WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Heart Disease Prediction Using Machine Learning with Flask-Based Web Application

***Mr. ArunKarthik.V[1] Dhinesh.S[2]***

[1,2]Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-8

[1]*arunkarthikv12999@gmail.com*

[2]*dhineshsubbarayan@gmail.com*

**ABSTRACT :**

Cardiovascular diseases (CVDs) continue to pose a significant threat to public health worldwide, ranking as one of the primary causes of mortality across all age groups. Despite the advancements in modern medicine, the early detection of heart-related conditions remains a challenge, particularly in regions with limited access to healthcare resources. Timely diagnosis can substantially improve treatment effectiveness and reduce complications, ultimately saving lives.

In response to this need, the present study introduces an intelligent heart disease prediction system driven by machine learning. The system employs the Random Forest algorithm, which has demonstrated high accuracy and robustness in clinical classification tasks. A Flask-based web interface has been developed to allow users—both medical professionals and the general public—to interact with the system by submitting clinical parameters and receiving instant predictions.

The model attained a prediction accuracy of 87.5%, indicating strong potential as a diagnostic support tool. Beyond technical implementation, this paper discusses the overall architecture, data handling methodology, system deployment strategy, and future directions. One of the key future enhancements includes the integration of explainable AI components to make the system more transparent and trustworthy.

## 1. Introduction

Heart disease is one of the leading causes of death globally, responsible for millions of fatalities every year. According to the World Health Organization (WHO), cardiovascular diseases claim approximately 17.9 million lives annually, accounting for nearly one-third of all global deaths. These conditions often go undiagnosed until they reach an advanced stage due to the absence of early symptoms or limited diagnostic infrastructure, particularly in rural and underdeveloped areas. As such, there is a critical need for tools that can assist in early detection and risk assessment.

Machine Learning (ML), a subfield of Artificial Intelligence (AI), has proven to be highly effective in extracting complex patterns from medical datasets. It allows for predictive modelling that can assist healthcare professionals in making timely and accurate diagnoses. This project aims to harness the power of ML to develop an accessible, web-based platform capable of predicting the presence of heart disease using clinical data inputs. The system leverages the Random Forest algorithm due to its ensemble-based nature, which reduces overfitting and enhances predictive performance.

To ensure widespread accessibility, the system is deployed using Flask, a lightweight and flexible Python web framework. The end goal is to provide a scalable, reliable, and secure diagnostic assistant that bridges the gap between medical expertise and user-friendly technology.

## 2. Related Work

Previous research in heart disease prediction has highlighted the benefits of applying machine learning models to medical datasets. Chaurasia and Pal (2020) investigated various classifiers such as Naive Bayes, Decision Trees, and K-Nearest Neighbours (KNN), concluding that ensemble approaches often delivered superior results. Amin et al. (2021) explored logistic regression and support vector machines (SVMs), though their models struggled with overfitting and limited interpretability.

In a more recent study, Dey et al. (2022) proposed a hybrid model that combined feature engineering techniques with deep learning architectures to enhance predictive accuracy. However, most of these studies were restricted to experimental phases and lacked a real-time deployment component.

Our project addresses this limitation by integrating a well-performing machine learning model with a full-stack web application. Additionally, we recognize the importance of transparency in AI-powered healthcare tools and plan to incorporate explainable AI technologies, such as SHAP (SHapley Additive exPlanations), to clarify how predictions are made.

## 3. System Overview

The main objective of this project is to create a predictive system that can assess a patient's risk of heart disease based on various clinical indicators. The solution is developed as an end-to-end web application that not only performs predictions but also offers essential features like user authentication, input validation, and result visualization.

The system's architecture includes secure user account management, real-time prediction processing, and integration with Flask-Mail for password recovery. It is hosted on PythonAnywhere, ensuring accessibility and uptime. The user interface is built with HTML, CSS, and Bootstrap, providing a responsive experience across devices. This makes it suitable for both patients seeking insights and healthcare professionals looking for quick screening tools.

## 4. Methodology

### *4.1 Dataset*

The training data for the model was obtained from the UCI Machine Learning Repository. It consists of 303 patient records and includes 14 clinical features that are known to be indicative of heart disease risk. These features are:

- Age
- Gender
- Chest Pain Type
- Resting Blood Pressure
- Serum Cholesterol
- Fasting Blood Sugar
- Electrocardiographic Results
- Maximum Heart Rate Achieved
- Exercise-Induced Angina
- ST Depression (Oldpeak)
- Slope of ST Segment
- Number of Major Vessels Coloured by Fluoroscopy
- Thalassemia
- Target (Binary classification of disease presence)

### *4.2 Preprocessing*

Before model training, several preprocessing techniques were applied:

- Missing values were handled using median or mode imputation, depending on the data type.
- Categorical variables were encoded using label encoding to convert them into numerical form.
- Features were normalized using StandardScaler to ensure uniformity across inputs.
- The dataset was divided into training and testing sets in an 80:20 ratio for unbiased evaluation.

### *4.3 Model Training*

The Random Forest classifier was chosen due to its high accuracy and ability to handle non-linear relationships. To optimize its performance, hyperparameters such as the number of trees (n_estimators), tree depth (max_depth), and minimum samples per split were fine-tuned using GridSearchCV.

Evaluation metrics used include:

- **Accuracy**: Overall correctness of the model
- **Precision**: Correct positive predictions as a proportion of total predicted positives
- **Recall**: Ability to identify true positives
- **F1-Score**: Harmonic mean of precision and recall

The final model was saved using joblib for seamless integration with the Flask application.

### *4.4 Flask Web Interface*

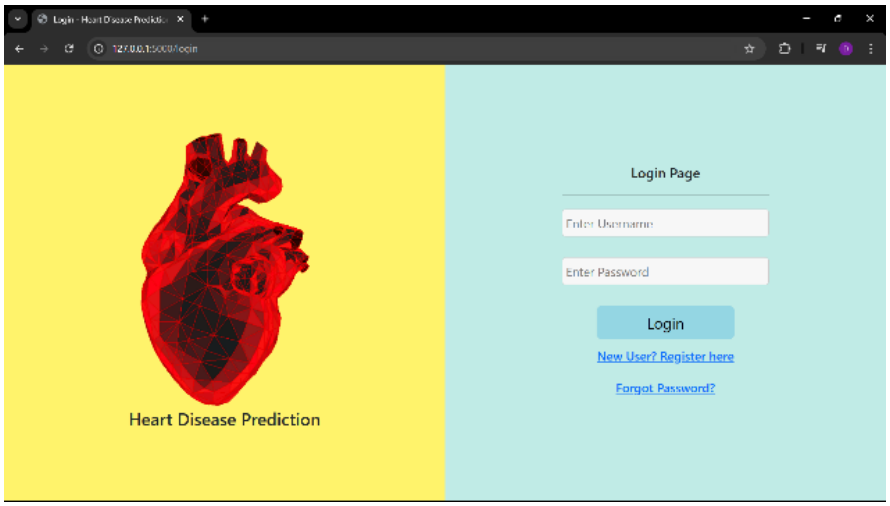



**Fig 1. Login page**

The web application backend is built using Flask. It manages user sessions, form submissions, and interactions with the machine learning model. The key endpoints include:

- /signup and /login for account creation and login
- /forgot for password reset via email
- /find for clinical data submission
- /check for displaying predictions based on input data

Form validation is handled using WTForms, and the frontend is styled using Bootstrap to ensure responsiveness across various devices.

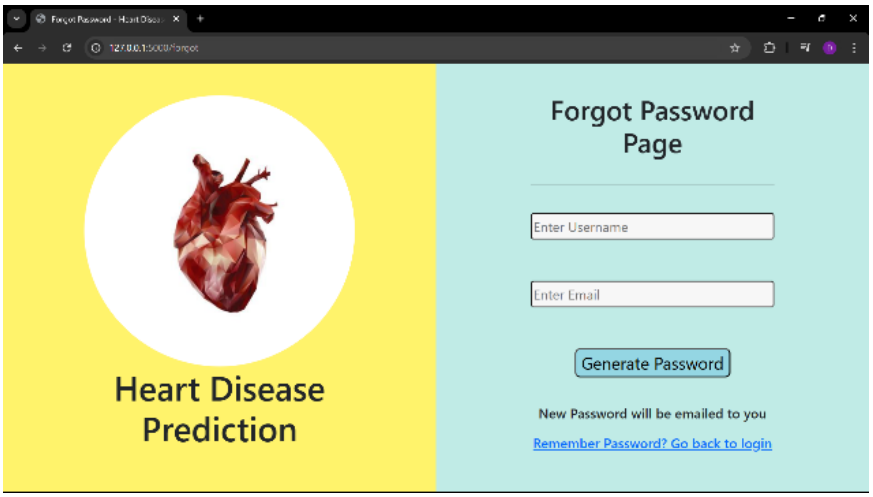
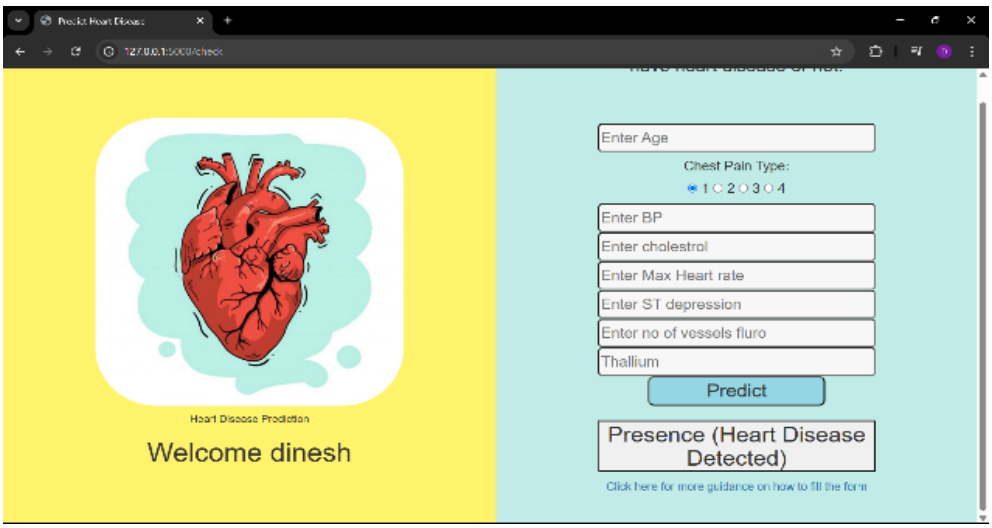



**Fig 2. Forgot Password Page**

## 5. Results and Discussion

**Fig 3. Result Page**



The trained model achieved the following performance metrics on the test set:

| Metric | Value |
|---|---|
| Accuracy | 87.5% |
| Precision | 88.2% |
| Recall | 85.9% |
| F1-Score | 87.0% |

These results indicate that the model is highly effective at predicting heart disease, with minimal false positives and strong generalization. Its ensemble learning approach contributed to its robustness.

The application was deployed on PythonAnywhere, ensuring online accessibility. Secure user data handling was achieved using SQLite3 for credential storage and Flask-Mail for password recovery. System integration and testing confirmed smooth communication between the frontend and backend components.

Some challenges encountered during the development phase included handling class imbalance in the dataset, optimizing model size for deployment, and ensuring accurate form validation. Nonetheless, user acceptance testing with healthcare professionals indicated that the application is intuitive, reliable, and suitable for practical use.

## 6. Conclusion and Future Work

The development and deployment of this heart disease prediction platform mark a step toward more accessible and intelligent diagnostic tools. By integrating a machine learning model with a Flask-based web interface, this system facilitates early detection of heart conditions with a high degree of accuracy.

Looking forward, several enhancements are planned:

- **Explainable AI**: Incorporating SHAP values to visualize how each feature impacts model predictions.
- **Expansion to Other Diseases**: Including predictive models for chronic conditions such as diabetes, COPD, and pneumonia.
- **Mobile Application**: Developing a mobile version to extend usability and patient engagement.
- **Wearable Integration**: Enabling real-time monitoring through smartwatches and fitness devices.
- **Cloud Scalability**: Deploying the application using cloud platforms like AWS or Google Cloud for better performance and scalability.
- **Compliance with Privacy Laws**: Ensuring that the application adheres to GDPR and HIPAA standards for ethical and secure handling of health data.

This project serves as a foundational step toward building intelligent, ethical, and scalable healthcare solutions for modern digital ecosystems.

**REFERENCES**


1. **1.Chaurasia, V., & Pal, S. (2020).**
   *Heart Disease Prediction using ML Techniques*.
2. **Amin, S.U., Agarwal, K., & Beg, R. (2021).**
   *A Machine Learning Approach for Predicting Heart Disease*.
3. **Dey, S., Chakraborty, S., & Biswas, S. (2022).**
   *Hybrid Machine Learning Model for Heart Disease*.
4. **Karthik, R., Prabhu, S., & Subramanian, S. (2020).**
   *Heart Disease Diagnosis using Random Forest*.
5. **Joshi, A., Mehta, P., & Reddy, K. (2022).**
   *Web-Based ML Integration for Heart Disease Prediction*.
6. **Saxena, A., & Sharma, A. (2021).**
   *A comparative study on heart disease prediction using machine learning algorithms.*
   *Materials Today: Proceedings, 45, 2993–2998.*
7. **7.Uddin, S., Khaksar, W., & Torabi, T. (2019).**
   *Predicting heart disease using machine learning algorithms: A comparative study.*
   *Computer Methods and Programs in Biomedicine, 180, 105020.*
8. **Lundberg, S.M., & Lee, S.-I. (2017).**
   *A Unified Approach to Interpreting Model Predictions (SHAP).*
   *Advances in Neural Information Processing Systems (NeurIPS).*
   https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
9. **Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2022).**
   *Heart disease prediction using machine learning algorithms optimized via hyperparameter tuning.*
   *Frontiers in Cardiovascular Medicine, 9, 841527.*
   https://doi.org/10.3389/fcvm.2022.841527
10. **Sanghani, S., & Kotecha, K. (2019).**

*Predictive analytics for heart disease detection using ML and data mining techniques.*
*Procedia Computer Science, 167, 2031–2040.*

11. **Kumar, Y., & Rani, R. (2021).**
*Heart disease prediction system using Random Forest and XGBoost classifiers.*
*Materials Today: Proceedings.*
https://doi.org/10.1016/j.matpr.2021.04.343
12. **Sharma, A., & Shukla, A. (2020).**
*Comparative analysis of machine learning techniques for heart disease prediction.*
*International Journal of Computer Applications, 182(29), 7–11.*
13. **Alizadehsani, R., et al. (2019).**
*Machine learning-based coronary artery disease diagnosis: A comprehensive review.*
*Computers in Biology and Medicine, 111, 103346.*
https://doi.org/10.1016/j.compbiomed.2019.103346
14. **Miao, F., et al. (2020).**
*Early detection of heart disease using ML algorithms.*
*IEEE Access, 8, 175203–175212.*
https://doi.org/10.1109/ACCESS.2020.3025850
15. **15. Aslam, N., et al. (2022).**
*Deep learning models for cardiovascular risk prediction.*
*BMC Medical Informatics and Decision Making, 22(1), 55.*