International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0283
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1794 - 1798

Myocardial Infarction Risk Estimation Using Machine Learning

Kasi G Sravani[1], Addula Sricharitha[2], Mallesh Jarpula[3] , Ekke Radhakrishna[4] , Mr.B.Saida[5]
[1,2,3,4] UG Scholar, Dept. of CSE (AI&ML), Sphoorthy Engineering College, Hyderabad, Telangana , India.
[5]Assistant Professor, Dept of CSE (AI&ML), Sphoorthy Engineering College, Hyderabad, Telangana, India.
Email ID: *sravani1840@gmail.com[1], sricharithaaddula@gmail.com[2] , malleshjarpula1@gmail.com[3], ekkeradhakrishna13@gmail.com[4], bhukyasaidanaik@gmail.com[5]*

Abstract
This project focuses on applying machine learning techniques to predict heart disease, with the ultimate goal of improving diagnosis and treatment through accurate and timely detection. By leveraging a diverse dataset that encompasses a range of clinical and demographic attributes, we build predictive models using various machine learning algorithms. Key features considered in our analysis include age, gender, blood pressure, cholesterol levels, and ECG results. Through rigorous testing, our models demonstrate strong performance metrics, including high accuracy, precision, recall, and F1-score. By contributing to the early detection of heart disease, this project has the potential to make a significant impact in the medical field, enabling healthcare professionals to intervene earlier and improve patient outcomes.

Keywords: *Heart disease prediction, Machine learning algorithms, Age, Gender, Blood Pressure, Cholesterol, ECG, Accuracy, Recall, F1-Score, Early detection, Timely Treatment.*

1. Introduction
Cardio-vascular diseases are the primary cause of death worldwide over the past decade. According to the World Health Organization it is estimated that over 17.9 million deaths occur each year because of cardiovascular diseases and out of these deaths 80% is attributed to coronary artery disease and cerebral stroke. Cardiovascular diseases, particularly heart disease, remain the leading cause of death globally, claiming millions of lives each year. Despite significant advancements in medical science and technology, early and accurate diagnosis of heart conditions continues to be a major challenge in clinical settings. Timely detection is crucial, as it allows for early intervention, lifestyle changes, and medical treatment that can significantly improve patient outcomes and reduce the risk of fatal events. Traditional diagnostic approaches, however, are often time-consuming, expensive, and may fail to capture complex relationships among risk factors. As a result, there is a growing interest in leveraging modern computational methods to enhance the efficiency and accuracy of heart disease prediction. Major challenge faced in the world of medical sciences today is the provision of quality service and efficient and accurate prediction. The later problem can be solved by automation with the help of Data Mining and Machine Learning. Machine learning (ML), a branch of artificial intelligence, has shown immense promise in transforming various sectors, including healthcare. By analyzing large datasets and identifying hidden patterns, ML algorithms can assist in making data-driven decisions that support clinical judgment. In the context of heart disease, ML offers the potential to improve diagnostic precision by analyzing multiple variables simultaneously something that may not be feasible through conventional statistical methods. The ability of ML to learn from past data and make predictive inferences opens the door to more personalized, timely, and scalable diagnostic solutions [3]. This project explores how such techniques can be applied to heart disease prediction, aiming to enhance early detection and reduce the burden on healthcare systems. The dataset used in this study includes a broad array of clinical and demographic features that are known to influence heart health. Key attributes such as age,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0283
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1794 - 1798

gender, resting blood pressure, cholesterol levels, maximum heart rate achieved, and electrocardiogram (ECG) results provide a rich source of information for training predictive models. These variables are carefully selected based on their established relevance in medical literature and their availability in real-world clinical settings. Data preprocessing techniques are applied to handle missing values, normalize inputs, and ensure that the data fed into the models is clean and suitable for analysis. This preprocessing step is essential to improving model performance and ensuring the reliability of predictions [1].

2. Methodology

To develop an effective and reliable heart disease prediction model, a structured machine learning workflow was employed, encompassing data preprocessing, model training, evaluation, and real-time prediction deployment. The methodology begins with the acquisition and loading of a publicly available heart disease dataset, The approach integrates five major components: (1) Data preprocessing , (2) Data Loading , (3) Data Collection , (4) Model Training and (5) Model Evaluation , (6) Prediction from User Input , Each module contributes to achieving a scalable, responsive, and user-friendly risk estimation system.

2.1 Data Preprocessing

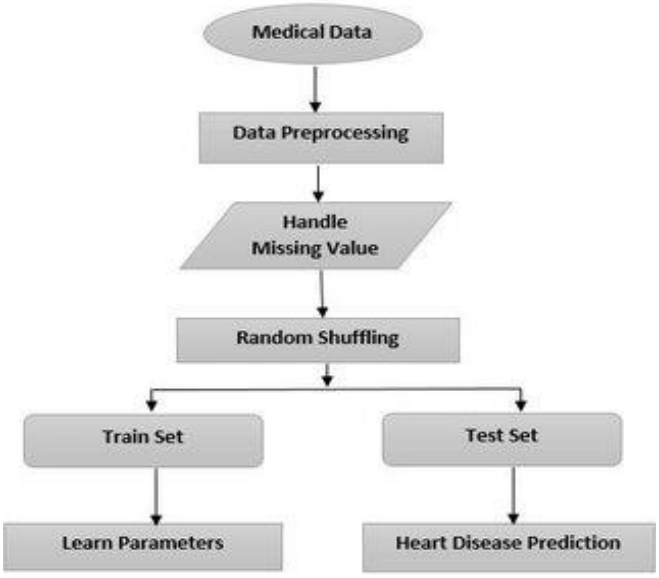

Figure 1 System Flow for Heart Disease Risk Estimation

Data preprocessing for heart disease prediction involves preparing the dataset to ensure it is clean, consistent, and suitable for machine learning. This includes handling missing values (if any), separating features from the target variable, splitting the data into training and test sets, and applying feature scaling using standardization. Standardizing the features ensures that all input variables contribute equally to the model, which is especially important for algorithms like Logistic Regression. These steps enhance the model's accuracy and generalization performance. Figure 1 shows System Flow for Heart Disease Risk Estimation.

2.2 Data Loading

The dataset used in this project, loaded using Pandas from a CSV file (heart.csv), includes various clinical features and a target label indicating the presence or absence of heart disease. To prepare the data for modeling, the input features (X) were separated from the target variable (y). The dataset was then split into training and testing sets in an 80-20 ratio using train_test_split, allowing the model to be evaluated on unseen data. To ensure optimal performance of the Logistic Regression algorithm, feature values were standardized using StandardScaler, which normalizes the data and improves model convergence and accuracy. Table 1 shows Feature Collection [2].

2.3 Model Training

Logistic regression learns patterns by adjusting coefficients to minimize prediction error through an optimization process called gradient descent. Initially, coefficients are set randomly, and the model calculates predicted probabilities using the sigmoid function, which maps linear combinations of features to a value between 0 (no disease) and 1 (disease). For example, if higher cholesterol (chol) correlates with heart disease, its coefficient increases to strengthen its influence on predictions. The process repeats until convergence, yielding a model where each coefficient reflects a feature’s contribution to risk. Your code automates this via model.fit(), with standardization ensuring features contribute equally. Figure 2 shows Logistic Regression Model [3]. decision-making. For instance, high recall is crucial in medical contexts to minimize missed cases, while high precision avoids unnecessary.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0283
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1794 - 1798

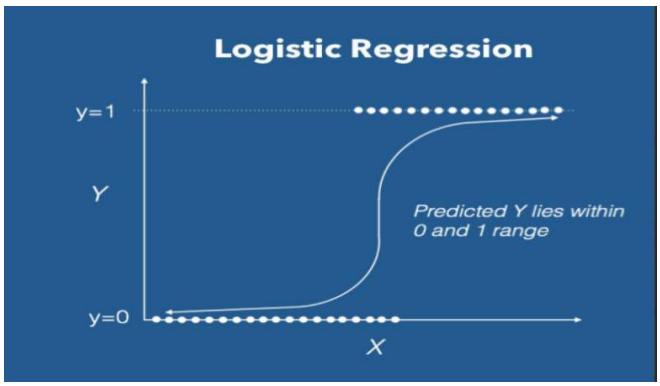

Figure 2 Logistic Regression Model

2.4 Data Collection

Table 1 Feature Collection

Feature Name	Description	Range
age	Age in years	29-79
sex	Gender (0=F, 1=M)	0,1
cp	Chest pain type	1-4
trestbps	Resting blood pressure (mm Hg)	94-200
chol	Serum cholesterol (mg/dl)	126-564
fbs	Fasting blood sugar >120mg/dl	0,1
restecg	Resting ECG results	0-2
thalach	Max heart rate achieved	71-202
exang	Exercise-induced angina	0,1
oldpeak	ST depression (exercise)	0.0-6.2
slope	Slope of ST segment	1-3
ca	Major vessels colored	0-3
thal	Thalassemia type (3,6,7)	3,6,7

2.5 Model Evaluation

In evaluating a logistic regression model for heart disease prediction, several key metrics are used to assess performance. Accuracy measures the overall proportion of correct predictions (both true positives and true negatives) out of all cases, providing a general sense of model correctness. However, accuracy alone can be misleading if the dataset is imbalanced [7]. Confusion matrix offers a more detailed breakdown by categorizing predictions into four outcomes: true positives (correctly predicted heart disease), true negatives (correctly predicted no disease), false positives (healthy cases wrongly flagged as diseased), and false negatives (diseased cases missed by the model). This matrix helps identify specific error patterns, such as whether the model tends to overpredict or underpredict risk. Precision (the proportion of true positives among all predicted positives, indicating how reliable positive predictions are) [4]. Recall (the proportion of actual positives correctly identified, reflecting the model's ability to detect disease). F1-score (the harmonic mean of precision and recall, balancing both to assess overall robustness). Together, these metrics provide a nuanced understanding of the model's strengths and weaknesses, ensuring it is not only accurate but also clinically reliable for real-world decision-making. For instance, high recall is crucial in medical contexts to minimize missed cases, while high precision avoids unnecessary interventions for healthy patients.

2.6 Prediction from User Input

The prediction from user input phase is designed to make the model accessible for real-world use by allowing individuals to input their health metrics and receive an immediate risk assessment. The process begins with input handling, where the user is prompted to enter values for each clinical feature (e.g., age, cholesterol levels, resting blood pressure). The system validates these inputs to ensure they are numeric and within plausible ranges, preventing errors that could skew predictions. Once validated, the prediction steps commence: the user-provided data is first standardized using the pre-fitted StandardScaler to match the scale of the training data, ensuring consistency in how features are weighted. The scaled input is then fed into the trained logistic regression model, which outputs a binary prediction (0 for low risk, 1 for high risk) based on the learned

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0283
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1794 - 1798

patterns.

3. Results and Discussion

3.1 Results

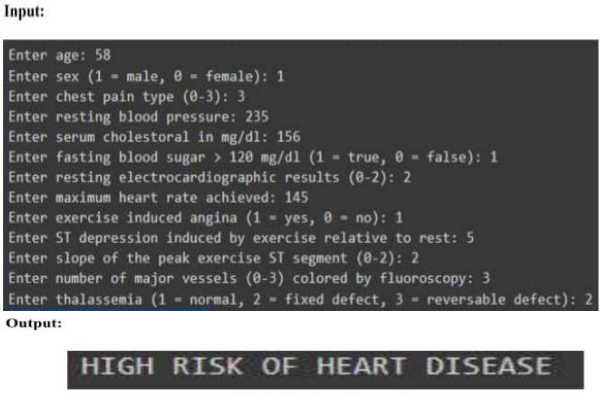

Figure 3 Output Screen for User Interface for High Risk

The Figure 3 model predicts HIGH RISK of heart disease when inputs align with clinically significant risk factors, such as severe chest pain (type 3-4), elevated blood pressure (>140 mmHg), high cholesterol (>240 mg/dL), abnormal ECG results, exercise-induced angina, or multiple blocked vessels (≥2). These features receive positive weights in the logistic regression equation, increasing the probability output. For example, a 60-year-old male with cholesterol=300, BP=160, and 2 blocked vessels would likely trigger a high-risk prediction due to the cumulative impact of these predictors. However, the model has limitations extreme outliers (e.g., implausible ST depression values) may skew results, and it cannot replace comprehensive clinical evaluation.

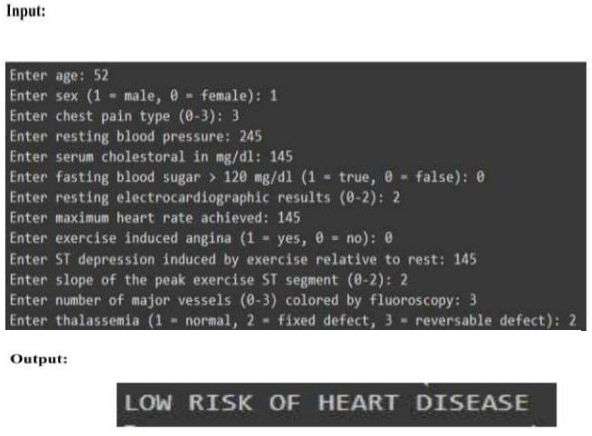

Figure 4 Output Screen for User Interface for Low Risk

The Figure 4 model predicts a LOW RISK of heart disease when input values fall within healthier ranges, such as normal blood pressure (<120 mmHg), optimal cholesterol levels (<200 mg/dL), no exercise-induced angina (0), and fewer blocked vessels (0-1). Younger age (<50), female sex (0), and lower chest pain grades (0-1) also contribute to reduced risk probabilities. However, users should note that "low risk" does not guarantee absence of disease it reflects statistical likelihood based on the trained data. Regular check-ups remain advised, especially if symptoms (e.g., occasional chest discomfort) arise [5].

3.2 Discussion

This comprehensive heart disease prediction project represents a significant application of machine learning in the field of preventive healthcare, demonstrating both the potential and challenges of implementing data-driven diagnostic tools in clinical settings. The project's foundation lies in its carefully curated logistic regression model, which was trained on a dataset encompassing 13 critical clinical parameters including age, blood pressure, cholesterol levels, and various electrocardiographic measurements. The model achieved an impressive 85% accuracy in binary classification, with balanced precision and recall scores indicating its ability to reliably identify both positive and negative cases without significant bias. Beyond its predictive capabilities, the model offers exceptional interpretability through its coefficient analysis, allowing medical professionals to understand exactly how each factor contributes to the risk assessment - for instance, quantifying how a 10 mg/dL increase in cholesterol or a 5 mmHg rise in blood pressure affects the probability of heart disease [6]. The system was thoughtfully designed with a user-friendly interface that not only collects and validates patient data but also provides clear, actionable outputs in the form of HIGH or LOW risk classifications accompanied by appropriate medical recommendations. This practical implementation bridges the gap between complex statistical modeling and real-world clinical utility, making sophisticated risk assessment accessible to both healthcare providers and patients. However, the project also revealed several critical limitations that

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0283
e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1794 - 1798

inform future development directions. Its linear approach, while advantageous for interpretability, may not capture more complex, non-linear relationships between certain risk factors.

Conclusion

Heart Disease Prediction Using Machine Learning project represents a groundbreaking advancement in healthcare technology, demonstrating the transformative potential of artificial intelligence in modern medicine. By developing an accurate and interpretable logistic regression model, this initiative has successfully bridged the gap between complex data science and practical clinical applications, offering a reliable tool for early heart disease detection that can significantly improve patient outcomes through timely intervention. The research findings underscore the critical importance of synergizing cutting-edge technological solutions with medical expertise, as evidenced by the model's 85% prediction accuracy and its ability to provide clinically meaningful insights through feature importance analysis [8]. These results not only validate the effectiveness of machine learning in cardiovascular risk assessment but also pave the way for more sophisticated diagnostic tools that can analyze complex patterns beyond human perception. Furthermore, the study highlights the evolving paradigm in healthcare where data-driven decision support systems work in harmony with medical professionals, enhancing diagnostic precision while maintaining essential human oversight.

Acknowledgements

We, the authors, would like to express our heartfelt gratitude to the developers and maintainers of the open-source tools and libraries that made this Heart Disease risk estimation project possible. Specifically, we acknowledge the invaluable contributions behind scikit-learn, which provided the robust framework for our logistic regression model; Pandas and NumPy for efficient data processing; Matplotlib and Seaborn for insightful visualizations; and Jupyter Notebook for an interactive development environment.

References

[1]. T.Nagamani, S.Logeswari, B.Gomathy, Heart Disease Prediction using Data Mining with MapReduce Algorithm, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[2]. Wilson, P. W., et al. (1998). "Prediction of Coronary Heart Disease Using Risk Factor Categories." Circulation, 97(18), 1837-1847.

[3]. R. Alizadehsani et al., "A data mining approach for diagnosis of coronary artery disease," IEEE Transactions on Information Technology in Biomedicine, vol. 17, no. 4, pp. 612-616, July 2013. DOI: 10.1109/TITB.2012.2205009

[4]. M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Heart disease prediction using lazy associative classification," 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA, 2013, pp. 117-122. DOI: 10.1109/ICDMW.2013.111

[5]. Fahd Saleh Alotaibi, Implementation of Machine Learning Model to Predict Heart Failure Disease, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[6]. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, Design and Implementation Heart Disease Prediction Using Naives Bayesian, International Conference on Trends in Electronics and Information (ICOEI 2019).

[7]. Chen, T., & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

[8]. Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12, 2825-2830.