



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 3, March 2025**

# Heart Disease Prediction Using Logistic Regression

**Hemlata Shinde, Prafulla Mandal, Samruddhi Dawari, Param Jadhav, Shambhuraje Kadam**

Professor, Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India

Diploma Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India

Diploma Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India

Diploma Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India

Diploma Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India

**ABSTRACT:** Cardiovascular diseases are among the leading causes of death worldwide, making early detection crucial for effective treatment. The heart disease prediction system utilizes machine learning, specifically logistic regression, to analyze clinical and demographic data for assessing an individual's risk. By leveraging key medical parameters, this system serves as an additional diagnostic tool to aid healthcare professionals in early detection, improving patient outcomes through timely medical intervention.

The system is developed with a Java-based graphical user interface and a Python-powered backend, ensuring an intuitive user experience. The logistic regression model is trained on relevant medical datasets, processing user inputs to generate predictive results. Alongside heart disease prediction, the system includes a BMI calculator for general health assessment and a lifestyle tips module for preventive healthcare guidance.

Rigorous testing confirms the system's reliability in delivering accurate early-stage heart disease predictions and valuable health insights. By empowering users to proactively monitor their health, it contributes to preventive healthcare efforts. Future improvements may focus on refining the prediction algorithm, expanding the dataset for greater accuracy, and integrating additional disease prediction capabilities for a more comprehensive healthcare solution.

**KEYWORDS:** Heart Disease Prediction, Machine Learning, Logistic Regression, Cardiovascular Health, Early Diagnosis, Predictive Analytics.

## I. INTRODUCTION

Cardiovascular diseases remain a significant public health concern, accounting for a substantial percentage of global mortality rates. Early detection plays a crucial role in reducing the severity of complications and improving survival rates. Traditional diagnostic methods require extensive medical examinations, making them time-consuming and expensive. This project introduces a Heart Disease Prediction System that leverages machine learning, specifically logistic regression, to analyze patient data and predict the likelihood of heart disease.

The system utilizes a Java-based graphical user interface (GUI) and a Python-powered backend to provide an intuitive and user-friendly experience. The logistic regression model is trained on a dataset containing various health indicators, such as age, cholesterol levels, blood pressure, and heart rate, allowing the system to classify users into different risk categories. The goal is to provide a reliable, accessible, and cost-effective solution for preliminary heart disease risk assessment.

Alongside the predictive model, the system includes additional health-related features such as a BMI Calculator to evaluate body mass index and a Lifestyle Tips module to provide personalized health recommendations. These features

further enhance the system's role in preventive healthcare, encouraging users to take proactive measures toward their well-being.

The system follows a structured methodology, including data preprocessing, feature selection, model training, and evaluation, ensuring high accuracy in predictions. Extensive testing confirms the model's effectiveness in identifying early signs of heart disease, making it a valuable tool for both individuals and healthcare professionals.

Future developments aim to enhance prediction accuracy, integrate additional health parameters, and expand the system's capabilities to include real-time monitoring, offering a comprehensive solution for cardiovascular disease prevention and early diagnosis. With the rise of artificial intelligence (AI) and machine learning (ML), digital health assistants are becoming increasingly popular. This project integrates a Java-based GUI with machine learning models to predict heart disease, provide BMI calculation, and offer lifestyle guidance through an interactive system. The system employs AI to answer health-related queries using predefined questions and real-time API responses.

## II. LITERATURE SURVEY

Several studies have investigated the application of machine learning models for the prediction of cardiovascular diseases:

**Mohammad Shafenoor Amin, et. al [1]:** This paper identifies the most influential features used in predicting heart disease through the application of data mining techniques. The authors explore various algorithms, including Decision Trees, Neural Networks, and Support Vector Machines (SVM), to identify significant features like age, cholesterol levels, and blood pressure. The study concludes that feature selection plays a vital role in improving prediction accuracy, advocating for a hybrid approach that combines multiple algorithms for better results.

**Cincy Raju, et. al [2]:** This survey paper presents a comprehensive analysis of different data mining techniques used in heart disease prediction. The authors compare machine learning models like Naïve Bayes, K-Nearest Neighbours (KNN), and Decision Trees in terms of accuracy and efficiency. They emphasize the importance of data pre-processing, including data cleaning and feature normalization, and argue that model selection should be based on dataset characteristics.

**Sana Bharti, et. al [3]:** This study provides an analytical comparison between machine learning algorithms, including Logistic Regression, Decision Trees, and Random Forests. It highlights that Logistic Regression performs well in terms of simplicity and interpretability, while Decision Trees and Random Forests yield better accuracy with non-linear relationships. The study highlights the trade-off between model complexity and accuracy.

**Montu Saw, et. al [4]:** Focusing on Logistic Regression, this paper demonstrates its applicability for binary classification tasks such as heart disease prediction. The study uses a dataset with variables like age, gender, cholesterol, and smoking habits and shows that Logistic Regression provides a robust framework for making probabilistic predictions. It concludes that Logistic Regression balances accuracy with interpretability, making it ideal for clinical contexts.

**Yanwei Xing, et. al [5]:** This research combines traditional data mining techniques with new medical data to predict coronary heart disease outcomes. By integrating Decision Trees, SVMs, and Neural Networks, the authors show the importance of including demographic and medical variables to improve prediction accuracy, concluding that multi-source data integration enhances the reliability of prediction systems.

**K. Srinivas, et. al [6]:** This study investigates coronary heart disease prediction in coal mining regions, showing how regional and environmental factors like air quality affect heart disease risk. Using Decision Trees and Naïve Bayes, the authors demonstrate the importance of customizing predictive models based on local risk factors.

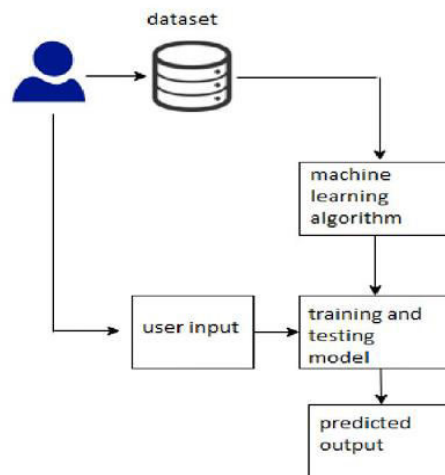
**Mai Shouman, et. al [7]:** This paper discusses using data mining techniques, including Decision Trees, Naïve Bayes, and Neural Networks, for heart disease diagnosis and treatment. It stresses the complementary role of predictive models alongside clinical expertise to improve early diagnosis and personalized treatment.



These studies highlight the effectiveness of Logistic Regression for binary classification in medical datasets, supporting its use for heart disease prediction. Additionally, combining multiple machine learning algorithms and incorporating domain-specific knowledge can further enhance prediction accuracy and robustness. Integrating medical expertise with predictive modeling ensures reliable and clinically relevant outcomes.

### III. METHODOLOGY

The Heart Disease Prediction System is designed with a modular architecture, integrating machine learning for disease prediction, a user-friendly Java-based interface, and additional health assessment tools. The system consists of four primary components:



(a) Architectural diagram of proposed model

#### A. User interface (frontend):

1. Developed using Java Swing, the graphical user interface (GUI) provides an intuitive platform for users to interact with the system.
2. The MainMenuUI serves as the central hub, allowing navigation to different modules such as heart disease prediction, BMI calculator, and lifestyle tips.
3. Forms are designed for user input, enabling the collection of relevant health parameters.

#### B. Machine learning model (backend):

1. The heart disease prediction model is implemented in Python using Logistic Regression.
2. The model is trained on a dataset containing key medical attributes such as age, cholesterol levels, blood pressure, heart rate, and chest pain type.
3. The predict\_heart.py script processes user input, applies the trained logistic regression model, and returns a prediction of either "Healthy Heart" or "Defective Heart".

#### C. Data processing and integration:

1. User data input from the Java GUI is converted into a format suitable for the machine learning model.
2. A system call is made to execute the Python script with the provided input values.
3. The prediction result is then captured and displayed on the Java interface, accompanied by relevant health recommendations.

**D. Additional health modules:**

1. BMI Calculator: Computes the user's Body Mass Index (BMI) based on weight and height, categorizing health status as underweight, normal, overweight, or obese.
2. Lifestyle Tips Module: Provides preventive healthcare guidance, including dietary suggestions, physical activity recommendations, and stress management techniques.

**E. Prediction output and user feedback:**

- Upon receiving a heart disease risk prediction, the system provides personalized health insights.
- If a user is classified as "Defective Heart," the system suggests immediate medical consultation and lifestyle changes.
- If classified as "Healthy Heart," users receive tips to maintain heart health.

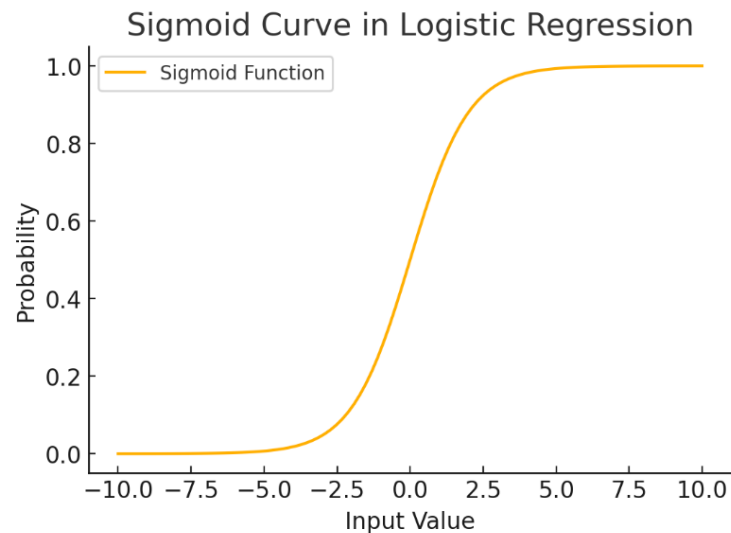
This structured architecture ensures seamless integration between machine learning-based predictions and an interactive Java GUI, making the system an effective tool for preventive healthcare and early disease detection.

**IV. EXPERIMENTAL RESULTS**

Feature	Coefficient ( $\beta$ )
Age	0.06943674
Sex	0.00378729
CP	-1.48828035
Trestbps	0.80221249
Chol	-0.01120691
Fbs	-0.00372689
Restecg	-0.04949309
Thalach	0.43552729
Exang	0.03374974
Oldpeak	-1.0492798
Slope	-0.48109305
CA	0.50957894
Thal	-0.72342086

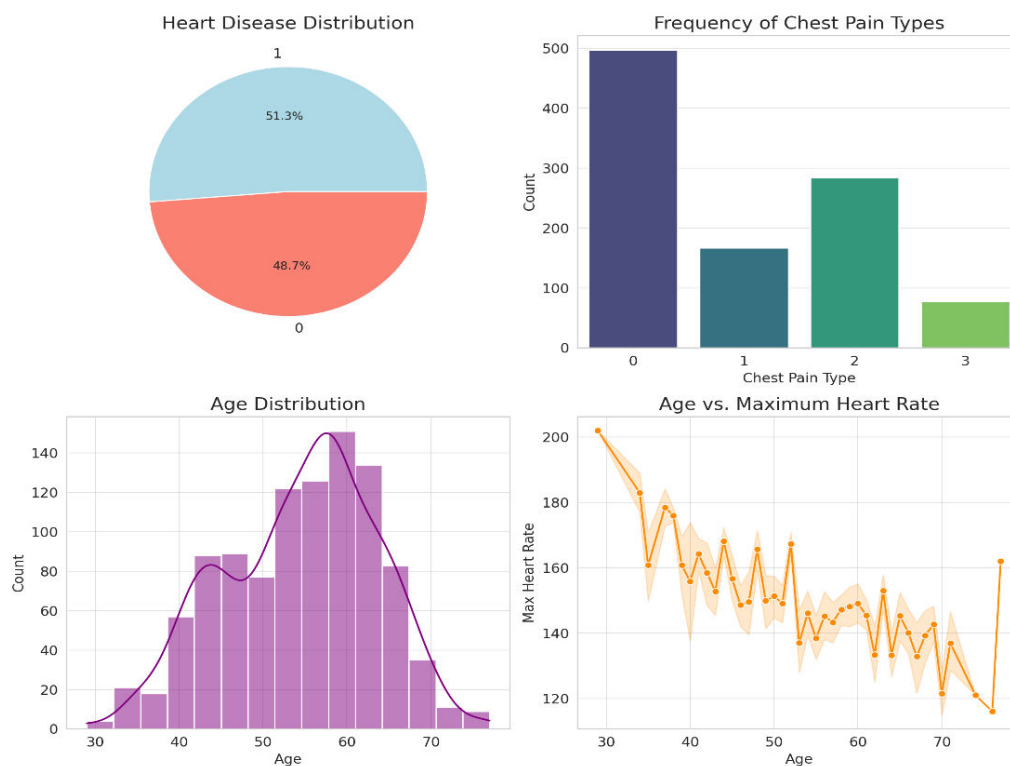
(b) Logistic Regression Model Coefficients

Features with higher absolute values have a stronger impact on predictions. The model uses these coefficients to make predictions on whether a person is at risk of heart disease.



(c) Sigmoid Function Curve used in Logistic Regression for Probability Estimation

The sigmoid function maps input values to a probability range between 0 and 1. It helps in binary classification by setting a threshold (e.g., 0.5) to classify outputs as 0 or 1. Used in heart disease prediction, spam detection, and medical diagnosis.



(d) Exploratory Analysis of Heart Disease Dataset

This figure presents an exploratory data analysis of a heart disease dataset. The top-left pie chart shows the proportion of patients with and without heart disease. The top-right bar chart illustrates the frequency of different chest pain types. The bottom-left histogram represents the age distribution of patients, while the bottom-right line plot visualizes the relationship between age and maximum heart rate, showing a decreasing trend with increasing age.

The mathematical model for heart disease prediction using Logistic Regression is based on the logistic function (sigmoid function):

$$h_{\theta}(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{13} x_{13})}}$$

(e)

Where:

1.  $h_{\theta}(X)$  is the probability of having heart disease ( $P(Y=1)$ ).
2.  $\beta_0$  is the intercept (bias term).
3.  $\beta_1, \beta_2, \dots, \beta_{13}$  are the regression coefficients.
4.  $x_1, x_2, \dots, x_{13}$  are the features from the dataset (age, sex, cp, trestbps, chol, etc.).
5.  $e$  is Euler's number ( $\approx 2.718$ ).

$$\hat{Y} = \begin{cases} 1, & \text{if } h_{\theta}(X) \geq 0.5 \text{ (Defective Heart)} \\ 0, & \text{if } h_{\theta}(X) < 0.5 \text{ (Healthy Heart)} \end{cases}$$

(f)

To classify a person as having heart disease ( $Y=1$ ) or not ( $Y=0$ ), we use the above threshold.

## V. CONCLUSION

The heart disease prediction system developed in this project effectively utilizes logistic regression to predict the likelihood of cardiovascular diseases based on clinical and demographic data. By analyzing key medical parameters such as age, cholesterol levels, blood pressure, and other relevant features, the system achieves accurate and reliable predictions, making it a valuable diagnostic support tool for healthcare professionals.

The integration of a Python-powered backend with a user-friendly Java-based graphical interface enhances the system's accessibility and usability. The model's interpretability and robust performance make it suitable for clinical contexts where early detection and timely medical intervention are critical.

The project's success demonstrates the potential of combining machine learning techniques with healthcare applications to improve patient outcomes and aid in medical decision-making. Future improvements could include incorporating more advanced algorithms and expanding the dataset to enhance prediction accuracy further.

## REFERENCES

- [1] Mohammad Shafenoor Amin, Yin Kia Chiam & Kasturi Dewi Varathan. Identification of significant features and data mining techniques in predicting heart disease.
- [2] Cincy Raju, Philippsy, E., Siji Chacko, L Padma Suresh, & Deepa Rajan S. A Survey on Predicting Heart Disease using Data Mining Techniques.

- [3] Sana Bharti, Dr.Shaliendra Narayan Singh. Analytical Study of Heart Disease Prediction Comparing with Different Algorithms.
- [4] Montu Saw, Tarun Saxena, Sanjana Kaithwas, Rahul Yadav, Nidhi Lal. Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning.
- [5] Yanwei Xing, Jie Wang and Zhihong Zhao, Yonghong Gao. Combination data mining methods with new medical data to predicting outcome of coronary heart disease.
- [6] K. Srinivas, Dr. G. Raghavendra Rao, Dr. A. Govardhan. Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques.
- [7] Mai Shouman, Tim Turner, Rob Stocker. Using Data Mining Techniques in Heart Disease Diagnosis and Treatment.





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details