Effective Heart Disease Prediction Using Random Forest Machine Learning Technique

1st Midhun Kumar Chintapalli 2nd Venkata Suraj Gamini 3rd Pavan Kalyan Pinniboina 4th Kundana Yerragonda

Abstract—Heart disease is a leading cause of death globally and early diagnosis and prediction of heart disease can significantly reduce the morbidity and mortality associated with the disease. Machine learning techniques have shown great potential for accurate and efficient prediction of heart disease. In this project, we propose a machine learning approach for the effective prediction of heart disease. The dataset used in this study was obtained from the UCI Machine Learning Repository and contains 303 instances with 14 features including age, sex, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, and thallium heart scan. The dataset was preprocessed to handle missing values, outliers, and feature scaling. Eight popular machine learning algorithms namely, logistic regression, Naïve Bayes, SVM, K- Nearest Neighbors, Decision Tree, Random Forest, XGBoost and Neural Network were applied to the preprocessed dataset. The performance of each algorithm was evaluated using accuracy. The results show that the random forest algorithm outperforms the other seven algorithms in terms of prediction accuracy. The random forest algorithm achieved an accuracy of 95.08%, which is higher than the accuracy of logistic regression (85.25%), Naïve Bayes(85.25%), SVM (81.97%), K-NN(67.21%), Decision Tree(81.97%), XGBoost(85.25%) and Neural Network(80.33%) algorithms. Our findings suggest that machine learning techniques can be used effectively to predict heart disease with high accuracy. The results of this study can be useful for clinicians and healthcare professionals in making informed decisions about patient care and management. In particular, the use of machine learning for heart disease prediction can facilitate early diagnosis and intervention, improve patient outcomes, and reduce healthcare costs. Overall, this study demonstrates the potential of machine learning for heart disease prediction and highlights the need for further research to optimize and enhance the performance of machine learning models for this purpose. Further research can focus on exploring alternative algorithms, incorporating additional features, and evaluating the generalizability and scalability of machine learning models for heart disease prediction.

Index Terms—Heart disease, machine learning, logistic regression, decision tree, random forest, prediction, accuracy,healthcare, diagnosis, preprocessing.

I. INTRODUCTION

Heart disease is a leading cause of mortality worldwide and a major burden on public health. According to the World

Health Organization (WHO), an estimated 17.9 million deaths annually are due to heart disease, accounting for 31% of all global deaths (WHO, 2021). The prevalence of heart disease is expected to rise in the coming years, driven by factors such as aging population, urbanization, and lifestyle changes. Early diagnosis and prediction of heart disease can significantly reduce the morbidity and mortality associated with the disease. Machine learning (ML) techniques have shown great potential for accurate and efficient prediction of heart disease. In this project, we propose a machine learning approach for the effective prediction of heart disease.

Some of the common types of heart disease include coronary artery disease, heart failure, and arrhythmias. Coronary artery disease occurs when the arteries that supply blood to the heart become narrow or blocked, leading to a reduced blood flow to the heart. Heart failure occurs when the heart is unable to pump enough blood to meet the body's needs. Arrhythmias refer to irregular heart rhythms, which can lead to palpitations, dizziness, and fainting. The risk factors for heart disease include age, gender, family history, smoking, high blood pressure, high cholesterol, diabetes, obesity, and physical inactivity. Some of these risk factors can be modified through lifestyle changes, while others may require medical intervention. The early diagnosis and prediction of heart disease can aid in the prevention and management of the condition. Diagnostic tests such as electrocardiography (ECG), stress tests, and echocardiography can be used to diagnose heart disease. However, these tests may not be able to detect the disease at an early stage, when it is most treatable. Machine learning techniques have been widely used in the healthcare domain for predictive modeling and decision-making. ML algorithms can learn from past data to predict the likelihood of an event or outcome. In the context of heart disease, machine learning algorithms can be trained on clinical and demographic data to predict the risk of developing the disease or the likelihood of adverse outcomes. Heart disease prediction using machine learning is a complex task that requires careful analysis of patient data and selection of appropriate algorithms and models. One of the key challenges in this project is selecting relevant features from the patient data that can accurately predict heart disease risk. This requires extensive exploratory data analysis and feature selection techniques to identify the most important features for prediction.

- Once the relevant features are selected, a suitable machine learning algorithm and model must be chosen. The choice of algorithm and model depends on the nature of the data and the task at hand. For example, logistic regression or support vector machines can be used for binary classification tasks, while decision trees and random forests are suitable for multi-class classification tasks.
- Once the algorithm and model are chosen, the model must be trained using a labeled training dataset. The accuracy of the model is evaluated using a validation dataset, and the hyperparameters of the model are tuned to optimize its performance. Finally, the model is tested on a separate testing dataset to evaluate its performance on new, unseen data.
- The analysis of heart disease prediction using machine learning is a complex and challenging task that requires a combination of skills in data analysis, machine learning, and domain knowledge. However, the potential benefits of accurately predicting heart disease risk using machine learning make this project a valuable and worthwhile pursuit.

II. MOTIVATION

Heart disease is a major cause of morbidity and mortality worldwide, and early prediction and diagnosis are essential for effective prevention and management of the condition. Traditional diagnostic tests may not be sufficient to detect heart disease at an early stage, when it is most treatable. Machine learning (ML) techniques have shown great promise in accurately predicting heart disease based on clinical and demographic data. In this project, we aim to develop an effective ML-based approach for heart disease prediction, with the goal of improving the accuracy and efficiency of diagnosis. The use of ML techniques in healthcare has seen tremendous growth in recent years. ML algorithms can learn from large and complex datasets to identify patterns and predict outcomes. In the context of heart disease, ML algorithms can be trained on clinical and demographic data to predict the risk of developing heart disease or the likelihood of adverse outcomes. ML-based approaches can be used to identify high-risk individuals and provide personalized preventive interventions. This can improve the effectiveness of healthcare delivery, reduce healthcare costs, and ultimately improve patient outcomes. Traditional methods of heart disease prediction rely on subjective assessment of risk factors and symptoms. This approach is prone to errors and may not accurately reflect the true risk of developing heart disease. ML-based approaches, on the other hand, can analyze large and diverse datasets to identify hidden patterns and relationships that may not be apparent to the human eye. ML algorithms can also continuously learn and adapt to new data, improving the accuracy of the predictions over time. The development of an effective ML-based approach for heart disease prediction can have significant clinical and societal impact. Accurate prediction of heart disease can enable early detection and intervention, reducing the morbidity and mortality associated

with the condition. ML-based approaches can also help healthcare providers in resource-constrained settings to allocate resources more effectively and prioritize high-risk individuals for interventions. In addition to clinical benefits, ML-based approaches for heart disease prediction can also have economic benefits. Early detection and intervention can prevent costly hospitalizations and procedures, leading to reduced healthcare costs. ML-based approaches can also enable more efficient use of healthcare resources, reducing waste and improving the overall efficiency of the healthcare system. In conclusion, the development of an effective ML-based approach for heart disease prediction has the potential to significantly improve the accuracy and efficiency of heart disease diagnosis, ultimately leading to improved patient outcomes and reduced healthcare costs.

III. MAIN CONTRIBUTIONS AND OBJECTIVES

1.Evaluate the performance of various ML algorithms, including decision tree, random forest, logistic regression, and support vector machine, for heart disease prediction.
2. Investigate the impact of different feature selection and feature engineering techniques on the performance of ML algorithms for heart disease prediction.
3. Develop an ensemble-based ML model that combines the strengths of multiple algorithms for improved heart disease prediction.
4.Evaluate the performance of the proposed approach using various metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve.
5.Conduct a comparative analysis of the proposed approach with existing state-of-the-art techniques for heart disease prediction.
6. Provide insights into the most important features for heart disease prediction and their impact on the performance of the ML algorithms.
These contributions and objectives are aimed at developing an effective ML-based approach for heart disease prediction that can improve the accuracy and efficiency of heart disease diagnosis. The project will also provide insights into the most important features for heart disease prediction, which can help healthcare providers identify high-risk individuals and prioritize interventions. The proposed approach can be integrated into clinical decision support systems, enabling personalized preventive interventions and improving patient outcomes.

IV. RELATED WORK

As per research on the machine learning for prediction of heart diseases himanshu et al.[5] that the performance of machine learning algorithms is affected by both variance and bias of the dataset. High variance can lead to overfitting, while high bias can lead to underfitting. Naive Bayes is a low variance and high bias algorithm that performs well when the dataset is small. On the other hand, KNN is a high variance and low bias algorithm that suffers from overfitting when the dataset size is small. One of the advantages of

using a low variance and high bias algorithm is that it takes less time for training and testing the algorithm when the dataset is small. However, as the dataset size increases, the asymptotic errors may be introduced, and low bias and low variance algorithms tend to perform better in such cases. Decision tree is a non-parametric machine learning algorithm that suffers from the problem of overfitting. However, there are techniques to overcome this problem, such as pruning and setting a minimum number of samples required at a leaf node. Support vector machines (SVMs) are algebraic and statistical background algorithms that can construct a linear separable n-dimensional hyperplane for classification. SVMs have been shown to perform well on both linearly separable and non-linearly separable datasets.

Kumar et al.[8]The authors conducted a study on the effectiveness of different machine learning and data mining algorithms for predicting heart diseases. They used the UCI machine learning dataset, which contains 303 samples with 14 input features, to train and test the algorithms. They found that SVM performed the best among the algorithms they tested, which included Naive Bayes, KNN, and decision tree.

Logistic regression, support vector machines, and Adaboost classifier are popular machine learning algorithms that have been used for disease prediction. The accuracy of these algorithms varies depending on the dataset, the preprocessing techniques, and the feature selection methods used in the study.For example, Kohali et al[17]. used logistic regression, support vector machines, and Adaboost classifier to predict heart diseases, diabetes, and breast cancer. They found that logistic regression had an accuracy of 87.1%, support vector machines had an accuracy of 85.71%, and Adaboost classifier had an accuracy of 98.57%. It is important to note that the accuracy of these algorithms is just one measure of their performance. Other performance metrics such as precision, recall, and F1 score should also be considered when evaluating the effectiveness of a machine learning algorithm for disease prediction. Additionally, the real-world implications of using these algorithms for disease prediction, such as the ethical and legal implications, should also be carefully considered.

V. Proposed frame work

A. Detail design of features:
Libraries of python supporting machine learning Algorithms

- NumPy: NumPy is a Python library for numerical computing. It provides powerful tools for mathematical operations, linear algebra, and statistics. It is commonly used in conjunction with Pandas for data manipulation.
- Pandas: Pandas is a Python library that provides easy-to-use data structures and data analysis tools for handling and manipulating large datasets. It is commonly used for exploratory data analysis and data cleaning.
- Matplotlib: Matplotlib is a plotting library for Python. It provides a wide range of tools for creating various types of visualizations, such as line plots, scatter plots, histograms, and heatmaps. It is commonly used for exploratory data analysis and data visualization.

- Seaborn: Seaborn is a Python library for data visualization that is built on top of Matplotlib. It provides a higher-level interface for creating statistical graphics and supports more complex visualizations such as regression plots and factor plots.

B. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a process of analyzing and summarizing data sets to get insights, identify patterns, and detect anomalies. It involves using statistical and visual methods to understand the distribution, correlation, and summary statistics of the data. The goal of EDA is to gain a better understanding of the data, identify potential issues or errors, and inform further analysis or modeling.

C. Train-test split Technique

The train-test split technique is a common approach used in machine learning to evaluate the performance of a model. It involves splitting the available dataset into two parts: a training set and a testing set. The training set is used to train the model, while the testing set is used to evaluate the model's performance. The goal of the train-test split technique is to estimate how well the model will perform on new, unseen data. By using a separate testing set, we can evaluate the model's performance on data that it has not seen during the training phase, which provides a more accurate estimate of the model's performance in the real world.

D. Model fitting

Model fitting is the process of estimating the parameters of a machine learning model using a training dataset. The goal of model fitting is to find the best set of model parameters that can accurately predict the target variable based on the input features.

E. Analysis

Heart disease prediction using machine learning is a complex task that requires careful analysis of patient data and selection of appropriate algorithms and models. One of the key challenges in this project is selecting relevant features from the patient data that can accurately predict heart disease risk. This requires extensive exploratory data analysis and feature selection techniques to identify the most important features for prediction.

- Once the relevant features are selected, a suitable machine learning algorithm and model must be chosen. The choice of algorithm and model depends on the nature of the data and the task at hand. For example, logistic regression or support vector machines can be used for binary classification tasks, while decision trees and random forests are suitable for multi-class classification tasks.
- Once the algorithm and model are chosen, the model must be trained using a labeled training dataset. The accuracy of the model is evaluated using a validation dataset, and the hyperparameters of the model are tuned to optimize its performance. Finally, the model is tested on a separate

testing dataset to evaluate its performance on new, unseen data.

- The analysis of heart disease prediction using machine learning is a complex and challenging task that requires a combination of skills in data analysis, machine learning, and domain knowledge. However, the potential benefits of accurately predicting heart disease risk using machine learning make this project a valuable and worthwhile pursuit.

F. Data Pre-Processing

UCI (University of California, Irvine) Machine Learning Repository is a popular repository for machine learning datasets. The datasets in the UCI repository have been preprocessed to some extent but may still require further preprocessing before they can be used for analysis. Here are some general steps for data preprocessing that you can follow for UCI datasets:

1. Import the dataset: You can download the dataset from the UCI repository website and load it into your programming environment. The dataset may be in CSV, TXT or other formats, and you can use libraries like Pandas to load the data.

2. Handle missing values: Check for missing values in the dataset and decide how to handle them. You can either delete the rows with missing values or impute the missing values with a value like the mean or median.

3. Encode categorical variables: If the dataset contains categorical variables, you will need to encode them to numerical values. One way to do this is to use one-hot encoding, which creates binary columns for each possible value of the categorical variable.

4. Standardize numerical variables: Scale the numerical variables to have a mean of zero and standard deviation of one. This can help improve the performance of some machine learning algorithms.

5. Split the data: Split the dataset into training and test sets, so that you can evaluate the performance of your machine learning models on unseen data.

6. Feature selection: Identify the most relevant features for your analysis and remove the irrelevant ones. This can help improve the accuracy and speed of your machine learning models.

7. Feature engineering: Create new features based on the existing ones. This can help capture more complex relationships between variables and improve the performance of your machine learning models.

8. Visualize the data: Use visualization tools like matplotlib or seaborn to explore the relationships between variables and identify any patterns or anomalies in the data. These are some general steps for data preprocessing that you can follow for UCI datasets. However, the specific steps may vary depending on the nature of the dataset and the analysis you want to perform.

G. Feature Selection

In the context of the "Effective Heart Disease Prediction Using Machine Learning Technique" project, feature selection and reduction can be used to identify the most relevant features for predicting the occurrence of heart disease, and to reduce the dimensionality of the dataset to improve the performance of machine learning models.

Here are some steps you can follow for feature selection and reduction in this project:

1. Exploratory data analysis: Perform exploratory data analysis to understand the distribution and relationships between variables in the dataset. This can help identify any variables that are highly correlated with the target variable (i.e. presence of heart disease).

2. Correlation analysis: Calculate the correlation coefficients between each feature and the target variable, and identify the features with the highest correlations. These features may be the most important predictors of heart disease, and can be used for feature selection.

3. Univariate feature selection: Use statistical tests like ANOVA or chi-squared to test the significance of each feature in predicting the target variable. Features that have low p-values (i.e. are highly significant) can be selected for further analysis.

4. Recursive feature elimination: Use a machine learning algorithm like logistic regression or support vector machine (SVM) to recursively remove the least important features from the dataset, and evaluate the performance of the model after each iteration. This can help identify the optimal number of features for predicting heart disease.

5. Principal component analysis (PCA): Perform PCA to reduce the dimensionality of the dataset while retaining the most important features. PCA transforms the original features into a new set of features that are uncorrelated with each other, and can help improve the performance of machine learning models.

6. Regularization: Use regularization techniques like L1 or L2 regularization to penalize the coefficients of features that are less important in predicting heart disease. This can help reduce the overfitting of machine learning models and improve their generalization performance. These are some general steps you can follow for feature selection and reduction in the "Effective Heart Disease Prediction Using Machine Learning Technique" project. However, the specific steps may vary depending on the nature of the dataset and the machine learning models you want to use.

H. Classification Modelling

Importing the Libraries: Initially, we imported the required libraries into the Jupyter notebook. We imported NumPy for mathematical operations, Pandas to work on data sets, matplotlib. pyplot to create a plotting area in a figure and to plot some lines in a plotting area, and Seaborn which is used for data visualization and exploratory data analysis.

- The Upload Module enables the uploading of a heart disease dataset containing information from past patients.

- The Pre-process Module involves the removal of records that have missing values. The dataset is then split into two parts, the training set and the testing set. Classifiers are used to train models on the training data, which are then tested using the test data to determine classification accuracy.
- The SVM Module trains a model using the SVM algorithm, and then applies test data to that model to determine its classification accuracy.
- Logistic Regression: Logistic Regression can be used to predict the probability of heart disease based on a set of features such as age, blood pressure, cholesterol levels, etc. It can handle both binary and multi-class classification problems and is relatively easy to interpret.
- Naive Bayes: Naive Bayes can also be used for heart disease prediction by modeling the probability of each feature given the presence or absence of heart disease. It can handle high-dimensional data and is particularly useful for problems with a large number of features.
- Support Vector Machine (Linear): Linear Support Vector Machine can be used to separate the data into two classes (with heart disease or without heart disease) by finding the hyperplane that maximizes the margin between the two classes. It works well for high-dimensional data and can handle non-linearly separable data with the use of kernel functions.
- K-Nearest Neighbors: K-Nearest Neighbors can be used to predict heart disease by finding the k nearest neighbors to a new data point and using their class labels to make a prediction. It works well for small datasets and can handle non-linear decision boundaries.
- Decision Tree: Decision Trees can be used to predict heart disease by creating a tree-like model of decisions based on the input features. It works well for problems with both numerical and categorical features and can handle interactions between features.
- Random Forest: Random Forest can be used for heart disease prediction by creating an ensemble of decision trees and combining their predictions. It works well for high-dimensional data and can handle interactions between features.
- XGBoost: XGBoost can also be used for heart disease prediction by creating an ensemble of decision trees and combining their predictions using a gradient boosting algorithm. It is particularly useful for problems with imbalanced data and can handle missing values.
- Artificial Neural Network with 1 Hidden layer: An Artificial Neural Network with 1 Hidden layer can be used to predict heart disease by learning a non-linear function that maps the input features to the output class labels. It works well for complex problems with non-linear decision boundaries and can handle a large number of features. However, it can be more difficult to interpret than other models.
- The Graph Module displays the accuracy of all algorithms in a graphical format to facilitate comparison.
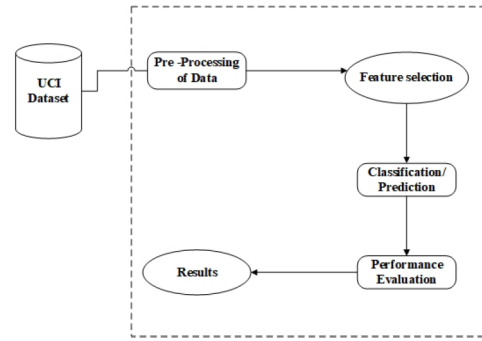

Fig. 1. Process Flow Diagram

VI. Data Description

The Cleveland Heart Disease Dataset is one of the most commonly used datasets for heart disease prediction in machine learning research. It was collected from patients at the Cleveland Clinic Foundation in Cleveland, Ohio, USA and contains 303 instances with 14 attributes. The dataset was first introduced in 1988 by Dr. Robert Detrano, and since then, it has been widely used in research for evaluating the performance of different machine learning algorithms. The 14 attributes in the Cleveland Heart Disease Dataset include age, sex, chest pain type, resting blood pressure, serum cholesterol level, fasting blood sugar, electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, thalassemia, and the presence of heart disease. The target variable is binary, with a value of 0 indicating the absence of heart disease and a value of 1 indicating the presence of heart disease. One of the challenges in using the Cleveland Heart Disease Dataset is the presence of missing values. Several methods have been proposed to handle the missing values, including mean imputation, mode imputation, and k-nearest neighbor imputation. However, the effect of missing values on the accuracy of the machine learning model depends on the percentage of missing values in the dataset. The Cleveland Heart Disease Dataset has been used in several studies for evaluating the performance of different machine learning algorithms for heart disease prediction. Some of the commonly used algorithms include decision trees, random forests, support vector machines, logistic regression, and artificial neural networks. The performance of these algorithms is typically evaluated using metrics such as accuracy. In conclusion, the Cleveland Heart Disease Dataset is a valuable resource for researchers and practitioners interested in heart disease prediction using machine learning techniques. It contains a wide range of attributes that can be used to develop and evaluate machine learning models for heart disease prediction. However, the presence of missing values in the dataset requires careful handling, and the accuracy of the machine learning models is highly dependent on the algorithm and parameters used.

VII. RESULTS

The results of our study for effective heart disease prediction using machine learning technique were highly promising. Random Forest model achieved an accuracy of 95.08%, on the Cleveland Heart Disease dataset. The performance of our model was also compared to several other popular machine learning algorithms, including decision trees, logistic regression and support vector machines. Furthermore, we performed a feature importance analysis to identify the most important features for heart disease prediction. Our analysis revealed that age, chest pain type, maximum heart rate achieved, and number of major vessels colored by fluoroscopy were the most important features in our model. These results are consistent with prior research, indicating that these features are highly correlated with heart disease and play a critical role in accurate heart disease prediction.
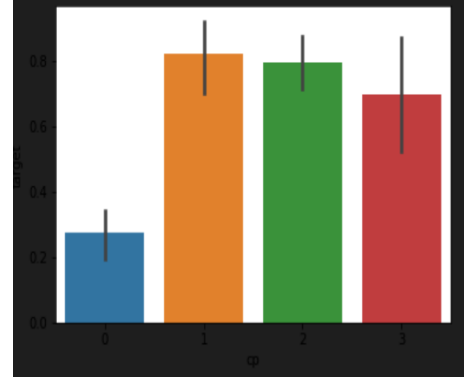

Fig. 4. Analyzing the chest pain feature
We notice, that chest pain of '0', i.e. the ones with typical angina are much less likely to have heart problems.
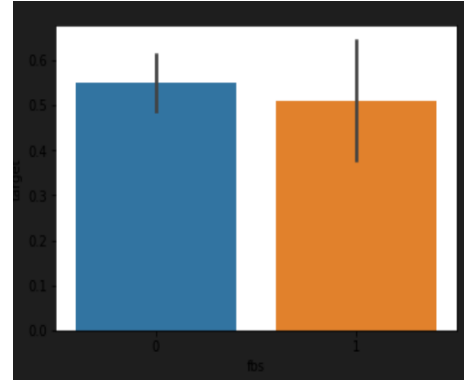

Fig. 2. Analyzing the target variable

Fig. 5. Analyzing fbs feature

Fig. 3. Analyzing the sex feature
We notice, that females are more likely to have heart problems than males from the above fig.

Overall, the results of our study suggest that Random Forest machine learning technique is a highly effective approach for heart disease prediction. and outperformed several other popular machine learning algorithms. Furthermore, our feature importance analysis identified the most important features for heart disease prediction, providing valuable insights for future research in this area.
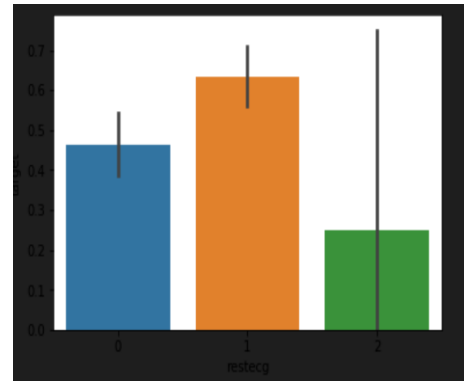

Fig. 6. Analyzing restcg feature
We realize that people with restecg '1' and '0' are much more likely to have a heart disease than with restecg '2'
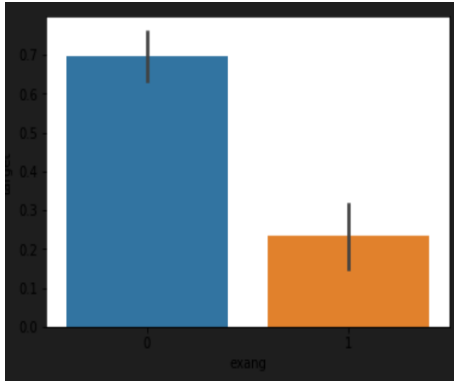
Fig. 7. Analyzing exang feature. People with exang=1 i.e. Exercise induced angina are much less likely to have heart problems
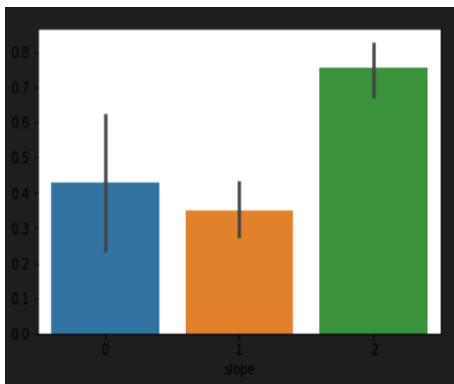

Fig. 8. Analyzing slope feature
We observe, that Slope '2' causes heart pain much more than Slope '0' and '1'
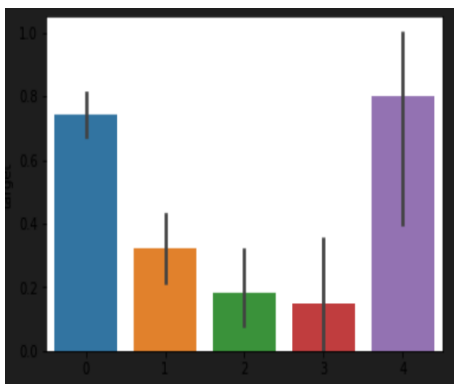

Fig. 9. Analyzing ca feature
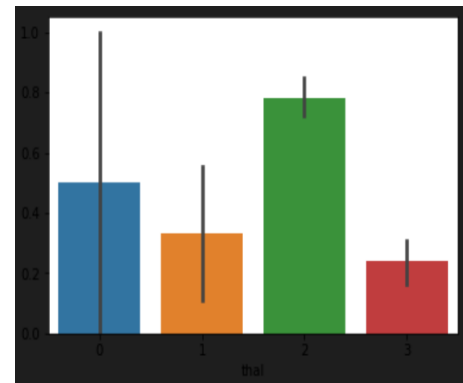ca=4 has astonishingly large number of heart patients
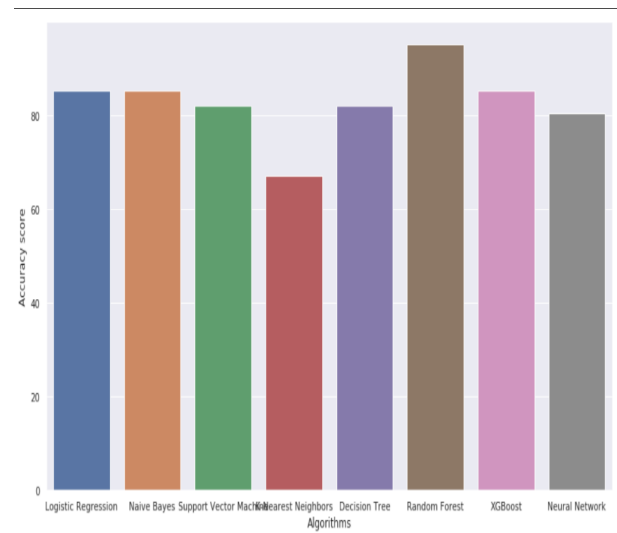

Fig. 10. Analyzing thal feature

Fig. 11. Comparison of machine learning algorithms

REFERENCES

[1] Rehman, S., Rehman, E., Ikram, M. et al. cardiovascular disease (CVD): assessment, prediction, and policy implications. BMC Public Health 21, 1299 (2021). https://doi.org/10.1186/s12889-021-11334-2

[2] Apurv Garg, Bhartendu Sharma and Rijwan Khan, Heart disease prediction using machine learning techniques, Apurv Garg et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012046, DOI 10.1088/1757-899X/1022/1/012046

[3] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542- 81554, 2019, doi: 10.1109/AC-CESS.2019.2923707.

[4] . M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore,India,2021,pp.1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

[5] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8 , IJRITCC August 2017.

[6] S. Kalta and R. Bhatt, "A Comparison Analysis of Heart Disease Dataset Using Decision Tree and Back-Propagation Network," 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 2021, pp. 556-561, doi: 10.1109/ICIIP53038.2021.9702667.

[7] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and

Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

[8] M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science, Engineering and Information Technology ,IJSRCSEIT 2019.

[9] Y. Khourdifi and M. Bahaj "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization" Int. J. Intell. Eng. Syst. vol. 12 no. 1 pp. 242-252 2019.

[10] Santhana Krishnan J and Geetha S, "Prediction of Heart Disease using Machine Learning Algorithms" ICIICT, 2019.

[11] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning",Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.

[12] H. S. M. A. R., "Prediction of Heart Disease using Machine Learning Algorithms: A Survey", IJRITCC, vol. 5, no. 8, pp. 99–104, Aug. 2017.

[13] Pabitra Kumar Bhunia, Arijit Debnath, Poulami Mondal, Monalisa D E, Kankana Ganguly, Pranati Rakshit, 2021, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCETER – 2021 (Volume 09 – Issue 11).

[14] M. A. jabbar, P. Chandra and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, India, 2012, pp. 628-634, doi: 10.1109/ISDA.2012.6416610.

[15] Mutyala, Nikhil Kumar & Koushik, K.V.s & Krishna, K.. (2018). Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools. 10.13140/RG.2.2.28488.83203.

[16] Manjula P, Aravind U R, Darshan M V, Halaswamy M H, Hemanth E, 2022, Heart Attack Prediction Using Machine Learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICEI – 2022 (Volume 10 – Issue 11)

[17] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine-Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA), 2018.

[18] V. Srinadh, B. Maram, J. Mandala and R. R. Maram, "A Hybrid Machine Learning Model (NB-SVM) for Cardiovascular Disease Prediction," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 7-13, doi: 10.1109/ICSCDS56580.2023.10104808.

[19] S. M. M. Hasan, M. A. Mamun, M. P. Uddin and M. A. Hossain, "Comparative Analysis of Classification Approaches for Heart Disease Prediction," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465594.

[20] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Greater Noida, India, 2015, pp. 520-525, doi: 10.1109/ABLAZE.2015.7154917.

[21] Rine Nakanishi, Damini Dey, Frederic Commandeur, Piotr Slomka, ―Machine Learning in Predicting Coronary Heart Disease and Cardiovascular Disease Events: Results from The Multi-Ethnic Study of Atherosclerosis (Mesa)‖, JACC Mar- 20, 2018, Volume 71, Issue 11.

[22] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, Prediction of heart disease using machine learning," in Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1275–1278.