Diagnosing Breast Cancer using Julia

Dipta Paul

Student,

Department of Computer Science and Engineering,

United International University.

Dr. Dewan Md. Farid

Associate Professor,

Department of Computer Science and Engineering,

United International University.

Problem Description

The problem is to diagnose the breast cancer.

- Early diagnosing breast cancer can promote timely clinical treatment to patients.
- Correct classification of the benign tumor can prevent patients from undergoing unnecessary treatments.

Thus, diagnosing breast cancer and accurate prediction of malignant and benign tumors is an important.

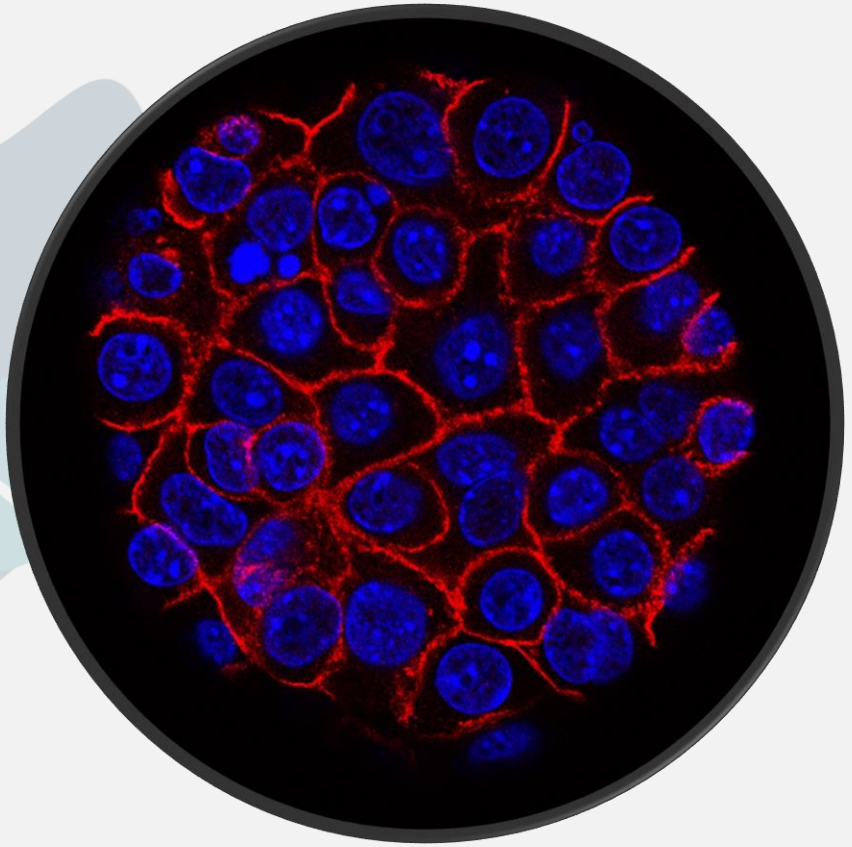
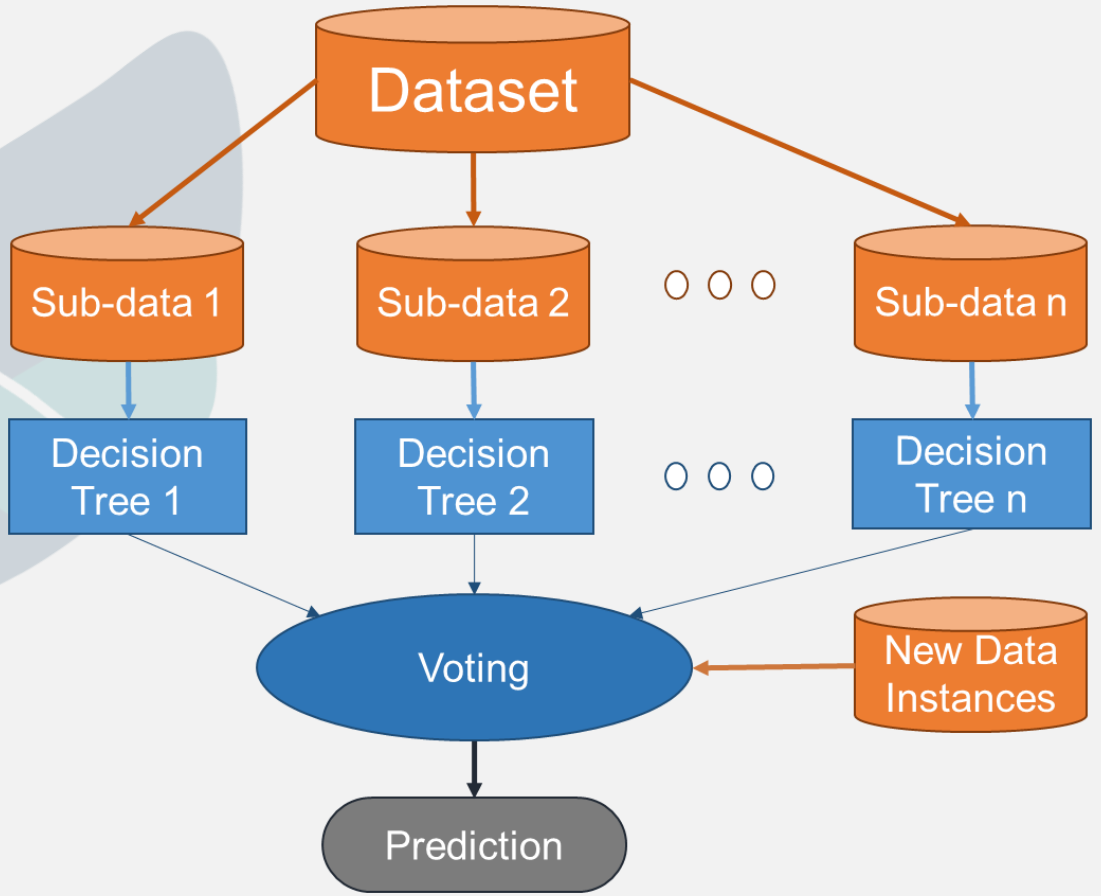


Image Source: USC Norris Comprehensive Cancer Center

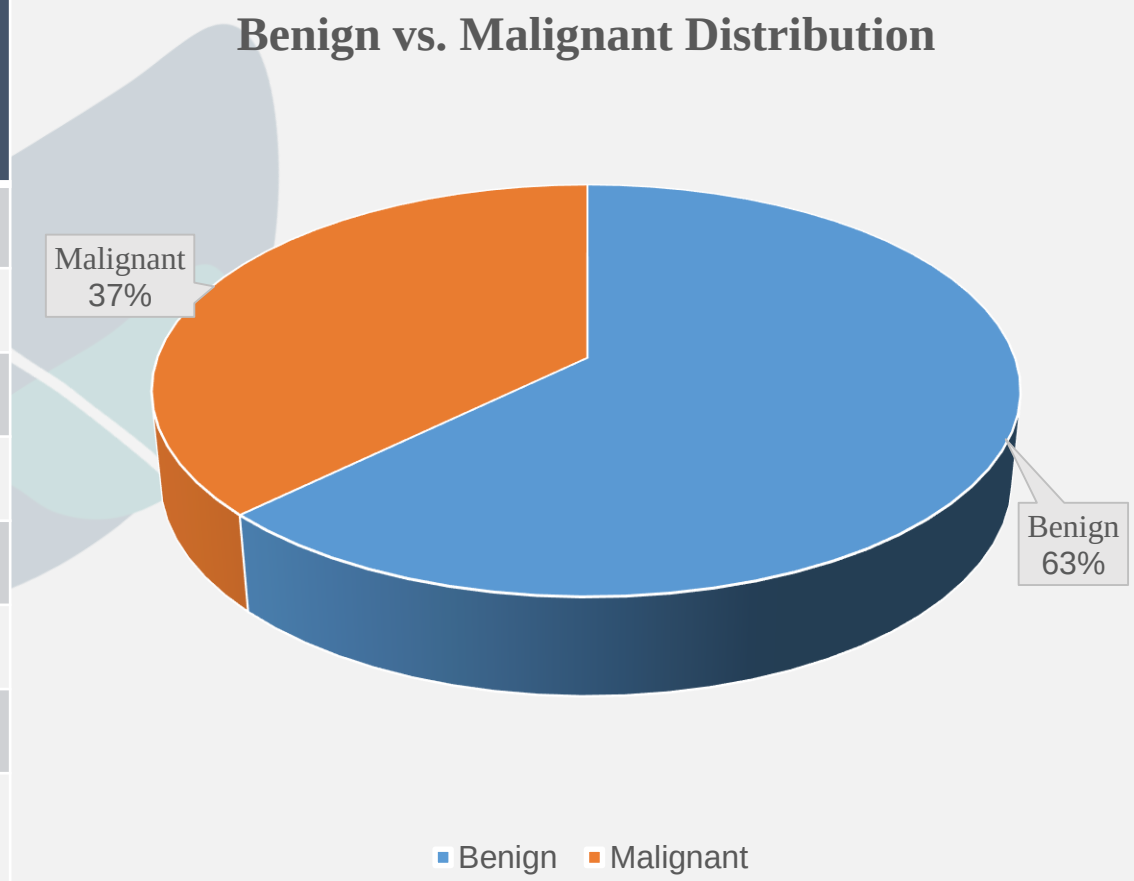
Classification Method

Random Forest Classifier



Dataset Description

Dataset Name	Breast Cancer Wisconsin(Diagnostic) Dataset
Source	Kaggle
Number of Instances	569
Features	31
Feature Type	Numeric and Categorical
Missing Values	No
Class Value	2
Learning Type	Classification
Class Instances	Benign => 357 Malignant => 212



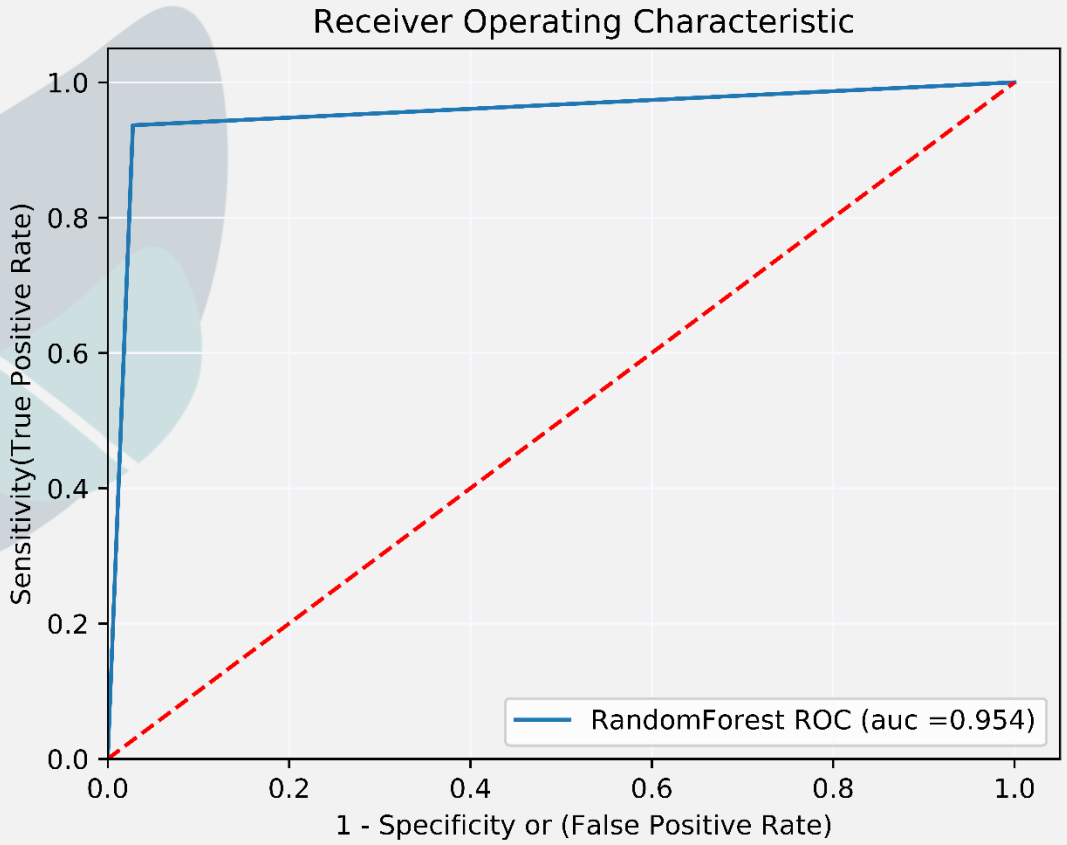
Performance Evaluation Criteria

- **Accuracy** $= \frac{TP+TN}{TP+FP+FN+TN}$
- **TP rate or Recall** $= \frac{TP}{TP+FN}$
- **FP rate** $= \frac{FP}{FP+TN}$
- **TN rate or Specificity** $= \frac{TN}{FP+TN}$

- **FNrate** $= \frac{FN}{FN+TP}$
- **Precision** $= \frac{TP}{TP+FN}$
- **F1_score** $= \frac{2\times TP}{2\times TP+FP+FN}$
- **AUC** $= \frac{TP\text{rate}+\text{TNrate}}{2}$
- **Gmean** $= \sqrt{TPrate + TNrate}$

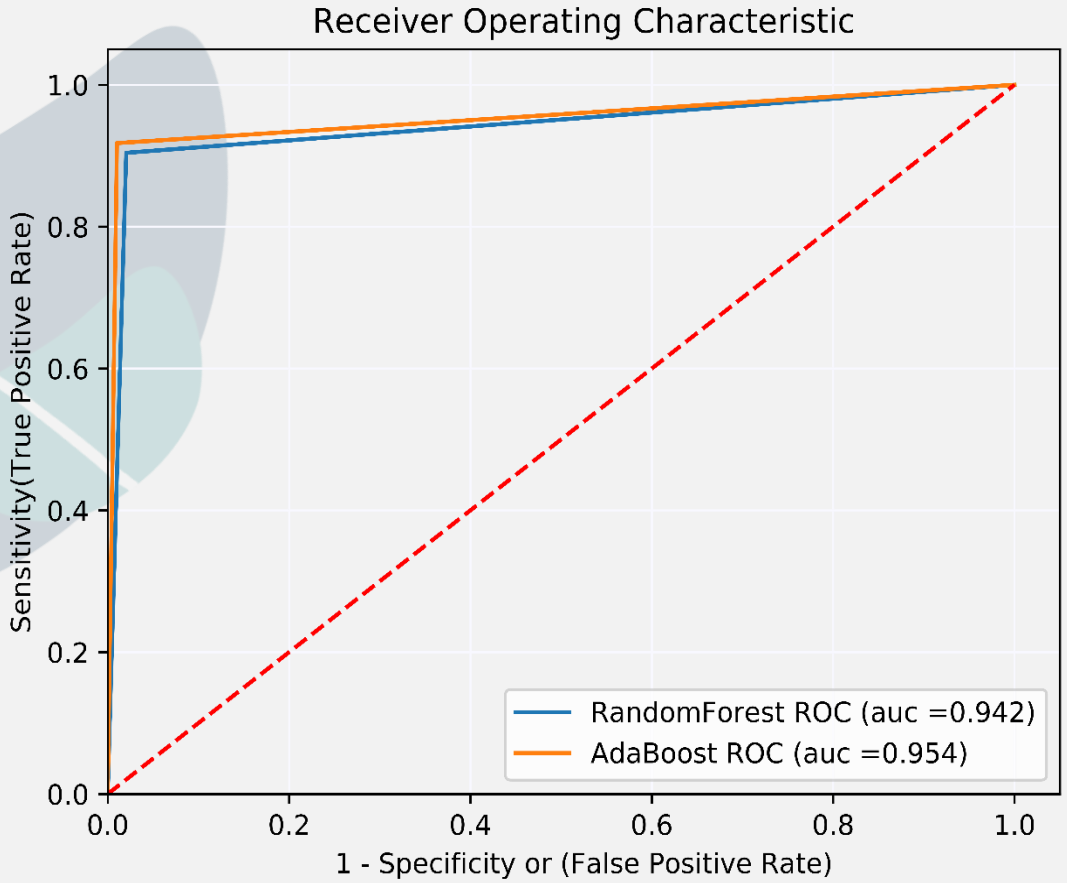
Results

Performance Metric	Score
Accuracy	96.10%
True Positive Rate or Sensitivity or Recall	92.96%
Specificity or True Negative Rate	98.09%
AUC	95.54%
Gmean	95.49%
Precision	96.81%
F1_score	94.83%
False Positive Rate	01.90%



Results

Metrics	Random Forest	AdaBoost
Accuracy	0.961014	**0.964912**
True Positive Rate or Sensitivity or Recall	0.929631	**0.933741**
Specificity or True Negative Rate	0.980988	**0.98439**
AUC	0.95531	**0.959065**
Gmean	0.954923	**0.958665**
Precision	0.968115	**0.973164**
F1_score	0.948361	**0.952829**
False Positive Rate	0.019012	**0.01561**



Julia Packages

- DataFrames
- CSV
- Random
- FreqTables
- PyPlot
- MLLabelUtils
- MLDataUtils
- DecisionTree
- ScikitLearn
- PyCall
- MLBase

Any Question ?

Thank You