

# DOCUMENTATION

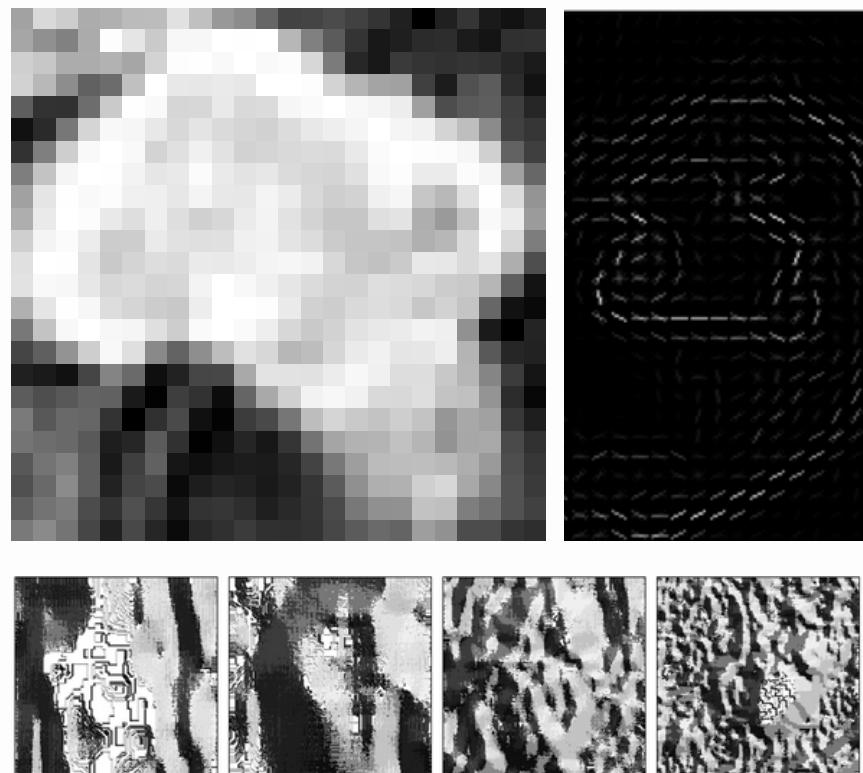
TumorTrace: MRI-Based AI for Breast Cancer Detection

J RAKSHA

# TABLE OF

# CONTENTS

INTRODUCTION	03
DATASET	05
DATASET PREPARATION	05
FEATURE EXTRACTION	06
MODEL DEVELOPMENT	13
TRAINING AND VALIDATION	13
MODEL EVALUATION	14
DEPLOYMENT AND SCALABILITY	17



# INTRODUCTION

Breast cancer is a major global health concern, ranking as the most common cancer among women and a leading cause of cancer-related deaths worldwide. Early and accurate diagnosis is critical for improving patient outcomes, as it enables timely intervention and treatment. Traditional diagnostic approaches, such as mammography, ultrasound, and magnetic resonance imaging (MRI), rely heavily on radiologists' expertise. While effective, these methods are prone to subjective interpretation, which can lead to variability in diagnosis and missed detections, especially in complex or ambiguous cases.

Recent advancements in artificial intelligence (AI) and machine learning have demonstrated significant potential in augmenting diagnostic accuracy and consistency in medical imaging. By leveraging large datasets and computational power, AI-driven models can identify subtle patterns and features that may be imperceptible to the human eye. Among these approaches, convolutional neural networks (CNNs) have emerged as a powerful tool for image-based classification tasks, including medical applications.

This study focuses on MRI imaging, which is widely regarded as a reliable diagnostic modality for detecting breast cancer due to its superior contrast resolution and ability to visualize soft tissues. Despite its advantages, interpreting MRI scans remains challenging due to the high variability in image characteristics, such as tumor size, shape, and intensity. Addressing this challenge, we propose TumorTrace, a robust AI-based framework that integrates traditional image processing techniques with state-of-the-art deep learning models to classify breast tumors as Benign or Malignant.

The dataset used in this study is structured into training, validation, and test sets, with a balanced distribution of benign and malignant cases. We enhance the dataset further through extensive augmentation techniques to simulate real-world variability. The feature extraction process combines handcrafted features, such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Gray Level Co-occurrence Matrix (GLCM), with deep learning-based feature extraction using customized VGG16 and ResNet18 architectures.

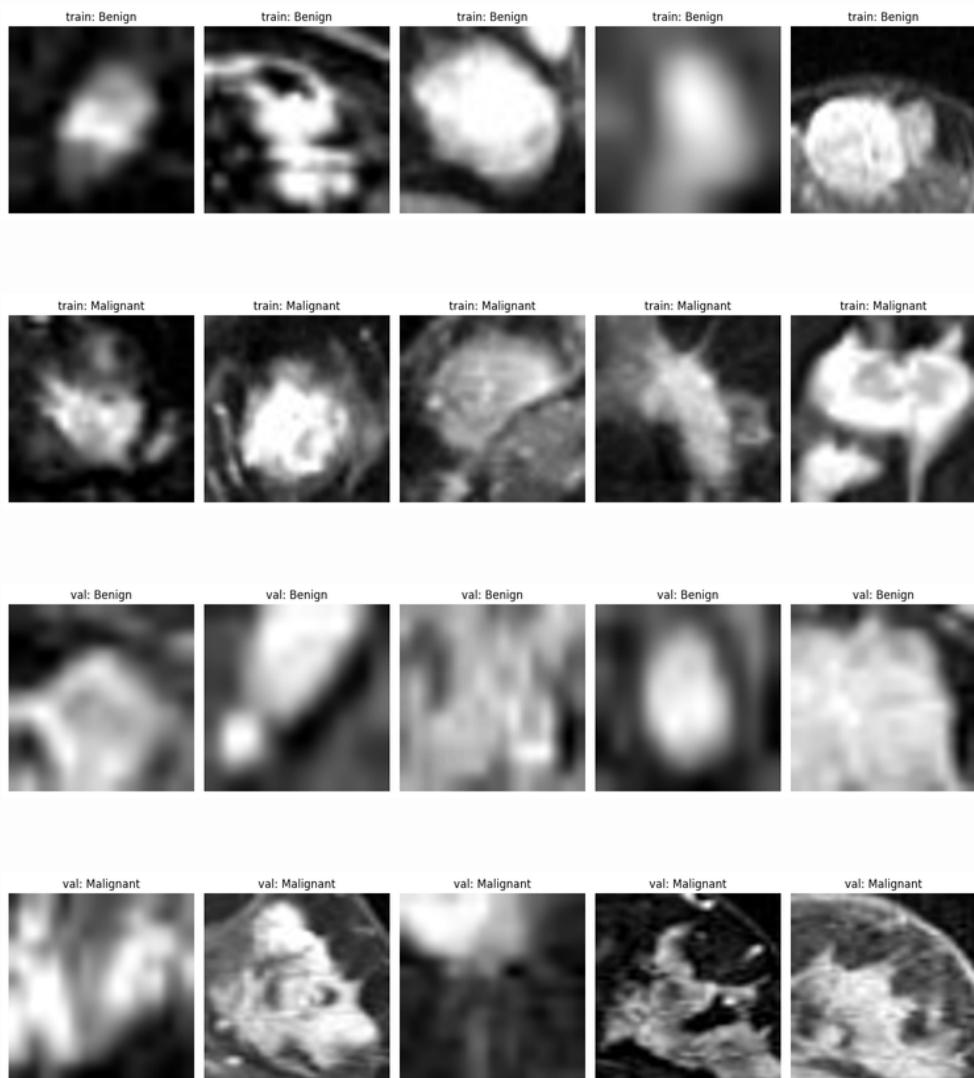
By combining classical and deep learning approaches, this study aims to create a hybrid framework that not only achieves high accuracy but also retains interpretability—an essential factor in medical AI applications. The proposed model is evaluated on various performance metrics, with a focus on its ability to generalize across different datasets and maintain sensitivity to both benign and malignant cases. This introduction sets the stage for exploring the synergy between traditional and modern computational techniques, paving the way for robust and interpretable

## PROBLEM STATEMENT

Breast cancer diagnosis is a critical yet challenging task in medical imaging due to the complexity and variability of tumor characteristics observed in MRI scans. Conventional diagnostic methods often rely on human expertise, which may be prone to inter-observer variability and subjective biases, especially in borderline or ambiguous cases.

The high false-positive and false-negative rates associated with traditional approaches can lead to delays in treatment or unnecessary interventions, adversely impacting patient outcomes. Furthermore, the lack of automated and scalable diagnostic tools increases the burden on radiologists, especially in resource-constrained settings.

Given these challenges, there is a pressing need for a robust, accurate, and interpretable automated system that can assist radiologists by classifying breast tumors as Benign or Malignant with high reliability. Such a system should effectively leverage large-scale imaging datasets while addressing variability in tumor morphology and imaging quality.



# METHODOLOGY

## DATASET

The dataset used in this study is meticulously organized into three primary directories: train, val (validation), and test, each containing subdirectories for Benign and Malignant tumor images. The complete dataset consists of 29,274 images spread across 4,185 folders. The training set includes 5,559 Benign images and 14,875 Malignant images, while the validation set contains 408 Benign images and 1,581 Malignant images. The test set comprises 1,938 Benign images and 4,913 Malignant images. For standardization and compatibility with deep learning models, all images are resized to 224x224 pixels, a size commonly used for CNNs like VGG16 and ResNet, ensuring optimal input dimensions for the models.

To further enhance the robustness and generalization capability of the model, several data augmentation techniques are applied. These include random horizontal and vertical flips, random rotations up to 30 degrees, affine shear transformations, and color jittering (adjusting brightness, contrast, saturation, and hue). These augmentations simulate real-world variations and help the model generalize better, reducing the risk of overfitting to specific patterns in the training data. Such transformations improve the model's ability to detect and classify breast cancer in a diverse range of MRI images, which may vary in tumor size, shape, and contrast.

However, the dataset exhibits class imbalance, with a significantly larger number of Malignant images compared to Benign images, which could lead to biased learning. To address this, techniques such as class weighting, oversampling the Benign class, or utilizing balanced batches during model training can be implemented to mitigate this issue. The dataset, enriched with augmented images and preprocessed for consistency, provides a strong foundation for training a reliable and accurate breast cancer classification model. Additionally, the incorporation of such diverse transformations ensures the model's ability to handle real-world imaging scenarios and improves its diagnostic capabilities in clinical settings.

## DATASET PREPARATION

Effective dataset preparation is a cornerstone of developing a robust and accurate breast tumor classification model. The process begins by resizing all MRI images to a standard size of 224x224 pixels, which aligns with the input requirements of deep learning models like VGG16 and ResNet. Resizing ensures uniformity across all images, making them compatible with pre-trained architectures and enabling efficient batch processing during training. This standardization is crucial for maintaining consistency while preserving essential features required for classification.

To enhance the model's generalization capability and make it robust to real-world variations, a series of data augmentation techniques is applied. These augmentations artificially expand the training dataset by introducing variations that mimic real-world imaging conditions. The Random Horizontal Flip and Random Vertical Flip transformations create flipped versions of images, simulating different tumor orientations and mirrored views that may occur in clinical imaging.

Random Rotation, applied within a range of  $\pm 30$  degrees, allows the model to learn from tumors at various angles, improving its ability to recognize tumors regardless of orientation. Further, Affine Shear Transformations distort the images by shifting one axis relative to another, imitating irregularities that might occur during MRI scans. Color Jittering, which modifies the brightness, contrast, saturation, and hue of images, accounts for variability in MRI scan quality caused by differences in imaging equipment or conditions. These augmentation techniques not only increase the diversity of the training dataset but also mitigate overfitting by exposing the model to a broader range of scenarios. This ensures that the model learns generalizable patterns rather than memorizing specific features.

Once the dataset is augmented, it is split into training, validation, and test subsets to facilitate the model development process. To streamline data loading, PyTorch DataLoaders are employed. A transformation pipeline, implemented using torchvision.transforms.Compose, integrates resizing, augmentation, and normalization. Normalization scales the pixel values to the range [-1, 1] using a mean and standard deviation of 0.5. This step is essential for ensuring numerical stability during training and improving the convergence rate of the model.

The DataLoader efficiently batches the images (commonly with a batch size of 32), shuffles the training data to prevent the model from learning the sequence of samples, and utilizes multiprocessing to accelerate data loading. Validation and test DataLoaders are created without shuffling to ensure consistent evaluation. This structured pipeline not only optimizes data handling but also ensures that the model is exposed to a balanced mix of real-world variations and standardized inputs. By integrating these preprocessing steps, the dataset is prepared to support the training of a highly accurate and reliable breast tumor classification model, capable of distinguishing between benign and malignant tumors with confidence.

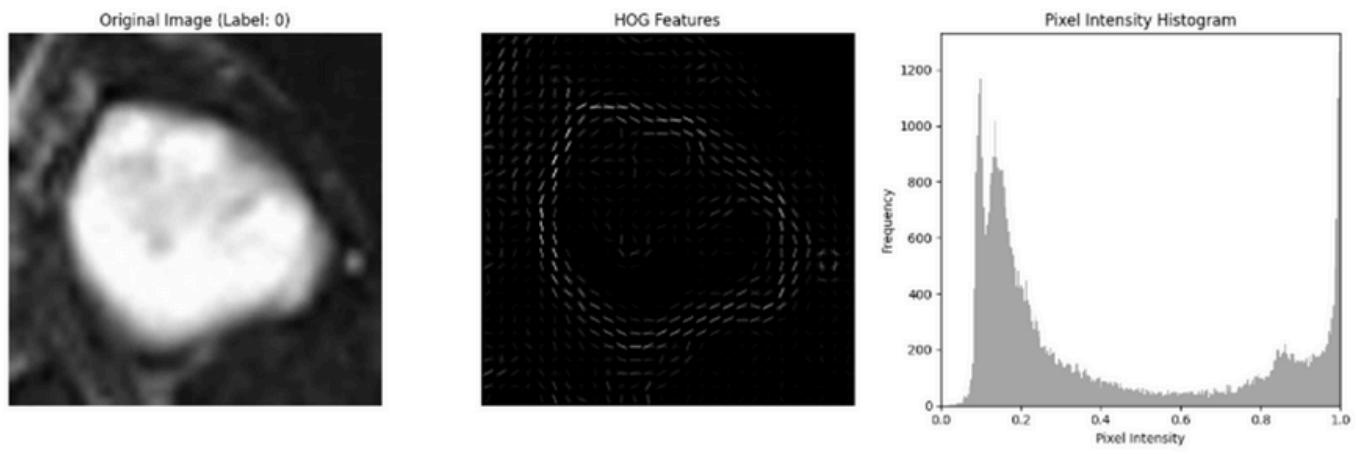
## FEATURE EXTRACTION

Feature extraction is a pivotal process in image analysis, designed to capture meaningful patterns and attributes from images that can be leveraged by machine learning models for classification tasks. This step transforms raw image data into a set of numerical representations, or features, that describe important characteristics such as edges, textures, and spatial relationships. The extracted features enable the model to understand the underlying structure and texture of the tumor, aiding in the differentiation between benign and malignant cases. Both handcrafted techniques and deep learning-based methods are used in this project to provide a comprehensive understanding of tumor characteristics.

## HISTOGRAM OF ORIENTED GRADIENTS (HOG)

The Histogram of Oriented Gradients (HOG) is a feature extraction technique that captures the distribution of gradient directions within localized regions of an image. By analyzing gradients, HOG emphasizes edges and contours, which are fundamental in identifying the structural details of objects, such as the boundaries and shapes of tumors.

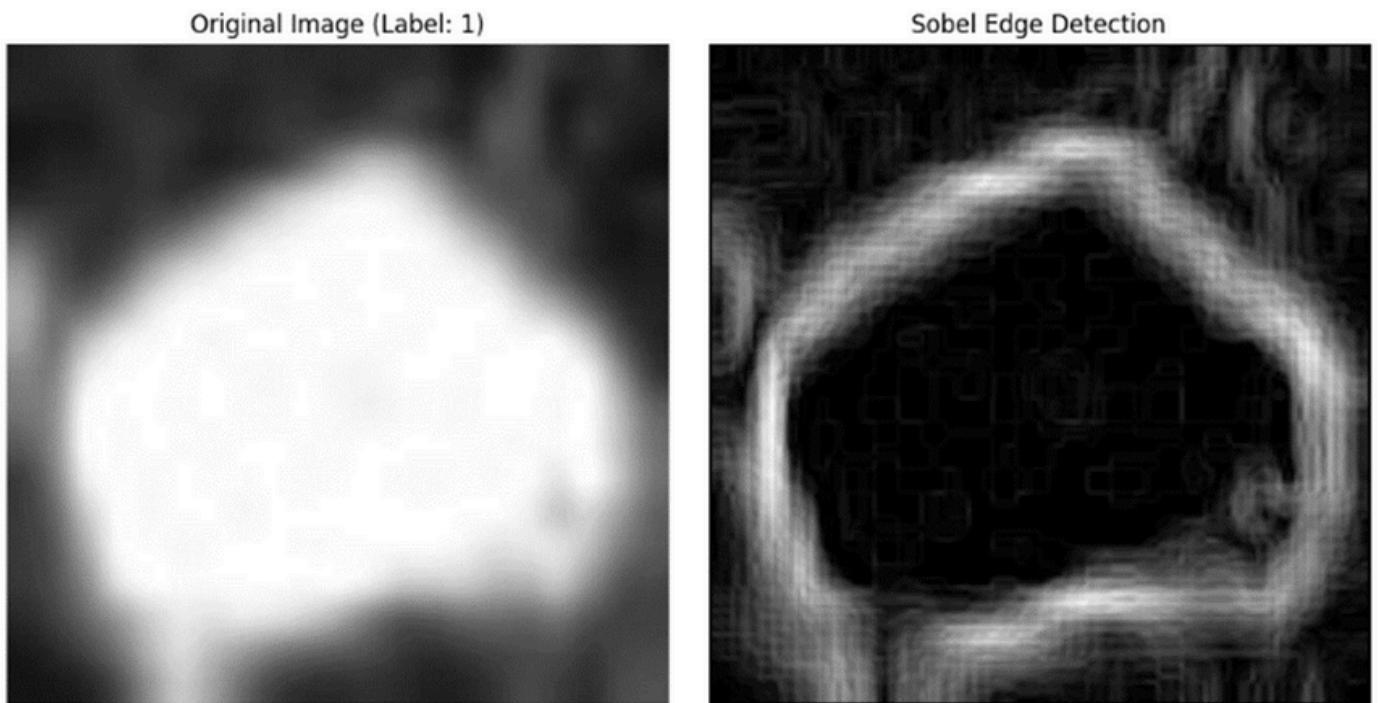
This method is particularly effective in detecting and highlighting the structural patterns of tumors, enabling the differentiation between benign and malignant regions based on their distinct edge and contour characteristics.

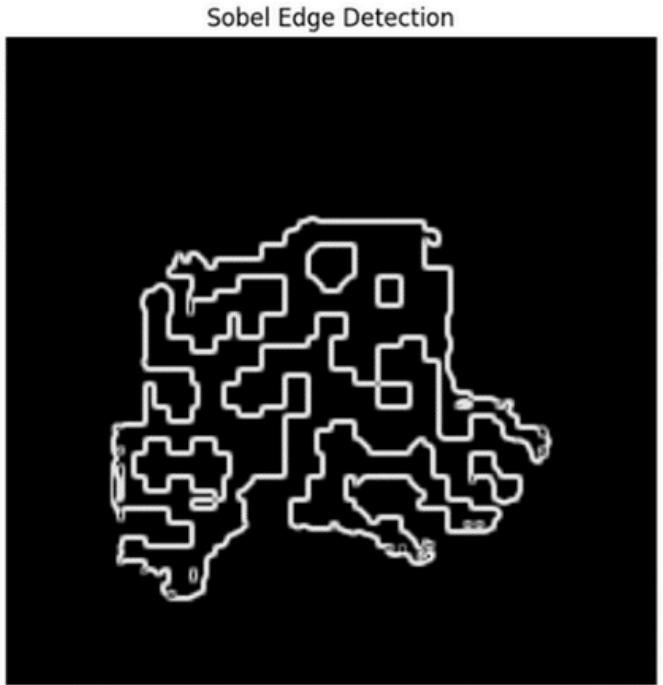
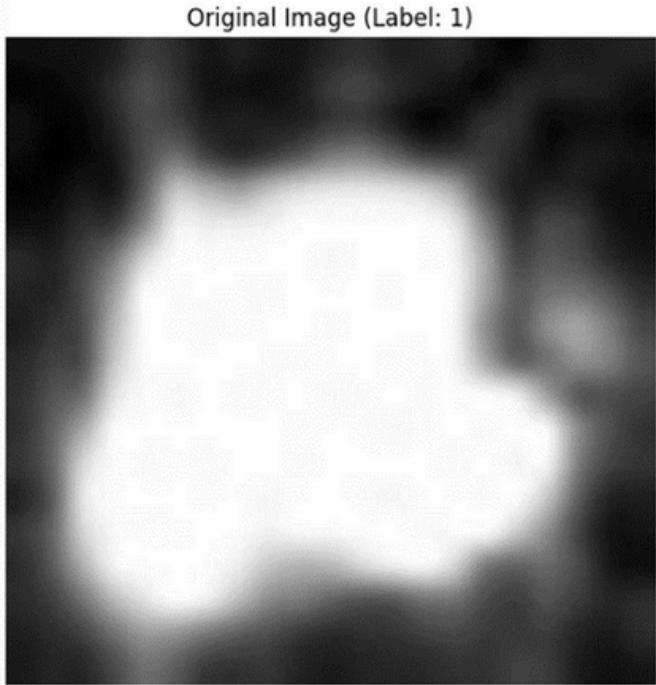


```
High pixel values: [0.8039215 0.8196079 0.8392157 ... 0.8039215 0.8039215 0.8039215]  
Number of high pixel values: 11745
```

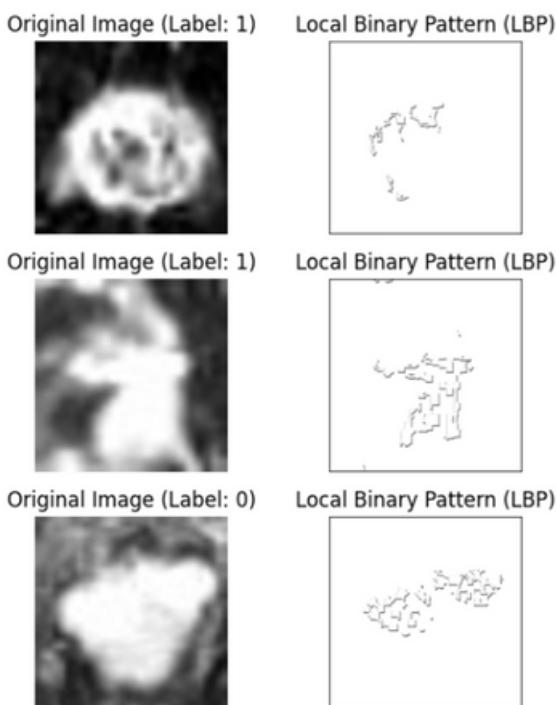
## SOBEL EDGE DETECTION

Sobel edge detection utilizes Sobel filters to compute image gradients by applying convolutional kernels. This process enhances the detection of edges and transitions in an image, focusing on areas where intensity changes are most pronounced. By emphasizing prominent edges and sharp intensity variations, Sobel edge detection is particularly effective in accentuating tumor boundaries. These boundaries often represent critical features for identifying and delineating tumor regions. This technique is highly useful for highlighting abrupt pixel intensity changes, which are commonly associated with tumor structures, aiding in accurate classification and analysis.



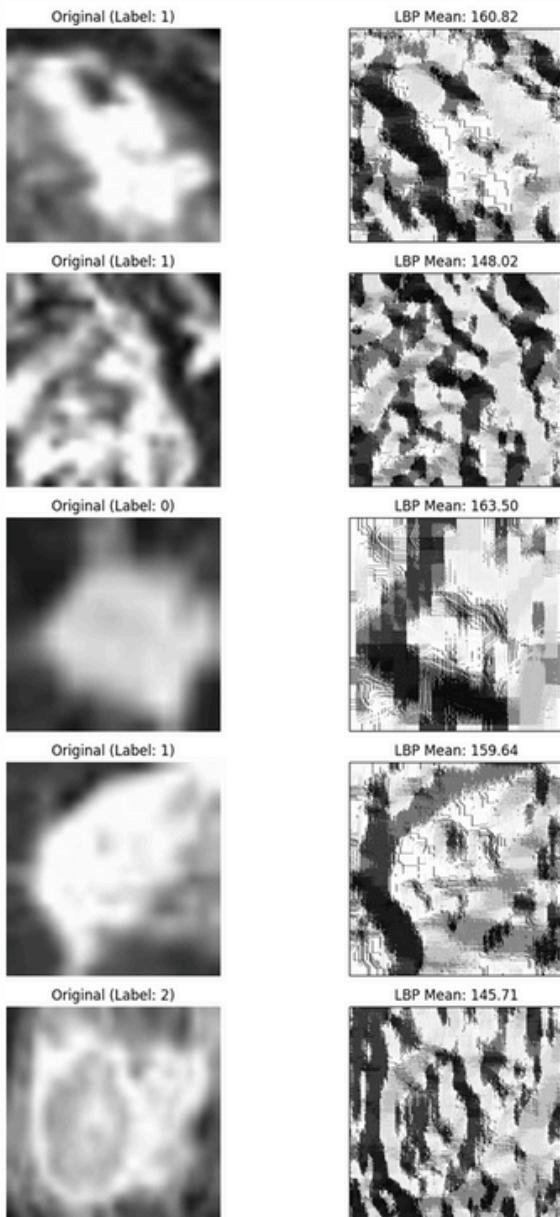
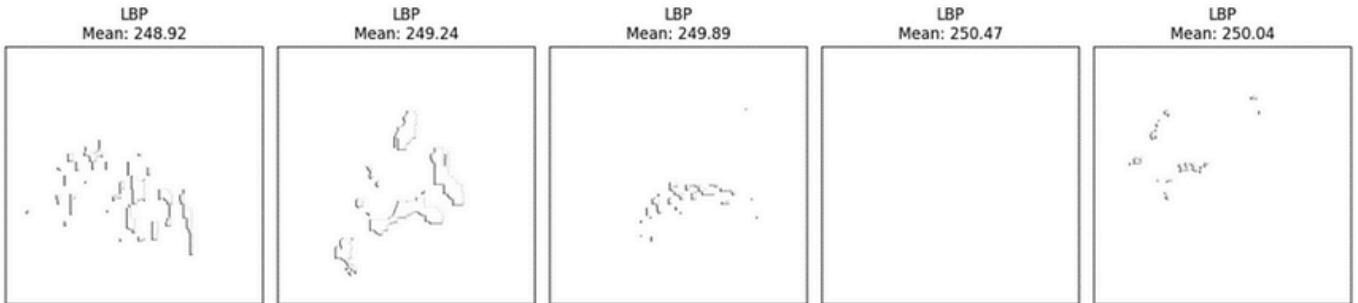
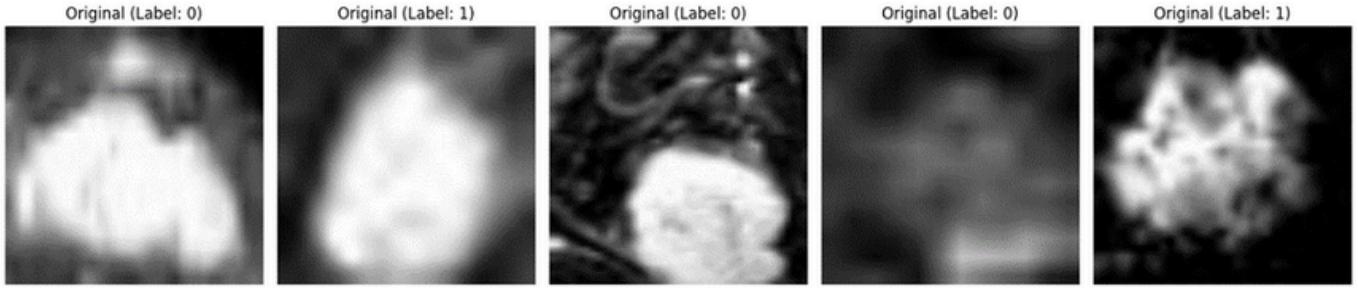


In the context of breast cancer analysis, Sobel edge detection is instrumental in highlighting the structural details and irregularities that distinguish benign from malignant tumors. Malignant tumors typically exhibit sharp, irregular boundaries, while benign tumors often have smoother, more regular edges. By accentuating these critical features, Sobel edge detection enhances the visualization of tumor morphology, supporting accurate segmentation, classification, and analysis. This technique not only improves the clarity of tumor regions but also aids radiologists and machine learning models in detecting and understanding key characteristics, ultimately contributing to more precise and reliable breast cancer diagnoses.



## **LOCAL BINARY PATTERNS (LBP)**

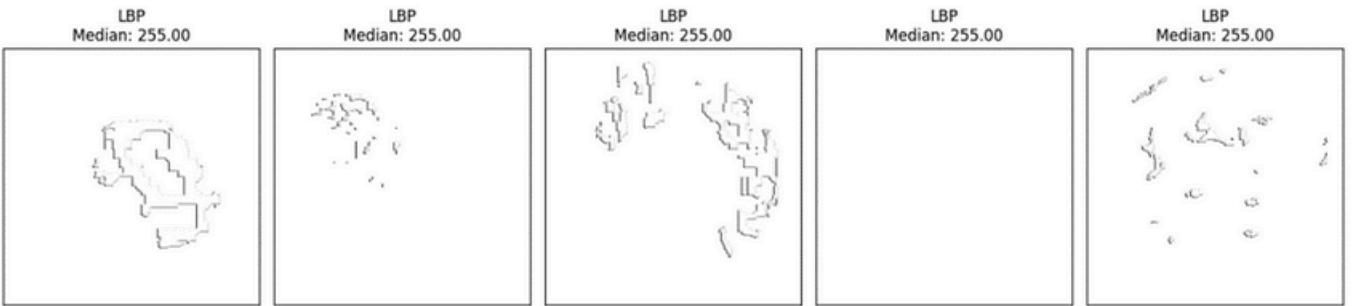
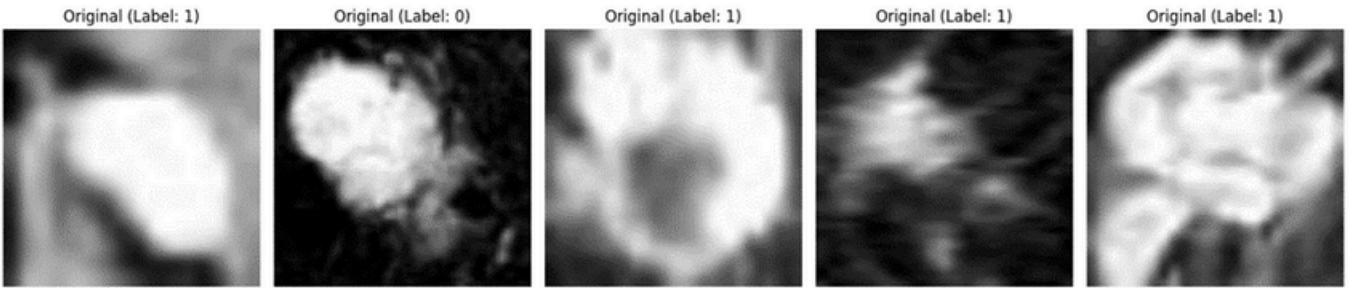
Local Binary Patterns (LBP) is a texture analysis technique that compares the intensity of each pixel with its surrounding neighbors in a local neighborhood. The result of these comparisons is encoded into a binary pattern that represents the local texture. This method effectively captures fine-grained texture features, making it particularly useful for identifying textural differences between tumor types. By analyzing variations in texture, LBP helps in distinguishing between smooth and irregular patterns, which are often indicative of benign or malignant tumor characteristics.



## LBP MEAN

The mean of Local Binary Pattern (LBP) feature values serves as a key statistical descriptor, reflecting the average texture intensity across an image. This metric provides a holistic assessment of the predominant texture characteristics, offering insights into the roughness or smoothness of a tumor's surface. A higher LBP mean typically signifies a more complex and heterogeneous texture, which is often associated with the irregular and chaotic structural patterns of malignant tumors. Conversely, a lower mean indicates a simpler, smoother texture, commonly observed in benign tumors with more homogeneous tissue structures.

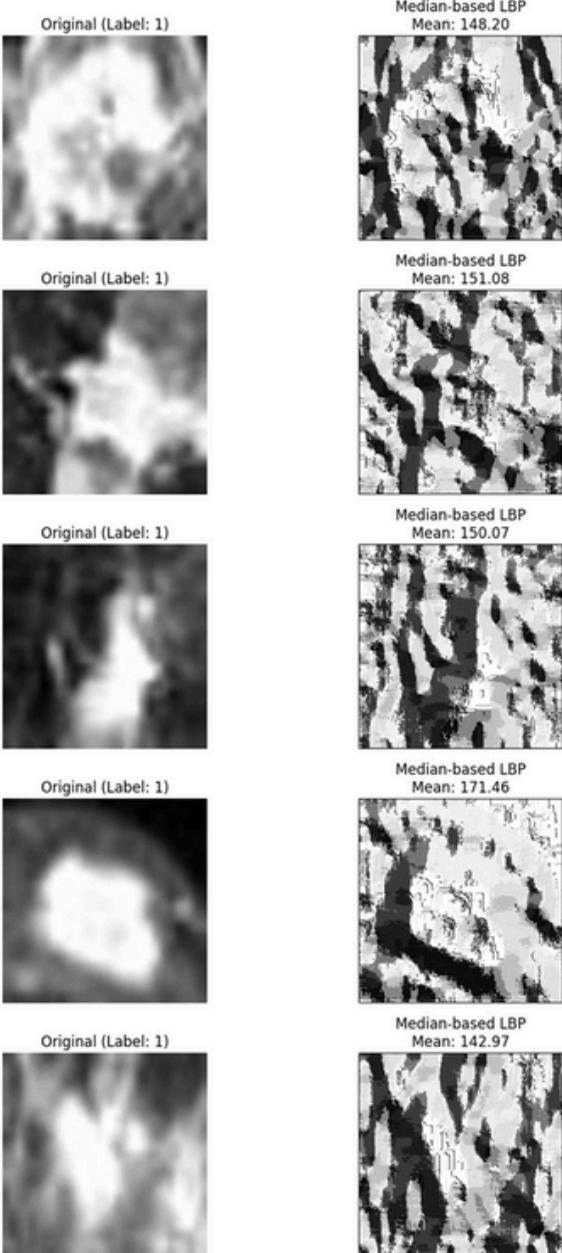
In the context of medical imaging, particularly breast cancer MRI analysis, the LBP mean helps summarize the intricate texture details that differentiate tumor types. This feature is crucial for distinguishing between benign and malignant cases, as it captures the subtle variations in texture that may not be immediately apparent through visual inspection. By quantifying these variations, the LBP mean enhances the diagnostic capabilities of machine learning models, aiding in more accurate tumor classification and supporting clinicians in identifying critical patterns indicative of malignancy.

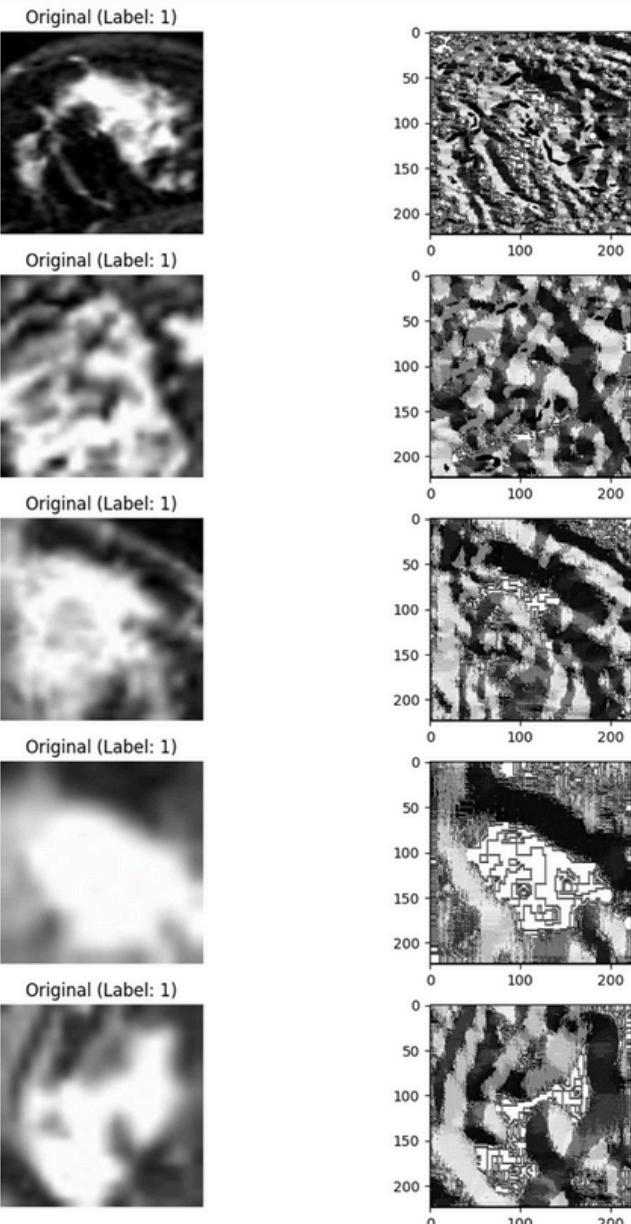
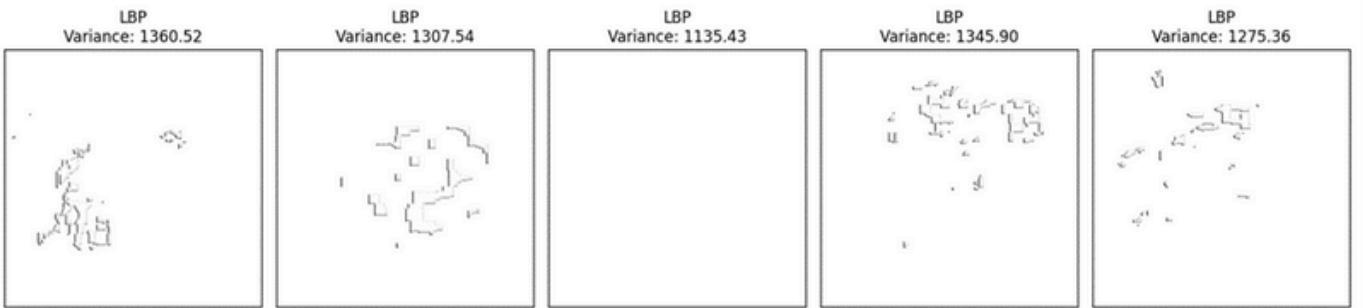
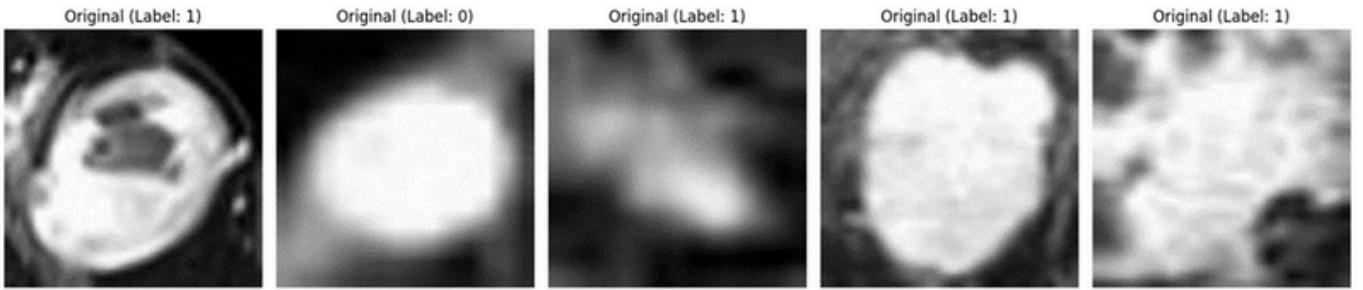


## LBP MEDIAN

The median of Local Binary Pattern (LBP) feature values represents the central value in the distribution of texture patterns, providing a robust and reliable measure of the predominant texture characteristics in an image. Unlike the mean, the median is unaffected by extreme values or outliers, making it particularly valuable for analyzing tumor regions where texture variability can be significant. By focusing on the middle value, the LBP median highlights the most frequent or dominant texture patterns, offering a clearer understanding of the tumor's overall structural properties.

In medical imaging, especially in breast cancer MRI analysis, the LBP median is instrumental in distinguishing between benign and malignant tumors. Benign tumors, often characterized by uniform and smooth textures, tend to have a median that reflects these consistent patterns. Malignant tumors, on the other hand, typically exhibit more irregular and heterogeneous textures, which are captured by variations in the median value. By summarizing the central tendency of texture features, the LBP median helps improve the interpretability of texture analysis and enhances the accuracy of tumor classification models, aiding clinicians in making more informed diagnostic decisions.



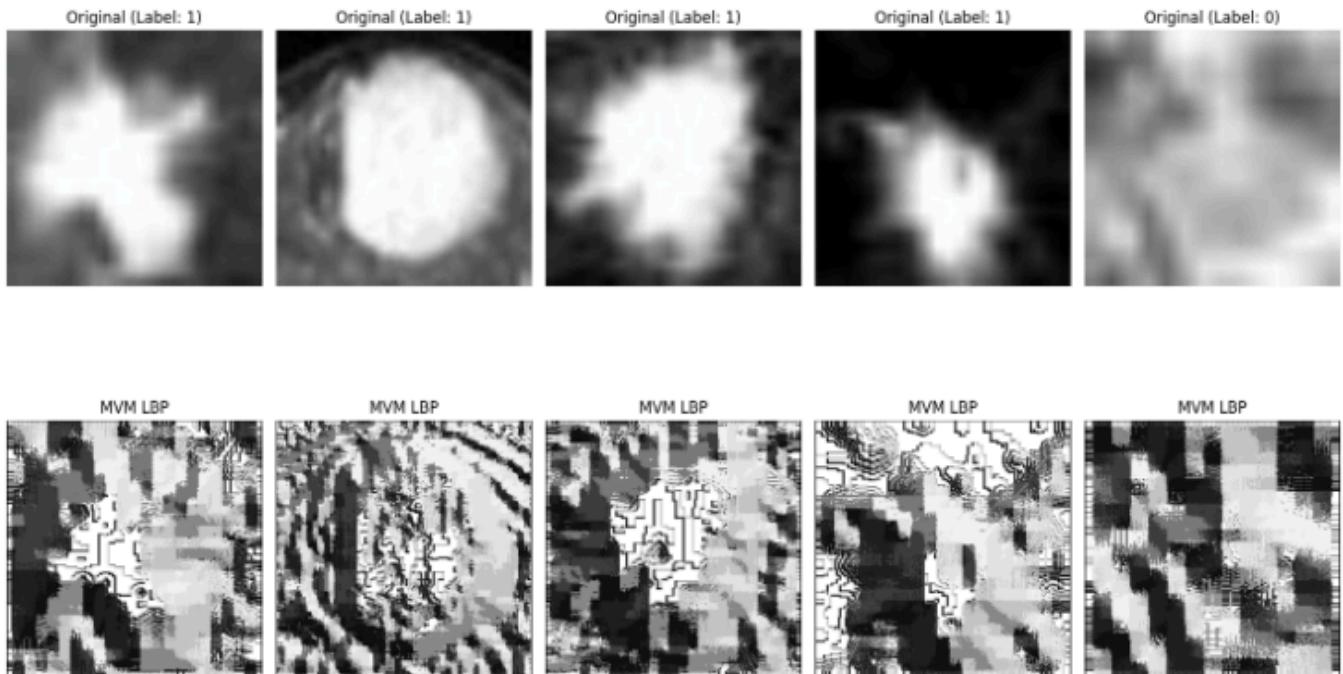


## LBP VARIANCE

The variance of Local Binary Pattern (LBP) feature values measures the degree of spread or variability in texture patterns across an image, offering critical insights into the complexity and diversity of the tumor's surface. A high LBP variance indicates substantial differences in local textures, which often correspond to the heterogeneous and irregular structures characteristic of malignant tumors. These irregularities reflect the chaotic growth patterns commonly associated with malignancy.

LBP variance serves as a valuable metric for distinguishing between benign and malignant tumors. It captures the subtle textural differences that are not easily discernible through direct visual inspection, highlighting the degree of irregularity in the tumor region. This measure of texture variability complements other statistical descriptors like the mean and median, providing a comprehensive analysis of texture patterns. By quantifying the complexity of tumor surfaces, LBP variance enhances the ability of diagnostic models to classify tumors accurately, supporting more precise and reliable clinical decision-making.

## MVM LBP



The mean, variance, and median of Local Binary Pattern (LBP) feature values are key statistical descriptors that provide a comprehensive understanding of texture patterns in breast cancer MRI analysis. The LBP mean represents the average texture intensity across the image, offering insights into the overall roughness or smoothness of the tumor surface. A higher mean typically indicates more complex and irregular textures, often associated with malignant tumors, while a lower mean reflects simpler and smoother textures, characteristic of benign tumors. The LBP variance, on the other hand, measures the spread or variability in texture patterns. High variance highlights significant heterogeneity and irregularity, which are commonly observed in malignant tumors, whereas low variance suggests uniform textures, generally found in benign tumors. Lastly, the LBP median provides the central value in the texture distribution, serving as a robust measure of the dominant texture patterns. Unlike the mean, the median is less influenced by outliers or extreme values, making it particularly useful for capturing the most frequent texture characteristics in a tumor region. Together, these metrics—mean, variance, and median—offer a detailed analysis of tumor texture, helping to distinguish between benign and malignant cases. This combination enhances the accuracy of classification models and supports precise diagnostic decisions in breast cancer imaging.

## GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM)

It is a technique that captures the spatial relationships between pixel intensities within an image. It analyzes how often pairs of pixels with specific values occur in a specified spatial relationship, providing a detailed description of texture by examining the patterns of intensity changes. GLCM is particularly valuable in extracting key texture properties such as contrast, which measures the difference in intensity between neighboring pixels; correlation, which quantifies the relationship between pixel values; energy, which reflects the uniformity of texture; and homogeneity, which indicates how similar neighboring pixel values are. These texture features are crucial for understanding the tissue structure, as they help in distinguishing between different types of tissues based on their distinctive texture patterns.

## MODEL DEVELOPMENT

The model development for tumor classification involves training a custom deep learning model that is fine-tuned on pre-existing architectures like VGG16, ResNet18, and ResNet50. These models are initialized with pre-trained weights, which are then adapted for the binary classification task by adding additional layers. The final model is designed to predict the likelihood of a tumor being benign or malignant. For efficient optimization, the Adam optimizer is employed, ensuring effective and adaptive weight updates during training. The binary nature of the classification task makes binary cross-entropy loss the most appropriate loss function, as it calculates the error between predicted probabilities and actual tumor classifications. Additionally, a learning rate scheduler is implemented to dynamically adjust the learning rate, which enhances convergence and improves the model's performance by ensuring efficient training throughout the process.

## TRAINING AND VALIDATION

The training process involves iteratively training the model over multiple epochs, with continuous validation to monitor performance. During each epoch, the model learns from the training set while its effectiveness is validated using the validation set. Metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) are tracked to ensure balanced performance across both classes (Benign and Malignant). Regularization techniques like early stopping and dropout are applied to prevent overfitting and ensure the model generalizes well.

### VGG16 Model:

- Epochs 48–50: The model achieved 99.97% training accuracy, with validation accuracy fluctuating between 83.56% and 83.86%. The AUC remained stable around 0.8795, but the best recorded AUC was 0.9037, indicating that the model's performance plateaued after a certain point.

### ResNet18 Model:

- Epochs 48–50: The model showed 100% training accuracy, but validation accuracy remained around 83.6% with a slight decrease in performance over the final epochs. The confusion matrix highlighted 302 benign (false negatives) and 106 malignant (false positives) predictions, with improved results in the later epochs. The model also had a learning rate scheduler, which helped adjust the learning rate in the final epoch.

### ResNet50 Model:

- Epochs 48–50: Like ResNet18, ResNet50 reached 100% training accuracy, but validation accuracy was in the range of 83.11% to 84.21%. The confusion matrix was similar, showing false negatives and false positives. Despite improvements in accuracy, AUC remained stable and reached a peak of 0.9298.

The best models from each architecture were saved, and the training showed that while all three models achieved near-perfect training accuracy, the validation accuracy and AUC values indicated some room for improvement. The ResNet50 model showed the best performance in terms of AUC, reaching 0.9298, suggesting it may have slightly better generalization capabilities compared to the others.

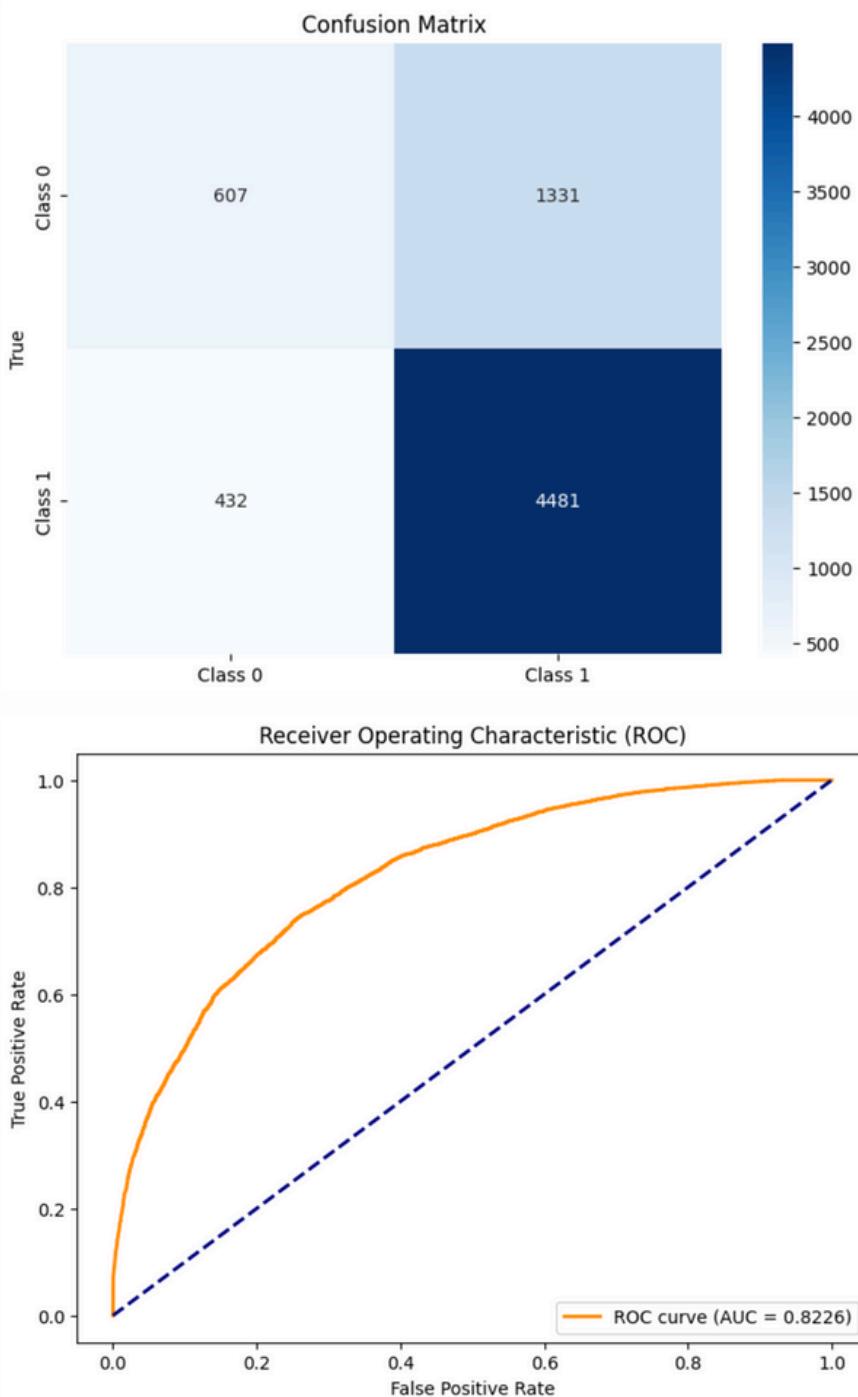
## MODEL EVALUATION

The trained models were evaluated on a separate test dataset consisting of 1,938 Benign and 4,913 Malignant images. The models' diagnostic performance was assessed using a range of metrics, including sensitivity, specificity, precision, recall, F1-score, and ROC-AUC. Special focus was given to sensitivity and specificity to ensure the reliable detection of both tumor types.

### VGG16 Model:

- Test Accuracy: 74.27%
- Test Loss: 1.2546
- Test AUC: 0.8226

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:

Test Accuracy: 74.34%

Test Loss: 1.1192

Test AUC: 0.8345

Confusion Matrix:

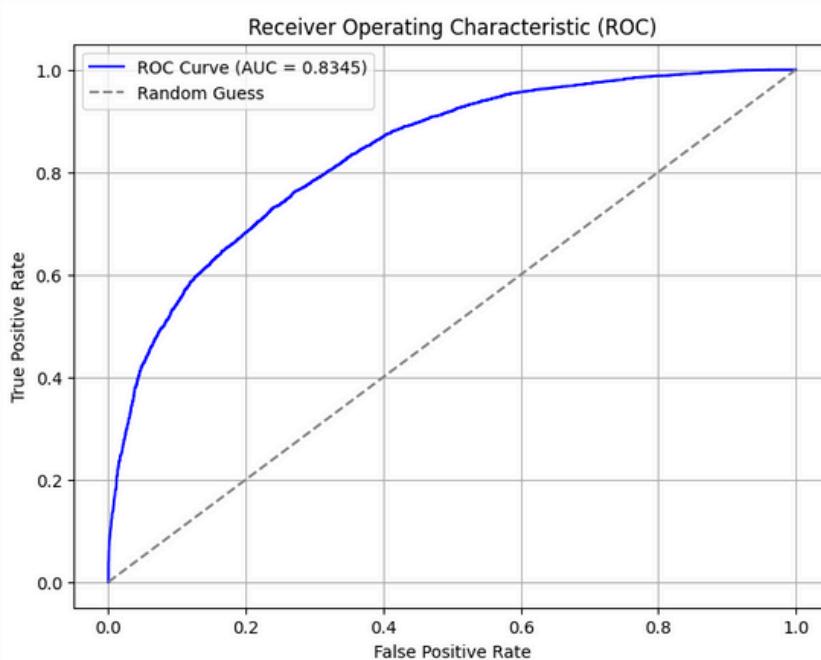
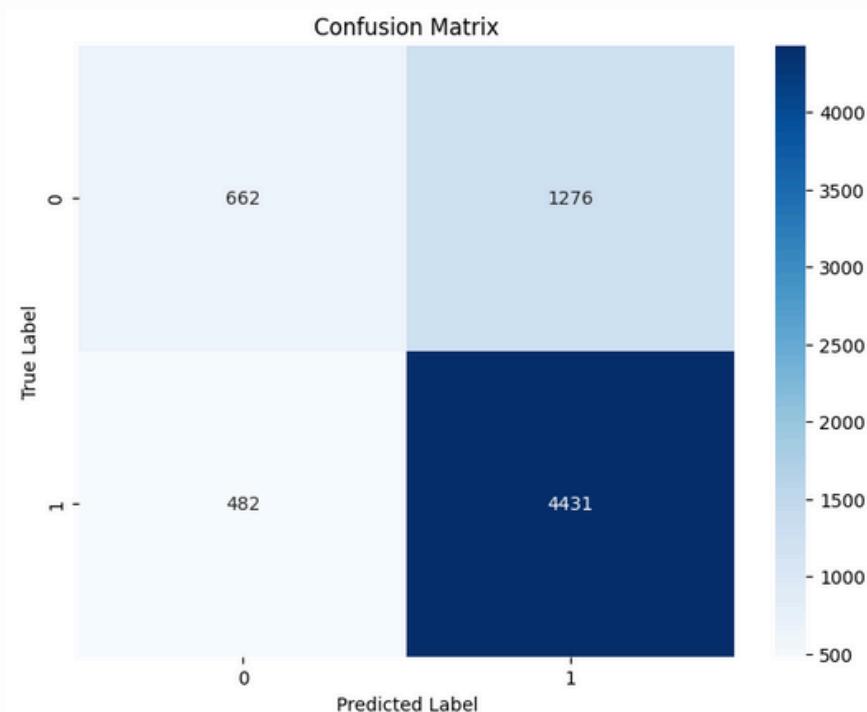
Benign (Predicted as Benign): 662

Malignant (Predicted as Malignant): 4431

False Negatives: 482

False Positives: 1276

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:

Test Accuracy: 77.54%

Test Loss: 4439.4057

Test AUC: 0.8549

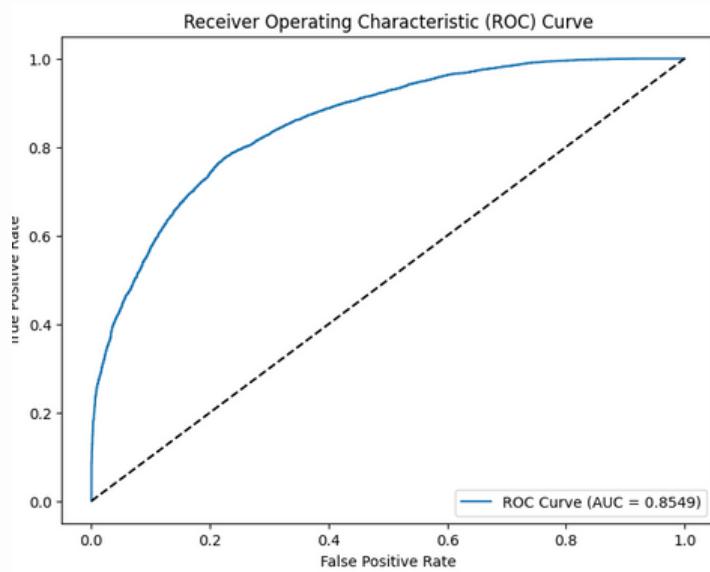
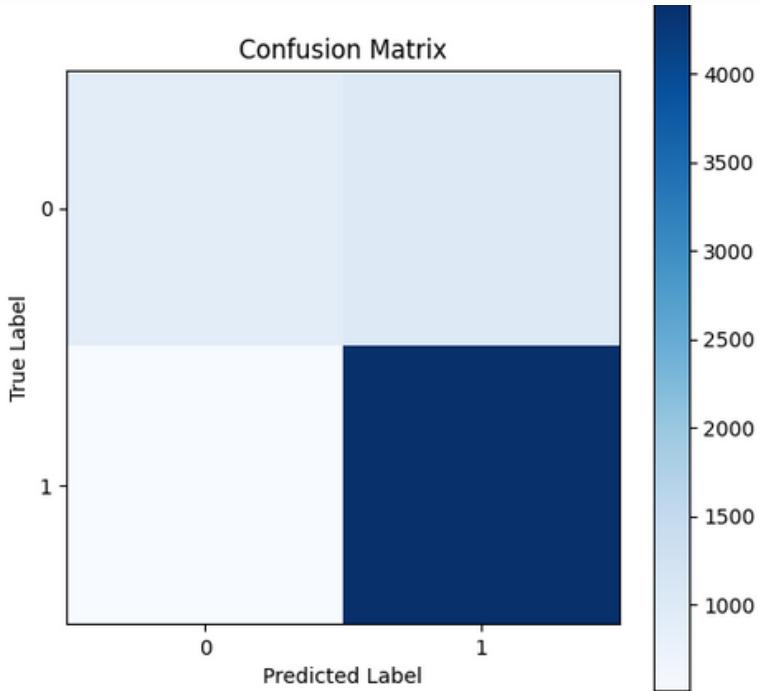
Confusion Matrix:

Benign (Predicted as Benign): 916

Malignant (Predicted as Malignant): 4396

False Negatives: 517

False Positives: 1022



The ResNet50 model achieved the highest accuracy and AUC among the three models, with the test accuracy reaching 77.54% and the AUC improving to 0.8549. 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 prepared for deployment in clinical settings to enable real-time classification of breast tumor MRI scans. This deployment focuses on ensuring the model's capability to operate efficiently on modern GPUs and edge devices, enabling rapid, on-site diagnosis in healthcare environments. To enhance diagnostic accuracy, provisions are included for future integration with multi-modal imaging data, such as mammograms and ultrasounds, allowing the system to make more comprehensive and informed classifications. The deployment is implemented using the ResNet50 model, fine-tuned for binary classification (Benign vs. Malignant). The model is optimized for inference on GPUs, allowing for fast processing of MRI images. The Gradio interface provides an easy-to-use web application where users can upload MRI scan images for real-time prediction. The model uses a pre-trained ResNet50 architecture with a final fully connected layer modified to output predictions for two classes (Benign or Malignant). The images are processed with standard transformations, including resizing to 224x224 pixels and normalization based on ImageNet's statistics, ensuring compatibility with the model's input requirements.

This interface allows users to upload an image of a breast tumor MRI scan, and the model will return a classification result with the predicted tumor type and the associated confidence level. The system is optimized for scalability, ensuring that it can handle multiple requests efficiently when deployed in clinical environments. This approach not only facilitates real-time decision-making but also enables the integration of future technologies and imaging modalities to enhance the diagnostic process.

### Benign or Malignant Classifier

Upload an image to classify it as benign or malignant.

 X  
Clear Submit Flag

output

Malignant

