**DEEP LEARNING**

**PROJECT REPORT**



**PROJECT TITLE:** BREAST CANCER PREDICTION

**SUPERVISED BY:** DR SADAF HUSSAIN

**SUBMITTED BY:** MUHAMMAD ABDULLAH

FA-2020/BSCS/**014-(A)**

MUHAMMAD NABEEL AMJAD

FA-2020/BSCS/**221-(A)**

TOUSEEF ABBAS

FA-2020/BSCS/**525-(A)**

**SUBMISSION DATE:** JULY 14TH, 2024

**DEPARTMENT OF COMPUTER SCIENCE**

**LAHORE GARRISON UNIVERSITY**

**Table of Contents**

[1: ABSTRACT](#_Toc157961909) **[Error! Bookmark not defined.](#_Toc157961909)**

[2: INTRODUCTION](#_Toc157961913) **[Error! Bookmark not defined.](#_Toc157961913)**

[3: LITERATURE REVIEW](#_Toc157961914) **[Error! Bookmark not defined.](#_Toc157961914)**

[4: METHODOLOGY](#_Toc157961915) **[Error! Bookmark not defined.](#_Toc157961915)**

[4.1 Dataset Description and Visualization](#_Toc157961916) **[Error! Bookmark not defined.](#_Toc157961916)**

[4.2 Data Cleaning](#_Toc157961917) **[Error! Bookmark not defined.](#_Toc157961917)**

[4.3 Data Augmentation](#_Toc157961918) **[Error! Bookmark not defined.](#_Toc157961918)**

[4.4 Algorithm Selection and Usage](#_Toc157961918) **[Error! Bookmark not defined.](#_Toc157961918)**

[5: RESULTS](#_Toc157961919) **[Error! Bookmark not defined.](#_Toc157961919)**

[5.1 Discussion](#_Toc157961920) **[Error! Bookmark not defined.](#_Toc157961920)**

[5.2 Performance Matrices](#_Toc157961921) **[Error! Bookmark not defined.](#_Toc157961921)**

[5.3 Comparison with Explanation](#_Toc157961922) **[Error! Bookmark not defined.](#_Toc157961922)**

[6: CONCLUSION](#_Toc157961924) **[Error! Bookmark not defined.](#_Toc157961924)**

1. **Abstract**

This document presents a comprehensive exploration of breast cancer prediction using deep learning algorithms. Our study investigates the performance of diverse algorithms, including Convolutional Neural Networks (CNN), Transfer Learning, Logistic Regression, Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Decision Trees.

*Keywords: Deep Learning, Breast Cancer, CNN, Transfer Learning*

1. **Introduction**

With the rising importance of predictive healthcare, this study focuses on breast cancer prediction using deep learning techniques. Utilizing datasets from renowned repositories, our investigation delves into the performance of various algorithms as well as pre-trained models. The introduction outlines the significance of accurate breast cancer prediction, emphasizing its potential impact on early diagnosis and intervention.

1. **Literature Review**

A comprehensive literature review highlights key studies in Breast Cancer Prediction using machine learning and deep learning.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr.** | **Title Author** | **Year** | **Algorithm** | **Dataset** | **Accuracy** |
| **1** | Breast Cancer Detection and Classification using Deep Learning Xception Algorithm | 2022 | Xception Algorithm | Kaggle.com | 99.78% |
| **2** | Convolution Neural Network for Breast Cancer Detection and Classification Using Deep Learning | 2023 | Xception, InceptionV3, VGG16, MobileNet and ResNet50 | Kaggle.com | Xception: 97.54%, InceptionV3: 95.33%, ResNet50:98.14%,  VGG16: 97.67%, MobileNet: 93.98%, BCCNN: 98.28% |
| **3** | Prediction of Breast Cancer using Machine Learning Approaches | 2022 | random forest (RF), neural network (MLP), gradient boosting trees (GBT), and genetic algorithms (GA) | DCIS: Ductal carcinoma in situ, IDC: Invasive ductal carcinoma, ILC: Invasive lobular carcinoma | RF: 80%  MLP: 73%  GBT: 74% |

1. **Methodology**

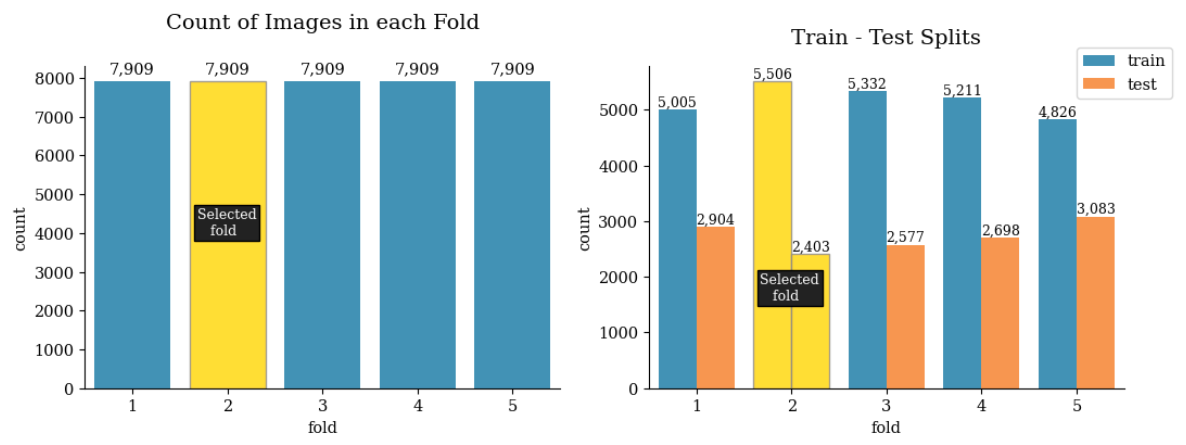
The aim of this system is to predict whether a patient has a malignant or benign case of breast cancer. Here is the methodology employed in our project.

* 1. **Dataset Description and Visualization**

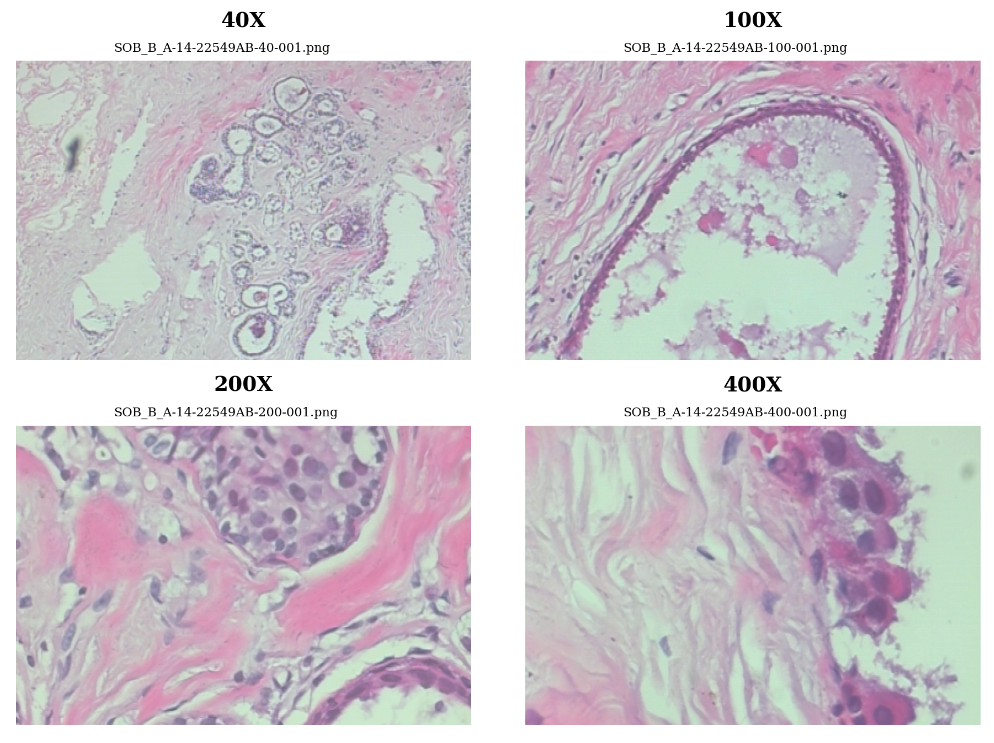
**Link:** <https://www.kaggle.com/datasets/tathagatbanerjee/breakhis-breast-cancer-histopathological/data>

This dataset was created for use in deep learning tasks for both binary and multiclass classification problems. It consists of 7,909 microscopic images of breast tumor tissue, captured using different magnification factors (40X, 100X, 200X, and 400X). The images in this dataset were resized to 224x224 pixels and organized according to binary and multiclass classification tasks. Below is the visualization of the dataset, showing the appearance of malignant and benign tissues in breast cancer.

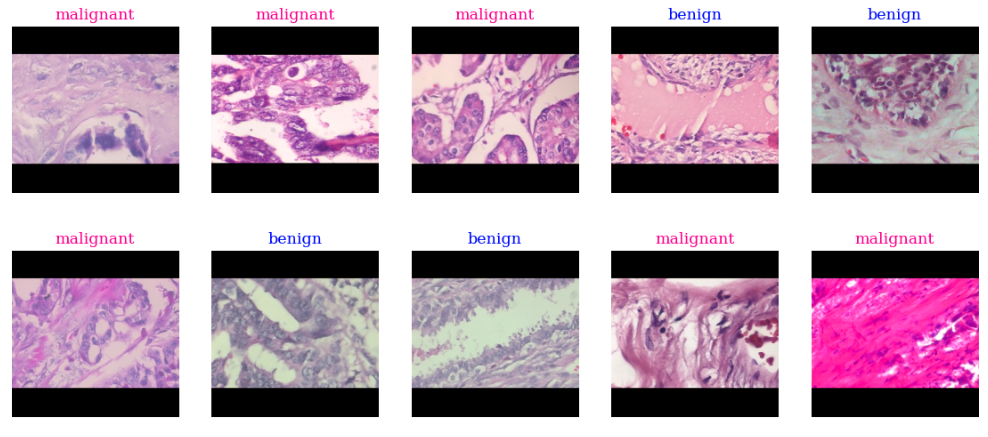
Here is the stats of dataset shown in this graph.



Below is the visualization of the dataset.



Here we can see how the malignant and benign tissues look in breast cancer.

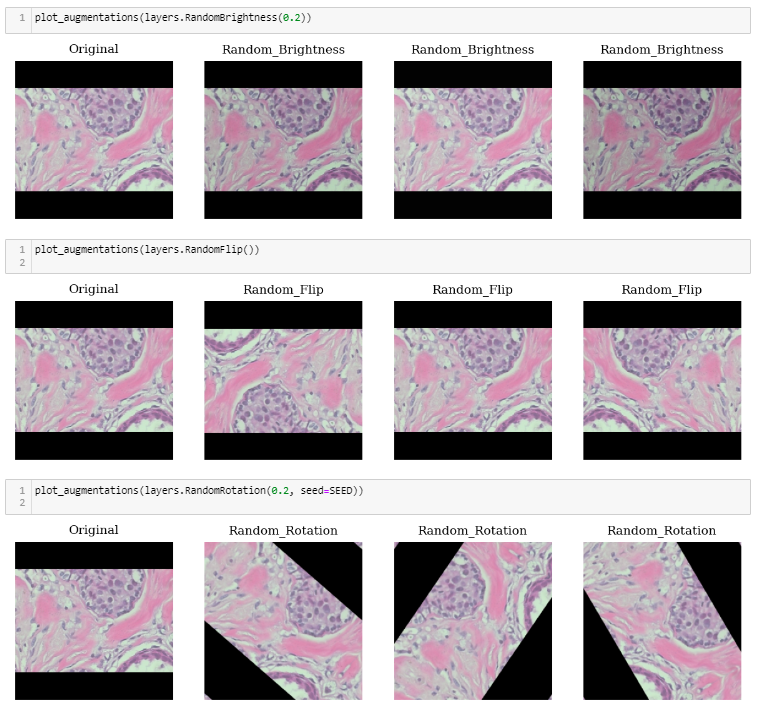


* 1. **Data Cleaning**

Upon examining the dataset, we identified 250 exact duplicate images and 7 nearly duplicate images. To ensure each data point is represented only once, we removed these duplicates to prevent model bias.

* 1. **Data Augmentation**

We applied data augmentation techniques to create variations that can help improve model accuracy. Random brightness, random flip, and random rotation were used for augmentation.



* 1. **Algorithm Selection and Usage**

These are the Algorithms We Selected:

**CNN:**

CNN is the most popular deep-learning technique that has been utilized in several studies for breast cancer detection. The CNN is a deep learning model which represents hierarchical abstraction and consists of various layers which accept features as raw data.

**LSTM:**

Recent research has explored the integration of LSTM network with feature selection techniques to improve breast cancer recurrence prediction. This approach combines the capabilities of deep learning in modelling sequential data with the benefits of feature selection in enhancing model interpretability.

**Transfer Learning:**

A powerful technique in deep learning, transfer learning reuses existing models and their knowledge for new problems, enabling the training of deep neural networks even with limited data.

We utilized the following pre-trained CNN models:

* **MobileNetV3**

MobileNet is a type of convolutional neural network designed for mobile and embedded vision applications. They are based on a streamlined architecture that uses depthwise separable convolutions to build lightweight deep neural networks that can have low latency for mobile and embedded devices.

* **EfficientNetB1**

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient.

* **VGG16**

VGG16 is a convolutional neural network model that's used for image recognition. It's unique in that it has only 16 layers that have weights, as opposed to relying on a large number of hyper-parameters. It's considered one of the best vision model architectures.

* **ResNet50V2**

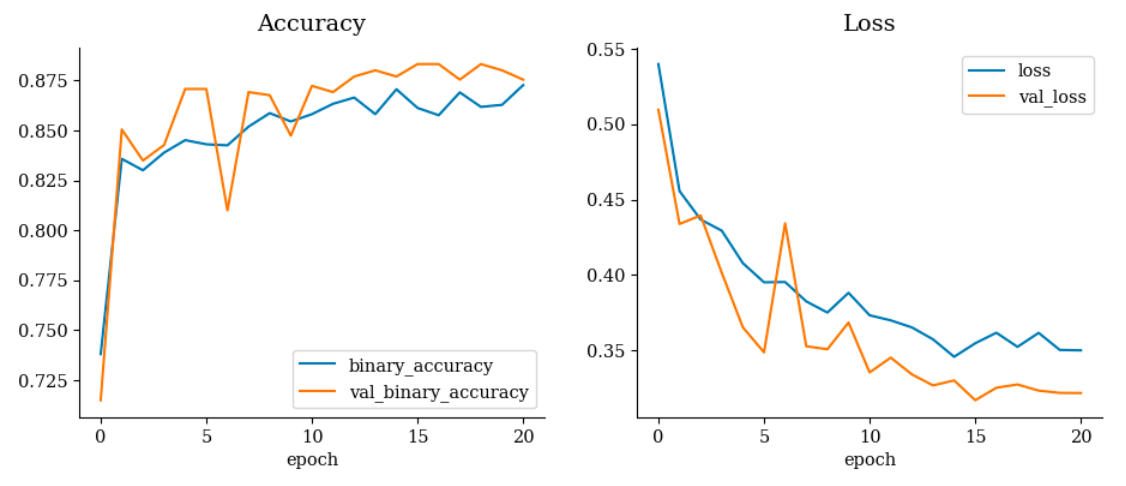
ResNet enables the construction of deeper neural networks, with more than a hundred layers, which was previously impossible due to the vanishing gradient problem. The residual connections allow the Network to learn better representations and optimize the gradient flow, making it easier to train deeper networks.

1. **Results**

The result is shown and discussed.

Below you can see the visualized accuracy of each trained model and also the loss.

**CNN:**

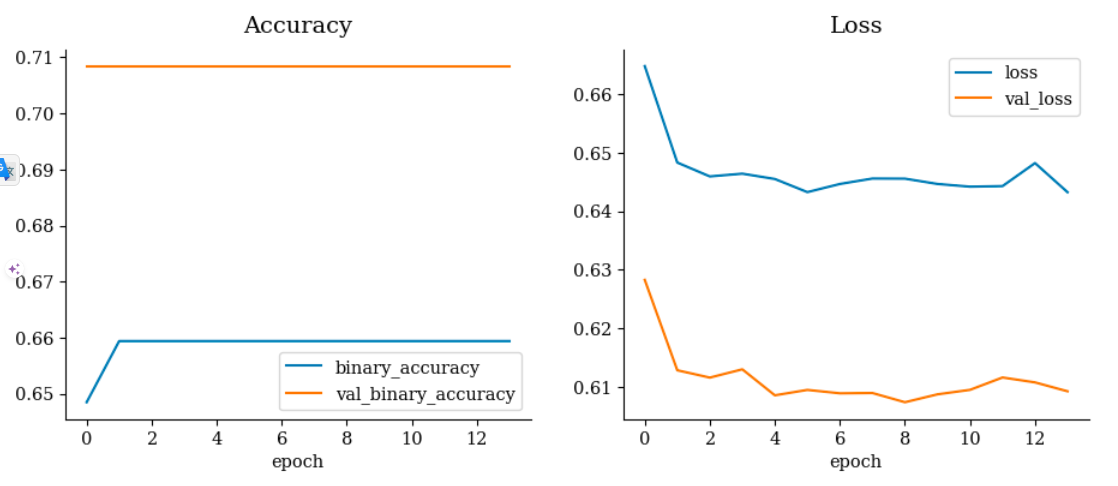


ROC-AUC: 0.82570

Accuracy: 0.82768

Loss: 0.55176

**LSTM:**



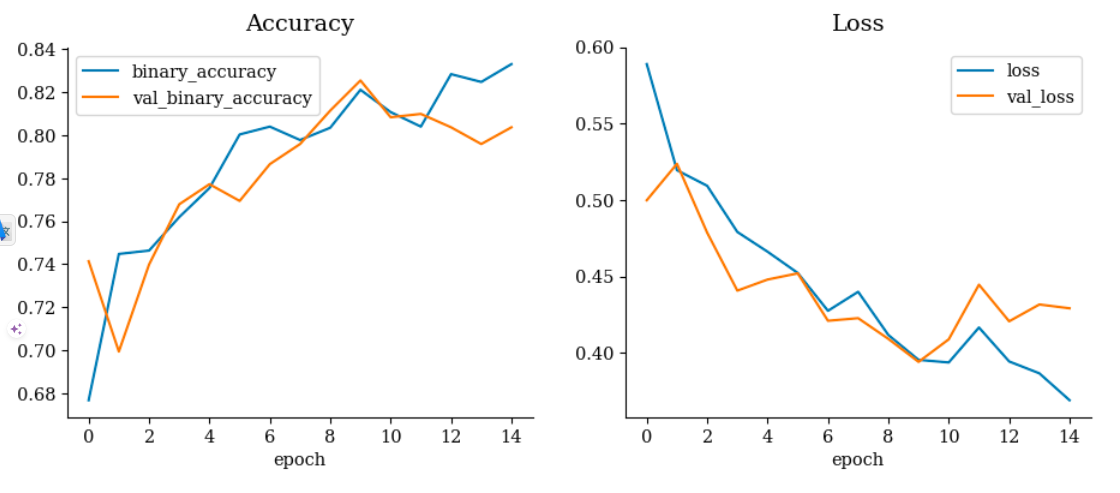
ROC-AUC: 0.50000

Accuracy: 0.70844

Loss: 0.60736

**Transfer Learning:**

* **MobileNetV3:**

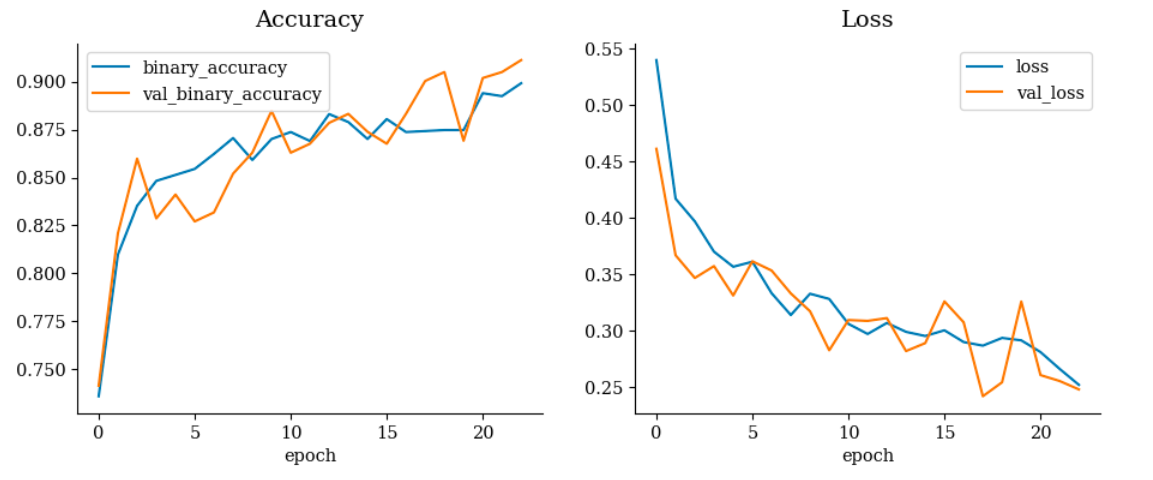


ROC-AUC: 0.74243

Accuracy: 0.72672

Loss: 0.55209

* **EfficientNetB1:**

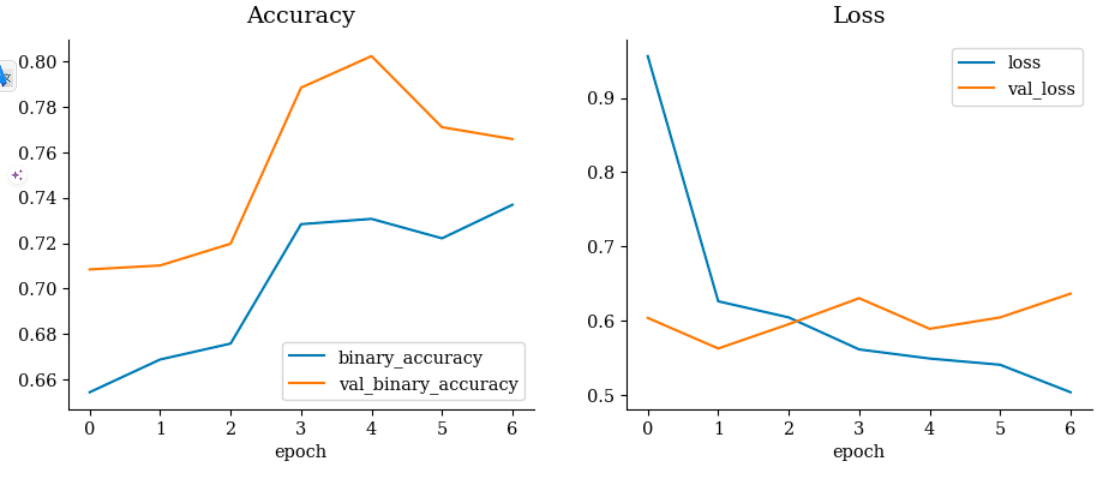


ROC-AUC: 0.87666

Accuracy: 0.79634

Loss: 0.44258

* **VGG16:**

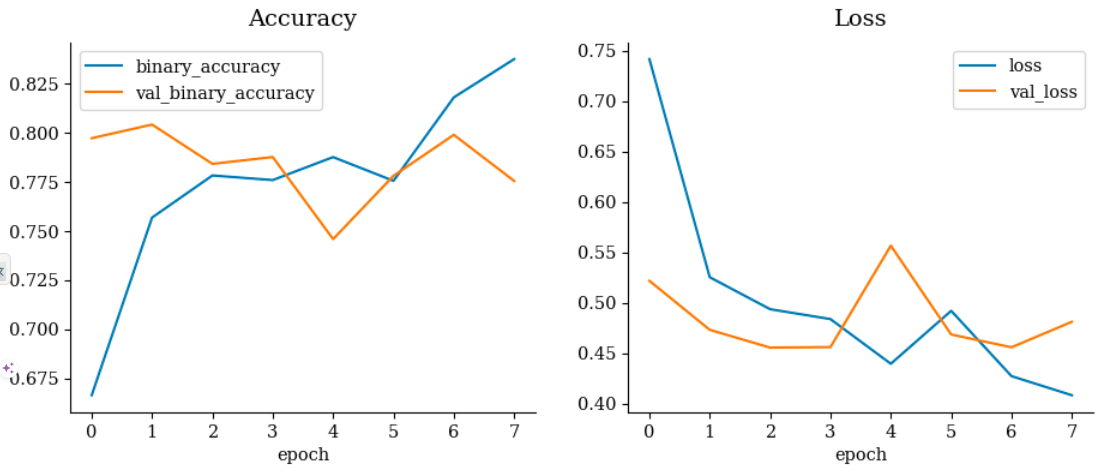


ROC-AUC: 0.75920

Accuracy: 0.71018

Loss: 0.56242

* **ResNet50V2:**



ROC-AUC: 0.82299

Accuracy: 0.78416

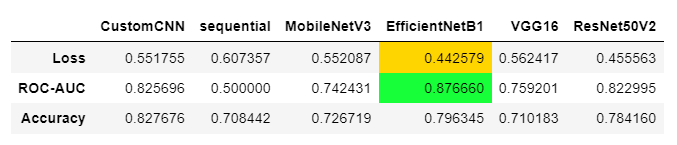
Loss: 0.45556

* 1. **Discussion**

we applied the algorithms which we selected above and the highest accuracy we got was 87% which was on our Pre trained model name EfficientNetB1.

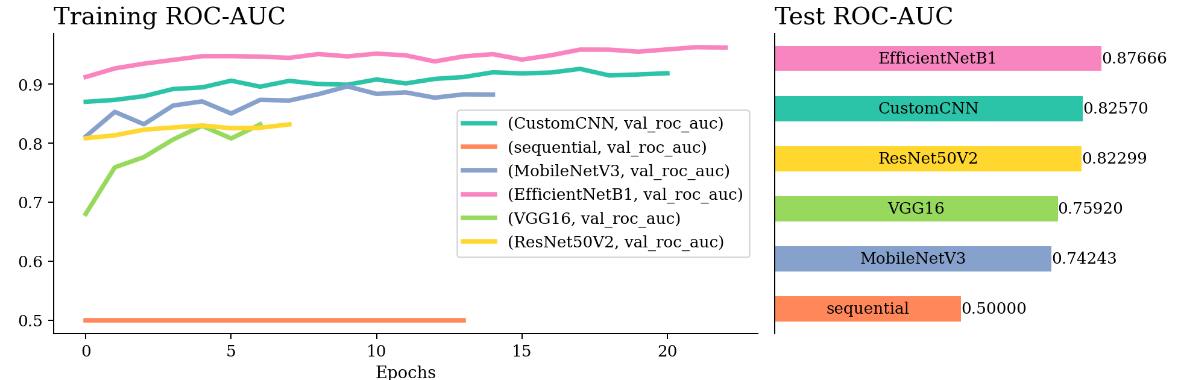
* 1. **Performance Matrices**

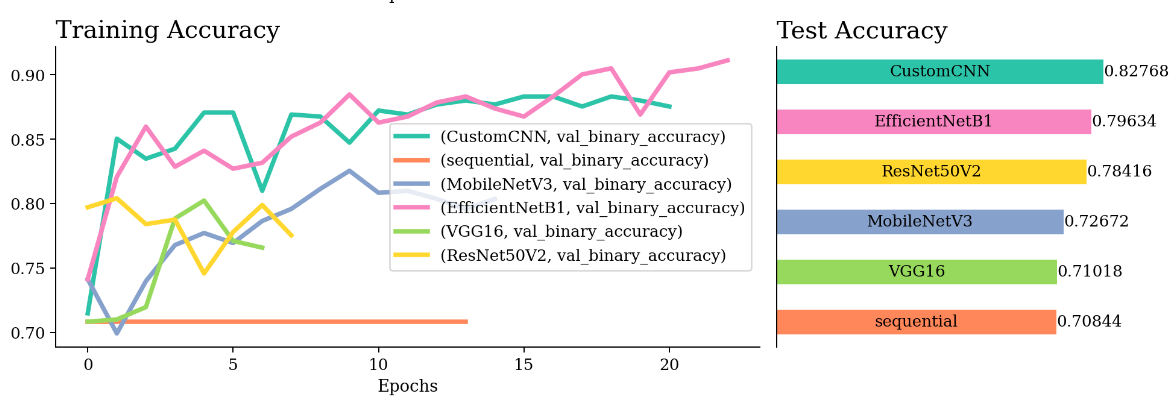
Here clearly we can see that EfficientNetB1 is the Model with Highest Accuracy.

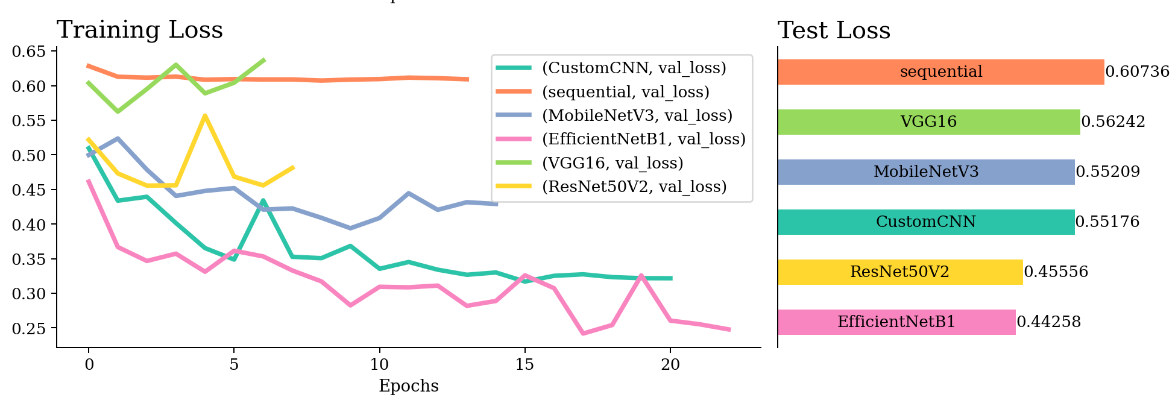


* 1. **Comparison**

Given below is the comparison of each training model.







1. **Conclusion**

The project concludes with the successful implementation of deep learning algorithms, particularly Pre trained EfficientNetB1, for Breast Cancer prediction, achieving accuracy rate of 87% respectively. The initial steps, including data augmentation significantly contributed to model performance.