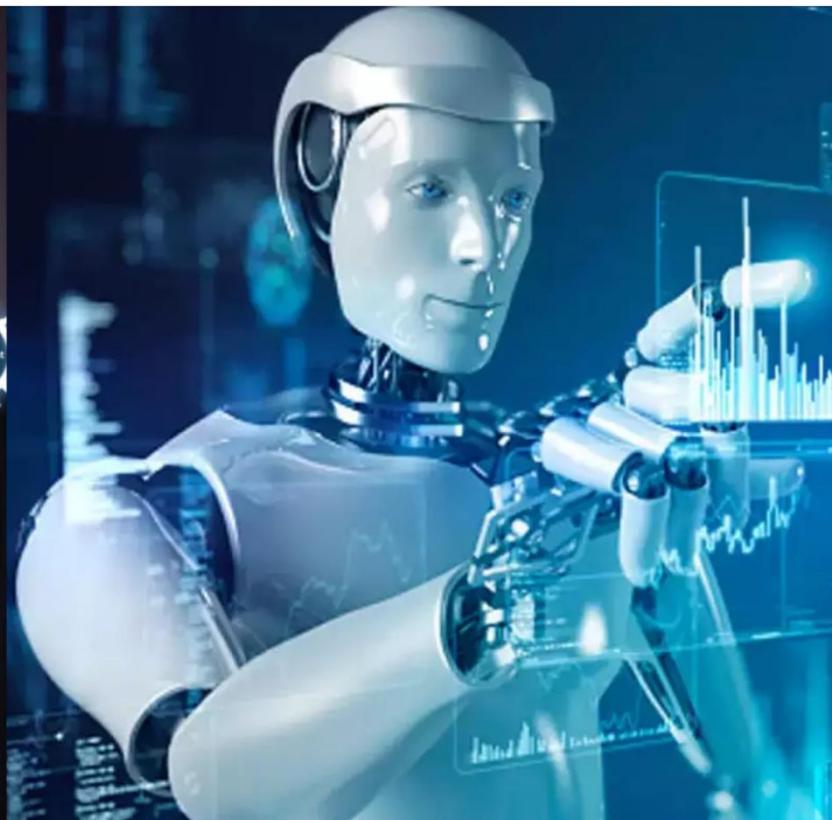




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

Cardiovascular Risk Analysis

Yerragalla Divya¹, Yarlagadda Bhavani Shankar², Shail Khadar Basha³, Munaga Rama Sai⁴,
E. Samba Siva Rao⁵

UG Student, Department of CSE (Artificial intelligence & Machine Learning), Kallam Haranadhareddy Institute of
Technology, Guntur, Andhra Pradesh, India¹⁻⁴

Asst. Professor, Kallam Haranadhareddy Institute of Technology, Guntur, Andhra Pradesh, India⁵

ABSTRACT: Cardiovascular diseases are among the leading causes of health problems worldwide and are strongly influenced by lifestyle habits and medical conditions. Understanding cardiovascular risk factors is important for early prevention and better healthcare planning. This project presents a data-driven framework for analyzing cardiovascular risk using Tableau visual analytics and machine learning techniques. The dataset used in this study was downloaded from Kaggle and is based on the CDC Behavioral Risk Factor Surveillance System (BRFSS) health indicators dataset used for heart disease research. The dataset contains attributes such as BMI, smoking habits, alcohol consumption, diabetes, physical activity, sleep time, and other health indicators that help analyze cardiovascular risk patterns.

Tableau is used to create interactive dashboards and visualizations that highlight relationships between different health factors and cardiovascular risk. These visual insights help identify important trends and patterns in the dataset. To enhance analytical capability, a Random Forest machine learning model is implemented to predict cardiovascular risk based on health parameters.

The analytical components are integrated into a web application developed using the Flask framework, allowing users to input health information and receive cardiovascular risk predictions through an interactive interface. The proposed system demonstrates how the integration of data visualization and machine learning can transform raw health data into meaningful insights that support better understanding of cardiovascular risk factors and data-driven healthcare decisions.

KEYWORDS: Cardiovascular Risk Analysis, Tableau Visualization, Machine Learning, Random Forest, Health Data Analytics, Interactive Dashboards, Kaggle Dataset, Flask Web Application.

I. INTRODUCTION

Cardiovascular diseases are among the leading causes of health problems worldwide and pose a major challenge to public health systems. Factors such as unhealthy lifestyle habits, lack of physical activity, smoking, obesity, and medical conditions like diabetes significantly increase the risk of cardiovascular diseases. As the global population continues to grow and lifestyle patterns change, the need for effective methods to analyze and predict cardiovascular risk has become increasingly important. Early identification of risk factors can help healthcare professionals and individuals take preventive measures and improve overall health outcomes. Traditionally, cardiovascular risk analysis relied on medical reports and basic statistical methods. Although these approaches provide general insights, they often fail to capture complex relationships between different health indicators. With the advancement of **Machine Learning (ML)**, healthcare data can now be analyzed using data driven techniques that reveal hidden patterns and predictive relationships within large health datasets. Machine learning algorithms help in predicting potential health risks, while visualization tools such as Tableau transform complex data into interactive and easy to understand visual insights.

This project focuses on developing a data-driven framework for analyzing cardiovascular risk factors using health indicator data. The dataset used in this study was downloaded from **Kaggle** and is based on the **CDC Behavioral Risk Factor Surveillance System (BRFSS) health indicators dataset**, which is widely used for heart disease research. The dataset includes several health-related attributes that influence cardiovascular risk.

Key health indicators analyzed in this project include:

- **Body Mass Index (BMI)** – a measure used to evaluate obesity and weight-related health risks.
- **Smoking Habits** – identifying whether individuals smoke or not.
- **Alcohol Consumption** – analyzing alcohol drinking behavior.
- **Physical Activity** – understanding the role of exercise and active lifestyle.
- **Diabetes** – identifying individuals with diabetic conditions.
- **General Health Status** – overall health condition reported by individuals
- **Sleep Time** – average number of hours of sleep.
- **Age Category and Gender** – demographic factors affecting cardiovascular risk.

By applying a **Random Forest machine learning model**, to predict heart disease risk based on health indicators, improving accuracy by combining multiple decision trees. It also includes dashboards for interactive visualization of health patterns, helping users understand relationships between different factors. Overall, the project demonstrates how machine learning and visual analytics can transform health data into meaningful insights for better healthcare decisions.

II. LITERATURE SURVEY

Recent studies in healthcare analytics highlight the importance of **data visualization and machine learning techniques** in understanding cardiovascular risk factors. Visualization tools such as **Tableau** help transform large healthcare datasets into meaningful graphical insights. Interactive dashboards enable analysts and healthcare professionals to observe health patterns, compare risk factors, and support better decision-making in preventive healthcare.

Researchers have also explored **machine learning algorithms** for predicting cardiovascular diseases using health indicator datasets. Algorithms such as **Random Forest, Logistic Regression, and Decision Trees** are commonly used to analyze relationships between health parameters and disease risk. These models help identify important health indicators and improve prediction accuracy based on historical health data.

Many previous studies focus either on **health data visualization** or **machine learning-based prediction** separately. This project integrates both **Tableau visual analytics and machine learning prediction** into a unified analytical framework. By combining predictive modeling with interactive dashboards, the system enables users to analyze cardiovascular risk patterns and predict potential health risks using an integrated decision-support platform.

FLASK

Flask is used as the **backend framework** to develop the Cardiovascular Risk Analysis web application. It connects the machine learning model with the user interface, processes user requests, and generates cardiovascular risk predictions dynamically. Flask routes such as **/dashboard, /prediction_page, and /result** manage navigation and prediction functionality within the system.

TABLEAU PUBLIC:

Tableau Public is used to create **interactive dashboards and visualizations** that analyze cardiovascular risk factors. These dashboards present insights such as **age-wise risk distribution, lifestyle factors, BMI trends, and relationships between different health indicators**. The visualizations help users easily explore patterns in health data and better understand cardiovascular risk.

MYSQL DATABASE:

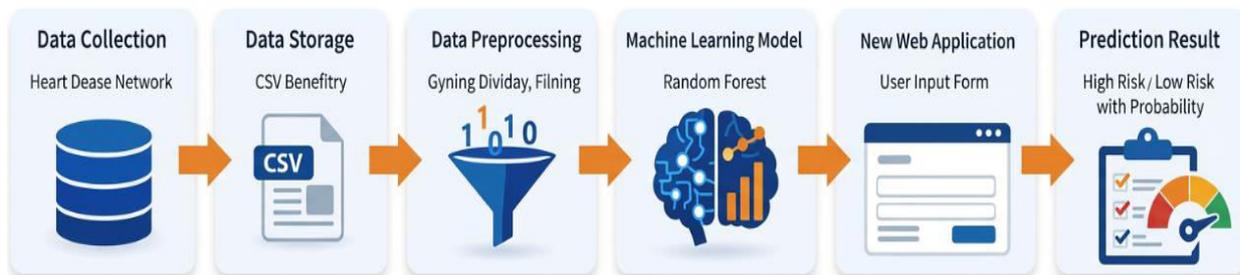
The dataset used in this study was downloaded from **Kaggle** and is based on the **CDC Behavioral Risk Factor Surveillance System (BRFSS) health indicators dataset** used for heart disease research. The dataset contains attributes such as **BMI, smoking habits, alcohol consumption, diabetes, physical activity, sleep time, age category, and general health status**, which are important indicators of cardiovascular risk.

HTML, CSS AND JAVASCRIPT:

HTML, CSS, and JavaScript are used to design the **front-end interface** of the web application. HTML structures the web pages, CSS provides styling and layout, and JavaScript adds interactivity. These technologies allow users to navigate the application, enter health data, and view cardiovascular risk prediction results.

SCIKIT-LEARN:

Scikit-Learn is used to implement the **Random Forest machine learning model** for cardiovascular risk prediction. The model is trained using health indicators such as **BMI, smoking, diabetes, physical activity, sleep time, and age category**. The Random Forest algorithm analyzes patterns in the dataset and predicts the likelihood of cardiovascular risk based on these factors.



Cardiovascular Risk Analysis Using Machine Learning

Fig 1: System Workflow

III. PROPOSED METHODOLOGY

The proposed system combines data preprocessing, feature engineering, machine learning techniques, and web-based deployment to predict the risk of heart disease. The methodology is designed to transform raw health data into meaningful predictions that can assist in early diagnosis and preventive care.

Data Collection:

The dataset used in this project is sourced from publicly available healthcare repositories, containing patient-related information associated with heart disease risk. The dataset includes various medical and lifestyle attributes that influence cardiovascular health.

Each record in the dataset contains the following features:

- **BM**
- **Smoking Habits**
- **Alcohol Consumption**
- **Stroke History**
- **Gender**
- **Age Category**
- **Diabetic Condition**
- **Physical Activity Level**
- **Sleep Duration**
- **Asthma**
- **Kidney Disease**
- **Heart Disesase(target value)**

These attributes are chosen as they collectively represent both physiological conditions and lifestyle behaviors that contribute to heart disease. The dataset reflects real-world health patterns by incorporating multiple risk factors.

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503256|

The data is handled using Python-based tools, where it is stored and processed efficiently for further analysis and modeling.

Data Preprocessing:

Before applying machine learning algorithms, the dataset is cleaned and transformed to improve data quality.

- Text-based categorical values are converted into numerical form for compatibility with machine learning models.
- Invalid or missing entries are removed to ensure consistency.
- The dataset is divided into training and testing sets in an 80:20 ratio.
- Only relevant features contributing to heart disease prediction are retained for model training.

Feature Engineering:

To improve prediction capability, additional meaningful features are derived from the existing dataset.

- A Risk Score is computed by combining multiple health risk indicators such as smoking, alcohol intake, diabetes, and lack of physical activity.
- BMI values are grouped into categorical ranges representing different weight conditions.
- A Sleep Risk factor is introduced to indicate insufficient sleep duration. These newly created features help in capturing deeper relationships within the data that are not directly visible from raw attributes.

Machine Learning Model:

A Random Forest classifier is utilized to predict the probability of heart disease. This algorithm is chosen due to its **robustness, ability to handle nonlinear relationships, and reduced risk of overfitting**. To address class imbalance in the dataset, the SMOTE technique is applied, which generates synthetic samples for underrepresented classes and improves model learning.

Model Evaluation Metrics:

The effectiveness of the model is measured using multiple evaluation metrics:

- **Accuracy** – Indicates overall prediction performance
- **Recall** – Ensures correct identification of heart disease cases
- **Precision and F1-score** – Provide balanced evaluation of the model. These metrics are particularly important in healthcare applications where identifying true positive cases is critical.

Cardiovascular Risk Analysis:

The dataset is further explored to understand patterns and trends associated with heart disease. The analysis includes:

- Relationship between lifestyle habits and heart disease
- Distribution of cases across different age groups
- Impact of BMI and sleep duration • Impact of Smoking & alcohol drinking on Heart Disease. This analysis provides insights into key factors affecting cardiovascular health.

Tableau Integration:

The prediction model is deployed using a Flask web framework to create an interactive user interface. The application includes multiple sections such as dashboard, prediction module, and informational pages. Users can input their personal health details, and the system processes the data to generate a risk level categorized as:

- **Low Risk**
- **Medium Risk**
- **High Risk**

This makes the system practical and user-friendly for real-time predictions.

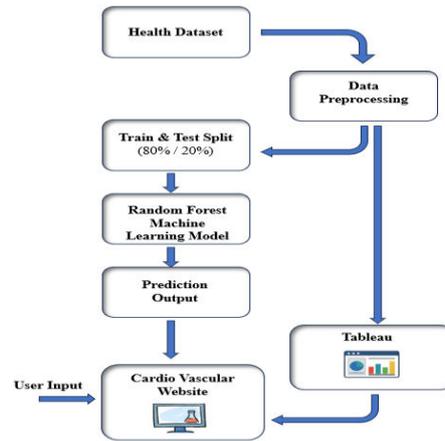


Fig 2: Block Diagram of Heart Disease Prediction Using Random Forest

IV. EXPERIMENTAL RESULTS

The proposed heart disease prediction system is evaluated using classification-based performance metrics to measure its effectiveness in identifying patients at risk. Since the problem involves predicting categorical outcomes, metrics such as accuracy, precision, recall, and F1-score are used.

1. Model Performance Metrics:

The performance of the **Random Forest model** is evaluated using standard regression evaluation metrics. These metrics measure how accurately the model predicts electricity consumption and quantify the difference between actual and predicted values.

Metric	Value
Accuracy	82.22%
Precision (Heart Disease = 1)	28%
Recall (Heart Disease = 1)	53.61%
F1-Score (Heart Disease = 1)	37%

Table 1: Classification Model Performance Metrics

2. Risk Level Analysis:

The prediction results are categorized into different risk levels based on probability thresholds. This classification helps in understanding patient risk distribution.

Risk Level	Description
Low Risk	Low probability of heart disease

Medium Risk	Moderate probability requiring attention
High Risk	High probability indicating potential heart disease

Table 2: Heart Disease Risk Classification

3. Health Pattern Analysis:

The dataset is analyzed to understand how different health and lifestyle factors influence heart disease risk.

Factor Type	Observation
Lifestyle Habits	Smoking and alcohol increase risk
Age Groups	Risk increases with age
Physical Activity	Reduces likelihood of heart disease
Sleep Patterns	Insufficient sleep increases risk

Table 3: Health Risk Factor Analysis

4. Feature Contribution Analysis:

This analysis highlights the most important features affecting the model's prediction.

Feature	Impact
Age Category	Strong influence on risk
Risk Score	Combined effect of multiple conditions
BMI	Indicates obesity-related risk
Diabetic	Major contributing factor
Physical Activity	Protective factor

Table 4: Important Features Influencing Prediction

5. Web Application Insights:

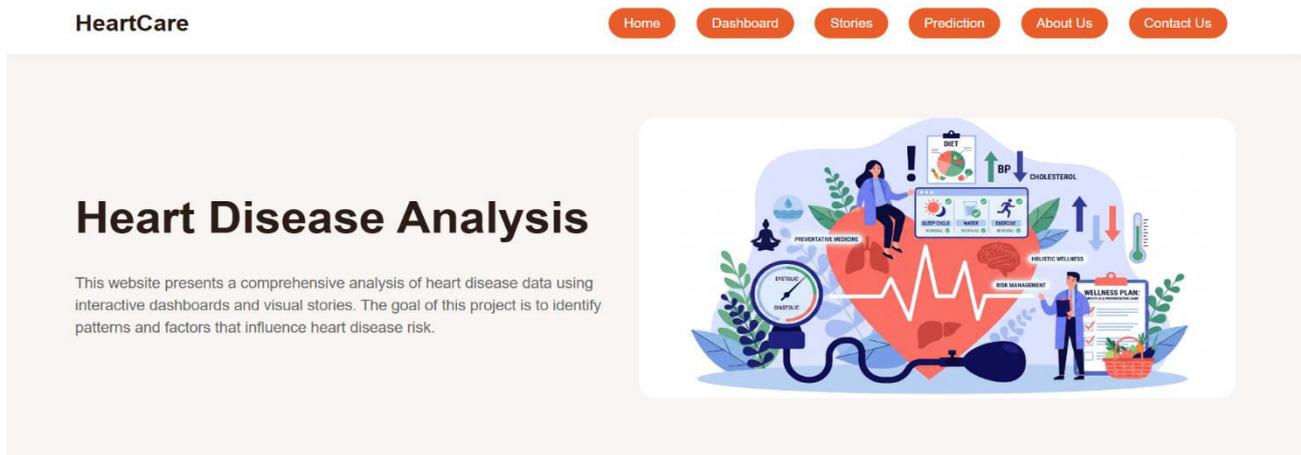
The system provides an interactive platform for users to input health data and receive predictions. The interface presents results in a simple and understandable format.

Component	Functionality
Prediction Module	Generates heart disease risk
Dashboard	Displays health insights
Monthly Consumption	Electricity demand changes across different months
User Interface	Enables easy data input and navigation
Result Page	Shows risk level and probability

Table 5 :System Interface and Insights

The user interface is developed using HTML, CSS, and JavaScript and integrated with a Flask backend for smooth interaction. The application provides easy navigation to access different sections such as prediction and dashboard.

Users can enter their health details, and the system displays the heart disease risk along with probability in a clear format. The interface is simple, user-friendly, and helps users easily understand the prediction results.



About This Project

Fig 1: Home Page

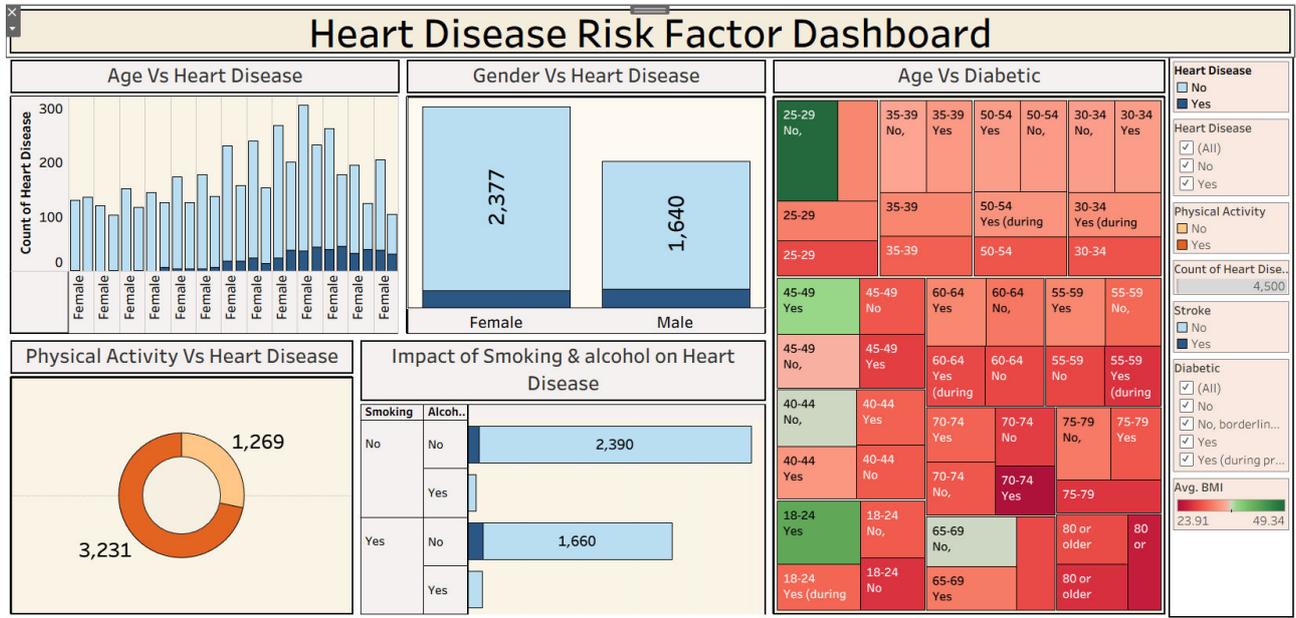
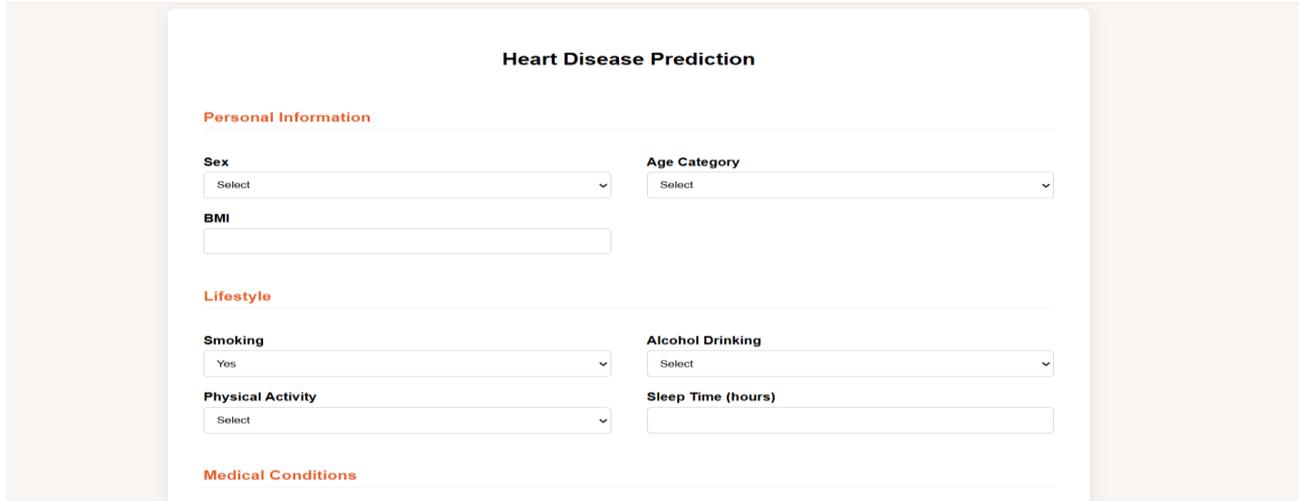


Fig 2: Tableau Dashboard Page

HeartDisease_Story



Fig 3: Tableau Story Page



Heart Disease Prediction

Personal Information

Sex: Age Category:

BMI:

Lifestyle

Smoking: Alcohol Drinking:

Physical Activity: Sleep Time (hours):

Medical Conditions



Lifestyle

Smoking: Alcohol Drinking:

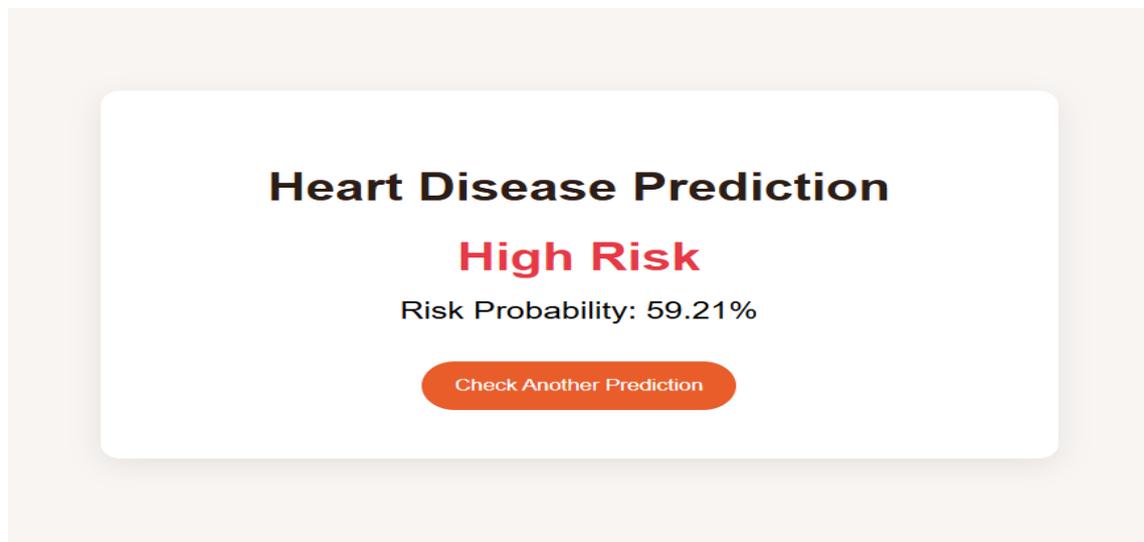
Physical Activity: Sleep Time (hours):

Medical Conditions

Stroke: Diabetic:

Asthma: Kidney Disease:

Fig 4: User Input Page



Heart Disease Prediction

High Risk

Risk Probability: 59.21%

Fig 5: Prediction Page

Volume 15, Issue 3, March 2026**|DOI: 10.15680/IJIRSET.2026.1503256|****V. DISCUSSION**

The results show that the Random Forest model is effective in identifying patterns related to heart disease risk. By using health and lifestyle features such as age, BMI, and medical conditions, the model is able to predict the likelihood of heart disease with reasonable accuracy. The use of SMOTE and threshold tuning helps improve the detection of high-risk patients.

VI. CONCLUSION

This project presents a heart disease prediction system by combining machine learning and web technologies. A Random Forest model is used to predict the risk of heart disease based on patient health data. The model is integrated into a Flask-based web application, allowing users to input their details and receive risk predictions in real time. The system provides a simple and user-friendly interface for understanding health risks. Overall, the project demonstrates how machine learning can support early detection of heart disease and assist in better healthcare decision-making.

VII. ACKNOWLEDGMENT

We would like to express our sincere thanks to our project guide and faculty members for their continuous support and guidance. We also thank the **Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning)** for providing the necessary resources to complete this project successfully.

REFERENCES

- [1] Kaggle, "Heart Disease Dataset." Available: <https://www.kaggle.com>.
- [2] L. Breiman, "Random Forests," Machine Learning Journal, 2001.
- [3] Scikit-learn Documentation. Available: <https://scikit-learn.org/>
- [4] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," 2002.
- [5] Flask Documentation. Available: <https://flask.palletsprojects.com/>
- [6] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," 2011.
- [7] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, 2019. Available: <https://archive.ics.uci.edu>
- [8] S. Rajkomar et al., "Machine Learning in Medicine," The New England Journal of Medicine, 2019.



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