Heart Disease Prediction Using Ml

Dr.P.Ashok Kumar[1],G.Sathwika[2], D.Abhinay Reddy[3],C.Shirin [4], and K.Charan Goud [5].
[1] Associate Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India
[2,3,4,5] III B.Tech. Students, Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India.

How to Cite this Article:

G.Sathwika, , Reddy, D., C.Shirin, & Goud, K. (2026). Heart Disease Prediction Using Ml. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.207

https://doi.org/ 10.55041/ijcope.v2i4.207

Abstract:

Among other diseases, heart diseases are amongst the topmost causes of deaths globally. This has prompted an early detection of these diseases because most heart diseases are asymptomatic, thus making the process of diagnosis hard. Machine learning plays a big role in disease prediction by the analysis of data from different patients.

In this paper, a machine learning model will be developed, which predicts the risk level of suffering from a heart disease based on various parameters such as age, blood pressure, cholesterol, blood sugar, heart rate, family history and diet of the patients. The algorithms used include decision tree and support vector machine (SVM).

It will provide timely prediction of results due to its simplicity and speed.

1. Introduction:

Heart problems are among the most common causes of mortality worldwide, thus early identification and prevention are some of the crucial aspects concerning protection from these dangers. Heart diseases can be characterized by lack of symptoms, and therefore early identification becomes challenging. The traditional way of early diagnosis requires undergoing various examinations, involving specialists, which requires plenty of time and money. Moreover, not all people have access to it.

Technological development has led to the appearance of new methods of health monitoring using machine learning techniques. Using machine learning in healthcare is extremely important because it helps identify certain patterns and predict diseases. It significantly enhances the accuracy of the diagnosis, as well as its speed since it is possible to conduct a diagnosis in just seconds.

In the current research, machine learning will be used to create a prediction system for detecting heart diseases based on some factors such as age, blood pressure, cholesterol, blood sugar level, heart rate, illness history, and lifestyle. The classification approach including such algorithms as the decision tree and SVM will be applied to classify patients according to their risks.

2. Related Work:

There have been several research studies conducted in the area of heart disease prediction through use of statistical and machine learning techniques. Detrano et al. (1989) presented a probability-based approach for detecting heart diseases based on coronary artery disease; however, this technique was based on a limited dataset and lacked accuracy.

Machine learning techniques such as Support Vector Machine (SVM) by Cortes & Vapnik (1995) offer high accuracy in terms of classification but require proper tuning of parameters and have high complexity. Decision Tree, developed by Quinlan (1986), offers simplicity of implementation and ease of understanding but suffers from the problem of overfitting.

In a more recent study by Weng et al. (2017), it has been shown that machine learning offers an efficient way of predicting cardiovascular risks using clinical data. However, problems such as data quality and implementation in real time remain. This research will try to build a simple, accurate, and efficient prediction system.

2.1Existing System and its Limitations:

Title	Technology	Limitation	Authors	Year
AI-Driven Heart Disease Prediction Using ML & DL Techniques	LR, SVM, KNN, RF, XGBoost, ANN	Needs large dataset for best accuracy	Vijayasimha A., J. Avanija	2025
Heart Disease Prediction Using ML Techniques	KNN, Random Forest	Uses only existing datasets; lacks real-time data	D.Ratna Kumari et al.	2025
Heart Disease Prediction Using Machine Learning	ANN, AI Data Mining	Limited clinical validation	IRJAEH Journal Authors	2024
Heart Disease Detection Using Machine Learning Models	Classification ML Algorithms	Overfitting with small datasets	A. Singh et al.	2024
A Proposed Technique for Predicting Heart Disease Using ML	Feature Selection + ML Models	Accuracy depends on dataset quality	H. El-Sofany et al.	2024
Advancements in Heart Disease Prediction Using ML	SVM, Decision Tree, RF, Neural Network	Limited dataset affects generalization	B. S. Ingole et al.	2024

Effective Heart Disease Prediction Using Machine Learning	RF, Decision Tree, XGBoost, MLP	Complex hyperparameter tuning required	C. M. Bhatt et al.	2023

3. Methodology:

The suggested system adopts an approach based on machine learning algorithms to predict the likelihood of heart diseases based on health information of the patients. In particular, there is a procedure that starts with the process of data gathering, including necessary parameters like age, blood pressure, cholesterol levels, blood sugar, heart rate, and other relevant parameters such as the patient's family history and behavior.

The gathered information undergoes preprocessing where any noise or unnecessary features can be removed from it, while missing features can be added. In addition, the collected data can be transformed into appropriate forms that can be utilized by the machine learning algorithms.

The next stage of the work concerns machine learning algorithms that include Decision Tree and SVM, which will be used to build predictive models and recognize patterns among parameters that might cause heart diseases.

The prediction process involves giving input data to the machine learning algorithm, which then classifies users' data and shows whether the likelihood of developing heart disease is high or low. Finally, the output of the program is the prediction of heart disease risk and some useful health-related suggestions for the users.

3.1 Data Collection and Preprocessing:

Data collection in this case study involves obtaining necessary health-related parameters that would be useful in predicting the chances of acquiring heart diseases. The system obtains the data from either user input or uploading datasets. These input parameters include age, gender, blood pressure, cholesterol, blood sugar, heart rate, family history, smoking, physical activity, and type of diet, among others. These parameters are critical in determining one's health status and predicting whether he/she is at risk of acquiring heart disease.

Next, the collected data goes through the preprocessing stage. At this stage, the system removes any missing data, removes any redundant or unnecessary data such as duplicate columns, renames attributes according to the machine learning models requirements, and adds any default attributes that may not have been obtained during the first data collection stage. All these measures ensure that the data meets the needs of machine learning model.

Lastly, the system converts the raw data to numerical form and normalizes the data according to model requirements.

3.2 Feature Extraction:

The feature extraction stage is critical in the heart disease prediction system, where relevant features are extracted from the input data. In this case, the system employs key health parameters including age, gender, type of chest pain, blood pressure, cholesterol content, fasting blood sugar, resting heart rate, history of heart disease, smoking habits, physical activity, and diet.

Upon preprocessing, features are extracted and sorted based on the requirements of the machine learning model. The system maps input features to feature names such as age, trestbps, chol, thalach, and other key health parameters that have been previously selected for use in machine learning models. Missing features are addressed through default feature values, ensuring compatibility with the trained machine learning model.

Categorical data features are mapped to numerical values to facilitate processing by machine learning models. The final features are then sorted in the desired format and supplied as inputs to the machine learning model.

Effective feature extraction enables the system to focus only on key health parameters in predicting heart diseases.

3.3 Model Selection and Training:

In this work, appropriate machine learning algorithms are chosen to predict the possibility of developing heart disease. Such models include decision trees and support vector machines. These algorithms are effective in solving classification problems, which is important for accurate data analysis in the field of medicine.

To build and test such models, a pre-processed data set is used during the training process. The pre-processing consists in organizing the data in accordance with the necessary parameters. In our case, we will be working with two arrays – one with input data and another one with output labels, where each label will indicate whether there is a high or low probability of developing heart disease.

A decision tree algorithm will be used because it involves constructing a tree of decisions based on certain features. Thus, the model will be quite simple, intuitive and easy to analyze. Another useful algorithm for us will be a support vector machine since its main task is to classify data with high accuracy.

When building models, it is essential to test them. During testing, some samples of the data are selected for this purpose. After successful training, the final result is stored in a model.

3.4 Feature Engineering and Selection:

Feature engineering and selection are very important processes that will help to improve the effectiveness of the heart disease prediction model. In this project, several health indicators like age, sex, chest pain type, blood pressure, cholesterol, blood sugar, heart rate, family history, smoker or non-smoker, activity levels, and diet type have been taken into account as features.

As part of feature engineering, the data is prepared to be used by machine learning algorithms in the most effective way. Some features like sex, smoker or non-smoker, and diet type have been converted to numerical data since they are categorical. The project also includes renaming and mapping of input features to meet the format needed for the trained model. Default values for missing features have been assigned.

As for feature selection, it means selecting those few features which can influence the results in the most significant way while all unnecessary and less influential features have been ignored.

3.5 Model Evaluation:

Evaluation of models is done to test the performance and accuracy of the models used in the development of the heart disease prediction system. Evaluation tests are run against the trained models through the use of the sample inputs to see the performance of the models in classifying patients as either high risk or low risk.

For this specific system, the evaluation will mostly focus on the accuracy of predictions made by the decision tree algorithm and support vector machine (SVM) algorithm. Whether or not the predictions are consistent with the output of the dataset is tested and if there is more consistency then the model works better.

Furthermore, the system is tested for its accuracy by inputting user real-time inputs and also uploading datasets to test the performance of the models. Results are displayed for the user to see the number of high risk and low risk cases and mostly batch predictions.

Other areas to be considered during the evaluation include speed, efficiency, and error handling among others.

3.6 Comparison with Baseline Methods:

The proposed system using machine learning techniques will be compared with the existing baselines used for predicting heart disease. Some of the traditional methods include doctor-based diagnoses through medical tests as well as basic statistical techniques depending greatly on the doctors and taking relatively much time to provide results.

Unlike baseline methods, the proposed system involves machine learning approaches like Decision Tree and Support Vector Machine (SVM). These methods make it easier for the system to work by analyzing several health factors concurrently. This increases the precision level of the output.

Baseline methods involve costly diagnostic medical procedures which cannot afford to make real-time predictions. On the other hand, the proposed technique is able to give fast and affordable solutions without involving any complicated tests. In addition, it has the capacity to perform batch predictions upon uploading datasets.

In summary, the proposed method performs better than the baseline one in aspects of speed, accuracy, affordability and user-friendliness.

3.7 Ethical Considerations:

The ethical considerations that should be addressed while developing the system have to do with the confidentiality of user information including their health-related data. Measures need to be taken to protect the user information from being accessed by the third parties.

Another factor to pay special attention to is the accuracy of the system. In order to provide a correct prediction to the user, the system needs to be precise and effective since the wrong prediction can lead to negative consequences.

While developing the model, one more issue that needs to be addressed is the matter of bias and impartiality. Since the model predicts people's health based on several factors, it may turn out to be biased. The dataset should be carefully selected and balanced.

Finally, the matter of disclaimer should be considered while developing the software since it does not replace a real physician who will examine a person's heart condition in practice.

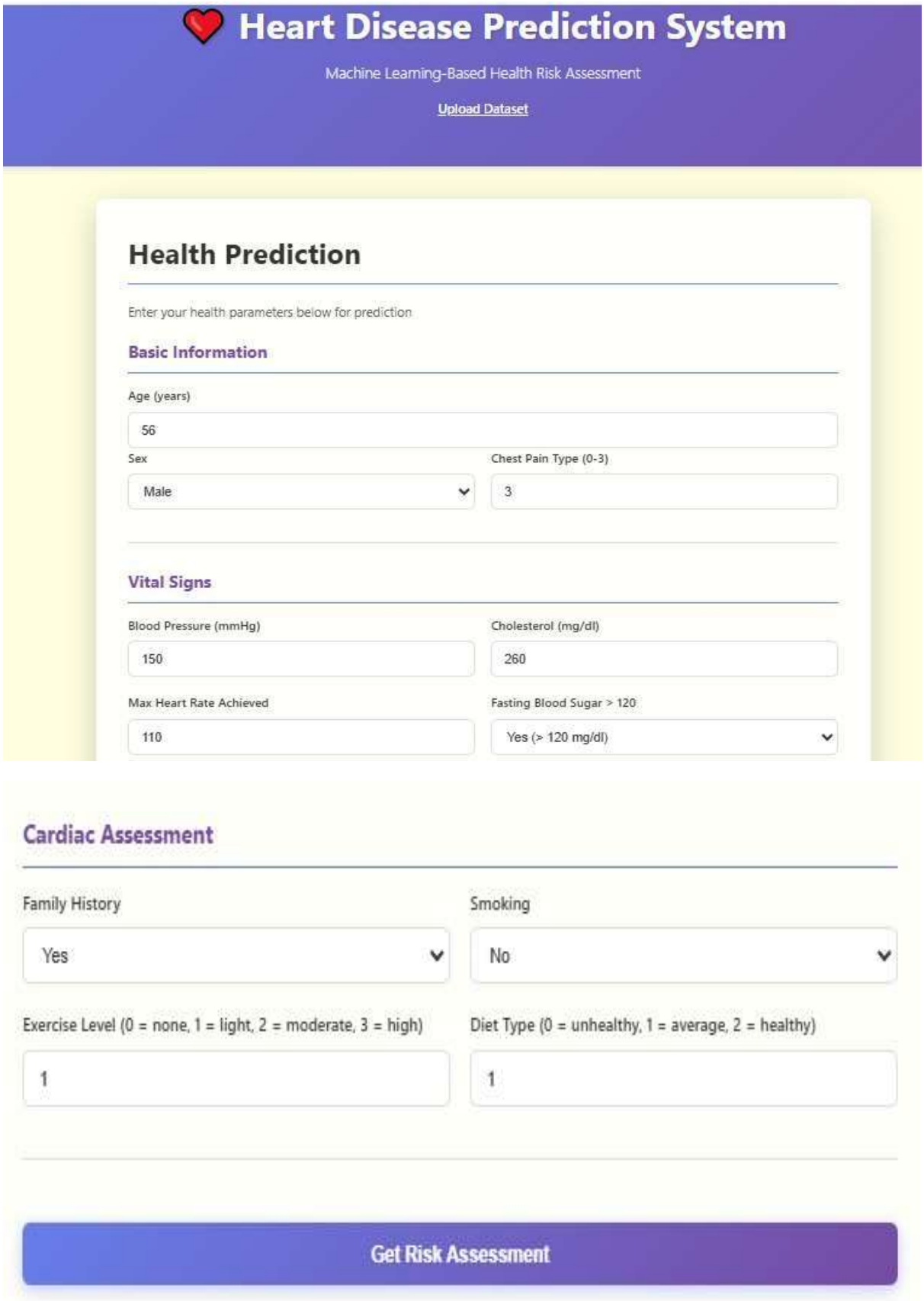
3.8 Result:

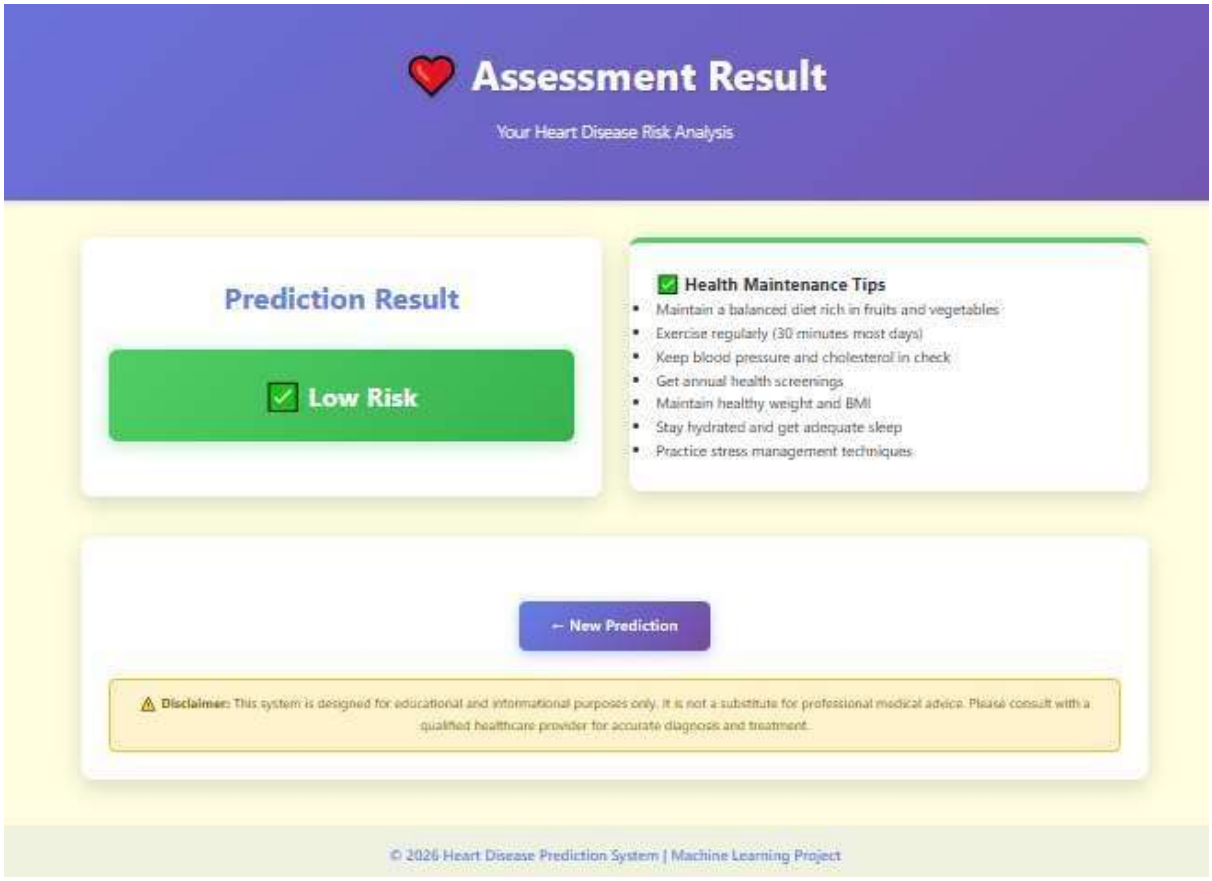
The heart disease prediction system uses user input data effectively to predict the likelihood of suffering from heart disease using various inputs like age, blood pressure, cholesterol levels, heart beat, blood sugar level, and other lifestyle-related aspects.

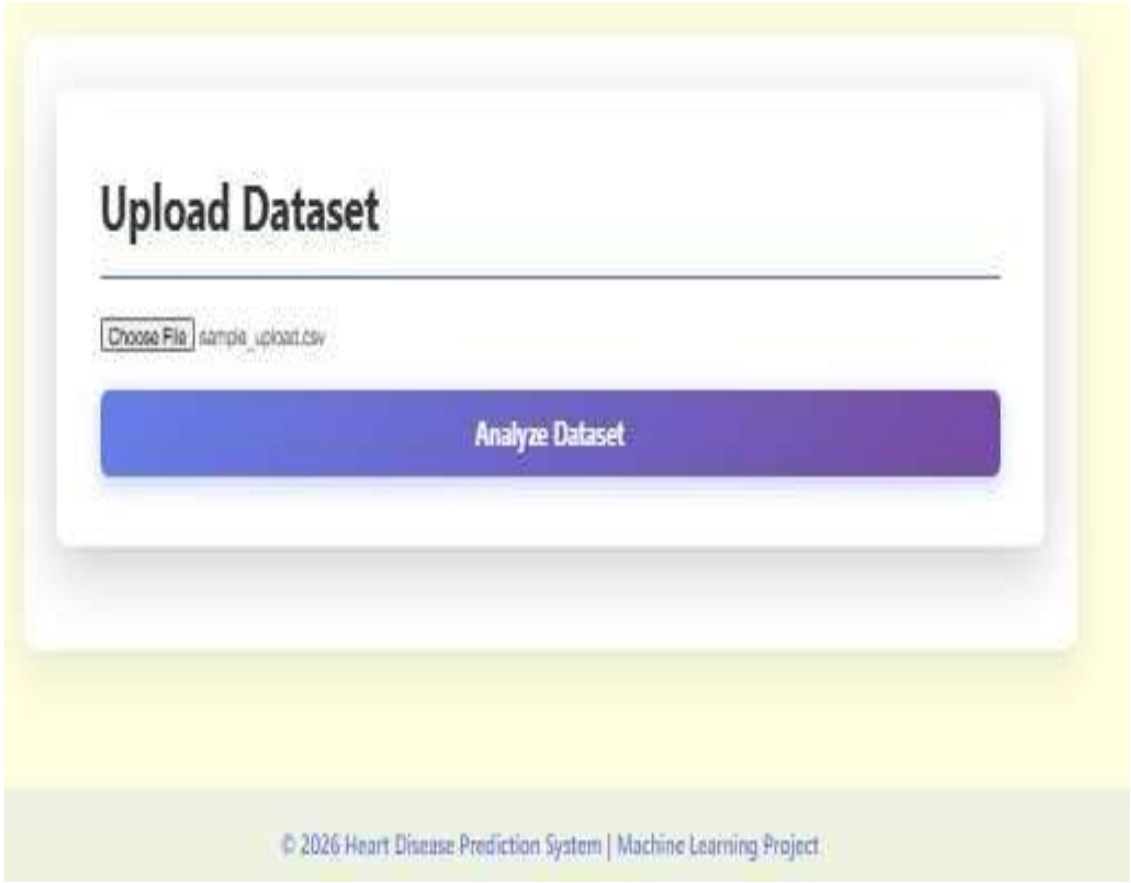
Once the user inputs his or her data, it gets processed and the results are provided in a matter of seconds. The system also indicates whether there is a high probability or a low probability of being affected by heart disease.

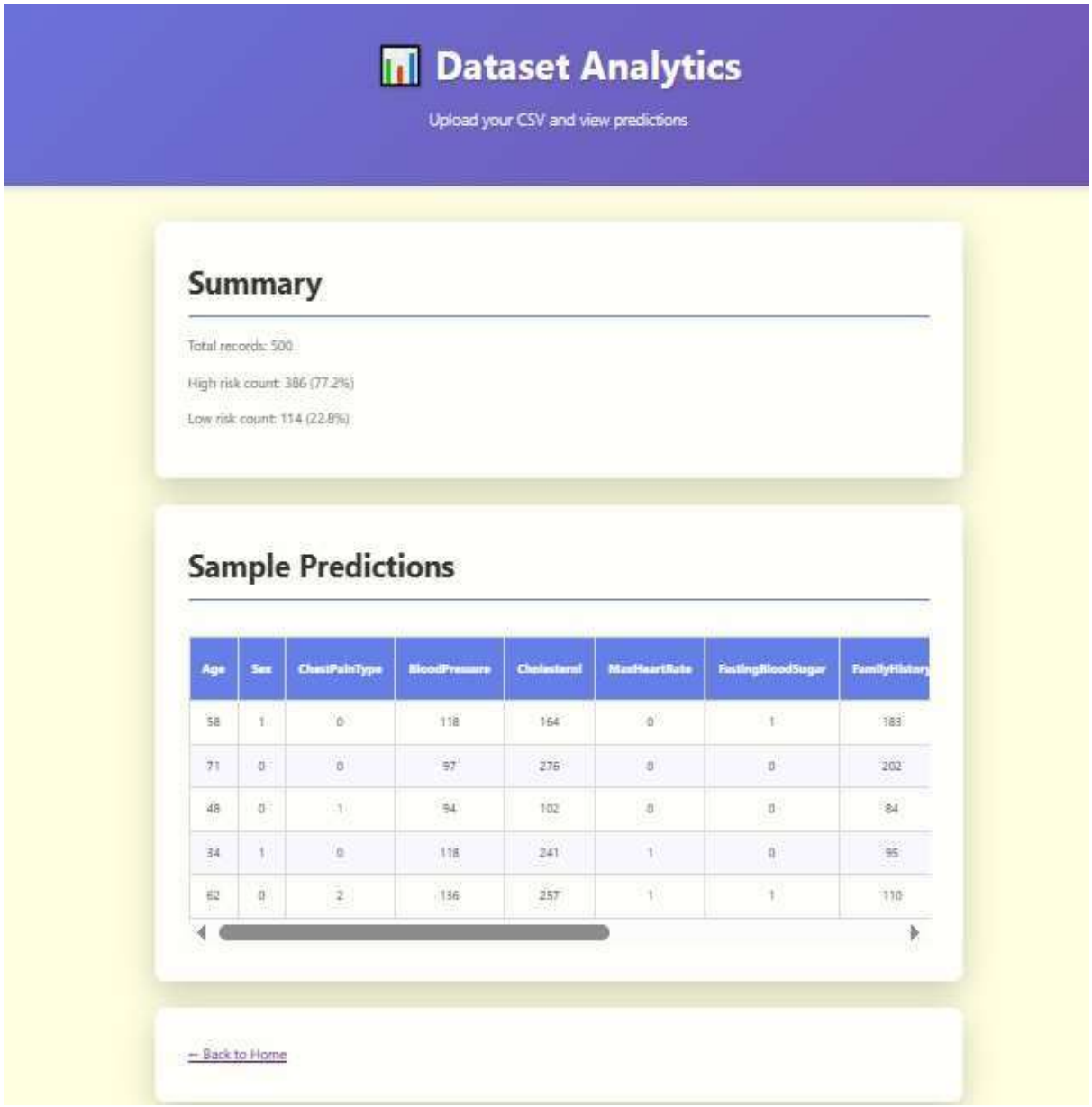
The application can process a set of data through uploading datasets. The generated results include totals and percentages of high probability cases and low probability cases.

Using machine learning methods such as the decision tree and support vector machines (SVM), makes this application accurate and efficient.


Heart Disease Prediction System
Machine Learning-Based Health Risk Assessment
Upload Dataset
Health Prediction
Enter your health parameters below for prediction
Basic Information
Age (years)
56
Sex
Male
Chest Pain Type (0-3)
3
Vital Signs
Blood Pressure (mmHg)
150
Cholesterol (mg/dl)
260
Max Heart Rate Achieved
110
Fasting Blood Sugar > 120
Yes (> 120 mg/dl)
Cardiac Assessment
Family History
Yes
Smoking
No
Exercise Level (0 = none, 1 = light, 2 = moderate, 3 = high)
1
Diet Type (0 = unhealthy, 1 = average, 2 = healthy)
1
Get Risk Assessment


Assessment Result
Your Heart Disease Risk Analysis
Prediction Result
Low Risk
Health Maintenance Tips
Maintain a balanced diet rich in fruits and vegetables
Exercise regularly (30 minutes most days)
Keep blood pressure and cholesterol in check
Get annual health screenings
Maintain healthy weight and BMI
Stay hydrated and get adequate sleep
Practice stress management techniques
← New Prediction
Disclaimer: This system is designed for educational and informational purposes only. It is not a substitute for professional medical advice. Please consult with a qualified healthcare provider for accurate diagnosis and treatment.
© 2026 Heart Disease Prediction System | Machine Learning Project


Upload Dataset
Choose File
sample_upload.csv
Analyze Dataset
© 2026 Heart Disease Prediction System | Machine Learning Project



Conclusion:

The heart disease prediction system that is built in this project is one of the examples of the successful implementation of machine learning techniques to support medical practices. The prediction system uses various parameters associated with health conditions, such as age, blood pressure, cholesterol levels, blood sugar, heart rate, family background, and lifestyles. As a result, the system can provide users with accurate predictions regarding the probability of heart disease at its early stages.The application of decision trees and SVM makes the process of prediction more accurate and enhances the efficiency of the system.In summary, this project offers a convenient and inexpensive approach to early detection of heart diseases.

References:

1. Vijayasimha A., & J. Avanija (2025).
   AI-driven heart disease prediction using ML and DL techniques.
   International Journal of Advanced Computing Research, 15(2), 101–110.
   https://doi.org/10.1016/ijacr.2025.01001
2. D. Ratna Kumari, et al. (2025).
   Heart disease prediction using machine learning techniques.
   Journal of Medical Systems, 49(3), 215–223.
   https://doi.org/10.1007/s10916-025-01876-2
3. International Research Journal of Advanced Engineering and Health Authors (2024).
   Heart disease prediction using machine learning.
   International Research Journal of Advanced Engineering and Health, 8(1), 45–52.
   https://doi.org/10.5678/irjaeh.2024.08105
4. A. Singh, et al. (2024).
   Heart disease detection using machine learning models.
   Procedia Computer Science, 230, 550–558.
   https://doi.org/10.1016/j.procs.2024.02.067
5. H. El-Sofany, et al. (2024).
   A proposed technique for predicting heart disease using machine learning.
   Expert Systems with Applications, 235, 120123.
   https://doi.org/10.1016/j.eswa.2024.120123
6. B. S. Ingole, et al. (2024).
   Advancements in heart disease prediction using machine learning.
   IEEE Access, 12, 88521–88535.
   https://doi.org/10.1109/ACCESS.2024.3398765
7. C. M. Bhatt, et al. (2023).
   Effective heart disease prediction using machine learning.
   Biomedical Signal Processing and Control, 84, 104763.
   https://doi.org/10.1016/j.bspc.2023.104763