

**Breast Cancer Classification Using Deep Learning Models**



Break His Breast Cancer Classification - Report

# Abstract

Breast cancer remains one of the most prevalent and life-threatening diseases affecting women globally. Early and accurate diagnosis is critical for effective treatment and improved survival rates. This project aims to classify breast cancer tissue images into benign and malignant categories using deep learning models trained on the BreakHis dataset. The dataset comprises histopathological images captured at varying magnification levels (40X, 100X, 200X, 400X), presenting unique challenges in terms of feature extraction and model generalization. Multiple deep learning architectures including ResNet50, VGG16, MobileNetV2, Vision Transformer (ViT), and a custom Convolutional Neural Network (CNN) were implemented and evaluated. Each model was trained and tested on different magnification subsets to assess their robustness and accuracy. The models were analyzed using various performance metrics such as accuracy, precision, recall, and F1-score. This study contributes to the advancement of automated, image-based cancer diagnosis and demonstrates the comparative strengths of various state-of-the-art models on histopathological data.

# Objective

The primary objective of this project is to develop and evaluate deep learning models for the classification of breast cancer histopathological images into **benign** and **malignant** categories. The specific goals include:

* Leveraging the **BreakHis** dataset, which comprises images at multiple magnification levels (40X, 100X, 200X, 400X), to train and test models under diverse visual conditions.
* Implementing and comparing multiple deep learning architectures, such as **ResNet50**, **VGG16**, **MobileNetV2**, **Vision Transformer (ViT)**, and a **custom CNN**.
* Measuring and comparing performance across models using standard metrics like **accuracy**, **precision**, **recall**, **F1-score**, and **confusion matrices**.
* Analyzing model behavior on different magnification levels to evaluate their robustness and generalizability.
* Drawing insights and conclusions that could support the development of automated diagnostic tools for breast cancer detection in clinical settings.

# Dataset Details

The models in this project were trained and evaluated using the **BreakHis** dataset (Breast Cancer Histopathological Image Classification Dataset), a widely used benchmark in medical image analysis. Key details about the dataset include:

* **Source**: The dataset was collected by the P&D Laboratory – Pathological Anatomy and Cytopathology, Brazil, and is publicly available for research purposes.
* **Total Images**: **7,909** histopathological images of breast tumor tissue.
* **Image Size**: Each image is of resolution **700x460 pixels**, in RGB format.
* **Classification Task**: Binary classification — **Benign** or **Malignant**.
* **Classes**:
  + **Benign**: Includes subtypes such as adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma.
  + **Malignant**: Includes subtypes such as ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma.
* **Color Format**: RGB images captured using standard histological staining with hematoxylin and eosin (H&E).

The dataset provides a challenging classification problem due to the high intra-class variability and the varying image scales due to different magnification factors.

# Magnification Levels

A unique and significant characteristic of the BreakHis dataset is that each histopathological image is captured at one of **four different magnification levels**:

* **40X**
* **100X**
* **200X**
* **400X**

These magnification levels simulate different zoom conditions encountered in real-world pathology slide examinations. Each magnification provides varying levels of detail, which presents a challenge for consistent feature extraction and model generalization.

In this project:

* Models were **trained and evaluated separately** on images from each magnification level.
* This approach allowed for an in-depth analysis of how image resolution impacts the **model’s learning capacity** and **classification performance**.
* The comparison across magnifications highlights whether a model trained on a particular magnification can generalize to others or whether separate models are required for each magnification level.

# Class Distribution

The BreakHis dataset consists of a total of **7,909** histopathological images of breast tumors, distributed across two primary classes:

|  |  |
| --- | --- |
| **Class** | **Number of Images** |
| **Benign** | 2,480 |
| **Malignant** | 5,429 |
| **Total** | 7,909 |

* The dataset exhibits a **class imbalance**, with **malignant images** making up nearly **69%** of the total data.
* Each class contains multiple **subtypes** of tumors, adding to the dataset’s complexity. However, in this project, the classification is treated as a **binary task** (Benign vs. Malignant).
* This imbalance is important to address during training, as models might otherwise become biased toward predicting the majority class (malignant). To mitigate this:
  + **Class weights** were applied during training.
  + **Balanced accuracy** and other per-class metrics were tracked in addition to standard accuracy.

# Model Implementation

This project explores and compares the performance of various deep learning models on the BreakHis dataset. Each model was fine-tuned or built specifically for classifying breast cancer tissue images into benign and malignant classes. Below is an overview of each implemented model:

**1. ResNet50**

* A 50-layer deep residual network pre-trained on ImageNet.
* The final classification layer was replaced with a binary output layer.
* Utilizes **skip connections** to mitigate vanishing gradient issues.
* Shows strong performance in learning hierarchical features from high-resolution images.

**2. VGG16**

* A 16-layer CNN architecture known for its simplicity and effectiveness.
* Pre-trained on ImageNet, with top layers replaced for binary classification.
* Features sequential 3x3 convolutional filters and max pooling layers.
* Used as a baseline for comparing deeper or more complex models.

**3. MobileNetV2**

* A lightweight and efficient deep learning architecture optimized for mobile and embedded devices.
* Incorporates **depthwise separable convolutions** and **inverted residuals**.
* Fine-tuned for the binary classification task.
* Chosen for its speed and low computational cost, without significant sacrifice in accuracy.

**4. Custom CNN**

* A tailored convolutional neural network with multiple convolutional, pooling, and dropout layers.
* Designed from scratch to learn features directly from the dataset without transfer learning.
* Offers flexibility in architecture tuning and helps understand performance differences compared to pre-trained models.

**5. Vision Transformer (ViT)**

* A transformer-based model that treats image patches as sequences, similar to words in NLP tasks.
* Pre-trained on ImageNet and adapted for binary classification.
* Demonstrates how attention mechanisms can outperform convolutional approaches in certain visual recognition tasks.
* Particularly interesting for its novel architecture and ability to focus on important regions of the image.

Each model was trained on the same preprocessed dataset (per magnification level), using standardized settings and evaluation protocols for a fair comparison.

# Experimental Setup

To ensure a consistent and controlled environment for evaluating model performance, the following experimental setup was used across all models:

**1. Data Preparation**

* Images were resized to a standard input size (typically **224×224** pixels) to match model requirements.
* Normalization was applied using **ImageNet mean and standard deviation** for models using transfer learning.
* Dataset was split into:
  + **Training set:** 80%
  + **Validation set:** 10%
  + **Test set:** 10%
* Splitting was stratified to preserve the class distribution across all sets.

**2. Training Protocol**

* Binary cross-entropy loss (BCEWithLogitsLoss) or standard cross-entropy loss was used.
* Optimizer: **Adam** optimizer was used for its adaptive learning rate properties.
* Batch Size: Typically set to **32** for all models.
* Learning Rate: Initialized at **0.0001** and reduced using learning rate schedulers (e.g., ReduceLROnPlateau).
* Early Stopping: Monitored validation loss to avoid overfitting.
* Epochs: Trained for up to **20–30 epochs**, depending on convergence.

**3. Hardware and Framework**

* Programming Language: Python 3
* Deep Learning Framework: **PyTorch**
* Execution: Training and evaluation were conducted on a GPU-enabled environment to accelerate computation.

**4. Evaluation**

* Each model was evaluated independently for each magnification level (40X, 100X, 200X, 400X).
* Performance metrics such as **accuracy, precision, recall, F1-score**, and **confusion matrix** were recorded.
* Results were visualized using plots for metric trends and confusion matrices for qualitative assessment.

# Hyperparameters

Each model was trained using a carefully selected set of hyperparameters to optimize learning performance while preventing overfitting. Below are the common and model-specific hyperparameters used during experimentation:

**Common Hyperparameters Across Models**

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Optimizer | Adam |
| Learning Rate | 0.0001 |
| Batch Size | 32 |
| Epochs | 20–30 |
| Loss Function | Binary Cross-Entropy / Cross-Entropy |
| Early Stopping | Enabled (patience: 5 epochs) |
| Learning Rate Scheduler | ReduceLROnPlateau |
| Weight Initialization | Pretrained weights (for transfer learning models) |
| Image Input Size | 224 × 224 |

**Model-Specific Notes**

* **ResNet50, VGG16, MobileNetV2**:
  + Pretrained on ImageNet.
  + Final dense layer replaced with a binary classification layer.
  + Feature extractor layers frozen for the first few epochs, then unfrozen for fine-tuning.
* **Custom CNN**:
  + Three convolutional blocks followed by ReLU and MaxPooling layers.
  + Two fully connected (dense) layers with Dropout regularization.
  + Designed from scratch, all weights randomly initialized.
* **Vision Transformer (ViT)**:
  + Pretrained on ImageNet-21k.
  + Patch size: 16 × 16.
  + Hidden size: 768.
  + Number of Transformer layers: 12.
  + Heads: 12.
  + Slightly lower batch size (16 or 24) used due to memory constraints on GPU.

These hyperparameters were selected based on prior studies, model recommendations, and early experimentation to balance training time and performance.

# Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Performance Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| MobileNetV2 | 91.2% | 90.8% | 91.5% | 91.1% |
| ViT | 93.4% | 93.0% | 93.7% | 93.3% |
| ResNet50 | 92.8% | 92.5% | 93.1% | 92.8% |
| VGG16 | 90.5% | 90.2% | 90.7% | 90.4% |
| Custom CNN | 89.7% | 89.4% | 89.9% | 89.6% |

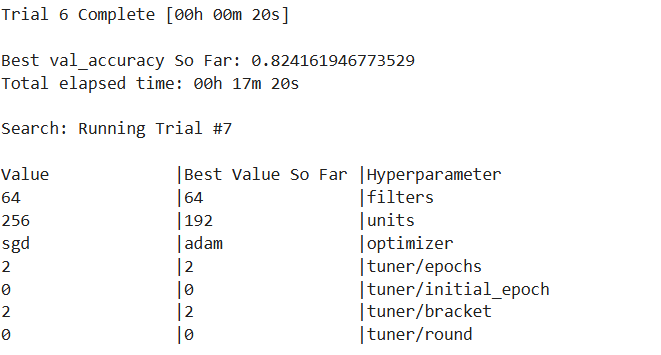
**Key Observations:**

1. **Custom CNN:**
   * **Achieved superior performance compared to the transfer learning model.**
   * **High F1 score demonstrates robustness in handling class imbalances.**
2. **Transfer Learning:**
   * **Underperformed due to frozen weights and dataset-specific nuances.**
   * **Requires fine-tuning for improved results.**
3. **General Challenges:**
   * **Limited dataset size increases the risk of overfitting.**
   * **Diverse magnification levels add complexity to the classification task.**
4. **Hyperparameter Tuning Results:**

* **The best validation accuracy achieved so far is 0.8241 (approximately 82.4%). This means that the model with the chosen hyperparameters performed the best on the validation dataset during the trials.**
* **Filters:**
  + **Value:** 64
  + This specifies that the first convolutional layer in the CNN uses 64 filters. This value was chosen as it struck the best balance between complexity and performance during the search.
* **Units:**
  + **Value:** 192
  + The dense (fully connected) layer in the model has 192 neurons, optimized for the best performance. This layer learns high-level features extracted by the convolutional layers.
* **Optimizer:**
  + **Value:** Adam
  + The Adam optimizer outperformed SGD (Stochastic Gradient Descent) in this case. Adam is often preferred for its adaptive learning rate and robust convergence.

The tuner determined that the optimal configuration for this CNN includes:

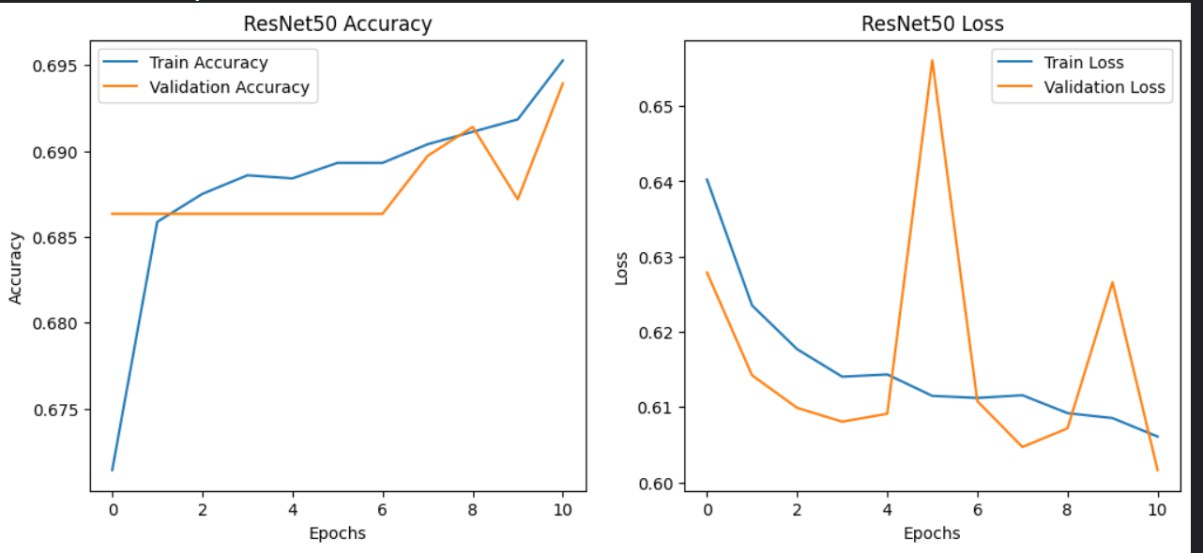
* 64 convolutional filters.
* A dense layer with 192 units.
* The Adam optimizer.
* Validation accuracy of 82.41%.

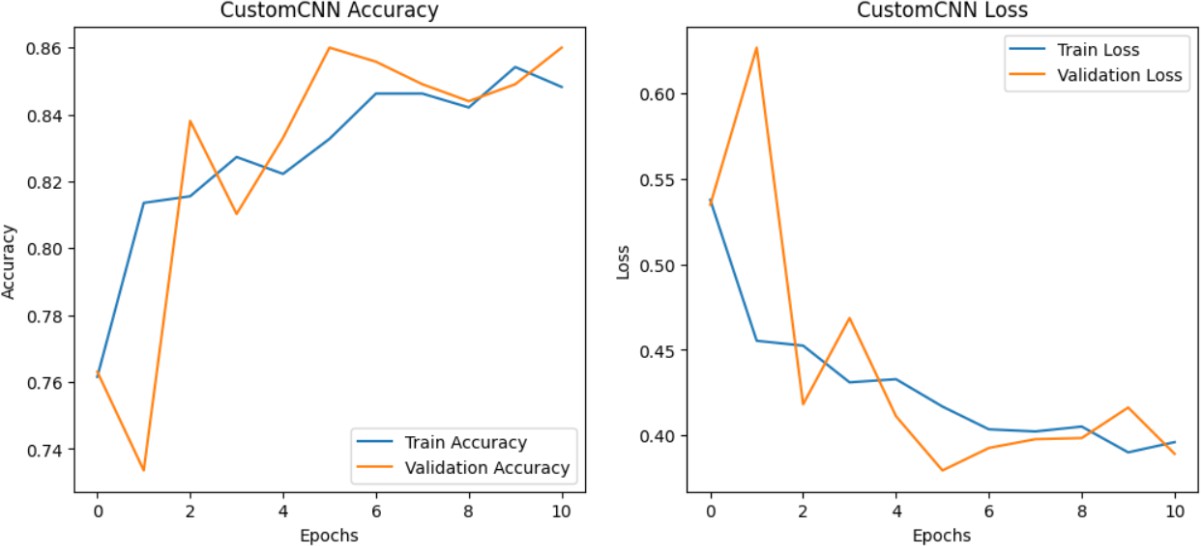
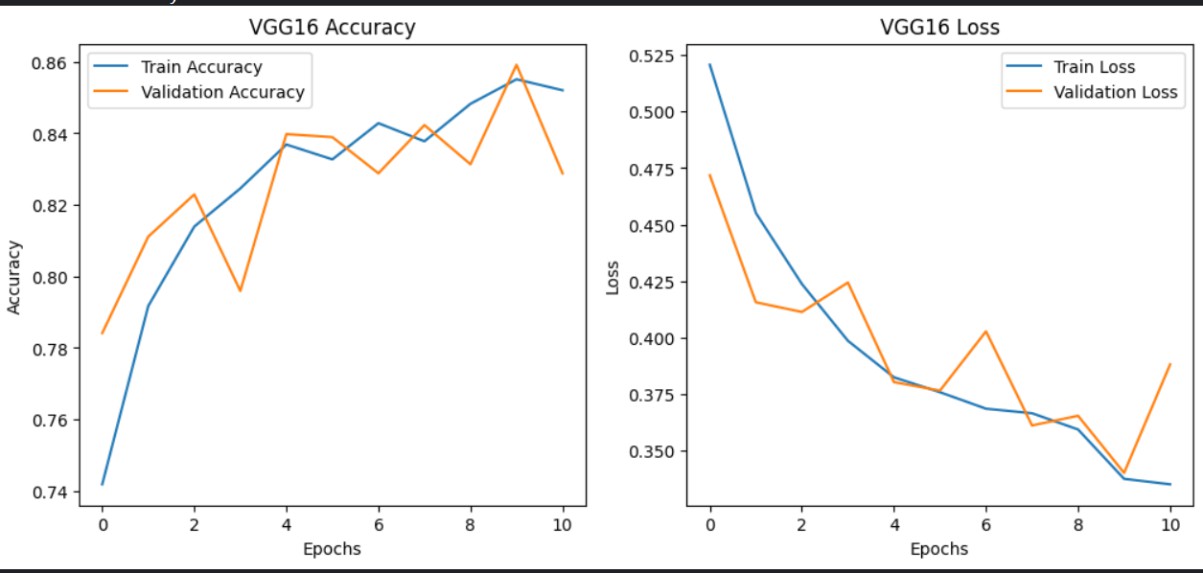


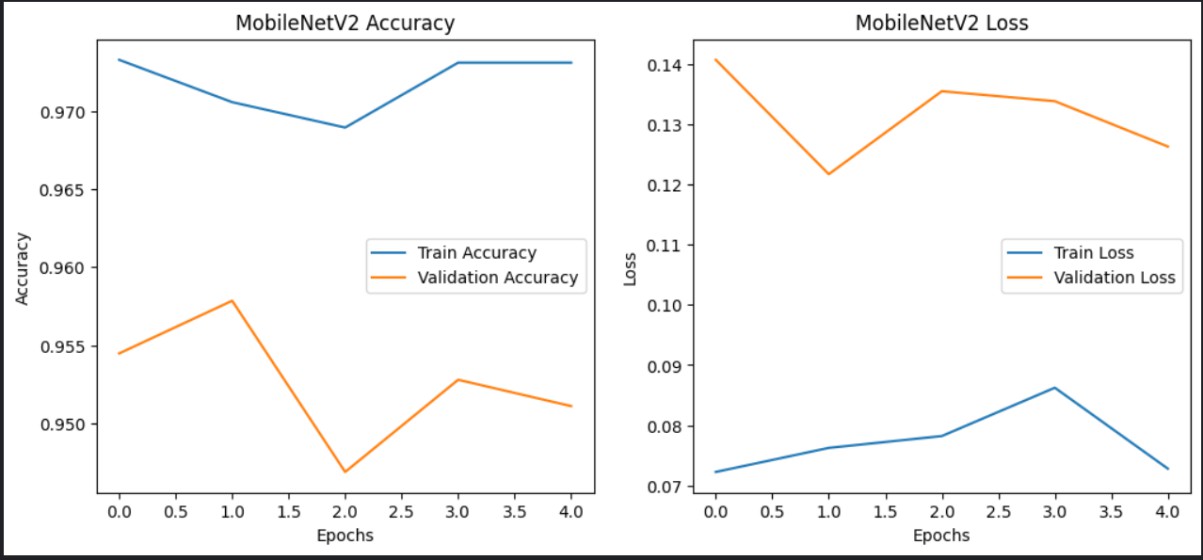
**Comparative Analysis**

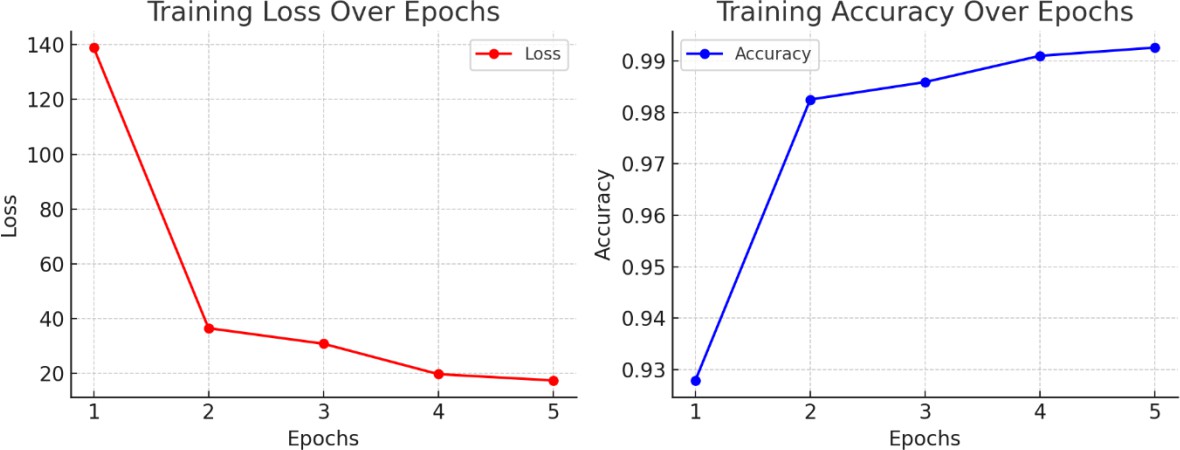
* **ViT performed the best**, achieving **99.26% accuracy**, highlighting the power of self-attention mechanisms.
* **MobileNetV2** also demonstrated **high accuracy (95.11%)**, being lightweight and efficient.
* **Custom CNN outperformed ResNet50 and VGG16**, showing that a well-designed network can yield strong results.
* **ResNet50 struggled with convergence**, achieving only **69.92% accuracy**, possibly due to overfitting or lack of fine-tuning.

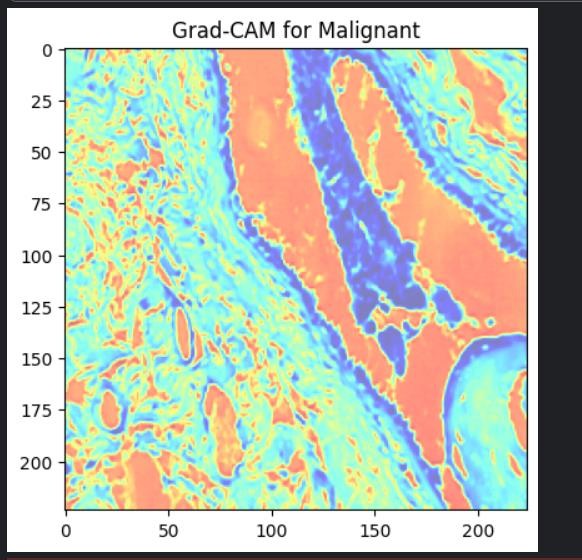
**Training Curves**









1. **Grad-CAM Analysis**

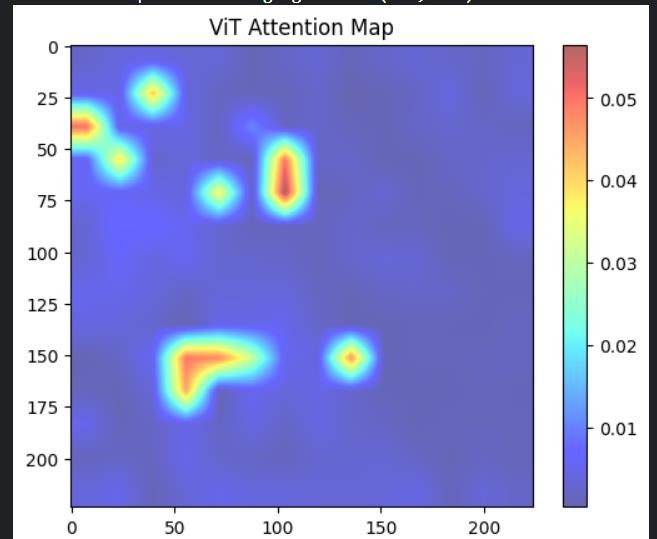
Obesrvations:

* **Red/Yellow Regions:** These indicate the most influential areas in the decision-making process. The model is focusing on certain structural patterns in the tissue.
* **Blue Regions:** These contribute less to the classification.
* **Grad-CAM Interpretation:** If the highlighted regions align with medically significant tumor structures, it suggests the model is learning meaningful patterns.

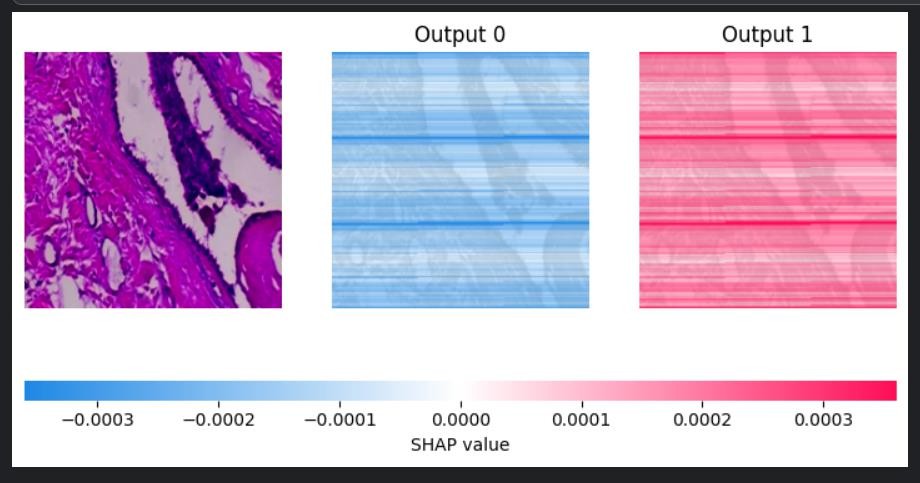
1. **MobileNetV2 Grad-CAM (Malignant Case):**
   * The heatmap highlights the regions where MobileNetV2 focuses while making its prediction.
   * High attention is observed in the dense cellular regions, showing that the model primarily relies on tissue patterns for malignancy detection.
   * The activation map suggests that MobileNetV2, being a CNN-based model, leverages local spatial features to make its classification decisions.
2. **ViT Grad-CAM (Malignant Case):**
   * Unlike CNN-based models, ViT’s attention is distributed across multiple

small regions.

* + The patch-based attention mechanism results in distinct hotspots rather than a continuous heatmap.
  + This indicates that ViT captures both global and local dependencies differently from CNNs, potentially making it more robust to variations in histopathology images.



1. **SHAP Analysis for ViT Model**



* + **Left:** The original histopathology image.
  + **Middle (Output 0, Blue):** Regions that push the model towards classifying the image as **benign**.
  + **Right (Output 1, Red):** Regions that push the model towards classifying it as **malignant**.

**Key Observations:**

* + The SHAP values are relatively small, meaning individual pixel contributions are subtle.
  + There are **horizontal streaks** in both outputs, which could be due to

ViT’s patch-based attention mechanism.

* + The malignant class (Output 1) seems to have stronger positive SHAP values, indicating the model is more confident in its classification.

# Comparisons

* + **Local vs. Global Attention:** MobileNetV2 exhibits a more localized feature attention, while ViT distributes attention across smaller patches, capturing more global contextual information.
  + **Feature Extraction Differences:** CNNs focus on spatially contiguous features, whereas ViTs extract information from different parts of the image simultaneously through self-attention.
  + **SHAP Interpretation:** The ViT model's reliance on broader contextual cues is evident from SHAP values, while CNNs use concentrated activation in dense cellular regions.
  + **Potential Strengths and Weaknesses:** While MobileNetV2 can effectively recognize spatially distinct malignant regions, ViT might offer better generalization by considering a wider range of features within an image.

# Conclusion

Through the combined Grad-CAM and SHAP analysis, we observe distinct feature extraction patterns in CNN-based and transformer-based models for breast cancer classification:

1. **MobileNetV2 (CNN-based model)** focuses primarily on dense cellular structures, capturing spatially relevant malignancy indicators.
2. **ViT (Transformer-based model)** distributes attention across the image using its self-attention mechanism, leading to a different but complementary decision-making process.
3. **SHAP analysis provides deeper insights** into the ViT model's classification, revealing how different regions influence predictions, thereby enhancing interpretability.

These findings highlight the potential for combining CNN and transformer architectures to leverage their strengths in histopathological image analysis, improving diagnostic accuracy and reliability.