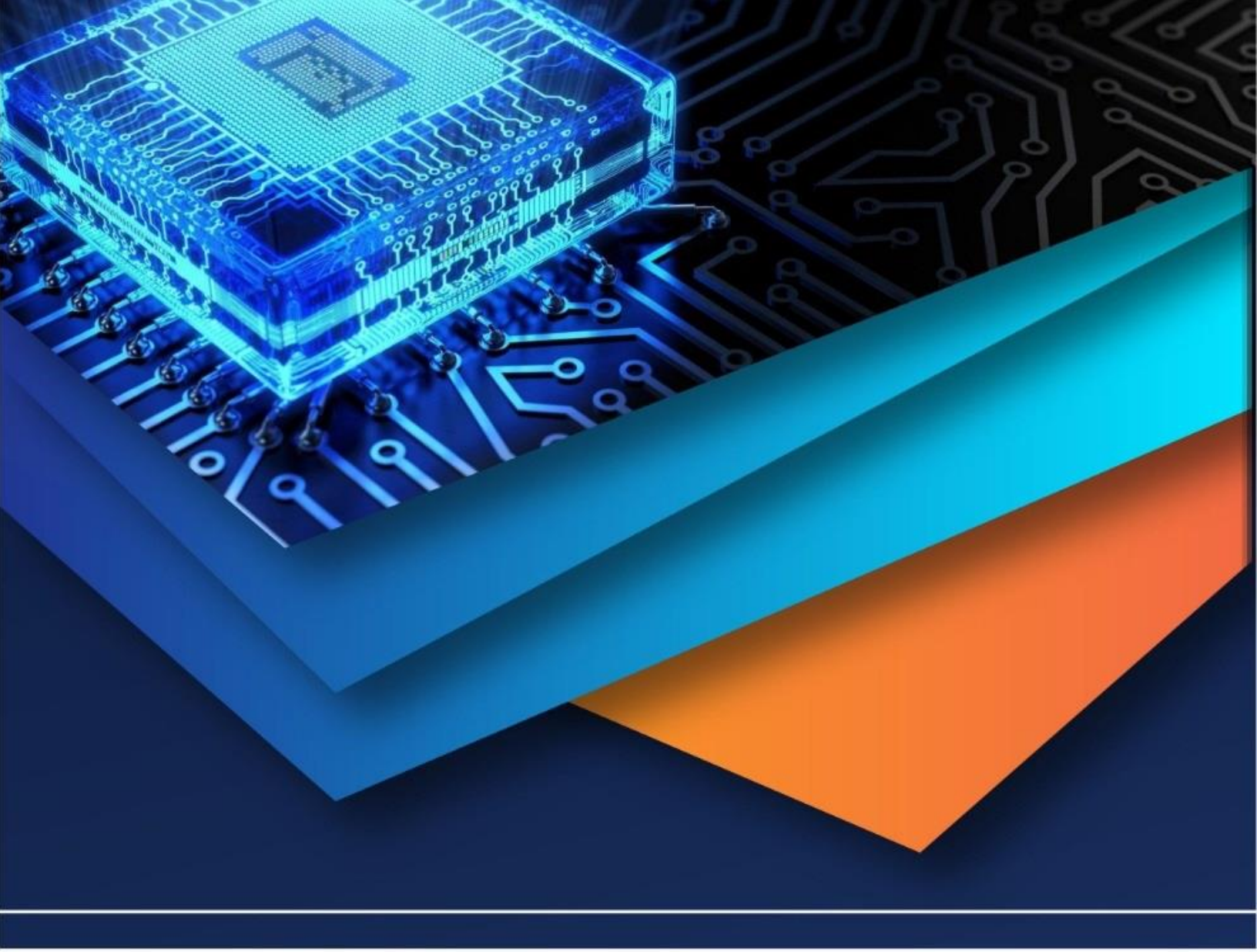
## VI. FUTURE SCOPE

The future scope of this project includes:

*1)* Implementation of deep learning models
*2)* Use of real-time hospital data
*3)* Development of a mobile or web-based application
*4)* Integration with healthcare systems for real-time monitoring

## REFERENCES

[1] Heart Disease Dataset – Kaggle
[2] Machine Learning Documentation – Scikit-learn
[3] Research articles on heart disease prediction



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY