International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

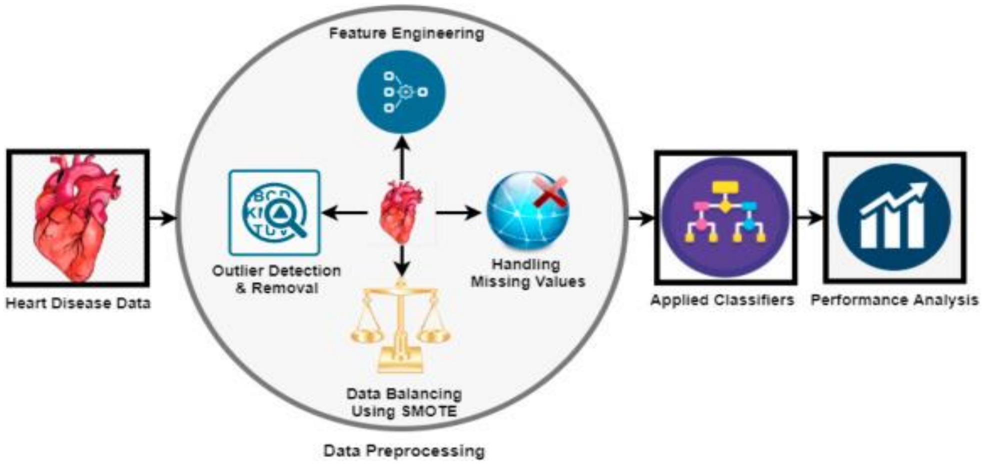
Implementation of Disease Detection and Prediction of Heart Using Machine Learning

Darsi Keerthi[1], Javvadi Shanmuka[2], P. Hima Bindu[3], Venkata Surendra[4], Lakshmi Prasanna[5]

[1,2,3,4]Student, Kl University
[5]Professor, Kl University

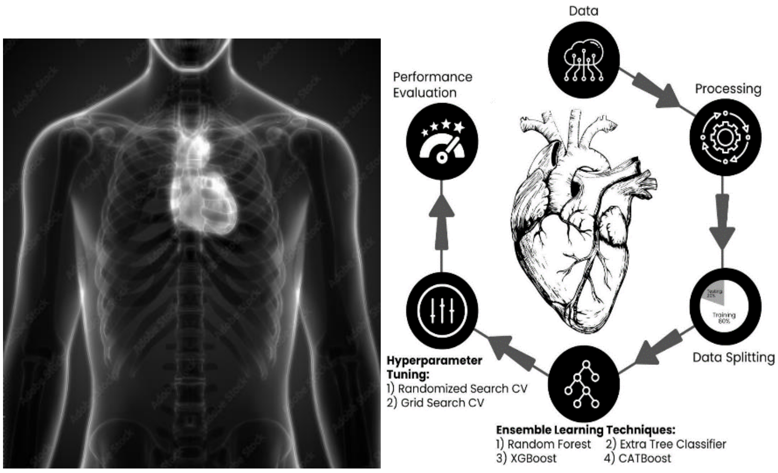
Abstract:
Heart disease detection and prediction is an important task for entrusting whether the person is healthy or not. In recent years, there has been an increasing interest in the use of feature extraction methods for this purpose. Feature extraction methods are techniques used to identify relevant information from images and other forms of data. This information can then be used to train machine learning models to classify and predict diseases.Various feature extraction methods have been proposed for heart disease detection and prediction, including color-based features, texture-based features, and shape-based features. Color-based features involve analyzing the color of blood vessels in angiography. Texture-based features involve analyzing the texture patterns within medical images, including those related to heart disease and helps to identify the texture of heart tissue in an MRI scan. Shape-based features can be used to analyze the shapes and outline of structures within medical images. Predicting heart disease using machine learning involves developing a system that can analyze various health-related data to determine the likelihood of a person having heart disease.

Keywords: Heart, Machine learning, Random Forest, Feature extraction, detection, Disease prediction.



1. Introduction

Heart disease detection and prediction using feature extraction methods is a technique used to identify and predict diseases in heart based on certain features extracted from images of the x-rays. This method uses computer vision and machine learning algorithms to analyze the images and identify any signs of disease.Feature extraction involves identifying and extracting important features from images, such as color, texture, and shape, that can be used to distinguish healthy from diseased ones. These features are then used to train machine learning models to classify images as healthy or diseased. By analyzing a person's medical history, lifestyle factors, and various health parameters, machine learning models can provide early warnings and personalized insights to help healthcare professionals and individuals make informed decisions for better heart health.



The following steps are used in predicting heart disease using relevant data:

Data Collection: The first step in predicting heart disease is gathering relevant data. This includes information about a person's age, gender, medical history, lifestyle factors (like smoking and exercise), and various health measurements such as blood pressure, cholesterol levels, and ECG readings).

Data Preprocessing: Once the data is collected, it needs to be cleaned and organized. This step involves removing errors or inconsistencies, handling missing values, and scaling or normalizing the data for accurate analysis.

Feature Selection: In this phase, we decide which factors (features) from the collected data are most important for predicting heart disease. Not all data is equally valuable, so we select the most relevant features to improve prediction accuracy.

Machine Learning Model: Now, we introduce the machine learning part. A model is trained using historical data where we know whether individuals had heart disease.

Training: The model learns from historical data, identifying patterns and relationships between the selected features and the presence or absence of heart disease.

Testing and Validation: To ensure the model's accuracy and generalization, we test it on new, unseen data that it hasn't been trained on. This helps us evaluate its performance and make improvements if necessary.

Predictive Power: Once the model is validated, it can be used to predict whether a person is at risk of heart disease based on their individual data.

Feedback Loop: The model can continuously improve itself by gathering more data and incorporating it into the training process. This ensures that it remains up-to-date and accurate.

Final Prediction: The model provides a prediction, indicating whether an individual is at low, medium, or high risk of developing heart disease.

Healthcare Intervention: Based on the prediction, healthcare professionals can take appropriate actions, such as recommending lifestyle changes, medication, or further diagnostic tests, to help prevent

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

or manage heart disease.

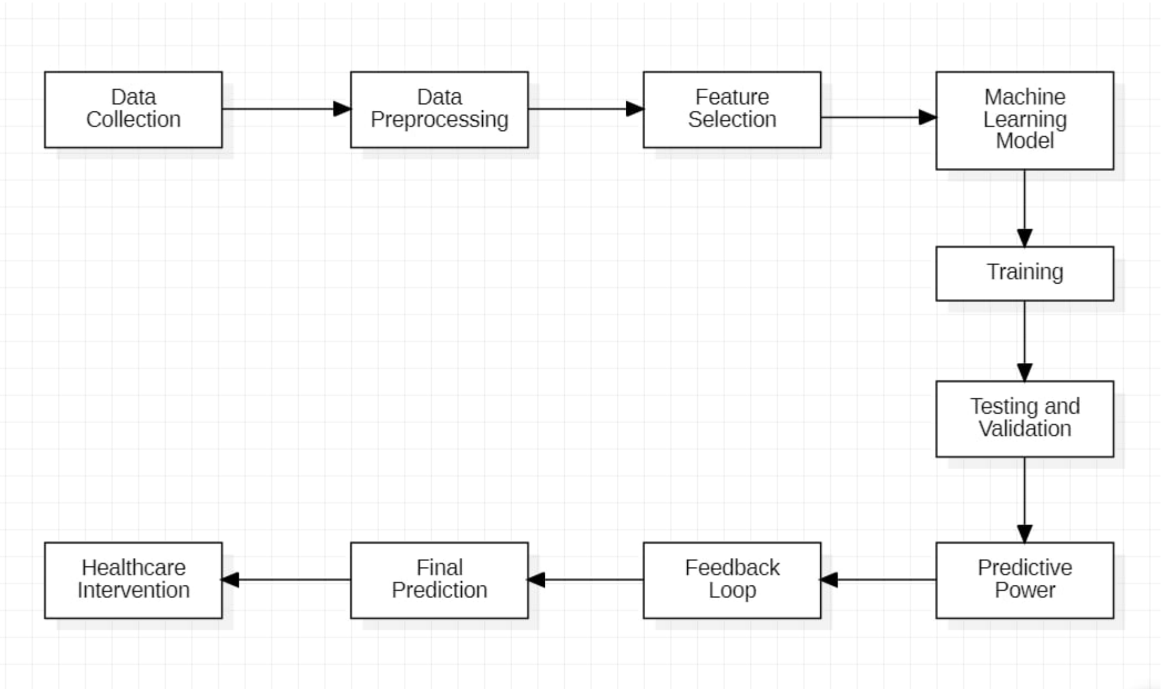


Fig.1 Flowchart Heart disease Prediction

FEATURE EXTRACTION METHODS

- Principal Component Analysis (PCA).
- Independent Component Analysis (ICA).
- Linear Discriminant Analysis (LDA).
- Wavelet Transform.
- Convolutional Neural Networks (CNN).

FEATURE EXTRACTION METHODS FOR HEART DISEASE PREDICTION AND DETECTION

- Color-Based
- Texture-Based
- Shape-Based
- Machine Learning Based

PREDICTION AND DETECTION ALGORITHMS

PREDICTION ALGORITHMS	DETECTION ALGORITHMS
Decision Tree	Regional Based Convolutional Neural Networks (R-CNN)
Random Forest Tree	Regional Based Fully Convolutional Neural Network (R-FNN)
Naïve Bayes	Single Shot Detector (SSD)
KNN	Alex Net
Artificial Neural Networks	Squeeze Net

Logistic Regression	CNN

Table 1 Data of prediction and detection Algorithms

2. Literature Survey

TITLE OF THE PAPER	AUTHOR	TECHNIQUES	YEAR OF PUBLISHED	ACCURACY
Heart Disease Prediction using Machine Learning Techniques	Devansh Shah Otoom et al	Naïve Bayes Random forest	2020	Naïve Bayes: 84.5% Random forest: 91.6%
Designing Disease Prediction Model Using Machine Learning Approach	Dhiraj Dah wade Ektaa Meshram	KNN CNN	2019	KNN: 87.5% CNN: 90.41%
Research of Heart Disease Prediction Based on Machine Learning	Shuge Ouyang	SVM Logistic regression	2022	SVM: 82% Logistic regression:85%
Heart diseaseprediction	Harshit Jindal Sarthak Agrawal	Random Forest KNN	2021	Random Forest: 87.5% KNN: 88.52%
Effective Heart Disease Prediction	Parth Patel Tarang Ghetia Pier Luigi mazzeo	Multilayer perceptron Random Forest	2023	Multilayer perceptron: 81.48% Random Forest: 87.23%
Improving Heart Disease Prediction Using Feature Selection Approaches	Saba Bashir Aitzaz Anjum Khurram Bashir	Decision Tree LR – SVM Naïve Bayes	2019	Decision Tree: 82.22% LR – SVM: 84.85% Naïve Bayes: 84.24%
An Analysis of Heart Disease Prediction using Different Data Mining Techniques	Afrin Haider Marjia Sultana	Decision Tree KNN Naïve Bayes	2016	Decision Tree: 52% KNN: 45.67% Naïve Bayes: 52.33%

3. Problem Statement

Heart disease is a significant problem worldwide, causing a lot of suffering and even death. The big issue here is that we often don't catch it early enough. When people first start showing signs of heart problems, it's usually quite advanced, and it's harder to treat effectively. Our solution is to use machine learning to predict heart disease before it gets bad, so we can tackle it earlier. Right now, one of the main problems is that we struggle to detect heart issues in our early stages. By the time symptoms show up, the disease has often progressed a lot, making it tough to take action quickly. So, our plan is to harness the power of machine learning to foresee heart problems before they become severe. Our challenge lies in the complexity of the human cardiovascular system and the vast amount of data involved. We need to develop robust algorithms that can handle diverse data types, from ECG readings to patient histories, and integrate them seamlessly. Moreover, ensuring the privacy and security of patient data is a paramount concern throughout this process. Ultimately, our goal is to create a user-friendly, accessible tool that can assist healthcare providers in making informed decisions and educate individuals about their cardiovascular health risks.

FEATURE EXTRACTION METHODS:

Feature extraction is the process of selecting and transforming relevant information from raw data into a set of features that can be used as input to a machine learning model. Here are some common feature extraction methods in machine learning:

Principal Component Analysis:

PCA is a feature extraction method used in heart disease prediction to simplify and reduce the dimensionality of the data while retaining as much relevant information as possible. PCA in heart disease prediction is like condensing a lot of information about patients into a smaller, more manageable set of variables that capture the most important aspects of their health. This streamlined data can then be used to build a more effective predictive model for identifying individuals at risk of heart disease.

Linear Discriminant Analysis:

LDA helps you find the best combination of features that can differentiate between people with heart disease and those without it. It's like finding the most critical pieces of information (features) in a puzzle that can help a doctor make a diagnosis. By reducing the dimensionality and focusing on these key features, LDA can improve the accuracy of heart disease prediction models.

Independent Component Analysis:

ICA is a feature extraction method used in heart disease prediction to uncover hidden patterns and sources of variation within complex biomedical data. In simpler terms, ICA helps us exact meaningful information from a mixture of signals and data collected from patients. For example, ICA can be used to separate the color, texture, and shape features of heart, and extract features that are most informative for disease detection and prediction.

Wavelet Transform:

Wavelet Transform serves as a valuable feature extraction technique for heart disease prediction. It helps capture important temporal and frequency domain information from physiological data, allowing machine learning models to make accurate predictions based on these extracted features. This approach can improve the accuracy and effectiveness of heart disease prediction models by considering both short-term and long-term variations in the data.

Convolutional Neural Networks:

CNN are primarily designed for image data and are not typically used as a direct feature extraction method for heart disease prediction from structured clinical data (e.g., patient demographics, lab results). However, CNNs can be utilized in combination with other techniques to extract features from medical images, such as X-rays, CT scans, or MRIs, which may contribute to heart disease prediction. CNNs are typically used to extract features from medical images, which can then be integrated with other clinical data for heart disease prediction. This hybrid approach leverages the CNN's ability to capture visual patterns in images and combines it with traditional machine learning techniques for a comprehensive prediction model.

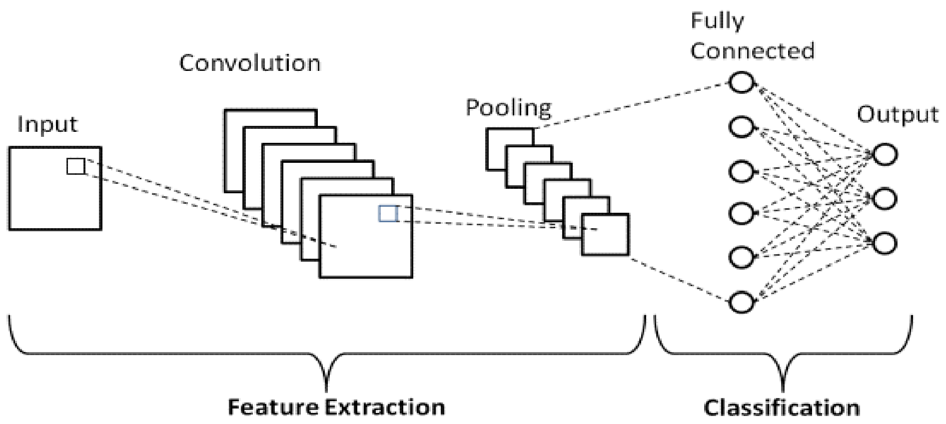


Fig. 2 Layers of Convolutional Neural Network

4. Methodology

Predicting heart disease using machine learning involves several steps. Here's step-by-step methodology in simple terms:

Data Collection:

Gather data from various sources. This includes information about patients, like age, gender, blood pressure, cholesterol levels, and medical history. Medical test results like electrocardiograms (ECGs) and imaging data (e.g., X-rays, CT scans).

Data Preprocessing:

Clean the data by handling missing values and ensuring consistency. You may need to convert categorical data (like gender) into numerical values for the machine learning model to understand.

Feature Extraction:

Identify which information in the data is most relevant for predicting heart disease. This could involve techniques like Principal Component Analysis (PCA) or Independent Component Analysis to reduce the dimensionality of the data and extract important features.

Data Splitting:

Divide your dataset into two parts: a training set and a testing set. Training set is used to teach machine learning model, while the testing set is used to evaluate how well it can make predictions.

Machine Learning Model Selection:

Choose an appropriate machine learning algorithm for your task. Common choices include logistic regression, decision trees, random forests, or more advanced techniques like neural networks.

Training the Model:

Feed the training data into your chosen model so that it can learn the patterns and relationships in the data. The model will adjust its internal parameters to make accurate predictions.

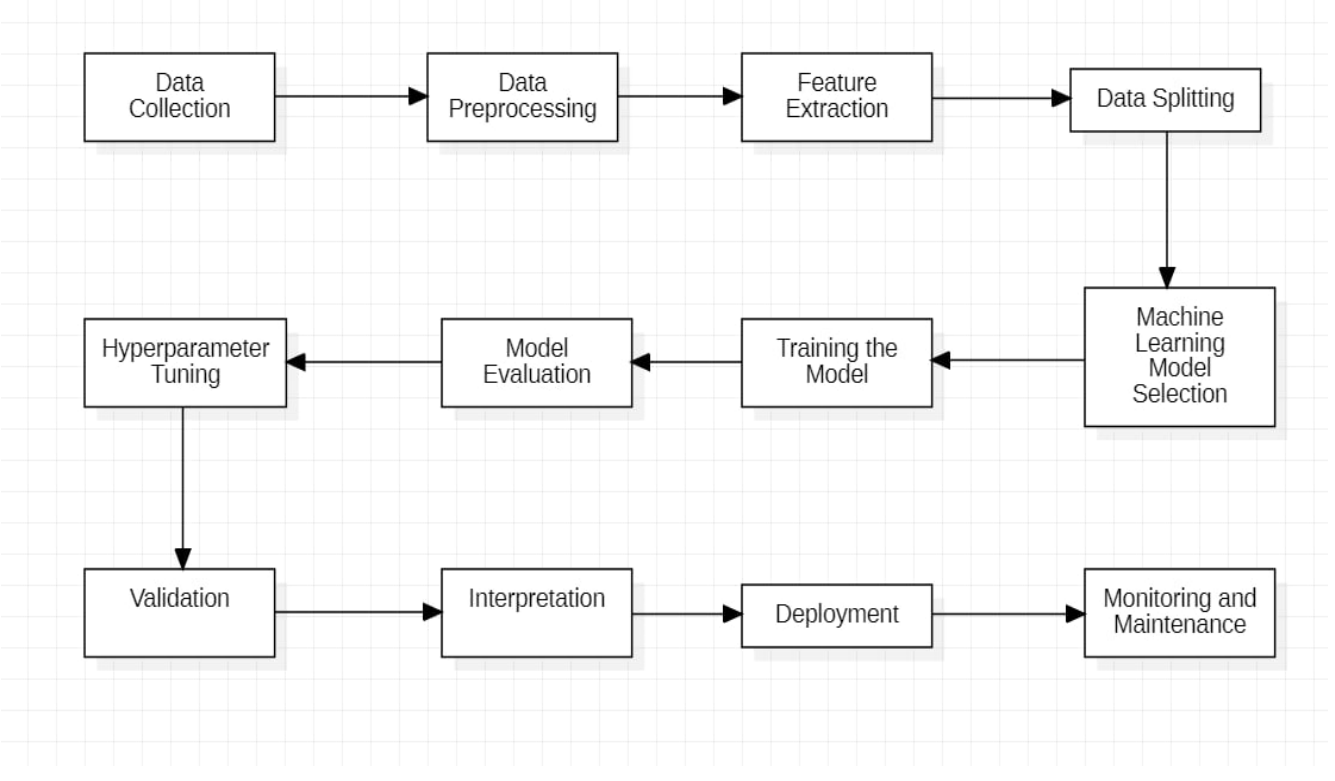
Model Evaluation:

Use the testing set to assess how well your trained model performs. You'll want to measure its accuracy, precision, recall, and other relevant metrics to see how good it is at predicting heart disease.

Hyperparameter Tuning:

Fine-tune your model by adjusting its settings (hyperparameters) to improve its performance. This might involve trying different settings for your chosen algorithm to find the best combination.

Validation:
Validate your model on an independent dataset to ensure it can generalize well to new, unseen data. This step helps confirm that your model isn't just good at predicting heart disease for the data it was trained on but can work with new patient data as well.
Interpretation:
If your model performs well, you can use it to make predictions about heart disease for new patients. It's also essential to understand which features are most influential in making predictions. This can provide insights into the risk factors for heart disease.
Deployment:
If your model is accurate and reliable, you can deploy it in a clinical setting to assist healthcare professionals in diagnosing and predicting heart disease.
Monitoring and Maintenance:
Continuously monitor the model's performance in the real world, and if necessary, update it to account for changes in the data or patient population.



5. Results
The height of the bars shows how many individuals fall into each category and the graph helps us understand the gender-based differences in the occurrence of heart disease, with males being more affected than females.

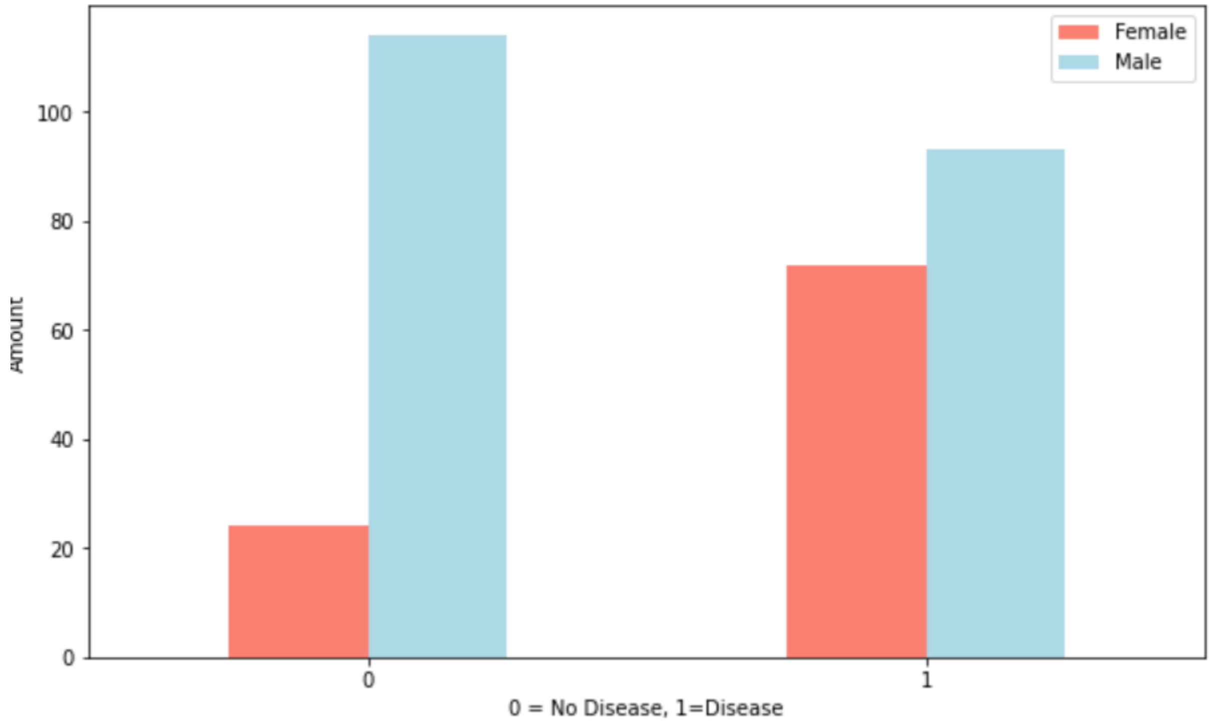


Fig. 3 Heart Disease Frequency of Sex

This figure is a scatter plot that helps us visualize the relationship between age and maximum heart rate concerning heart disease. The data points are represented as dots on the graph. The red dots indicate individuals with heart disease (target=1), and the blue dots represent individuals without heart disease (target=0).

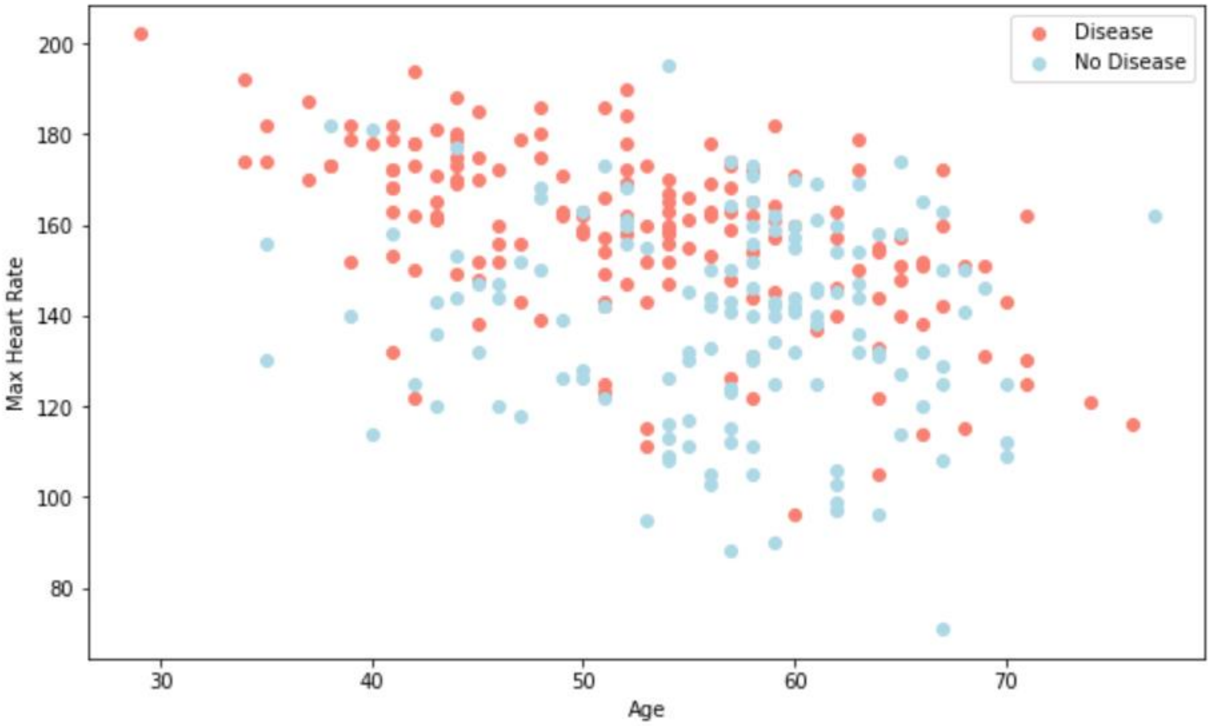


Fig. 4 Age vs Max Heart Rate

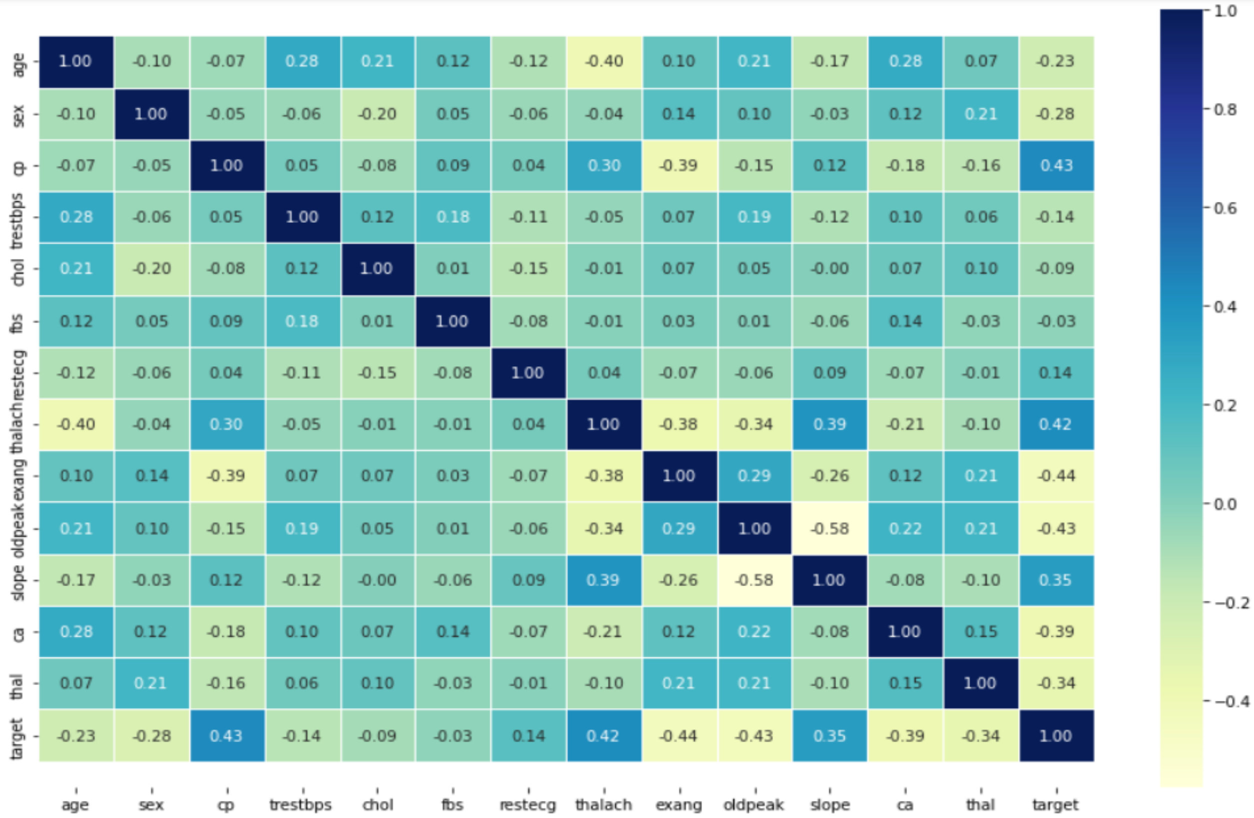
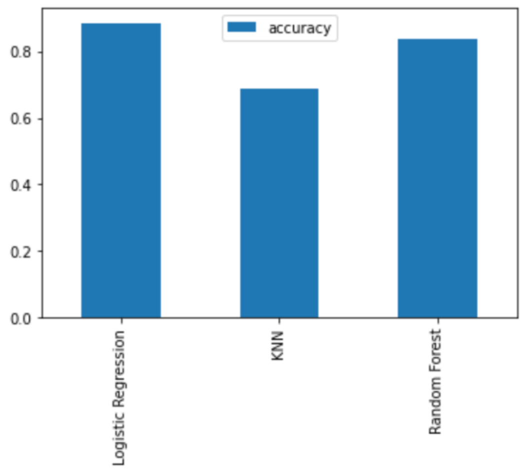


Fig. 5 Plot of Correlation matrix



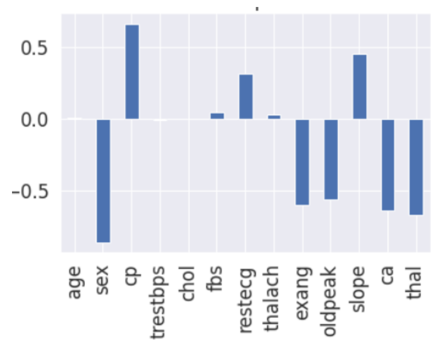


Fig. 6 Plot of Feature Importance

6. Conclusion

Predicting heart disease using machine learning is a promising approach that can help us identify individuals at risk of heart-related problems earlier and more accurately. Machine learning models analyze a wide range of patient data, suchas medical history, lifestyle factors, and test results, to make predictions. They can highlight potential risk factors that doctors might overlook, enabling more personalized and proactive healthcare. However, it's important to remember that machine learning models are not infallible. They rely on the quality and quantity of data they're trained on, and there's always the possibility of false positives or false negatives. Therefore, while they can provide valuable insights, they should be used as a complementary tool rather than a definitive diagnosis. Ultimately, the use of machine learning in heart disease predictionholds great promise for improving patient outcomes and reducinghealthcare costs, but it should be integrated into a broader healthcare approach that includes regular check-ups, lifestyle changes, and medical expertise to ensure the best possible care forindividuals at risk. As technology continues to advance and more healthcare data becomes available, machine learning-based heart disease prediction is expected to play an increasingly vital role in improving patient outcomes and reducing the burden of cardiovascular diseases.

References

1. R S Ledley and L B Lusted, "Reasoning foundations of medical diagnosis",Science, vol. 130, pp. 9-21, 2009.
2. G Kumar, "A survey on machine learning techniques in health care industry",International Journal of Recent Research Aspects, vol. 3, no. 2, pp. 128-132, 2016.
3. Tie Zheng Sun and Zehao Yu, "A study of heart disease case classification prediction based on machine learning", vol. 17, no. 26, pp. 439-448, 2021.
4. Shu tong Liang, Maozu Guo and Lingling Zhao, "Survey on medical decision support systems based on machine learning", *Computer Engineering and Applications*, vol. 55, no. 19, pp. 1-11, 2019.
5. K. Thenmozhi and P. Deepika, "Heart Disease Prediction Using Classification with Different Decision Tree Techniques", International Journal of Engineering Research and General Science, vol. 2, no. 6, 2014.
6. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", ," IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.
7. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity-based method for interactive patient risk prediction," Springer Data Mining Knowle. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.
8. Palaniappan S, Awang R, "Intelligent heart disease prediction System using data mining techniques," IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2008., vol., no., pp.108115, March 31, 2008-April 4 2008.
9. S. Leoni Sharmila, C. Dharuman and P. Venkatesan "Disease Classification Using Machine Learning Algorithms - A Comparative Study", International Journal of Pure and Applied Mathematics Volume 114 No. 6 2017, 1-10.
10. Barry Leventhal, "An introduction to data mining and other techniques for advanced analytics", *Journal of Direct Data and Digital Marketing Practice*, vol. 12, no. 2, pp. 137-153, 2010.
11. Barry Leventhal, "An introduction to data mining and other techniques for advanced analytics", Journal of Direct Data and Digital Marketing Practice, vol. 12, no. 2, 2010.

12. Guy McCann, David Drachman, Marshall Folstein, Robert Katzman, Donald Price and Emanuel M. Stadlan, "Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease", Neurology, vol. 34, no. 7, pp. 939-939, 1984.
13. Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle and Yi-Ping Phoebe Chen, "Association rule mining to detect factors which contribute to heart disease in males and females", Expert Systems with Applications, vol. 40, no. 4, pp. 1086-1093, 2013.
14. Mohammed Abdul Khaled, Sateesh Kumar Pradhan and G. N. Dash, "A survey of data mining techniques on medical data for finding locally frequent diseases", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, pp. 149-153, August 2013.
15. R. Chitra and V. Seenivasagam, "Review of heart disease prediction system using ata mining and hybrid intelligent", *ICTACT Journal on Soft Computing*, vol. 03, no. 04, July 2013.