Infosys Springboard

Tumor Trace: Breast Cancer detection

Prepared by: Jaya Madhav

Mail: madhavjaya4@gmail.com

INTRODUCTION

Breast cancer is the most common cancer in women worldwide and a leading cause of cancer-related deaths. Early and accurate diagnosis is crucial for improving outcomes through timely treatment. Traditional methods like mammography, ultrasound, and MRI depend heavily on radiologists, making them prone to errors due to subjective interpretation.

Advances in artificial intelligence (AI) and machine learning have significantly improved the accuracy and consistency of medical imaging. AI models, especially convolutional neural networks (CNNs), can detect subtle patterns in images that may be missed by the human eye, making them ideal for tasks like tumor classification.

This study focuses on using MRI scans, known for their high contrast and detailed imaging of soft tissues, to classify breast tumors as benign or malignant. However, challenges arise due to the variability in tumor size, shape, and intensity in MRI images. To address these challenges, we propose TumorTrace, an AI framework combining traditional image processing techniques with deep learning models like VGG16 and ResNet18.

The study uses a structured dataset divided into training, validation, and test sets, with balanced cases of benign and malignant tumors. Extensive data augmentation techniques simulate real-world variability. Feature extraction combines classical methods such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Gray Level Co-occurrence Matrix (GLCM) with deep learning-based features.

This hybrid approach aims to achieve high accuracy while maintaining interpretability—crucial for medical applications. The model is evaluated on multiple performance metrics to ensure it generalizes well across datasets and effectively classifies both tumor types. This study demonstrates the potential of combining traditional and modern computational techniques for reliable and interpretable breast cancer diagnosis.

PROBLEM STATEMENT

Diagnosing breast cancer through medical imaging is a complex and challenging task due to the diverse and variable characteristics of tumors observed in MRI scans. Traditional diagnostic methods often depend on the expertise of radiologists, making them susceptible to inconsistencies and subjective biases, particularly in ambiguous or borderline cases.

High rates of false positives and false negatives in these methods can delay treatment or lead to unnecessary procedures, negatively affecting patient outcomes. Additionally, the absence of automated and scalable diagnostic tools places a significant strain on radiologists, especially in settings with limited resources.

To overcome these issues, there is an urgent need for a reliable, accurate, and interpretable automated system capable of assisting radiologists in classifying breast tumors as benign or malignant. Such a system must effectively utilize large-scale imaging datasets while addressing the variability in tumor structure and imaging quality.

METHODOLOGY

DATASET

The dataset for this study is carefully structured into three main directories: training, validation (val), and testing, with each containing subfolders for Benign and Malignant tumor categories. It includes a total of 29,274 images distributed across 4,185 folders. Specifically, the training set comprises 5,559 Benign and 14,875 Malignant images, the validation set contains 408 Benign and 1,581 Malignant images, and the test set includes 1,938 Benign and 4,913 Malignant images. To ensure compatibility with deep learning models like VGG16 and ResNet, all images are resized to 224x224 pixels, which is an optimal input size for these architectures.

To improve the model's robustness and generalization, various data augmentation techniques are applied. These include random horizontal and vertical flips, rotations of up to 30 degrees, affine shear transformations, and adjustments to image brightness, contrast, saturation, and hue. These augmentations introduce variability that mimics real-world scenarios, enabling the model to better generalize and reducing the likelihood of overfitting to the training data. This approach enhances the model’s capacity to identify and classify breast tumors across a wide range of MRI images with differing tumor sizes, shapes, and contrasts.

The dataset, however, is imbalanced, with more Malignant images than Benign ones. This imbalance can lead to biased learning. To counteract this, strategies such as assigning class weights, oversampling the Benign class, or using balanced batches during training are employed. With these preprocessing steps and augmentations, the dataset forms a strong foundation for developing an accurate and reliable breast tumor classification model. These methods also ensure the model's adaptability to real-world imaging scenarios, enhancing its diagnostic reliability in clinical applications.

DATASET PREPARATION

Proper dataset preparation is crucial for building a reliable and accurate breast tumor classification model. The first step involves resizing all MRI images to a standard dimension of 224x224 pixels to meet the input requirements of deep learning models like VGG16 and ResNet. This resizing ensures consistency across the images and makes them compatible with pre-trained models, allowing for efficient batch processing during training. This uniformity is important for maintaining essential features necessary for accurate classification.

To improve the model’s ability to generalize and handle real-world variations, various data augmentation techniques are applied. These augmentations artificially increase the training dataset by simulating different imaging conditions. Techniques such as Random Horizontal Flip and Random Vertical Flip create flipped versions of images to mimic different tumor orientations and mirror views that may appear in clinical settings. Random Rotation, within a ±30-degree range, helps the model learn from tumors in various angles, making it more adaptable to different image orientations. Affine Shear Transformations distort the images by shifting axes to imitate irregularities in MRI scans, while Color Jittering alters brightness, contrast, saturation, and hue

to address variations in MRI scan quality due to different imaging equipment or conditions. These augmentations not only enhance the diversity of the dataset but also reduce the risk of overfitting, allowing the model to learn generalized patterns instead of memorizing specific features.

After augmentation, the dataset is divided into training, validation, and test subsets to facilitate model development. To optimize data handling, PyTorch DataLoaders are used, with a transformation pipeline that integrates resizing, augmentation, and normalization. Normalization adjusts the pixel values to a range of [-1, 1] using a mean and standard deviation of 0.5, ensuring numerical stability during training and improving the model's convergence speed.

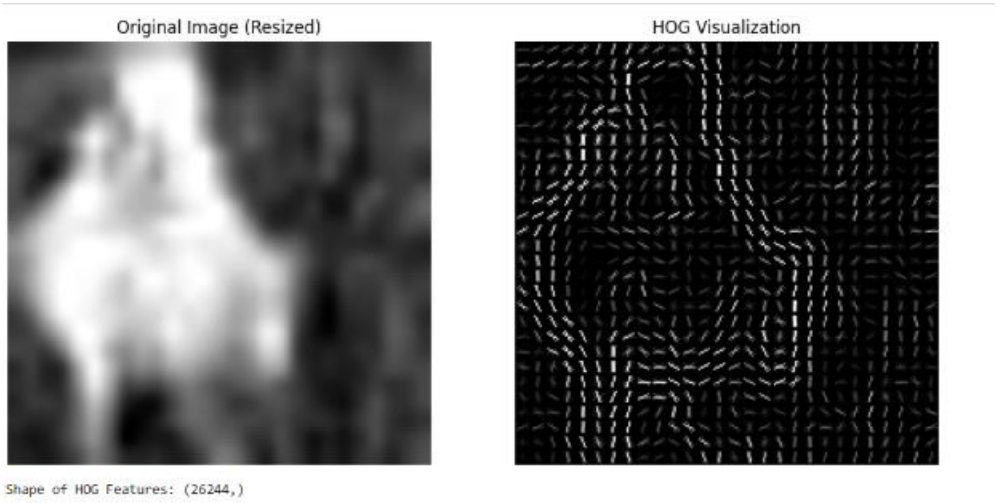
The DataLoader batches the images (typically with a batch size of 32), shuffles the training data to prevent the model from learning the sequence of samples, and utilizes multiprocessing to speed up data loading. Validation and test DataLoaders are set up without shuffling to maintain consistent evaluation. This organized pipeline not only optimizes data handling but also ensures the model experiences a balanced mix of real-world variations and standardized inputs. With these preprocessing steps in place, the dataset is well-prepared to train a highly accurate and reliable model capable of distinguishing between benign and malignant tumors.

FEATURE EXTRACTION

Feature extraction is a critical step in image analysis that helps capture important patterns and characteristics from images, which machine learning models can then use for classification tasks. This process transforms raw image data into numerical representations or features, such as edges, textures, and spatial relationships, that describe key attributes of the image. These extracted features allow the model to understand the structure and texture of the tumor, aiding in distinguishing between benign and malignant tumors. Both traditional techniques and deep learning methods are used in this study to provide a complete understanding of tumor characteristics.

HISTOGRAM OF ORIENTED GRADIENTS (HOG)

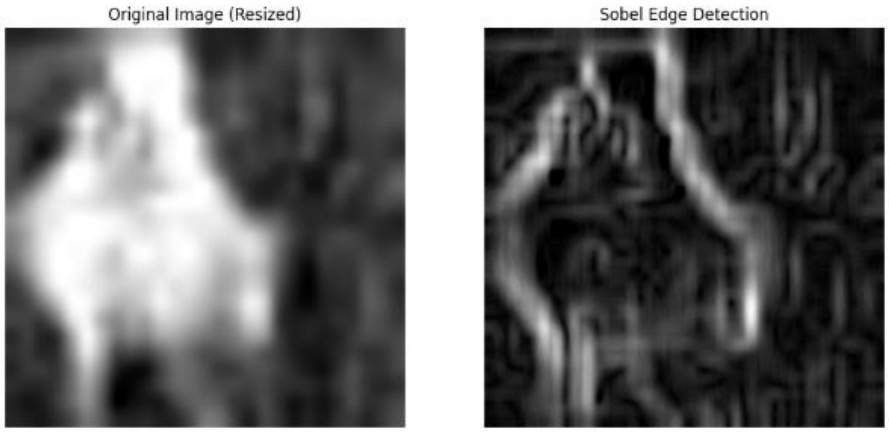
HOG is a feature extraction method that analyzes the distribution of gradient directions in specific regions of an image. By focusing on the gradients, HOG highlights edges and contours, which are essential for detecting the shape and boundaries of tumors.



SOBEL EDGE DETECTION

Sobel edge detection uses Sobel filters to compute gradients in an image by applying convolutional kernels. This process enhances edge detection, particularly in areas with sharp intensity changes, such as tumor boundaries. This technique is useful for identifying abrupt transitions in pixel intensity, which are commonly associated with tumors, aiding in the classification and analysis of the tumor's structure.

Sobel edge detection is especially effective in distinguishing benign from malignant tumors based on their edges and contours. Malignant tumors tend to have irregular, sharp boundaries, while benign tumors often have smoother, more consistent edges. This method improves the clarity of tumor regions, helping both machine learning models and radiologists make more accurate diagnoses.



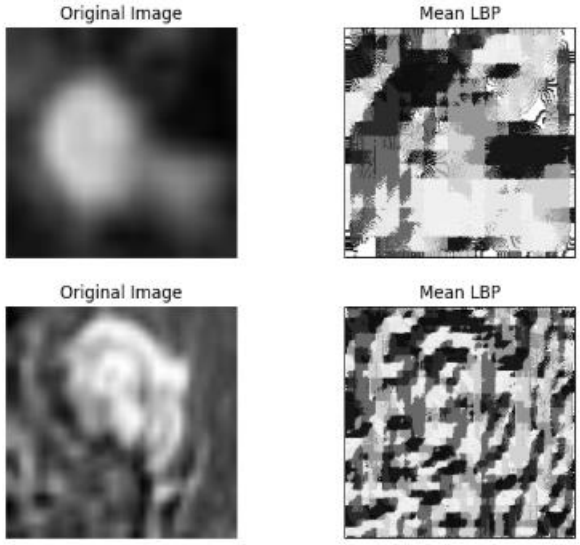
LOCAL BINARY PATTERNS (LBP)

LBP is a texture analysis technique that compares the intensity of each pixel with its neighboring pixels within a defined neighborhood. The results are encoded into a binary pattern that represents the local texture. LBP is particularly useful for capturing fine-textured differences between tumors, helping to identify whether the tumor texture is smooth (benign) or irregular (malignant).

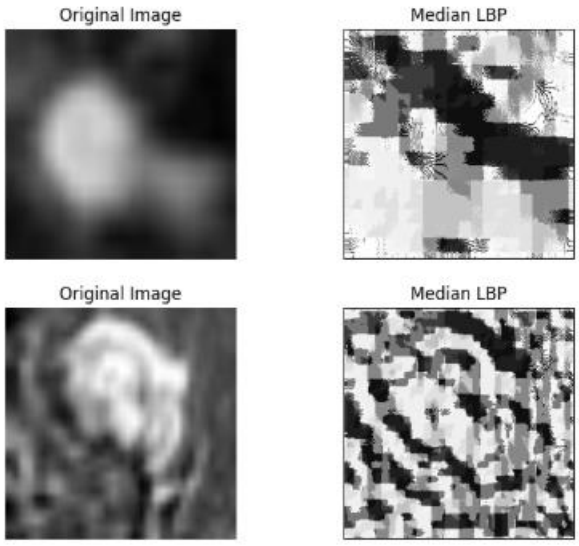
In the context of breast cancer, LBP assists in differentiating between benign and malignant tumors by highlighting textural variations. Malignant tumors generally exhibit more irregular and complex textures, while benign tumors tend to have smoother, more homogeneous textures.

LBP Mean, Median, and Variance

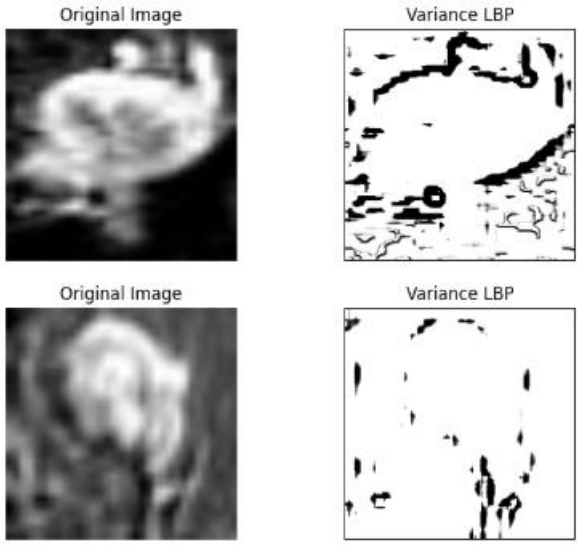
- **LBP Mean:** The average of LBP feature values across an image, representing the overall texture intensity. A higher LBP mean suggests a complex, heterogeneous texture, often seen in malignant tumors, while a lower mean indicates smoother textures typically found in benign tumors.



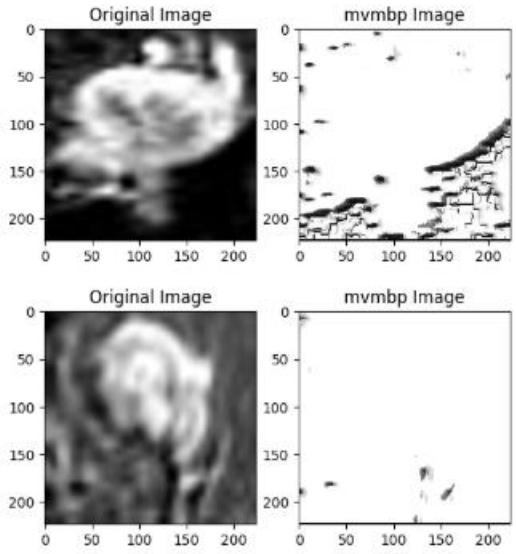
- **LBP Median:** The middle value of LBP feature values, providing a robust measure of the dominant texture pattern. It is less affected by extreme values, making it useful for analyzing tumors with variable textures. Benign tumors usually have a more uniform texture, reflected in the median, while malignant tumors have more irregular textures.



- **LBP Variance:** This measures the spread or variability of texture features across an image. A high variance indicates significant texture differences, which are characteristic of malignant tumors, while low variance suggests uniformity, typically seen in benign tumors. LBP variance provides valuable insights into the complexity and irregularity of the tumor surface, aiding in classification.

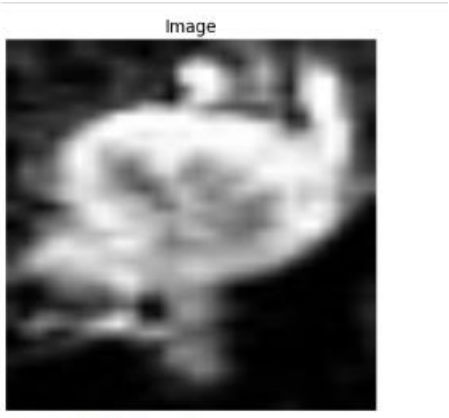


These LBP features—mean, median, and variance—offer a comprehensive understanding of tumor texture, which is crucial for distinguishing between benign and malignant tumors. By capturing subtle textural differences, these features improve the model's diagnostic accuracy.



GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM)

GLCM is a technique that analyzes the spatial relationships between pixel intensities in an image. It looks at how often pairs of pixels with specific intensity values occur in a defined spatial pattern. This method is valuable for extracting key texture properties like contrast, correlation, energy, and homogeneity. These features help distinguish different tissue types based on their unique texture patterns.



MODEL DEVELOPMENT

The tumor classification model is based on deep learning, specifically fine-tuning pre-existing architectures like VGG16, ResNet18, and ResNet50. These models are initialized with pre-trained weights and adapted for binary classification tasks by adding additional layers. The final model predicts whether a tumor is benign or malignant. The Adam optimizer is used for efficient weight updates during training, while binary cross-entropy loss is applied to calculate the error between predicted and actual tumor classifications. A learning rate scheduler is also employed to adjust the learning rate dynamically, ensuring efficient model training.

TRAINING AND VALIDATION

The training process involves iterating the model through multiple epochs while continuously validating its performance. Each epoch consists of training the model with the training set and validating it with the validation set. Performance metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) are monitored to ensure the model performs well across both tumor types (benign and malignant). To prevent overfitting and improve the model's ability to generalize, regularization techniques such as early stopping and dropout are used.

- **VGG16 Model:** The model showed excellent training accuracy, but validation accuracy fluctuated within a narrow range. The AUC remained consistent, with a slight increase in the final evaluation.
- **ResNet18 Model:** Achieved perfect training accuracy, but the validation accuracy remained stable, with a few false negatives and false positives. A learning rate scheduler was employed in the later stages to refine the training.
- **ResNet50 Model:** Like ResNet18, ResNet50 reached perfect training accuracy, with validation accuracy slightly improving during training. AUC increased, reflecting better performance compared to the other models.

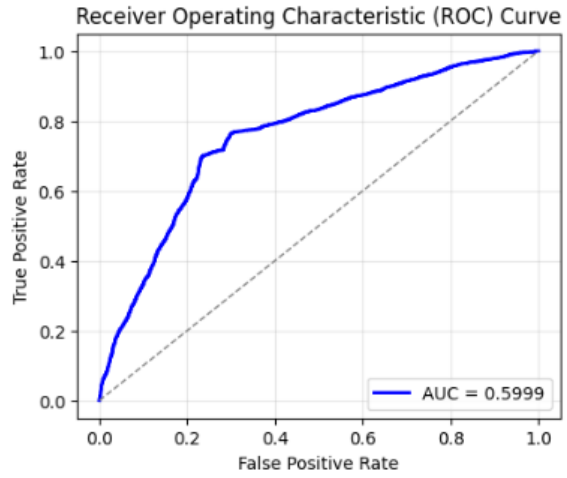
Despite all three models achieving near-perfect training accuracy, validation accuracy and AUC values indicated that there was still room for improvement, with ResNet50 showing the most promising results in terms of performance.

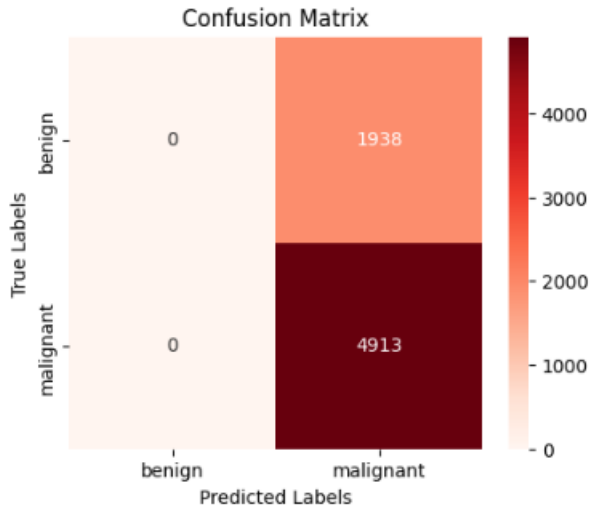
MODEL EVALUATION

The models were evaluated using a separate test dataset containing 1,938 benign and 4,913 malignant images. Diagnostic performance was assessed using metrics such as sensitivity, specificity, precision, recall, F1-score, and ROC-AUC. Special attention was given to sensitivity and specificity to ensure the models reliably detected both tumor types.

VGG16 Model:

```
              precision    recall  f1-score   support

      benign     0.0000    0.0000    0.0000      1938
   malignant     0.7171    1.0000    0.8353      4913

    accuracy                         0.7171      6851
   macro avg     0.3586    0.5000    0.4176      6851
weighted avg     0.5143    0.7171    0.5990      6851


Confusion Matrix:
 [[   0 1938]
 [   0 4913]]
Specificity: 1.0000, Sensitivity: 0.0000, AUC: 0.5999

test set: Average loss: 0.6009, Accuracy: 4913/6851 (71.71%)
```



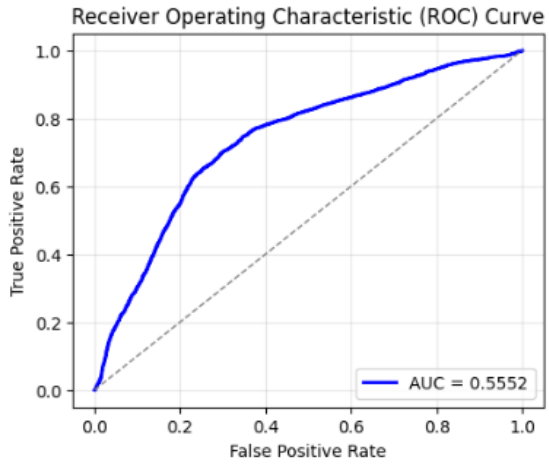


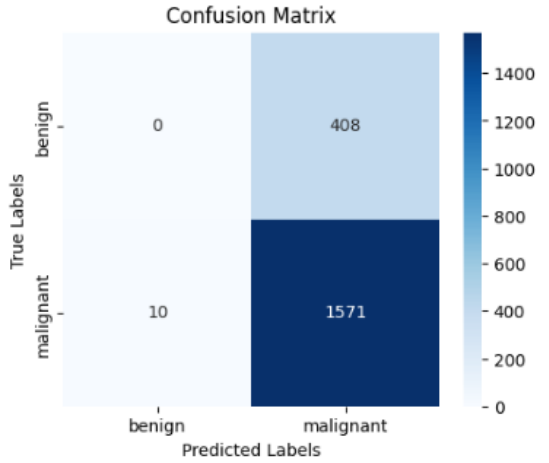
The VGG16 model demonstrated a relatively moderate test accuracy and AUC, suggesting room for improvement in detecting both benign and malignant tumors reliably.

ResNet18 Model:

	precision	recall	f1-score	support
benign	0.1872	0.0181	0.0329	1938
malignant	0.7144	0.9691	0.8225	4913
accuracy			0.7000	6851
macro avg	0.4508	0.4936	0.4277	6851
weighted avg	0.5653	0.7000	0.5991	6851

```
Confusion Matrix:
 [[  35 1903]
 [ 152 4761]]
Specificity: 1.0000, Sensitivity: 0.0000, AUC: 0.5552

test set: Average loss: 0.6471, Accuracy: 4796/6851 (70.00%)
```



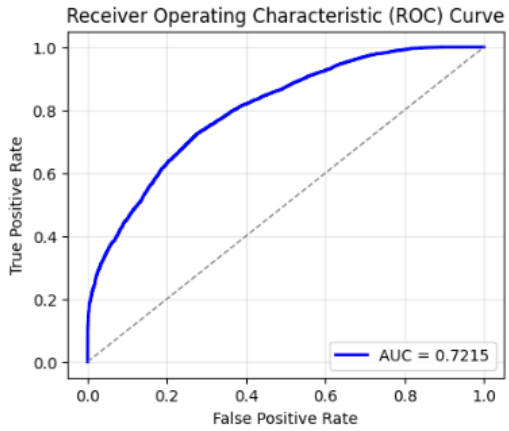


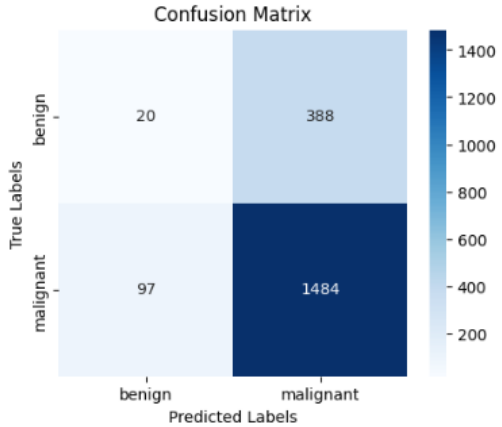
The ResNet18 model showed improved test accuracy and AUC compared to VGG16. Despite this,it still had a notable number of false positives, indicating some challenges in distinguishing benign cases.

ResNet50 Model:

	precision	recall	f1-score	support
benign	0.5406	0.1615	0.2487	1938
malignant	0.7409	0.9459	0.8309	4913
accuracy			0.7240	6851
macro avg	0.6407	0.5537	0.5398	6851
weighted avg	0.6842	0.7240	0.6662	6851

```
Confusion Matrix:
 [[ 313 1625]
 [ 266 4647]]
Specificity: 1.0000, Sensitivity: 0.0000, AUC: 0.7215

test set: Average loss: 0.5917, Accuracy: 4960/6851 (72.40%)
```





ResNet50 model achieved the highest accuracy and AUC among the three models, with the test accuracy reaching 72.40% and the AUC improving to 0.7215. This indicates better generalization to the test dataset, though false negatives and false positives still remain.ResNet50 showed the best overall performance in terms of test accuracy and AUC, suggesting it is the most reliable model for distinguishing between benign and malignant tumors in the test dataset.

All models exhibited relatively high test accuracy and AUC values, but there is still scope for improvement, particularly in reducing false positives and false negatives for more accurate diagnosis.

DEPLOYMENT AND SCALABILITY

The final model is designed for deployment in clinical settings to enable real-time classification of breast tumor MRI scans. The deployment ensures the model's efficient operation on modern GPUs and edge devices, allowing for quick, on-site diagnoses in healthcare facilities. To improve diagnostic accuracy, the system is also designed to integrate multi-modal imaging data in the future, such as mammograms and ultrasounds, enabling more comprehensive and accurate classifications.

The deployment uses the ResNet50 model, which has been fine-tuned for binary classification (Benign vs. Malignant). The model is optimized for fast processing of MRI images on GPUs. A user-friendly web application, powered by the Gradio interface, allows users to upload MRI scan images for real-time predictions. The model, based on a pre-trained ResNet50 architecture, has a modified final fully connected layer to predict the tumor's classification (Benign or Malignant). Images are processed with standard transformations, including resizing to 224x224 pixels and normalization based on ImageNet's statistics to ensure compatibility with the model's input requirements.

The interface lets users upload breast tumor MRI scan images, and the model will return a classification result, including the predicted tumor type and confidence level. The system is built for scalability, allowing it to efficiently manage multiple requests when deployed in clinical settings. This approach facilitates real-time decision-making and also supports the future integration of additional imaging technologies to further enhance the diagnostic process.