

**A Project Report on**

# Breast Cancer Classification Model

*Submitted by,*

|  |  |
| --- | --- |
| Shreenivas Dudhate | 202201070041 |
| Aditi Kulkarni | 202201070046 |
| Nikita Shelar | 202201070047 |

*Guided by,*

Dr. Smita Kulkarni

# School of Electronics & Telecommunication Engineering MIT Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

# Alandi (D), Pune – 412105 (2024–2025)



**CERTIFICATE**

It is hereby certified that the work which is being presented in the Report entitled **“Breast Cancer Classification Model”**, in partial fulfillment of the requirements for the award of the Bachelor of Technology in Electronics & Telecommunication Engineering and submitted to the **School of Electronics & Telecommunication Engineering** of **MIT Academy of Engineering, Alandi(D), Pune, affiliated to Savitribai Phule Pune University (SPPU), Pune**, is an authentic record of work carried out during Academic Year **2024–2025** Semester **V**, under the supervision of **Dr. Smita Kulkarni**, **School of Electronics & Telecommunication Engineering.**

|  |  |
| --- | --- |
| Shreenivas Dudhate | 202201070041 |
| Aditi Kulkarni | 202201070046 |
| Nikita Shelar | 202201070047 |
|  |  |

# Abstract

The Breast Cancer Classification Model leverages machine learning to classify tumors as malignant or benign. This model aims to assist in early detection, improving patient outcomes and aiding medical professionals in making informed decisions. The solution processes tumor characteristics such as radius, texture, and smoothness, applying machine learning techniques like Logistic Regression, Random Forest, and Support Vector Machine. The model is evaluated using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC. MLOps practices are employed for deployment, scalability, and continuous monitoring.

# Content

|  |  |  |
| --- | --- | --- |
| **Sr.no** | **Topic** | **Page** |
| 1. | Overview | 5 |
| 2. | Data Preprocessing | 6 |
| 3. | Model Building and Evaluation | 11 |
| 4. | MLOps Flow for Deployment | 13 |
| 5. | Tool Framework Utilized | 14 |
| 6. | Conclusion | 15 |

## Overview

### Problem Statement:

Traditional methods for detecting breast cancer can be time-consuming and may lack accuracy. This project aims to build a machine learning model to classify breast cancer as malignant or benign based on tumor characteristics, assisting in faster and more reliable diagnosis.

### Dataset Description:

**Source :**

[**https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset**](https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset)

**Features**

Tumor characteristics (e.g., radius, texture, smoothness)

**Target**

Malignant or benign classification

### Objective

To develop a robust model using machine learning techniques to:

* Analyze and preprocess tumor-related data.
* Build and evaluate classification models.
* Deploy the best-performing model for real-time use.

### Data Preprocessing

* **Data Cleaning:** Handled missing values and ensured consistency in the dataset.
* **Feature Scaling:** Scaled numerical features for balanced model performance.
* **Encoding:** Converted categorical variables (if any) into a numerical format for compatibility with ML algorithms.
* **Exploratory Data Analysis (EDA):** Conducted visualizations and correlation analysis to understand feature relationships and insights.

### Model Building and Evaluation

**Algorithms Implemented:**

* Logistic Regression
* Random Forest
* Support Vector Machine (SVM)

**Performance Metrics:**

* **Accuracy:** Overall correctness of the model.
* **Precision & Recall:** Measures for class-specific performance.
* **F1-Score:** Balanced measure between Precision and Recall.
* **ROC-AUC:** Area under the ROC curve, indicating model discrimination power.

**Model Performance Summary:**

* **Logistic Regression:** Performed well for linear relationships.
* **Random Forest:** Captured complex patterns, showing high performance in the classification task.
* **Support Vector Machine:** Showed high accuracy but required careful tuning for optimal performance.

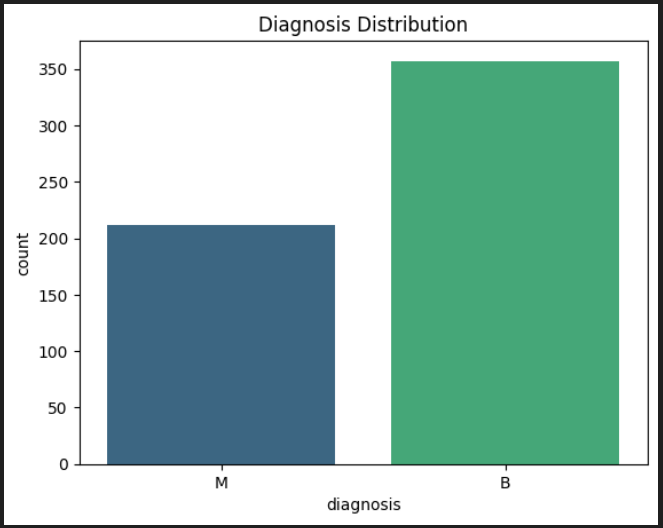
### MLOps Workflow for Deployment

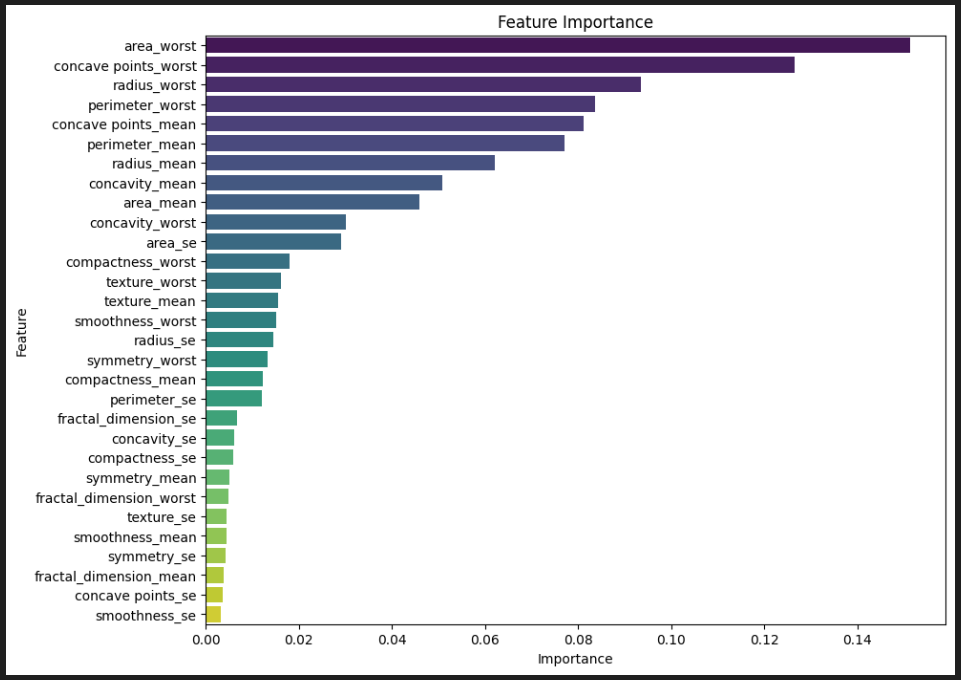
1. **Data Preprocessing:** Cleaned and scaled the dataset for consistency.
2. **Model Training:** Trained the model using selected algorithms (Logistic Regression, Random Forest, SVM).
3. **Model Evaluation:** Identified the best-performing model based on evaluation metrics.
4. **Deployment:** Used Docker/Kubernetes for scalable model deployment, enabling API access for real-time predictions.
5. **Model Monitoring:** Monitored model performance, tracking concept drift and logging for continuous improvement.
6. **Continuous Integration & Delivery (CI/CD):** Automated retraining and deployment for consistent performance.

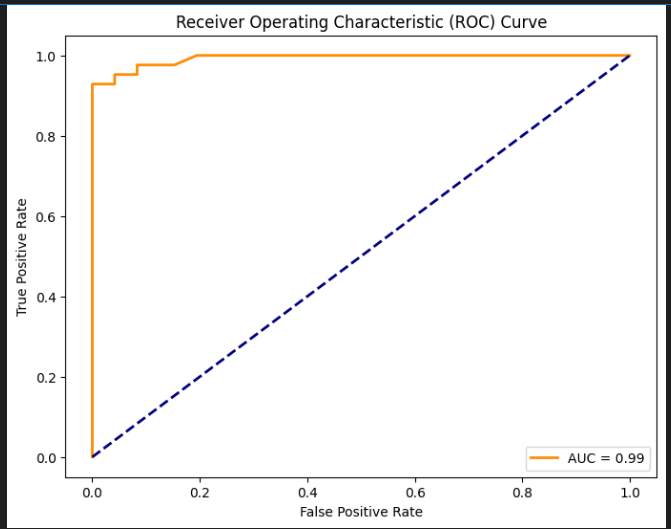
### 5. Tool Framework Utilized

* **Programming Language:** Python
* **Libraries & Tools:**
  + **Pandas & NumPy:** Data manipulation and numerical computations
  + **Matplotlib & Seaborn:** Data visualization
  + **Scikit-learn:** Model building and evaluation
  + **Docker/Kubernetes:** Deployment and scalability
  + **Git & Jenkins (for CI/CD):** Version control and continuous deployment

**Result**

****



**

### 6. Conclusion

The Breast Cancer Classification project demonstrates the effectiveness of machine learning in early diagnosis of breast cancer. The Random Forest model showed the best performance, balancing accuracy and robustness. MLOps practices enabled scalable deployment, while CI/CD ensured consistent updates and reliability.

**Future Enhancements:**

* **Advanced Models:** Experiment with ensemble methods for improved performance.
* **Larger Datasets:** Integrate more extensive datasets to enhance generalization.
* **New Features:** Incorporate additional tumor-related features for better predictions.