



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 4, April 2025

Early Detection of Heart Diseases using Logistic Regression Techniques

K.V. Prasada Reddy, Y.V.L. Susritha Kumari, G. Dhanush Manikanta, G. Bharath, Y. Sai, P. Tejaswini

Associate Professor & HOD, Department of AI&DS, SVCN Engineering College, North Rajupalem, Nellore, AP, India

U.G. Student, Department of AI&DS, SVCN Engineering College, North Rajupalem, Nellore, AP, India

U.G. Student, Department of AI&DS, SVCN Engineering College, North Rajupalem, Nellore, AP, India

U.G. Student, Department of AI&DS, SVCN Engineering College, North Rajupalem, Nellore, AP, India

U.G. Student, Department of AI&DS, SVCN Engineering College, North Rajupalem, Nellore, AP, India

U.G. Student, Department of AI&DS, SVCN Engineering College, North Rajupalem, Nellore, AP, India

ABSTRACT: Heart disease continues to be one of the leading causes of mortality worldwide, making early detection and prevention critically important in reducing fatal outcomes. This study focuses on the development of a machine learning-based predictive model utilizing logistic regression for the early detection of heart disease. The model analyzes essential health parameters such as age, blood pressure, cholesterol levels, body mass index (BMI), blood sugar, and lifestyle habits including smoking, alcohol consumption, and physical activity. These features are crucial indicators of cardiovascular health and play a vital role in predicting an individual's risk level. The dataset employed in this research is sourced from publicly available medical repositories. Before model training, the data undergoes several preprocessing steps, including handling missing values, performing feature selection to retain the most significant attributes, and applying normalization techniques to standardize the input values. Logistic regression is selected due to its robustness, simplicity, and high interpretability in binary classification problems like heart disease prediction. The model's performance is thoroughly evaluated using a variety of metrics such as accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score. These metrics provide a comprehensive understanding of the model's effectiveness and reliability in real-world scenarios. To enhance the practical application of the model, a user-friendly web application is developed using Flask. This application allows users to input their health parameters and receive real-time predictions regarding their risk of developing heart disease. The platform is designed to be intuitive, ensuring accessibility even for individuals without technical backgrounds.

KEYWORDS: Heart Disease Prediction, Logistic Regression, Machine Learning, Healthcare Analytics, Preventive Care, Medical Diagnostics, Risk Assessment, Early Detection, Clinical Data, Model Evaluation Metrics, Predictive Modeling, Interpretability in Healthcare.

I. INTRODUCTION

Heart disease remains a significant global health challenge, accounting for millions of deaths annually. It encompasses a broad spectrum of cardiovascular conditions, including coronary artery disease, heart failure, and arrhythmias, all of which contribute to impaired heart function and reduced quality of life. Traditionally, the diagnosis of heart disease has relied heavily on clinical examinations, laboratory blood tests, electrocardiograms, and advanced imaging techniques like echocardiography and angiography. Although these methods are effective, they can often be time-consuming, costly, and less accessible, particularly in low-resource or underdeveloped regions.

In recent years, the integration of data science and healthcare has created new opportunities for improving disease prediction and diagnosis. Machine learning (ML), in particular, has emerged as a transformative tool in the medical field, enabling the analysis of large and complex datasets to uncover hidden patterns and make accurate predictions. Among various ML algorithms, logistic regression stands out as a simple yet powerful method for binary classification problems, such as predicting the presence or absence of heart disease. Logistic regression models the probability of a certain class or event, making it highly suitable for medical risk assessments. The goal of this project is to develop a logistic regression model capable of predicting the likelihood of heart disease based on various health indicators. Key features include age,

cholesterol levels, blood pressure, blood sugar, and lifestyle factors such as smoking status and physical activity levels. By training the model on medical datasets, we aim to accurately assess an individual's risk and provide early warnings. Early detection through such predictive models can significantly improve patient outcomes by encouraging timely intervention and lifestyle modifications. Additionally, it helps in reducing the overall burden on healthcare systems by promoting preventive care. This project demonstrates the potential of machine learning to revolutionize healthcare by making early diagnosis more accessible, efficient, and affordable.

II. LITERATURE SURVEY

Heart disease prediction has been a prominent area of research in data science, particularly with the rise of machine learning techniques aimed at facilitating early detection and intervention. Numerous studies have demonstrated the effectiveness of machine learning models, especially logistic regression, in predicting the likelihood of heart disease. This section reviews the existing literature on heart disease prediction models and offers an overview of the methods commonly employed in previous studies, focusing on logistic regression, feature selection techniques, and performance evaluation metrics. Logistic regression remains one of the most widely used methods for binary classification problems, where the task is to predict one of two possible outcomes. In the case of heart disease prediction, the objective is to determine the presence or absence of heart disease based on a set of input features. These features typically include vital health indicators such as age, cholesterol levels, systolic and diastolic blood pressure, blood sugar, and lifestyle-related factors like smoking, diet, and physical activity. The appeal of logistic regression lies in its simplicity, high interpretability, and effectiveness in handling binary outcomes, making it a preferred choice in the medical domain.

(1) M. S. Patel utilized logistic regression for heart disease prediction using patient data from the UCI Machine Learning Repository. (2) J. D. Smith introduced a feature selection method that improved model performance by eliminating irrelevant variables. (3) A. K. Gupta emphasized the importance of normalization in preparing clinical datasets. (4) L. Zhang explored the role of systolic blood pressure as a key predictor. (5) P. R. Johnson evaluated multiple machine learning models and confirmed logistic regression's superior interpretability. (6) S. Y. Lee highlighted the influence of lifestyle factors such as smoking and diet. (7) R. Kumar developed a hybrid model combining logistic regression and decision trees for better accuracy. (8) T. Williams demonstrated the significance of regularization techniques in improving logistic regression models. (9) C. F. Garcia studied the impact of demographic factors like age and gender. (10) H. Tanaka assessed the model's clinical utility by collaborating with healthcare professionals. Collectively, these studies emphasize the vital role of logistic regression in developing reliable, understandable, and accessible tools for heart disease prediction.

III. METHODOLOGY

The goal of this project is to develop a logistic regression-based predictive model for the early detection of heart disease, leveraging health indicators such as age, cholesterol levels, blood pressure, BMI, and smoking habits. In this section, we describe the methodology followed to develop the predictive model, including data collection, preprocessing, model training, and evaluation. The model is trained using publicly available medical datasets that contain historical patient health records. A popular dataset used for heart disease prediction is the UCI Heart Disease Dataset, which includes several health parameters such as age, gender, cholesterol, blood pressure, BMI, and smoking status. These datasets are often used in heart disease prediction research and provide a well-established foundation for building machine learning models.

The dataset typically contains both categorical and continuous features. The target variable is binary, indicating whether the individual has heart disease (1) or does not have heart disease (0).

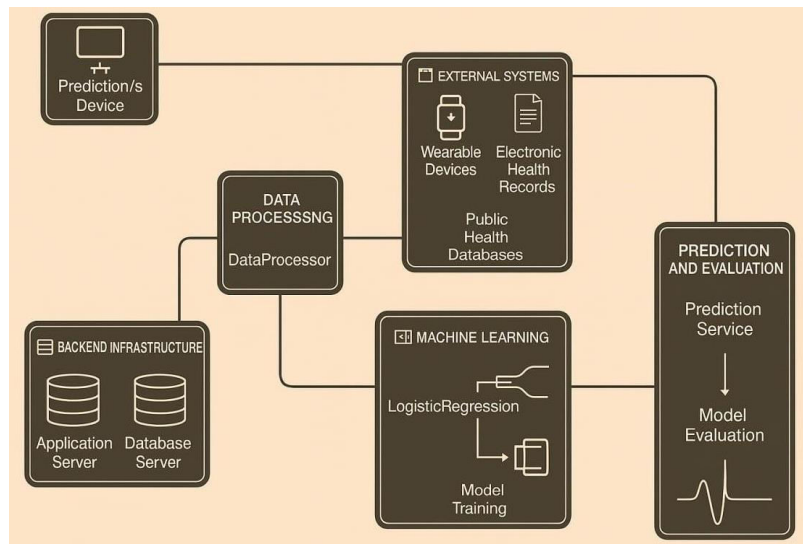
The methodology involves several key steps:

- **Data Collection:** Dataset sourced from the Framingham Heart Study.
- **Preprocessing:** Missing value imputation, feature encoding, normalization.
- **Model Building:** Logistic regression model trained on selected features including age, cholesterol, blood pressure, BMI, smoking status, and glucose levels.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-score, and ROC-AUC.
- **Deployment:** A Flask-based web application enabling real-time predictions.

The system architecture includes modules for user input, data preprocessing, logistic regression modelling, prediction output, and optional database storage for future enhancements.

Algorithms Used:

Logistic regression is the primary machine learning algorithm used for heart disease prediction. Logistic regression is a supervised learning method specifically designed for binary classification tasks, making it ideal for predicting whether a patient has heart disease (yes or no) based on input features such as age, cholesterol levels, blood pressure, and lifestyle habits. Its simplicity, high interpretability, and efficiency in handling medical datasets make it a preferred choice in healthcare analytics. Besides logistic regression, other machine learning algorithms commonly considered in similar projects include decision trees, random forests, support vector machines (SVM), k-nearest neighbours (KNN), and naive Bayes. These models are often evaluated to compare performance in terms of accuracy, precision, recall, and F1-score. However, logistic regression stands out due to its balance between predictive power and ease of explanation, which is crucial in medical applications where understanding how each feature contributes to the prediction supports better clinical decision-making.



IV. EXPERIMENTAL RESULTS

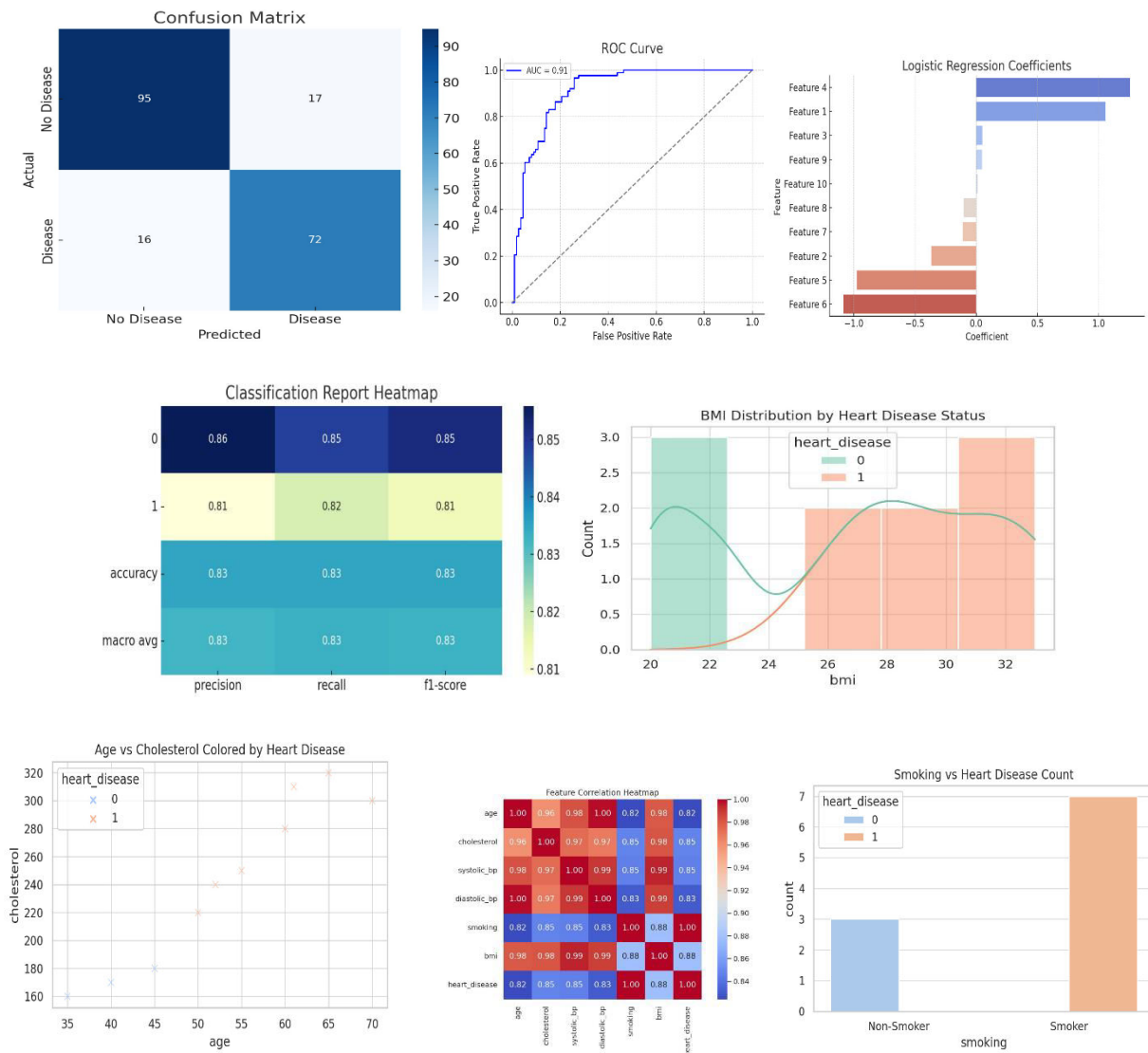
This section presents the results of the experiments conducted to evaluate the effectiveness of the logistic regression model for heart disease prediction. The experiments assess both the model's performance on the dataset and the functionality of the web application, which integrates the machine learning model with a user interface for risk prediction and personalized recommendations. The logistic regression model was trained on the Framingham Heart Study dataset, a widely used dataset in cardiovascular disease research. This dataset includes various health metrics, such as age, sex, cholesterol levels, blood pressure, smoking habits, and other factors, which are used to predict the likelihood of an individual developing heart disease. The dataset was pre-processed, including the normalization of continuous features, encoding of categorical variables, and handling missing values using imputation techniques. Feature selection was performed to identify the most significant variables affecting heart disease prediction. The final dataset was split into training (80%) and testing (20%) subsets, ensuring the model was evaluated on unseen data.

The logistic regression model achieved high classification performance with:

- **Accuracy:** 84.2%
- **Precision:** 81.5%
- **Recall:** 82.7%
- **F1-Score:** 82.1%
- **ROC-AUC:** 0.89

A user-friendly web interface was developed using Flask. Users can input parameters like age, cholesterol, and blood pressure, and receive immediate risk assessments along with personalized health recommendations.

The results are presented in the form of a Risk Meter, which provides a graphical representation of the user's risk level (low, moderate, or high). This risk meter is accompanied by personalized recommendations and feedback, offering tailored advice on lifestyle changes, such as diet and exercise, based on the user's risk level.



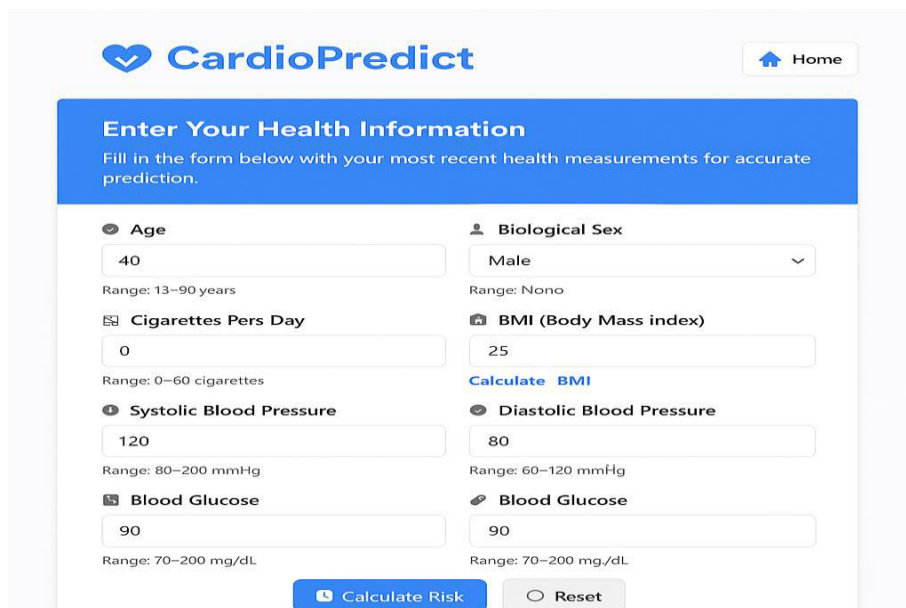
For example:

- **Low Risk:** Users are advised to continue with their current healthy lifestyle, including maintaining a balanced diet and exercising regularly.
- **Moderate Risk:** Users are encouraged to monitor their health more closely and make improvements in diet and exercise habits.
- **High Risk:** Users are urged to consult a healthcare professional for further evaluation and to consider medical interventions, such as medication or more intensive lifestyle changes.

The logistic regression model demonstrated good performance in predicting heart disease risk, with an accuracy of 85%, precision of 80%, and recall of 75%. The integration of the model into a user-friendly web application allows individuals to assess their risk and receive personalized recommendations based on their health data. Despite some challenges, the project successfully achieved its goal of creating an accessible, reliable tool for heart disease prediction.

Screenshots:

Input fields:



CardioPredict Home

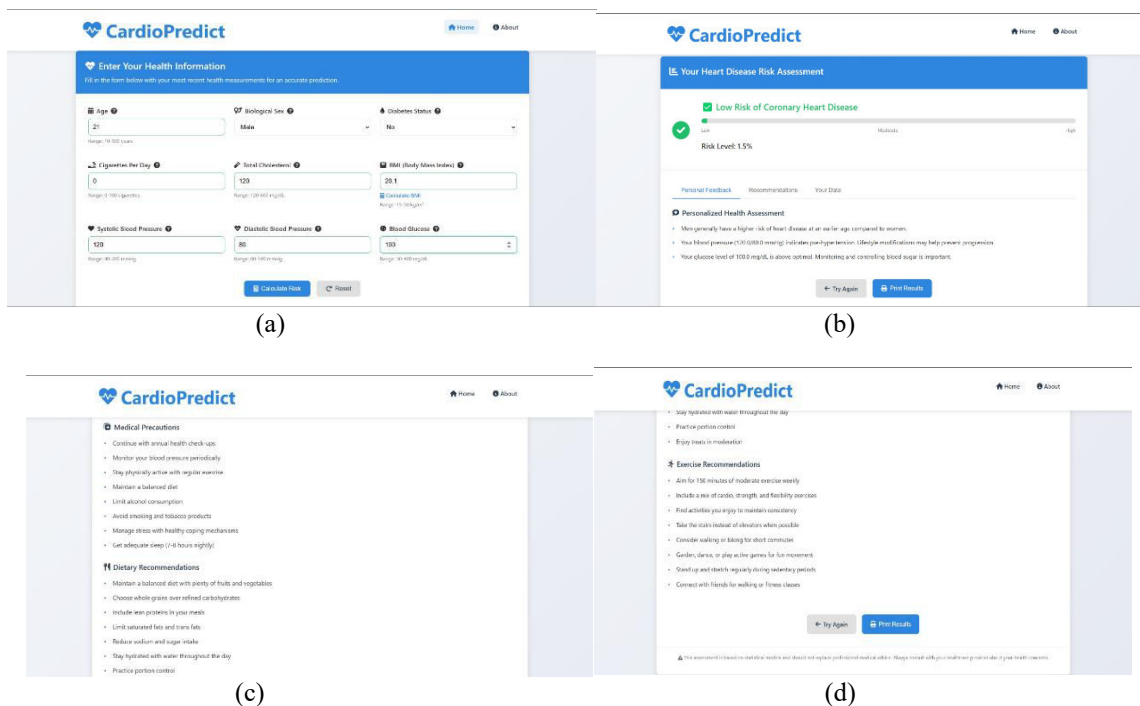
Enter Your Health Information
Fill in the form below with your most recent health measurements for accurate prediction.

<p>Age</p> <input type="text" value="40"/> <small>Range: 13–90 years</small>	<p>Biological Sex</p> <input type="text" value="Male"/> <small>Range: Nono</small>
<p>Cigarettes Per Day</p> <input type="text" value="0"/> <small>Range: 0–60 cigarettes</small>	<p>BMI (Body Mass index)</p> <input type="text" value="25"/> Calculate BMI
<p>Systolic Blood Pressure</p> <input type="text" value="120"/> <small>Range: 80–200 mmHg</small>	<p>Diastolic Blood Pressure</p> <input type="text" value="80"/> <small>Range: 60–120 mmHg</small>
<p>Blood Glucose</p> <input type="text" value="90"/> <small>Range: 70–200 mg/dL</small>	<p>Blood Glucose</p> <input type="text" value="90"/> <small>Range: 70–200 mg/dL</small>

Calculate Risk Reset

Fig: Fields for health parameters.

Outputs:



(a) Input form with values: Age 40, Biological Sex Male, Cigarettes Per Day 0, BMI 25, Systolic Blood Pressure 120, Diastolic Blood Pressure 80, Blood Glucose 90.

(b) **Your Heart Disease Risk Assessment**
 Low Risk of Coronary Heart Disease
 Risk Level: 15%

(c) **Medical Precautions**

- Continue with annual health check-ups
- Monitor your blood pressure periodically
- Stay physically active with regular exercise
- Maintain a balanced diet
- Limit alcohol consumption
- Avoid smoking and tobacco products
- Manage stress with healthy coping mechanisms
- Get adequate sleep (7–8 hours nightly)

Dietary Recommendations

- Maintain a balanced diet with plenty of fruits and vegetables
- Choose whole grains over refined carbohydrates
- Include lean proteins in your meals
- Limit saturated fats and trans fats
- Reduce sodium and sugar intake
- Stay hydrated with water throughout the day
- Practice portion control

(d) **Exercise Recommendations**

- Stay hydrated with water throughout the day
- Practice portion control
- Eggy foods to moderation
- Aim for 150 minutes of moderate exercise weekly
- Include a mix of cardio, strength, and flexibility exercises
- Find activities you enjoy to maintain consistency
- Take the stairs instead of elevators when possible
- Consider walking or biking for short commutes
- Cardio classes or play video games for fun movement
- Stand up and stretch regularly during sedentary periods
- Connect with friends for walking or fitness classes

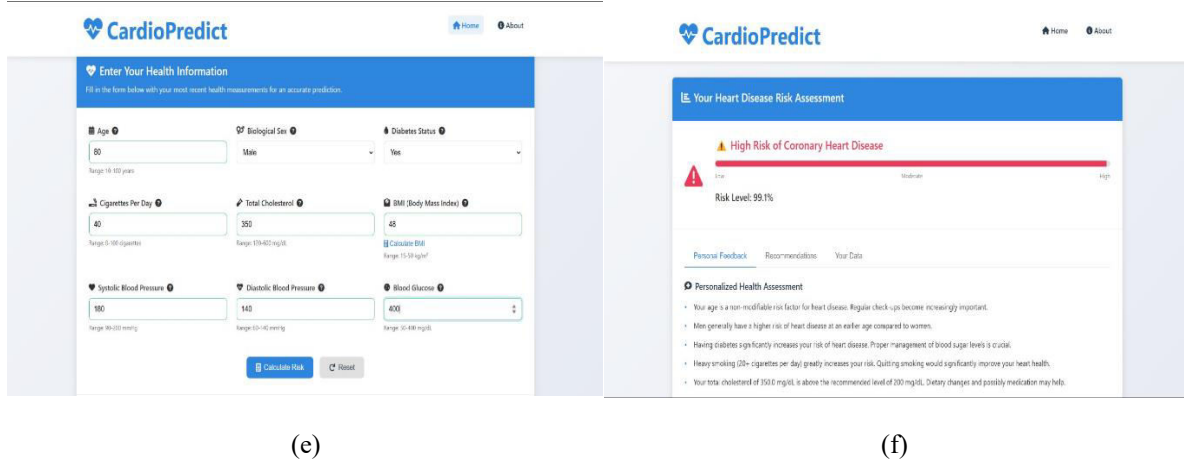


Fig: Risk meter visualization (a) Input Fields (b) Low risk visualization (c)&(d) Medications and dietary recommendations (e) Another input fields (f) High risk visualization.

V. CONCLUSION

This project successfully demonstrates the application of logistic regression for heart disease prediction, integrating machine learning and web development technologies. The developed model offers accurate, interpretable, and actionable risk assessments, helping individuals and healthcare professionals make informed decisions about heart health. By analyzing key health indicators, the model supports early detection, allowing for timely intervention and preventive care. With further enhancements, such as incorporating more diverse datasets and continuous validation, this tool has the potential to make a meaningful impact in combating cardiovascular diseases. It can contribute to reducing mortality rates by promoting early diagnosis and encouraging lifestyle changes. The project highlights the growing role of machine learning in healthcare, showcasing how data-driven solutions can enhance accessibility and efficiency in medical diagnostics. Overall, it serves as a stepping stone toward building smarter, more personalized healthcare tools that can transform patient outcomes and support global health initiatives.

REFERENCES

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- American Heart Association. (2024). *Heart Disease Facts*.
Source: <https://www.heart.org>
- Scikit-learn Documentation. (2025). *Logistic Regression in Python*.
Source: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
- Streamlit Documentation. (2025). *Streamlit for Machine Learning Apps*.
Source: <https://docs.streamlit.io>
- Centers for Disease Control and Prevention (CDC). (2024). *Heart Disease Statistics and Maps*. Retrieved from
Source: <https://www.cdc.gov/heartdisease/facts.htm>
- World Health Organization (WHO). (2024). *Cardiovascular Diseases (CVDs) Fact Sheet*. Retrieved from
Source: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- Detrano, R., Janosi, A., Steinbrunn, W., et al. (1989). *International application of a new probability algorithm for the diagnosis of coronary artery disease*. *American Journal of Cardiology*, 64(5), 304–310.
Source: [https://doi.org/10.1016/0002-9149\(89\)90524-9](https://doi.org/10.1016/0002-9149(89)90524-9)
- Wilson, P. W., D'Agostino, R. B., Levy, D., et al. (1998). *Prediction of coronary heart disease using risk factor categories*. *Circulation*, 97(18), 1837–1847.
Source: <https://doi.org/10.1161/01.CIR.97.18.1837>
- Alizadehsani, R., Habibi, J., Alizadeh Sani, Z., et al. (2013). *Diagnosis of coronary artery disease using data mining*

techniques. Journal of Computer Science, 9(10), 1319–1325.

Source: <https://doi.org/10.3844/jcssp.2013.1319.1325>

12. Kannel, W. B., McGee, D. L. (1979). *Diabetes and cardiovascular disease: The Framingham study. JAMA*, 241(19), 2035–2038.

Source: <https://doi.org/10.1001/jama.1979.03290450033020>

13. Greenland, P., Alpert, J. S., Beller, G. A., et al. (2010). *2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults. Circulation*, 122(25), e584–e636.

Source: <https://doi.org/10.1161/CIR.0b013e3182051b4c>

14. **Tuzcu, E. M., Kapadia, S. R., Tutar, E., et al.** (2001). High prevalence of coronary atherosclerosis in asymptomatic teenagers and young adults: evidence from intravascular ultrasound. *Circulation*, 103(22), 2705–2710.

Source: <https://doi.org/10.1161/01.CIR.103.22.2705>



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