Heart Care Heart Attack Possibility

Shahajadi Sadia Afsana
ID: 180104138
Group: C2

01

INTRODUCTION

The heart seems to be a very complicated organ in the human body.If some part of the heart has been seriously damaged, the remaining part of the heart will remain functioning. But as a result of the injury, the heart can be weakened and unable to pump as much blood as normal.With timely detection of those injuries can save life.Various types machine learning algorithm including random forest, MLP, KNN, and Adaboost can be used for this detection.
My attempt is to detect that possibility of heart attract using Machine Learning approach in this study.This paper determines the optimum algorithm and illustrates the confusion matrices, features, and ROC curve. It is clear from this research that KNN is the best model, with an accuracy of roughly 82%.

01

02

Literature Reviews

Cardiovascular disease had become the leading cause of death in the United States by 1940[1]. When President Franklin D. Roosevelt died in 1945 of hypertensive heart disease,Americans became interested in heart disease research [2]. The Framingham Heart Study was founded in the United States in 1948 to study heart disease and its various risk factors[3]. Since then, there has been continuous research on heart failure and its various risk factors, with the goal of better preventing it with modern technology. In his paper, Fizar Ahmed [4] explained the architecture of heart attack rates and used the IoT concept to predict future heart attack patients. He also used one of the most popular machine learning algorithms, kNN (k Nearest Neighbour), to complete his work and achieve better accuracy.Prince Kansal et al. [5] focused on how to predict heart disease early using various data mining techniques, and they used nearly four machine learning algorithms. They used age,gender, blood pressure, and blood sugar levels in their dataset. They got better results from the decision tree. Manually diagnosing heart disease, as Asha Rajkumar [2] points out, takes a long time and requires the assistance of experts. They were particularly interested in her paper on fast diagnosis of heart disease using data mining techniques. Three algorithms and tanagra tools were used. The Naive Bayes algorithm took 609ms to diagnose the heart disease.

03

Data Collection & Processing

Dataset Overview

For datasets, I have used dataset from Heart Disease Data Set of UCI Machine Learning Repository [6]. This database contains 76 attributes, but here I used only 14 of them.attributes are :

1. Age
2. Sex
3. Chest Pain Type
4. Resting Blood Pressure (in mm)
5. Serum Cholesterol in mg/dl
6. Fasting Blood Sugar > 120 mg/dl
7. Resting Electrocardiographic Results
8. Maximum Heart Rate Achieved
9. Exercise Induced Angina
10. ST depression induced by exercise relative to rest
11. The Slope of the Peak Exercise ST Segment
12. Number of Major Vessels (0-3) Colored by Fluoroscopy
13. Thal
14. Diagnosis of Heart Disease (Angiographic Disease Status)

Data sample and Distribution

Sample Dataset

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Distribution

	0 Labeled (Less Chance)	1 Labeled (More Chance
Train.csv = 242	110	132
Test.csv = 61	28	33
Total = 303		

Data Splitting

I divided the final data for training and testing in an 8:2 ratio. 80% of the data for training, while 20% were set aside for testing.

Total Data	Train Data	Test Data
303	242	61
100% data	80% data	20% data

Data Splitting

04

Methodology

Methodology

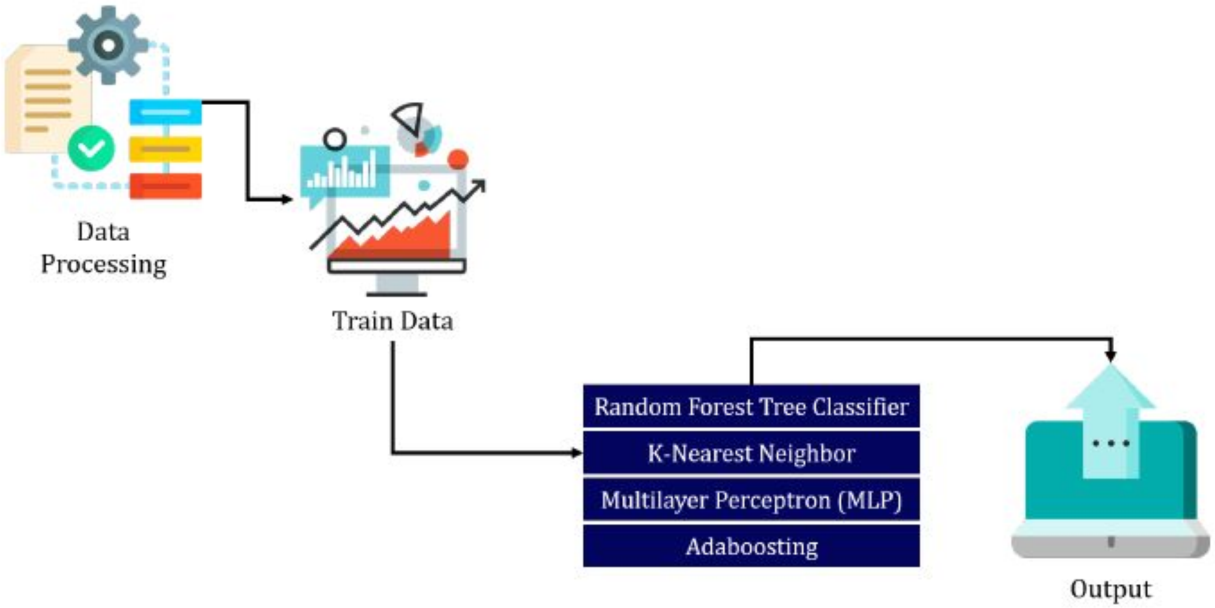


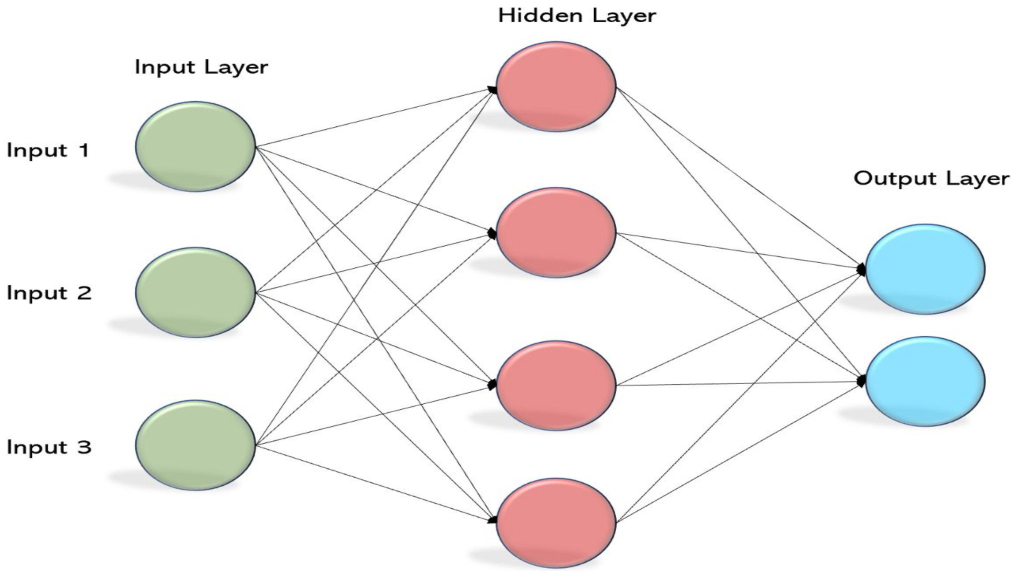
Figure 4.1: Working Methodology

K-Nearest Neighbours

Here is an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

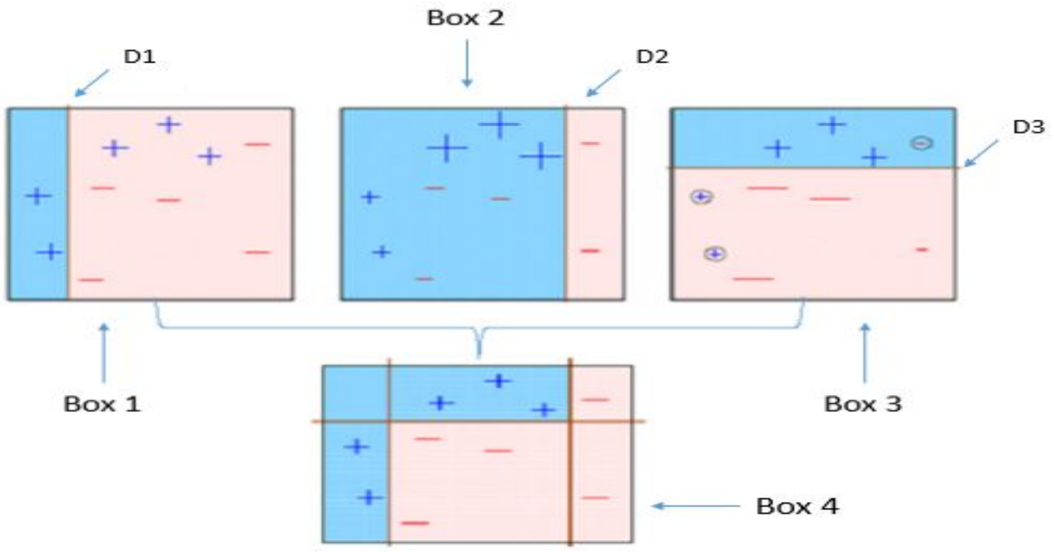
Multi layer Perceptron (MLP)

A Multilayer Perceptron has input and output layers, and one or more **hidden layers** with many neurons stacked together. And while in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a Multilayer Perceptron can use any arbitrary activation function.



Adaboosting

AdaBoost also called Adaptive Boosting is a technique in Machine Learning used as an Ensemble Method.here the weak learners are produced parallelly during the training phase. The performance of the model can be increased by parallelly training a number of weak learners on data sets.All weak learners make a strong output together.



04

Random Forest Tree Classifier

N samples are drawn from a fruit basket, and an individual decision tree is built for each sample. The final result is determined by majority voting.Here, the majority decision tree produces an apple rather than a banana, so the final output is an apple.

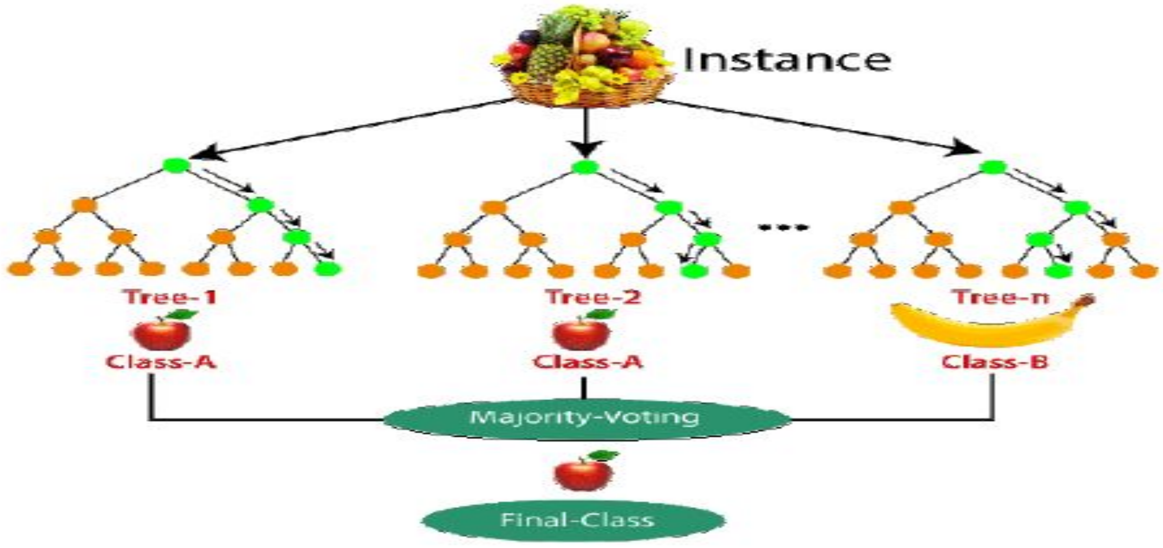


Figure 4.2: Random Forest Example

05

Experiments and Results

KNN

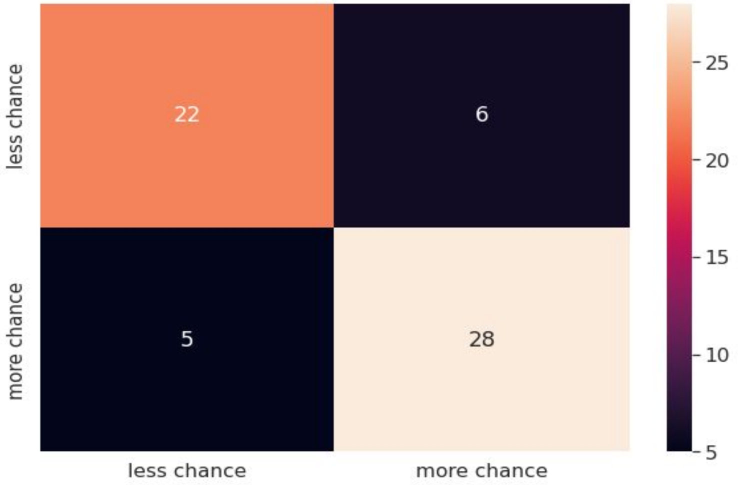
Train Average Accuracy is 82.21%

Experimental Results

	precision	recall	f1-score	support
less chance	0.81	0.79	0.80	28
more chance	0.82	0.85	0.84	33
accuracy			0.82	61
macro avg	0.82	0.82	0.82	61
weighted avg	0.82	0.82	0.82	61

Table 5.1: K-Nearest Neighbor Classification Report

Confusion Matrix



MLP

Train Average Accuracy is 80.97%

Experimental Results

	precision	recall	f1-score	support
less chance	0.72	0.75	0.74	28
more chance	0.78	0.76	0.77	33
accuracy			0.75	61
macro avg	0.75	0.75	0.75	61
weighted avg	0.76	0.75	0.75	61

Table 5.2: Multilayer Perceptron Classification Report

Confusion Matrix

AdaBoosting

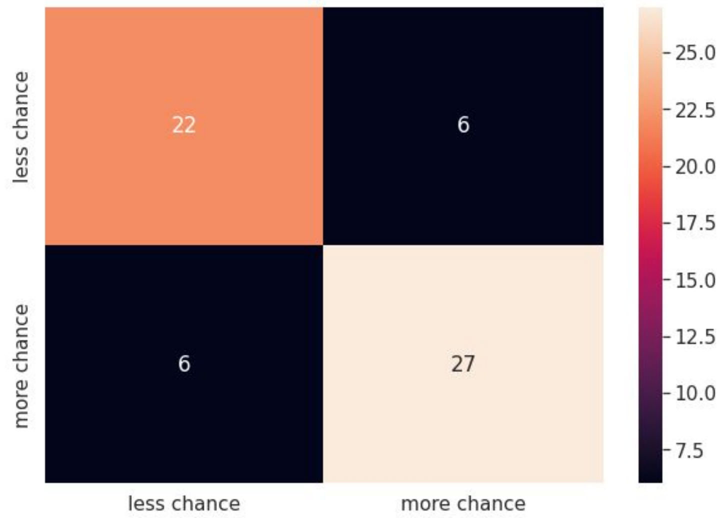
Train Average Accuracy is 82.63%

Experimental Results

	precision	recall	f1-score	support
less chance	0.79	0.79	0.79	28
more chance	0.82	0.82	0.82	33
accuracy			0.80	61
macro avg	0.80	0.80	0.80	61
weighted avg	0.80	0.80	0.80	61

Table 5.3: Adaboosting Classification Report

Confusion Matrix



RF

Train Average Accuracy is 84.68%

Experimental Results

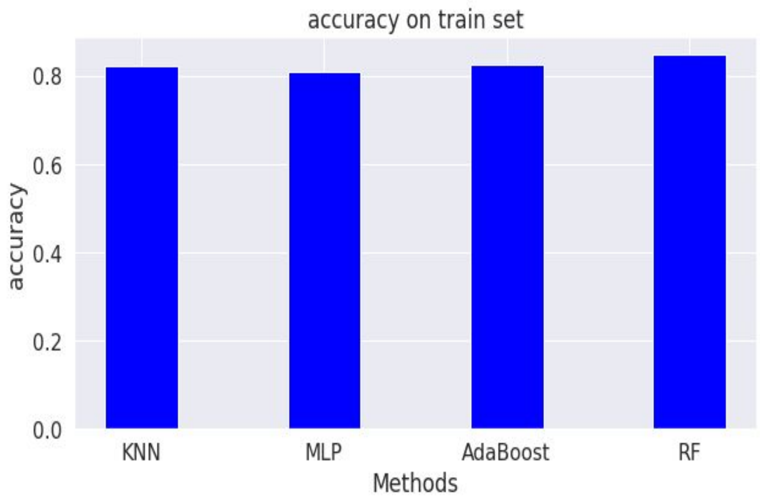
	precision	recall	f1-score	support
less chance	0.76	0.79	0.77	28
more chance	0.81	0.79	0.80	33
accuracy			0.79	61
macro avg	0.79	0.79	0.79	61
weighted avg	0.79	0.79	0.79	61

Table 5.4: Random Forest Classification Report

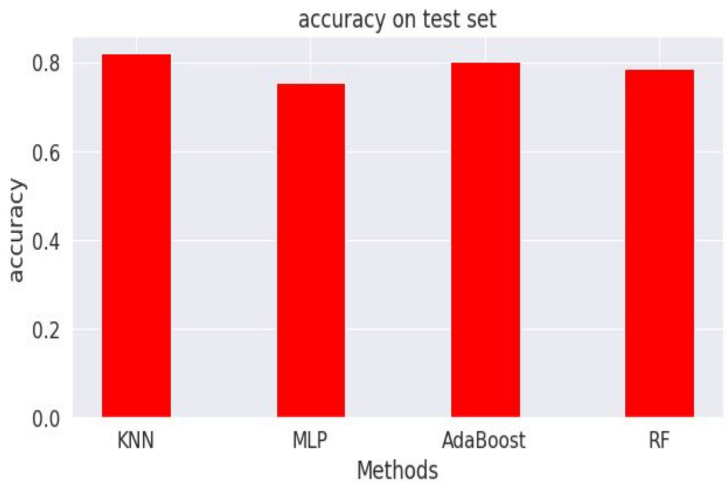
Confusion Matrix

BAR CHART

Accuracy Bar chart for train set



Accuracy Bar chart for test set



05

Accuracy Comparison of models

MODELS

KNN

TRAINING ACC-82.2%
TESTING ACC- 82%

MLP

TRAINING ACC-81%
TESTING ACC- 75.4%

AdaBoost

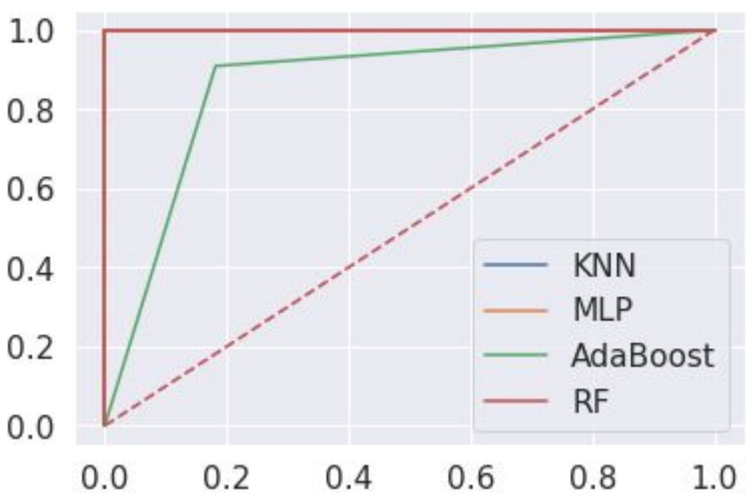
TRAINING ACC-82.6%
TESTING ACC- 80.3%

RF

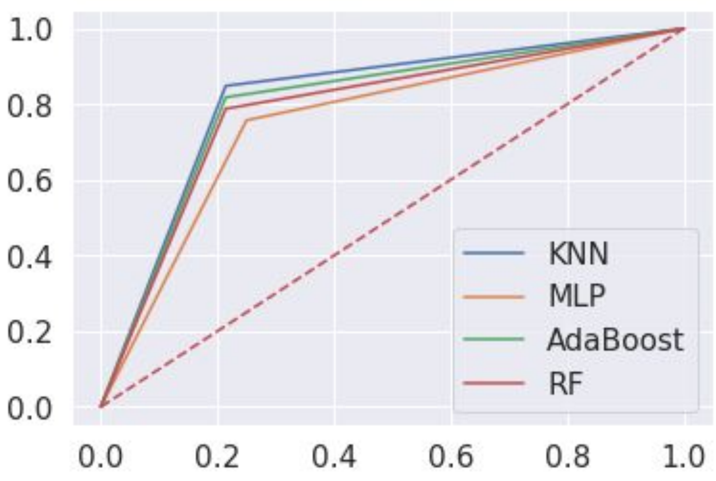
TRAINING ACC-84.7%
TESTING ACC- 78.7%

ROC curve Analysis

ROC curve for traindata



ROC curve for test data



05

ROC curves display the performance of a classification model.The greater the AUC for the ROC Curve, the better the performance. We can see from the constructed curves that the KNN has the highest AUC, indicating that it is the best model for this classifier.

06

Future Work and Conclusion

Future Work and Conclusion

Future Works

My experiment produced good results, but the accuracy may be increased by testing with more data and more sophisticated ways. Using several models can also assist to enhance performance.

Conclusion

The heart attack possibility performance is shown in this research analysis. The performance analysis identified several categories, including confusion metrics, precision, recall,f measure, and auc. In terms of overall performance, KNN outperforms the others, with an accuracy of 82%. To determine the best performance from our dataset, some machine learning techniques are used. In the future, this study will incorporate deep learning and artificial intelligence methods to analyze and predict the possibility of a heart attack. I also added more information as needed.

References

- W. B. Kannel, “Contribution of the framingham study to preventive cardiology,” Journal of the American College of Cardiology, vol. 15, no. 1, pp. 206–211, 1990.
- A. Rajkumar and G. S. Reena, “Diagnosis of heart disease using data mining algorithm,” Global journal of computer science and technology, vol. 10, no. 10, pp. 38–43, 2010.
- S. S. Mahmood, D. Levy, R. S. Vasan, and T. J. Wang, “The framingham heart study and the epidemiology of cardiovascular disease: a historical perspective,” The lancet, vol. 383, no. 9921, pp. 999–1008, 2014.
- F. Ahmed, “An internet of things (iot) application for predicting the quantity of future heart attack patients,” International Journal of Computer Applications, vol. 164, no. 6,pp. 36–40, 2017.
- A. Methaila, P. Kansal, H. Arya, P. Kumar, et al., “Early heart disease prediction using data mining techniques,” Computer Science & Information Technology Journal, vol. 28,pp. 53–59, 2014.
- “Heart disease data set.” https://archive.ics.uci.edu/ml/datasets/Heart+Disease. Accessed: 2022-08-15.

THANK YOU!