WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

HeartGuard Smart Safe Disease Prediction System

Dr. P. Ashok Kumar[1], Prashanth[2], Ram Charan[3], Pradeep[4], Vivek[5].

[1]Assistant Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India
[2,3,4,5]Students, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India
ashokkumar502@gmail.com[1], mail2sprashanth@gmail.com[2], ramcharanb2005@gmail.com[3], pradeepaele14@gmail.com[4], vivekhendige03@gmail.com[5]

ABSTRACT:

Chronic diseases such as heart disease remain one of the leading causes of death worldwide after COVID-19, placing a significant burden on healthcare systems and patient quality of life. Early detection and timely intervention are critical to managing these conditions effectively. This project presents **HeartGuard**, a Heart Disease Prediction System designed to assist in the early identification of individuals at risk of heart disease by analysing key clinical parameters. Leveraging machine learning techniques, specifically the Random Forest algorithm, **HeartGuard** evaluates various patient attributes such as age, blood pressure, cholesterol levels, chest pain type, and other relevant medical indicators to predict the likelihood of heart disease. The model is trained on a well-established dataset, demonstrating high accuracy, robustness, and interpretability. By automating the risk assessment process, **HeartGuard** aims to support healthcare professionals in making informed decisions and enhancing preventive care. The project highlights the potential of data-driven approaches in combating chronic diseases and improving patient outcomes through early detection.

Keywords: chronic diseases, heart disease, prevention, prediction, health, Machine Learning, Random Forest algorithm

I. Introduction:

Heart Complaints remain one of the leading causes of death worldwide, claiming millions of lives each time. Beforehand discovery and timely medical intervention are essential to reducing the inflexibility and long-term impact of heart-related conditions. Still, traditional threat assessment styles frequently fall suddenly directly landing the complexity of individual case biographies and the relations among colorful threat factors. This design, named HeartGuard, aims to bridge these gaps by using advanced machine-literacy ways to enhance the delicacy, speed, and availability of heart complaint threat vaticination. Despite the eventuality of ML in healthcare, several challenges hamper its wide relinquishment numerous ML models warrant external confirmation and standardization, limiting their capability to perform well across different populations. Deficient or prejudiced datasets reduce model trustability and fairness. The complexity of some ML algorithms makes their prognostications delicate for clinicians to interpret or trust. There's a need for ML systems that can help croakers in making informed opinions, especially in complex medical cases. By addressing these challenges, HeartGuard seeks to ameliorate clinical decision-timber, reduce individual crimes, and ensure that no critical medical information is overlooked — eventually leading to better issues for cases.

II. Literature Survey:

The project titled **"Heart Disease Prediction Using Machine Learning Algorithms"** by Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath proposed a hybrid approach to predict heart disease. This model integrates various machine learning algorithms such as Decision Trees, Support Vector Machines (SVM), Gradient Boosting, XGBoost, Artificial Neural Networks (ANN), and Logistic Regression. By combining these techniques, the goal was to leverage the unique strengths of each algorithm to improve predictive accuracy. Careful preprocessing steps, including feature selection, handling of missing data, and class imbalance correction, were implemented. The model's performance was evaluated using metrics such as accuracy, precision, recall, and AUC, demonstrating its robustness and reliability in clinical applications.

In the study titled **"An Active Learning Machine Technique Based Prediction of Heart Disease"**, the authors used the UCI heart disease dataset and applied a wide array of machine learning classifiers. These included K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machines (SVM), and Decision Trees (DT). In addition, more advanced neural network models such as Naive Bayes, Radial Basis Function networks, and Learning Vector Quantization were tested. The project placed strong emphasis on feature engineering, preprocessing techniques, and comparative evaluation of models to identify the most accurate and efficient approach for predicting heart disease.

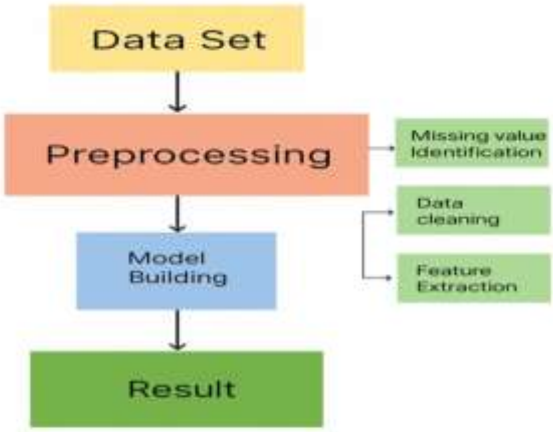
Several other notable studies have explored heart disease prediction using machine learning. One project employed a Random Forest classifier as the primary model, focusing on training it with heart disease datasets for accurate prediction. This approach highlighted the benefits of supervised learning in medical diagnosis and showed promising results in predicting potential heart attacks. Another study by **Richards et al. (2001)** and **Sung et al. (2015)** applied models like K-Nearest Neighbors and Multi-Linear Regression to predict stroke severity index (SSI), concluding that KNN outperformed linear models in accuracy.

Arslan et al. (2016) explored the use of Support Vector Machines and Penalized Logistic Regression for heart stroke prediction. Their findings showed that SVM offered superior performance compared to other models, confirming its suitability for complex classification problems. Similarly, **Boshra Brahmi et al.** evaluated a variety of classification techniques including J48 Decision Tree, KNN, and Naïve Bayes to measure their effectiveness in heart disease diagnosis. Metrics like accuracy, precision, sensitivity, and specificity were used to assess the models.

Finally, **Senthilkumar Mohan et al.** developed a hybrid algorithm called the Hybrid Random Forest Linear Method (HRFLM), which combines Random Forest with a Linear Method (LM). Using the Cleveland heart disease dataset, the authors chose to omit personal attributes like age and sex, arguing they were not essential for prediction. Their model integrated four algorithms and demonstrated that the hybrid method yielded better results than using individual classifiers. The study emphasized continuous improvement through hybridization of models to enhance diagnostic accuracy.

III. Methodology:

The methodology outlines the step-by-step approach followed to design, develop, and evaluate a heart disease prediction system using the Random Forest classification algorithm. The objective is to predict whether a person has heart disease based on a set of clinical and demographic features.



a) Data Collection

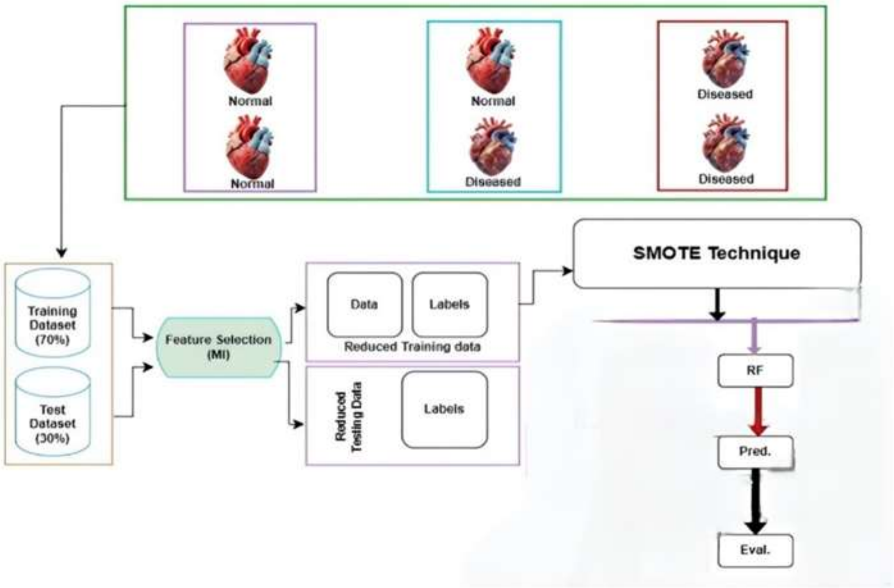
The dataset used for this study is sourced from the UCI Machine Learning Repository, a widely recognized public repository for machine learning datasets. It contains comprehensive medical records of patients, including multiple attributes related to heart health. These attributes provide essential clinical and demographic information critical for predicting the presence or absence of heart disease. Some of the key features included in the dataset are age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the ST segment, number of major vessels coloured by fluoroscopy, thalassemia, and the target variable indicating heart disease presence (1) or absence (0)

b) Data Preprocessing

To ensure data quality and model effectiveness, several preprocessing steps are undertaken. Missing data points in the dataset are either removed or imputed to maintain completeness and avoid biases during training. Categorical variables such as chest pain type and sex are converted into numerical formats using encoding techniques like dummy variables or one-hot encoding, allowing the model to process these features effectively. Numeric attributes are standardized or normalized to align their scales, improving the model's ability to learn from the data consistently.

c) Proposed Model

System Architecture



System architecture is a comprehensive blueprint that defines the structure, behavior, and interactions of various components within a system—whether it's a software application, a computer system, or a complex network of systems. It provides a high-level view of how the system is organized and how different parts such as hardware, software, data storage, processing units, communication protocols, and user interfaces interact to perform specific functions. In software systems, architecture describes how modules or services are divided, how they communicate (e.g., via APIs or message queues), and how data flows through the system. In hardware systems, it includes the design of processors, memory units, input/output devices, and how they are connected. System architecture also includes considerations for scalability (handling growth in users or data), security (protecting data and operations), maintainability (ease of updates and debugging), and performance (speed and efficiency).

d) Model Evaluation:

The trained model's performance is evaluated using metrics like accuracy, precision, recall, F1-score, and confusion matrix. To ensure stability, k-Fold Cross-Validation is applied, dividing the dataset into k subsets for iterative training and testing.

Evaluation Metrics:

The heart disease prediction system was built using the Random Forest algorithm, trained on a dataset containing various patient attributes such as age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting ECG results, maximum heart rate, exercise-induced angina, ST depression, and other relevant features After preprocessing the data, including handling missing values, normalization, and encoding categorical variables, the dataset was split into training and testing sets in an 80:20 ratio. The Random Forest model was then trained on the training set.

The model’s performance on the test set is summarized below:

Total Instances (N): 41000

True Positives: 20200

True Negatives: 20200

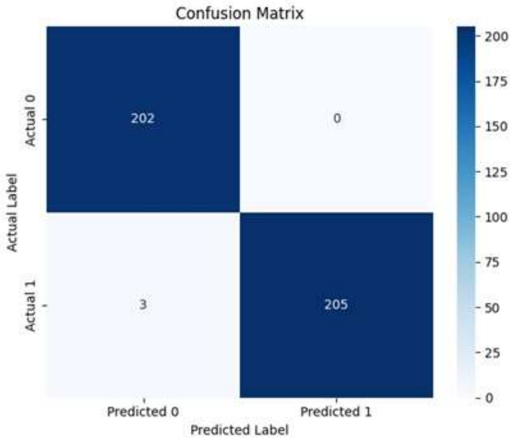
False Positives: 0

False Negatives: 600

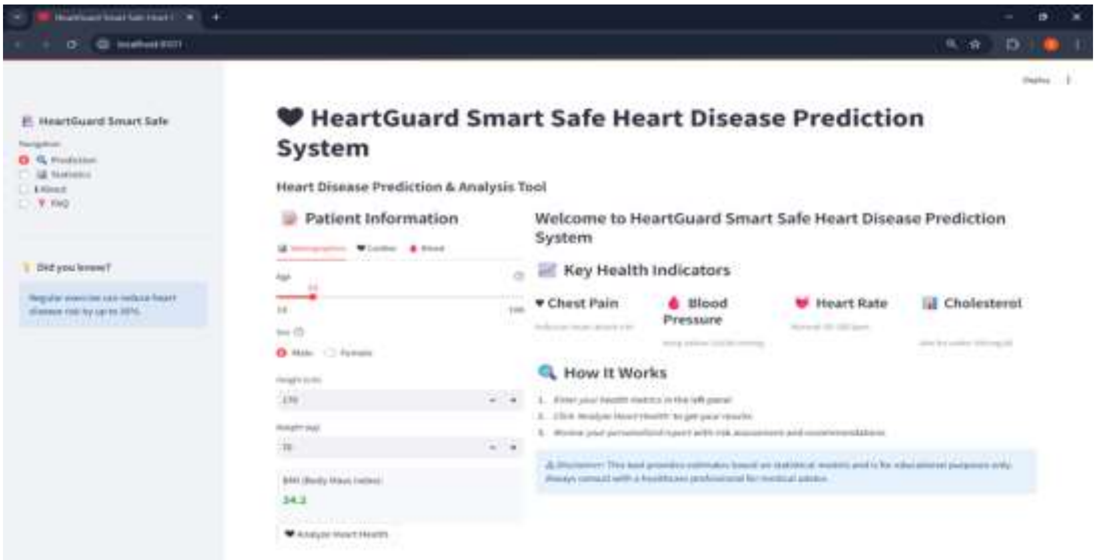
Accuracy = ((20200+20200)/ 41000) = 98.53%

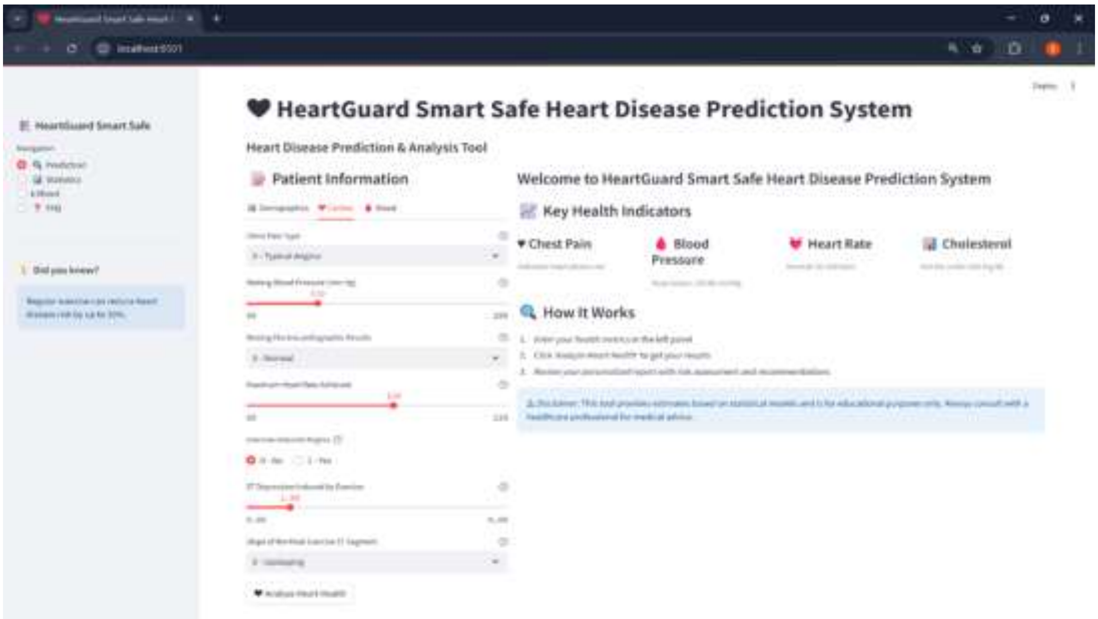
Precision = (20200/(20200+0)) ≈ 100%

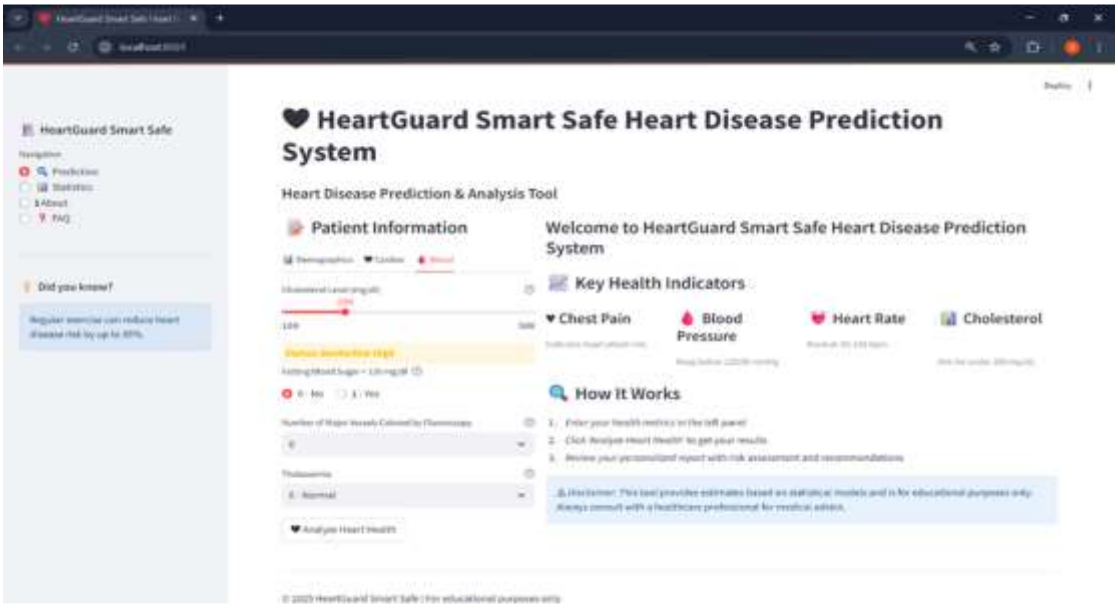
Recall = (20200/(20200+600)) ≈ 97.1%

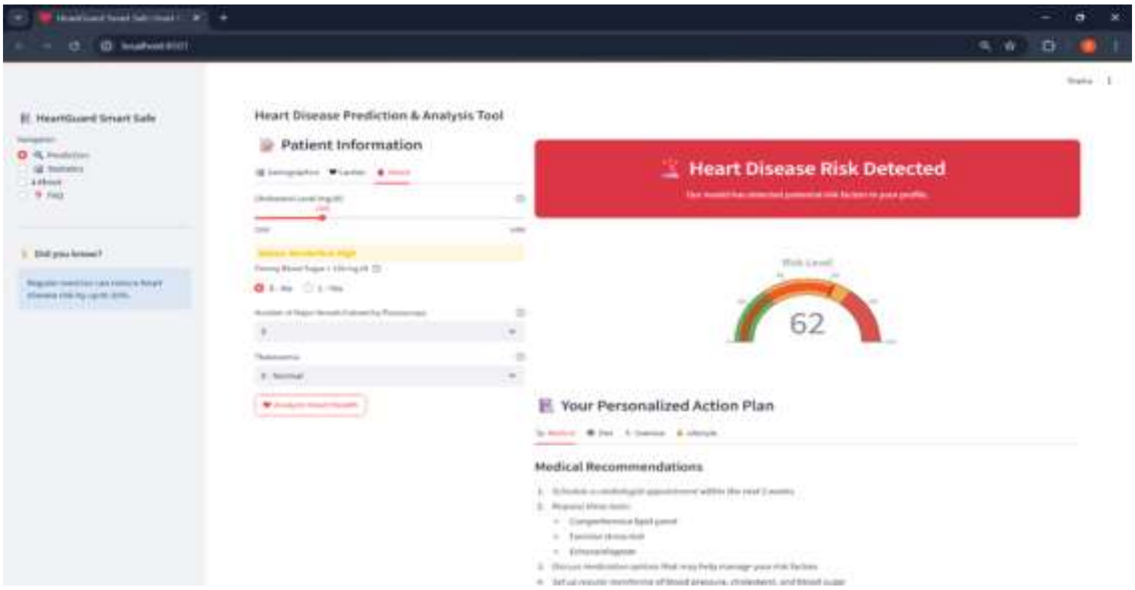

Confusion Matrix
Actual 0
Actual 1
Actual Label
202
0
3
205
Predicted 0
Predicted 1
Predicted Label
200
175
150
125
100
75
50
25
0

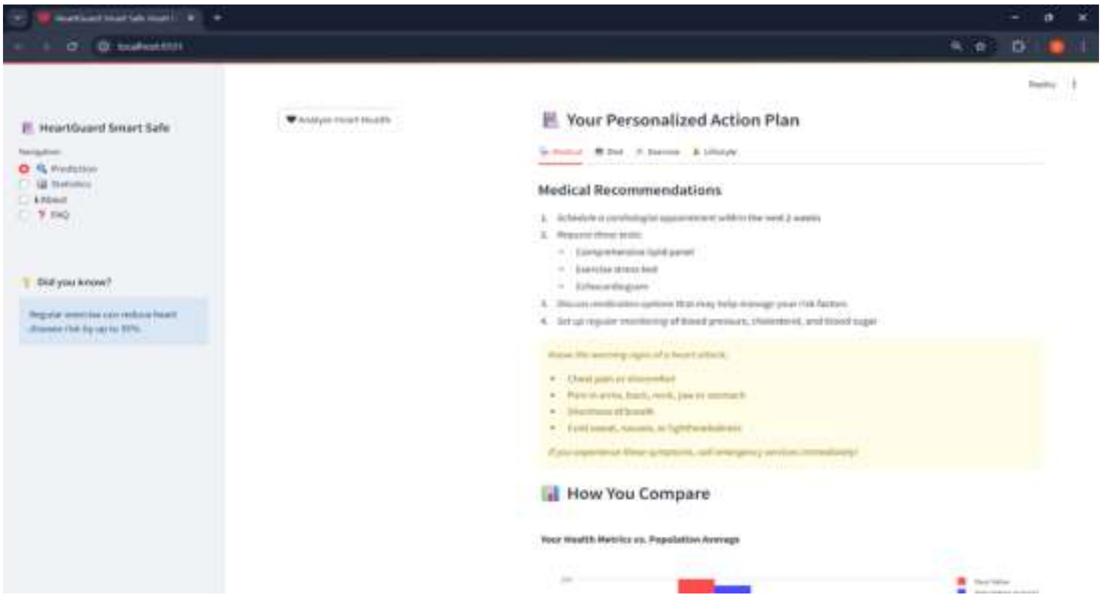
IV. Results:

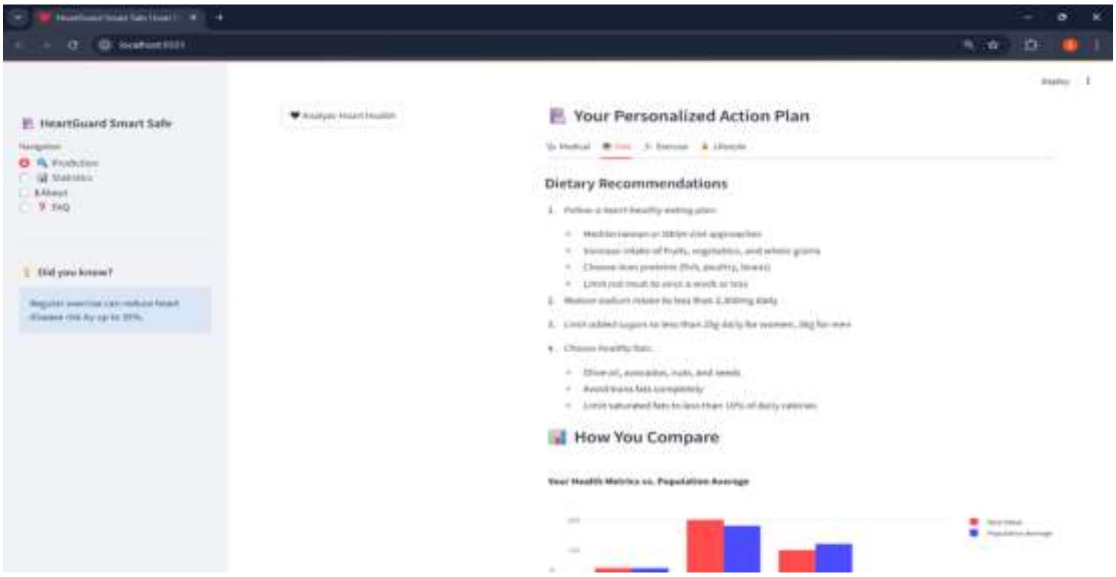

HeartGuard Smart Safe Heart Disease Prediction System
Heart Disease Prediction & Analysis Tool
Patient Information
Welcome to HeartGuard Smart Safe Heart Disease Prediction System
Key Health Indicators
Chest Pain
Blood Pressure
Heart Rate
Cholesterol
How It Works

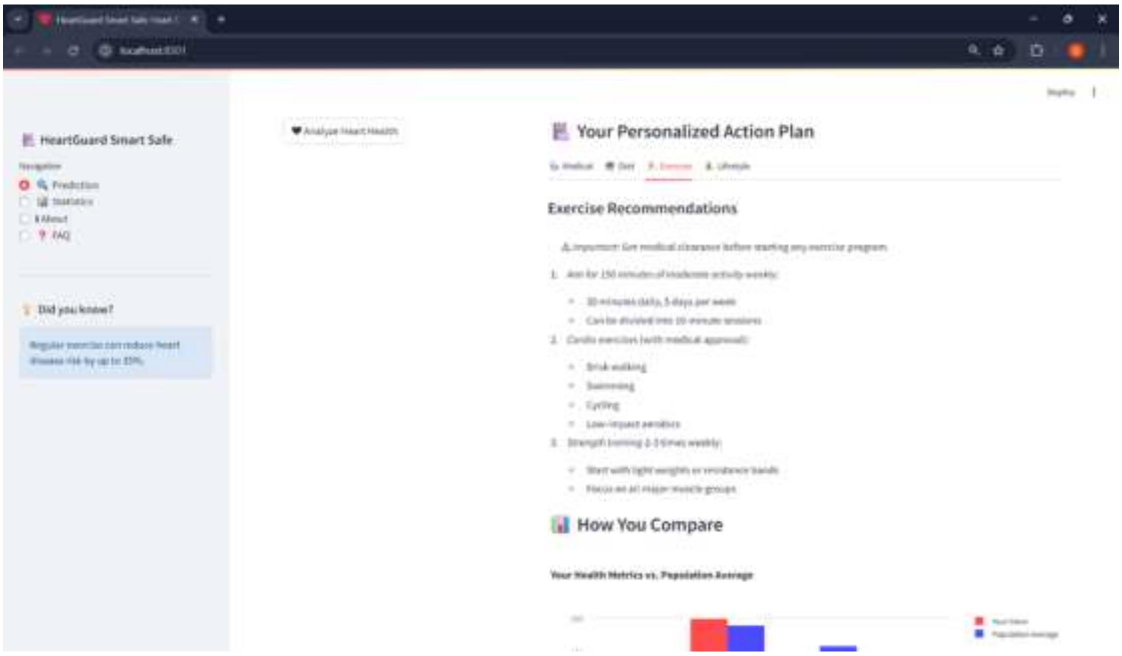

HeartGuard Smart Safe Heart Disease Prediction System
Heart Disease Prediction & Analysis Tool
Patient Information
Welcome to HeartGuard Smart Safe Heart Disease Prediction System
Key Health Indicators
Chest Pain
Blood Pressure
Heart Rate
Cholesterol
How It Works

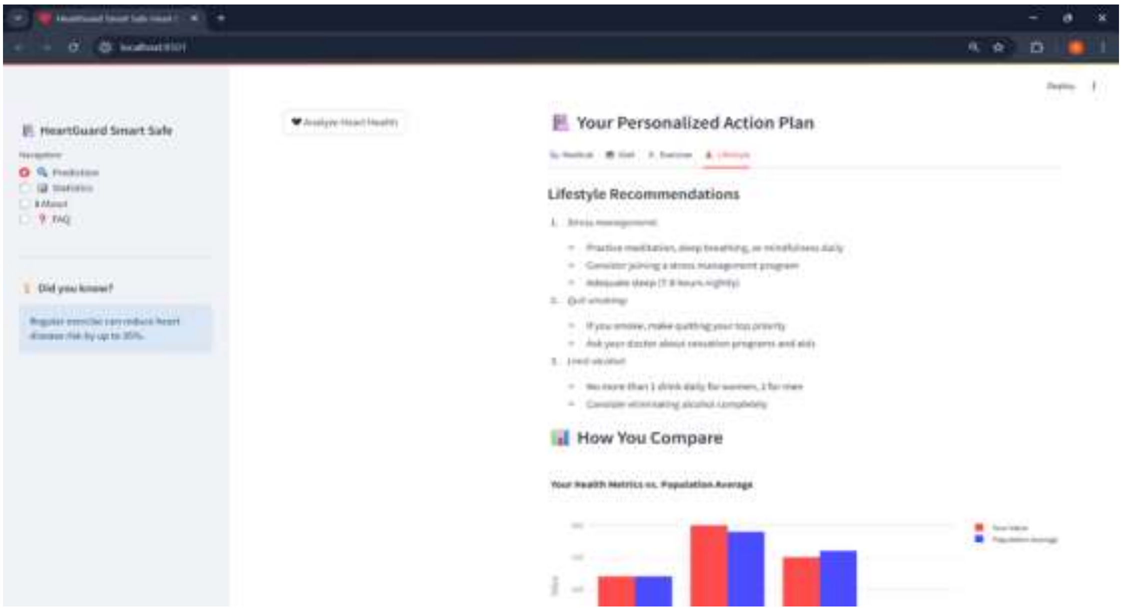

HeartGuard Smart Safe
Did you know?
HeartGuard Smart Safe Heart Disease Prediction System
Heart Disease Prediction & Analysis Tool
Patient Information
Welcome to HeartGuard Smart Safe Heart Disease Prediction System
Key Health Indicators
Chest Pain
Blood Pressure
Heart Rate
Cholesterol
How It Works


HeartGuard Smart Safe
Did you know?
Heart Disease Prediction & Analysis Tool
Patient Information
Heart Disease Risk Detected
62
Your Personalized Action Plan
Medical Recommendations


HeartGuard Smart Safe
Did you know?
Your Personalized Action Plan
Medical Recommendations
How You Compare


HeartGuard Smart Safe
Did you know?
Your Personalized Action Plan
Dietary Recommendations
How You Compare
Your Health Metrics vs. Population Average


HeartGuard Smart Safe
Did you know?
Your Personalized Action Plan
Exercise Recommendations
How You Compare
Your Health Metrics vs. Population Average


HeartGuard Smart Safe
Did you know?
Your Personalized Action Plan
Lifestyle Recommendations
How You Compare
Your Health Metrics vs. Population Average

V. Conclusion:

A heart disease prediction system was successfully developed using the Random Forest algorithm. The model was trained on a medical dataset containing various clinical features and achieved high accuracy and reliability in predicting the presence of heart disease. The ensemble nature of Random Forest allowed for improved performance by reducing overfitting and enhancing generalization across unseen data. The results demonstrate that Random Forest is a suitable choice for medical classification tasks, providing a good balance between accuracy, sensitivity, and precision. Key features such as chest pain type, maximum heart rate, and ST depression were found to significantly influence the predictions. While the model performed well, future work can focus on improving interpretability, incorporating real-time data, and expanding the dataset to include more diverse patient profiles. Integrating this system into healthcare environments could assist medical professionals in making quicker, data-driven decisions, ultimately contributing to early diagnosis and better patient outcomes.

VI. References:

1. Dai, W.Y., Brisimi, T.S., Adams, W.G.: Prediction of hospitalization due to heart disease by supervised learning methods. Int. J. Med. Inform. **84**(3), 189–197 (2015).
2. Kedar, S., Bormane, D.S., Nair, V.: Computational Intelligence in Data Mining, vol. 1, pp. 69–56. Springer, New Delhi, India (2016).
3. Chen, J.H., Wan, H.J., Zhang, Q.H.: diagnosis of heart disease based on support vector machine with missing data. Math. Pract. Knowl. **47**(02), 130–135 (2017).
4. Wen, B.W., Dong, W.H., Xie, W.J., Ma, J.: parameter optimization of random forest based on improved grid search algorithm. Comput. Eng. Appl. **54**(10), 154–157 (2018).
5. Zhao, J.C., Li, Y., Wang, D., Zhang, J.H.: An optimized random forest heart disease prediction algorithm. J. Qingdao Univ. Sci. Technol. (Nat. Sci. Edition) **42**(02), 112–118 (2021).
6. Wang, Y.S., Xia, S.T.: Overview of random forest algorithms for ensemble learning. Inf. Commun. Technol. **12**(01), 49–55 (2015).
7. KUMAR, P. ASHOK, GSatish KUMAR, and SETTI NARESH KUMAR. "Improve the Capacity of Uniform Embedding for Efficient JPEG Steganography Based on DCT." (2015).
8. Kumar, P. Ashok, B. Vishnu Vardhan, and Pandi Chiranjeevi. "Investigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." In XVIII International Conference on Data Science and Intelligent Analysis of Information, pp. 13-26. Cham: Springer Nature Switzerland, 2023.
9. Sunkavalli, Jayaprakash, B. Madhav Rao, M. Trinath Basu, Harish Dutt Sharma, P. Ashok Kumar, and Ketan Anand. "Experimentation Analysis of VQC and QSVM on Sentence Classification in Quantum Paradigm." In 2024 International Conference on Computing, Sciences and Communications (ICCSC), pp. 1-5. IEEE, 2024.
10. Kumar, P. Ashok. "Event Based Time Series Sentiment Trend Analysis.".
11. PANDI, CHIRANJEEVI, THATIKONDA SUPRAJA, P. ASHOK KUMAR, and RALLA SURESH. "A SURVEY: RECOMMENDER SYSTEM FOR TRUSTWORTHY.".
12. Kumar, P. Ashok, B. Vishnu Vardhan, and Pandi Chiranjeevi. "Correction to: Investigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." In XVIII International Conference on Data Science and Intelligent Analysis of Information, pp. C1-C1. Cham: Springer Nature Switzerland, 2023.