An Interpretable and Explainable Logistic Regression Based Model for Heart Disease Prediction

Disha Loya[1]

[1]Department of Computer Science and Engineering

Publication Date: 2026/02/14

Abstract: Cardiovascular disorders remain a dominant contributor to global mortality, emphasizing the importance of timely risk assessment and preventive intervention. The expansion of electronic medical data combined with advancements in computational intelligence has enabled the development of predictive systems that assist clinical evaluation. This work presents a clinical decision support framework that applies logistic regression to estimate the probability of heart disease using patient-specific clinical indicators. Due to its statistical foundation, low computational demand, and interpretability, logistic regression is particularly suitable for medical environments where explainability is essential. Experimental evaluation demonstrates that the model provides consistent predictive performance and identifies influential risk variables, supporting its applicability in early-stage cardiovascular risk screening.

Keywords: Binary Classification, Machine Learning Techniques, Heart Disease Prediction, Logistic Regression, Healthcare Analytics, Medical Data Analysis, Supervised Learning.

I. INTRODUCTION

Cardiovascular conditions pose a persistent public health burden and are responsible for a substantial fraction of preventable deaths worldwide. Detecting risk at an early stage plays a critical role in minimizing complications and improving treatment outcomes. Conventional diagnostic procedures often depend on multiple laboratory tests and specialist evaluations, which may be expensive and time-intensive. These challenges create demand for computational systems capable of assisting clinicians in risk estimation using available patient information.

With the rapid accumulation of digital medical records, data-driven approaches have become increasingly relevant in healthcare analytics. Machine learning techniques allow patterns to be extracted from clinical attributes and translated into predictive insights. Logistic regression, a probabilistic classification method, offers a balance between mathematical rigor and interpretability. This study leverages logistic regression to construct a predictive model that supports clinical assessment of heart disease risk.

II. LITERATURE REVIEW

The application of machine learning to cardiovascular diagnosis has attracted growing research attention due to improved data availability. Earlier investigations frequently employed statistical modeling techniques, particularly logistic regression, owing to their transparency and well-established theoretical basis. These models provided insight into how individual physiological indicators relate to disease likelihood.

More recent studies have explored complex architectures such as neural networks and ensemble approaches. However, despite the rise of advanced algorithms, logistic regression continues to serve as a dependable baseline because of its robustness, ease of interpretation, and suitability for clinical environments. Research consistently highlights variables including age, blood pressure, cholesterol concentration, heart rate, and chest pain characteristics as influential predictors of cardiac conditions.

III. METHODOLOGY

The development of the proposed prediction system followed a sequential analytical workflow designed to ensure reliability and reproducibility.

➤ *Data Acquisition*

Patient records were obtained from an open-access cardiovascular dataset containing demographic and diagnostic attributes. Variables include age, sex, chest pain classification, resting blood pressure, serum cholesterol, fasting blood sugar status, electrocardiographic observations, peak heart rate, and exercise-related angina. A binary target variable denotes disease presence.

➤ *Data Preparation*

Data refinement procedures were applied to improve model input quality. These steps involved addressing incomplete entries, eliminating redundant samples, encoding categorical variables numerically, and applying normalization to prevent scale imbalance among attributes.

➤ *Exploratory Data Analysis (EDA)*

Statistical summaries and graphical techniques were used to observe distributions and relationships among features. Correlation analysis assisted in understanding interdependencies within the dataset.

➤ *Model Selection*

Logistic regression was selected for its probabilistic interpretation of class membership. Model parameters were estimated through maximum likelihood optimization, enabling disease probability.

➤ *Performance Assessment*

Model effectiveness was measured using classification metrics including accuracy, precision, recall, F1-score, and confusion matrix evaluation. Emphasis was placed on recall to reduce undetected positive cases, a critical requirement in medical screening systems.

➤ *Risk Factor Interpretation*

The trained model is evaluated using standard performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Accuracy measures overall prediction correctness, while precision and recall assess the model's ability to correctly identify patients with heart disease. Reducing false negatives is particularly important in medical diagnosis to avoid missed detections quantification of how each feature Influences.

➤ *Feature Importance Analysis*

The magnitude and direction of logistic regression coefficients are used to analyze feature importance. The most significant factors influencing the prediction of heart disease, such as age, blood pressure, cholesterol level, and type of chest pain, are identified by this analysis.
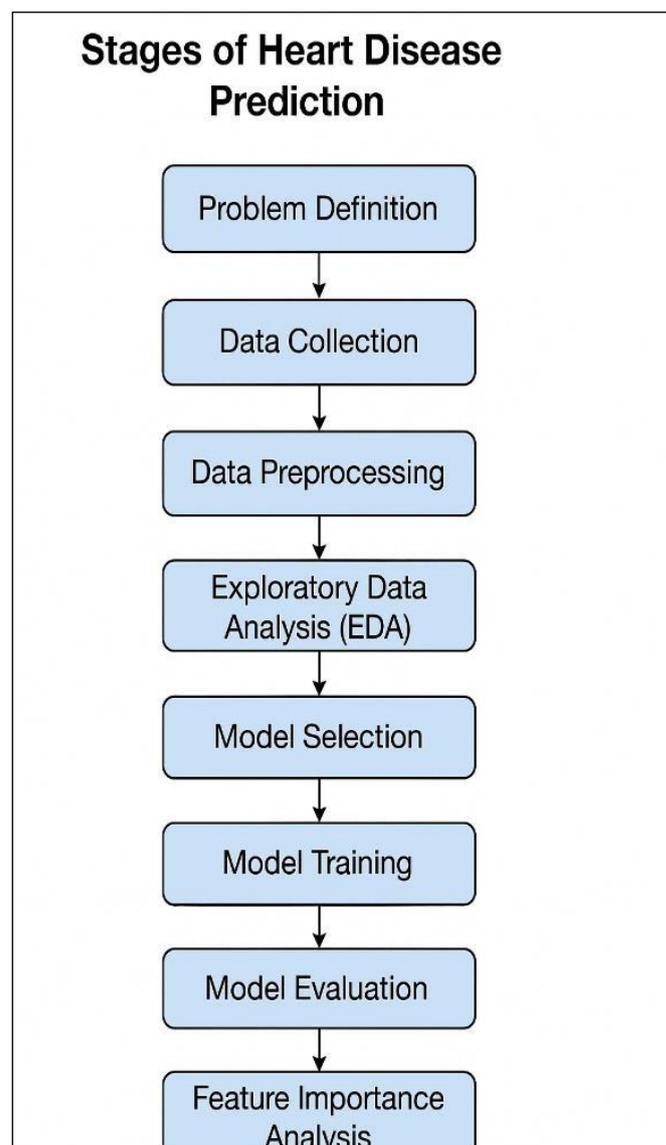

Fig 1: Stages of Predicting Heart Disease.

IV. KEY STRATEGIES

Logistic regression demonstrated strong predictive capability for heart disease by effectively modeling binary outcomes and providing clear.

Logistic regression demonstrated strong predictive capability for heart disease by effectively modeling binary outcomes and providing clear interpretability. The model successfully captured relationships between critical medical attributes such as age, blood pressure, cholesterol level, chest pain type, and heart rate with the likelihood of heart disease.

Evaluation results based on accuracy, precision, recall, F1-score, and confusion matrix confirm that the model achieved dependable classification performance with minimal error rates. A key contribution of this study is the feature importance analysis, which highlights major risk indicators and supports informed clinical decision-making.

V. CONCLUSION

These findings demonstrate that logistic regression is a practical and efficient approach for heart disease prediction. The model successfully captured relationships between critical medical attributes such as age, blood pressure, cholesterol level, chest pain type, and heart rate with the likelihood of heart disease.

Evaluation results based on accuracy, precision, recall, F1-score, and confusion matrix confirm that the model achieved dependable classification performance with minimal error rates. A key contribution of this study is the feature importance analysis, which highlights major risk indicators and supports informed clinical decision-making.

REFERENCES

[1]. Abdul wahab Ali Almazroi, Eman A. Aldhahri, Saba Bashir, Sufyan Ashfaq (2023) A clinical decision support system for heart disease prediction using deep learning .

[2]. Inam Ullah , Tariq Inayat , Naeem Ullah, Faris Alzahrani ,Muhammad Ijaz Khan (2023), A clinical decision support system for heart disease diagnosis and prediction by machine learning algorithms: A systematic literature review.

[3]. Shuai Niu , Jing Ma , Xian Yang(2024) Modelling Patient Longitudinal Data for clinical Decision Support A case Study on emerging AI health care Technologies

[4]. Siva Teja Reddy(2025) Kandula Comparsion and performance Assessment of intelligent ML Models for forecasting CardioVascular .

[5]. Mohamed S . Elgendy , Hossam El – Din Moustafa, Hala B . Nafea(2025) Utilizing voting classifiers for enhanced analysis and diagnosis of cardiac conditions.

[6]. Katy E Trinkley, Steven T Simon, Michael(2025) A Rosenberg Impact of an Alert – Based Inpatient Clinical Decision Support Tool to prevent Drug Induced long QT syndrome: Large- Scale, System – Wide observational study.

[7]. Abhinav Sharma , Sanjay Dhanka , Ajat Shatru Arora(2025), A systematic review on machine learning intelligent System for Heart Disease Diagnosis

[8]. Advancing Clinical Decision(2025) – Making using Artificial Intelligence and Machine Learning for Accurate Disease Diagnosis.

[9]. Wivica Kauppi, Henrik Imberg, Carl Magnusson (2025), advancing machine learning based decision support tool for pre hospital assessment of dyspnea by emergency medical service clincians: a retrospective observational study.

[10]. Enhancing heart disease diagnosis(2025) Leveraging classification and ensemble machine learning techniques in health care decision making.

[11]. Muhammet Alkan , Idris Zakariyya , Samuel Leighton,n Kaushik Bhargav Sivangi(2025), Artificial Intelligence – Driven Clinical Decision Support System.

[12]. Macolm Risk, Jeannie Callum , Kevin Trentino , Kevin Murray, Lil Zhao, Xu shi , Amol Verma , Fahad Razak, Sheharyar Raza(2025) Tranfusion Probability as an alterative measure of lab guided medical Decision making.

[13]. Haowei Yang, Ziyu Shen , Junli Shao, Luyao Men , Xinyue Han , jing Dong (2025) LLM – Augumented Sympton Analysis for cardiovascular Disease Risk prediction : A Clinical NLP.

[14]. Idris Zubairu Sadiq , Fatima Sadiq Abuakar , MuhammadAuwalSaliu,BabangidaSansui(2025),Machine learning algorithms for predictive modeling of dyslipidemia associated cardiovascular disease risk in Pregnancy : a comparsion of boosting , random forest , decision tree regression.

[15]. Arhath Kumar , Prathiksha Raj , Mark Maben(2025) Heart Disease prediction Using Machine learning algorithm

[16]. Mohammad Nauman , Ahmad S. Almadhor , Mohammed Albekaiiri , Ali R . Ansari , Muhammad A. B. Fayyaz, Raheel Nawaz(2025), The Role of Big DataAnalytics in Revolutionizing Diabetes Managnement HealthCare Decision Making.

[17]. Mohammad Dehghani Saryazdi, Ali Mostafaeipour(2025) Identification and validation of key predictive factors for heart attack diagnosis using machine learning and fuzzy clustering.

[18]. P Rajendra Kumar , Prasun Chakrabarti , Julika Chakrabarti, Bhuvan Unhelkar , Martin Margala(2025)Heart disease prediction using spark

[19]. architecture with fused feature set and hybrid Squeezenet – Linknet model.

[20]. Shahid Mohammad Ganie , Pijush Kanti Dutta Pramanik , Zhongming Zhao(2025), Ensemble learning with explainable AI for improved heart disease prediction based on multiple datasets.

[21]. S.Ramchandra Reddy & G. Vishnu Murthy(2025) Cardio vascular Disease prediction Using particle Swarm optimization and neural Network based an integrated Framework