

BREAST CANCER CLASSIFIER



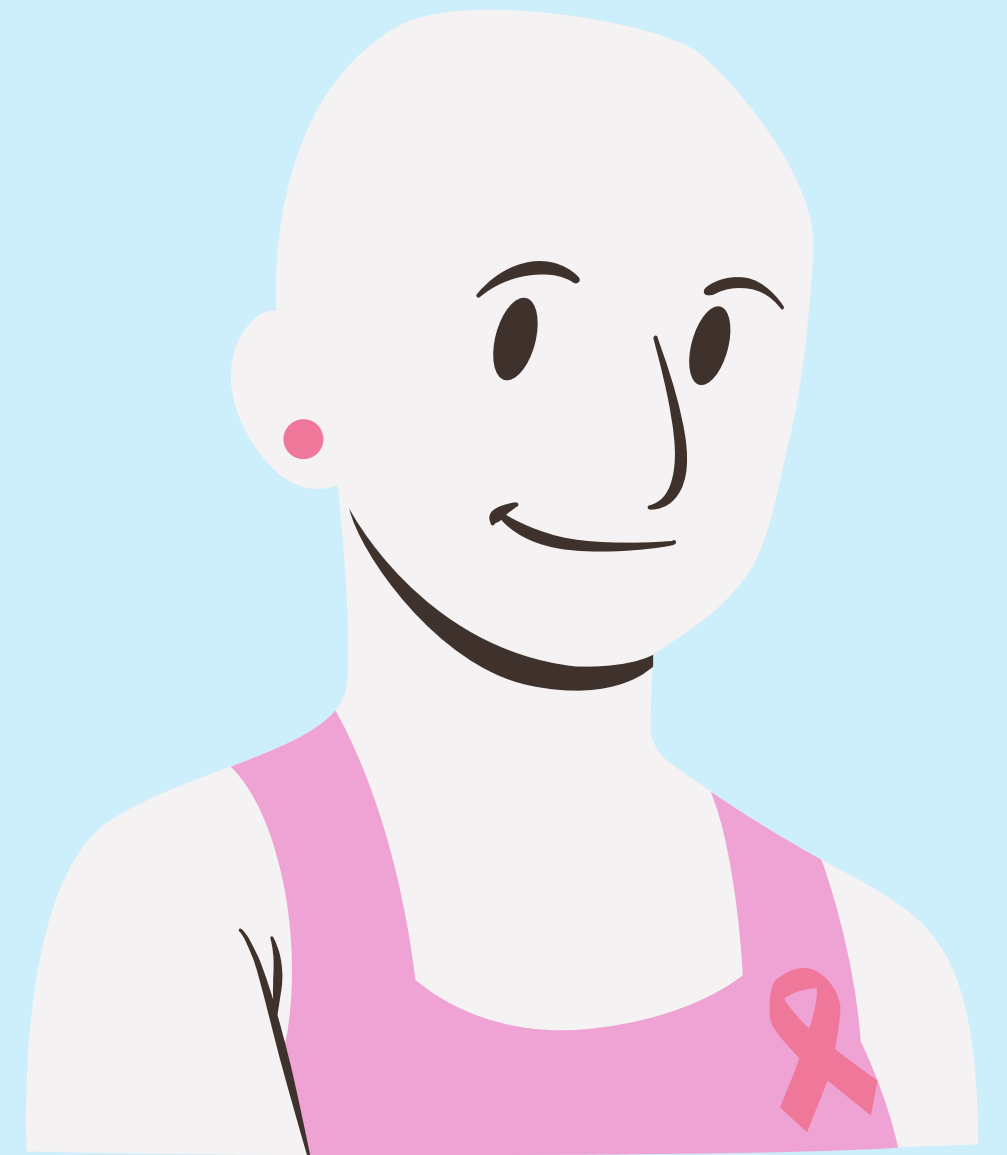
AGENDA

- **Introduction**
- **Problem Statement**
- **Related Work**
- **Proposed Methodology**
- **Results**
- **Conclusions**
- **Team members**



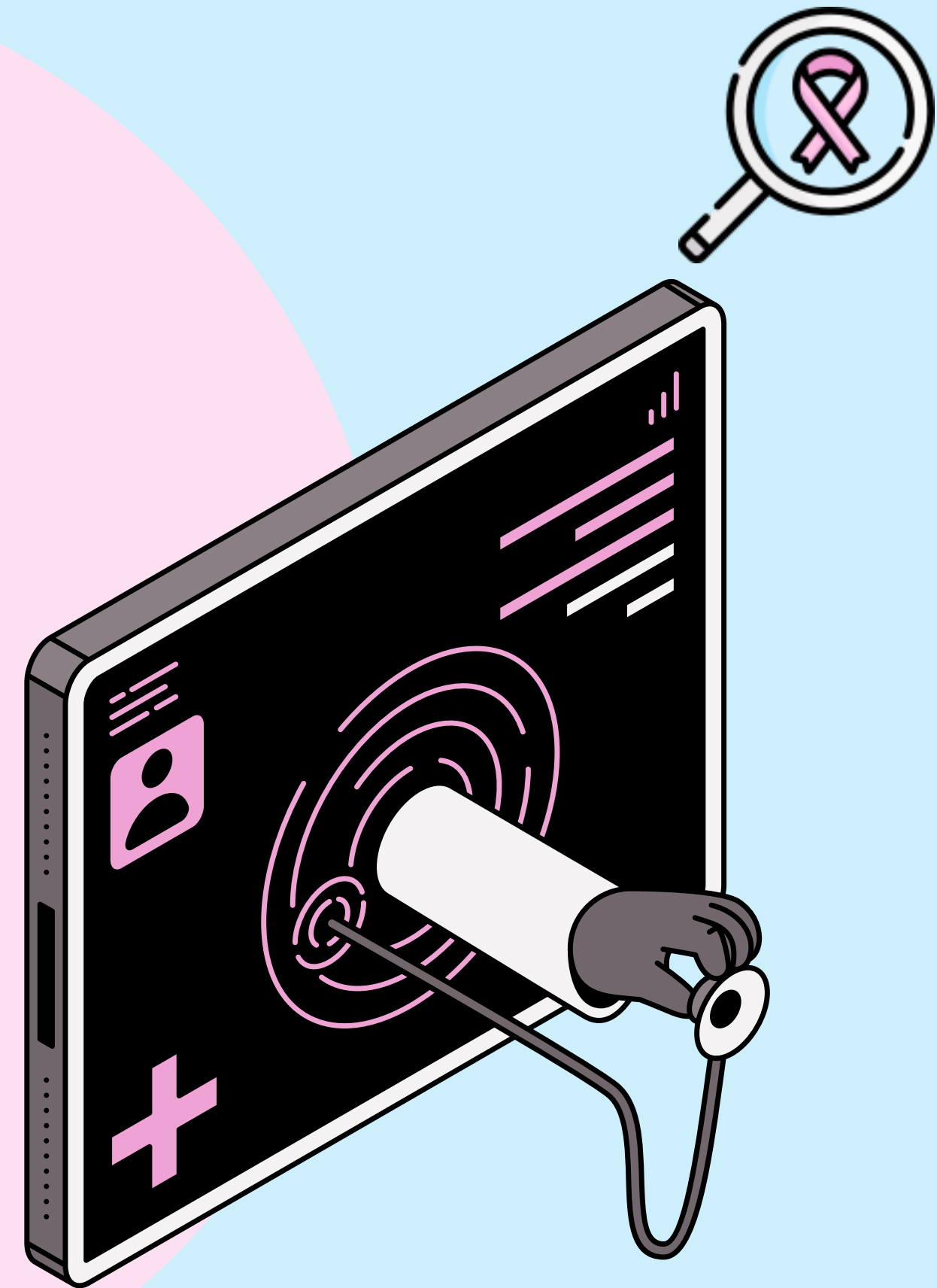
INTRODUCTION

Breast cancer is a type of cancer that begins in the cells of the breast. It is the most common cancer among women globally. While the exact cause is not always clear, factors like genetics, lifestyle, and age can increase the risk. Early detection and prompt treatment are key to improving survival rates, making the role of advanced diagnostic tools even more important.



Problem Statement

Breast cancer is a leading cause of death among women, accounting for a significant number of cancer-related fatalities annually. Early detection significantly improves patient outcomes by enabling timely interventions. However, current diagnostic processes, which heavily rely on manual evaluation of cell characteristics, are prone to subjectivity and human error. This project aims to create a machine learning-based model that can accurately classify tumors as malignant or benign based on their characteristics, thereby assisting medical professionals in making faster and more reliable diagnoses.





Related Work

1. <https://www.mdpi.com/2313-433X/6/6/39>
2. <https://ieeexplore.ieee.org/abstract/document/8391453>
3. <https://www.mdpi.com/2075-4426/11/2/61>



Proposed Methodology



1- Dataset Overview:

- Data comprises 683 records of breast cancer diagnoses with features such as clump thickness, cell size uniformity, and chromatin texture.
- The target variable indicates whether the tumor is malignant (1) or benign (0).



Proposed Methodology



2-Data Preprocessing and Visualization:

- **Visualization:** Representing data in graphs to understand the distribution of values and the relationship of features to each other, such as: histogram, boxplot, and heatmap.
- **Feature Selection:** The Mitoses feature was removed due to its weak effect and relationship with the rest of the features.



Proposed Methodology



3-Model Training:

- **Data Splitting:** The dataset was divided into 80% training data and 20% testing data.
- **Models:**
 1. Logistic regression for binary classification.
 2. Random forest with 300 decision tree.
 3. Naive Bayes of type: GaussianNB.



Proposed Methodology



4-Evaluation Metrics:

The model's performance was evaluated using:

- **Accuracy:** Measures overall correctness of the predictions.
- **Precision:** Focuses on the model's ability to correctly identify malignant cases.
- **Recall:** Ensures no malignant cases are missed.
- **F1 Score:** Balances precision and recall for a comprehensive evaluation.
- **Cross-Validation Score:** Provides an average performance metric by evaluating the model on multiple subsets of the data, ensuring its robustness and generalizability.



Proposed Methodology



5-Prediction:

We tested the three models on data from one patient from each class and all three models made correct predictions.

Results



	test accuracy	cross validation score	f1-score
Logistic regression	98.5%	96.3%	97.9%
Random Forest	97.8%	97.0%	96.9%
Naive Bayes	97.8%	96.3%	96.9%

Conclusions

1-Effectiveness: The model provides a reliable and accurate solution for breast cancer diagnosis.

2-Practicality: By reducing reliance on manual analysis, it speeds up the diagnostic process while maintaining precision.

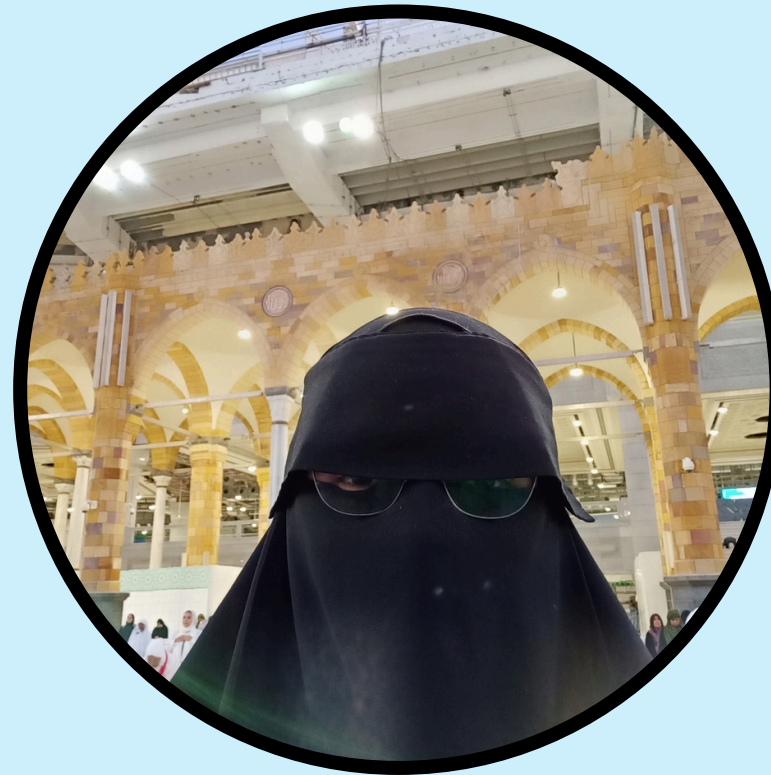
3-Future Directions:

- Incorporate additional datasets from diverse populations to enhance generalizability.
- Explore advanced ensemble techniques or deep learning for further improvements.
- Develop a user-friendly application to integrate the model into clinical workflows.





Team Members



**Safia Nashaat Ali
Mohammed**

Sec 4



**Sara Atef Othman
Mohammed**

Sec 3



Thank you...

