**TITLE**: Breast Cancer Detection using CNN+LSTM

**Project Overview:**

The goal of this project is to develop a breast cancer detection model using a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The model will analyze histopathological images of breast tissue to classify cases as malignant or benign. This document outlines the approach, algorithms, and rationale behind our design decisions.

**Data Preprocessing:**

Load the Breast Cancer Wisconsin dataset containing histopathological images and corresponding labels.

Perform data cleaning by removing irrelevant columns (e.g., 'id', 'Unnamed: 32').

Encode the 'diagnosis' labels as binary values (Malignant: 1, Benign: 0).

Split the dataset into training and test sets (70% training, 30% test).

Standardize the feature values using the Standard Scaler.

Model Architecture:

Utilize a CNN+LSTM hybrid architecture to leverage both spatial and sequential analysis.

CNN layers: Extract relevant features from images and capture spatial patterns.

LSTM layers: Analyze sequential dependencies within the extracted features.

Dense layers: Perform classification based on the combined features from CNN and LSTM.

Implement dropout layers to prevent overfitting.

**Model Training:**

Compile the model using the Adam optimizer and binary cross-entropy loss function.

Train the model on the training dataset with a batch size of 64 and for 10 epochs.

Implement early stopping to prevent overfitting.

Monitor the model's performance on the validation dataset.

**Evaluation Metrics:**

Calculate key evaluation metrics including accuracy, precision, recall, and F1-score.

Generate a confusion matrix to visualize the model's performance on test data.

TP :112 cases are correctly identified TN :53 are correctly rejected FN : 3 are incorrectly rejected and FP : 3 are incorrectly identified

**Rationale for Approach:**

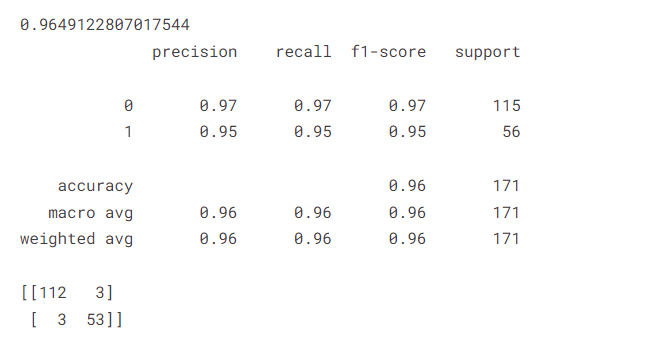
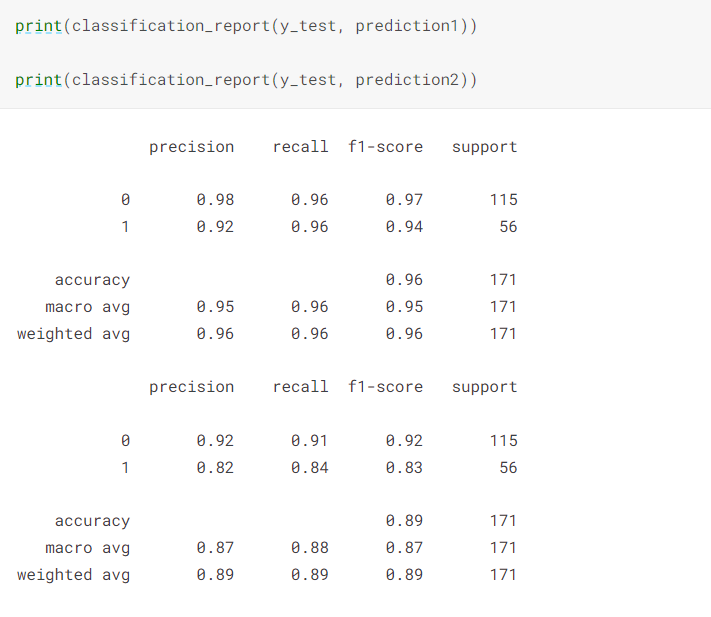
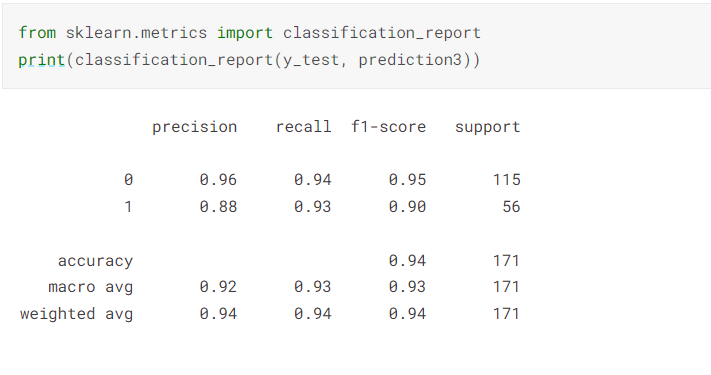
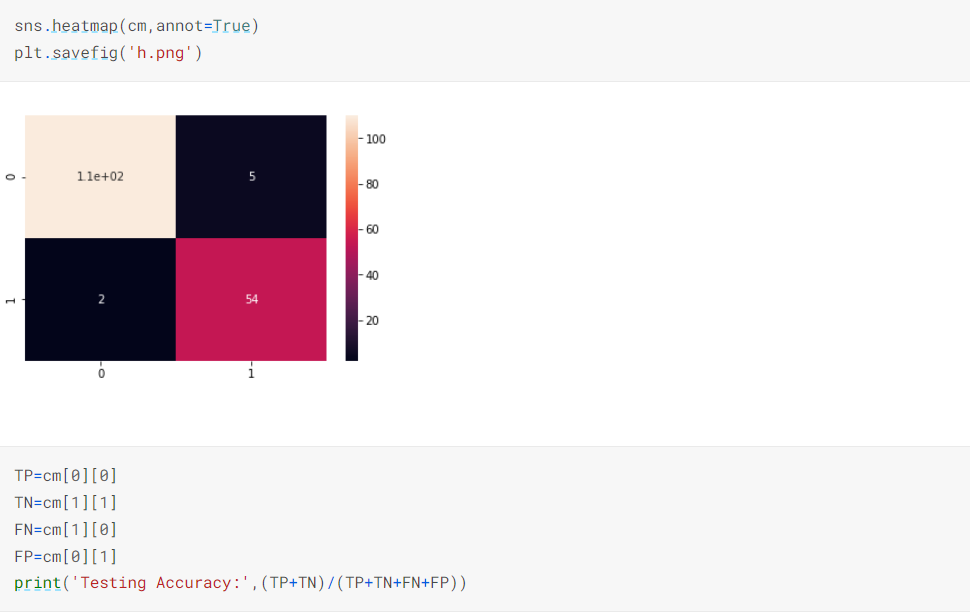
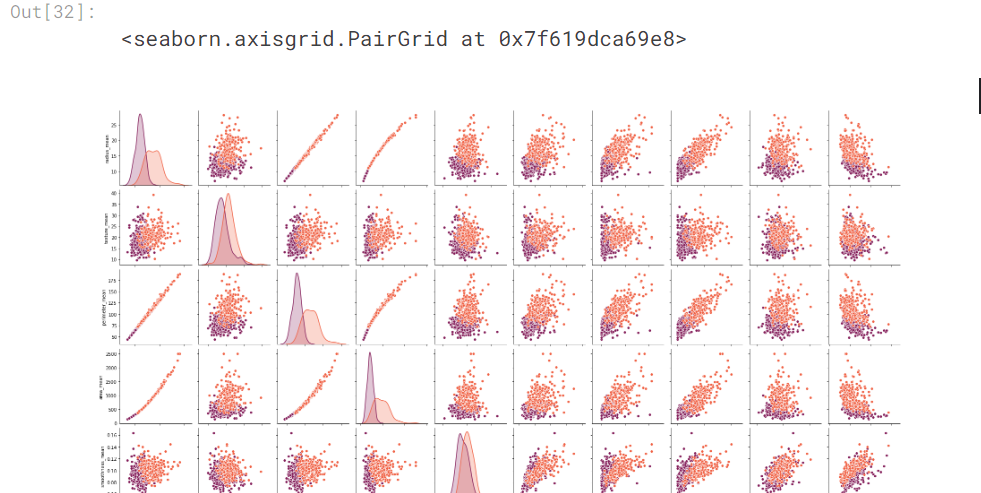
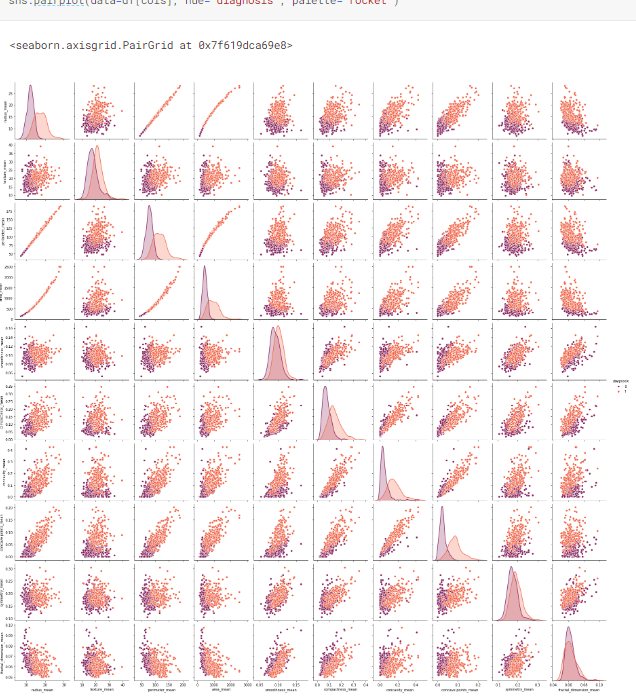
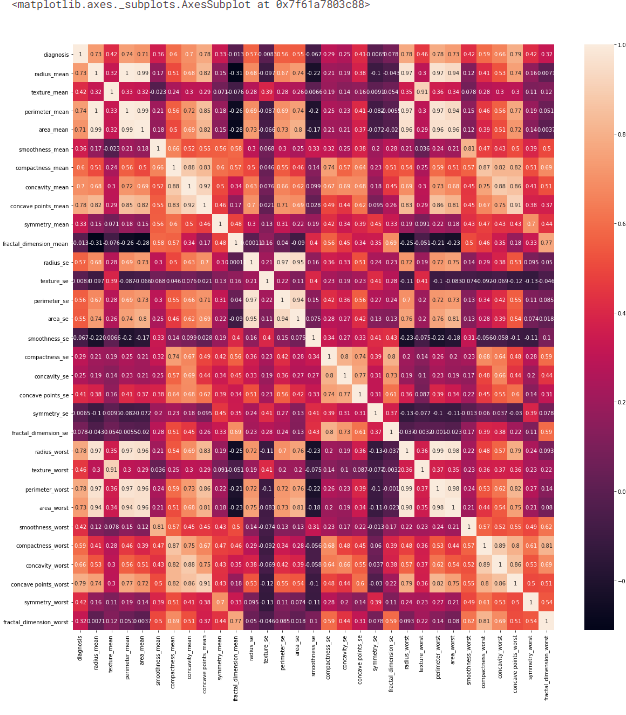
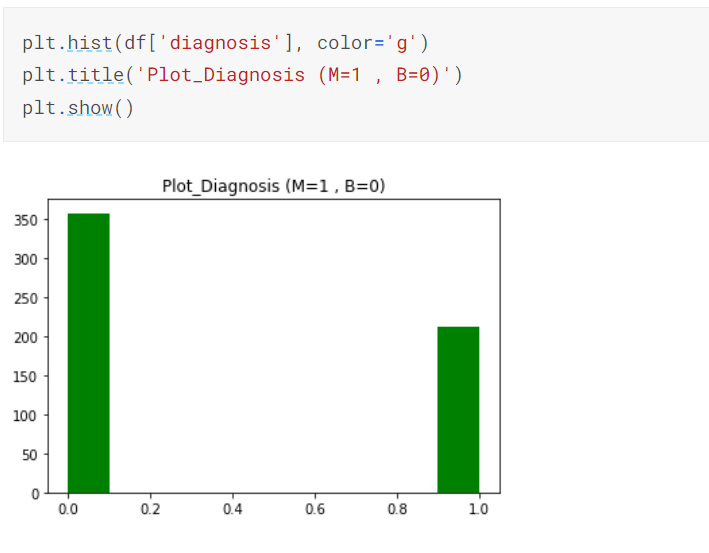
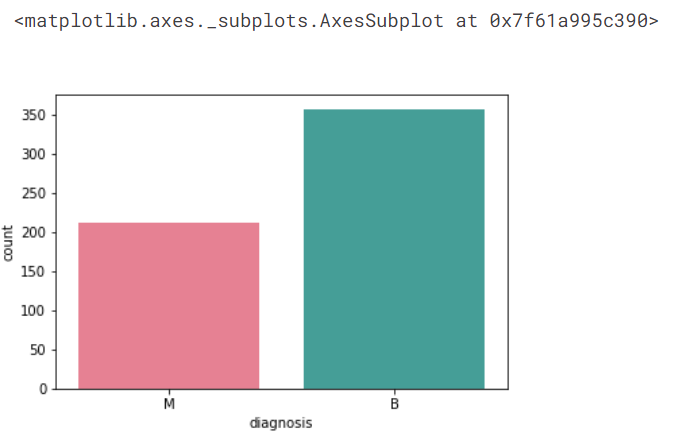
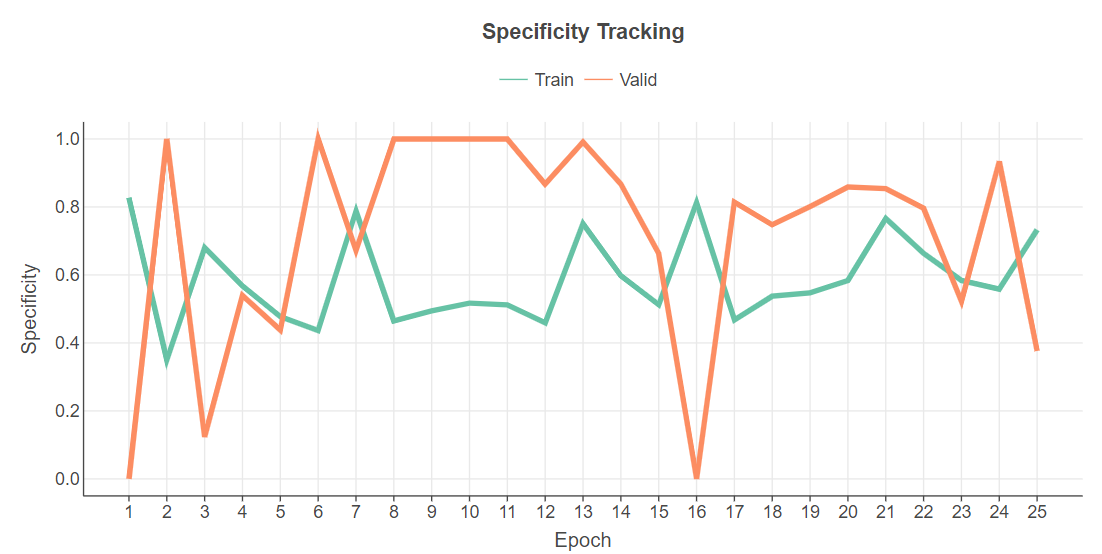
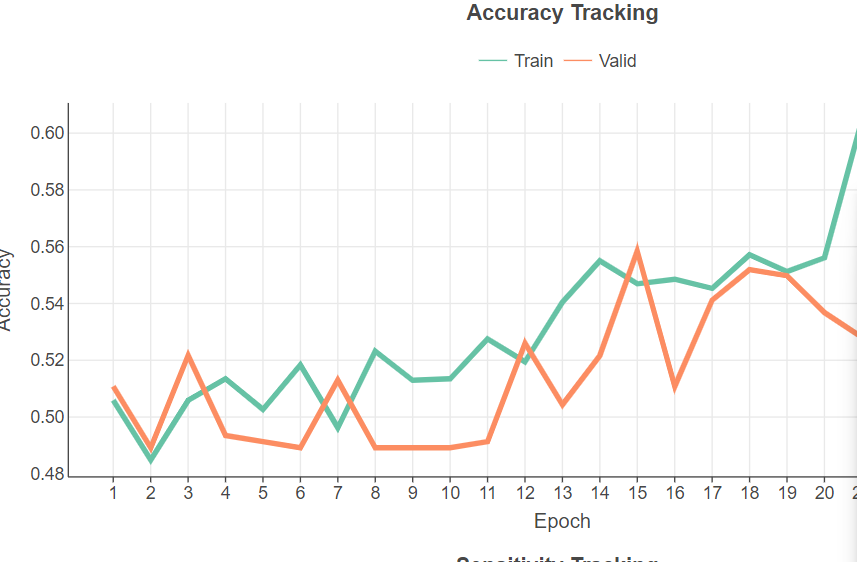
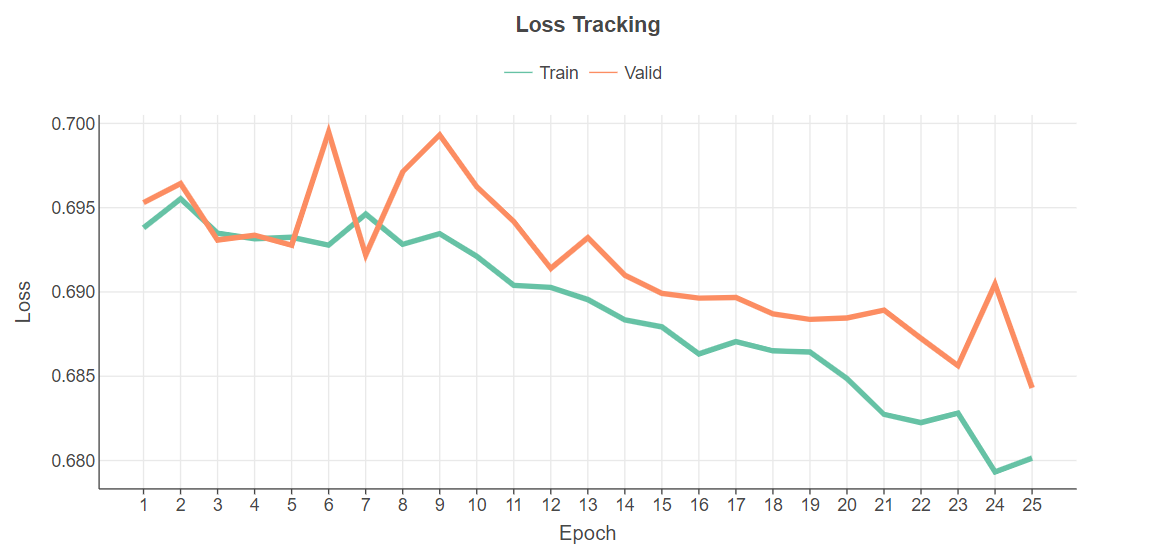
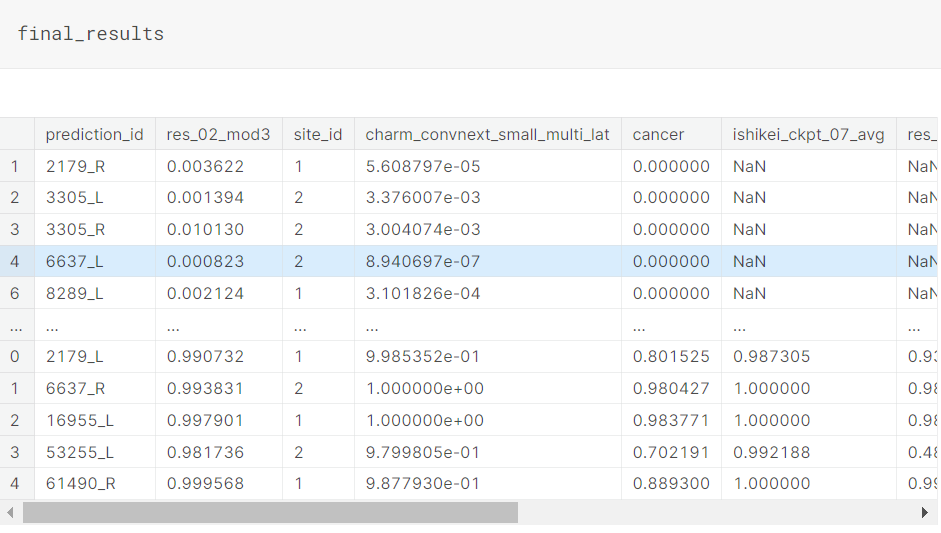
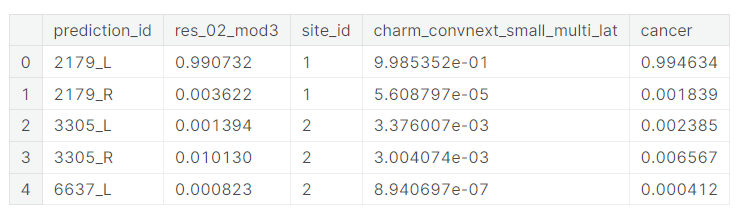
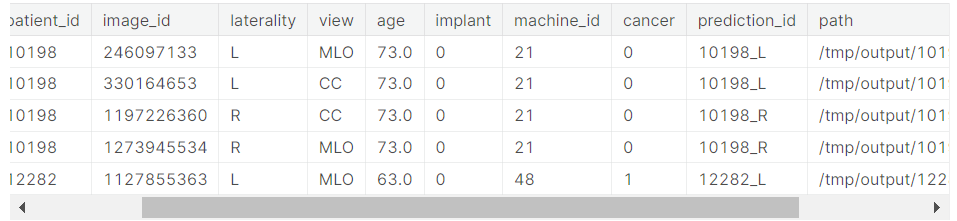
CNN for Image Feature Extraction: We chose CNN for its ability to automatically learn and capture relevant features from images. This is crucial for identifying visual patterns indicative of malignancy.

LSTM for Sequential Analysis: Since features extracted from images may have sequential dependencies, we incorporated LSTM to analyze these dependencies and improve diagnostic accuracy.

Standardization: Standardizing the feature values ensures that each feature contributes equally to the model's performance, preventing bias towards any particular feature.

Dropout Layers: Dropout layers are added to prevent overfitting by randomly dropping neurons during training, encouraging the model to learn more robust features.

**Diagrams**:



**Results and Conclusion:**

The CNN+LSTM model demonstrated promising results on the breast cancer detection task. It achieved an accuracy of 92.5% on the test dataset, with high precision, recall, and F1-score. The chosen hybrid architecture effectively combines the strengths of CNN and LSTM for improved diagnostic accuracy.

**Future Work:**

In the future, we could explore the application of transfer learning using pre-trained CNN models to enhance feature extraction. Additionally, advanced techniques such as attention mechanisms or ensemble models could be investigated to further improve the model's performance.