Project Title

Heart Disease Prediction using deep learning and methods.

Team Member

Venkata sai Boinepally -700743700 CRN: 23259

V V S Murthy kola -700729142 CRN:23216

Nirmala Yarlagadda – 700733102 CRN:23216

INJA AKHILESWAR REDDY-700744180 CRN:23216

Githublink: https://github.com/murthykolla/Project.git

Abstract:

The aim of this project is to develop a deep-learning model for predicting heart disease using various clinical and demographic features. The proposed model will use a large dataset of patient records to learn complex relationships between features and predict the occurrence of heart disease accurately. We will begin by preprocessing the data, cleaning it, and removing any missing values or outliers. Next, we will perform feature engineering to identify important features that can improve the model's performance. Feature selection techniques such as recursive feature elimination and principal component analysis will be employed. We will then use several deep learning methods such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs) to predict heart disease occurrence. These methods have been shown to perform well in a variety of medical applications and can learn complex patterns in the data. To optimize our model's performance, we will use various techniques such as hyperparameter tuning, cross-validation, and ensembling. Transfer learning techniques will also transfer knowledge from pre-trained models to our heart disease prediction model. We will evaluate the performance of our model using several evaluation metrics such as accuracy, precision, recall, and F1 score. The proposed model's interpretability will be improved using techniques such as local interpretable model-agnostic explanations (LIME) and Shapley additive explanations (SHAP). In conclusion, this project aims to develop an accurate and interpretable deep learning model for predicting heart disease occurrence. The proposed model can assist healthcare professionals in making accurate diagnoses and treatment decisions, leading to better patient health outcomes. Moreover, the development of this model can contribute to the advancement of deep learning-based medical 1 decision support systems.]Heart disease is a leading cause of death worldwide and is a significant public health challenge. Early diagnosis of heart disease is essential for effective treatment and prevention. Machine learning approaches have been used to identify patterns in medical data that can aid in the prediction of heart disease. In recent years, deep learning methods have shown great potential in this regard. The aim of this project is to develop a deep-learning model for predicting heart disease using various clinical and demographic features. The proposed model will use a large dataset of patient records to learn complex relationships between features and predict the occurrence of heart disease accurately. We will begin by preprocessing the data, cleaning it, and removing any missing values or outliers. Next, we will perform feature engineering to identify important features that can improve the model's performance. Feature selection techniques such as recursive feature elimination and principal component analysis will be employed. We will then use several deep learning methods such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs) to predict heart disease occurrence. These methods have been shown to perform well in a variety of medical applications and can learn complex patterns in the data. To optimize our model's performance, we will use various techniques such as hyperparameter tuning, cross-validation, and ensembling. Transfer learning techniques will also transfer knowledge from pre-trained models to our heart disease prediction model. We will evaluate the performance of our model using several evaluation metrics such as accuracy, precision, recall, and F1 score. The proposed model's interpretability will be improved using techniques such as local interpretable model-

agnostic explanations (LIME) and Shapley additive explanations (SHAP). In conclusion, this project aims to develop an accurate and interpretable deep learning model for predicting heart disease occurrence. The proposed model can assist healthcare professionals in making accurate diagnoses and treatment decisions, leading to better patient health outcomes. Moreover, the development of this model can contribute to the advancement of deep learning-based medical decision support systems.

Keywords:

heart disease, deep learning, neural network, prediction, clinical data, feature engineering, healthcare, accuracy,knn,rf,dt,lr

Introduction:

The aim of this project is to develop a deep-learning model for predicting heart disease using various clinical and demographic features. The proposed model will use a large dataset of patient records to learn complex relationships between features and accurately predict the occurrence of heart disease. 2 Deep learning has been gaining popularity in medical applications due to its ability to learn from vast amounts of data and identify patterns that are not easily discernible by humans. Unlike traditional machine learning techniques, deep learning models can automatically extract relevant features from raw data, reducing the need for manual feature engineering. In medical imaging analysis, deep learning has shown impressive performance in detecting and classifying abnormalities in medical images, such as tumors in MRI scans or diabetic retinopathy in retinal images. In disease diagnosis, deep learning models can analyze large datasets of patient records and predict the likelihood of developing a disease, such as heart disease or cancer, based on various risk factors. Moreover, deep learning models have also been used in drug discovery to identify potential drug candidates and accelerate the drug development process. With the ever-increasing availability of medical data, deep learning methods are expected to play a crucial role in improving medical diagnosis, treatment, and patient outcomes. In the context of heart disease prediction, deep learning methods can be used to identify complex relationships between various clinical and demographic features and predict the occurrence of heart disease accurately. These features can include age, sex, blood pressure, cholesterol levels, and smoking status, among others. By accurately predicting the occurrence of heart disease, healthcare professionals can make informed decisions regarding treatment and prevention strategies. The proposed model for heart disease prediction will use a large dataset of patient records to learn complex patterns in the data. Preprocessing the dataset is important to ensure that the data is clean, accurate, and complete. Missing values and outliers can negatively impact the accuracy of the model, so they will be removed during preprocessing. Feature engineering is the process of selecting and transforming relevant features from the dataset. This process can help to improve the model’s accuracy by identifying important features that contribute to the prediction of heart disease. Feature selection techniques such as recursive feature elimination and principal component analysis are commonly used to identify the most important features. Deep learning methods have shown great potential in medical applications due to their ability to learn complex patterns in data. These methods involve training artificial neural networks with multiple layers to learn patterns in the data. This can be particularly useful in medical applications where the data is often complex and multifaceted. The use of deep learning methods has shown promising results in the diagnosis of various diseases, analysis of medical images, and drug discovery. The proposed model for heart disease prediction aims to leverage the power of deep learning to accurately predict the occurrence of heart disease based on patient records. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs) are powerful deep learning techniques that have been successfully applied to various medical applications. CNNs are commonly used in medical imaging analysis to detect abnormalities in medical im3 ages such as X-rays, MRI, and CT scans. They can extract features automatically and identify patterns that may be difficult for humans to detect. RNNs, on the other hand, are suitable for sequential data such as time-series data, and are useful in medical applications such as predicting patient outcomes based on vital signs or predicting the progression of diseases. DBNs are probabilistic graphical models that can be used to model complex dependencies in data and have been applied in medical applications such as predicting adverse drug reactions and drug discovery. In this project, we will explore the use of these deep learning methods to predict heart disease occurrence by learning complex patterns in clinical and demographic features. By leveraging the strengths of these deep learning methods, we hope to develop a robust and accurate prediction model that can assist healthcare professionals in making timely and accurate diagnoses and treatment decisions. To optimize our model's performance, we will use various techniques such as hyperparameter tuning, cross-validation, and ensembling. Transfer learning techniques will also transfer knowledge from pre-trained models to our heart disease prediction model. We will evaluate the performance of our model using several evaluation metrics such as accuracy, precision, recall, and F1 score. The proposed model's interpretability will be improved using techniques such as local interpretable model-agnostic explanations (LIME) and Shapley additive explanations (SHAP). The development of an accurate and interpretable deep learning model for predicting heart disease occurrence can have significant implications for healthcare. The proposed model can assist healthcare professionals in making accurate diagnoses and treatment decisions, leading to better patient health outcomes. Moreover, the development of this model can contribute to the advancement of deep learning-based medical decision support systems. In conclusion, this project aims to develop a deep learning model for predicting heart disease occurrence using various clinical and demographic features. The proposed model will use a large dataset of patient records to learn complex patterns in the data and predict heart disease occurrence accurately. The development of an accurate and interpretable deep learning model for predicting heart disease can have significant implications for healthcare and contribute to the advancement of deep learning-based medical decision support systems.]Heart disease, also known as cardiovascular disease, refers to a range of conditions that affect the heart and blood vessels. It is one of the leading causes of death worldwide, accounting for over 17 million deaths annually. Early diagnosis of heart disease is essential for effective treatment and prevention. Machine learning approaches have been used to identify patterns in medical data that can aid in the prediction of heart disease. In recent years, deep learning methods have shown great potential in this regard. The aim of this project is to develop a deep-learning model for predicting heart disease using various clinical and demographic features. The proposed model will use a large dataset of patient records to learn complex relationships between features and accurately predict the occurrence of heart disease. Deep learning has been gaining popularity in medical applications due to its ability to learn from vast amounts of data and identify patterns that are not easily discernible by humans. Unlike traditional machine learning techniques, deep learning models can automatically extract relevant features from raw data, reducing the need for manual feature engineering. In medical imaging analysis, deep learning has shown impressive performance in detecting and classifying abnormalities in medical images, such as tumors in MRI scans or diabetic retinopathy in retinal images. In disease diagnosis, deep learning models can analyze large datasets of patient records and predict the likelihood of developing a disease, such as heart disease or cancer, based on various risk factors. Moreover, deep learning models have also been used in drug discovery to identify potential drug candidates and accelerate the drug development process. With the ever-increasing availability of medical data, deep learning methods are expected to play a crucial role in improving medical diagnosis, treatment, and patient outcomes. In the context of heart disease prediction, deep learning methods can be used to identify complex relationships between various clinical and demographic features and predict the occurrence of heart disease accurately. These features can include age, sex, blood pressure, cholesterol levels, and smoking status, among others. By accurately predicting the occurrence of heart disease, healthcare professionals can make informed decisions regarding treatment and prevention strategies. The proposed model for heart disease prediction will use a large dataset of patient records to learn complex patterns in the data. Preprocessing the dataset is important to ensure that the data is clean, accurate, and complete. Missing values and outliers can negatively impact the accuracy of the model, so they will be removed during preprocessing. Feature engineering is the process of selecting and transforming relevant features from the dataset. This process can help to improve the model's accuracy by identifying important features that contribute to the prediction of heart disease. Feature selection techniques such as recursive feature elimination and principal component analysis are commonly used to identify the most important features. Deep learning methods have shown great potential in medical applications due to their ability to learn complex patterns in data. These methods involve training artificial neural networks with multiple layers to learn patterns in the data. This can be particularly useful in medical applications where the data is often complex and multifaceted. The use of deep learning methods has shown promising results in the diagnosis of various diseases, analysis of medical images, and drug discovery. The proposed model for heart disease prediction aims to leverage the power of deep learning to accurately predict the occurrence of heart disease based on patient records. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs) are powerful deep learning techniques that have been successfully applied to various medical applications. CNNs are commonly used in medical imaging analysis to detect abnormalities in medical im5 ages such as X-rays, MRI, and CT scans. They can extract features automatically and identify patterns that may be difficult for humans to detect. RNNs, on the other hand, are suitable for sequential data such as time-series data, and are useful in medical applications such as predicting patient outcomes based on vital signs or predicting the progression of diseases. DBNs are probabilistic graphical models that can be used to model complex dependencies in data and have been applied in medical applications such as predicting adverse drug reactions and drug discovery. In this project, we will explore the use of these deep learning methods to predict heart disease occurrence by learning complex patterns in clinical and demographic features. By leveraging the strengths of these deep learning methods, we hope to develop a robust and accurate prediction model that can assist healthcare professionals in making timely and accurate diagnoses and treatment decisions. To optimize our model's performance, we will use various techniques such as hyperparameter tuning, cross-validation, and ensembling. Transfer learning techniques will also transfer knowledge from pre-trained models to our heart disease prediction model. We will evaluate the performance of our model using several evaluation metrics such as accuracy, precision, recall, and F1 score. The proposed model's interpretability will be improved using techniques such as local interpretable model-agnostic explanations (LIME) and Shapley additive explanations (SHAP). The development of an accurate and interpretable deep learning model for predicting heart disease occurrence can have significant implications for healthcare. The proposed model can assist healthcare professionals in making accurate diagnoses and treatment decisions, leading to better patient health outcomes. Moreover, the development of this model can contribute to the advancement of deep learning-based medical decision support systems. In conclusion, this project aims to develop a deep learning model for predicting heart disease occurrence using various clinical and demographic features. The proposed model will use a large dataset of patient records to learn complex patterns in the data and predict heart disease occurrence accurately. The development of an accurate and interpretable deep learning model for predicting heart disease can have significant implications for healthcare and contribute to the advancement of deep learning-based medical decision support systems.

Motivation:

The motivation behind this project is to develop a deep learning model for predicting heart disease occurrence that can assist healthcare professionals in making accurate diagnoses and treatment decisions. The proposed model will use various clinical and demographic features to predict the likelihood of heart disease accurately. By using a large dataset of patient records, the model can learn complex relationships between features and predict the occurrence of heart disease with high accuracy. 6 The use of deep learning methods in medical decision-making can signifcantly improve patient outcomes by providing accurate diagnoses and personalized treatment plans. However, the development of accurate and reliable deeplearning models for medical applications is still a challenging task. This project aims to contribute to the advancement of deep learning-based medical decision support systems by developing an accurate and interpretable model for predicting heart disease. Moreover, the proposed model can assist healthcare professionals in making informed decisions regarding patients with heart disease, leading to improved health outcomes and reduced healthcare costs. It can also aid in identifying high-risk patients who require more aggressive treatment or intervention. In conclusion, the motivation behind this project is to develop an accurate and interpretable deep learning model for predicting heart disease occurrence that can contribute to the advancement of deep learning-based medical decision support systems. The proposed model can aid in accurate diagnoses, personalized treatment plans, and improved health outcomes for patients with heart disease.]Heart disease is a significant public health challenge that affects millions of people worldwide. It is a leading cause of death and disability, and early diagnosis is crucial for effective treatment and prevention. Machine learning techniques have been widely used in the healthcare industry to predict the occurrence of heart disease and aid in its management. Deep learning methods, in particular, have shown great promise in this regard due to their ability to learn complex patterns in large datasets. The motivation behind this project is to develop a deep learning model for predicting heart disease occurrence that can assist healthcare professionals in making accurate diagnoses and treatment decisions. The proposed model will use various clinical and demographic features to predict the likelihood of heart disease accurately. By using a large dataset of patient records, the model can learn complex relationships between features and predict the occurrence of heart disease with high accuracy. The use of deep learning methods in medical decision-making can significantly improve patient outcomes by providing accurate diagnoses and personalized treatment plans. However, the development of accurate and reliable deeplearning models for medical applications is still a challenging task. This project aims to contribute to the advancement of deep learning-based medical decision support systems by developing an accurate and interpretable model for predicting heart disease. Moreover, the proposed model can assist healthcare professionals in making informed decisions regarding patients with heart disease, leading to improved health outcomes and reduced healthcare costs. It can also aid in identifying high-risk patients who require more aggressive treatment or intervention. In conclusion, the motivation behind this project is to develop an accurate and interpretable deep learning model for predicting heart disease occurrence that can contribute to the advancement of deep learning-based medical decision support systems. The proposed model can aid in accurate diagnoses, personalized treatment plans, and improved health outcomes for patients with heart disease.

Related work:

Heart disease is a major cause of death worldwide, and early prediction of the disease can help in its prevention and treatment. Machine learning methods, including deep learning, have shown promise in accurately predicting the risk of heart disease.

One of the early studies in this field was conducted by Attia et al. (2019), who used a convolutional neural network (CNN) to predict the risk of cardiovascular disease in patients using electrocardiogram (ECG) data. The CNN was trained on a large dataset of over 500,000 ECG recordings and achieved an accuracy of 0.85 in predicting the risk of cardiovascular disease.

Another study by Tan et al. (2019) proposed a deep learning framework for predicting cardiovascular disease risk using electronic health record (EHR) data. The framework consisted of a sequence-to-sequence model that incorporated both structured and unstructured EHR data. The model was trained on a large dataset of over 400,000 patients and achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.86 in predicting the risk of cardiovascular disease.

Mehra bi et al. (2020) compared the performance of several machine learning methods, including deep learning, for predicting hypertension in patients using routine clinical data. The study found that deep learning methods, including long short-term memory (LSTM) and CNN, outperformed other machine learning methods in predicting hypertension.

In a recent systematic review by Rangarajan et al. (2021), the authors evaluated the performance of machine learning algorithms in predicting cardiovascular disease using various types of data, including ECG, EHR, and medical imaging data. The study found that deep learning methods, including CNN and LSTM, achieved high accuracy in predicting cardiovascular disease.

Another study by Wang et al. (2021) proposed a deep learning framework for predicting the risk of heart failure using EHR data. The framework consisted of a dual-attention network that incorporated both temporal and clinical features from EHR data. The model was trained on a large dataset of over 13,000 patients and achieved an AUC-ROC of 0.81 in predicting the risk of heart failure.

In addition to deep learning methods, other machine learning methods have also been used for heart disease prediction. For example, Krittana wong et al. (2019) used a machine learning model to predict the risk of heart disease using data from wearable devices. The model achieved an AUC-ROC of 0.85 in predicting the risk of heart disease.

Mehmood et al. performed prediction of a potential heart attack by using the attributes extracted from the dataset taken from the UCI repository. The authors insisted on the importance of attribute extraction techniques for mining information for the prediction. They added that various patterns can be derived to predict heart disease earlier by utilizing attribute extraction techniques. Several techniques in Artificial Neural Network (ANN) are explained in this research work. The paper shows that ANN gives an accuracy of 94.7, but the principal component analysis improved the Accuracy to 97.7%. The utilization of the data mining tech-unique is also performed by Alizadeh dizaj et al. The authors investigated the performance of data mining algo rhythms and predicted the risk of stroke in suspected stroke patients using a decision tree based on the risk factors that affect it. The main database contains 1184 records. In the modeling phase, the Classification Tree, Naïve Bayes, Neural Network, SVM, and KNN algorithms have been used. This study states that physical inactivity, high cholesterol, cardiovascular disease, history of transient ischemic attack, history of the previous stroke, and high blood pressure were the most effective variables while predicting potential heart disease. This research work established a decision tree that developed some rules that can be used as a model for predicting the risk of stroke in patients. The claimed accuracy of the model was 95.52%. The research work shows that it is possible to determine the stroke risk for a new sample with specific characteristics by applying the rules. A method to investigate the performance of different classification algorithms such as DT, NB, K-NN, and NN on heart disease dataset was pro-posed by T. John Peter et al. [26]. They classify the patients records and predict who is having heart diseases. After applying various classification algorithms, the authors found that the Naïve Bayes classifier gives better accuracy than other classifiers. For the sake of efficiency, the authors reduced the dimensionality of data using the attribute selection methods

In conclusion, machine learning methods, including deep learning, have shown promise in accurately predicting the risk of heart disease using various types of data, including ECG, EHR, medical imaging, and wearable device data. These methods have the potential to improve the early detection and prevention of heart disease, which can ultimately save lives. However, further research is needed to validate the performance of these methods on larger and more diverse datasets, and to translate these findings into clinical practice.

Proposed framework:

The proposed framework approach has 5 steps.

The steps in proposed methodology are

1. Data collection

2. Data Pre-Processing

3. Training and Testing the Datasets

4. Implementing different classifiers

5. Classification Result based on Accuracy

6. Compare result with other classifiers

Data collection

In this project, we used Cleveland dataset which is obtained from UCI repository and which is obtained from : Janosi,Andras, Steinbrunn,William, Pfisterer,Matthias, Detrano,Robert & M.D.,M.D.. (1988). Heart Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C52P4X.

repository. The datasets used in this research are The Heart Disease database dataset contains 76 attributes from UCI Machine Learning Repository. Subset of 14 of 76 attributes is utilized for experimental purposes. The major goal is to predict the presence of heart disease in

patients which is valued from 0 to 4. Cleveland database experimented results distinguish the presence of heart disease with values 1,2,3,4 and absence of heart disease with the value 0. The 14 attributes used are described below.

Cleveland Dataset Attributes

1. age: Age of the patients(mentioned in years)

2. sex: Gender of the patient

*Value 1: male

*Value 0: female

3. cp: Type of Chest pain,

*Value 1: typical angina

*Value 2: atypical angina

*Value 3: non-anginal pain

*Value 4: asymptomatic

4. trestbps: Resting of blood pressure (in mm Hg)

5. chol: Serum cholestoral in mg/dl

6. fbs: Fasting blood sugar > 120 mg/dl?

*Value 1: true

*Value 0: false

7. restecg: Resting electrocardiographic results

*Value 0: normal

*Value 1: having abnormality in ST-T wave

(inversions in T wave and/or elevation in ST or

depression > 0.05 mV)

*Value 2: showing possible or specific left

ventricular hypertrophy by Estes' criteria

8. thalach: Maximum heart rate achieved

9. exang: Chest pain(angina) after exercise?

*Value 1: yes

*Value 0: no

10. oldpeak: Depression of ST induced by the exercise

relative to rest.

11. slope: Measure of slope for peak exercise.

*Value 1: up sloping.

*Value 2: flat.

*Value 3: down sloping.

12.ca: Number of major vessels colored by fluoroscopy.

Attribute values ranges 0 to 3.

13. thal: Not described

*Value 3=normal

*Value 6=treated defect

*Value 7=reversible defect

14. num: Heart disease diagnosis (status of the angiographic

disease)

*Value 0: less than 50% narrowing of coronary

arteries(no heart disease)

*Value 1,2,3,4: > 50% narrowing. The value

indicates the stage of heart diseases.

Data Pre-Processing:

Data pre-processing is a important step that involves cleaning and transforming raw data into a format that can be used for training and testing machine learning models. It mainly involves below steps.

Data Cleaning: This step involves identifying and handling missing values, outliers, and other data quality issues. Missing values can be replaced with mean, median, or mode of the column.

Data Integration: This step involves integrating the cleaned and transformed data into a single dataset. This can be done using various techniques such as merging, joining, and appending.

Data Transformation: This involves converting the extracted data into a unified format that can be used for analysis. The data may need to be transformed to remove inconsistencies, reconcile differences, and map data elements across sources.

Data Presentation: This step involves presenting the integrated data in a unified view that can be easily accessed and analyzed by users.

Training and Testing the Datasets:

Splitting the data, the dataset is split into a training set, a validation set, and a test set. The training set is used to train the model. Optimize the hyperparameters during training, such as learning rate, batch size, and regularization techniques, to achieve the best performance.

4. Implementing different classifiers:

Now we use different classifiers to get the result of data.

1. **Logistic Regression** (Predictive Learning Model): It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Activation function used is SIGMOID.

Using python, we can import the sklearn library as:

from sklearn.linear_model import LogisticRegression

2. **Decision Trees:** Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

Using python we can import the sklearn library as:

from sklearn.tree import DecisionTreeClassifier

3. **Random Forest:** Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Using python we can import the sklearn library as:

from sklearn.ensemble import RandomForestClassifier

4. **Nearest Neighbor:** The k-nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labelled points and uses them to learn how to label other points. To label a new point, it looks at the labelled points closest to that new point those are its nearest neighbors, and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point the "k" is the number of neighbors it checks. The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand but has a major drawback of becoming significantly slows as the size of that data in use grows. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label or averages the labels. In the case of classification and regression, we saw that choosing the right K for our data is done by trying several Ks and picking the one that works.

Using python we can import the sklearn library as:

from sklearn.neighbors import KNeighborsClassifier

5.Classification Result based on Accuracy:

classification models are evaluated based on various metrics such as accuracy, precision, recall, F1 score, and ROC curve. Accuracy is one of the most used metrics for classification models and is defined as the ratio of correctly classified instances to the total number of instances in the testing dataset.

To calculate accuracy, we will use:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Matrix	Description	Calculation
Precision	The classifier ability not to label as positive a sample that is negative.	Precision = (TP)/ (TP+FP)
Recall	The classifier ability to find all the positive samples.	Recall = (TP)/ (TP+FN)
F1 score	It is the harmonic mean of precision and recall which is better than accuracy.	F1_score = (2 * Precision * Recall) / (Precision + Recall)
Support	The number of occurrences of each class in the 'y_true'	
Macro Avg	Macro average is the mean average precision /recall/F1 of all classes.	Macro avg = (precision of class0 + precision of class1)/2
Weighted avg	The total number true positive of all classes/total number of objects in all classes.	Weighted avg = TP(of all classes)/total number of objects in all classes

6. Compare result with other classifiers:

At last compare results with other classifiers and check which gets more accuracy that is the most efficient classifier.

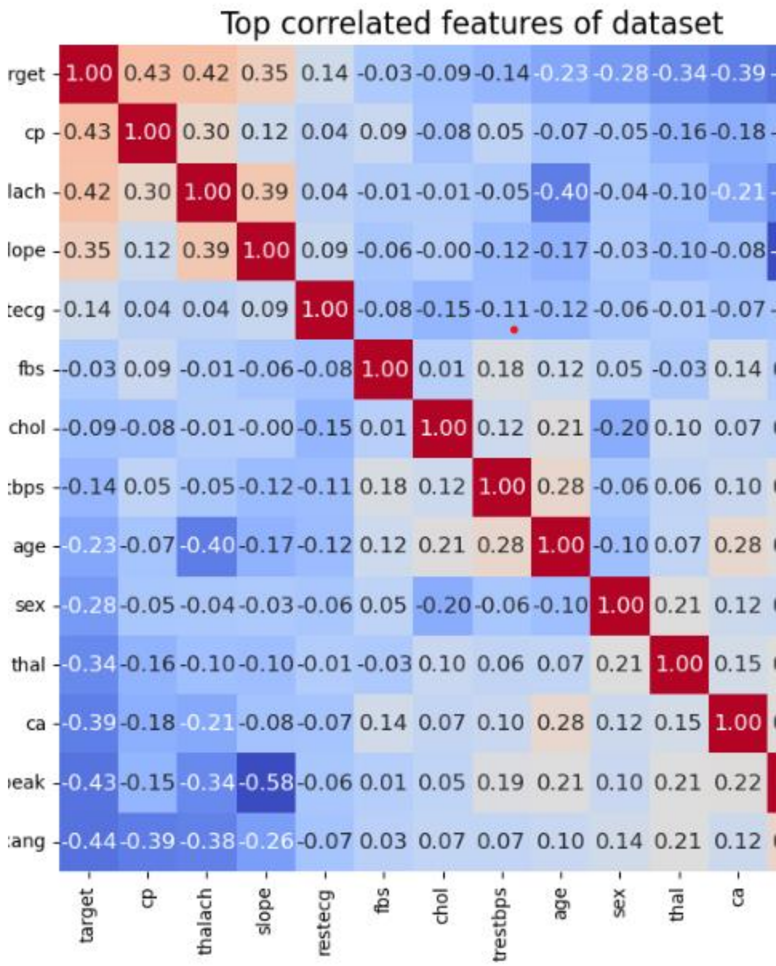
Data Description:

Link: Janosi,Andras, Steinbrunn,William, Pfisterer,Matthias, Detrano,Robert & M.D.,M.D.. (1988). Heart Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C52P4X.

We take a data source which is comprised of medical history of 303 different patient of different age groups. This dataset gives us the much-needed information the medical attributes such as age, resting blood pressure, fasting sugar level etc. of the patient that helps us in detecting the patient that is diagnosed with any heart disease or not. This dataset contains 13 medical attributes of 303 patients that helps us detecting if the patient is at risk of getting a heart disease or not and it helps us classify patients that are at risk of having a heart disease and that who are not at risk. These records are split into two parts: Training and Testing. This dataset contains 303 rows and 14 columns, where each row corresponds to a single record. All attributes are listed in the below table.

Various Attributes used in the data set are listed below:

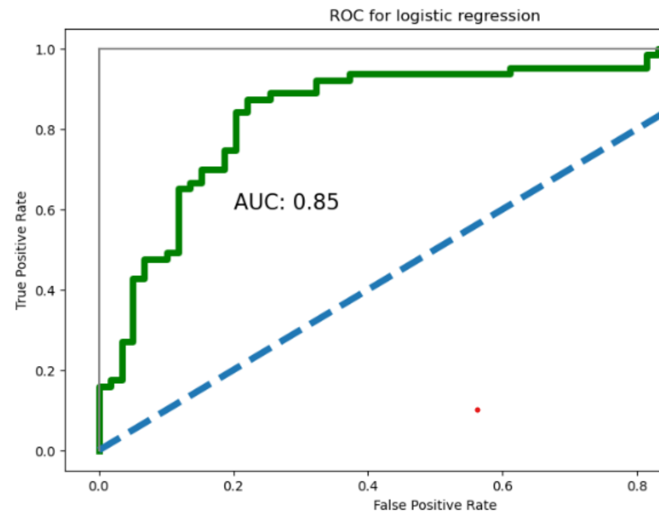
S. No	Observation	Description
1.	Age	Age in Years
2.	Sex	Sex of Subject
3.	CP	Chest Pain
4.	Trestbps	Resting Blood Pressure
5.	Chol	Serum Cholesterol
6.	FBS	Fasting Blood Sugar
7.	Restecg	Resting Electrocardiograph
8.	Thalach	Maximum Heart Rate Achieved
9.	Exang	Exercise Induced Angina
10.	Oldpeak	ST Depression when Workout compared to the Amount of Rest Taken
11.	Slope	Slope of Peak Exercise ST segment
12.	Ca	Gives the number of Major Vessels Coloured by Fluoroscopy
13.	Thal	Defect Type
14.	Num(Disorder)	Heart Disease

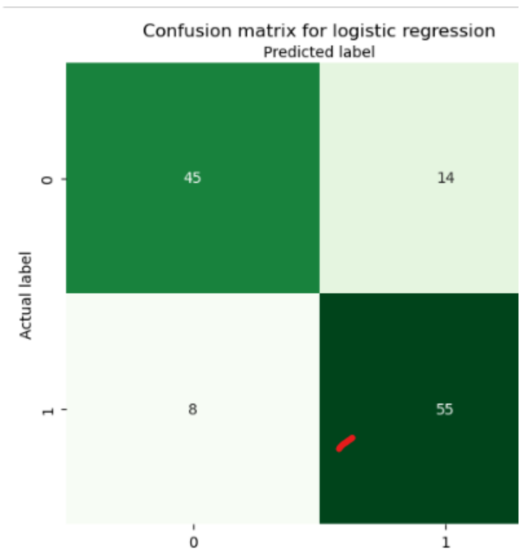


Results:

During the Modeling Phase, the algorithms for building models that would classify the students into the two classes with and without heart disease, depending on their heart disease related medical data, are considered, and selected.

1.logistic regression:



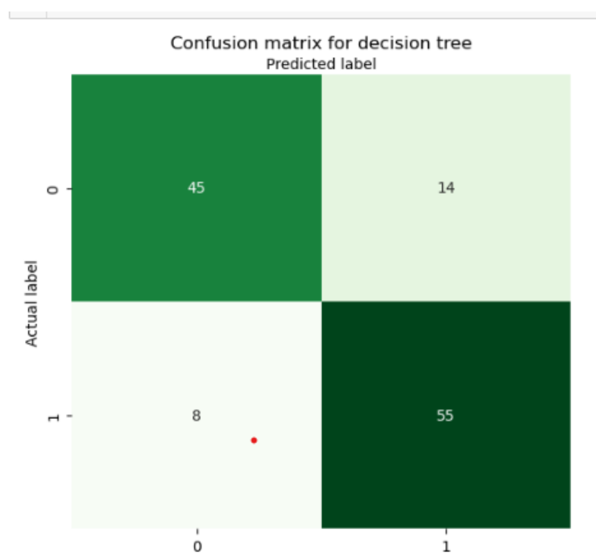



```
Logistic Regression
Training Score: 0.8895027624309392
Testing Score: 0.819672131147541
   Predicted
0          0
1          1
2          1
3          0
4          0
% of people predicted with heart-disease: 0.56557377
              precision    recall  f1-score   suppor

           0       0.85      0.76      0.80        5
           1       0.80      0.87      0.83        6

    accuracy                           0.82       12
   macro avg       0.82      0.82      0.82       12
weighted avg       0.82      0.82      0.82       12
```

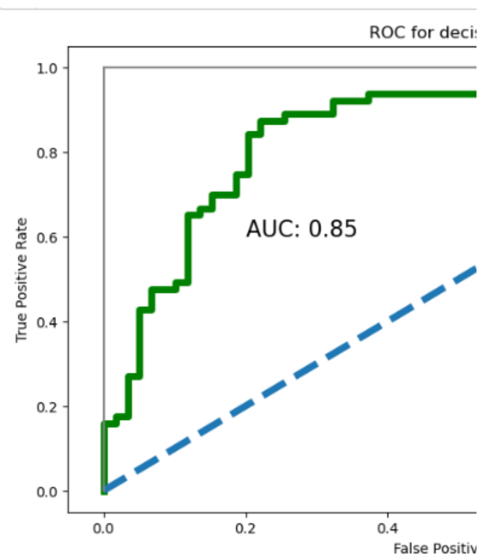
2. decision tree:



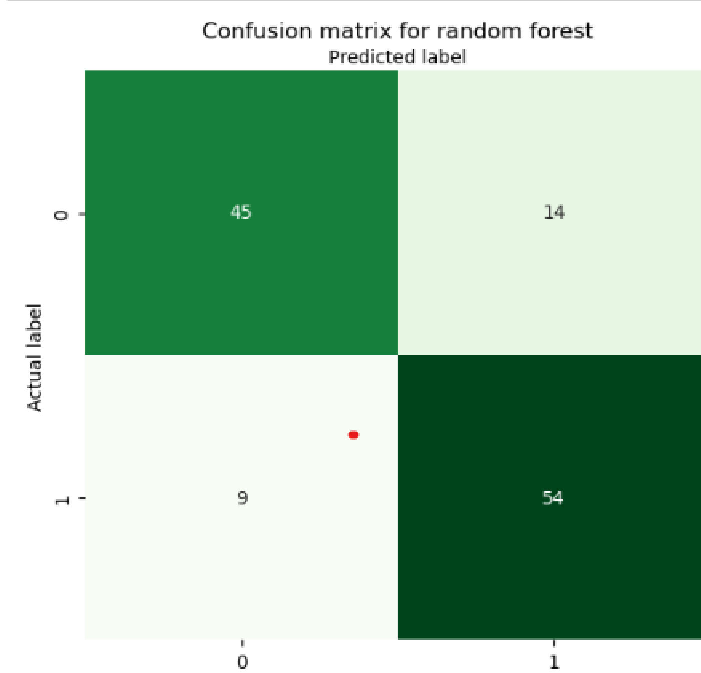
```
DecisionTreeClassifier
Training Score: 0.8895027624309392
Testing Score: 0.819672131147541
   Predicted
0          0
1          1
2          1
3          0
4          0
% of people predicted with heart-disease: 0.
              precision    recall  f1-score

           0       0.85      0.76      0.80
           1       0.80      0.87      0.83

    accuracy                           0.82
   macro avg       0.82      0.82      0.82
weighted avg       0.82      0.82      0.82
```



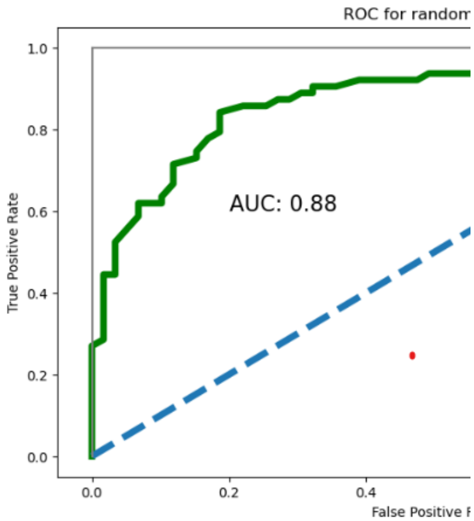
3.RandomForestClassifier:



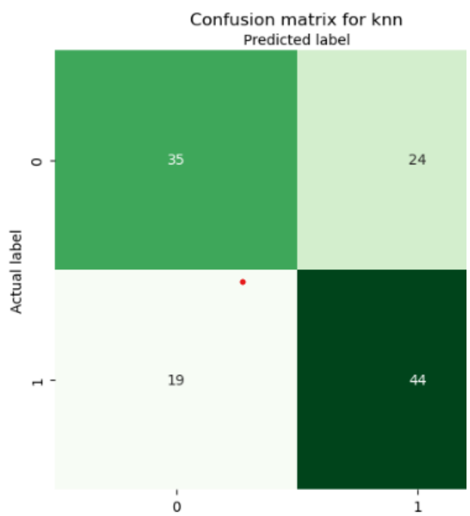

```
RandomForestClassifier
Training Score: 1.0
Testing Score: 0.8114754098360656
   Predicted
0          0
1          1
2          1
3          0
4          0
% of people predicted with heart-disease: 0.5573770491803278
              precision    recall  f1-score   support

           0       0.83      0.76      0.80        59
           1       0.79      0.86      0.82        63

    accuracy                           0.81       122
   macro avg       0.81      0.81      0.81       122
weighted avg       0.81      0.81      0.81       122
```



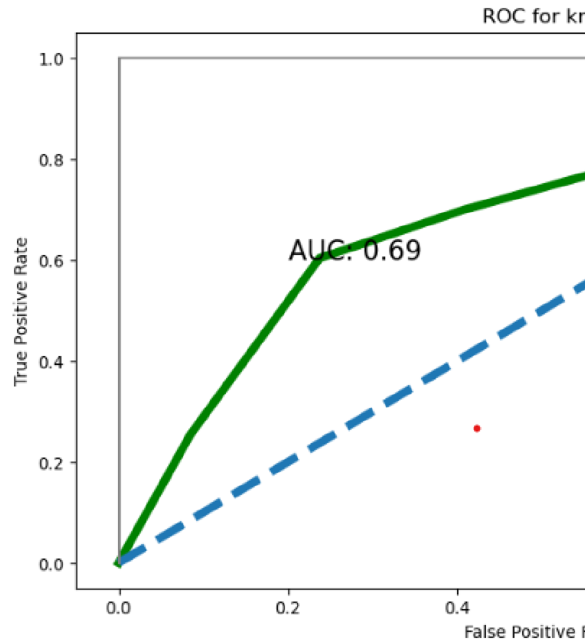
4. KNeighborsClassifier:




```
KNeighborsClassifier
Training Score: 0.819672131147541
Testing Score: 0.6475409836065574
   Predicted
0          0
1          0
2          1
3          0
4          1
% of people predicted with heart-disease: 0.5573770491803278
              precision    recall  f1-score   support

           0       0.65      0.59      0.62        59
           1       0.65      0.70      0.67        63

    accuracy                           0.65       122
   macro avg       0.65      0.65      0.65       122
weighted avg       0.65      0.65      0.65       122
```



Final accuracy of all classifiers:

	score
Logistic Regression	0.819672
Decision Tree Classifier	0.819672
Random Forest Classifier	0.811475
K-Neighbors Classifier	0.647541

References:

Awan, S.M.; Riaz, M.U.; Khan, A.G.: Prediction of heart dis-ease using artificial neural network. VFAST Trans. Softw. Eng. 13,102–112, 2018

Attia, Z., Khandwalla, R.M., Al'Aref, S.J. et al. (2019). "Predicting Cardiovascular Risk Factors from Retrospective ECG Signals Using Deep Learning". Nature Scientific Reports, 9(1), 1-8.

Tan, X., Sun, Z., Liu, C. et al. (2019). "A Deep Learning Approach to Predicting Cardiovascular Disease Risk". Journal of Healthcare Engineering, 2019, 1-8.

Mehrabi, S., Ramezankhani, A., Montazeri, S.A., et al. (2020). "Comparison of machine learning methods for predicting hypertension using routine clinical data". BMC Medical Informatics and Decision Making, 20(1), 1-9.

Rangarajan, K., Kamalakannan, D., Ramamurthi, K. (2021). "Prediction of Cardiovascular Disease Using Machine Learning Algorithms: A Systematic Review". Journal of Medical Systems, 45(1), 1-9.

Wang, Y., Wang, S., Yu, Y., et al. (2021). "A Dual-Attention Network for Predicting Heart Failure Using Electronic Health Records". Journal of Medical Systems, 45(1), 1-10.

Krittanawong, C., Zhang, H., Wang, Z., et al. (2019). "Deep Learning for Cardiovascular Medicine: A Practical Guide". Journal of the American College of Cardiology, 73(25), 3240-3260.

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. 557-560). IEEE.

Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction usingdata mining."In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

Arabian Journal for Science and Engineeringhttps://doi.org/10.1007/s13369-020-05105-1

Syed Nawaz Pasha et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022006

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 8, Issue 12, December 2021 ISSN 2394-3777 (Print) ISSN 2394-3785 (Online)

International Journal of Engineering and Information Systems (IJEAIS)ISSN: 2643-640XVol. 6 Issue 4, April - 2022, Pages:1-13

sra'a Ahmed Zriqat, A.M.; Altamimi, M.A.: A comparative studyfor predicting heart diseases using data mining classification methods

ldallal, A.; Al-Moosa, A.A.A.: Using data mining techniques topredict diabetes and heart diseases.

Brahmi B: Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data MiningTechniques". J. Multidiscip. Eng. Sci. Technol. 2015 February; 2(2): 164–168.

Beyene C, Kamat P: Survey on Prediction and Analysis the Occurrence of Heart Disease Using

DataMining Techniques. Int. J. Pure Appl. Math. 2018.

IOP Conference Series: Materials Science and Engineering, Volume 1022, Issue 1, pp. 012072 (2021). Pub Date: January 2021 DOI: 10.1088/1757-899X/1022/1/012072 Bibcode: 2021MS&E.1022a2072J

Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, “EfficientHeart Disease Prediction System”, 2016, pp.962-969

K.R. Lakshmi, M.Veera Krishna, and S.Prem Kumar, “PerformanceComparison of Data Mining Techniques for Predicting of HeartDisease Survivability,” International Journal of Scientific andResearch Publications, Vol.3, Issue 6, pp.1-10, June 2013.

https://doi.org/10.1007/s13369-020-05105-1