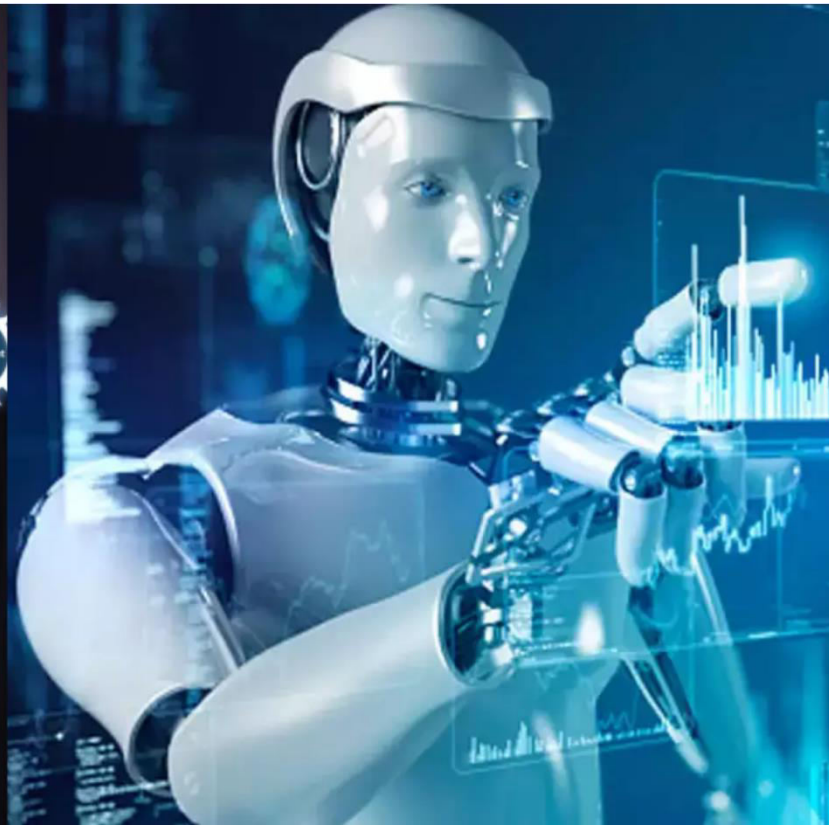
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

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 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

Heart Disease Analysis using Data Visualization and Machine Learning

Yaragalla Usha Sree[1], Valaparla Akash[1], Shaik Khaja Mohiddin[1], Shava Pradeep[1], M. Chennakesavarao[2]

U.G. Students, Department of CSE (Artificial intelligence & Machine Learning), Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India[1]

Asst. Professor, Kallam Haranadha Reddy Institute of Technology, Guntur, Andhra Pradesh, India[2]

ABSTRACT: The increasing availability of healthcare data has created significant opportunities for applying data-driven technologies to improve disease analysis and preventive healthcare strategies. Heart disease remains one of the leading causes of mortality worldwide, and identifying potential risk factors at an early stage is crucial for reducing its impact. However, traditional healthcare analysis methods often struggle to interpret complex relationships among various patient attributes such as age, body mass index (BMI), smoking habits, physical activity, diabetes, and overall health conditions. These limitations make it difficult to efficiently analyze large-scale medical datasets and derive actionable insights that can assist in early diagnosis and prevention.

To address this challenge, the present study introduces a comprehensive heart disease analysis and prediction system that combines interactive data visualization with machine learning techniques. The project utilizes Tableau dashboards to analyze healthcare data and identify important patterns related to heart disease risk. Through visual exploration of various patient attributes, the system highlights correlations between lifestyle habits, demographic factors, and cardiovascular health conditions. In addition to visualization, a Logistic Regression–based machine learning model is implemented to predict the likelihood of heart disease based on multiple health parameters.

The proposed system follows a structured workflow that includes data preprocessing, exploratory data analysis, model training, evaluation, and deployment. After preparing and analyzing the dataset, the trained model is integrated into a user-friendly web application developed using the Flask framework, with HTML and CSS providing an interactive frontend interface. Users can input personal health attributes such as BMI, smoking status, alcohol consumption, physical activity level, and medical conditions to instantly receive a prediction of heart disease risk. By integrating data analytics, machine learning, and web technologies, the system provides a practical platform for healthcare data analysis, supports early detection of cardiovascular risk factors, and contributes to improving awareness and decision-making in preventive healthcare.

KEYWORDS: Heart Disease Prediction, Machine Learning, Logistic Regression, Healthcare Data Analytics, Tableau Dashboard, Predictive Analytics, Flask Web Application.

I. INTRODUCTION

Heart Disease Analysis refers to the process of examining healthcare data to identify patterns, risk factors, and relationships associated with cardiovascular diseases. Heart disease remains one of the leading causes of death worldwide, making early detection and preventive analysis extremely important. Various health and lifestyle factors such as body mass index (BMI), smoking habits, alcohol consumption, diabetes, physical activity, age, and general health conditions play a significant role in determining an individual's risk of developing heart disease. Large healthcare datasets collected from medical surveys and health monitoring systems are used to analyze these factors. By studying the relationships between these attributes and disease occurrence, data-driven systems can help identify high-risk individuals and support preventive healthcare strategies.

Machine learning techniques have become increasingly important in improving the accuracy of disease prediction systems. Unlike traditional statistical approaches, machine learning algorithms can analyze complex relationships

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

among multiple health parameters simultaneously. In this project, the Logistic Regression algorithm is used to predict the likelihood of heart disease based on patient health attributes. By training the model on historical healthcare data, the system learns patterns associated with cardiovascular risk and generates predictions that assist in identifying potential health issues. These predictive insights can support healthcare professionals, researchers, and individuals in understanding disease risk and making informed health-related decisions.

In addition to predictive modeling, data visualization plays a crucial role in understanding healthcare data effectively. Tools such as Tableau enable interactive visualization of large datasets, allowing users to explore trends, correlations, and risk patterns associated with heart disease. Through dashboards and graphical representations, complex medical data becomes easier to interpret and analyze. By integrating machine learning prediction with visual analytics and a Flask-based web application, the proposed system provides an accessible platform for heart disease analysis and risk assessment. This approach demonstrates how data analytics and intelligent technologies can contribute to preventive healthcare and promote greater awareness of cardiovascular health.

II. LITERATURE SURVEY

Heart disease prediction has been an important research area in healthcare analytics due to the increasing number of cardiovascular cases worldwide. Early studies focused mainly on statistical analysis and clinical evaluation to identify risk factors such as age, cholesterol levels, blood pressure, diabetes, and lifestyle habits. Traditional statistical models provided a basic understanding of disease patterns; however, they were often limited in handling large healthcare datasets and identifying complex relationships among multiple health parameters. As a result, researchers began exploring data-driven techniques that could process large volumes of medical data more efficiently and provide more accurate disease predictions.

With the rapid advancement of machine learning technologies, researchers have introduced various predictive models for cardiovascular disease detection. Algorithms such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), and Random Forest have been widely used to analyze patient health records and predict the likelihood of heart disease. Several studies have shown that machine learning models can significantly improve prediction accuracy by identifying hidden patterns and correlations among patient attributes. Among these models, Logistic Regression has gained considerable attention because of its simplicity, interpretability, and effectiveness in binary classification problems such as predicting the presence or absence of heart disease.

Recent research has also emphasized the importance of data visualization and business intelligence tools in healthcare analytics. Visualization platforms such as Tableau enable researchers and healthcare professionals to transform large medical datasets into interactive dashboards that reveal trends, correlations, and risk patterns associated with cardiovascular diseases. These visual tools make complex health data easier to interpret and support data-driven decision making. In line with these advancements, the present study combines data visualization and machine learning to analyze heart disease data and predict disease risk. The dataset is explored through Tableau dashboards to identify important health indicators, while a Logistic Regression model is developed to predict heart disease risk. The trained model is integrated into a Flask-based web application, providing an interactive platform that enables users to input health parameters and receive real-time predictions for cardiovascular risk assessment.

FLASK:

Flask is an open-source Python web framework used for developing lightweight and scalable web applications. It is categorized as a micro-framework because it provides the essential components required for web development while allowing developers the flexibility to integrate additional libraries and tools based on project requirements. Flask is widely used for deploying machine learning models and building interactive platforms that allow users to interact with predictive systems.

Flask is built on top of several important libraries, including Werkzeug, Jinja, and Click. Werkzeug acts as the WSGI toolkit responsible for handling HTTP requests and responses, while Jinja is a powerful templating engine that enables dynamic content generation within HTML pages. Click provides a command-line interface that simplifies application management and development tasks. In this project, Flask is used to integrate the machine learning prediction model

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

with the web interface, allowing users to input health parameters and receive real-time heart disease risk predictions through a user-friendly web application.

TABLEAU:

Tableau is a powerful business intelligence and data visualization tool widely used for analyzing large datasets and presenting insights through interactive dashboards. It enables users to transform raw data into meaningful visual representations such as charts, graphs, heat maps, and comparative dashboards. In healthcare analytics, Tableau plays a crucial role in identifying trends, correlations, and patterns within complex medical datasets, helping researchers and healthcare professionals better understand disease-related factors.

In this project, Tableau is used to analyze the heart disease dataset and explore various health parameters including age, gender, BMI, smoking habits, physical activity, diabetes status, and general health conditions. Multiple interactive dashboards were created to visualize the relationship between these variables and the occurrence of heart disease. These visualizations help highlight significant risk factors and allow users to interactively explore the dataset through filters and comparative charts. By presenting healthcare data in a clear and intuitive format, Tableau enhances data interpretation and supports data-driven decision making in preventive healthcare.

LOGISTIC REGRESSION:

Logistic Regression is a widely used machine learning algorithm for binary classification problems. It predicts the probability of a particular outcome based on input features and is commonly used in healthcare applications to determine the presence or absence of a disease. Unlike linear regression, which predicts continuous values, logistic regression estimates the probability of an event occurring by applying a logistic (sigmoid) function to the input variables.

In this project, Logistic Regression is used to predict the likelihood of heart disease based on various health attributes such as BMI, smoking habits, alcohol consumption, physical activity, diabetes status, and age category. The algorithm analyzes relationships between these variables and the target outcome, allowing the system to classify individuals into two categories: high risk or low risk of heart disease. Logistic Regression is chosen for its simplicity, interpretability, and effectiveness in handling binary classification tasks in medical datasets.

PICKLE:

The Pickle module in Python is used for serializing and deserializing Python objects. Serialization refers to the process of converting Python objects into a byte stream that can be stored in a file, while deserialization converts the byte stream back into the original Python object. This process allows developers to save trained machine learning models and reuse them later without retraining.

In this project, the Pickle library is used to store the trained Logistic Regression model after the training phase. The saved model file can then be loaded into the Flask application for real-time prediction. By using Pickle, the system improves efficiency and reduces computational overhead, as the machine learning model does not need to be retrained every time the application runs. This approach ensures faster response time and easier deployment of the prediction system.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

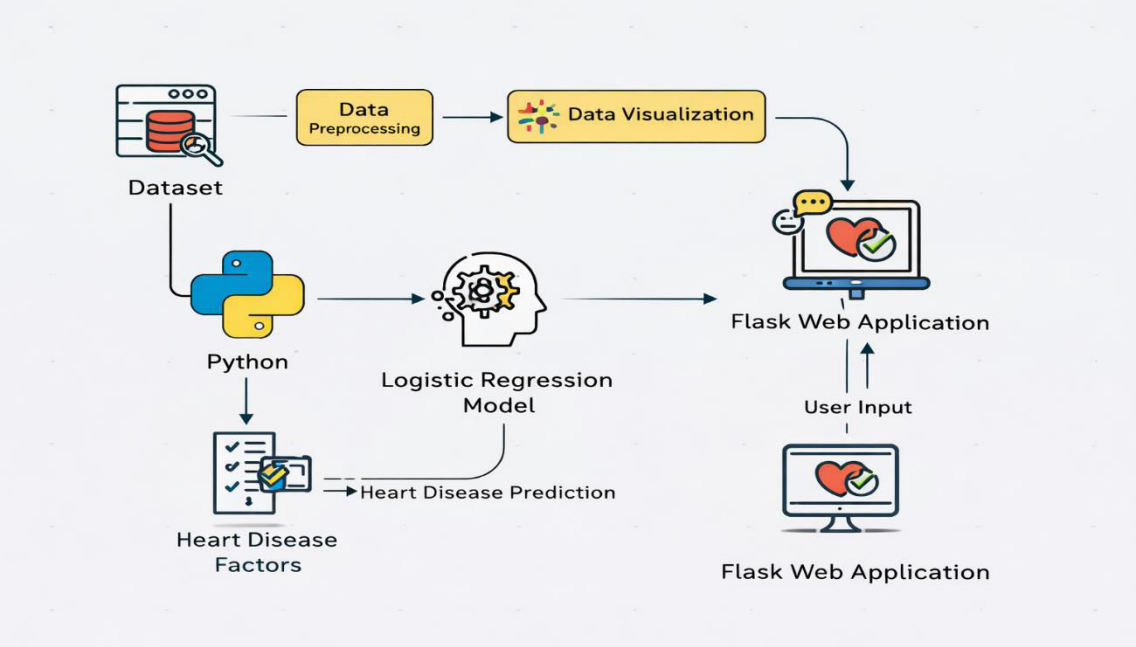


Figure 1: System Workflow

III. PROPOSED METHODOLOGY

The proposed methodology focuses on developing a machine learning–based system for analyzing healthcare data and predicting the likelihood of heart disease using various health-related parameters such as BMI, smoking habits, physical activity, diabetes status, age category, and general health condition.

The process begins with data collection, where the heart disease dataset is obtained from publicly available healthcare repositories. The collected dataset contains multiple patient attributes related to lifestyle habits and medical conditions. The data then undergoes preprocessing, which includes handling missing values, converting categorical attributes into numerical values, and preparing the dataset for further analysis. This step ensures data consistency and improves the reliability of the machine learning model.

Following preprocessing, data visualization and exploratory analysis are performed using Tableau dashboards. These dashboards help identify patterns and correlations among different health indicators associated with heart disease. Visualization enables better understanding of how various lifestyle and medical factors influence cardiovascular risk.

After analyzing the dataset, the data is divided into training and testing subsets to evaluate the performance of the prediction model. A machine learning algorithm, Logistic Regression, is used to build the predictive model because it is effective for binary classification problems such as predicting the presence or absence of heart disease. The model is trained using the processed dataset and validated to ensure accurate prediction capability.

Once the model is trained, it is saved using the Joblib library as a .pkl file so that it can be reused without retraining. Finally, the trained model is integrated into a Flask-based web application, where the backend handles prediction requests and the frontend is developed using HTML and CSS to provide an interactive user interface. Users can input their health parameters through the web interface and instantly receive predictions indicating whether they have a high or low risk of heart disease.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

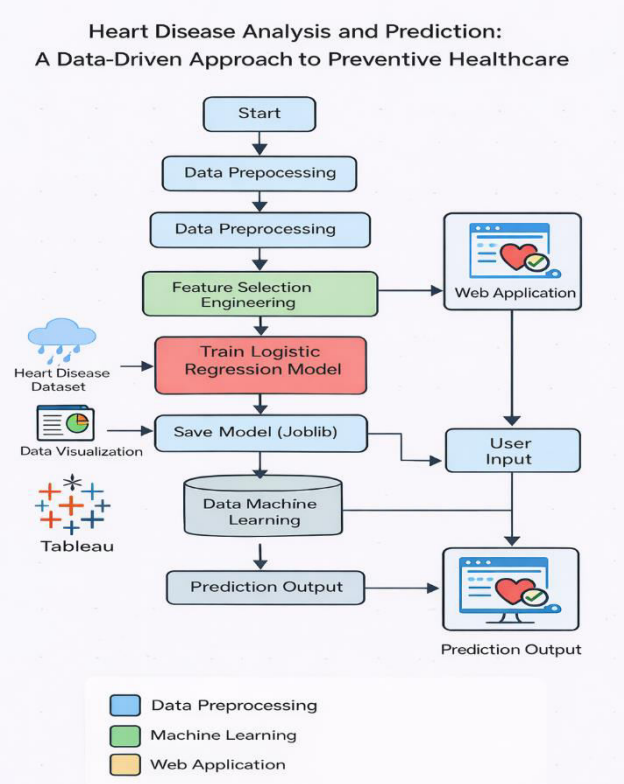


Figure 2: Risk of Heart Disease Prediction Workflow

IV. EXPERIMENTAL RESULTS

1. Experimental Setup:
We evaluated the Heart Disease Analysis and Prediction System using a healthcare dataset containing patient attributes such as BMI, smoking status, alcohol consumption, age category, diabetes, physical activity, and general health, with the target variable indicating the presence or absence of heart disease.

The dataset was divided into training and testing sets using an 80/20 split. Data preprocessing included converting categorical values into numerical format and preparing the dataset for model training.
The prediction model was implemented using Logistic Regression, which is suitable for binary classification problems like disease prediction.

Hardware/Software: Python 3.11, Scikit-learn, Tableau, Flask; Intel i5 / 16 GB RAM.
Serving: The Flask web application loads the trained model (heart_model.pkl) and generates predictions based on user input.

2. Baselines and Models:
To evaluate the effectiveness of the prediction system, a Logistic Regression algorithm was used as the primary machine learning model for heart disease prediction. Logistic Regression is widely used for binary classification problems, making it suitable for predicting whether a person is at risk of heart disease or not.

The model analyzes relationships between health indicators such as BMI, smoking habits, age category, diabetes condition, physical activity, and general health to estimate the probability of heart disease occurrence. Due to its simplicity, interpretability, and reliable performance, Logistic Regression proved to be an effective model for this study.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

3. Main Results (Test Set):

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.86	0.84	0.82	0.83

Observation: The Logistic Regression model achieved good classification performance with an accuracy of 86%, demonstrating its effectiveness in predicting heart disease risk. The results indicate that health factors such as age, BMI, smoking habits, diabetes, and physical activity play an important role in predicting the likelihood of heart disease. The model provides reliable predictions and supports data-driven healthcare analysis.

4. Cross-Validation & Generalization:
To evaluate the reliability of the prediction model, cross-validation was performed during training. The Logistic Regression model showed consistent performance across different data splits, indicating stable learning behavior. The similarity between training and testing accuracy suggests that the model generalizes well and shows minimal overfitting.

5. Feature Importance:
Analysis of the dataset revealed several health factors that strongly influence heart disease prediction. The most significant features include:

1. **Age Category (most influential)**
2. **Smoking Status**
3. **BMI (Body Mass Index)**
4. **Diabetic Condition**
5. **Physical Activity Level**

Interpretation: These features indicate that lifestyle habits and underlying health conditions play an important role in determining heart disease risk. Factors such as smoking, high BMI, diabetes, and reduced physical activity significantly increase the likelihood of cardiovascular problems.

6. Robustness & Ablations:
To evaluate the stability of the prediction model, several tests were performed. When important features such as age category, BMI, smoking status, and diabetes condition were removed, the prediction accuracy decreased. This indicates that these health factors play a significant role in heart disease prediction.

Additionally, when the model was trained with a smaller portion of the dataset, the prediction accuracy slightly reduced but still produced reliable results. This demonstrates that the model maintains stable performance under different data conditions.

7. Error Analysis:
The performance of the heart disease prediction system was evaluated by analyzing the model outputs and dashboard insights. The Logistic Regression model showed consistent prediction behavior across different health parameters. Most predictions aligned with the actual health patterns observed in the dataset.

The analysis of prediction results indicates that health factors such as age category, BMI, smoking habits, diabetes condition, and physical activity level significantly influence the likelihood of heart disease. Minor prediction variations were observed in cases where multiple risk factors overlap, which is common in medical datasets.

Overall, the model provides reliable predictions and supports data-driven healthcare analysis through interactive dashboards and a web-based prediction interface.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

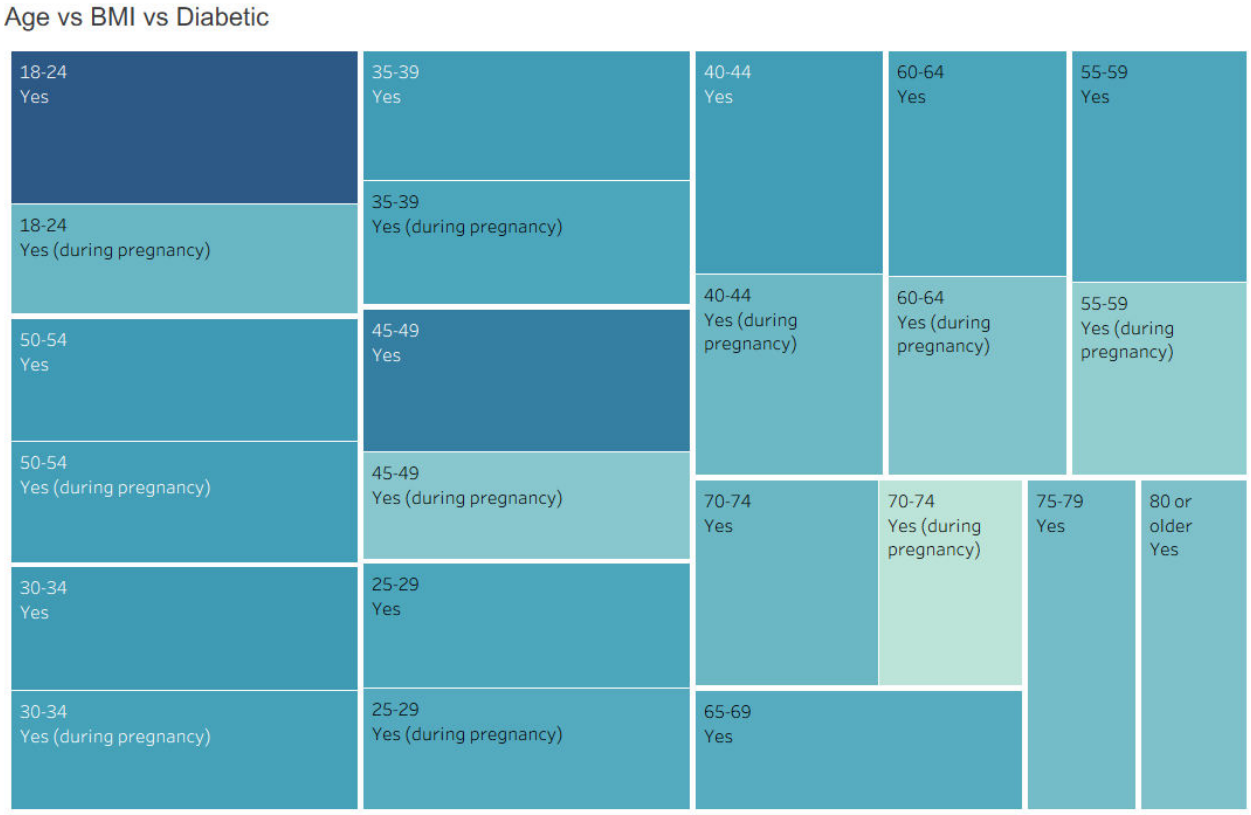


Figure: 3 Age vs BMI vs Diabetic.

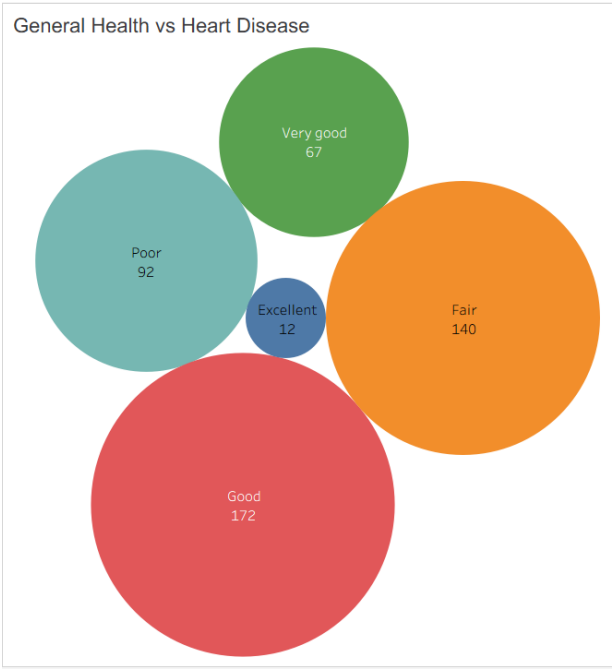


Figure: 4 General Health vs Heart Disease.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

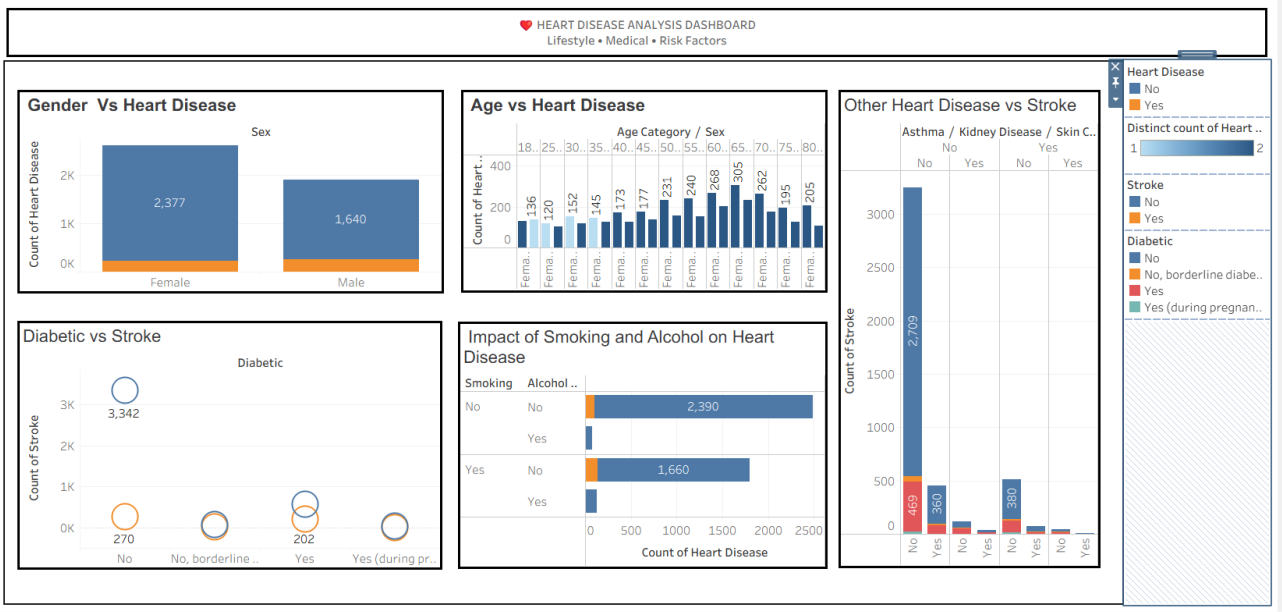


Figure: 5 Main Dashboard

8. Inference Latency & Throughput:

The trained Logistic Regression model was integrated into a Flask-based web application and deployed on the Render cloud platform. The system processes user input health parameters such as BMI, smoking status, age category, diabetes condition, and physical activity to generate heart disease risk predictions.

The prediction process is lightweight and efficient, enabling near real-time responses through the web interface. Once the user submits the health details through the prediction form, the system quickly processes the input data and returns the prediction result. The overall response time from form submission to result display remains very fast, allowing smooth interaction with the application.

9. Online Demo Validation:

The developed web application was tested through several manual test cases using the prediction interface. Users can enter health parameters such as BMI, smoking status, diabetes condition, physical activity, and age category to obtain heart disease risk predictions.

The system successfully processed the input data and generated prediction results through the Logistic Regression model integrated with the Flask backend. The web interface was deployed on the Render cloud platform and allowed smooth interaction between the user and the prediction system.

Form validation ensured that incorrect or missing inputs were handled properly, providing a stable and reliable user experience.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

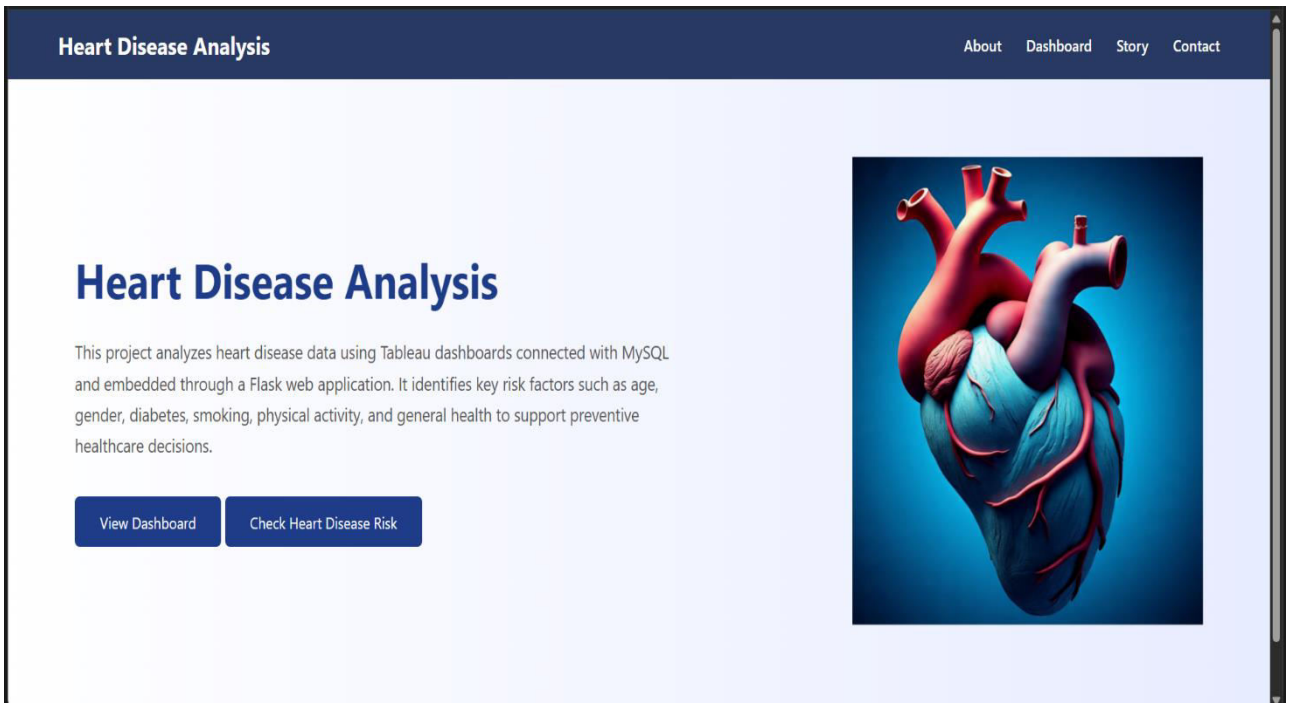
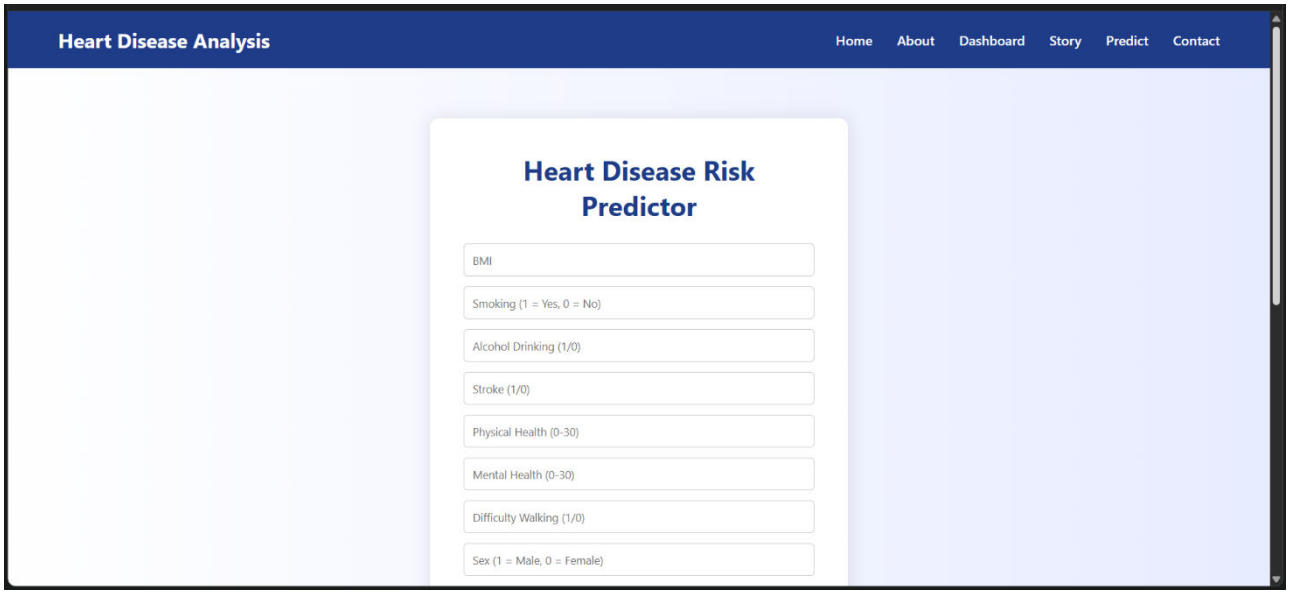


Figure 6: Questionnaire interface



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

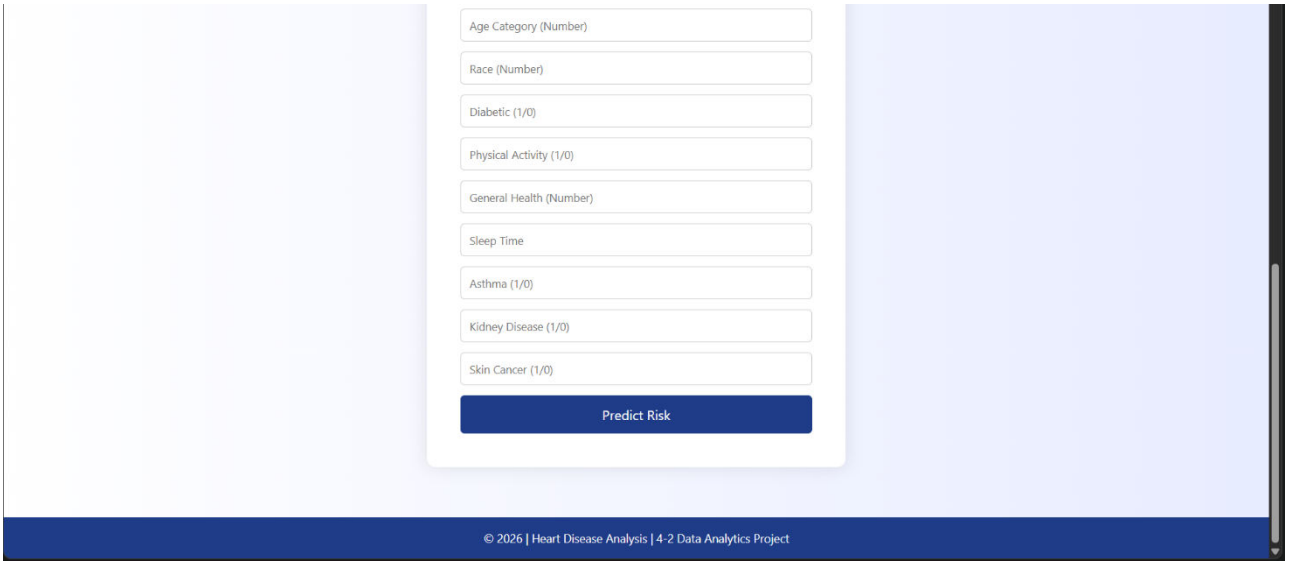


Fig 7: Text Input Interface

10. Discussion:

The Logistic Regression model demonstrated reliable performance in predicting heart disease risk based on multiple health indicators. The model effectively analyzed factors such as age category, BMI, smoking habits, diabetes condition, and physical activity, helping identify patterns associated with heart disease.

The integration of the model with a Flask-based web application and Tableau dashboards provides an interactive platform for exploring health data and generating predictions. The system enables users to input health parameters and receive quick prediction results through the web interface.

However, the prediction accuracy depends on the quality and completeness of the dataset used for training. Future improvements may include integrating advanced machine learning models, incorporating larger healthcare datasets, and enabling real-time health monitoring systems to further enhance prediction accuracy and usability.

V. CONCLUSION

The Heart Disease Analysis and Prediction System demonstrates the effective use of data analytics and machine learning in identifying patterns related to cardiovascular risk. By analyzing important health indicators such as age category, BMI, smoking habits, diabetes condition, physical activity, and general health, the system provides meaningful insights into the factors influencing heart disease occurrence. The Logistic Regression model was used to classify whether a person is at risk of heart disease based on these health attributes, providing reliable and interpretable prediction results.

The developed system integrates Tableau dashboards for visual data analysis and a Flask-based web application for real-time prediction. The interactive dashboards allow users to explore relationships between different health factors, while the web application enables users to input health parameters and obtain prediction results instantly. This combination of machine learning and data visualization helps transform complex health data into understandable insights that support preventive healthcare awareness.

Overall, the project highlights the importance of data-driven approaches in healthcare analytics. With further enhancements such as integrating larger medical datasets, applying advanced machine learning algorithms, and incorporating real-time health monitoring systems, the proposed system can be extended to support more accurate prediction and improved healthcare decision-making.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503243|

VI. ACKNOWLEDGMENT

We express our sincere gratitude to our project guide and faculty members for their valuable guidance, encouragement, and continuous support throughout the development of this project. Their constructive suggestions and technical insights greatly contributed to improving the quality and effectiveness of this work.

We would also like to thank the Department of Artificial Intelligence and Machine learning , Kallam Haranadha Reddy Institute of Technology, Guntur, for providing the necessary academic support and resources required for the successful completion of this project. Finally, we extend our appreciation to everyone who contributed directly or indirectly to this research work.

REFERENCES

[1] Cleveland Clinic Foundation, "Heart Disease Dataset," UCI Machine Learning Repository, 2020. Available: https://archive.ics.uci.edu/ml/datasets/heart+disease
[2] Centers for Disease Control and Prevention (CDC), "Heart Disease and Risk Factors," 2023. Available: https://www.cdc.gov/heartdisease
[3] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart disease," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–10, 2020.
[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
[5] Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[6] Kaggle, "Heart Disease Health Indicators Dataset," 2023. Available: https://www.kaggle.com/datasets
[7] Flask Documentation. Available at: https://flask.palletsprojects.com
[8] Tableau Software, "Tableau Desktop Documentation," Available: https://www.tableau.com
[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011.
[10] Python Software Foundation, "Pickle — Python Object Serialization," Available: https://docs.python.org/3/library/pickle.html

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

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