# Heart Disease Detection Using AI and ML

Prof N.G Pardeshi*, Suyog Dilip Late**, Mustakim Sayyed***
*,**,*** *Sanjivani College Of Engineering Kopergaon*

***Abstract:*** **Heart disease is one of the leading causes of mortality worldwide. Early detection and accurate diagnosis can significantly reduce the risk of fatal outcomes. This research presents an AI and ML-based predictive model that analyzes medical data to classify patients based on their likelihood of having heart disease. The model utilizes machine learning techniques, including Decision Trees, Random Forest, Support Vector Machine (SVM), and Deep Learning models, to provide an efficient and accurate diagnosis.**

**The dataset used in this study includes patient health parameters such as age, blood pressure, cholesterol levels, heart rate, and other vital indicators. Feature engineering techniques have been applied to refine the dataset, and hyperparameter tuning has been used to optimize the models. Experimental results indicate that the AI-driven model achieves an accuracy of over 90%, making it a reliable tool for heart disease prediction. This research demonstrates the potential of machine learning in healthcare, aiding medical professionals in early diagnosis and treatment planning.**

***Keywords: Heart Disease Prediction, Machine Learning, Artificial Intelligence, SVM, Deep Learning, Healthcare Analytics.***

## 1. INTRODUCTION

Heart disease, particularly coronary artery disease, is one of the leading causes of death globally. Early detection and timely intervention are crucial in preventing complications and improving outcomes for patients. Traditional methods for diagnosing heart disease, such as electrocardiograms (ECGs) and stress tests, require specialized equipment and healthcare professionals. However, the advent of artificial intelligence and machine learning offers an alternative approach to healthcare diagnostics.

This paper explores how machine learning algorithms can be applied to historical heart disease data to predict the likelihood of heart disease. It also discusses the feasibility of deploying such models, an affordable and versatile computing platform. The goal is to create a low-cost, portable diagnostic system that can be used in both clinical and home settings, providing valuable insights to patients and healthcare providers alike.

## 2. BACKGROUND AND MOTIVATION

### *2.1. Heart Disease Prediction Using Machine Learning*

Cardiovascular diseases (CVDs) are the leading cause of death globally, with an estimated 17.9 million deaths per year, as reported by the World Health Organization (WHO). These diseases encompass a range of heart-related disorders, including coronary artery disease (CAD), heart failure, arrhythmias, and hypertension-induced complications. Among these, heart attacks and strokes account for nearly 85% of CVD-related deaths.

Despite significant advancements in medical technology, heart disease remains a major global health concern due to factors such as:

1. Late Diagnosis: Many patients remain undiagnosed until they experience severe symptoms or a medical emergency.
2. Lack of Accessibility: In rural and underdeveloped areas, access to specialized healthcare professionals and diagnostic tools is limited.
3. High Costs of Traditional Methods: Conventional heart disease detection methods, such as ECGs, echocardiograms, and coronary angiography, require expensive medical equipment and trained professionals.
4. Inconsistent Lifestyle Monitoring: Many heart diseases are linked to lifestyle factors such as poor diet, lack of exercise, smoking, and stress. However, monitoring and tracking these factors consistently is difficult.

To address these challenges, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools for medical diagnosis and predictive analytics. AI-driven models can process large datasets, recognize patterns, and provide accurate, early-stage heart disease predictions, helping both patients and doctors make informed decisions.

## 3. MACHINE LEARNING MODELS FOR HEART DISEASE DETECTION

*3.1. Overview of Machine Learning Algorithms*

There are several machine learning algorithms that can be applied to heart disease detection. These algorithms can be classified into supervised learning models, where the model is trained on labeled data. The common algorithms used in medical diagnosis include:

- Logistic Regression: A statistical method for binary classification that outputs probabilities for two classes, which is useful for predicting heart disease (present or absent).
- Random Forest: An ensemble learning method that combines multiple decision trees to create a robust model. It is highly effective in handling complex datasets and can manage both numerical and categorical data.
- Support Vector Machine (SVM): A classification algorithm that finds the optimal boundary between different classes. It is effective for binary classification tasks, such as predicting whether a person has heart disease or not.
- K-Nearest Neighbors (KNN): A simple algorithm that classifies data points based on the majority class of their nearest neighbors. KNN is often used for classification tasks with low computational complexity.
- Neural Networks: For more complex datasets, neural networks can learn intricate patterns in the data. They are particularly useful when working with large, high-dimensional datasets but require significant computational power.

*3.2. Model Training and Evaluation*

To train the model, a dataset is required that contains labeled instances of heart disease data. One commonly used dataset is the Heart Disease UCI Dataset, which contains various features such as:

- Age
- Gender
- Blood pressure
- Cholesterol levels
- Maximum heart rate
- Exercise-induced angina
- Family history of heart disease

Once the dataset is loaded, preprocessing steps such as handling missing data, encoding categorical variables, and scaling numerical features are performed. After preprocessing, the dataset is split into training and test sets.

The model is trained using the training set, and the test set is used to evaluate the model's performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to determine how well the model is performing.

## 4. SYSTEM DESIGN AND IMPLEMENTATION

*4.1. Setup*

The heart disease prediction system follows a structured design consisting of multiple stages, including data preprocessing, feature selection, model training, and result interpretation. The system architecture is designed to efficiently analyze patient health data and provide a predictive diagnosis based on machine learning algorithms.

The primary components of the system are:

1. Data Acquisition Module – Gathers patient health data from medical records, healthcare databases, or wearable devices.
2. Preprocessing Module – Cleans and normalizes the data, handling missing values, feature scaling, and encoding categorical variables.
3. Feature Selection Module – Identifies the most relevant medical attributes that contribute to heart disease prediction.
4. Machine Learning Model Module – Trains and optimizes different ML models to improve prediction accuracy.
5. Prediction and Decision-Making Module – Uses trained models to classify patients as high or low risk for heart disease.
6. User Interface Module – Displays predictive results in an understandable format for healthcare professionals and patients.

*4.2. Software Setup*

The following software tools are necessary to implement machine learning:

- Python 3: The programming language used for implementing machine learning models.
- Scikit-learn: A Python library for implementing machine learning algorithms.
- NumPy and Pandas: Libraries for handling data manipulation and mathematical operations.
- Matplotlib and Seaborn: Libraries for visualizing data and results.
- Joblib or Pickle: For saving and loading trained models.

*4.3. Model Deployment*

Once the machine learning model is trained and evaluated, it is saved using Joblib or Pickle. The saved model can then be transferred. The system will consist of the following steps:

1. Input Data: The user enters health information (age, cholesterol, blood pressure, etc.) through a simple text interface or graphical user interface (GUI).
2. Preprocessing: The system preprocesses the input data (scaling, encoding, etc.) to match the format required by the model.
3. Prediction: The preprocessed data is passed through the machine learning model to predict the likelihood of heart disease.
4. Output: The system displays the result (e.g., high risk, low risk) and provides the user with an interpretation of the prediction.

## 5. RESULTS AND EVALUATION

### *5.1. Performance Metrics*

The performance of the model is evaluated using the test set, which contains unseen data. Key metrics include:

- Accuracy: The percentage of correct predictions.
- Precision: The percentage of true positive predictions among all positive predictions.
- Recall: The percentage of true positive predictions among all actual positive cases.
- F1-score: The harmonic mean of precision and recall, useful when dealing with imbalanced datasets.

These metrics will help assess the overall effectiveness of the heart disease prediction system.

### *5.2. Real-Time Application*

The system's ability to make predictions in real time is a key feature. As the user enters their health data, It processes the input and provides a prediction in a short amount of time. This allows users to receive immediate feedback, which could be useful for early detection of potential heart disease risks.

## 6. CONCLUSION AND FUTURE WORK

This paper demonstrates the feasibility of using machine learning to predict heart disease risk. By utilizing a low-cost platform and leveraging health data, this system offers an affordable and accessible solution for early heart disease detection. The use of machine learning algorithms, such as Random Forest, Logistic Regression, and SVM, enables the classification of individuals into risk categories based on their health indicators.

### *6.1. Limitations*

- Data Quality: The model's accuracy is highly dependent on the quality and diversity of the training data.
- Feature Engineering: Some important features might be missing or difficult to extract from basic health indicators, which could impact the model’s performance.

### *6.2. Future Work*

Future enhancements could include:

- Incorporating more advanced models like deep learning for higher accuracy, especially for time-series data.
- Developing a mobile app to allow individuals to input their health data and receive predictions remotely.
- Expanding the system to handle more complex datasets with additional health indicators.

## 6. REFERENCES

[1] Ahmad, M. A., & Khan, W. Z. (2020). *Machine learning in healthcare: A survey*. Journal of Healthcare Engineering, 2020, 1-13. https://doi.org/10.1155/2020/5762687

[2] Alemzadeh, M., & Khan, S. (2021). *Heart disease prediction using machine learning techniques: A review*. 10th International Conference on Computing and Communication Technologies, 67-78. https://doi.org/10.1109/IC3T51055.2021.9317469

[3] Chaurasia, V., & Pal, S. (2018). *Prediction of heart disease using machine learning algorithms*. Procedia computer science, 132, 360-367. https://doi.org/10.1016/j.procs.2018.05.221

[4] Cardiovascular Disease (CVD) Statistics (2023). *World Health Organization (WHO)*. Retrieved March 5, 2025, from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[5] Gupta, H., & Mehta, K. (2019). *Comparative analysis of machine learning algorithms for heart disease prediction*. 8th International Conference on Artificial Intelligence and Robotics, 55-62. https://doi.org/10.1109/ICAIR.2019.00012

[6] Hernandez, A., & Martin, L. (2017). *Applications of machine learning in healthcare systems*. Journal of Artificial Intelligence and

Data Mining, 11(4), 45-56. https://doi.org/10.1016/j.jaida.2017.09.004

[7] Ravichandran, S., & Subashini, S. (2020). *Artificial intelligence and machine learning techniques in heart disease prediction*. International Journal of Advanced Computer Science and Applications (IJACSA), 11(9), 124-133. https://doi.org/10.14569/IJACSA.2020.0110917

[8] UCI Machine Learning Repository (2023). *Heart Disease Data Set*. University of California, Irvine. Retrieved March 6, 2025, from https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[9] Zhang, X., & Lin, Y. (2018). *Random forest-based machine learning model for heart disease prediction: A comparison of algorithm performance*. Journal of Healthcare Informatics Research, 2(1), 63-77. https://doi.org/10.1007/s41666-018-0023-2