

# **Welcome**

**To The Research Final Presentation**

---



## Department of Computer Science and Engineering

### Supervised by:

Md. Mijanur Rahman  
Assistant Professor, Dept. of CSE

### SUBMITTED BY:

**MASHRUFUL HASAN**

ID: 20200000000119

BATCH: 54

**MD SHAHADAT HOSSAIN**

ID: 20200000000117

BATCH: 54

**ORNAB BISWASS**

ID: 20200000000095

BATCH: 54



# **Research Topic**

**Breast Cancer Classification Using Machine  
Learning Algorithms**

# Contents

- ❖ **Introduction**
- ❖ **Objectives**
- ❖ **Related Works**
- ❖ **Methodology**
- ❖ **Result and Discussion**
- ❖ **Conclusion**
- ❖ **Reference**

# Introduction

- Breast cancer is the second most common cancer after lung cancer and one of the main causes of women's death worldwide.
- Difficulty in identifying breast cancer at early stage
- Two types of breast cancer : **1.malignant and 2. Benign**
- Aid of classification algorithms
- Algorithms applied:
  - *K-Nearest Neighbor*
  - *Decision Tree*
  - *Random Forest Classifier*
  - *Naive Bayes Classifier*
  - *Logistic Regression*
- Algorithms applied on the WDBC dataset.

# Objectives

- **Early prediction can slow down the progress of the disease and reduce the mortality rate through appropriate therapeutic interventions at the right time.**
- Develop a classification algorithm that can effectively predict and classify the breast cancer.
- Find out the highest accuracy.
- Calculate the performance of classification algorithms.
- Compare the effectiveness.
- Produce reliable results.

# Related Works

Ref. No.	Name And Period	Datasets	Algorithms	Accuracy
[1]	Rasool et al, 2022	WDBC and BCCD	<b>SVM</b> , LR, KNN, Ensemble Classifier	<b>99.3%</b> , 98.06%, 97.35%, 97.61%
[2]	Roy et al, 2021	WBC and WDBC	<b>LR</b> ,KNN,SVM,NB,DT, RF	<b>98.245%</b> ,97.368%,98.24 5%,92.105%,97.368%, 97.368%
[3]	Roy et al, 2020	WBC and WDBC	NB,SVM,KNN, <b>LR</b>	92%,96%,97%, <b>99%</b>
[4]	Rajaguru et al, 2019	WDBC and WBC	DT, <b>KNN</b>	91.23%, <b>95.61%</b>
[5]	Gupta et al, 2018	WDBC and WBC	<b>MLP</b> ,DT,SVM,KNN	<b>97.7%</b> ,90.9%,93.9%, 94.4%
[6]	Ertel Merouane, 2022	WBC and WDBC	<b>SVM</b> ,KNN,LR,NB	<b>90.6%</b> ,86.1%,80.6%, 51.7%

# Methodology

- ❖ Data Collection
- ❖ Data Pre-processing
- ❖ Algorithm Used

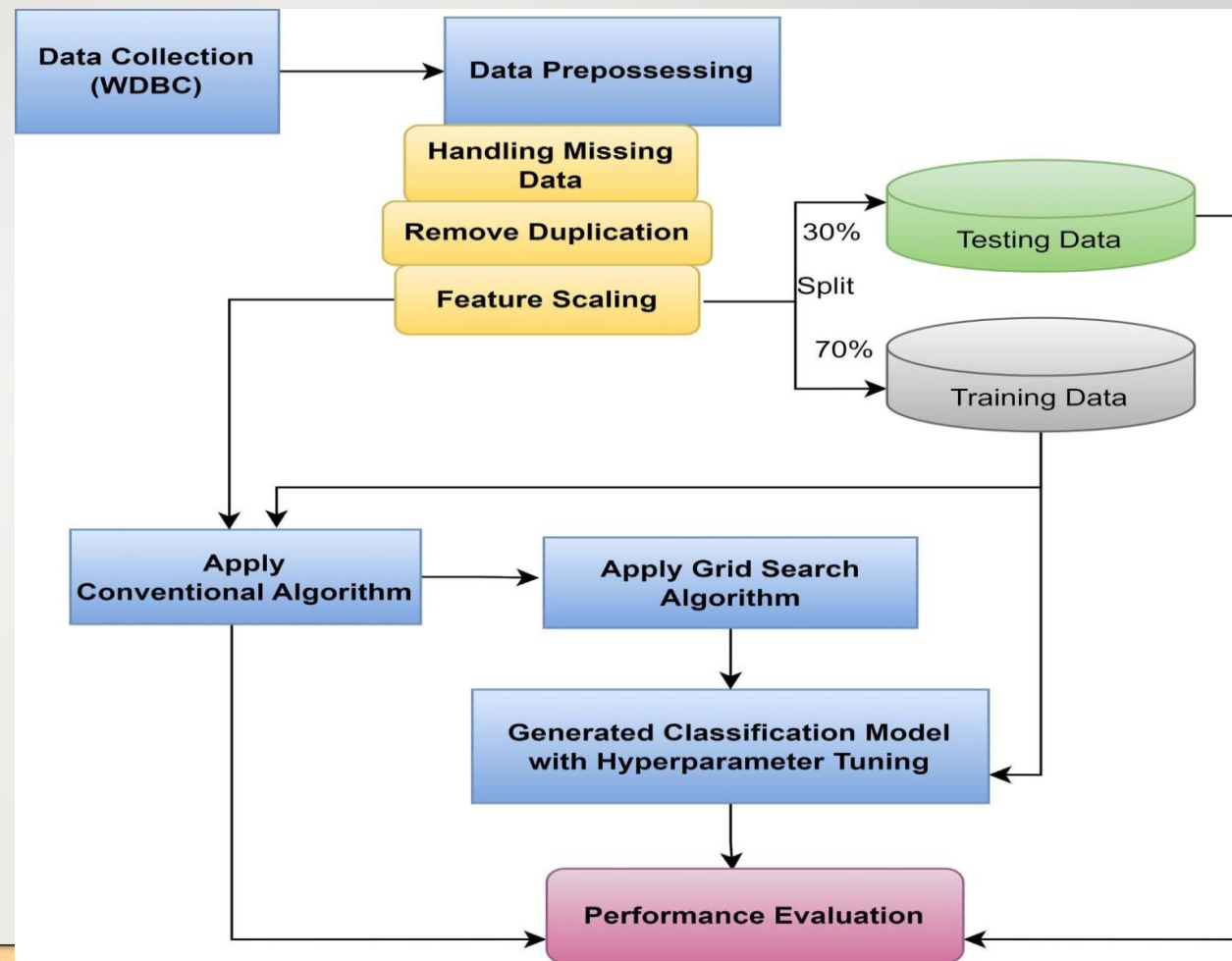


Figure 1. Model For Researched System

# Data Collection

## Wisconsin Breast Cancer Dataset (WDBC)

- ❖ Aggregated from Kaggle
- ❖ 569 instances
- ❖ 32 features.

# Dataset Description

Feature Name	Feature Description
Radius	The average distance between the spots at the circumference's center and edges.
Texture	Grayscale value's SD. Perimeter Gross separation exists between the snake's points.
Perimeter	Gross separation exists at the snake's tip and between.
Area	Total amount of pixels inside the snake, plus one-half of each pixel outside its body.
Smoothness	Measured locally by computing the length difference, the variation in radius length.

# Data Pre-processing

- **First checked the dataset**
- **Extracted unnecessary features**
- **Scaling dataset**

# Algorithm Used

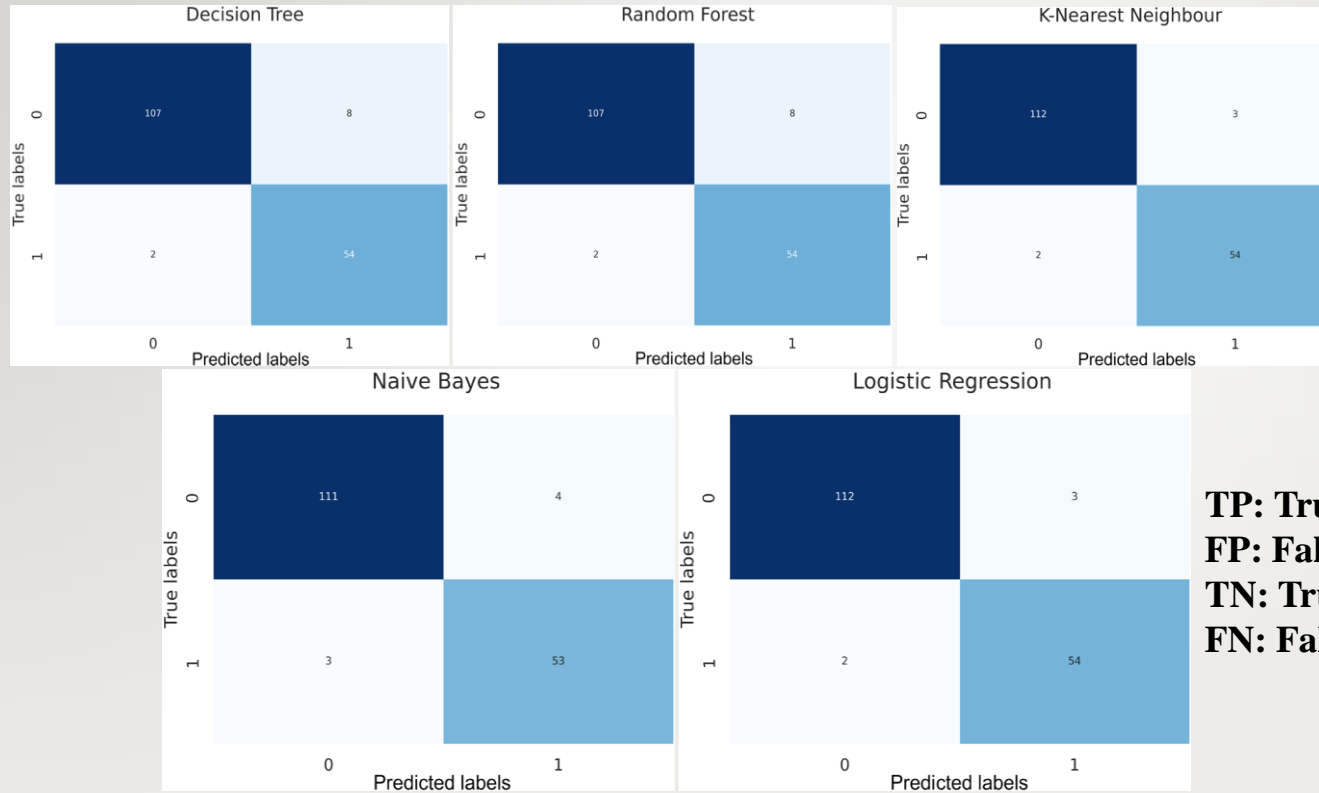
- ❖ *K-Nearest Neighbor*
- ❖ *Decision Tree*
- ❖ *Random Forest Classifier*
- ❖ *Naive Bayes Classifier*
- ❖ *Logistic Regression*

# Experimental Result

## Performance Evaluation Metrix

- ✓ Confusion Metrix
- ✓ AUC & ROC Curve
- ✓ Accuracy
- ✓ Precision
- ✓ Recall
- ✓ F1 Score

# Experimental Result (cont.)



**TP: True Positive (Correctly Identified)**  
**FP: False Positive (Correctly Rejected)**  
**TN: True Negative (Incorrectly Identified)**  
**FN: False Negative (Incorrectly Rejected)**

**Figure 3. Confusion matrix after tuning.**

## Experimental Result (cont.)

Table:3. Performance Evaluation

Performance	Hyperparameter Tuning	DT	RF	KNN	NB	LR
Accuracy	With	94.15%	97.08%	98.83%	95.91%	97.08%
	Without	93.56%	97.08%	96.49%	95.91%	96.49%
Precision	With	95%	97%	99%	96%	97%
	Without	94%	97%	96%	96%	96%
Recall	With	94%	97%	99%	96%	97%
	Without	94%	97%	96%	96%	96%
F1 Score	With	94%	97%	99%	96%	97%
	Without	94%	97%	96%	96%	96%

# Experimental Result (Cont.)

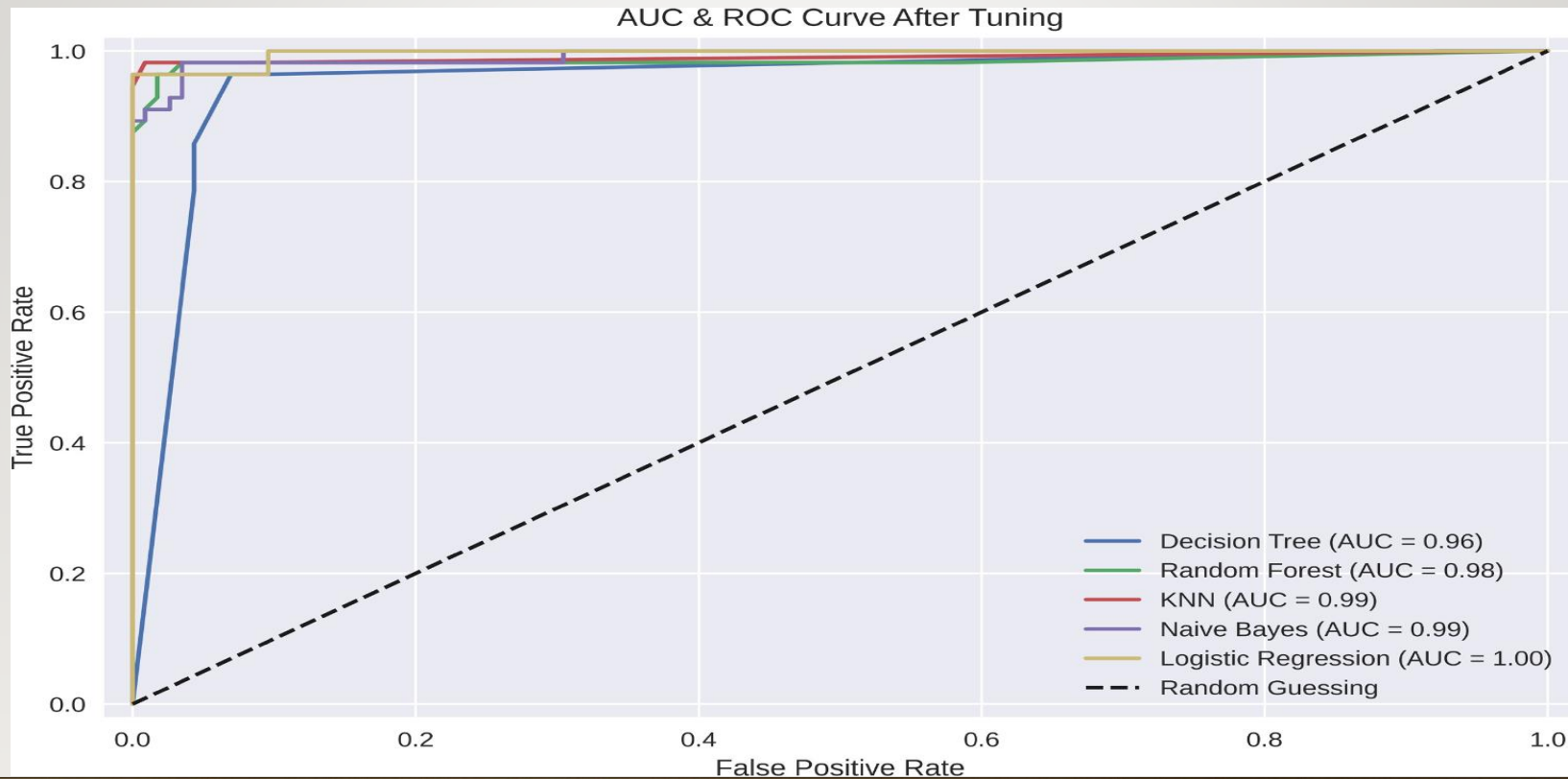
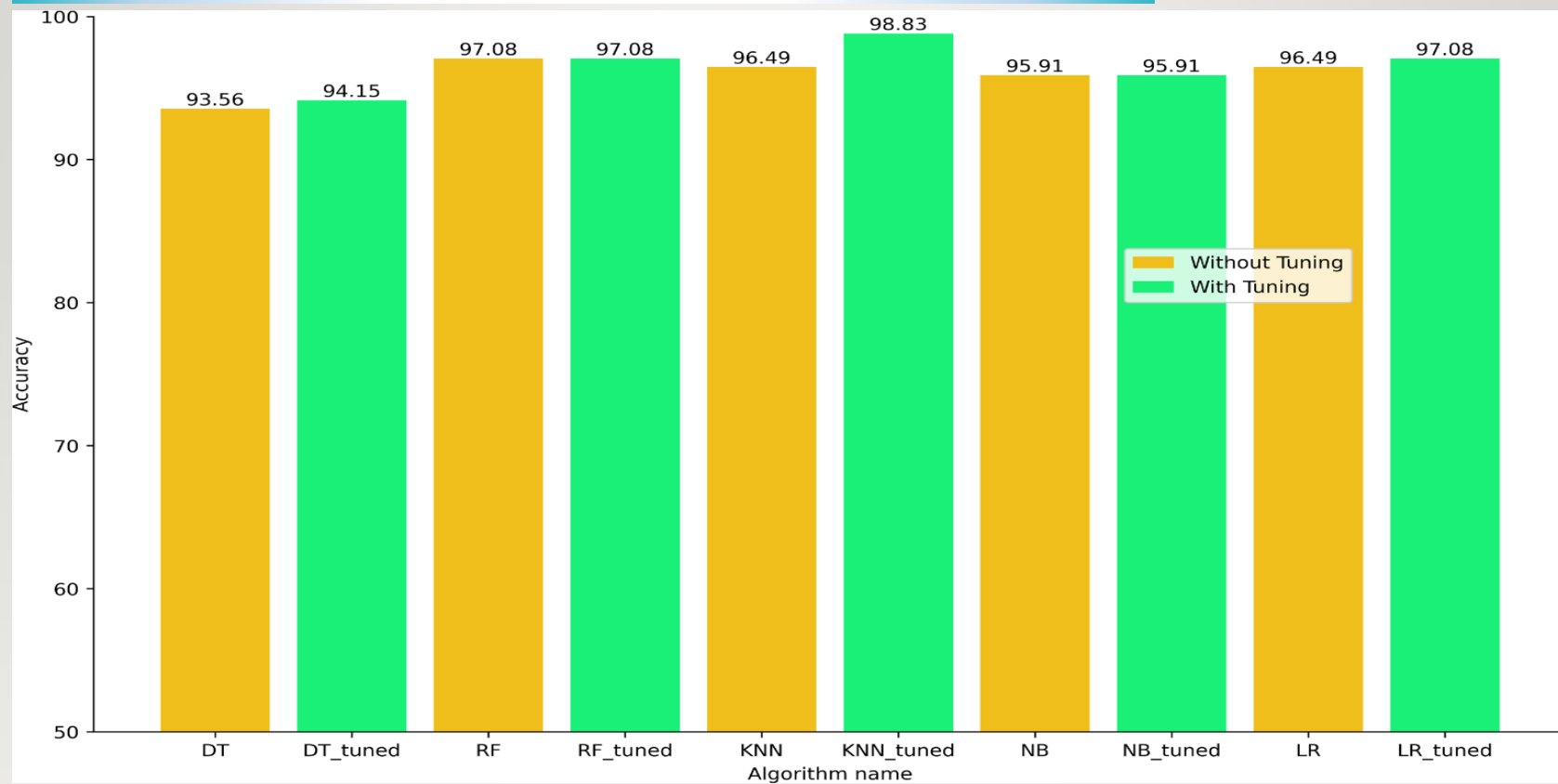


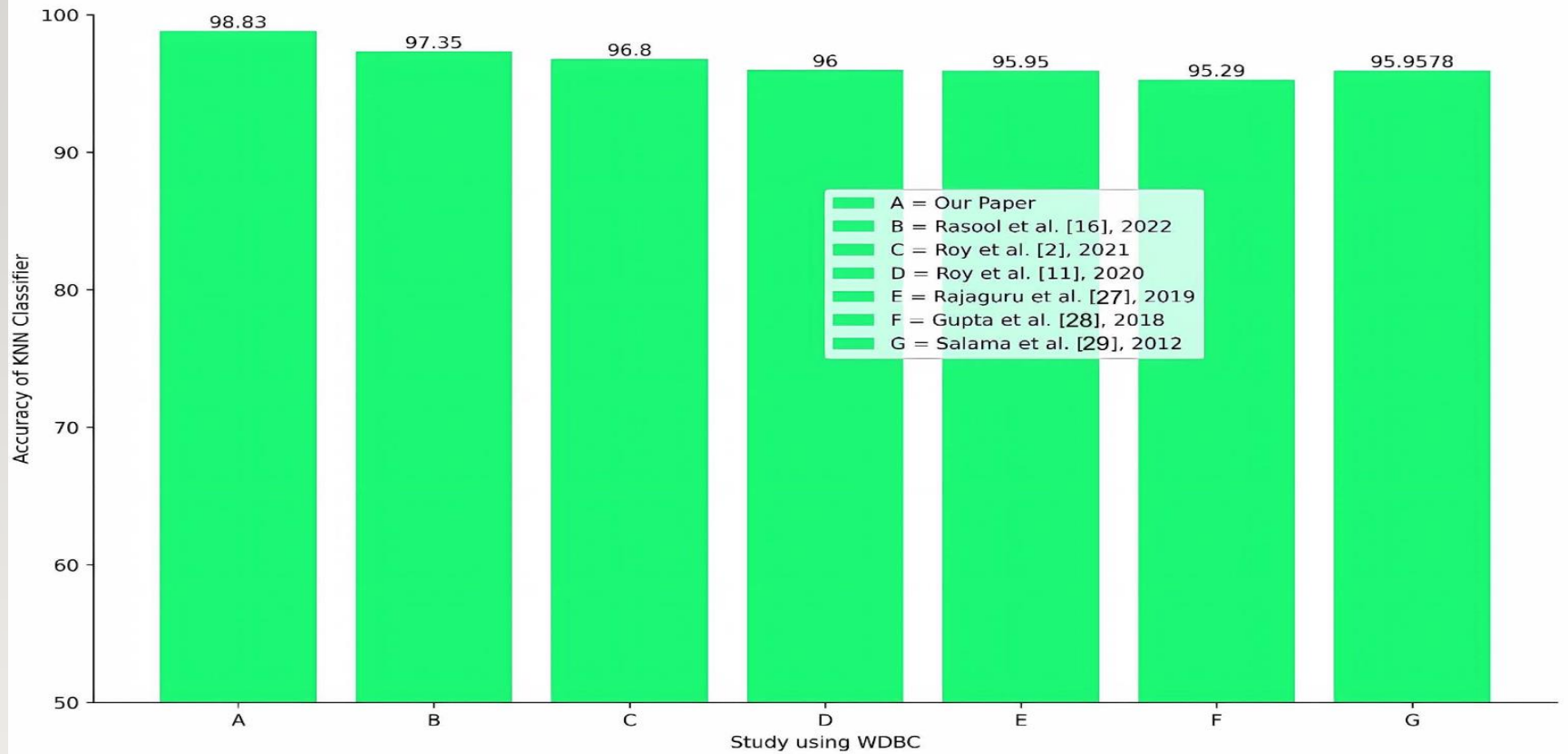
Figure 4. AUC and ROC curve after tuning.

# Experimental Result (cont.)



**Figure 5. Result analysis on accuracy.**

# FINAL OUTPUT (cont.)



**Figure 6.** Result in comparison with existing work.

# CONCLUSION

**This study will help:**

- The healthcare professional
- To predict breast cancer early
- Got 98.83% accuracy in KNN
- Scope for Future work

# References

1. Rasool, A., Bunternghit, C., Tiejian, L., Islam, M.R., Qu, Q. and Jiang, Q. (2022) Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis. International Journal of Environmental Research and Public Health, 19, Article 3211. <https://www.mdpi.com/1660-4601/19/6/3211> <https://doi.org/10.3390/ijerph19063211>
2. Roy, S., Gawande, R., Nawghare, A. and Mistry, S. (2021) Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer. International Journal of Computer Science Trends and Technology (IJCST), 9, 103-111.
3. Roy, B.R., Pal, M., Das, S. and Huq, A. (2020) Comparative Study of Machine Learning Approaches on Diagnosing Breast Cancer for Two Different Dataset. Pro-ceedings of 2020 2nd International Conference on Advanced Information and Com-munication Technology (ICAICT), Dhaka, 28-29 November 2020, 29-34. <https://ieeexplore.ieee.org/abstract/document/9333507> <https://doi.org/10.1109/ICAICT51780.2020.9333507>
4. Rajaguru, H. and Chakravarthy, S.R. (2019) Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer. Asian Pacific Journal of Cancer Prevention, 20, 3777-3781.
5. Gupta, M. and Gupta, B. (2018) A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques. Proceedings of 2018 Second In- ternational Conference on Computing Methodologies and Communication (ICCMC), Erode, 15-16 February 2018, 997-1002. <https://doi.org/10.1109/ICCMC.2018.8487537>
6. Salama, G.I., Abdelhalim, M.B. and Zeid, M.A. (2012) Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. International Journal of Computer and Information Technology, 1, 36-43.
7. Merouane, E. and Said, A. (2022) Prediction of Metastatic Relapse in Breast Cancer using Machine Learning Classifiers. International Journal of Advanced Computer Science and Applications, 13, 176-181. <https://doi.org/10.14569/IJACSA.2022.0130222>

*Thank You*

**Any Question?**