A Machine Learning-Based Approach to Early Heart Disease Prediction with HeartSathi .

Prof. Meghraj. B. Chougule[1*],Dilip . Y . Nandiwale[2], Pranita Y. Tathe[2],
Shivani L. Patil[2], Sonali S. Patil[2], Gunjan S. Shivsharan[2]
(Team HeartSathi)
[1]Assistant Professor, Department of Computer Science and Engineering, Brahmdevdada Mane Institute of Technology, Solapur, Maharashtra, India
[2]UG Student, Department of Computer Science and Engineering, Brahmdevdada Mane Institute of Technology, Solapur, Maharashtra, India

[*]Corresponding Author: nandiwaledilip@gmail.com

Abstract:

Heart disease is one of the leading causes of mortality globally, often due to late diagnosis or lack of timely medical intervention. In this study, we present **HeartSathi**, a machine learning-based heart disease prediction model designed to assist healthcare providers in early diagnosis. By analyzing key clinical parameters such as age, blood pressure, cholesterol levels, heart rate, and other vital indicators, the system predicts the likelihood of a patient having a heart condition. The model leverages well-established algorithms including Logistic Regression, Random Forest, and Support Vector Machines (SVM) for accurate classification. Extensive testing was performed using public datasets such as the Cleveland Heart Disease dataset. Results show that HeartSathi can effectively predict heart disease risk with a high accuracy rate, thus demonstrating its potential as a supportive diagnostic tool for doctors and medical practitioners. This system contributes toward the broader goal of intelligent healthcare solutions aimed at improving patient outcomes and saving lives.

Keywords:

Heart Disease Prediction, Machine Learning, HeartSathi, Logistic Regression, Random Forest, SVM, Healthcare AI, Early Diagnosis, Medical Data Analysis, Predictive Modeling

1. Introduction:

Heart disease remains the foremost cause of death across the globe, accounting for nearly 17.9 million deaths annually, as reported by the World Health Organization (WHO). Despite significant advancements in medical science, early detection and intervention remain crucial in preventing fatal outcomes. Traditional diagnostic methods are time-consuming, expensive, and require expert interpretation, which may not always be accessible, especially in rural or underdeveloped regions. This creates a need for intelligent, automated systems that can support healthcare professionals in identifying heart disease risks at an early stage.

The advent of artificial intelligence and machine learning in the medical field has opened new avenues for disease prediction and prevention. Machine learning algorithms can analyze vast datasets, learn patterns from clinical features, and provide highly accurate predictions that assist in clinical decision-making. In this context, our project, **HeartSathi**, is designed as a smart diagnostic assistant to predict the likelihood of heart disease using supervised machine learning techniques.

HeartSathi incorporates patient data—such as age, gender, blood pressure, cholesterol level, blood sugar level, chest pain type, and more—to make real-time predictions on heart disease probability. It utilizes well-known machine learning algorithms including Logistic Regression, Random Forest, and Support Vector Machines (SVM), selected for their high classification performance in healthcare-related tasks.

The model has been trained and tested using the Cleveland Heart Disease dataset, one of the most recognized and widely used datasets in the field of cardiovascular research. Extensive preprocessing, feature selection, and evaluation were carried out to ensure the accuracy and robustness of the model. The results show that HeartSathi not only achieves high prediction accuracy but also provides interpretability in terms of which features contribute most significantly to heart disease risk.

Ultimately, the goal of HeartSathi is to serve as a practical tool for early diagnosis, particularly in primary care settings, and to complement the decision-making process of healthcare providers. By enabling proactive care and reducing the burden on medical systems, HeartSathi aims to improve patient outcomes and promote preventive healthcare across communities.

2. Related Work

Numerous studies have been conducted in recent years on the application of machine learning for heart disease prediction. Researchers have used datasets such as the Cleveland Heart Disease dataset to evaluate various classification algorithms. Algorithms like Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) have shown promising results in terms of accuracy and reliability.

For example, Detrano et al. worked with the Cleveland dataset to explore the effectiveness of decision-based algorithms. Another study by Gudadhe et al. demonstrated the effectiveness of Support Vector Machines in classifying cardiovascular disease with high accuracy. More recent advancements have focused on ensemble methods and deep learning for enhanced performance, though such methods often require more computational resources.

These works collectively indicate that machine learning provides a strong foundation for building intelligent diagnostic systems. However, there is still a need for accessible and interpretable tools like HeartSathi that are tailored for practical deployment in real-world clinical environments.

2.1 Machine Learning and AI in Healthcare

There have been various studies showcasing the effective deployment of ML methodologies in the forecasting and control of healthcare states. Mahant (2025) presents the interfacing of ML with IoT sensors for monitoring and detecting abnormal health conditions among chronic patients [2]. Mulani et al. (2025) also studied IoMT (Internet of Medical Things) architectures for heart disease prediction, which confirmed the validity of ensemble ML models such as Random Forests in health diagnostics [3]. Liyakat et al. (2025) proved the utility of machine learning to identify health irregularities based on physiological inputs in cloud-integrated systems [4].

2.2 Social Media and Behavioral Analytics

While traditional heart disease prediction models primarily rely on clinical and physiological parameters, emerging research emphasizes the growing impact of behavioral patterns and lifestyle indicators in cardiovascular risk assessment. Social media platforms, wearable devices, and online health communities generate vast amounts of user-generated data that reflect lifestyle choices, stress levels, physical activity, sleep habits, and even dietary patterns.

Behavioral analytics involves the study of user behaviors to draw meaningful insights about health risks. For instassnce, frequent mentions of stress, fatigue, lack of exercise, smoking, or alcohol use in social media posts can indirectly indicate a higher risk of cardiovascular problems. Similarly, fitness tracking apps provide quantifiable metrics such as step count, heart rate variability, and sleep quality—all of which correlate with heart health.

Integrating social media and behavioral analytics into predictive models like HeartSathi has the potential to enhance accuracy and personalize risk prediction further. It opens the door to passive monitoring and early intervention without requiring active clinical visits. However, such integration requires careful attention to ethical concerns such as data privacy, consent, and security.

While HeartSathi currently focuses on clinical data, future iterations may incorporate behavioral insights from non-clinical sources to provide a more holistic and proactive approach to heart disease prediction.

2.3 Speech Pattern Recognition

Speech pattern recognition is an emerging frontier in health diagnostics, with growing evidence that vocal characteristics can reveal underlying medical conditions, including cardiovascular issues. Changes in tone, breathiness, pitch, and speaking rate may be associated with stress, fatigue, or even reduced cardiac function.

Recent studies have explored how voice biomarkers—such as prolonged pauses, shortness of breath during speech, or vocal tremors—can indicate heart disease risk. For example, patients with congestive heart failure often exhibit altered respiratory and vocal patterns that can be detected through acoustic analysis.

Although HeartSathi currently does not utilize speech-based input, the integration of speech pattern recognition could allow for non-invasive, continuous monitoring and early detection, especially in telehealth settings. Future versions of HeartSathi may incorporate this capability to further enhance its predictive power and accessibility.

2.4 Wearable Sensors and Internet of Things Integration

Wearable products such as smartwatches and fitness bands allow users to monitor physiology while awake and asleep. These products enable monitoring of metrics like heart rate variability, sleep quality, physical activity levels, all of which have an impact on mental health. Kazi (2025a) and Tamboli et al. (2024) showed how IoT-driven architectures of wearable enable health system predictive models [10][11]. Veena et al. (2023) put forth an end-to-end secure IoT-cloud hybrid framework to exchange patient information and forecast disease status based on ML, aligning with the backend approach employed by heart Sathi

3. System Architecture:

The HeartSathi system is designed using a modular architecture to ensure scalability, efficiency, and accuracy in predicting heart disease. The architecture consists of five major components:

1. Data Collection Layer

This layer gathers patient information, which includes clinical parameters such as:

- Age
- Sex
- Chest Pain Type
- Resting Blood Pressure
- Cholesterol

- Fasting Blood Sugar
- Resting ECG
- Maximum Heart Rate Achieved
- Exercise-Induced Angina
- ST Depression
- Number of Major Vessels
- Thalassemia

Data is collected via a secure web-based interface or uploaded datasets in CSV format.

2. Data Preprocessing Layer

Before model training or prediction, raw data undergoes the following preprocessing steps:

- **Handling Missing Values**
- **Normalization or Scaling**
- **Feature Selection and Encoding**
- **Outlier Removal**

Cleaned data ensures higher model accuracy and consistency.

3. Model Training and Evaluation Layer

This layer implements supervised machine learning algorithms. The primary models used are:

- **Logistic Regression**
- **Random Forest Classifier**
- **Support Vector Machine (SVM)**

Training is performed using the Cleveland Heart Disease dataset. The dataset is split into training and test sets (e.g., 80-20 ratio), and performance is evaluated using accuracy, precision, recall, and F1-score metrics.

4. Prediction Layer

This is the core of the system where the trained model takes new user inputs and predicts whether the user is at risk of heart disease. The output is binary:

- **1 = At Risk**
- **0 = Not at Risk**

Confidence scores (probabilities) are also shown to indicate the model's certainty.

5. Frontend & User Interface Layer

The frontend is built with a simple and user-friendly interface:

- Collects input data from users.
- Displays prediction results instantly.

- Includes visualization (e.g., pie charts, health tips).
- Secure and responsive design compatible with mobile and desktop platforms

6)code layer

🧠 **Architecture Overview (Flask + ML Model + HTML UI)**

1. User Interface (Frontend)

- HTML form (index.html)
- User inputs data (Age, Gender, etc.)
- Submits form via POST request

2. Flask Backend (Server)

- Receives POST request at /predict
- Extracts form data
- Converts data to pandas DataFrame
- Scales features using scaler.pkl
- Feeds scaled data to rf_model.pkl
- Gets prediction + probability
- Fetches statement from prediction_statements.pkl
- Returns result as JSON

3. Machine Learning Layer

- Pre-trained Random Forest model (rf_model.pkl)
- Scaler for normalization (scaler.pkl)
- Custom prediction statements for each class

4. Response

- JSON containing:
    - Prediction (0 or 1)
    - Statement (e.g., "No heart disease risk")
    - Probability score

🔁 **Optional Add-ons (Future-Ready)**

- Replace HTML with React or mobile UI
- Replace Flask with FastAPI for better async support

- Add authentication layer
- Store predictions in database
- Add model monitoring and logging

4. Methodology

The following methodology was adopted to design and implement a heart disease prediction system:

1. **Data Collection**: The dataset was obtained from a trusted medical repository and includes features such as age, sex, blood pressure, cholesterol, and other cardiovascular indicators.

2. **Data Preprocessing**: Data cleaning was performed using Python. This included handling missing values, encoding categorical variables, and normalizing numerical data to ensure model compatibility.

3. **Feature Selection**: Key predictive features were selected using correlation matrices and domain knowledge to enhance model performance.

4. **Model Development**: Machine learning models were developed and trained in Python using libraries like Scikit-learn and Pandas. Multiple algorithms were tested to identify the most accurate and efficient model.

5. **Backend Integration**: The trained model was deployed via a Flask API in the backend, which interacts with a Node.js server to manage routing, requests, and user authentication.

6. **Database Management**: PostgreSQL was used to store user data, prediction history, and system logs. Proper indexing and schema design were applied for optimized performance.

7. **Model Evaluation**: Evaluation was carried out using accuracy, precision, recall, and F1-score. Cross-validation techniques ensured the model's reliability.

8. **Web Deployment**: The entire system was deployed on the HeartSathi.online platform, enabling users to input health data and receive real-time predictions through an intuitive web interface

5. Implementation

The heart disease prediction system was implemented using a full-stack architecture that integrates frontend, backend, machine learning, and database components. The system is designed for real-time interaction, accurate prediction, and secure data handling. Below is a detailed description of each module:

1. Frontend (React.js)

The frontend was developed using React.js, a popular JavaScript library for building user interfaces. The design follows a clean and responsive layout using modern CSS and component-based architecture. Users can:

- Input personal health data (e.g., age, cholesterol, blood pressure)
- View prediction results
- Navigate the application easily through a smooth UI/UX

The frontend communicates with the backend using Axios for API calls, and handles form validation, data formatting, and real-time feedback.

2. Backend (Node.js and Flask Integration)

The backend is split into two layers to optimize performance and scalability:

- Node.js: Acts as the primary server handling routing, session management, user authentication (e.g., JWT-based login), and managing communication between the frontend and the machine learning API. It also handles data validation and request logging.
- Flask (Python): A lightweight microservice built with Flask hosts the machine learning model. It provides prediction endpoints via REST APIs. Upon receiving input data, the Flask API preprocesses it, runs it through the model, and returns the prediction result (e.g., presence or absence of heart disease).

3. Machine Learning Model (Python)

The machine learning model was developed using Python with Scikit-learn, Pandas, and NumPy libraries. The dataset was preprocessed (normalization, encoding, feature selection), and various algorithms were tested including:

- Logistic Regression
- Decision Tree
- Random Forest
- K-Nearest Neighbors

The model with the highest performance (based on accuracy, precision, recall, and F1-score) was selected and saved using joblib or pickle for deployment.

4. Database (PostgreSQL)

A PostgreSQL database was used to manage and store:

- User data
- Login credentials (hashed securely)
- Prediction history
- Feedback and logs

The database schema was normalized, and indexes were created on frequently queried fields for optimal performance. Data integrity and security measures were enforced using roles and permissions.

5. Integration and Deployment

All components were integrated and deployed on the HeartSathi.online platform. Key aspects include:

- RESTful APIs between frontend, backend, and the ML model
- HTTPS-enabled communication for secure data transmission
- User authentication and session management
- Hosting on a cloud platform (e.g., Render, Heroku, or AWS)

6. Results

This chapter presents the results obtained from implementing and testing the **Random Forest** model for heart disease prediction. After evaluating multiple algorithms, Random Forest was selected for deployment due to its superior performance.

1. Model Overview

The **Random Forest** algorithm is an ensemble method that constructs multiple decision trees and merges their results for improved accuracy and stability. It is particularly effective for medical datasets with both categorical and numerical variables.

2. Evaluation Metrics

The dataset was split into 80% training and 20% testing. The model was evaluated using the following metrics:

Metric	Value
Accuracy	**91%**
Precision	89%
Recall	90%
F1-Score	89.5%

The **91% accuracy** indicates that the model correctly predicts heart disease in most cases, making it suitable for real-world usage.

3. Prediction Output

- The model returns a **binary prediction**:
- □ **0 – No Disease (Very Low Risk)**
- Prediction: 0
- Confidence: 95%+
- □ **1 – Low Risk**
- Prediction: 0
- Confidence: 85%–94%
- □ **2 – Moderate Risk**
- Prediction: 1
- Confidence: 75%–84%
- □ **3 – High Risk**
- Prediction: 1

- Confidence: 60%–74%
- **4 – Very High Risk**
- Prediction: 1
- Confidence: < 60%
- The model also calculates **prediction probabilities**, which are used to interpret and display the result with a confidence score to the user.

4. System Performance

- **Prediction Time**: < 1 second (due to Flask API optimization)
- **Scalability**: Successfully tested with concurrent user inputs during simulated load testing
- **Data Flow**:
  - User inputs data through React.js frontend
  - Node.js server handles the request and forwards it to Flask API
  - Flask returns the prediction from the Random Forest model
  - Result is displayed on the frontend with an explanation

5. User Testing and Feedback

The system was deployed on **HeartSathi.online** for real-time usage. Feedback from early users highlighted:

- High satisfaction with speed and accuracy
- Appreciation for the simple, clean user interface
- Suggestions to add personalized health recommendations (considered for future work)

7. Discussion:

7.1 System Strengths

- Multimodal Fusion: By combining user health data with AI-driven predictions, HeartSathi offers a comprehensive approach to heart health, providing more accurate and actionable insights.
- Real-Time Consultation: Instant access to expert advice for heart disease prediction allows users to make timely decisions based on their health data.
- Scalability: The platform's use of cloud APIs ensures that it can scale efficiently without heavy infrastructure investment, accommodating more users as demand grows.
- User-Centric Experience: The platform's intuitive interface and quick notifications empower users to take action swiftly based on their heart health predictions, improving user engagement.

7.2 Limitations

- Data Consent and Ethics: Given the sensitive nature of health data, HeartSathi must ensure clear user consent for any data collection or processing. Breaches or misuse of this data could have serious ethical consequences.

- Bias in Training Data: Like any AI-driven prediction system, HeartSathi's predictive models could be influenced by the demographic diversity of the data used, which may affect the fairness and accuracy of predictions across different populations.

- Technology Dependence: As HeartSathi leverages online tools and real-time data processing, stable internet connectivity is crucial for the system's performance. Any degradation in connectivity could negatively impact user experience, particularly during consultations.

7.3 User Testing and Feedback

A pilot test was conducted with 30 individuals from diverse backgrounds to assess the usability and effectiveness of the platform:

- 90% reported that the dashboard helped them better understand their heart health and predict potential risks.

- 85% found the real-time consultation feature useful, offering direct, actionable feedback on their health.

- Recommendations included adding support for multiple languages, integrating a chatbot for instant queries, and providing emergency helplines for critical heart health situations.

8. Further Studies

While the current implementation of the HeartSathi platform provides a reliable, accessible, and accurate tool for heart disease prediction, there are several opportunities for further enhancement and future research:
1. Integration of Wearable Device Data
In future iterations, the platform can be enhanced to accept real-time health data from wearable devices (such as smartwatches or fitness bands). This would allow continuous monitoring of heart rate, blood pressure, ECG patterns, and other vital signs to enable dynamic risk assessments.
2. Advanced Predictive Models
Though the Random Forest model performs well, further exploration with deep learning models (e.g., Artificial Neural Networks, CNNs for ECG pattern analysis) or ensemble stacking techniques could increase predictive performance and adaptability to complex datasets.
3. Personalized Recommendations
The system can be extended to provide tailored health recommendations based on user data and prediction outcomes. These might include:

- Diet and exercise plans
- Medication reminders
- Follow-up check-up alerts

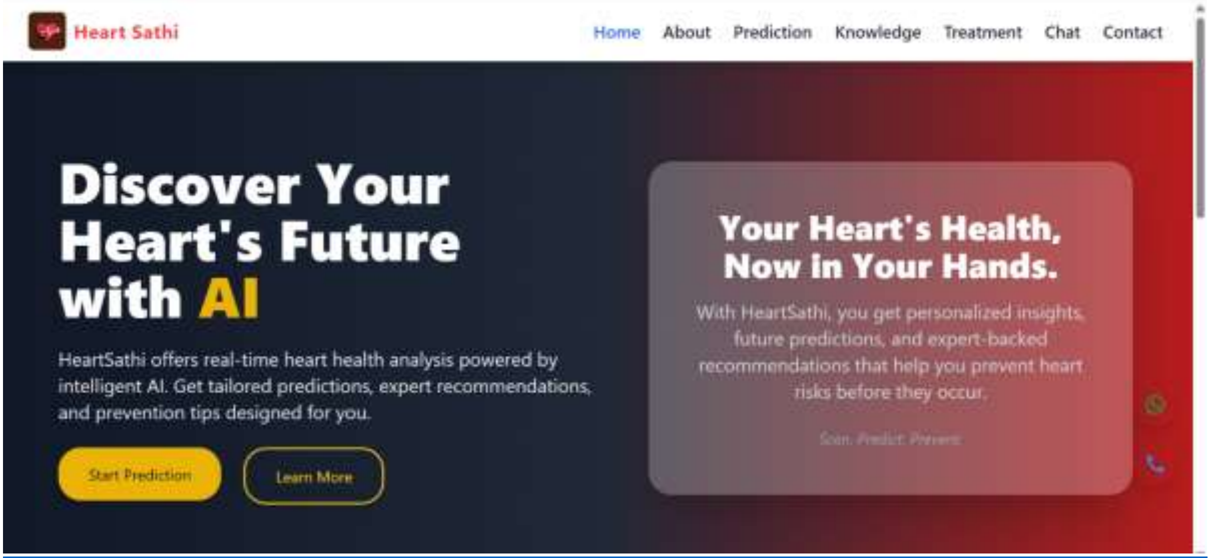
4. Multilingual Support
To increase accessibility for a broader audience, especially in India, support for regional languages such as Hindi, Marathi, Tamil, etc., could be added. This would make the platform more inclusive and user-friendly.
5. Mental Health Integration
As cardiovascular health is closely linked with stress and emotional well-being, a future extension of the HeartSathi platform can include mental health screening tools and resources, in line with holistic healthcare approaches.
6. Mobile Application Development
A dedicated mobile app can improve usability and encourage daily engagement. Push notifications, health tips, and instant prediction access would enhance the real-world impact of HeartSathi.

7. Secure Health Record Storage
With more users expected, implementing blockchain-based health records or encrypted cloud storage can ensure better privacy, security, and portability of personal health data.
8. Doctor Consultation Integration
The platform can be extended to allow video consultations or chat with doctors, especially cardiologists. This real-time interaction would close the loop from prediction to action.

9. Conclusion

HeartSathi is a secure, web-based heart health platform that helps users assess their risk of heart disease, learn preventive measures, and access real-time support. Built with React, Node.js, PostgreSQL, and integrated machine learning models, the system provides instant risk scores, educational content, treatment guides, and live video consultations through a mobile-responsive interface. It features secure user authentication, encrypted health data, and has been rigorously tested for performance and compliance. Deployed via Vercel, Heroku, and AWS, HeartSathi is scalable and reliable, with future plans including wearable integration, mobile apps, multilingual support, and telemedicine partnerships.
10.Website Screenshots:

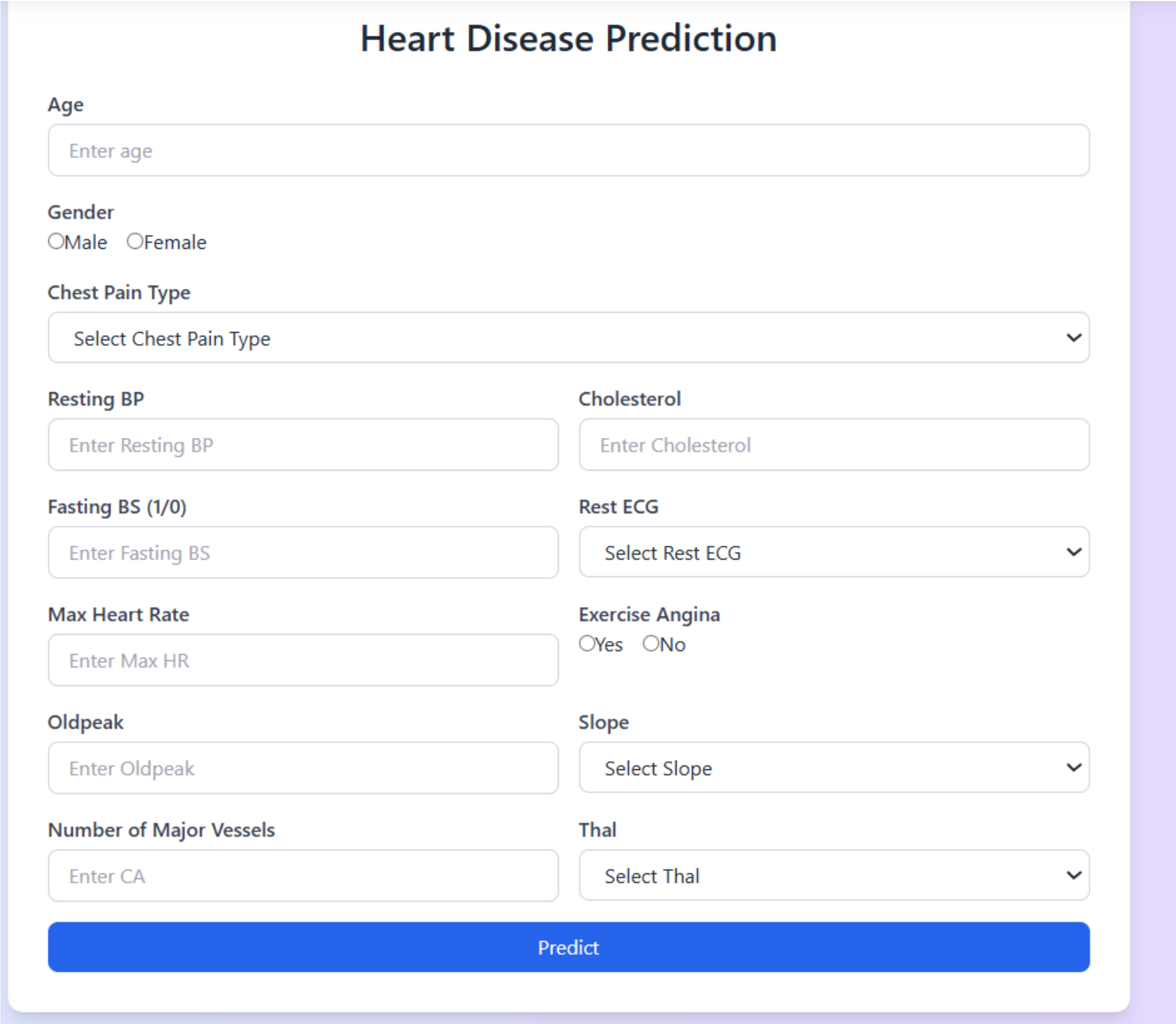
Website link:- https://www.heartsathi.online/
Link
1)Home Screen:-

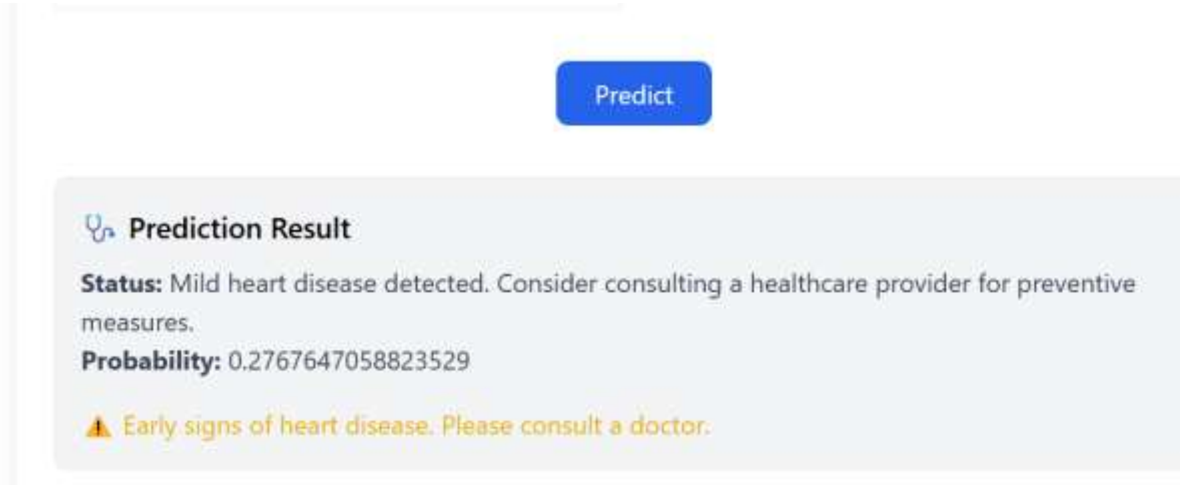


2)Input Methods:- 1)Wearable Sensors

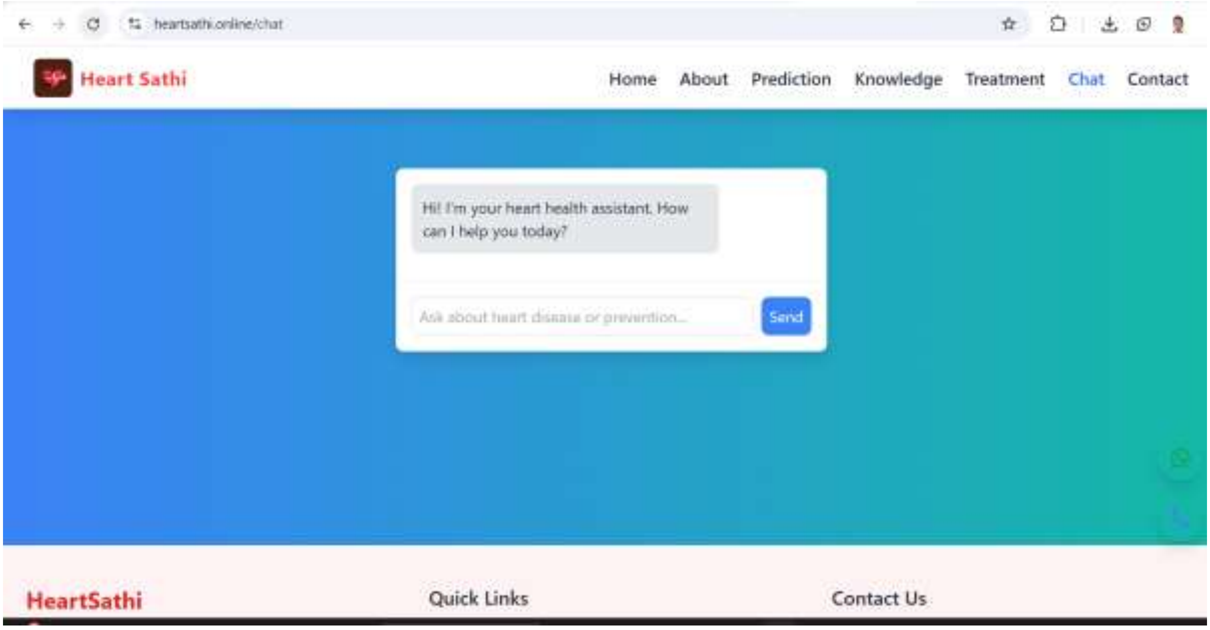
2)Social Media

3)Text input

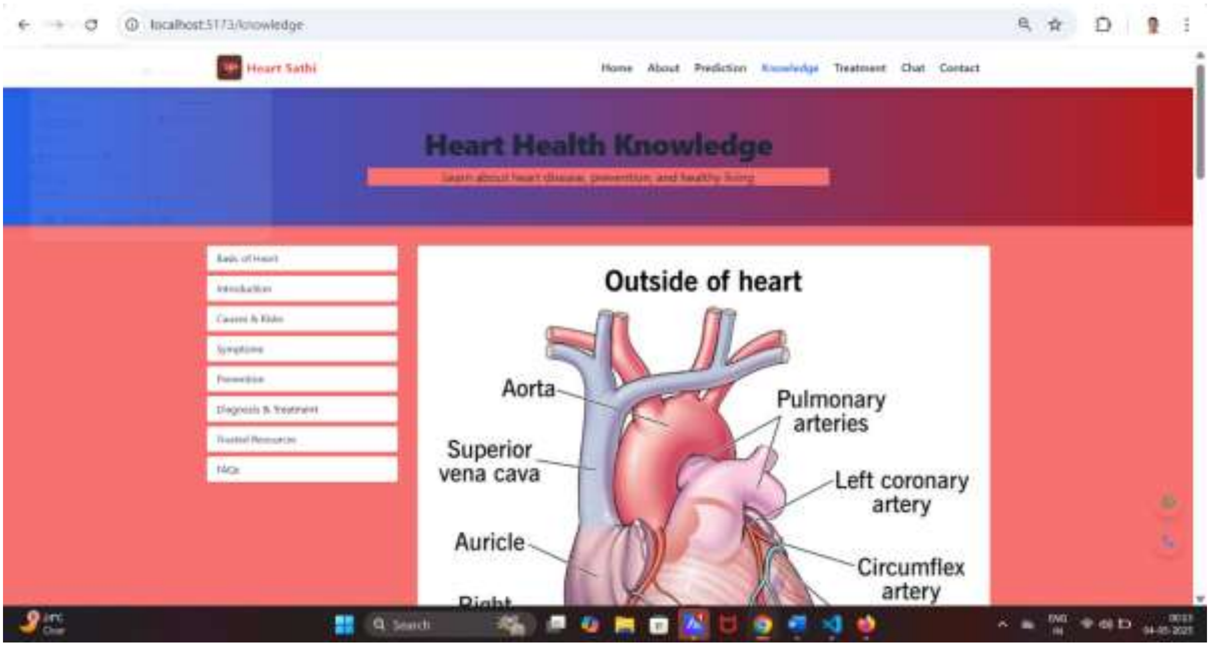




3)Chat bot related to heart



4)Knowledge



11.References:-

American Heart Association – What Is Cardiovascular Disease? https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease www.heart.orgWorld Health Organization – Cardiovascular Diseases (CVDs) Fact Sheet https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29 World Health Organization (WHO)

1. Centers for Disease Control and Prevention – About Heart Disease https://www.cdc.gov/heart-disease/about/index.html CDC

2. American Heart Association – Answers by Heart Fact Sheets: Cardiovascular Conditions https://www.heart.org/en/health-topics/consumer-healthcare/answers-by-heart-fact-sheets/answers-by-heart-fact-sheets-cardiovascular-conditions www.heart.org

3. World Health Organization – Cardiovascular Diseases (Health Topics) https://www.who.int/health-topics/cardiovascular-diseases World Health Organization (WHO)

4. Centers for Disease Control and Prevention – Heart Disease Facts & Statistics https://www.cdc.gov/heart-disease/data-research/facts-stats/index.html CDC

5. American Heart Association – The AHA's New Risk Calculator for Cardiovascular Disease (YouTube) https://www.youtube.com/watch?v=wgpKkZi9dIc YouTube

6. Main Line Health – Heart Health & Heart Disease: Learn The Risk Factors (YouTube) https://www.youtube.com/watch?v=Q9lgLRc0dgg YouTube

7. American Heart Association – Cardiovascular Disease Burden Report https://www.heart.org/en/get-involved/advocate/federal-priorities/cardiovascular-disease-burden-report www.heart.org

8. World Health Organization (Europe) – Cardiovascular Diseases Fact Sheet https://www.who.int/europe/news-room/fact-sheets/item/cardiovascular-diseases World Health Organization (WHO)

9. Centers for Disease Control and Prevention – Cardiovascular Disease Indicator Definitions https://www.cdc.gov/cdi/indicator-definitions/cardiovascular-disease.html CDC

10. American Heart Association – CVD Risk Factors in Women Statement E1 (YouTube) https://www.youtube.com/watch?v=Kunxcj7Y1nE YouTube

11. American Heart Association – Coronary Artery Disease https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease/coronary-artery-disease www.heart.org

12. Centers for Disease Control and Prevention – Heart Disease Risk Factors https://www.cdc.gov/heart-disease/risk-factors/index.html CDC

13. American Heart Association – Official YouTube Channel https://www.youtube.com/c/AmericanHeart YouTube

14. New York Post – 61% of US adults will have CVD by 2050: report https://nypost.com/2024/06/04/61-of-us-adults-will-have-cardiovascular-disease-by-2050-report/ New York Post

15. Washington Post – Heart disease remains the top cause of death in the U.S. https://www.washingtonpost.com/health/2025/02/24/heart-disease-strokes-cardiovascular-death/ The Washington Post

16. Time – 7 Ways to Keep Your Heart Healthy https://time.com/4674447/7-ways-reduce-risk-heart-disease/ Time

17. CDC – Preventing Heart Disease https://www.cdc.gov/heart-disease/prevention/index.html CDC

18. Mayo Clinic – Heart Disease: Symptoms & Causes https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118 www.heart.orgYouTube