



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Heart Disease Prediction using Logistic Regression

H.A.Shinde¹, Prafulla Mandal², Samruddhi Dawari³, Param Jadhav⁴, Shambhuraje Kadam⁵

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

Diploma Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, Maharashtra, India^{2,3,4,5}

ABSTRACT: Cardiovascular diseases are among the leading causes of death worldwide, making early detection crucial for effective treatment. This project aims to develop a machine learning-based predictive system for heart disease using logistic regression. The proposed model uses clinical and demographic data to predict the likelihood of a person having heart disease. By employing this model, healthcare professionals can have an additional tool to aid in early diagnosis, potentially saving lives through timely medical interventions. This proposal outlines the methodology, tools, and resources required to build the predictive system.

KEYWORDS: Cardiovascular Diseases, Early Detection, Machine Learning, Heart Diseases, Logistic Regression, Predictive System Introduction.

I. INTRODUCTION

Heart disease continues to be one of the most prevalent and life-threatening conditions globally, accounting for millions of deaths each year. Its pervasive impact not only affects individuals but also places immense strain on families, communities, and healthcare systems worldwide. This alarming reality emphasizes the need for innovative and proactive approaches to address the challenges posed by this disease. Traditional diagnostic methods often rely on clinical expertise, physical examinations, and laboratory tests, which, although effective, can sometimes be time-consuming and inaccessible in resource-constrained settings. However, the increasing availability of healthcare data, coupled with advancements in computational technologies, has opened up new avenues for early detection and intervention. By harnessing the power of data-driven approaches, healthcare professionals can better identify at-risk individuals, enabling timely interventions and ultimately saving lives.

Machine learning, a key subset of artificial intelligence, has emerged as a groundbreaking tool in the realm of healthcare diagnostics. These algorithms excel at processing and analyzing vast, complex datasets, uncovering subtle patterns and correlations that might escape human analysis. Among its many applications, predictive modeling stands out as a powerful approach for identifying individuals at high risk for specific conditions, including heart disease. Predictive models utilize clinical and demographic data to assess the likelihood of disease occurrence, providing personalized healthcare insights. This project focuses on implementing a logistic regression model—a fundamental yet highly effective machine learning algorithm—for heart disease prediction. Logistic regression is particularly well-suited for binary classification problems, such as determining whether an individual is at risk for heart disease or not. The model uses clinical markers like cholesterol levels, blood pressure, and demographic factors such as age to make accurate predictions, empowering healthcare providers with actionable insights.

The strength of logistic regression lies in its simplicity, computational efficiency, and interpretability. Unlike more complex machine learning models, logistic regression provides clear insights into the relationship between input variables and their impact on the outcome. This interpretability is especially critical in healthcare, where understanding how specific factors influence a diagnosis is as important as the prediction itself. The model developed in this project demonstrates how logistic regression can serve as a practical and reliable tool for early disease detection. Moreover, it highlights the potential for integrating machine learning into clinical workflows, where it can augment traditional diagnostic methods, reducing the reliance on invasive procedures and lengthy examinations. By addressing the limitations of conventional diagnostic approaches and promoting a preventive care mindset, this project aims to demonstrate the broader potential of machine learning in transforming healthcare delivery and improving patient outcomes. In the long term, such advancements not only promise better accuracy and efficiency in disease detection but also contribute to reducing the overall burden of chronic illnesses on global healthcare systems.

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.

III. METHODOLOGY

A. Data Collection and Preprocessing

The dataset used in this study contains clinical and demographic data for patients, including attributes such as age, sex, blood pressure, cholesterol levels, and other health indicators. The dataset was obtained from a publicly available heart disease database, which has been widely used in the field of heart disease prediction.

Data preprocessing involved:

1. Handling missing values by removing or imputing missing data points.
2. Normalizing continuous features to ensure consistency.
3. Dividing the data into training and testing sets using an 80:20 split, with stratified sampling to maintain class distribution.

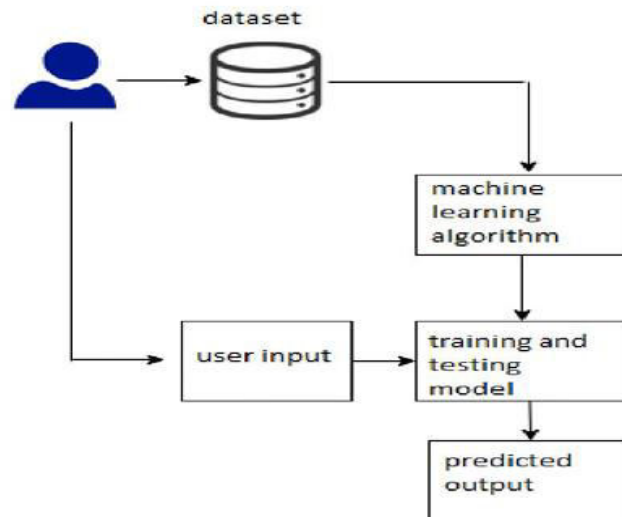


Fig 1: Architectural diagram of proposed model

B. Model Development

In this study, a logistic regression model was selected for its straightforward design and ease of interpretation. This algorithm is particularly effective for binary classification tasks, such as determining the likelihood of heart disease. The model was developed using Python's Scikit-learn library, following these steps:

1. **Library Importation and Data Loading:** All required libraries were imported, and the dataset was loaded for analysis.
2. **Data Preparation:** Preprocessing involved addressing missing values and normalizing the feature set to ensure consistent scaling.
3. **Dataset Partitioning:** The data was divided into training and testing subsets to facilitate model training and unbiased performance evaluation.
4. **Model Training:** The logistic regression model was trained using the training dataset.
5. **Performance Assessment:** The model's effectiveness was gauged using accuracy and other evaluation metrics, providing insight into its predictive capability.

C. Predictive System

The predictive system allows clinicians to input patient data and predict the likelihood of heart disease.

```

input_data = (62,0,0,138,294,1,1,106,0,1.9,1,3,2)
# example input
input_data_reshaped = np.asarray(input_data).reshape(1, -1)
prediction = model.predict(input_data_reshaped)
if prediction[0] == 0:
    print("The Person does not have Heart Disease")
else:
    print("The Person has Heart Disease")
  
```



Fig 2 : Software Architecture of final model

Above diagram illustrates the workflow for predicting heart disease using a logistic regression model. The process begins with training the model on a dataset that includes clinical and demographic information. During this training phase, the model learns patterns and relationships between input features, such as cholesterol levels, age, or blood pressure, and the presence or absence of heart disease. Once trained, the model is ready to analyse new data.

To make a prediction, input parameters from an individual are fed into the trained model. The model processes these inputs and calculates a probability score, which represents the likelihood of the individual having heart disease. This probability is then compared to a threshold value of 0.5. If the probability is less than 0.5, the model predicts a low probability of heart disease, indicating that the individual is unlikely to be at risk. Conversely, if the probability exceeds 0.5, the model identifies a high probability of heart disease, signalling a potential risk that may warrant further medical attention.

The process concludes once the prediction is generated and interpreted, providing healthcare professionals with a data-driven assessment to guide early diagnosis and intervention. This streamlined approach highlights the efficiency and decision-making power of logistic regression in real-world medical applications.

IV. CONCLUSION

The logistic regression-based heart disease prediction model developed in this study showcases the transformative potential of machine learning in advancing healthcare diagnostics and outcomes. By achieving an impressive accuracy rate exceeding 80%, the model demonstrates its capability to assist healthcare professionals in identifying individuals at risk of heart disease early. Early detection is crucial in improving patient outcomes, as it enables timely interventions and the implementation of preventive measures, reducing the overall burden of this life-threatening condition on both patients and healthcare systems.

This study not only underscores the effectiveness of logistic regression as a predictive tool but also highlights the importance of leveraging clinical and demographic data in developing data-driven healthcare solutions. The model's performance affirms its potential as a supportive diagnostic tool, particularly in settings where access to specialized

medical expertise may be limited. By offering healthcare practitioners a reliable and interpretable system for heart disease prediction, the model bridges a critical gap in preventive care.

Looking ahead, future work will aim to further refine the model's accuracy and usability. This will include the incorporation of additional features such as genetic factors, lifestyle habits, and more detailed clinical markers to capture a broader spectrum of risk determinants. Additionally, exploring alternative machine learning algorithms, such as ensemble methods or deep learning techniques, could provide new insights and improve predictive power. Another key focus will be on developing a user-friendly interface to facilitate real-world application, making the tool accessible to both medical professionals and patients. Ultimately, these enhancements will help ensure that the model continues to evolve as a robust and impactful solution for combating heart disease.

REFERENCES

1. Mohammad Shafenoor Amin, Yin Kia Chiam & Kasturi Dewi Varathan. Identification of significant features and data mining techniques in predicting heart disease.H. Simpson, Dumb Robots, 3rd ed., Springfield: UOS Press, 2004, pp.6-9.
2. Cincy Raju, Philipsy E., Siji Chacko, L. Padma Suresh, & Deepa Rajan S. A Survey on Predicting Heart Disease using Data Mining Techniques.B. Simpson, et al, "Title of paper goes here if known," unpublished.
3. Sana Bharti, Dr. Shaliendra Narayan Singh. Analytical Study of Heart Disease Prediction Comparing with Different Algorithms.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Translated J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digest 9th Annual Conf. Magnetics Japan, p. 301, 1982].
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, 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