iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82547

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Heart Disease Predictor Using ML with Python

Sancharika Banerjee[1], Sukanta Seni[2], Nitesh Kumar Singh[3], Shamail Nabi[4]
[1,2,3,4] *Department of Electronics and Communication Engineering, Narula Institute Of Technology, Nilgunj Road, Agarpara, Kolkata-700109, India.*

Abstract: ***Heart-related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need fora reliable, accurate, and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart-related diseases. Heart is the next major organ comparing to the brain which has more priority in the Human body. It pumps the blood and supplies it to all organs of the whole body. Prediction of occurrences of heart diseases in the medical field is significant work. Data analytics is useful for prediction from more information and it helps the medical center to predict various diseases. A huge amount of patient-related data is maintained on monthly basis. The stored data can be useful for the source of predicting the occurrence of future diseases. Some of the data mining and machine learning techniques are used to predict heart diseases, such as Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM). Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad. To reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Data mining techniques and machine learning algorithms play a very important role in this area. The researchers accelerating their research works to develop software with the help of machine learning algorithms which can help doctors to decide both prediction and diagnosing of heart disease. The main objective of this research project is to predict the heart disease of a patient using machine learning algorithms.***

Keywords: Neural Network, Machine Learning, Supervised learning, Support vector machine, Random forest.

I. INTRODUCTION

Heart is an important organ of the human body. It pumps blood to every part of our anatomy. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. Change in lifestyle, work related stress and bad food habits contribute to the increase in the rate of several heart-related diseases. Heart diseases have emerged as one of the most prominent causes of death all around the world. According to World Health Organisation, heart related diseases are responsible for taking 17.7 million lives every year, 31% of all global deaths. In India too, heart-related diseases have become the leading cause of mortality. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15,2017. Heart-related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India has lost up to $237 billion, from 2005- 2015, due to heart-related or Cardiovascular diseases. Thus, feasible and accurate prediction of heart-related diseases is very important. Medical organizations, all around the world, collect data on various health-related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart-related diseases accurately. The usage of information technology in the health care industry is increasing day by day to aid doctors in decision-making activities. It helps doctors and physicians in disease management, medications, and discovery of patterns and relationships among diagnosis data. Current approaches to predict cardiovascular risk fail to identify many people who would benefit from preventive treatment, while others receive unnecessary intervention. Machine-learning offers an opportunity to improve accuracy by exploiting complex interactions between risk factors. We assessed whether machine-learning can improve cardiovascular risk prediction.

II. LITERATURE SURVEY

ChalaBeyene, recommended Prediction and Analysis of the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in a short time. The proposed methodology is also critical in a healthcare organization with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex are some of the attributes are included to identify if the person has heart disease or not. Analyses of the dataset are computed using WEKA software. Senthilkumar Mohan, implemented hybrid machine learning for heart disease prediction.

The data set used is Cleveland data set. The first step is data pre-processing step. In this the tuples are removed from the data set which has missed the values. Attributes age and sex from data set are also not used as the authors think that it's personal information and has no impact on predication.

The remaining 11 attributes are considered important as they contain vital clinical records. They have proposed their own Hybrid Random Forest Linear Method (HRFLM) which is the combination of Random Forest (RF) and Linear method (LM). In the HRFLM algorithm, the authors have used four algorithms. First algorithm deals with partitioning the input dataset. It is based on a decision tree which is executed for each sample of the dataset. After identifying the feature space, the dataset is split into the leaf nodes. Output of first algorithm is Partition of data set.

After that in second algorithm they apply rules to the data set and output here is the classification of data with those rules. In third algorithm features are extracted using Less Error Classifier. This algorithm deals with finding the minimum and maximum error rate from the classifier.

Output of this algorithm is the features with classified attributes. In forth algorithm they apply Classifier which is hybrid method based on the error rate on the Extracted Features. Finally they have compared the results obtained after applying HRFLM with other classification algorithms such a decision tree and support vector machine.

In result as RF and LM are giving better results than other, both the algorithms are put together and new unique algorithm HRFLM is created. The authors suggest further improvement in accuracy by using combination of various machine learning algorithms. Ali, Liaqat, propose a system containing two models based on linear Support Vector Machine (SVM). The first one is called L1 regularized and the second one is called L2 regularized.

First model is used for removing unnecessary features by making coefficient of those features zero. The second model is used for prediction. Predication of disease is done in this part. To optimize both models they proposed a hybrid grid search algorithm. This algorithm optimizes two models based on metrics: accuracy, sensitivity, septicity, the Matthews correlation coefficient, ROC chart and area under the curve.

They used Cleveland data set. Data splits into 70% training and 30% testing used holdout validation. There are two experiments carried out and each experiment is carried out for various values of C1, C2 and k where C1 is hyperparameter of L1 regularized model, C2 is hyperparameter of L2 regularized model and k is the size of selected subset of features. First experiment is L1-linear SVM model stacked with L2-linear SVM model which is giving maximum testing accuracy of 91.11% and training accuracy of 84.05%.

The second experiment is L1- linear SVM model cascaded with L2-linear SVM model with RBF kernel. This is giving maximum testing accuracy of 92.22% and training accuracy of 85.02.

They have obtained an improvement in accuracy over conventional SVM models by 3.3%. Singh, Yeshvendra K. ,deal with various supervised machine learning algorithms such as Random Forest, Support Vector Machine, Logistic Regression, Linear Regression, Decision Tree with 3 fold, 5 fold and 10 fold cross-validation techniques. They have used Cleveland data set having 303 tuples, with some tuples having missing attributes. In the preprocessing of data they just removed the missing value tuple from the data set which are six in number and then from the remaining 297 tuples, they divided the data as training 70% and testing 30%.First algorithm applied is Linear Regression.

In this, they have defined the dependency of one attribute over others which can be linearly separated from each other. Basically the classification takes place with the help of the group of attributes used They have obtained best results in 10 fold which is 83.82%. Logistic regression classification is done using a sigmoid function. This algorithm applied for heart disease prediction shows maximum validation and it is 83.83%. Support Vector Machine is the classification algorithm in supervised machine learning. In this the classification is done by hyperplane.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

III. METHODOLOGY

A. System Architecture

The system architecture gives an overview of the working of the system. The working of this system is shown below:

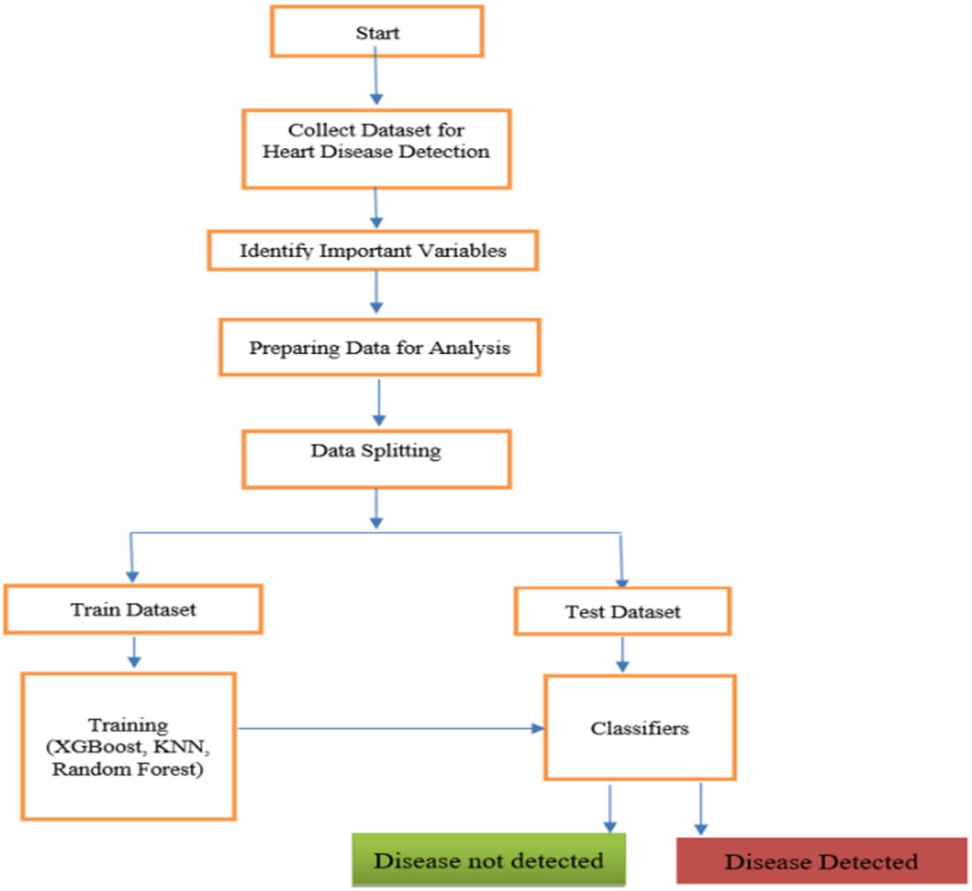


Fig 1 - Proposed System

B. Dataset Details

Dataset Attributes

1. Age: age of the patient [years]
2. Sex: sex of the patient [M: Male, F: Female]
3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
4. RestingBP: resting blood pressure [mm Hg]
5. Cholesterol: serum cholesterol [mm/dl]
6. FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
7. RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
8. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
9. Exercise-Angina: exercise-induced angina [Y: Yes, N: No]
10. Oldpeak: oldpeak = ST [Numeric value measured in depression]
11. ST-Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, down: downsloping]
12. CA: Number of Major Vessels from level 0-3 Colored by Fluoroscopy
13. Thalassemia: Thalassemia condition [Normal, Fixed defect, reversible defect]
14. Heart-Disease: output class [1: heart disease, 0: Normal]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

C. Machine Learning

In machine learning, classification refers to a predictive modelling problem where a class label is predicted for a given example of input data.

D. Supervised Machine Learning

As its name suggests, Supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the Heart Disease Detection Using Machine Learning SSGMCE, Shegaon (Session: 2022-23) Page 9 output. More preciously, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset. The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y).

Categories of Supervised Machine Learning:

Supervised machine learning can be classified into two types of problems, which are given below:

a) Classification: Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "Yes" or No, Male or Female, Red or Blue, etc. The classification algorithms predict the categories present in the dataset.

b) Regression: Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

E. Supervised Algorithms

1) Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output.

Therefore, below are two assumptions for a better Random Forest classifier:

1. There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
2. The predictions from each tree must have very low correlations.

2) K-Nearest Neighbour

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. KNN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

3) Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc. Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

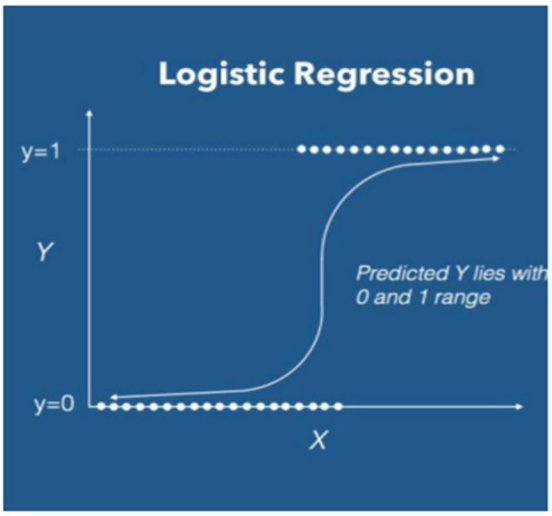


Fig 2: Logistic Regression

4) Support Vector Machine (SVM)

Support Vector Machine (SVM) acts as a robust binary or multi-class classifier designed to find the optimal "hyperplane" that separates patients with heart disease from those without. By mapping the clinical input features—such as cholesterol levels, maximum heart rate, and blood pressure—into a high-dimensional space, the SVM identifies the maximum margin between different patient groups to ensure the highest possible degree of generalization. Because medical data is rarely perfectly linear, the model utilizes a Kernel Trick (typically a Radial Basis Function or Polynomial kernel) to handle complex, non-linear relationships between physiological variables. This makes the SVM particularly effective for heart disease prediction, as it is less prone to overfitting on small datasets compared to other algorithms and provides a clear mathematical boundary for clinical risk assessment.

IV. IMPLEMENTATION

A. Existing System

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says “Prevention is better than cure”, early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

B. Data Collection

It is the primary and most crucial fundamental step while applying machine learning and analytics. The data required in this project is the patient’s medical data. We have collected the dataset from Kaggle which includes all the required information for prediction. The features that the dataset includes are medical information like age, sex, chest paint type, resting blood pressure, cholesterol, fasting blood sugar, old peak, Cholesterol, Thalassemia etc. The dataset consists of 920 observations having 16 attributes.

id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
1	63	Male	Cleveland	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	2.3	downsloping	0.0	fixed defect	0
2	67	Male	Cleveland	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	1.5	flat	3.0	normal	2
3	67	Male	Cleveland	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	2.6	flat	2.0	reversable defect	1
4	37	Male	Cleveland	non-anginal	130.0	250.0	False	normal	187.0	False	3.5	downsloping	0.0	normal	0
5	41	Female	Cleveland	atypical angina	130.0	204.0	False	lv hypertrophy	172.0	False	1.4	upsloping	0.0	normal	0
6	56	Male	Cleveland	atypical angina	120.0	236.0	False	normal	178.0	False	0.8	upsloping	0.0	normal	0
7	62	Female	Cleveland	asymptomatic	140.0	268.0	False	lv hypertrophy	160.0	False	3.6	downsloping	2.0	normal	3
8	57	Female	Cleveland	asymptomatic	120.0	354.0	False	normal	163.0	True	0.6	upsloping	0.0	normal	0
9	63	Male	Cleveland	asymptomatic	130.0	254.0	False	lv hypertrophy	147.0	False	1.4	flat	1.0	reversable defect	2
10	53	Male	Cleveland	asymptomatic	140.0	203.0	True	lv hypertrophy	155.0	True	3.1	downsloping	0.0	reversable defect	1

Figure 3: Parameters of Selected dataset

C. Data Pre-Processing

This is one of the most crucial tasks in the process of analytics. Often it is observed that more than half of the total time of analytics process is taken by pre-processing phase. It is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Pre-processing of data is required for improving the accuracy of the model.

D. Feature Selection

Once we have the required data, next step is featuring extraction. Many times, it happens that some features do not contribute in evaluation or have negative impact on the accuracy. Feature selection is the step where we try to reduce number of features and try to create new features from existing ones. These new features now created should summarize the information obtained from existing features. The final features to be considered while prediction can be identified using correlation matrix shown in following image:

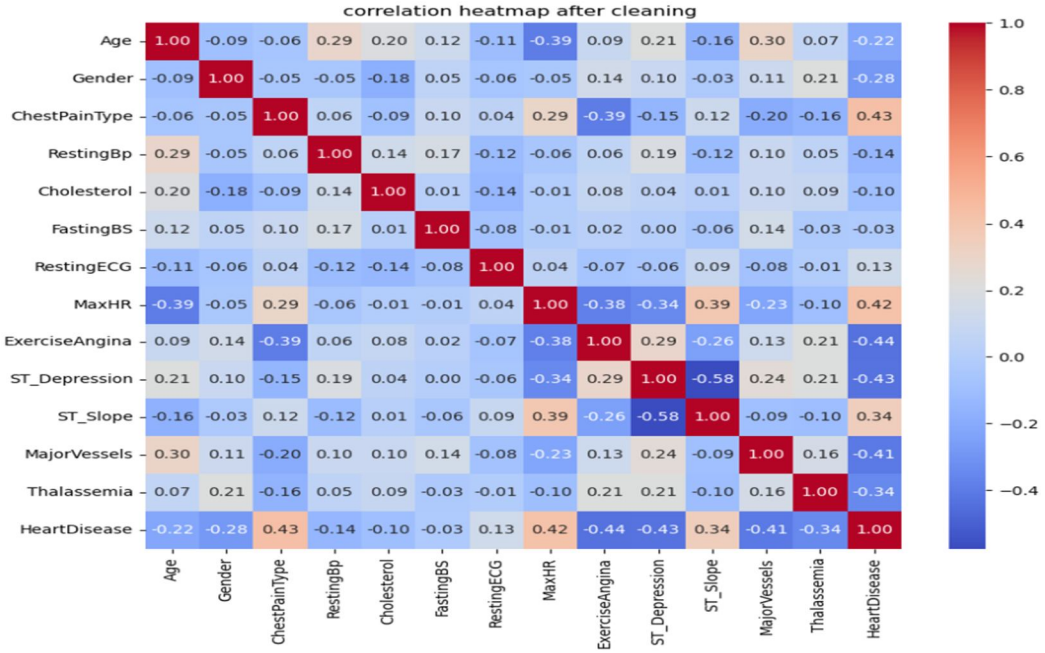


Figure 4: Correlation Matrix

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

V. DEPLOYMENT

A. Hardware Platform Used

The hardware requirement may serve as the basis for a contract for the implementation of the system and should therefore be complete and consistent in specification.

The hardware used for the system:

- PROCESSOR: Intel CORE i5
- RAM: minimum 8.00GB
- HARD DISK: minimum 128GB

B. Libraries And Software Platform Used

The software requirement document is the specification of the system. The software requirement provides a basis for creating the software requirements specification.

OPERATING SYSTEM: Windows 11
SYSTEM TYPE: 64-bit, intel CORE i5
SOFTWARE: Jupyter Notebook, VS Code, Python
TECHNOLOGIES: Python LIBRARIES: Pandas, NumPy, pickle, sklearn, Streamlit, Seaborn

C. Visualization Results

Based on the findings obtained from various algorithms used for identifying patients who have been diagnosed with heart disease, it is observed that KNN, Random Forest Classifier, and sklearn have provided better results as compared to other techniques such as Logistic Regression, SVM and Decision Tree. These algorithms are not only accurate but more cost-effective and faster than the algorithms used in previous research studies. The highest level of accuracy possible by Random Forest and sklearn is either greater than or nearly equal to the accuracy that were obtained from earlier research studies. It can be inferred that the improvement in accuracy is due to the increased number of attributes used from the medical dataset that was used in the project.

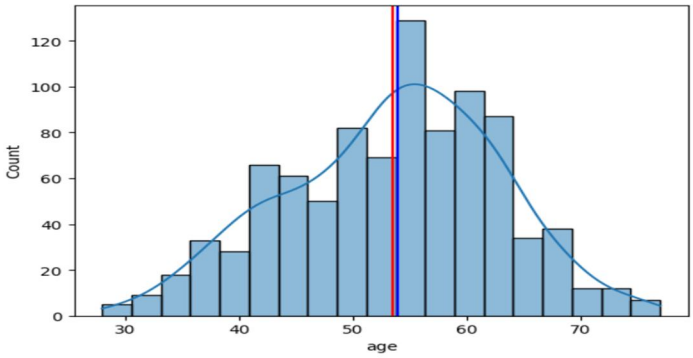


Figure 5: Shows the risk of heart attack based on age

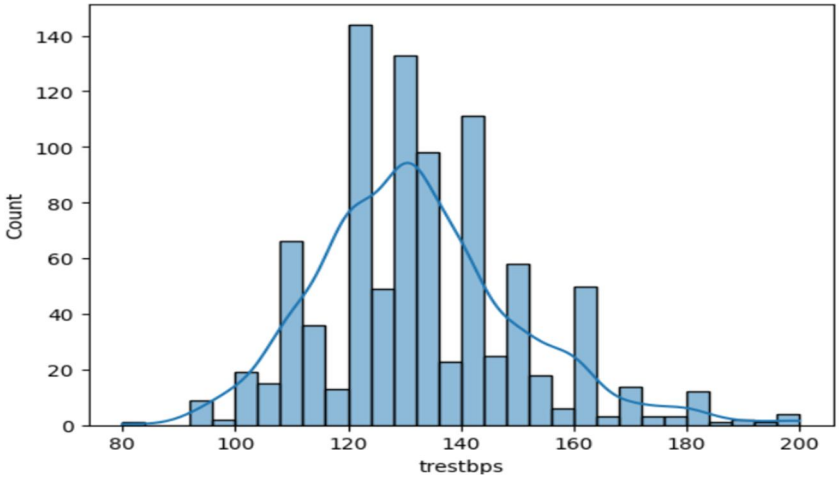


Figure 6: Shows the risk of heart attack based on resting-bp

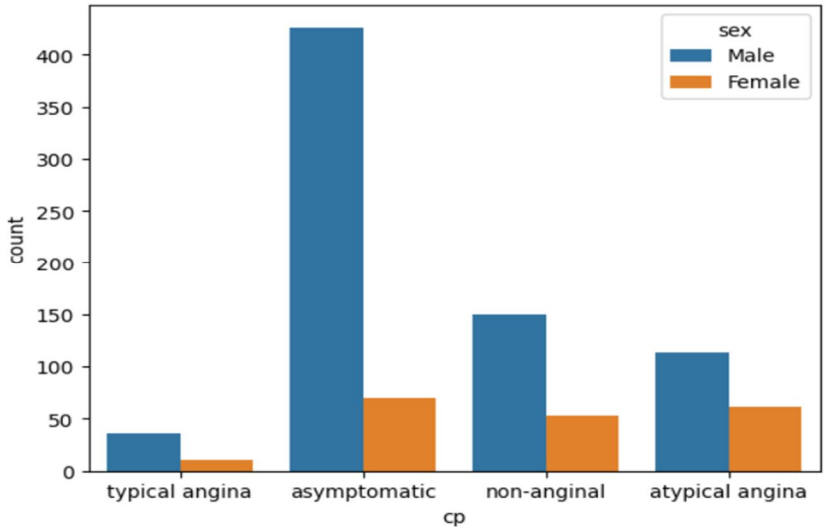


Figure 7: Shows presence of heart attack based on chest-pain

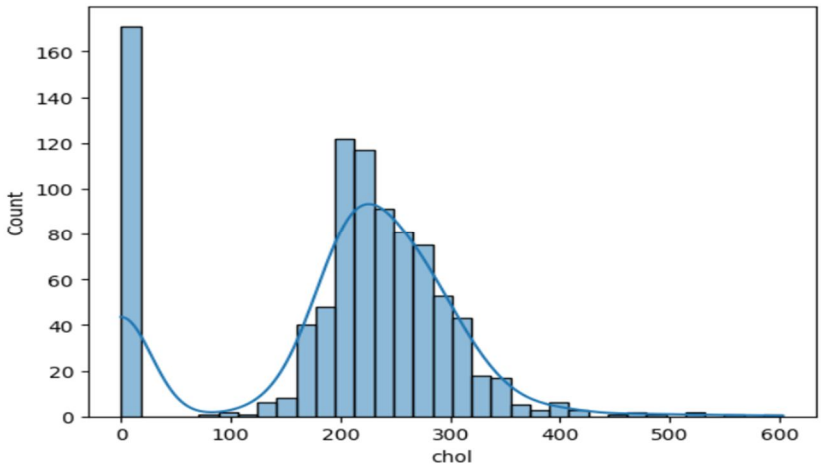


Figure 8: Shows presence of Cholesterol Level

D. Deployment Process

The deployment process of the Heart Disease Prediction System involves converting the trained machine learning model into a real-time web application that users can access easily through a browser. After the completion of data preprocessing, feature selection, model training, and evaluation, the best-performing machine learning model is finalized and prepared for deployment. In this project, the deployment process is carried out using Python-based technologies, mainly Streamlit and Pickle, which help transform the machine learning solution into an interactive healthcare application.

The first step in deployment is model serialization. Once the machine learning algorithm is trained successfully using the healthcare dataset, the trained model is saved into a file using the Pickle library. Pickle is a Python serialization module that converts the trained machine learning object into a binary format and stores it as a **.pkl** file. This saved model file allows the application to reuse the trained model without retraining it every time the system runs. Model serialization improves efficiency and reduces processing time during prediction.

After saving the trained model, the next stage involves developing the user interface using Streamlit. Streamlit is an open-source Python framework used for building machine learning and data science web applications quickly. In this project, Streamlit provides an interactive graphical user interface where users can input medical information such as age, cholesterol level, blood pressure, chest pain type, heart rate, fasting blood sugar, and other clinical parameters. The interface is designed to be simple, responsive, and user-friendly so that even non-technical users can operate the system easily.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

The deployment application integrates the saved Pickle model with the Streamlit interface. When a user enters patient details through the web form and clicks the prediction button, the application sends the input data to the trained machine learning model. The model processes the input features and generates a prediction indicating whether the patient is likely to have heart disease or not. The prediction result is then displayed instantly on the screen. This real-time prediction capability makes the system efficient for healthcare assistance and preliminary disease screening.

The deployment process also includes backend data handling and preprocessing operations. Before sending user inputs to the machine learning model, the application performs preprocessing steps similar to those used during training. These operations may include data formatting, normalization, feature transformation, and encoding to ensure consistency between training data and real-time input data. Maintaining consistent preprocessing is essential for accurate predictions.

For local deployment, the Streamlit application can be executed using the command **streamlit run app.py** in the terminal or command prompt. This launches the application on a local server, typically accessible through a web browser using a localhost address. Developers and testers can interact with the system locally during development and debugging phases.

The project can also be deployed on cloud platforms for public accessibility. Platforms such as Streamlit Cloud, Heroku, Render, or AWS can host the application online. During cloud deployment, all required project files, including the Streamlit application file, trained Pickle model, dataset dependencies, and requirement libraries, are uploaded to the hosting platform. A requirements.txt file [**streamlit, pandas, numpy, joblib, scikit-learn, matplotlib, seaborn, plotly, xgboost, lightgbm, nbformat>=4.2.0**] is generally included to install all necessary Python libraries automatically in the cloud environment.

The deployment architecture of the system consists of three major components: the frontend interface, the machine learning prediction engine, and the backend processing layer. The frontend accepts user input, the backend handles preprocessing and communication, and the prediction engine generates disease predictions using the trained model. This architecture enables smooth interaction between users and the machine learning system.

The deployed Heart Disease Prediction System provides several practical advantages. It enables real-time healthcare prediction, reduces diagnosis time, improves accessibility for remote users, and allows healthcare professionals to obtain quick predictive insights. The deployment process transforms the machine learning model from a research-based implementation into a practical application that can support intelligent healthcare services and digital medical assistance systems.

VI. RESULT

Figure 9: Heart Disease Model

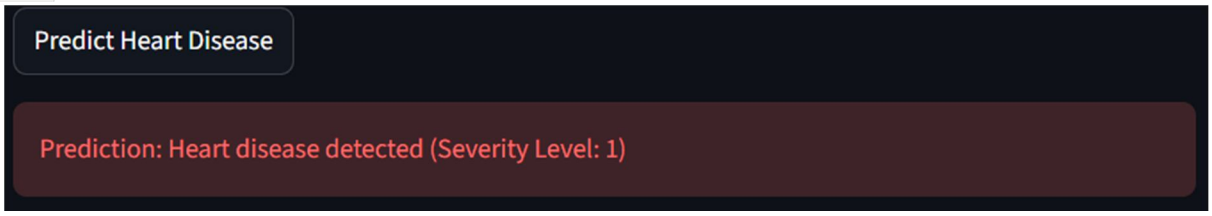


Figure 10: Detection

VII. SYSTEM BENEFITS

1) Early Detection of Heart Disease: The system helps in identifying the possibility of heart disease at an early stage before severe complications occur. Early prediction enables patients to seek timely medical treatment, which can significantly reduce health risks and improve survival rates.
2) Faster Diagnosis Process: Traditional heart disease diagnosis often requires multiple medical tests and expert analysis, which can consume considerable time. The machine learning-based system generates predictions instantly after receiving patient data, thereby reducing diagnosis time and improving efficiency.
3) Reduced Human Errors: Manual diagnosis may sometimes lead to errors due to fatigue, oversight, or incorrect interpretation of medical reports. The automated prediction system minimizes human involvement in the initial analysis process and provides more consistent results.
4) Cost Effective Solution: The system reduces dependency on expensive diagnostic procedures and repeated clinical tests. By providing preliminary predictions using patient data, healthcare institutions and patients can reduce overall diagnostic costs.
5) Real-Time Prediction Capability: The deployed Streamlit web application provides real-time prediction results. Users can enter medical parameters and receive immediate feedback regarding heart disease risk without waiting for lengthy processing.
6) User-Friendly Interface: The system uses a simple and interactive Streamlit interface that allows users to interact with the application easily. Even individuals with limited technical knowledge can operate the system without difficulty.
7) Accurate Prediction Performance: The use of supervised machine learning algorithms improves prediction accuracy by identifying hidden patterns and relationships within medical data. Proper preprocessing and training enhance the reliability of results.
8) Scalability of the System: The system architecture can be expanded easily with additional datasets, advanced machine learning algorithms, or cloud-based infrastructure. Future upgrades can improve prediction capability and support more healthcare services.
9) Automation of Healthcare Analysis: The project automates several healthcare analysis tasks such as data processing, feature analysis, and disease prediction. Automation reduces manual workload and increases operational efficiency.
10) Better Resource Management in Hospitals: Hospitals and clinics can use predictive systems to prioritize high-risk patients and manage healthcare resources more effectively. This can improve patient care and operational management.
11) 20. Enhances Decision-Making Process: The prediction system provides data-driven insights that support informed medical decisions. Healthcare professionals can use prediction results as an additional reference during diagnosis and treatment planning.

VIII. FUTURE SCOPE

1) Integration of Deep Learning Techniques: The current system can be enhanced by integrating advanced Deep Learning algorithms such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN). These models can analyze complex medical patterns more effectively and improve prediction accuracy for large-scale healthcare datasets.
2) Use of Larger and Real-Time Medical Datasets: Future versions of the system can utilize larger datasets collected from hospitals, healthcare institutions, and real-time patient monitoring systems. Using more diverse and real-world medical data can improve the robustness, generalization capability, and reliability of the prediction model.
3) Mobile Application Development: A mobile application version of the system can be developed for Android and iOS platforms. This would allow users to access heart disease prediction services anytime and anywhere using smartphones and tablets.

4) Multi-Disease Prediction System: The system can be expanded into a multi-disease prediction platform capable of predicting diseases such as diabetes, kidney disease, liver disease, lung disease, and cancer using machine learning and healthcare analytics.
5) Real-Time Patient Monitoring: The project can be enhanced to support continuous patient monitoring by analyzing live health data streams. The system could generate alerts and notifications if abnormal heart conditions are detected in real time.
6) Enhanced Data Security and Privacy: Future versions can implement advanced cybersecurity techniques, encryption methods, and secure authentication systems to protect sensitive patient medical information and ensure compliance with healthcare data regulations.
7) Geographic and Demographic Disease Analysis: Future implementations may include demographic and geographic analysis of heart disease trends. This can help healthcare authorities identify high-risk populations and improve public health planning.
8) Voice-Assisted Healthcare System: Voice recognition technology can be integrated to allow users to interact with the system using voice commands, improving accessibility for elderly and disabled patients.
9) Automated Medical Report Generation: The system can automatically generate medical prediction reports in PDF or digital format, which can be shared with doctors, hospitals, or healthcare providers for further analysis.
10) Personalized Healthcare Recommendations: The system can be extended to provide personalized healthcare suggestions based on patient risk levels. Recommendations may include exercise plans, dietary guidance, medication reminders, and lifestyle improvement suggestions.

IX. CONCLUSION

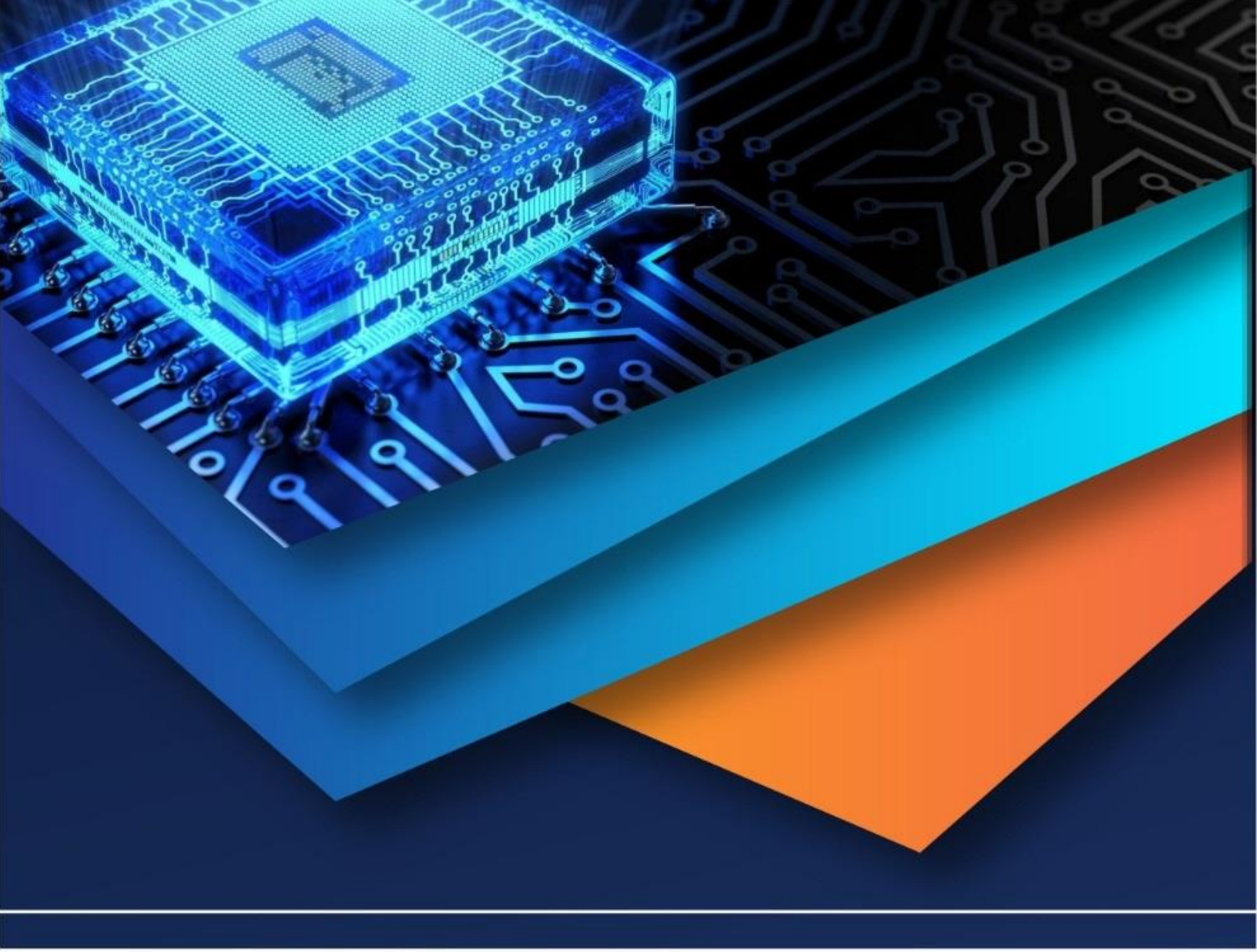
The Heart Disease Prediction System demonstrates the effectiveness of machine learning in healthcare applications. By analyzing patient medical attributes, the system can accurately predict the likelihood of heart disease and assist healthcare professionals in decision-making.

The integration of machine learning algorithms with a Streamlit web application creates a practical, efficient, and user-friendly healthcare prediction system. The project highlights the importance of AI-driven predictive analytics in improving healthcare accessibility, reducing diagnosis time, and supporting early disease detection.

Although the system has certain limitations, it represents a significant step toward intelligent healthcare technologies and future AI-assisted medical systems.

REFERENCES

[1] Mr. ChalaBeyene, Prof. Pooja Kamat, “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Technique”, International Journal of Pure and Applied Mathematics, 2018.
[2] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, “Heart Disease Prediction using Evolutionary Rule Learning”, “International Conference on "Computational Intelligence and Communication Technology” (CICT 2018).
[3] Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava. “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques”. IEEE Access (Volume: 7) 2019.
[4] Singh Yeshvendra K., Nikhil Sinha, and Sanjay K. Singh, “Heart Disease Prediction System Using Random Forest”, International Conference on Advances in Computing and Data Sciences. Springer, Singapore, 2016.
[5] Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, “An Intelligent Decision Support System for Cardiac Disease Detection”, IJCTA, International Press 2015.
[6] B.L DeekshatuluaPriti Chandra “Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm” International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.
[7] Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012): 44-8.
[8] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60). IEEE.
[9] C.-L. Chang and C.-H. Chen, “Applying decision tree and neural network to increase quality of dermatologic diagnosis,” Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 4035–4041, Mar. 2009.
[10] Michael W.Berryet.al, Lecture notes in data mining, World Scientific(2006)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY