

Heart Disease Prediction using Logistic Regression

Dr. A. N. Gedam¹, Pratik Bhandari², Srushti Chavre³, Revati Chougule⁴, Avanti Kadam⁵

Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India¹

Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India²

Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India³

Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India⁴

Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India⁵

ABSTRACT: Cardiovascular diseases continue to be one of the leading causes of mortality worldwide, emphasizing the importance of early detection and continuous monitoring. Early diagnosis plays a significant role in reducing health risks and improving patient outcomes. This research presents a web-based heart disease risk assessment system that integrates machine learning techniques with user-provided health data. The system utilizes a Logistic Regression model to analyze clinical parameters such as age, blood pressure, cholesterol level, heart rate, and other medical indicators to predict the likelihood of heart disease. The application provides a user-friendly interface where users can enter their health details and receive real-time prediction results. The backend system processes the data using preprocessing techniques and generates accurate predictions. The proposed system offers a cost-effective and efficient solution for early detection of heart disease. It supports healthcare professionals in decision-making and enables individuals to take preventive measures at an early stage.

KEYWORDS: Cardiovascular Diseases, Heart Disease Prediction, Machine Learning, Logistic Regression, Risk Assessment System, Clinical Data Analysis, Early Detection, Predictive Analytics

I. INTRODUCTION

Heart disease is one of the leading causes of death worldwide and poses a major challenge to public health. It often develops gradually and shows very few symptoms in its early stages, making early diagnosis difficult. As a result, many cases are detected only when the condition becomes severe, increasing health risks and mortality rates [1].

Traditional diagnostic methods depend on advanced medical equipment and expert knowledge, making them time-consuming, costly, and less accessible in rural or resource-limited areas [2]. Therefore, there is a need for efficient and accessible systems that can support early detection.

With the advancement of machine learning, it has become possible to analyze medical data and predict diseases accurately. Machine learning algorithms can identify patterns in healthcare data and assist in clinical decision-making [3]. Logistic Regression, in particular, is widely used for binary classification problems such as heart disease prediction due to its simplicity and effectiveness.

In this project, a web-based heart disease risk assessment system is developed using Logistic Regression. The system allows users to input health parameters and provides prediction results, helping in early detection and preventive healthcare.



Fig 1: AI-Powered Heart Health Prediction System

II. PROBLEM STATEMENT

Heart disease often remains undetected in its early stages due to the absence of noticeable symptoms. The lack of affordable and easily accessible diagnostic tools further leads to delayed detection and treatment, especially in resource-limited areas. As a result, many patients are diagnosed only at advanced stages, increasing health risks and treatment costs.

In addition, traditional diagnosis methods are time-consuming and depend heavily on medical experts, which may lead to human errors. Therefore, there is a need for an automated system that can accurately predict heart disease using easily available patient data. Such a system can assist healthcare professionals in early detection, improve diagnostic accuracy, and support timely medical intervention.

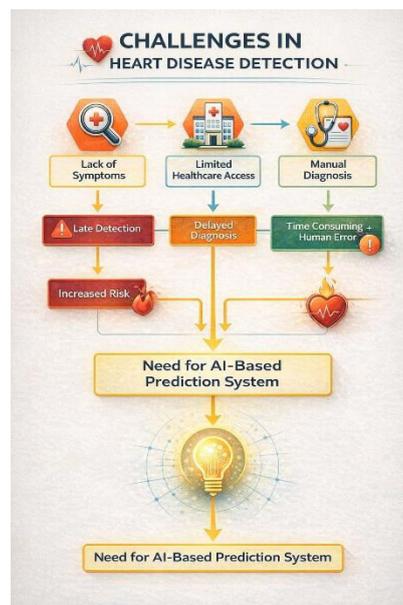


Fig 2: Problem Statement

III. LITERATURE SURVEY

Sr. No.	Publication	Inference
1	Mohammad Shafenoorr Amin, Yin Kia Chiam & Kasturi Dewi Varathan. Identification of significant features and data mining techniques in predicting heart disease.	Identifies key features and applies multiple classifiers. A hybrid model (Naive Bayes + Logistic Regression) achieved highest accuracy (87.4%).
2	Mai Shouman, Tim Turner, Rob Stocker. Using Data Mining Techniques in Heart Disease Diagnosis and Treatment.	Shows effectiveness of data mining in diagnosis. Limited focus on treatment prediction.
3	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.	Uses multiple attributes and data mining techniques for prediction. Automated systems can reduce healthcare cost.
4	Harshit Jindal, Sarthak Agarwal, Rishabh Khera, Rachna Jain and Preeti Nagrath. Heart Disease Prediction using Machine Learning Algorithms.	Uses Logistic Regression and KNN for prediction. Achieves better accuracy but limited to experimental setup.
5	Jayshril S. Sonawane, D. R. Patil. Prediction of Heart Disease using Learning Vector Quantization Algorithm.	Applies LVQ neural network for prediction. Highlights complexity of model implementation.
6	Diaa Salama Abdelminaam, Nada Mohamed, Hady Wael, Abdelrahman Khaled, Adham Moataz. MLHeartDisPrediction: Heart Disease Prediction using Machine Learning.	Uses Logistic Regression and Random Forest with high accuracy (>90%). Needs improvement for real-world use.

IV. METHODOLOGY

The proposed methodology provides a structured framework for predicting heart disease using machine learning techniques. The system integrates user-provided health parameters with historical medical data to generate accurate and reliable predictions, supporting early diagnosis and informed decision-making.

i. Data Collection and User Input

The process begins with the collection of health-related data from two primary sources: user input and existing health datasets. User input includes important medical parameters such as age, blood pressure, cholesterol level, heart rate, and other relevant clinical indicators. In addition, historical health data stored in the database is used to train the model and improve prediction accuracy.

ii. Data Processing and Feature Preparation

The collected data undergoes preprocessing to ensure quality and consistency. This step involves handling missing values, removing noise, normalizing numerical data, and selecting relevant features. Proper preprocessing ensures that the input data is clean and structured, which significantly improves the performance of the machine learning model.

iii. Application of Machine Learning Algorithm

The processed data is analyzed using the Logistic Regression algorithm to predict the presence or absence of heart disease. Logistic Regression is selected because it is well-suited for binary classification problems such as predicting the presence or absence of heart disease.

iv. Prediction Generation

The system processes the user input data using the Logistic Regression algorithm and calculates the probability of heart disease. Based on this probability, the system determines whether the user is at high risk or low risk.

v. Model Comparison and Evaluation

The performance of Logistic Regression is compared with Linear Regression using evaluation measures such as accuracy. The comparison shows that Logistic Regression provides higher accuracy and is more suitable for classification problems.

vi. Final Prediction Output

The final prediction is displayed to the user through the web interface in a simple and understandable format. The output helps users and healthcare professionals take timely decisions and adopt preventive measures for heart disease.

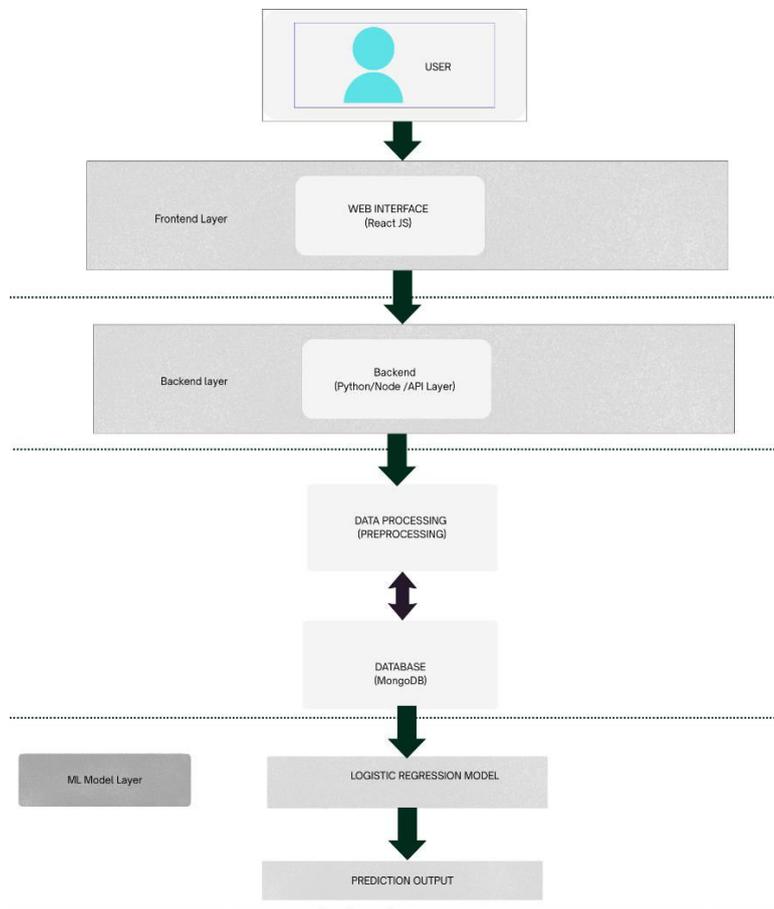


Fig 3: Proposed Methodology

V.DATASET DESCRIPTION

The system uses a structured heart disease dataset consisting of multiple patient records with various medical attributes. The dataset includes important features such as age, gender, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, ECG results, maximum heart rate, and exercise-induced angina.

The dataset also contains a target variable that indicates the presence (1) or absence (0) of heart disease. Data preprocessing techniques such as handling missing values, normalization, and feature selection are applied to improve

VI. KEY FEATURES

The Heart Disease Prediction System includes several important features designed to support early diagnosis, improve prediction accuracy, and assist in healthcare decision-making using machine learning techniques:

- **Early Disease Prediction:**

The system predicts the risk of heart disease at an early stage using patient health parameters, enabling timely medical intervention and preventive care.

- **Machine Learning–Based Analysis:**

Logistic Regression is used as the primary algorithm to analyze clinical data and classify patients as having heart disease or not, providing reliable and accurate results.

- **Model Comparison:**

The system compares Logistic Regression with Linear Regression to evaluate performance and select the most suitable algorithm for prediction.

- **Use of Clinical Health Data:**

The system utilizes important medical attributes such as age, blood pressure, cholesterol level, heart rate, and other clinical indicators to ensure accurate prediction.

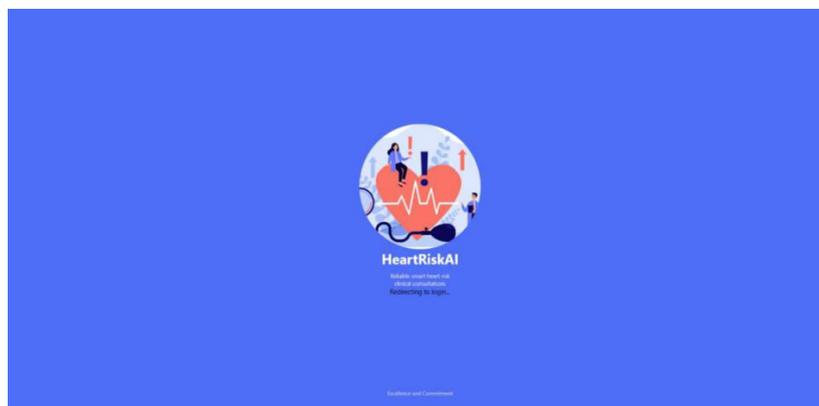
- **User-Friendly Interface:**

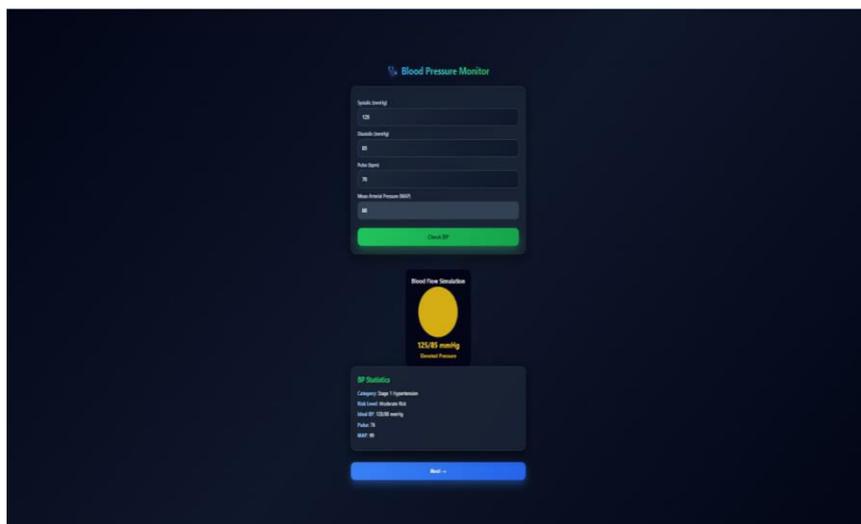
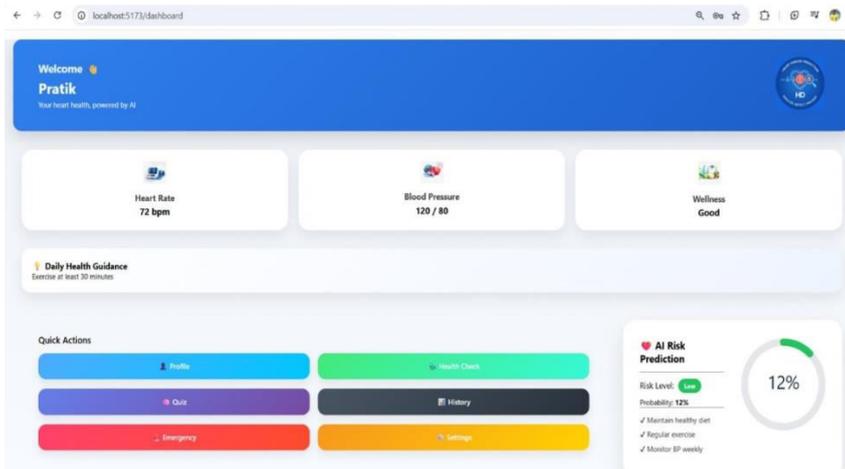
The application provides an easy-to-use interface where users can enter their health details and quickly receive prediction results.

- **Cost-Effective and Time-Efficient Diagnosis:**

The automated prediction process reduces dependency on manual diagnosis, saving time, lowering healthcare costs, and minimizing human error.

VII.SYSTEM IMPLEMENTATION





Health Summary

ECG Summary
Heart Rate: 85 BPM
RR Interval: 0.86 ms
Status: Normal

Blood Pressure
Systolic: 125 mmHg
Diastolic: 85 mmHg
Pulse: 76 BPM
MAP: 92 mmHg
Status: Stage 1 Hypertension

This summary is based on your entered values. Please continue for full health assessment and risk prediction.

[Proceed to Full Assessment](#)

Heart Health Quiz

What is your age group?

Below 30

30 - 45

46 - 60

Above 60

Question 1 of 6

Heart Health Quiz

Do you have diabetes?

No

Yes

Question 2 of 6

Heart Health Quiz

Do you smoke?

No

Occasionally

Regularly

Question 3 of 6

Heart Health Quiz

Do you feel chest pain?

Never

Sometimes

Often

Question 4 of 6

Heart Health Quiz

Do you get shortness of breath?

No

Yes

Question 5 of 6

Heart Health Quiz

How often do you exercise?

Daily

2-3 times/week

Rarely

Question 6 of 6

Quiz Result

67%

High Risk
Based on your answers, your estimated heart disease risk is **67%**.

[Continue to Quiz Result](#)

Emergency Assistance

Quick help during critical health situations

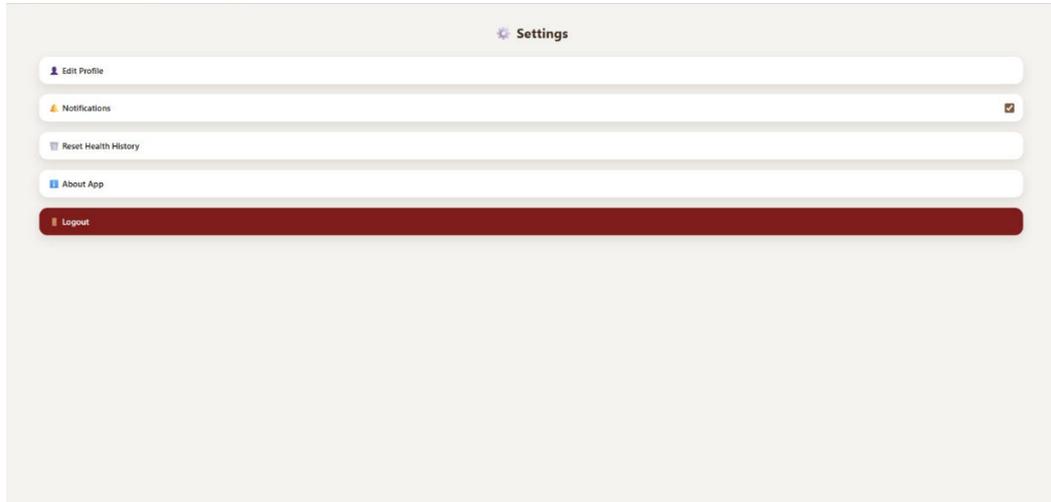
Ambulance
Call 108

National Emergency
Call 112

Police
Call 100

Fire Service
Call 101

[Back to Dashboard](#)



VIII. RESULT ANALYSIS

The performance of the system is evaluated by comparing Logistic Regression with Linear Regression for heart disease prediction. Since this is a binary classification problem, selecting the appropriate algorithm is important.

Linear Regression is not suitable for classification as it produces continuous values, which are difficult to interpret. In contrast, Logistic Regression provides probability-based outputs between 0 and 1, making it more suitable for predicting heart disease.

The comparison graph shows that Logistic Regression achieves higher accuracy and more reliable results than Linear Regression. Therefore, Logistic Regression is selected as the final algorithm for the proposed system.

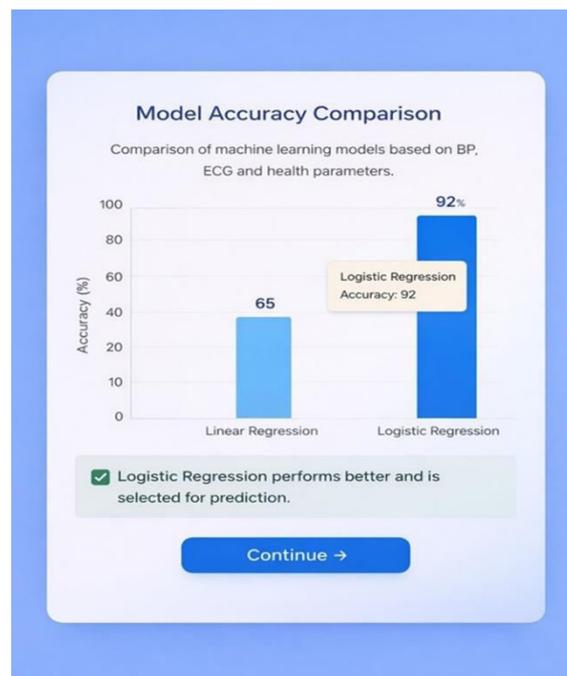


Fig 4: Accuracy Comparison of Logistic Regression and Linear Regression

Volume 15, Issue 3, March 2026**|DOI: 10.15680/IJRSET.2026.1503192|****IX. CHALLENGES AND RESEARCH GAPS**

Despite the growing use of machine learning in healthcare, several challenges still exist in heart disease prediction systems. One major issue is the limited availability of high-quality and diverse datasets, which affects the accuracy and generalization of models.

Another challenge is maintaining both accuracy and interpretability, as it is often difficult to understand how predictions are generated. Additionally, concerns related to data privacy, security, and algorithmic bias remain important.

Integrating these systems into real-world clinical environments also requires improvements in scalability and real-time performance.

X. CONCLUSION

The proposed system successfully predicts heart disease risk using Logistic Regression, providing an accurate and efficient solution for early detection. The web-based application is user-friendly and helps individuals monitor their health and take preventive measures. The system supports healthcare professionals in decision-making and reduces the limitations of manual diagnosis. Overall, it contributes to improving healthcare outcomes through timely and reliable prediction.

XI. FUTURE SCOPE

Future improvements can include the use of advanced machine learning algorithms such as Random Forest and Neural Networks to improve prediction accuracy. The system can also be enhanced by integrating real-time data from wearable devices for continuous health monitoring. Additionally, developing a mobile application and using larger, diverse datasets can further improve accessibility, performance, and real-world applicability of the system.

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



www.ijirset.com

Scan to save the contact details