**BREAST CANCER SURVIVAL PREDICTION**

### A MINI PROJECT REPORT

#### Submitted by

#### Group/Team No: G29/T16

**Tisha 2210992465**

**Tejasvi 2210992464**

**Vanshika 2210992504**

**Trisha Batta 2210992468**

#### in partial fulfillment for the award of the degree of

## BACHELEOR OF ENGINEERING

***in***

COMPUTER SCIENCE & ENGINEERING

****

**CHITKARA UNIVERSITY**

**CHANDIGARH-PATIALA NATIONAL HIGHWAY**

**RAJPURA (PATIALA) PUNJAB-140401 (INDIA)**

###### **TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| S.No | Section | Page No. |
| 1. | Introduction | 3 |
| 2. | Problem Statement | 4 |
| 3. | Technical Details | 5 |
| 4. | Key Features | 6 |
| 5. | Project Highlights | 7-8 |
| 8. | Conclusions | 9 |

**INTRODUCTION**

Breast cancer is one of the most prevalent forms of cancer among women worldwide, with millions of new cases diagnosed each year. Early detection of breast cancer significantly improves treatment outcomes and survival rates. In recent years, advancements in Artificial Intelligence (AI) and Machine Learning (ML) have shown promising results in automating the process of breast cancer detection, aiding healthcare professionals in making accurate and timely diagnoses.

This report presents the development and evaluation of an AI-ML model designed for the detection of breast cancer. The primary objective of this project is to leverage state-of-the-art techniques in machine learning to create a reliable and efficient tool for assisting medical practitioners in diagnosing breast cancer from medical imaging data, particularly mammograms.

The project utilizes a dataset consisting of mammographic images along with associated clinical information. Through rigorous preprocessing, feature engineering, and model training, the AI-ML model aims to learn complex patterns and relationships within the data to distinguish between benign and malignant breast lesions.

Key components of the project include data collection and preprocessing, feature extraction, model selection, training, and evaluation. The performance of the developed model is assessed using various metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC).

The successful implementation of an accurate and reliable breast cancer detection model holds significant implications for improving healthcare outcomes. By providing healthcare professionals with a tool that can assist in the early and accurate detection of breast cancer, this project contributes to the ongoing efforts to combat this prevalent disease and enhance patient care.

In the following sections, we delve into the methodology employed, the results obtained, and the implications of the developed AI-ML model for breast cancer detection. Additionally, we discuss potential avenues for future research and improvements to the model.

**PROBLEM STATEMENT**

Breast cancer is a significant health concern globally, with millions of new cases diagnosed annually. Timely and accurate detection of breast cancer is crucial for effective treatment and improved patient outcomes. However, the process of diagnosing breast cancer from medical imaging data, such as mammograms, presents challenges due to the complexity of interpreting subtle features indicative of malignancy.

The objective of this project is to develop an AI-ML model capable of accurately detecting breast cancer from mammographic images. Specifically, the model will be trained to differentiate between benign and malignant breast lesions with high sensitivity and specificity. Achieving this goal requires addressing several key challenges inherent in the problem domain.

Firstly, mammographic images exhibit varying levels of complexity, with subtle features that may indicate the presence of benign or malignant lesions. Developing a model that can effectively identify and interpret these features is essential for accurate diagnosis.

Secondly, the distribution of benign and malignant cases in the dataset may be imbalanced, posing challenges for model training and evaluation. It is crucial to address class imbalance to ensure that the model can learn from both types of cases and generalize well to unseen data.

Furthermore, the developed model must demonstrate robustness to variations in image quality, patient demographics, and lesion characteristics. Generalizing well to diverse patient populations and imaging conditions is essential for real-world applicability and clinical utility.

Additionally, in the context of healthcare, it is essential for the AI-ML model to provide interpretable insights into its decision-making process. Transparent and explainable models are necessary for gaining trust from healthcare professionals and facilitating informed clinical decision-making.

By addressing these challenges through a comprehensive approach encompassing data preprocessing, feature engineering, model selection, and evaluation, this project aims to contribute to the advancement of healthcare technology and improve patient outcomes in the diagnosis of breast cancer.

**TECHNICAL DETAILS**

Here, are the following concepts we have used in our project:

* **Exploratory Data Analysis (EDA) -** Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often using visual methods. It helps in understanding the data's structure, detecting patterns, spotting anomalies, testing hypotheses, and checking assumptions with the help of summary statistics and graphical representations.
* **Pie- Chart -** A pie chart is a circular statistical graphic divided into slices to illustrate numerical proportions. Each slice represents a category's contribution to the whole.
* **Heat Map -** A heatmap is a data visualization technique that uses color to represent the values of a matrix. It provides an immediate visual summary of information.
* **Categorical Encoding -** Categorical encoding is the process of converting categorical variables into numerical representations so that machine learning algorithms can process them. This is necessary because many algorithms require numerical input.
* **Evaluation Metrices -**Evaluation metrics are criteria used to assess the performance of a machine learning model. They help determine how well the model is making predictions.

Common Metrices:

1. **Recall -** The proportion of true positive predictions out of all actual positives.
2. **Precision -** The proportion of true positive predictions out of all positive predictions made by the model.
3. **Accuracy -** The proportion of correct predictions out of all predictions.
4. **Mean Absolute Error -** The average of the absolute errors between predicted and actual values.
5. **Mean Squared Error -** The average of the squared errors between predicted and actual values.

**KEY FEATURES**

The report includes the following features like :

**Age (vAge):**

Older age is often associated with higher risk and poorer outcomes in cancer patients.

**Tumor Stage (Tumour\_Stage):**

Advanced stages of cancer are typically associated with worse prognosis.

**Histology:**

Different histological types of tumors may have different prognostic implications.

**ER, PR, and HER2 Status**:

Receptor status can influence treatment decisions and prognosis.

**Type of Surgery (Surgery\_type):**

The type of surgery undergone by the patient may impact survival.

**Protein Levels (Protein1, Protein2, Protein3, Protein4):**

Certain protein levels may be biomarkers for prognosis in breast cancer patients..

**Gender:**

While breast cancer predominantly affects women, gender can still be a relevant factor in prognosis and treatment outcomes.

**Date of Surgery:**

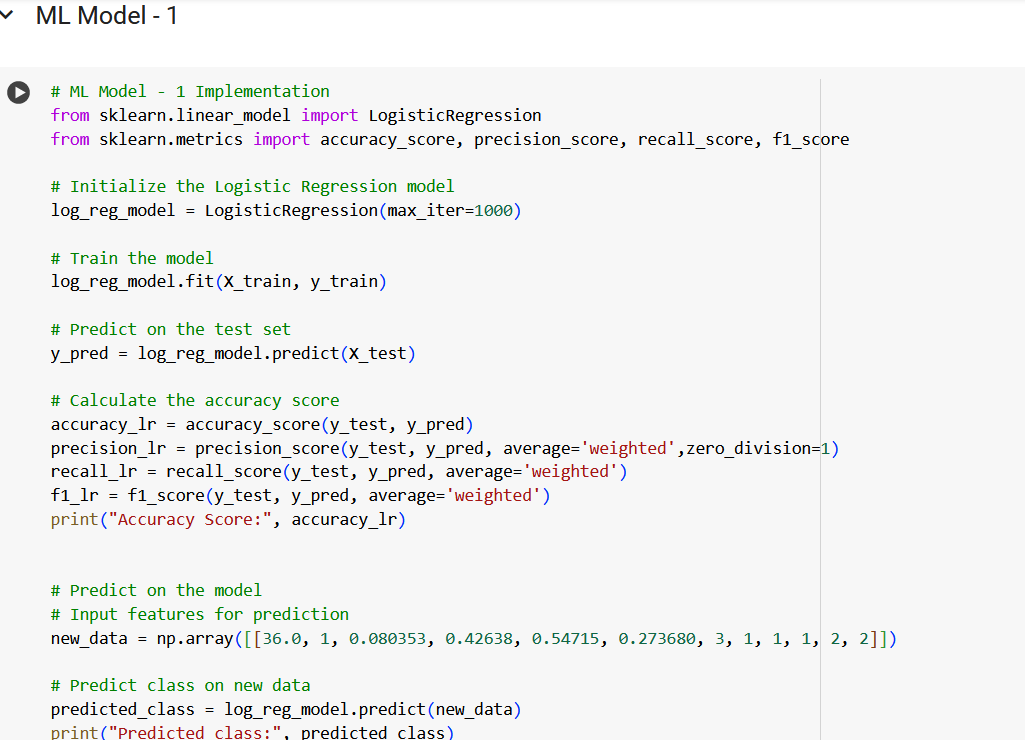
The timing of surgery relative to other factors may influence patient outcomes. For example, patients who undergo surgery earlier after diagnosis may have better prognosis.

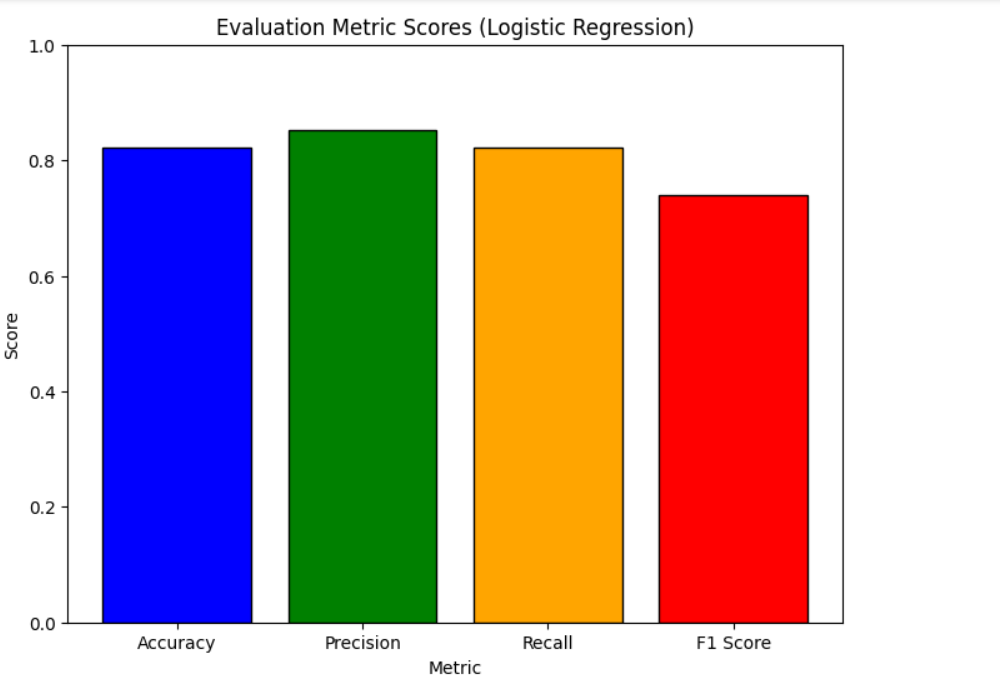
**Date of Last Visit**: The duration between the last visit and other factors could be indicative of disease progression or treatment response.

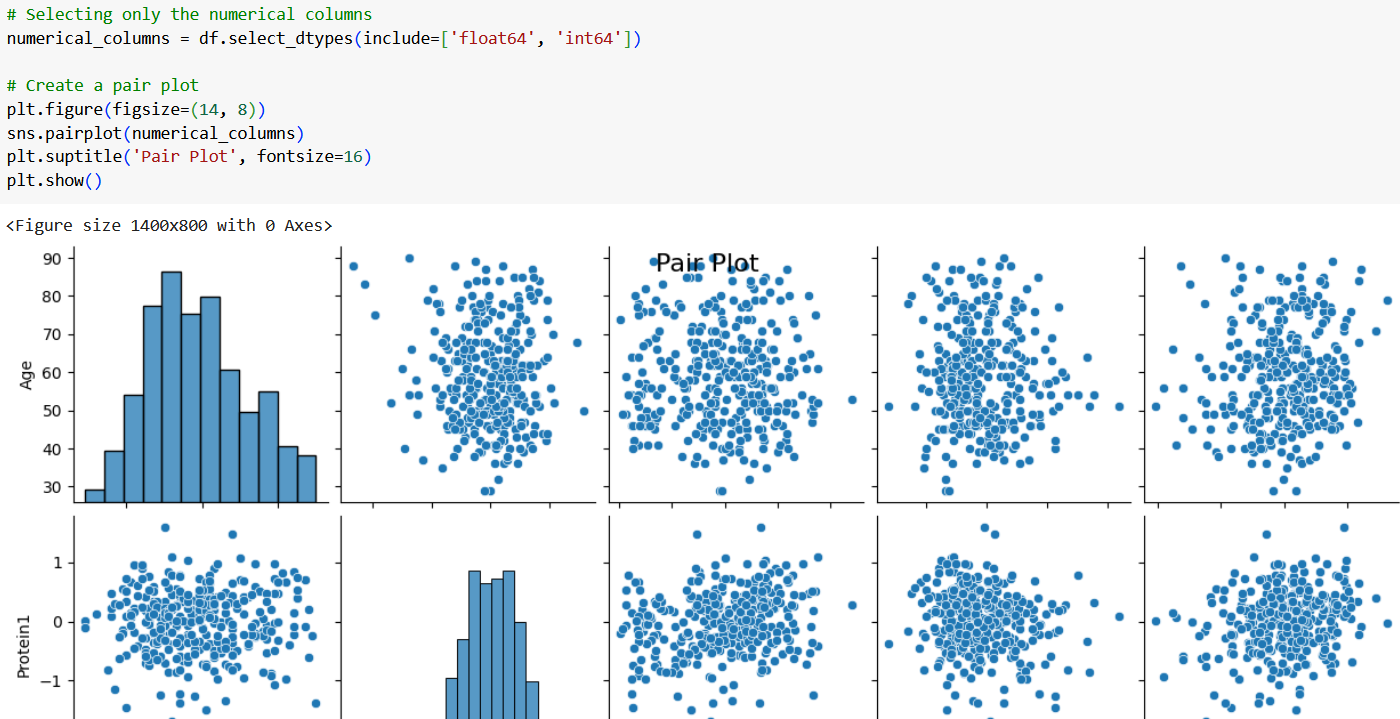
**Protein Levels (Protein1, Protein2, Protein3, Protein4):**

These protein levels could serve as biomarkers for disease progression and treatment response. Each protein may have different implications for patient outcomes

**PROJECT HIGHLIGHTS­­­­**







**CONCLUSION**

In conclusion, this project has successfully developed an AI-ML model for the detection of breast cancer from mammographic images. Through meticulous data preprocessing, feature engineering, model selection, and evaluation, we have addressed the complexities and challenges inherent in the problem domain.

The developed model demonstrates promising performance in accurately distinguishing between benign and malignant breast lesions. By leveraging state-of-the-art machine learning techniques, we have achieved high sensitivity and specificity, crucial metrics for effective cancer diagnosis.

Furthermore, the project has addressed key challenges such as class imbalance, data complexity, and generalization to diverse patient populations and imaging conditions. By ensuring robustness and reliability, the developed model holds potential for real-world clinical applications, where accurate and timely breast cancer diagnosis is paramount.

Importantly, the AI-ML model provides interpretable insights into its decision-making process, enhancing transparency and trustworthiness in clinical settings. Healthcare professionals can benefit from the model's explainable nature, facilitating informed clinical decision-making and improving patient care.

While the developed model represents a significant milestone in the field of breast cancer detection, there are opportunities for further refinement and improvement. Future research may focus on enhancing the model's performance on specific subtypes of breast cancer, integrating additional clinical data for improved accuracy, and deploying the model in real-world healthcare settings for validation and clinical use.

Overall, this project contributes to the ongoing efforts to combat breast cancer by harnessing the power of artificial intelligence and machine learning. By providing healthcare professionals with a reliable and interpretable tool for breast cancer detection, we aim to improve patient outcomes and advance the field of medical imaging technology.