Heaven's Light is Our Guide

Course No: 4000

Coronary Artery Disease(CAD) Prediction Using Machine Learning Methods

Presented By
Dibbo Barua Chamak
Roll: 1603117
Computer Science and Engineering,
Rajshahi University of Engineering and Technology

Supervised By
Professor Dr. Md. Ali Hossain
Professor
Computer Science and Engineering,
Rajshahi University of Engineering and Technology

Presentation Outline

Introduction

- Cardiovascular diseases (CVDs) are the leading cause of death globally.
- About 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack and stroke. (WHO report)
- Coronary Artery Disease(CAD) appears to be the most common Cardiovascular defect.
- Coronary artery disease is a narrowing or blockage of our coronary arteries usually caused by the buildup of fatty material called plaque.
- The major blood vessels that supply the heart (coronary arteries) struggle to send enough blood, oxygen and nutrients to the heart muscle.

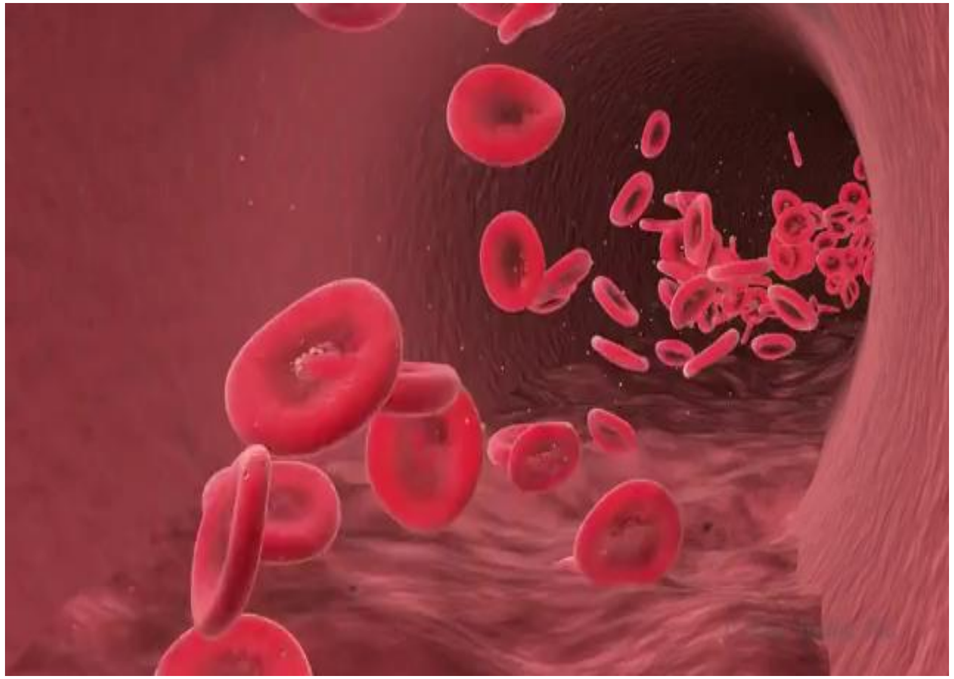


Fig 1:Plaque builds up in the arteries

Introduction(cont.)

- Coronary artery disease is also called coronary heart disease, ischemic heart disease and heart disease

- Symptoms:
- Chest pain or discomfort (angina)
- Weakness, light-headedness, nausea (feeling sick to your stomach)
- Fatigue
- Pain or discomfort in the arms or shoulder
- Shortness of breath

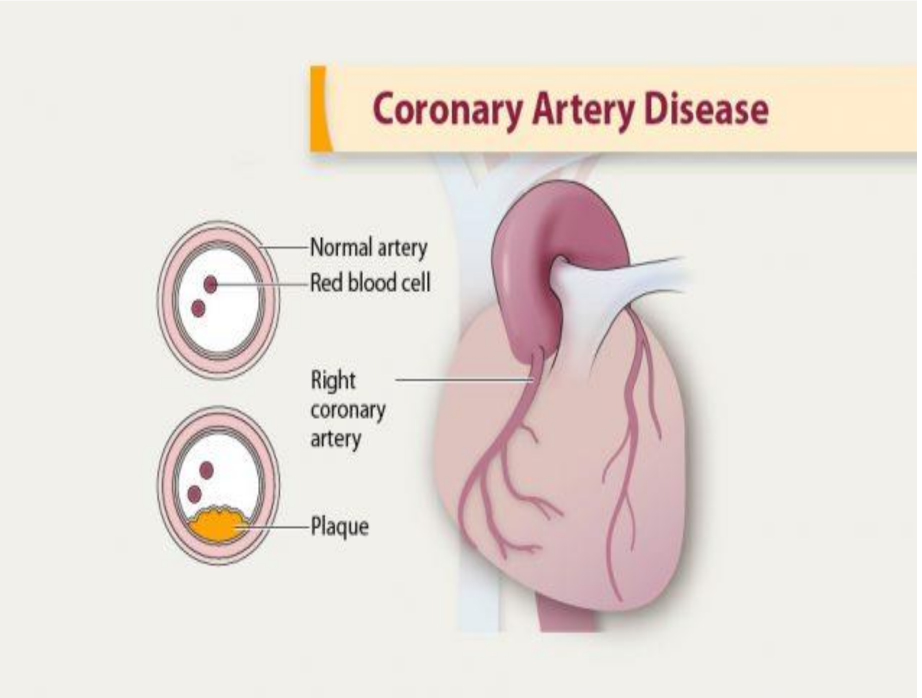


Fig 2: Coronary Artery Disease

Literature Review

Title : Coronary Artery Disease Prediction using Data Mining Techniques, Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020] IEEE Xplore Part Number: CFP20M19-ART; ISBN: 978-1-7281-7089-3

Review :

- This paper showed comparison among three machine learning methods- SVM ,Naïve bayes and Random Forest model.
- Random Forest showed better performance .
- It has considered 14 attributes to predict CAD.
- Here, the accuracy were low as compared to others.

Limitations:

- They have worked on a small dataset and their result accuracy were low as compared to others.

Literature Review(Cont.)

Title : Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method -A Comparative Study , Journal of Artificial Intelligence and Capsule Networks (2021) Vol.03/ No.01 Pages: 17-33

Review :

- This paper showed Comparison of two classifier- Naïve Bayes & SVM.
- SVM showed better result than the Naïve Bayes classifier.
- It proposed SVM classifier with pooling layer which showed the better result.
- But it has considered 80 features to predict the CAD disease.

Limitations:

- There is research gap in the early prediction of CAD based on clinical parameters.

Literature Review(Cont.)

Title : Predicting coronary artery disease: a comparison between two data mining algorithms, Ayatollahi et al. BMC Public Health (2019) 19:448,https://doi.org/10.1186/s12889-019-6721-5

Review :

- This paper Compared between SVM & ANN model.
- SVM model showed higher accuracy and better performance than the ANN model.
- It has considered 25 features.
- It has used multilayer perceptron with 3 hidden layer in ANN model.

Limitations:

- This paper suggested to use other Data Mining Algorithms to improve the positive predictive value of the disease prediction.

Motivation

- With millions of people dying from the disease each year, early predictions can prolong human deaths from heart disease problems.
- This motivates me to work on this topic to predict CAD in early Stage.

Objectives

- To predict Coronary Artery Disease with higher accuracy level .
- To predict CAD using less feature than the previous works.
- To show the Comparison between multiple classifiers.
- To compare the performance of my work and the previous works.

Datasets

- A dataset named 'heart_statlog_cleveland_hungary' is used in this work.
- The dataset consists of 1190 records of patients from US, UK, Switzerland and Hungary. It has 11 features and 1 target variable.

Datasets

	age	sex	chest pain type	resting bp s	cholesterol	fasting blood sugar	resting ecg	max heart rate	exercise angina	oldpeak	ST slope	target
0	40	1	2	140	289	0	0	172	0	0.0	1	0
1	49	0	3	160	180	0	0	156	0	1.0	2	1
2	37	1	2	130	283	0	1	98	0	0.0	1	0
3	48	0	4	138	214	0	0	108	1	1.5	2	1
4	54	1	3	150	195	0	0	122	0	0.0	1	0

Fig 3: Dataset Information

Chest pain Type:
0: typical angina
1: atypical angina
2: non-anginal pain
3: asymptomatic

sex :
1: male
0: female

fasting blood sugar:
> 120 mg/dl
1: true
0: false

Resting electrocardiographic results:
0: normal
1: having ST-T wave abnormality (T wave)

exercise induced angina:
1: yes
0: no

Datasets Analysis

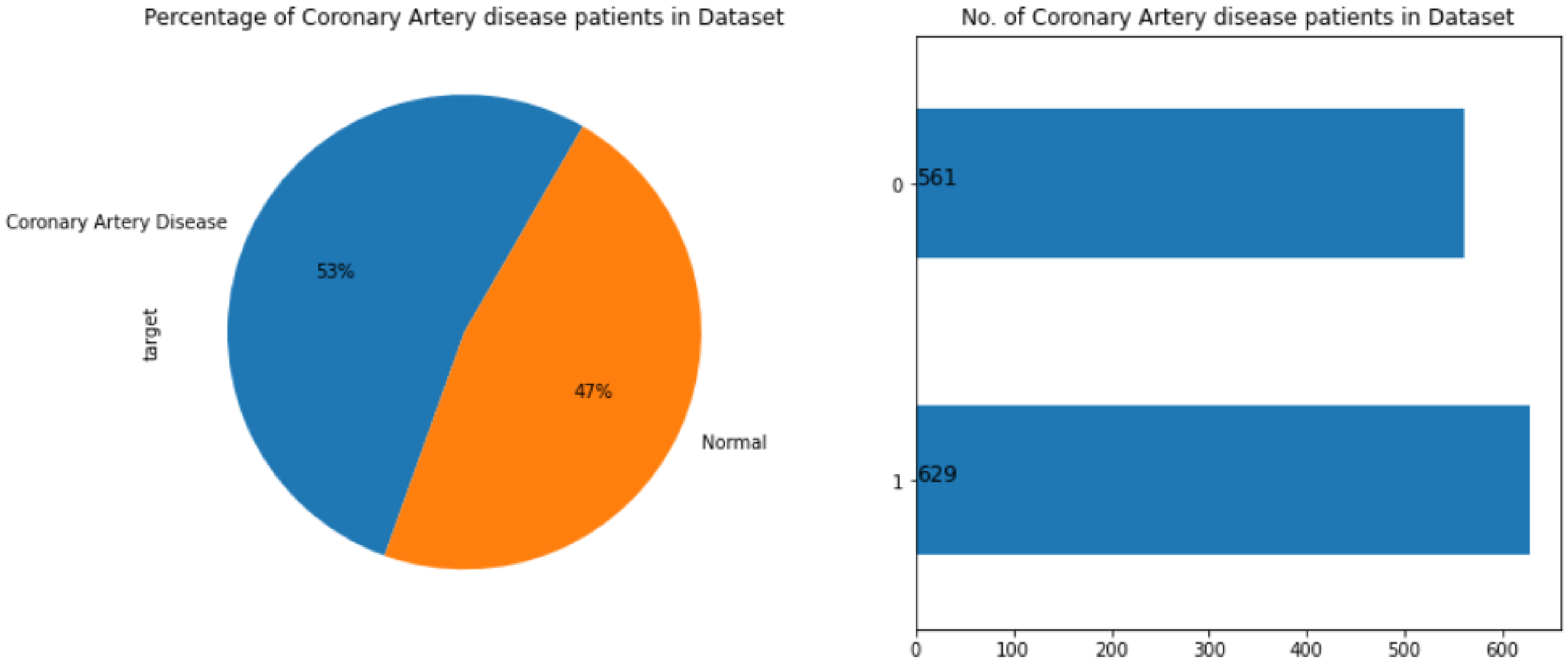


Fig 4: Number of CAD patients in Dataset

Datasets Analysis(cont.)

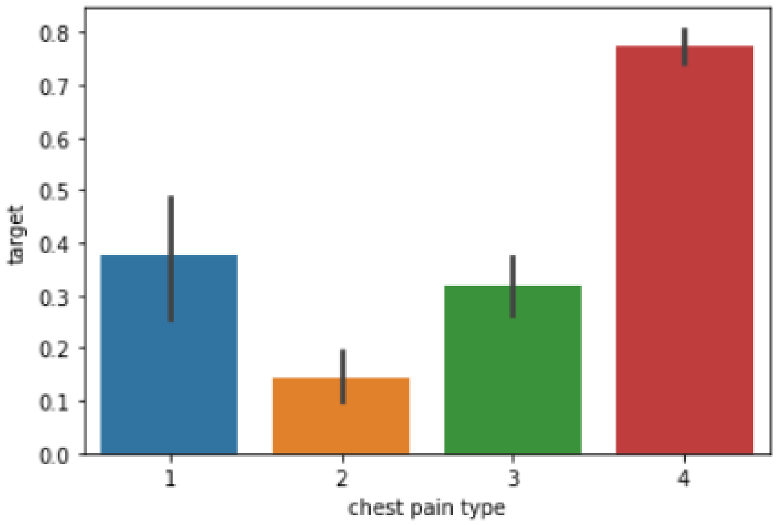


Fig 5: Chest pain type VS Target in Dataset

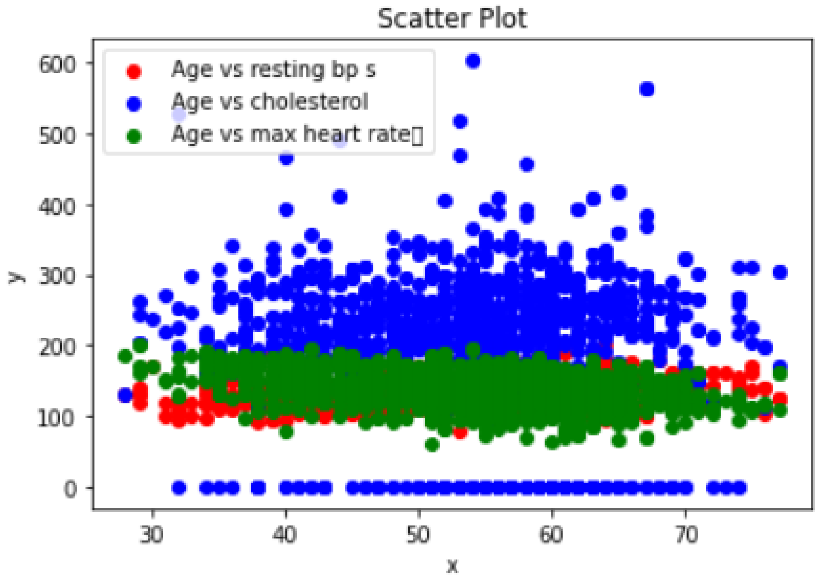


Fig 6: Scatterplot

Datasets Analysis(cont.)

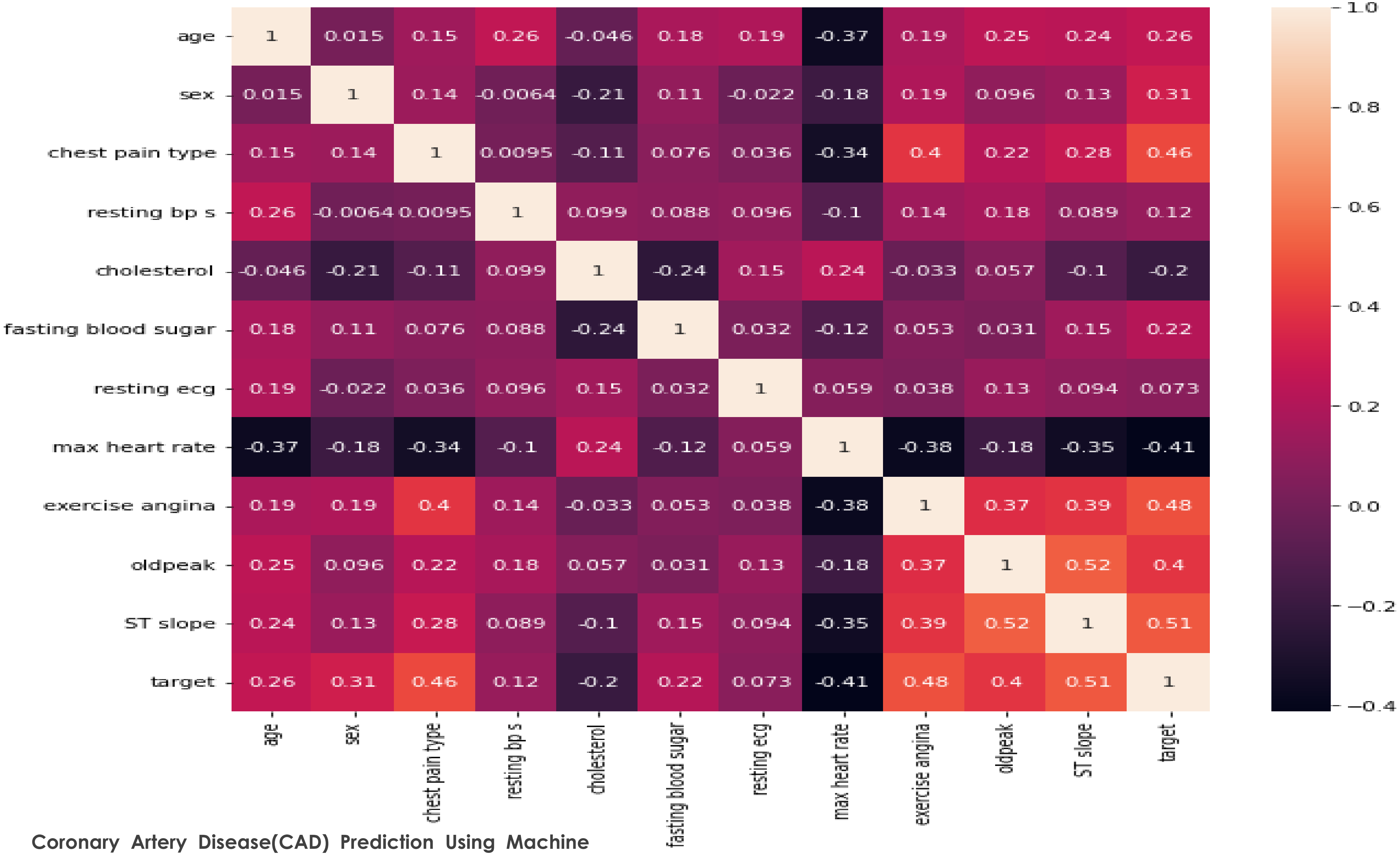


Fig 7: Heat map of Dataset

Methodology

- First, we will take raw Input Data
- Then we will preprocess those data & convert the categorical variables into nominal variables
- then we will apply classifiers(SVM , Naïve Bayes, Random forest , Neural Network) to classify CAD.
- After that we will evaluation the performance & see the result

Methodology(cont.)

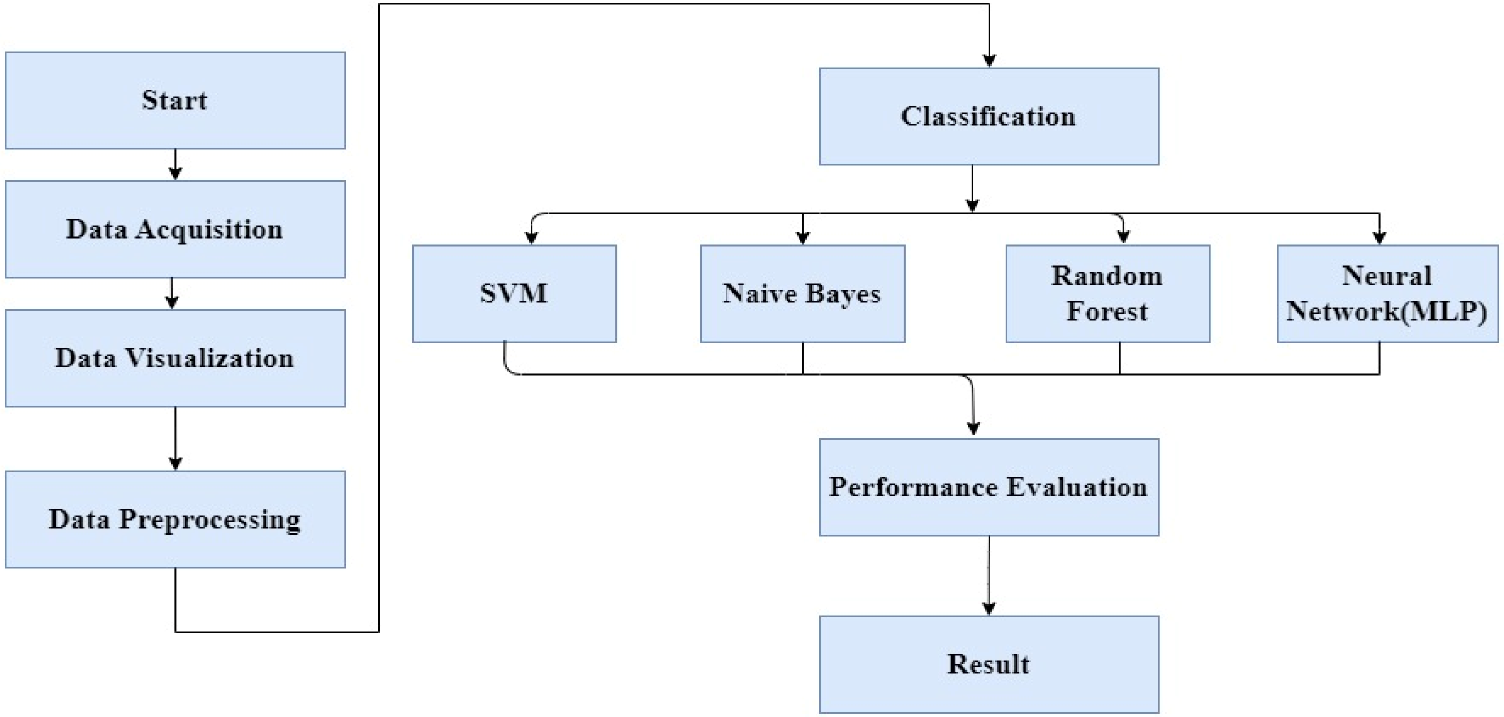


Fig 8: Block diagram of Methodology

Methodology(cont.)

- **SVM :**

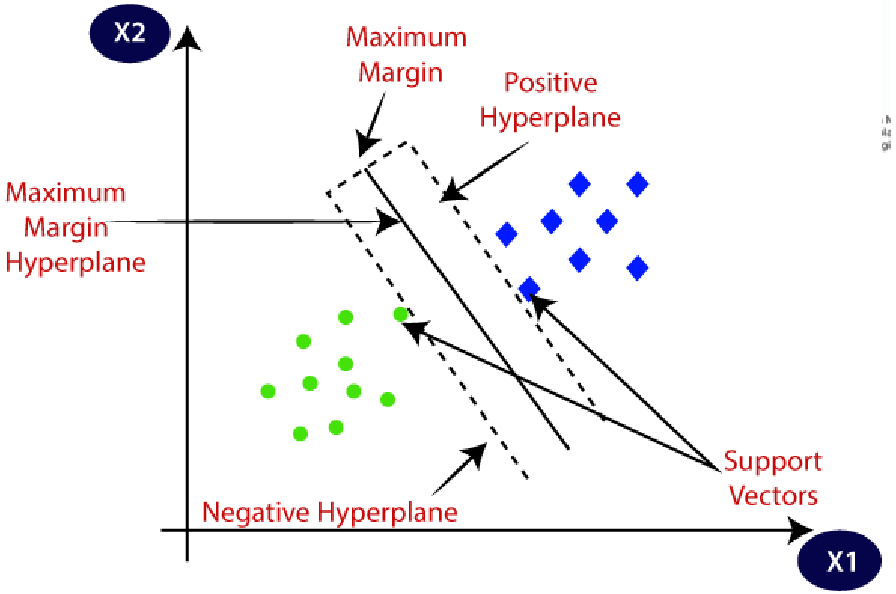


Fig 9: Support Vector Machine

Methodology(cont.)

➢ **SVM with RBF kernel**

Formula:

$$K(X1, X2) = \exp - \frac{\|X1 - X2\|^2}{2\sigma^2}$$

Where,

- σ' is the variance and our hyperparameter,
- ||X1 - X2|| is the Euclidean (L2-norm) Distance between two points X1 and X2

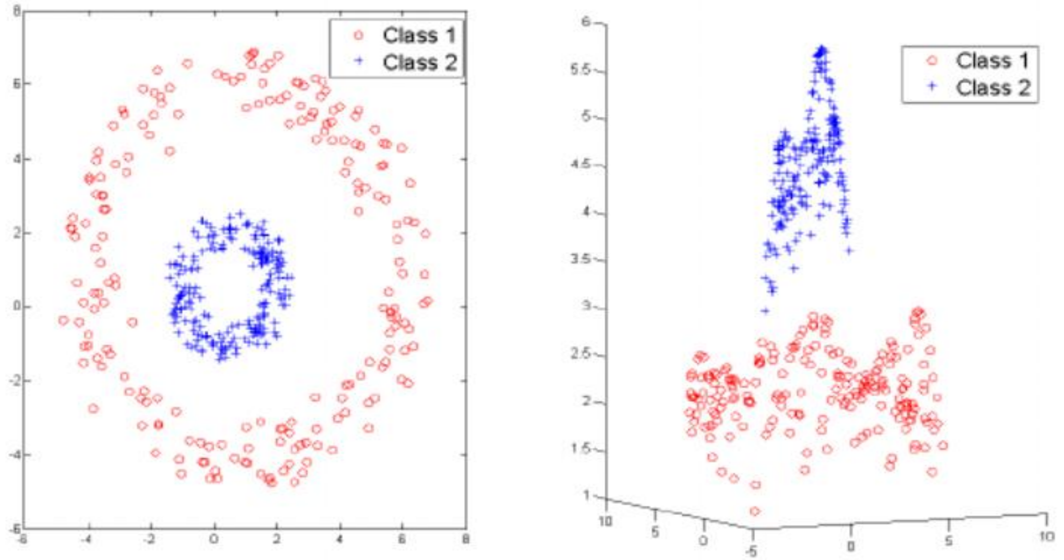


Fig 10: Example of Radial Basis Function (RBF) kernel mapping data from non linear separable space to high-dimensional separable space.

Methodology(cont.)

- **SVM with RBF kernel**

- C=1000, gamma=0.0001, kernel=rbf

- **Results:**

- Accuracy : 0.86
- Precision : 0.97
- Recall : 0.69
- F1 Score: 0.81

Accuracy = TP+TN/TP+FP+FN+TN	**Recall = TP/TP+FN**
Precision = TP/TP+FP	**F1 Score = 2*(Recall * Precision) / (Recall + Precision)**

Methodology(cont.)

- **Naïve Bayes:**
- **Formula:**
- **Bayes Theorem:**

  P(A|B) = P(B|A) * P(A)/P(B)

Methodology(cont.)

- **Naïve Bayes:**
- **Result:**

- Accuracy: 0.84
- Precision : 0.84
- Recall : 0.85
- F1 Score: 0.85

Accuracy = TP+TN/TP+FP+FN+TN	Recall = TP/TP+FN
Precision = TP/TP+FP	F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Methodology(cont.)

- **Random Forest Classifier:**

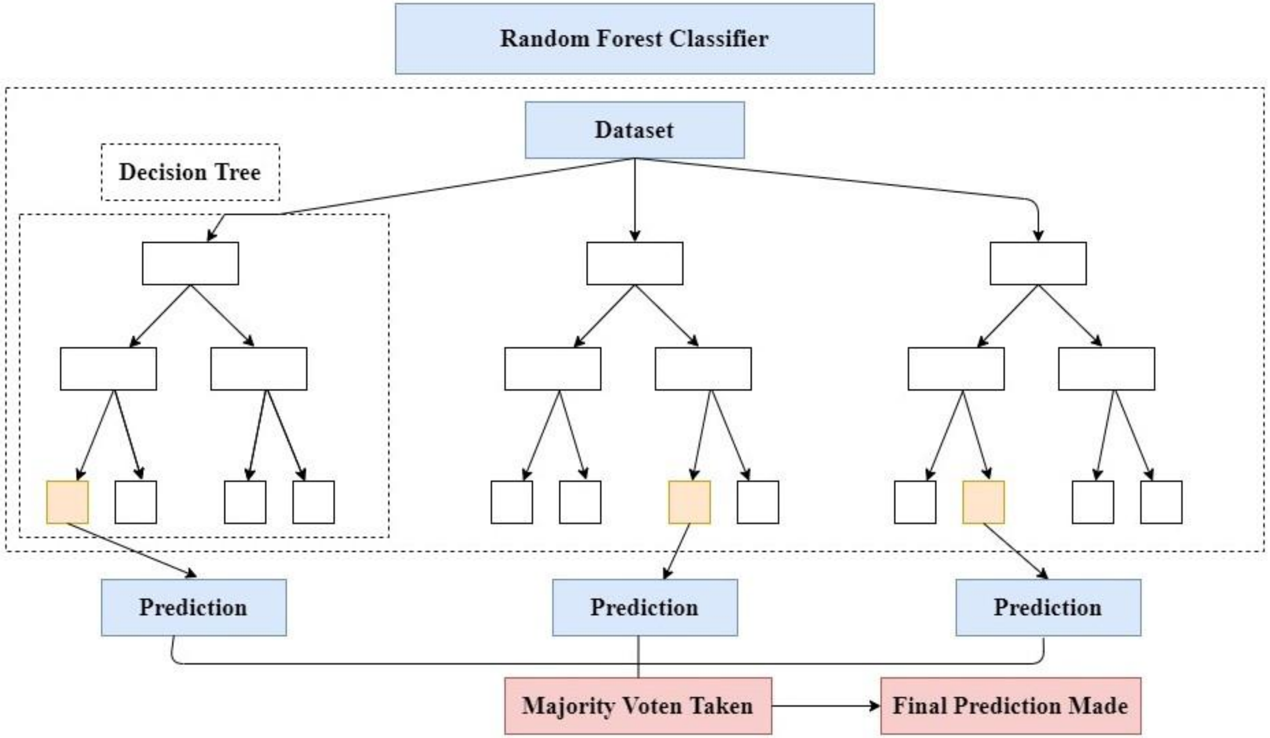


Fig 11: Overview of Random Forest Classifier

Methodology(cont.)

- **Random Forest Classifier:**
- **Random Hyperparameter Grid:**
- {'bootstrap': True,
  'max_depth': 70,
  'max_features': 'auto',
  'min_samples_leaf': 4,
  'min_samples_split': 10,
  'n_estimators': 400}

Methodology(cont.)

- **Random Forest Classifier:**

- **Results:**
  - Accuracy : 0.94
  - Precision : 0.97
  - Recall : 0.91
  - F1 Score: 0.94

Accuracy = TP+TN/TP+FP+FN+TN	Recall = TP/TP+FN
Precision = TP/TP+FP	F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Methodology(cont.)

- **Neural Network(Multi-Layer Perceptron):**
- **Hyperparameter tuning setup:**
- hidden_layer_sizes : (5,10,15,20,25,)
- solver='lbfgs'
- activation='relu'
- max_iter=200

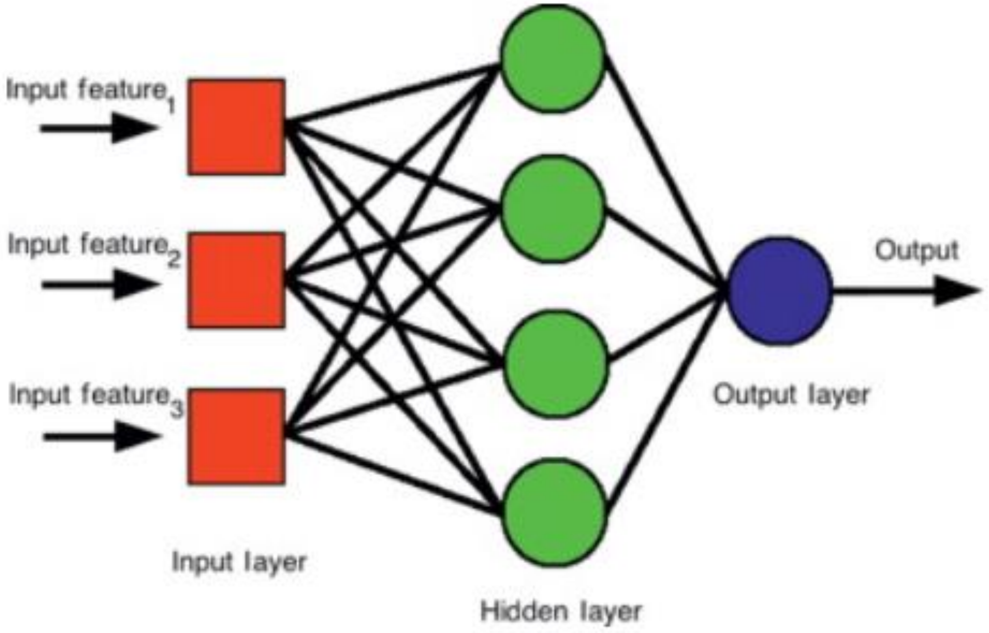


Fig12: Multi-Layer Perceptron

Methodology(cont.)

- **Neural Network(Multi-Layer Perceptron):**
- **ReLU(Rectified Linear Unit) Activation Function:**

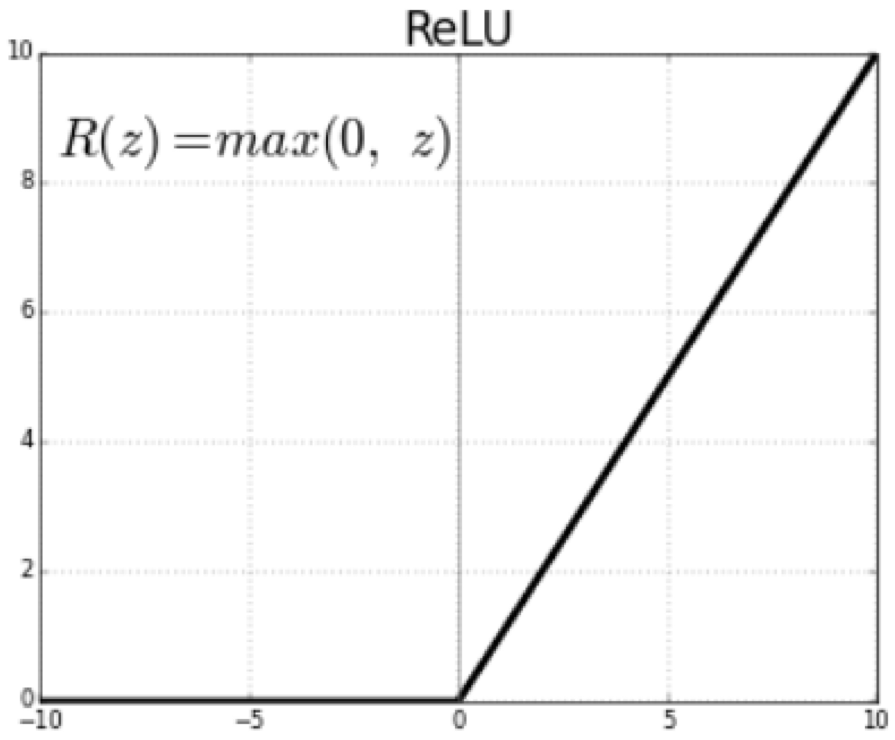


Fig13 : ReLU function

Methodology(cont.)

- **Neural Network(Multi-Layer Perceptron):**

- **Results:**

- Accuracy: 0.81
- Precision: 0.82
- Recall : 0.83
- F1 Score: 0.82

Accuracy = TP+TN/TP+FP+FN+TN	Recall = TP/TP+FN
Precision = TP/TP+FP	F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Results

➢ Result & Comparison:

Classifiers	Previous Accuracy	Present Accuracy
Support Vector Machine	0.86	0.85
Naive Bayes	0.82	0.84
Random Forest	0.90	0.94
Neural Network	-	0.81

Fig 14: Comparison of Accuracy

Classifiers	Precision	Recall	F1 Score
Support Vector Machine	0.97	0.69	0.81
Naive Bayes	0.84	0.85	0.85
Random Forest	0.97	0.91	0.94
Neural Network	0.82	0.83	0.82

Fig 15 : Precision, Recall,F1 score of classifiers

Results(cont.)

- Result & Comparison:

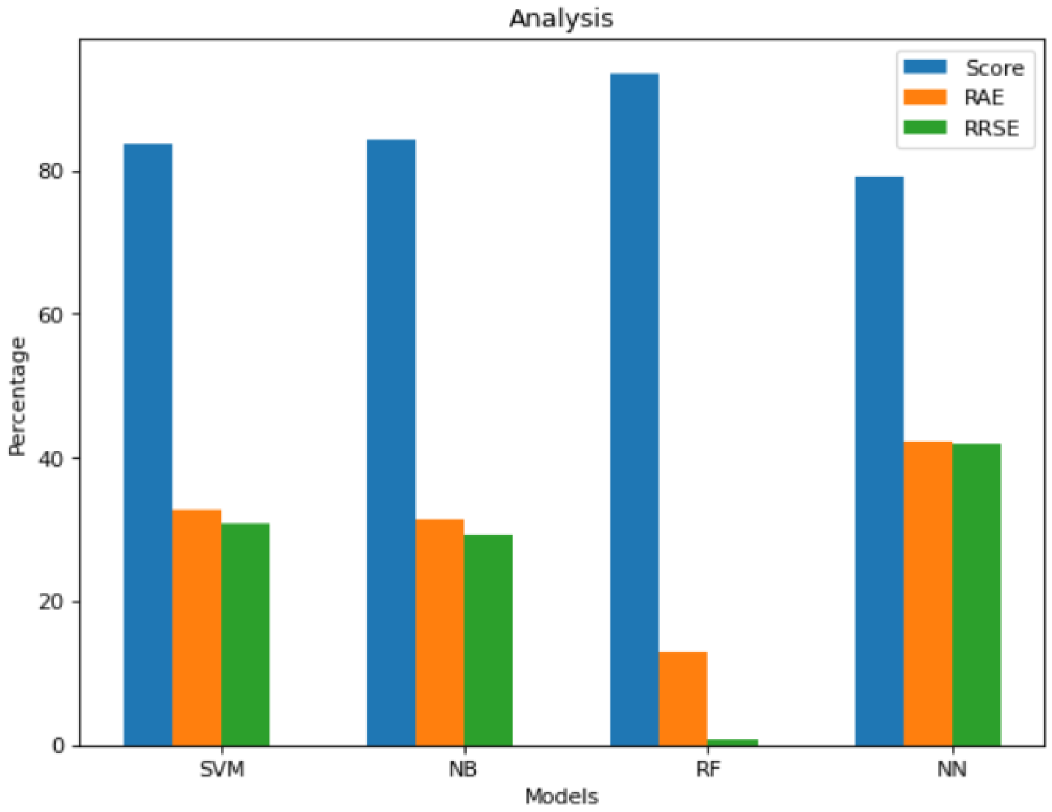


Fig 16: Result Analysis in my work

Results

- Result & Comparison:

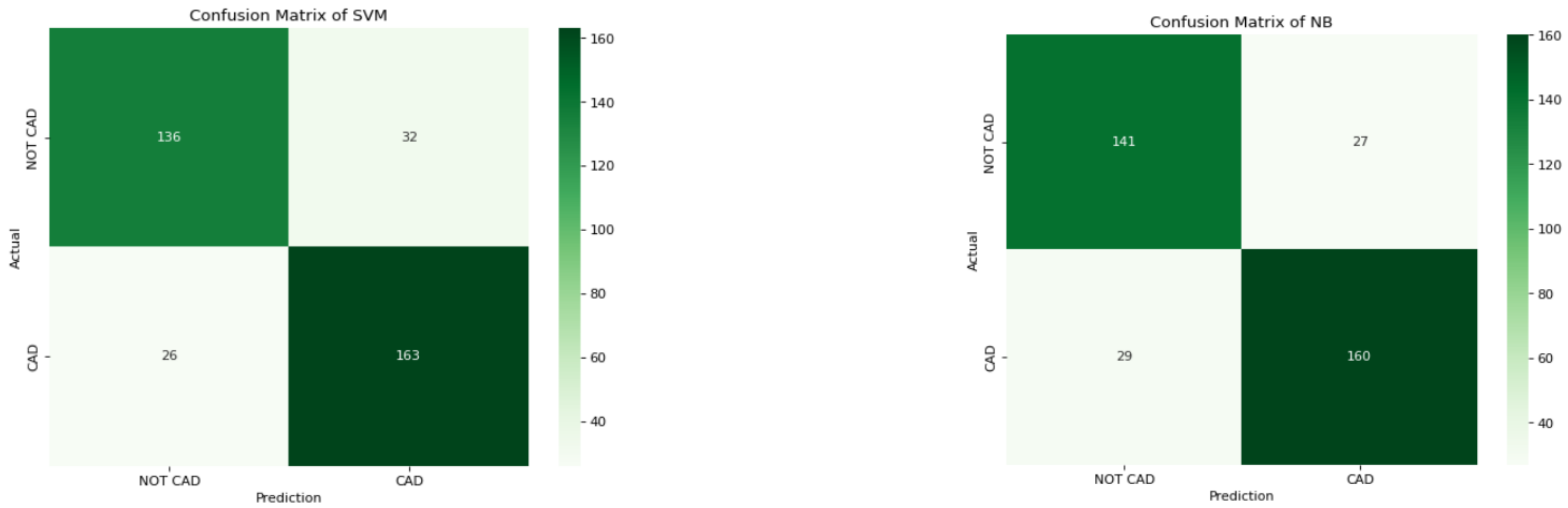
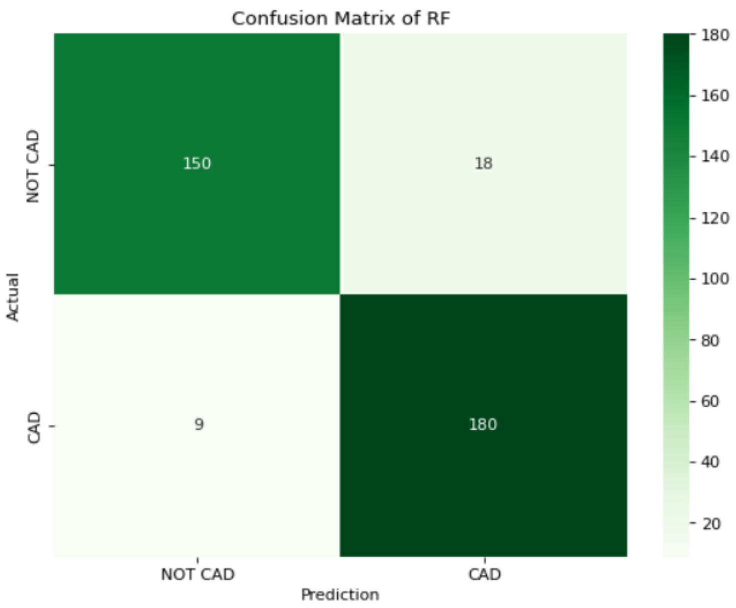


Fig 17: Confusion Matrix of SVM,NB

Results

➤ **Result & Comparison:**



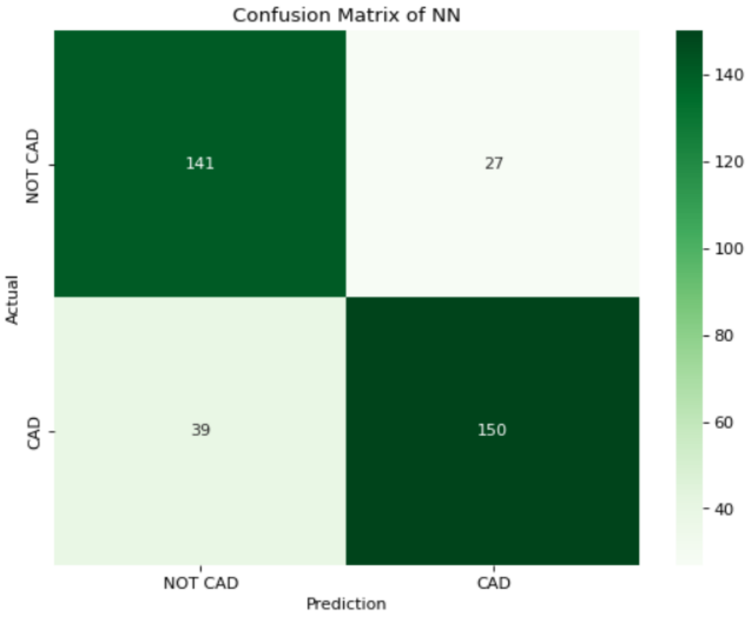


Fig 18: Confusion Matrix of RF & NN

Conclusion

- Four alternative Classification algorithms were used to classify CAD disease.
- A dataset with 1190 instances were used in this process.
- Random Forest Classifier showed the better accuracy.

Future Work

- Performing classification using these classifiers on new dataset with more instances and features
- Combination of other classifier may show better result
- New Dataset with more features on Neural Network may show better accuracy.

References

- Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method -A Comparative Study , Journal of Artificial Intelligence and Capsule Networks (2021) Vol.03/ No.01 Pages: 17-33
- Predicting coronary artery disease: a comparison between two data mining algorithms, Ayatollahi et al. BMC Public Health (2019) 19:448,https://doi.org/10.1186/s12889-019-6721-5
- Coronary Artery Disease Prediction using Data Mining Techniques, Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020] IEEE Xplore Part Number: CFP20M19-ART; ISBN: 978-1-7281-7089-3
- Modele de graphe et modele de langue pour la reconnaissance de scenes visuelles, Trong-Ton Pham

References(cont.)

- Venkataramana Veeramsetty*,* A. Thrishul Kumar*,* B. Navya*,* T. Bhavan*, and* Y. Hrishikesh "Heart disease prediction using machine learning algorithms", AIP Conference Proceedings 2418, 040013 (2022) https://doi.org/10.1063/5.0081782
- dataset link- https://www.kaggle.com/ronitf/heart-disease-uci
- WHO, "Cardiovascular diseases (cvds),available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)."
- https://www.cdc.gov/heartdisease/coronary_ad.htm
- https://gfycat.com/gifs/search/coronary+heart+disease

References(cont.)

- CDC, “Coronary artery disease, available at: https://www.cdc.gov/heartdisease/coronaryad.htm.”
- DeepAi, “Neural network,what is a neural network?, available: https://deepai.org/machine-learning-glossary-and-terms/neural-network.
- Javapoint, “Support vector machine algorithm,available at:https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm.”

Thank You, Any Questions?