IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.1411107

Heart Disease Prediction and Prevention

Swetha P[1], Bhavana S[2], Bhoomika B N[3], Gowri H R[4], Koyal M[5]

Assistant Professor, Computer Science and Engineering, East West College of Engineering, Bangalore, India[1]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[2]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[3]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[4]

Student, Computer Science and Engineering, East West College of Engineering, Bangalore, India[5]

Abstract: Heart disease remains one of the leading causes of global mortality, often due to late diagnosis and the absence of early risk assessment. Traditional diagnostic methods require clinical visits, medical equipment, and expert interpretation, which may not always be accessible. This paper presents a Machine Learning (ML)–based heart disease prediction system designed to evaluate an individual's likelihood of developing heart disease using key medical parameters. The model uses attributes such as age, cholesterol, blood pressure, fasting blood sugar, maximum heart rate, chest pain type, and other clinical indicators to generate accurate predictions. Several ML algorithms—Logistic Regression, Random Forest, KNN, and Support Vector Machine—were trained and evaluated, with Random Forest achieving the highest accuracy. A web-based interface built using Streamlit allows users to enter their health metrics and receive prediction results instantly, along with personalized health recommendations. This system is scalable, user-friendly, and promotes early prediction, ultimately supporting preventive healthcare.

Keywords: Heart Disease Prediction, Machine Learning, Random Forest, Health Monitoring System, Medical Diagnosis, Streamlit.

.

I. INTRODUCTION

Heart disease is responsible for millions of deaths globally and remains a major public health challenge. Early diagnosis is critical, yet many individuals fail to undergo timely medical evaluations due to limited awareness, accessibility issues, or the high cost of clinical examinations. Modern advancements in Machine Learning (ML) have demonstrated the potential to support medical decision-making by analyzing large datasets to detect patterns linked to cardiovascular risks.

Machine learning models can process various health-related parameters such as age, cholesterol levels, resting blood pressure, chest pain type, ECG results, and exercise-induced angina to predict the likelihood of heart disease. These models outperform conventional statistical techniques by learning complex feature correlations and improving prediction accuracy. The purpose of this study is to develop a user-friendly, web-based heart disease prediction system that assists users in understanding their cardiac risk and promoting timely preventive actions.

Heart disease prediction has become increasingly important as lifestyle changes, stress, and genetic factors continue to elevate cardiovascular risk across all age groups. With the rise of digital health technologies, machine learning offers a powerful approach for analyzing complex medical data and identifying patterns that traditional diagnostic methods may overlook. By leveraging predictive analytics, individuals can gain early insights into their heart health without requiring frequent clinical visits. This not only enhances awareness but also promotes proactive health management. The integration of machine learning into heart disease prediction thus represents a significant step toward accessible, preventive, and personalized healthcare.

The Data Flow Diagram (DFD) illustrates the overall flow of processes in the Heart Disease Prediction System, beginning from application launch to generating the final prediction results. The system ensures a smooth sequence of user authentication, data processing, machine learning–based prediction, and delivery of health precautions.
The process begins when the user starts the application and is prompted to choose between Login and Register.
If the user is new, the Register process collects details such as username, password, and email. The system validates and stores the user credentials. After successful registration, the user is redirected back to the login page.
If the user selects Login, the system verifies the credentials. When the login details are valid, the user proceeds to the health data input stage. If the login information is invalid, the system displays an error message and returns the user to the login page.

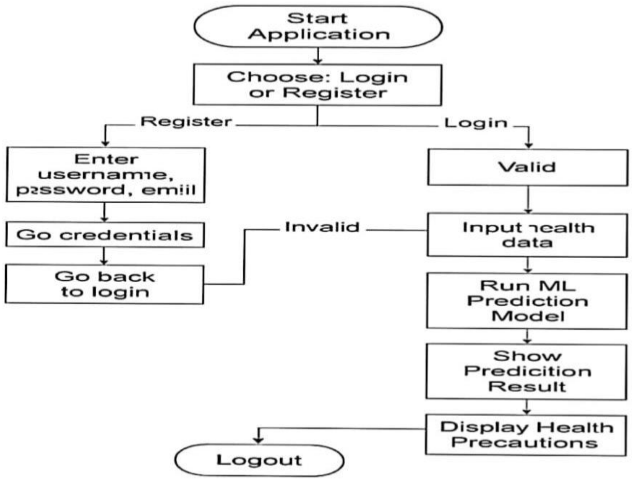
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.1411107

DATA FLOW DIAGRAM



After successful login, the user provides essential health parameters such as age, blood pressure, cholesterol, chest pain type, and other attributes required for prediction. This data is forwarded to the Machine Learning Prediction Model, which analyzes the input using trained algorithms to determine whether the user is at risk of heart disease.
Once the model processes the data, the system displays the Prediction Result to the user. Following this, the system provides Health Precautions tailored to the user's risk category. These may include lifestyle changes, diet recommendations, or preventive health measures.
Finally, the user can choose to Logout, which terminates the session and safely exits the application.
Overall, the DFD represents a structured and systematic flow of operations, ensuring secure user access, efficient data handling, accurate ML-based predictions, and helpful health guidance within the application.

II. PROBLEM DEFINATION

Heart disease remains one of the leading causes of mortality worldwide, primarily due to late detection, lack of awareness, and limited access to timely health assessments. Traditional diagnosis methods require clinical visits, medical tests, and professional evaluation, making early prediction difficult for many individuals, especially those in remote or underserved regions. Moreover, existing healthcare systems often lack personalized, data-driven tools that can evaluate an individual's health condition in real time.

Many individuals are unaware of their cardiac risk until symptoms become severe, leading to delayed intervention and increased chances of complications. Furthermore, conventional risk assessment tools are static and cannot integrate multiple health parameters dynamically to provide personalized predictions. There is a need for a system that allows users to easily register, securely log in, enter their personal health information, and receive an accurate heart disease prediction using machine learning techniques.

The proposed Heart Disease Prediction System addresses this challenge by providing a web-based, user-friendly platform that offers early risk assessment through a trained machine learning model. Users can input key health metrics such as age, blood pressure, cholesterol level, chest pain type, and other medical parameters, which are processed to determine the likelihood of heart disease. In addition to prediction, the system provides precautionary recommendations to help users take preventive measures.

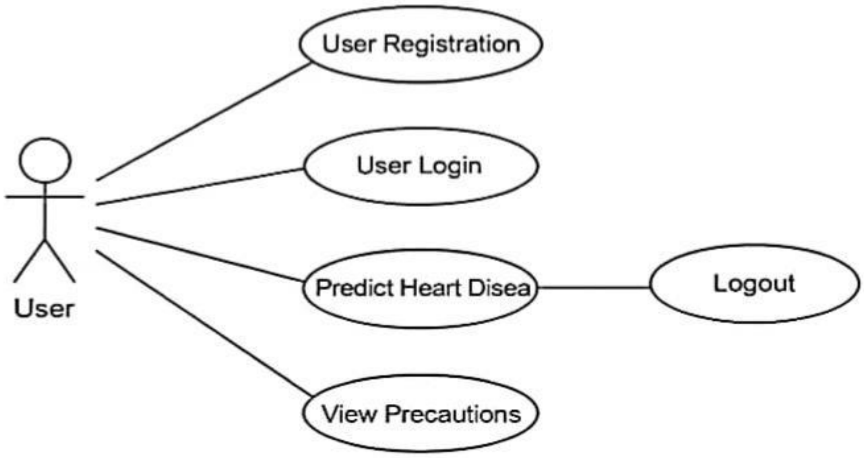
Thus, the core problem addressed by this system is the absence of an accessible, efficient, and personalized digital tool that can support early detection and awareness of heart disease risk. The proposed solution fills this gap by integrating user authentication, data input, machine learning prediction, and preventive guidance into a single seamless applications.

III. USE CASES AND USER SCENARIOS

Use Cases

The Use Case Diagram illustrates the interaction between the User and the core functionalities provided by the Heart Disease Prediction System. The user is the primary actor who engages with the system to perform various tasks related to registration, login, disease prediction, and accessing health precautions.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025
DOI: 10.17148/IJARCCE.2025.1411107



The first use case, User Registration, allows new users to create an account by providing essential details such as username, email, and password. Once registered, the user proceeds to the User Login use case, where they authenticate themselves using their credentials. Successful login enables the user to access the system's main functionalities.

After logging in, the user can utilize the Predict Heart Disease use case. In this process, the user inputs their health-related parameters, which are processed by the system's machine learning model to predict the likelihood of heart disease. Based on the prediction result, the system offers tailored health advice, which the user can access through the View Precautions use case. This includes recommendations on lifestyle changes, diet, exercise, and preventive measures.

Finally, the Logout use case allows the user to securely exit the system and end their session. Overall, the Use Case Diagram provides a structured view of how users interact with the system and the sequence of actions involved in performing key operations.

User Scenarios

Scenario 1: New User Registration
A new user opens the application and selects the User Registration option. They enter their username, password, and email address. The system validates the information and creates a new account. The user is then redirected to the login page to access the system.

Scenario 2: Returning User Login
An existing user starts the application and chooses the User Login option. They enter their registered credentials. If the information is valid, the system grants access to the main dashboard; otherwise, an error message is displayed, prompting the user to retry.

Scenario 3: Predicting Heart Disease
After logging in, the user navigates to the Predict Heart Disease section and inputs various medical parameters such as age, blood pressure, cholesterol level, chest pain type, and other required factors. The system processes the input using the trained ML model and displays the prediction—either indicating risk or no risk of heart disease.

Scenario 4: Viewing Health Precautions
Once the prediction is generated, the user selects the View Precautions option. The system presents personalized recommendations based on the prediction result. These may include diet suggestions, exercise advice, and general heart-healthy practices.

Scenario 5: Logout
After completing their tasks, the user chooses the Logout option. The system securely ends the user session and returns to the main application start page.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.1411107

IV. TECHNICAL IMPLEMENTATION

The technical implementation of the Heart Disease Prediction and Prevention System involves a structured pipeline that integrates machine learning, data processing, user authentication, and a web-based interface to deliver accurate and user-friendly predictions. The system is built using Python, Streamlit, and a trained machine learning model, ensuring reliability, accessibility, and real-time performance.

The implementation begins with the development of the machine learning model, which is trained using publicly available datasets such as the Cleveland Heart Disease Dataset. The dataset undergoes preprocessing steps including handling missing values, normalization of numerical attributes, and encoding of categorical variables. Several algorithms—such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine—are evaluated, and the model with the highest accuracy is chosen. The final trained model is serialized using pickle and stored for integration with the web application.

The frontend interface is developed using Streamlit, providing users with an intuitive platform to input health parameters, view results, and access health precautions. Streamlit components such as number inputs, dropdown menus, and buttons are used to collect user medical details including age, blood pressure, cholesterol levels, chest pain type, and ECG readings. Once the user submits their inputs, the system processes them in the backend and passes the data to the loaded machine learning model for prediction.

A user authentication system Is implemented to enhance security and personalization. New users can register by providing a username, password, and email, while returning users can log in using stored credentials. Passwords are hashed using SHA-256 to ensure data security, and user information is stored in a local file or database.

The backend logic handles data conversion, model integration, and output generation. Upon prediction, the system displays whether the user is at risk of heart disease and provides personalized precautionary measures such as diet advice, exercise recommendations, and lifestyle modifications. The application also includes a Healthy Heart Strategies section that guides users on preventive healthcare practices.

Finally, the entire system is deployed as a lightweight web application capable of running on local servers or cloud platforms. Its modular design ensures easy maintenance, future scalability, and integration of more advanced models or real-time sensors. Through this seamless integration of machine learning and interactive UI components, the system provides an efficient and accessible tool for early heart disease prediction.

V. LITERATURE REVIEW

Cardiovascular diseases (CVDs) remain the leading cause of death worldwide, making early prediction and prevention critically important. Traditional diagnostic methods such as ECG, stress tests, and blood examinations, although effective, are often time-consuming, expensive, and inaccessible to individuals in remote regions. Recent research has therefore shifted toward computational approaches, particularly machine learning (ML), which can analyze medical datasets to discover hidden patterns and improve prediction accuracy.

One of the earliest and most influential works in this domain is the Framingham Heart Study (1948–present), which identified major risk factors such as hypertension, high cholesterol, smoking, obesity, and physical inactivity. Machine learning algorithms including decision trees and neural networks have been applied to this dataset to predict long-term heart disease risk.

The Cleveland Heart Disease Dataset is one of the most widely used datasets for building predictive models. Numerous studies have applied algorithms such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forests, demonstrating high predictive accuracy for identifying high-risk patients. These studies validate the potential of ML-based decision support tools in clinical diagnosis.

According to K. Sharma et al. (2020), Logistic Regression, Decision Trees, Random Forest, KNN, and SVM have shown strong performance when predicting heart disease risk. Their comparative analysis highlights that ensemble techniques such as Random Forest and Gradient Boosting outperform traditional methods, providing better generalization and accuracy.

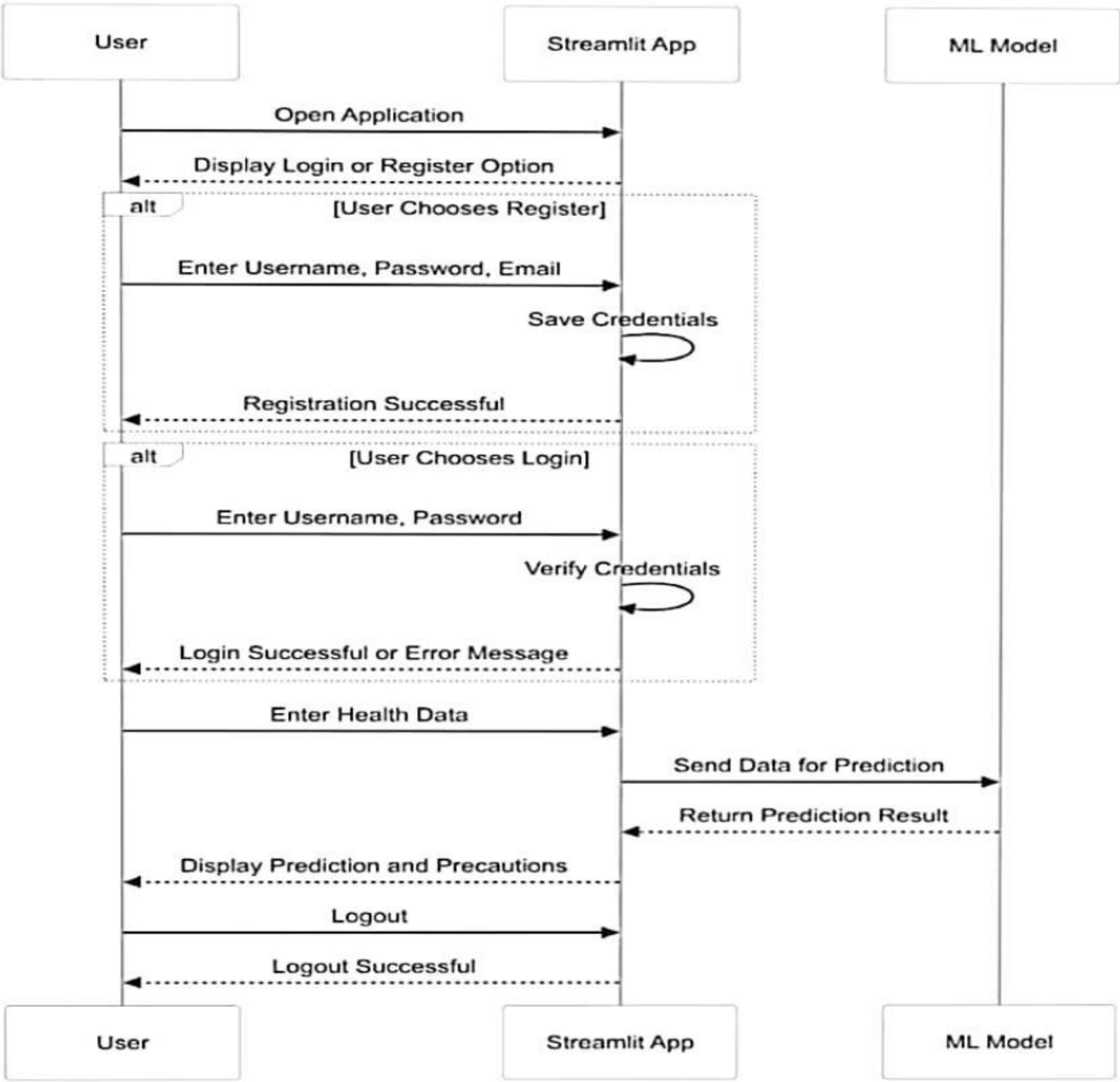
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.1411107

VI. EVALUATION AND RESULTS



The sequence diagram provides a comprehensive representation of the interactions and message flow between the User, the Streamlit Application, and the Machine Learning (ML) Model throughout the entire operational lifecycle of the Heart Disease Prediction System. It outlines the chronological execution of events, beginning from application launch, followed by authentication procedures, data entry, prediction processing, and finally system logout.

The sequence initiates when the user opens the application interface. Upon launch, the Streamlit application immediately displays the initial authentication screen, presenting two primary actions: Login or Register. At this point, the system waits for the user to make a choice. If the user selects the Register option, they are prompted to enter essential credentials such as a username, password, and email address. Once the user provides these details, the Streamlit application processes the registration request by storing the credentials securely—typically after hashing the password—and then confirms the successful registration back to the user.

Alternatively, if the user chooses the Login option, they are required to enter their existing username and password. The Streamlit application forwards these credentials to its authentication module, which verifies them against the stored user database. Based on the outcome, the system either grants access by displaying a “Login Successful” message or notifies the user of incorrect credentials through an error message. This decision-making flow is represented in the diagram using

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.1411107

an "alt" fragment to differentiate between the successful and unsuccessful login paths.
After successful authentication, the user proceeds to the next stage where they input their health-related data. The system prompts the user for all required medical parameters such as age, cholesterol values, resting blood pressure, chest pain type, fasting blood sugar, ECG readings, maximum heart rate, and other relevant features used in the prediction model. Once all data fields are filled, the Streamlit application compiles and structures the input into a format compatible with the ML model.

The Streamlit application then sends the structured health data to the Machine Learning Model for prediction. The ML model receives this data, processes it using the trained algorithm, and determines whether the user is likely to have heart disease. The model subsequently returns the prediction result to the Streamlit application. This back-and-forth communication between the application and the ML model is represented in the diagram through synchronous message exchanges.

Upon receiving the model's result, the Streamlit application displays the outcome in a clear and user-friendly manner. If the model predicts a potential risk of heart disease, the application also provides additional health precautions, including dietary recommendations, exercise guidelines, lifestyle modifications, and preventive measures. These recommendations are shown only after the prediction is generated, ensuring contextual relevance for users.

Finally, when the user chooses to log out, the Streamlit application processes the logout request, terminates the user session, and confirms a successful logout on the interface. This marks the completion of the system workflow. The sequence diagram thus captures each interaction meticulously, highlighting how user actions trigger system processes, and how the Streamlit application acts as the central coordinator between user input and ML-based decision-making. This detailed sequence improves understanding of the real-time operational behavior of the system and validates the correctness of the workflow implementation.

VII. CONCLUSION

The Heart Disease Prediction System developed in this project demonstrates the effectiveness of integrating machine learning techniques with a user-friendly web-based interface to provide early and accessible cardiac risk assessment. By utilizing key medical parameters and a trained predictive model, the system offers quick, reliable results that help users understand their likelihood of developing heart disease. The application not only identifies potential risk but also provides personalized health precautions, enabling individuals to take proactive steps toward improving their cardiovascular health.

The implementation of a secure authentication system, combined with an intuitive Streamlit interface, ensures smooth user interaction and enhances ease of use for individuals with diverse technical backgrounds. The model's real-time prediction capability, together with an organized presentation of health strategies, highlights the system's potential to support preventive healthcare. While traditional diagnostic methods can be costly and time-consuming, this system offers a cost-effective and readily accessible alternative for preliminary screening.

Overall, the project successfully fulfills its objective of offering an intelligent, efficient, and practical solution for heart disease prediction. Future enhancements such as integrating real-time sensor data, adopting more advanced deep learning models, or deploying the system on cloud platforms will further improve its accuracy, scalability, and applicability within real-world healthcare environments

REFERENCES

[1]. J. Nielsen, "Usability 101: Introduction to Usability," Nielsen Norman Group, 2023.
[2]. Interaction Design Foundation, "What is User Interface Design?" Interaction Design Foundation, 2023.
[3]. Flask Software Foundation, "Flask: A Python Microframework," Flask Documentation, 2023.
[4]. Streamlit Inc., "Streamlit: The Fastest Way to Build and Share Data Apps," Streamlit Documentation, 2023.
[5]. S. Patel, "Building a Heart Disease Prediction Model Using Machine Learning," Towards Data Science, 2020.
[6]. P. Shrivastava and M. Sharma, "A Review on Heart Disease Prediction System," ResearchGate, 2020.

