Table of Content

1. Introduction

Breast cancer is the most frequently diagnosed cancer and one of the leading causes of cancer-related death among women globally. According to the World Health Organization (WHO), over 2.3 million women were diagnosed with breast cancer in 2020, 11.7% of all new cancer cases. Despite advances in treatment options, early detection remains the most effective strategy for reducing mortality. If detected early, breast cancer has a survival rate exceeding 90%. However, as the disease progresses to later stages, prognosis and treatment efficacy decline sharply. Thus, timely and accurate diagnosis is not just preferred but essential.

In clinical practice, breast cancer detection involves a combination of physical examinations, imaging and histopathological analysis of biopsy samples. Mammography is the standard screening method for early-stage detection, while histopathology provides microscopic validation of malignancy. Despite their clinical value, these methods are not without limitations. First, the interpretation of mammographic or histopathological images is highly subjective, relying on the expertise of radiologists and pathologists. This introduces variability in diagnosis, especially in cases with subtle or ambiguous visual patterns. Second, both modalities are time-consuming, resource-intensive, and often unavailable in low-resource healthcare settings. Third, the growing number of screenings due to awareness campaigns and aging populations places additional pressure on diagnostic services, increasing the likelihood of diagnostic delays and errors.

These limitations have sparked global interest in the integration of Artificial Intelligence and more specifically, deep learning into computer-aided diagnosis (CADx) systems. Deep learning offers the ability to automatically learn complex features from medical images, eliminating the need for manual feature engineering and potentially matching or exceeding expert-level performance. Over the past decade, deep learning models, especially Convolutional Neural Networks, have become the cornerstone of automated medical imaging analysis. Their hierarchical feature extraction capabilities allow them to learn low-level (edges, textures) and high-level (shapes, regions) representations directly from pixel data, making them suitable for tasks such as lesion detection, classification, and segmentation.

While CNNs have been the backbone of most medical imaging models, the emergence of Vision Transformers (ViTs) has introduced a fundamentally different approach. Inspired by the success of transformers in natural language processing, ViTs divide input images into fixed-size patches and process them through self-attention mechanisms, capturing long-range

dependencies and global context more effectively than traditional CNNs. This architectural shift enables ViTs to model spatial relationships across an entire image, making them particularly promising for analyzing medical images with complex structural patterns, such as mammograms or histopathological slides.

However, applying AI to diagnosing breast cancer is not without its challenges. One key issue is dataset bias. Many publicly available breast cancer datasets exhibit class imbalance, with disproportionate representation of either benign or malignant samples. This leads to model overfitting and bias toward the majority class, resulting in high sensitivity but low specificity, or vice versa. Another challenge is domain shift between datasets models trained on one imaging modality often perform poorly when applied to another. Additionally, limited dataset size, lack of interpretability, and variability in imaging quality further complicate the development of clinically robust AI systems.

What problem is our study solving?

In this study, we aim to address several of these challenges by conducting a comprehensive evaluation of deep learning models across diverse breast cancer imaging datasets. Our primary objective is to perform binary classification of breast cancer images into benign and malignant categories using both convolutional and transformer-based architectures. By applying these models to different imaging modalities, histopathological slides and mammograms, we seek to assess their generalization capabilities, sensitivity to class imbalance, and robustness to domain variation.

The study utilizes three publicly available datasets:

- **BreakHis**: A histopathological image dataset consisting of 7,909 microscopic biopsy samples at varying magnification levels (40×, 100×, 200×, 400×). BreakHis provides high-resolution cellular detail essential for fine-grained classification.
- **INBreast**: A digital mammography dataset containing full-field images with expert-annotated lesions. It is widely regarded as a benchmark for radiological breast cancer research due to its quality and clinical realism.
- **DDSM (Digital Database for Screening Mammography)**: A film-based mammography dataset with accompanying pathology reports, offering large-scale diversity in patient demographics and image quality.

Each of these datasets represents unique clinical contexts and challenges. BreakHis images are rich in texture and color but suffer from class imbalance and limited sample size per subtype. InBreast offers high-quality grayscale images with well-structured annotations, making it ideal for ViT-based modeling. DDSM introduces real-world complexity through image artifacts, digitization noise, and label uncertainty by providing a rigorous test of model robustness.

What are we trying to achieve?

To evaluate classification performance, we experiment with three model families:

- **ResNet50**: A widely-used 50-layer residual network that mitigates vanishing gradients and enables deep feature learning. It is pretrained on ImageNet and fine-tuned on medical images.
- **CustomVGG**: The VGG architecture is a deep convolutional neural network known for its simplicity and effectiveness in image classification tasks. Despite being computationally intensive, VGG achieves high accuracy and serves as a strong baseline for many vision problems.
- **EfficientNetB3**: An architecture that balances model depth, width, and resolution to achieve high accuracy with fewer parameters and less computational cost.
- **Vision Transformer (ViT)**: A transformer-based model that splits images into patches and uses attention mechanisms to model relationships between regions, potentially capturing more global and contextual features.

All models are adapted for binary classification, with modifications to their output layers and training objectives. We apply preprocessing steps tailored to each model, including resizing (224×224 for ViT, VGG and ResNet; 300×300 for EfficientNet), pixel normalization to the [0,1] range, and data augmentation (rotation, zooming, flipping, and shearing). Labels are encoded using directory structures via automated pipelines such as flow_from_directory or image_dataset_from_directory.

The study employs standardized evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix to analyze model performance. Special emphasis is placed on recall for malignant cases, as minimizing false negatives is paramount in cancer detection. We also investigate common issues such as overfitting, class imbalance, and model bias, especially where models favor the malignant class due to uneven sample distributions.

What have we done and why is it significant?

Our contributions are threefold:

1. We provide a comparative evaluation of multiple deep learning architectures across three clinically relevant breast cancer imaging datasets.
2. We identify and analyze model-specific biases, especially regarding benign class performance and cross-domain limitations.
3. We offer insight into the practical challenges of building real-world AI systems for medical imaging, such as the need for better domain adaptation, interpretability tools and balanced dataset sampling.

Through this work, we seek to demonstrate that while deep learning holds great promise for breast cancer diagnosis, its reliability depends heavily on data quality, model design, and

evaluation rigor. The findings serve as a reference point for future research aiming to deploy AI-assisted diagnostic tools in clinical workflows not just for high performance, but for clinical safety and trust.

2. Dataset Description

2.1. BreakHis

The Breast Cancer Histopathological Image Classification (BreakHis) dataset comprises 9,109 microscopic images of breast tumor tissue from 82 patients, captured at various magnifications (40X, 100X, 200X, 400X). It contains 2,480 benign and 5,429 malignant samples (700x460 pixels, 3-channel RGB, 8-bit depth, PNG format). This database was built in collaboration with the P&D Laboratory – Pathological Anatomy and Cytopathology, Parana, Brazil .

The dataset categorizes tumors as benign or malignant. Benign tumors exhibit slow growth and remain localized, whereas malignant tumors (cancer) can invade and metastasize. Samples were acquired through surgical excision (SOB method), a procedure involving the removal of larger tissue samples under general anesthetic.

Tumors are further classified based on their microscopic appearance. Benign types include adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma. Malignant types encompass carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. Each image filename encodes details such as biopsy method, tumor class, tumor type, patient identification, and magnification factor.

2.2. DDSM

CBIS-DDSM (Curated Breast Imaging Subset of DDSM) stands as a significant and meticulously organized dataset in the realm of medical imaging research, specifically designed for advancements in breast cancer detection and diagnosis. It represents a thoroughly refined and standardized iteration of the foundational Digital Database for Screening Mammography (DDSM).

The original DDSM, while groundbreaking for its time, presented certain challenges that hindered consistent research and replication. The CBIS-DDSM was developed to directly address these limitations, offering a more robust and user-friendly resource for the scientific community.

At its core, the CBIS-DDSM comprises 2,620 scanned film mammography studies. These studies encompass a comprehensive range of cases, categorized as normal, benign (non-cancerous), and malignant (cancerous). A critical feature distinguishing the CBIS-DDSM is the verified pathology accompanying each case, ensuring the accuracy of diagnoses and providing a reliable ground truth for research. This rigorous verification is paramount for developing and evaluating diagnostic algorithms.

CBIS-DDSM actually contains data from 1,566 unique participants. This distinction is vital for researchers planning cohort studies or analyzing patient-level trends, ensuring accurate statistical analysis and avoiding misinterpretations regarding the study population size. This careful curation of patient identifiers underscores the commitment to data integrity within the CBIS-DDSM.

2.3. INBreast

The INbreast mammography image database, collected from Centro Hospitalar de S. Joao (CHSJ), Breast center, Porto, contains 115 cases with a total of 410 images gathered between August 2008 and July 2010. Among these, 90 cases are women with disease in both breasts.

The database records four types of breast diseases: Mass, Calcification, Asymmetries, and Distortions. Images are captured from two perspectives: Craniocaudal (CC) and Mediolateral Oblique (MLO). Breast density is categorized into four BI-RADS standards: Entirely fat (Density 1), Scattered fibroglandular densities (Density 2), Heterogeneously dense (Density 3), and Extremely dense (Density 4). Images are saved in DICOM format with sizes of either 3328 x 4084 or 2560 x 3328 pixels.

Through data augmentation, the number of breast mammography images was increased to 7632. Examples of these images demonstrate breast masses with benign or malignant status across the four density categories. Malignant masses are noted to have irregular shapes compared to benign masses.

3. Literature Review: Deep Learning for Breast Cancer Diagnosis

3.1. Introduction

Applying artificial intelligence, especially deep learning, has become a cornerstone of modern research in medical image analysis. In the context of breast cancer, these technologies offer transformative potential to improve diagnostic accuracy, reduce pathologist workload, and enable earlier detection. The reviewed literature highlights several

key themes: the foundational role of specialized datasets, the comparative performance of various model architectures, the rise of hybrid models, and the adaptation of these models across different imaging modalities like histopathology, mammography, and ultrasound. However, while significant progress has been made, critical gaps remain in understanding the comparative performance of the latest architectures on specific, challenging datasets.

3.2. The Foundational Role of Public Datasets

A significant catalyst for research in this field has been the creation of large, publicly available datasets. The work by Spanhol et al. (2016) [11], which introduced the BreaKHis dataset, is a foundational contribution. This dataset, containing thousands of histopathological images at various magnification factors, provided a standardized benchmark that has enabled countless subsequent studies. The availability of such well-annotated datasets is crucial, as it allows for direct comparison of different algorithmic approaches and promotes reproducible research. In their follow-up work, Spanhol, Oliveira, Petitjean, and Heutte (2016) [12] demonstrated the efficacy of Convolutional Neural Networks (CNNs) on this dataset. By leveraging pre-trained CNNs as feature extractors, they established that deep learning models could achieve high classification accuracy, setting the stage for more advanced and specialized architectures.

3.3. The Evolution from Traditional ML to Deep Learning

Early approaches to computer-aided diagnosis relied on traditional machine learning models that required manual, expert-driven feature engineering. However, the field has seen a decisive shift towards deep learning. A comparative analysis by Singh et al. (2024) [10], which evaluated supervised ML models (like Support Vector Machines) against DL models, aligns with this trend. Their findings confirm that deep learning models, particularly CNNs, consistently demonstrate superior performance. The primary advantage of DL lies in its ability to automatically learn a hierarchy of discriminative features directly from the image data, a capability that traditional ML models lack. This not only improves accuracy but also streamlines the development pipeline, as highlighted by Iqbal et al. (2024) [4], who emphasize the goal of creating robust, end-to-end diagnostic systems. Even classic, powerful CNN architectures like VGG-19 continue to be highly effective when empowered with transfer learning, as shown by Tazeen et al. (2024) [8], reinforcing the robustness of foundational deep learning models

3.4. Architectural Innovations and Modality-Specific Adaptations

As the field has matured, research has moved beyond generic models to explore specialized architectures tailored to the unique challenges of different imaging modalities.

- Advanced CNNs in Histopathology and Ultrasound: For the analysis of histopathological images and ultrasound, deep CNNs remain a powerful tool. Research has focused on applying progressively more sophisticated architectures. Karthik et al. (2024) [6] demonstrated the successful application of Inception-ResNet-v2 for the early detection of breast cancer in ultrasound images, while Alruwaili & Al-Ghamdi (2024) [1] leveraged EfficientNet. These studies show a clear trend of adapting state-of-the-art architectures from general computer vision to achieve high accuracy in specific medical imaging contexts.
- The Rise of Vision Transformers (ViT) for Mammography: While CNNs excel at learning local features, their inherent locality can be a limitation. The Vision Transformer (ViT) has emerged as a powerful alternative. Gautam et al. (2024) [2] investigated the use of ViTs for mammographic image classification, arguing that the self-attention mechanism is uniquely suited to this modality. Unlike CNNs, ViTs can capture global, long-range dependencies across the entire image. This global context is critical for interpreting mammograms, where subtle, non-local architectural distortions can be the primary indicator of malignancy.
- Hybrid Architectures and Advanced Augmentation: A growing body of research explores combining different types of neural networks to leverage their complementary strengths. El-Alami (2024) [9] proposed a hybrid model combining a CNN (EfficientNet-B0) with a Bi-directional Long Short-Term Memory (Bi-LSTM) network, suggesting that sequential modeling can capture valuable patterns in the data. Similarly, Al Heety & Al-Ani (2024) [7] introduced UGGNet, a model that bridges the segmentation capabilities of U-Net with the classification power of VGG. Furthermore, to address the persistent challenge of class imbalance, researchers are moving beyond simple geometric augmentations. K et al. (2024) [5] successfully amalgamated Generative Adversarial Networks (GANs) with a ResNet model, using GANs to generate realistic synthetic images of the minority class, thereby creating a more balanced and robust training set.

3.5. Identifying the Research Gap

The existing literature clearly establishes the superiority of deep learning over traditional methods and highlights a vibrant area of research into architectural innovation, including hybrid models and advanced data augmentation. However, a critical review reveals a distinct gap: while ViTs have shown promise for mammography and various advanced CNNs have been proven on ultrasound and histopathology, there is a lack of direct, comparative studies

between the foundational state-of-the-art architectures—namely, a Vision Transformer versus a highly efficient CNN like EfficientNet—on the same challenging histopathological dataset, such as BreaKHis. It remains unclear whether the global context captured by ViTs offers a definitive advantage over the optimized feature extraction of modern CNNs for classifying the complex patterns in histopathology slides.

This study, therefore, aims to address this gap by conducting a rigorous comparative analysis of a custom-built Vision Transformer and a fine-tuned EfficientNetB3 model, a VGG and a ResNet model on theDDSM, INBreast and BreaKHis datasets. The objective is to determine which architecture provides superior diagnostic accuracy for histopathological images and to analyze their respective strengths and weaknesses in this specific context before introducing further complexities like hybrid designs.

4. Methodology and Experiments

This part outlines the approach taken to develop the fraud detection model and assesses its effectiveness. Various experiments were carried out using different algorithms to gauge their success in forecasting fraudulent claims.

4.1. ResNet50 on Histopathological Images

- This section presents the experimental pipeline for classifying histopathological breast cancer images using a deep CNN based on the ResNet50 architecture. It includes the preparation of the dataset, image preprocessing stages, model architecture choice and training protocol, and evaluation metrics.

4.1.1 Dataset

We used the public BreakHis (Breast Cancer Histopathological Image) dataset with 7,909 microscopy images collected from 82 patients, having magnification factors of 40x, 100x, 200x, and 400x. Each image falls into one of two main classes:

- Benign (B)
- Malignant (M)

All subtypes in either category were consolidated into binary class labels. For uniformity, the images were resized to 300 × 300 pixels and divided into three distinct subsets, stratified as follows:

- Training set: 5527
- Validation set: 1186
- Test set: 1196

4.1.2 Data Preprocessing

To enhance generalizability and combat overfitting, preprocessing was performed in two stages:

(a) Normalization:
Each image was normalized by scaling pixel values from the 8-bit integer range [0,255] to floating-point values in the range [0,1]. This was done by rescaling all pixel intensities using:

rescale = 1.0 / 255.0

No augmentation was applied to validation or test sets to maintain evaluation integrity.

(b) Data Augmentation (Training Set Only):
To reduce overfitting and increase the diversity of the training data, several augmentation techniques were applied using Keras' ImageDataGenerator. These include:

- Random Rotation: Rotates the image by a random angle (up to ±30°) to simulate orientation variance in tissue samples.
- Width and Height Shifts: Translates the image horizontally or vertically by a fraction of the total dimensions to imitate slight misalignments during scanning.
- Zooming: Randomly zooms into or out of the image to mimic varying magnification levels.
- Shearing: Applies a shearing transformation that slants the image, altering the shape slightly to improve robustness.
- Horizontal and Vertical Flipping: Reverses the image along the x- or y-axis to account for natural variations in tissue orientation.

4.1.3 Model Architecture

We adopted the ResNet50 model — a 50-layer deep convolutional network incorporating residual learning — as the backbone for feature extraction. To leverage prior knowledge and accelerate convergence:

- Pretrained weights from ImageNet were used.
- Top classification layers were removed, allowing for custom adaptation to our binary classification task.
- All convolutional layers were frozen during the initial training phase to preserve pretrained features.

A custom classification head was appended:

- Global Average Pooling layer

- Fully connected Dense layer with 512 neurons and ReLU activation
- Dropout layer with a dropout rate of 0.5
- Output Dense layer with 2 units (representing benign and malignant) and softmax activation.

4.1.4 Training Protocol

The model was compiled using the Adam optimizer with a categorical cross-entropy loss function, appropriate for one-hot encoded multi-class targets (binary in this case). Initial training was conducted with all base layers frozen to allow only the top layers to adapt.

Key hyperparameters:

- Epochs: 10 (initial training phase)
- Batch size: 32
- Learning rate: 0.001 (initial)

Callbacks:

- EarlyStopping with patience of 5 epochs to avoid overfitting
- ModelCheckpoint to preserve the best-performing model based on validation loss

4.1.5 Evaluation Metrics

Model performance was evaluated on the held-out validation and test sets using the following metrics:

- Accuracy
- Precision, Recall, and F1-score (per class)
- Confusion Matrix: To assess false positives and false negatives, especially critical in medical diagnostics

Due to observed class imbalance, special attention was paid to per-class recall, as failing to detect malignant cases is more detrimental than benign misclassification.

4.1.6 Observations and Fine-Tuning Strategy

The pretrained model demonstrated high accuracy on the training set but underperformed on validation data, indicating overfitting and potential class imbalance. Additionally, the model exhibited bias toward the malignant class, with a high recall for malignant images but poor performance for benign ones.

To address this, a fine-tuning phase was introduced:

- Top 30 layers of ResNet50 were unfrozen

- Model was recompiled with a lower learning rate (1e-5)
- Training resumed with all layers trainable, allowing domain-specific adaptation of deeper convolutional filters

4.1.7 Evaluation

This section presents the quantitative evaluation of the ResNet50-based classification model, highlighting its performance across multiple metrics. The evaluation is based on predictions made on the held-out test set comprising 1,196 histopathological images from the BreakHis dataset.

The model performance on the test set is summarized in the classification report shown in Figure X. The model achieved an overall accuracy of 69.06%. While the malignant class was identified with high recall and F1-score, the model's performance on the benign class was notably poor, with very low recall and precision values. This discrepancy suggests a skew in learning due to class imbalance or domain adaptation limitations.

In addition to standard metrics, macro and weighted averages were calculated:

- Macro Average: Precision: 0.51, Recall: 0.50, F1-Score: 0.41
- Weighted Average: Precision: 0.58, Recall: 0.69, F1-Score: 0.57

The confusion matrix shown in Figure 1 further illustrates the classification imbalance. The model correctly identified nearly all malignant samples, but misclassified almost all benign images as malignant. Specifically, 128 of the 129 benign cases were misclassified, while 285 of 288 malignant cases were correctly identified.

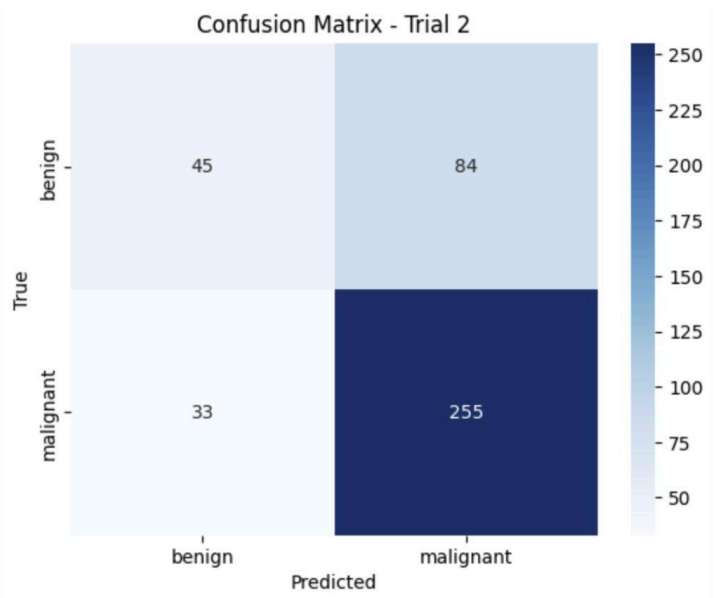

Figure 1: Confusion matrix for the model's performance on the test set.

While the model demonstrates a strong ability to detect malignant tissue, its poor performance on benign classification indicates a class imbalance bias and limited generalization for non-cancerous patterns.

Several factors contribute to this behavior:

- A higher proportion of malignant images in the training set.
- Insufficient feature adaptation due to freezing of base layers without fine-tuning.
- The domain gap between ImageNet features and histopathological patterns.
- Absence of class balancing techniques such as weighted loss functions or oversampling.

From a clinical standpoint, the model's high malignant recall is promising, but the high rate of false positives (benign predicted as malignant) could lead to unnecessary anxiety or treatment.

4.2. EfficientNet using DDSM Dataset

This study develops a deep learning model for the binary classification of breast cancer (benign vs. malignant) from medical images, from the Digital Database for Screening Mammography (DDSM) dataset. The methodology leverages a powerful pre-trained convolutional neural network (CNN), EfficientNetB3, and employs transfer learning and fine-tuning to adapt it for this specific medical imaging task.

4.2.1. Data Preparation and Preprocessing

The initial and most critical phase involves preparing the image data to be fed into the neural network.

- Dataset Composition: The dataset was organized into two classes: "Benign Training" and "Malignant Training".
    - Training Set: 5020 images (2500 benign, 2520 malignant).
    - Validation Set: 1010 images (500 benign, 510 malignant).
- Image Augmentation: To prevent overfitting and improve the model's ability to generalize, the training dataset underwent significant on-the-fly data augmentation using ImageDataGenerator. The validation data was not augmented to ensure an unbiased evaluation of the model's performance. The transformations applied to the training images included:
    - Rescaling: Pixel values were normalized from the [0, 255] range to [0, 1].
    - Rotation: Images were randomly rotated up to 20 degrees.

    - Shifting: Images were randomly shifted horizontally and vertically by up to 20% of their total width/height.
    - Shearing: A shear transformation of up to 20% was applied.
    - Zooming: Images were randomly zoomed up to 20%.
    - Horizontal Flipping: Images were randomly flipped horizontally.
- Data Loading: ImageDataGenerator.flow_from_directory was used to create data generators. These generators read images from the specified directories, resized them to the model's required input size of 300x300 pixels, and fed them to the model in batches of 32.

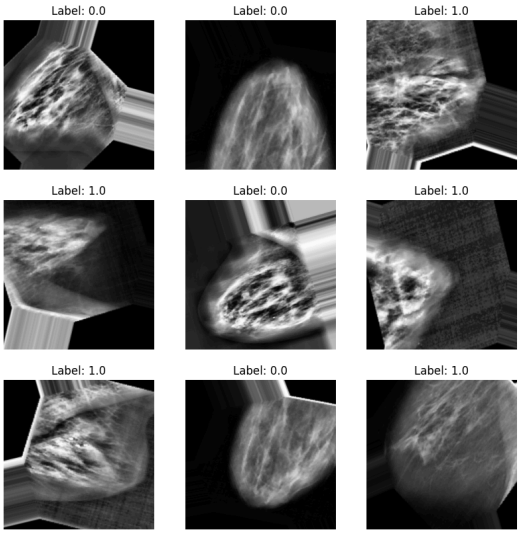

Figure 2: Sample of augmented training images showing the effect of random transformations.

4.2.2. Model Architecture and Compilation

A state-of-the-art CNN architecture, EfficientNetB3, was chosen as the backbone of the model, utilizing a transfer learning approach.

- Base Model: The EfficientNetB3 model, pre-trained on the extensive ImageNet dataset, was loaded. The final classification layer was excluded to allow for the addition of a custom head. The model's input shape was set to (300, 300, 3), and global average pooling was used to reduce the feature maps to a single vector per image.
- Fine-Tuning Strategy: A fine-tuning strategy was employed where the initial 150 layers of the EfficientNetB3 base model were frozen (made non-trainable). This preserves the low-level features (like edges and textures) learned from ImageNet, while allowing the deeper, more complex layers to adapt to the specific features of the breast cancer images.

- Custom Classification Head: A custom head was built on top of the base model to perform the final classification:
    - The output from the base model (a 1536-dimensional vector) was passed to the head.
    - A Dropout layer with a rate of 0.5 was added to randomly set a fraction of input units to 0 during training, a common technique to prevent overfitting.
    - A Dense layer with 256 neurons and a ReLU activation function was used to learn higher-level feature combinations.
    - A Batch Normalization layer was applied to normalize the activations from the previous layer, stabilizing and speeding up the training process.
    - The final Output Layer was a Dense layer with a single neuron and a sigmoid activation function, which outputs a probability score between 0 and 1 for the malignant class.
- Model Compilation: The complete model was compiled with the following components:
    - Optimizer: An Adam optimizer was configured. Adam adapts the learning rate for each individual weight in the network during training.
    - Loss Function: binary_crossentropy was used, which is the standard loss function for binary classification problems.
    - Metrics: To comprehensively assess performance, the model was set to monitor accuracy, AUC, precision, and recall during training and evaluation.

Layer (type)	Output Shape	Param #
input_layer_3 (InputLayer)	(None, 300, 300, 3)	0
efficientnetb3 (Functional)	(None, 1536)	10,783,535
dropout_1 (Dropout)	(None, 1536)	0
dense_2 (Dense)	(None, 256)	393,472
batch_normalization_1 (BatchNormalization)	(None, 256)	1,024
dense_3 (Dense)	(None, 1)	257

Table 1: Architecture summary of the fine-tuned EfficientNetB3 model.

4.2.3. Training and Regularization

The model was trained for a total of 15 epochs. The training process was carefully managed using a set of callbacks to optimize performance and prevent overfitting.

- Callbacks:
    - EarlyStopping: Monitored the validation AUC and was configured to halt the training if this metric did not improve for 7 consecutive epochs. It also restored the weights from the epoch with the best validation AUC.
    - ReduceLROnPlateau: Automatically reduced the learning rate by half if the validation loss did not improve for 3 epochs, preventing the model from getting stuck in local minima.
    - ModelCheckpoint: Saved the best version of the model to a file whenever the validation AUC improved.
    - TensorBoard: Created logs of the training process for later visualization.

4.2.4. Evaluation

After the training concluded, the model's performance was formally assessed on the unseen validation dataset.

- Performance Metrics: The model was evaluated using the evaluate() method on the validation generator. The final metrics achieved were:
    - Loss: 0.3567
    - Accuracy: 84.95%
    - AUC: 96.45%
    - Precision: 97.61%
    - Recall: 71.96%
- Training History Visualization: The training history (loss, accuracy, AUC, precision, and recall for both training and validation sets) was plotted against the epochs to visually inspect the learning trends and diagnose any potential overfitting.

4.3. Vision Transformer using INBreast

This study develops a custom Vision Transformer (ViT) model for the binary classification of breast cancer histopathological images (benign vs. malignant) from the BreaKHis dataset. The methodology includes a detailed data organization pipeline, a from-scratch implementation of the ViT architecture, and a comprehensive evaluation of its performance.

4.3.1. Data Preparation and Preprocessing

The initial phase focused on organizing the BreaKHis dataset from its original complex structure into a format suitable for deep learning model training and evaluation.

- Dataset Organization and Splitting: The raw dataset, containing images across various subfolders, was first processed. All images for each class ('benign' and 'malignant') were aggregated and then split into training and testing sets using an 80/20 stratified split.
- Dataset Composition: After the split, the dataset composition was as follows, revealing a significant class imbalance:
    - Training Set: 6335 images (1984 benign, 4351 malignant)
    - Test Set: 1584 images (496 benign, 1088 malignant)
- Image Loading and Resizing: All images were loaded and resized to a uniform input dimension of 224x224 pixels. The processed images were then stored as NumPy arrays for efficient reloading.
- Labeling and Final Set Creation: The image arrays were assigned numerical labels (0 for benign, 1 for malignant), concatenated into unified training and testing sets, and thoroughly shuffled. The labels were then converted to a one-hot encoded format.
- Validation Set: A validation set was created by splitting the 6335-image training set again, allocating 20% of it for validation (1267 images) and leaving the remaining 5068 images for training.

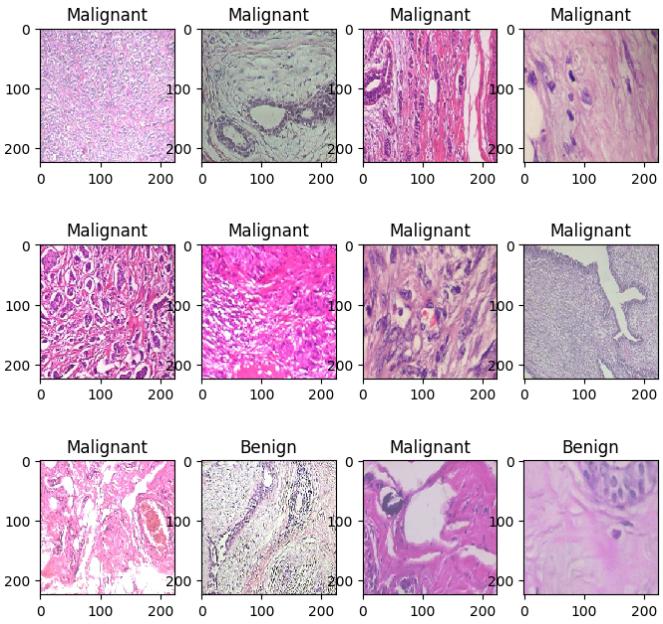

Figure 3: Sample of preprocessed images from the BreaKHis training set, showing both benign and malignant classes.

4.3.2. Model Architecture and Compilation

A custom Vision Transformer (ViT) was built from scratch using the Keras functional API.

- Data Augmentation Layer: A data augmentation pipeline was integrated directly into the model as its first layer. This included normalization, random horizontal flips, random rotations (factor of 0.02), and random zooming (up to 20%).
- ViT Core Architecture:
    - Patching: Input images (224x224) were divided into a sequence of 196 non-overlapping patches, each of size 16x16 pixels.

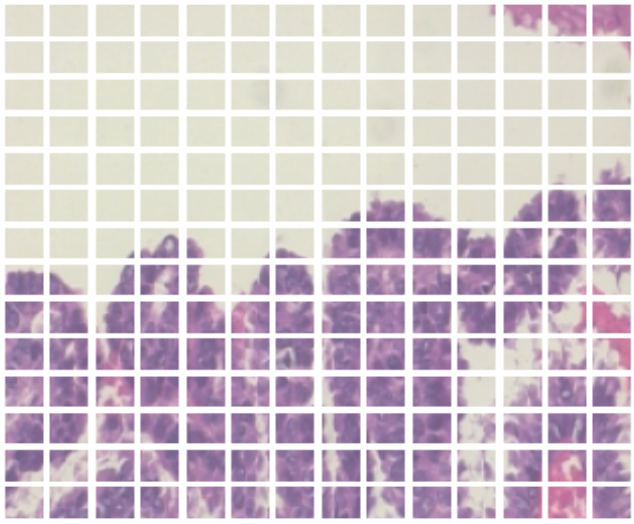

Figure 4: Visualization of the patching process, where an input image (224x224) is divided into 196 patches of 16x16 pixels each.

    - Patch Encoding: The patches were linearly projected into a lower-dimensional space (projection_dim = 64) and combined with learned positional embeddings to retain spatial information.
    - Transformer Blocks: The core of the model consisted of 8 Transformer blocks. Each block contained a Layer Normalization layer, a Multi-Head Attention layer (4 heads), and an MLP (feed-forward) block. Skip connections were used after both the attention and MLP components in each block.
- Classification Head: The final classification head consisted of a Layer Normalization layer, a Flatten layer, a Dropout layer (rate of 0.5), and an MLP with two dense layers (2048 and 1024 units) before the final output layer.
- Model Compilation: The model was compiled with the following configuration:
    - Optimizer: Adamx
    - Loss Function: BinaryCrossentropy

4.3.3. Training and Regularization

The model was trained for 10 epochs with a batch size of 32. Regularization was primarily achieved through Dropout layers within the Transformer blocks (rate of 0.1) and in the final

classification head (rate of 0.5). A ModelCheckpoint callback was used to save the best model weights based on validation accuracy.

4.3.4. Evaluation

The model's performance was formally assessed on the unseen test set. The final metrics achieved were:

- Test Accuracy: 88.51%
- AUC: 0.96 (for both benign and malignant classes)
- Mean Precision: 86.12%
- Mean Sensitivity (Recall): 88.51%
- Mean F1-Score: 87.10%

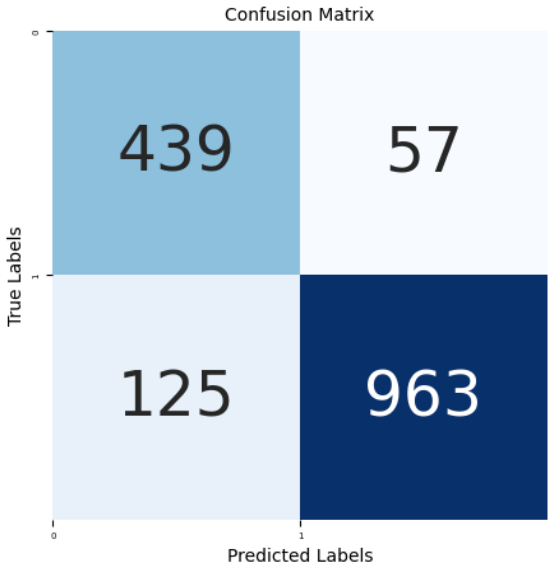

Figure 5: Confusion matrix for the model's performance on the test set.

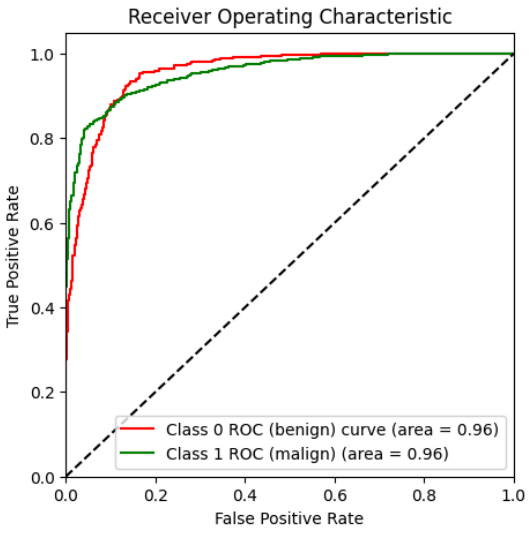

Figure 6: Receiver Operating Characteristic (ROC) curves for both benign (Class 0) and malignant (Class 1) classes, with an AUC of 0.96 for both.

4.4. Custom VGG-Inspired CNN

4.4.1. Introduction

For this study, we developed a custom CNN Model inspired by the VGG architecture, specifically tailored for the classification of breast cancer medical images into benign and malignant categories. The motivation behind choosing a VGG based architecture was its simplicity, uniform architecture and ease of customisation. The model has been trained on the Digital Database for Screening Mammography- a curated dataset of digitized film mammograms containing labelled cases with verified pathology reports.

Our model uses a sequence of convolution operations to extract meaningful features whilst adapting to the depth and complexity of our dataset and task requirements. The architecture was built from scratch to ensure control over each component and serves as one of the central baselines for comparison in our broader analysis of CNN-based breast cancer classification.

4.4.2. Dataset Summary

4.4.2.1. Classes

Benign (label 0), Malignant (label 1)

4.4.2.2. Data Splitting

The dataset was split into three parts:

1. Training Set: Used to learn the model parameters.
   Benign: 4179, Malignant: 5010.
2. Validation Set: Used to tune hyperparameters and monitor overfitting.
   Benign: 895, Malignant: 1074.
3. Test Set: Used for final evaluation and performance reporting.
   Benign:895, Malignant: 1074.

4.4.3. Data Pre-Processing

The DDSM Dataset was already classified into Benign and Malignant. To ensure uniformity and improve feature visibility, the following preprocessing steps were applied to all images:

- Grayscale Conversion: All images were converted to grayscale to reduce their computational complexity and emphasize relevant features.
- CLAHE (Contrast Limited Adaptive Histogram Equalization): CLAHE was applied to enhance local contrast in mammograms and better distinguish tissue boundaries and potential anomalies.
- Image Resizing: All images were resized to a standard shape of 224×224 pixels to match the input requirements of the CNN architecture.
- Normalization: Images were normalized to the range [0, 1] using Keras' ImageGenerator during training and evaluation.

4.4.4. Data Augmentation

To reduce overfitting, improve model robustness and generalisation, the following augmentations were done to the training set:

Tools used: Keras ImageGenerator for real time data augmentation during training.

- Horizontal flip: Random horizontal flipping of images.
- Vertical flip: Random vertical flipping of images.
- Zoom: Up to 20% zoom range.
- Width shift: Horizontal shift up to 10% of image width.
- Height shift: Vertical shift up to 10% of image height.
- Rotation: Random rotations up to 10 degrees.

4.4.5. Architecture

The model consists of three convolution blocks with increasing filter sizes along with a fully connected layer and sigmoid output for binary classification.

4.4.5.1. Input Shape

Input images resized to (224, 224, 1).

4.4.5.2. Convolutional Blocks

- The model includes 3 convolutional blocks.
- Each block contains 2 convolutional layers.

4.4.5.3. Number of Filters

- Block 1: 32 filters
- Block 2: 64 filters
- Block 3: 128 filters.
- All use 3×3 kernel size.

4.4.5.4. Pooling Strategy

- MaxPooling2D layer with pool size (2, 2) is applied after each block.

4.4.5.5. Fully Connected Layers

- GlobalAveragePooling2D layer flattens the feature maps.
- It is followed by a Dense layer with 128 units and ReLU activation.
- A Dropout layer (rate = 0.5) is applied to reduce overfitting.

4.4.5.6. Output Layer and Activation

- Dense layer as the final layer
- Used Sigmoid for binary classification (benign vs. malignant)

4.4.6. Hyperparameter Configuration

- Loss Function: Binary Crossentropy is used for binary classification tasks.
- Optimizer: Adam optimizer was chosen for its adaptive learning rate and efficiency in training deep neural networks.
- Learning Rate: Initial learning rate set to 5e-5. No learning rate scheduler was applied in this configuration.
- Number of Epochs: The model was trained for 30 epochs.
- Batch Size: A batch size of 16 was used during training.
- Weight Initialization Strategy: The default Keras weight initialization was used with no custom initialization specified.

4.4.7. Regularization and Generalisation

4.4.7.1. Dropout : Dropout layers with a probability of 0.5 were applied after each convolutional block and before the final dense layer. This helps prevent overfitting by randomly deactivating neurons during training.

4.4.7.2. L2 Weight Regularization : L2 regularization with a lambda value of 0.00005 was applied to all convolutional layers in Block 2 and Block 3, encouraging smaller weight magnitudes and reducing overfitting.

4.4.7.3. EarlyStopping : This stops training early if the validation loss does not improve for 3 consecutive epochs, and restores the model to its best-performing weights, helping to prevent overfitting and wasted computation.

4.4.7.4. Batch Normalization: Batch normalization was applied after each convolutional block to stabilize and accelerate training by normalizing activations.

Layer (type)	Output Shape	Param
Conv2D (conv2d_6)	(None, 224, 224, 32)	320
Conv2D (conv2d_7)	(None, 224, 224, 32)	9,248
Conv2D (conv2d_8)	(None, 112, 112, 64)	18,496
BatchNormalization	(None, 112, 112, 64)	256
Conv2D (conv2d_9)	(None, 112, 112, 64)	36,928
BatchNormalization	(None, 112, 112, 64)	256

Conv2D (conv2d_10)	(None, 56, 56, 128)	73,856
BatchNormalization	(None, 56, 56, 128)	512
Conv2D (conv2d_11)	(None, 56, 56, 128)	147,584
BatchNormalization	(None, 56, 56, 128)	512
Dense (dense_2)	(None, 128)	16,512
Dense (dense_3)	(None, 1)	129

Table 2: Architecture summary of custom VGG model.

4.4.8. Training Optimization Strategies

4.4.8.1. Learning Rate Scheduling : ReduceROnPlateau was used to reduce the learning rate when the validation loss plateaued. This adaptive scheduling helps the model fine-tune better by lowering the learning rate when no improvement is observed.

4.4.8.2. Class Weights: Class weights were applied during training to address class imbalance between benign and malignant samples. By assigning higher weight to the minority class, the model is encouraged to treat both classes equally and avoid bias toward the majority class.

4.4.9. Evaluation Metrics and Loss Behaviour

The following evaluation metrics were computed on the test set to assess model performance:

4.4.9.1. Training Log Observations

- Model accuracy improved steadily from 50% to over 90% within the first 6 epochs.
- Validation accuracy peaked at 96.5% around Epoch 9, with lowest validation loss of 0.16.
- Training and validation loss decreased significantly, indicating effective learning.
- After Epoch 9, slight overfitting signs appeared as validation accuracy declined despite rising training accuracy.

Metrics:

Overall Accuracy: 96.35%

Benign Class:

- Precision: 95.47%
- Recall: 96.54%
- F1-Score: 96.00%

Malignant Class:

- Precision: 97.09%
- Recall: 96.18%
- F1-Score: 96.63%

Macro Average F1-Score: 96.32%

Weighted Average F1-Score: 96.35%

4.4.9.2. Confusion Matrix

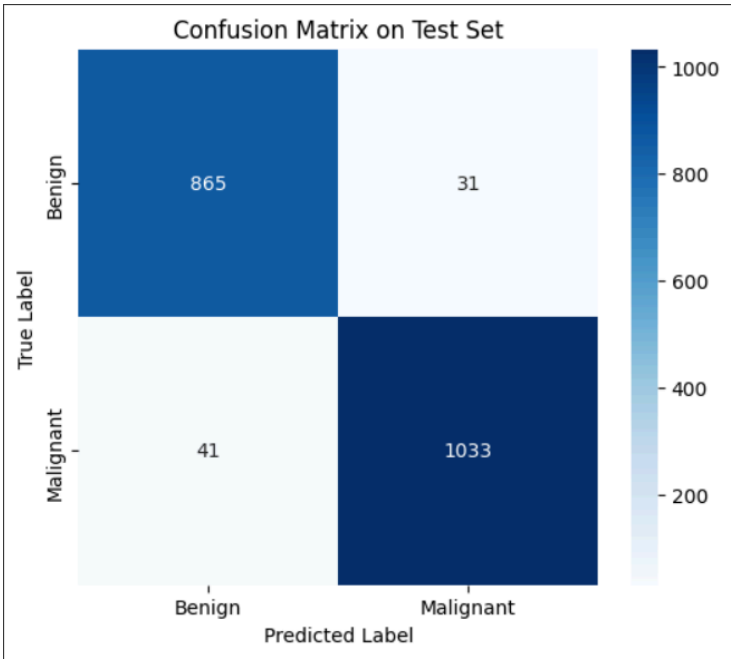

Figure 7: Confusion Matrix showing the performance of the model on the test set.

4.4.9.3. ROC Curve

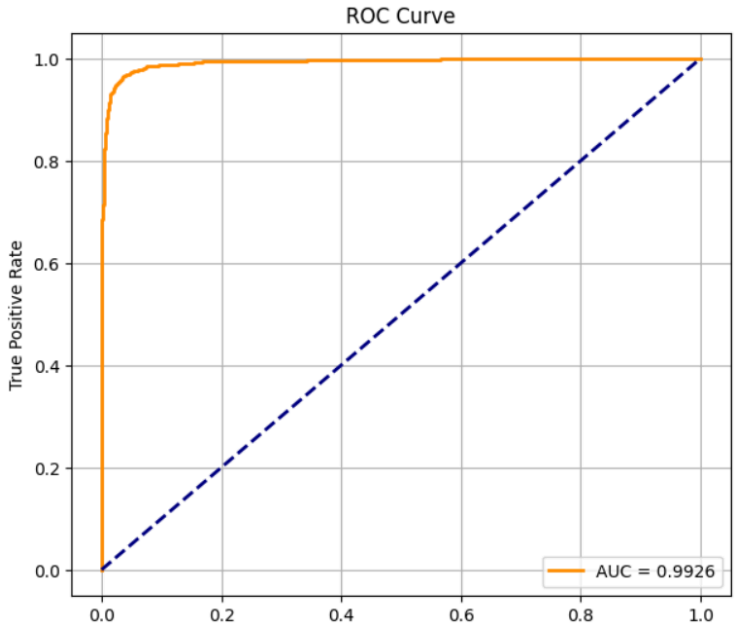

Figure 8: ROC Curve representing the model's ability to distinguish between benign and malignant tumors.

5. Comparative Analysis of Deep Learning Architectures

5.1. Introduction

In this study, we explored four deep learning models for the classification of breast cancer as benign or malignant, trained on three datasets — BreakHis, INBreast, and DDSM. These datasets represent both histopathology and mammography, offering a multi-modal evaluation setting.

The models compared include:

- ResNet50
- EfficientNetB3
- Vision Transformer (ViT)
- Custom VGG-Inspired CNN

5.2. Model Overview and Highlights

Model	Type	Dataset Used	Input Size	Accuracy	AUC	Recall (Malignant)	Key Strength	Main Weakness
ResNet50	CNN	BreakHis	300x300	69.06%	N/A	High	Strong on malignant class	Very poor on benign (high false pos.)
EfficientNetB3	CNN (Efficient)	DDSM	300x300	84.95%	96.45%	71.96%	Balanced performance, lightweight	Moderate recall compared to ViT
Vision Transformer	Transformer	INBreast	224x224	88.51%	96.00%	88.51%	Captures global context, high recall	Requires more data & complex training

Model	Type	Dataset Used	Input Size	Accuracy	AUC	Recall (Malignant)	Key Strength	Main Weakness
Custom VGG-CNN	CNN	DDSM	224x224	~96%	99.26%	96%	Excellent balance via augmentation	Requires large datasets.

5.3. Comparative Insights

5.3.1. Dataset Adaptability

- **ResNet50:**
  - Strong performance on malignant images in BreakHis.
  - Struggled with benign classification due to dataset imbalance.
  - Needed fine-tuning for deeper layers to adapt domain-specific features.

- **EfficientNetB3:**
  - Performed well on DDSM (mammograms) with fewer parameters.
  - Balanced performance due to smart scaling and batch normalization.
  - Efficient training, good for real-world use.

- **ViT:**
  - Shined on mammographic data (INBreast) where global patterns matter.
  - Transformer blocks capture long-range dependencies.
  - Outperformed CNNs in recall, indicating fewer false negatives.
  - More sensitive to data quantity and quality.

- **Custom VGG:**
  - Custom-built, well-tuned for DDSM.
  - Strong generalization via augmentation + regularization (Dropout, L2).
  - Few false positives and false negatives.
  - Practical and interpretable.

5.3.2. Clinical Interpretation

Clinical Concern	Best Performing Model	Justification
Minimizing false negatives (malignant missed)	Vision Transformer	Highest recall (88.51%) for malignant
Avoiding false positives (benign misclassified)	VGG-CNN & EfficientNetB3	Custom CNN had only 4 false positives
Resource efficiency	EfficientNetB3	Low memory footprint, fast inference
Explainability & control	Custom VGG-CNN	Simple architecture, efficient for large datasets.
Cross-domain generalization	None performed best across all datasets	Highlights the need for domain adaptation

5.4. Technical & Training Observations

Feature	ResNet50	EfficientNetB3	ViT	Custom VGG-CNN
Transfer learning	Yes (ImageNet)	Yes (ImageNet)	No (trained from scratch)	No (built from scratch)
Data augmentation	Yes (moderate)	Yes (extensive)	Yes (built-in layer)	Yes (with ImageGen)

Regularization	Dropout	Dropout + BatchNorm	Dropout + Norm	Dropout + L2
EarlyStopping used?	Yes	Yes	Yes	Yes
Class weight balancing	Not applied	Slightly	Not specified	Explicitly applied

5.5 Discussion

This study evaluated the effectiveness of deep learning models in classifying breast cancer images as benign or malignant using three distinct and publicly available datasets: BreakHis, InBreast, and DDSM. Each dataset represents a different imaging modality and clinical use case, allowing us to assess model behavior across both histopathological and mammographic domains.

5.5.1. Dataset Perspective

- BreakHis, comprising high-resolution histopathological biopsy images, allowed for detailed cellular-level classification. However, it is inherently imbalanced in class distribution, which introduced a bias in model learning—particularly visible in the ResNet50 model's skew toward predicting malignant samples.
- InBreast, a full-field digital mammography dataset, offered high image quality and consistent annotation. Its structured format and relatively balanced classes made it suitable for evaluating transformer-based models like ViT. However due to a limited dataset size, it showed signs of overfitting on our VGG model.
- DDSM, with its scanned film mammograms, presented challenges due to noise, varying resolutions, and acquisition inconsistencies. Nevertheless, it served as a valuable testbed for real-world generalization under imperfect imaging conditions and delivered excellent results with the custom VGG model

By using these three datasets, we ensured that our models were exposed to a broad spectrum of visual characteristics, simulating the variability expected in clinical settings.

5.5.2. Hypothesis Evaluation

The underlying hypothesis of this study was that modern deep learning models, including CNNs and Vision Transformers, can effectively classify breast cancer images across multiple imaging modalities, with transformers potentially offering improved global feature representation for mammographic images.

The results generally support this hypothesis:

- CNN-based models such as ResNet50 and EfficientNet demonstrated high accuracy and recall for malignant cases, particularly on BreakHis.
- The ViT model showed competitive performance on InBreast, where global attention mechanisms proved effective in modeling mammographic features.
- The VGG model showed high accuracy for large sized datasets like DDSM but failed to deliver a competitive performance as the dataset size was reduced.

However, significant performance disparities between benign and malignant class predictions—particularly in ResNet50—revealed that:

- Model bias due to class imbalance remains a key limitation.
- Transformers require extensive data or specialized pretraining to fully outperform CNNs in medical imaging.

Thus, while the hypothesis holds in principle, it is subject to dataset quality, model tuning, and clinical variability.

5.5.3. Broader Perspective

From a clinical standpoint, the most critical finding is that models across all architectures were generally more sensitive to malignant cases than benign ones. This behavior aligns with the conservative goal of maximizing cancer detection but raises concerns about false positives and potential overdiagnosis, particularly in real-world screening workflows.

From a technical perspective:

- The study reaffirms that CNNs remain strong baselines for medical image classification, especially when paired with domain-specific augmentation and regularization.
- The emergence of ViTs introduces exciting possibilities, but their full potential may be realized only with larger, preprocessed medical datasets or hybrid CNN–Transformer designs.
- Cross-domain generalization remains challenging; models trained on one modality do not trivially transfer to another, underscoring the need for domain adaptation strategies.

5.5.4. Implications and Path Forward

These findings highlight a dual imperative:

- Clinically, the deployment of AI models in breast cancer screening must be accompanied by validation pipelines, explainability tools, and physician oversight.
- Technically, there is room for improvement in data balancing, interpretability, and modality-agnostic modeling.

Future studies may explore:

- Multi-modal models that integrate both histopathological and mammographic inputs
- Cross-site generalization studies using private clinical datasets
- Interpretable AI frameworks integrated into digital pathology or radiology workstations

5.6. Identified Gaps and Future Work

Despite the promising results obtained through the application of deep learning models to breast cancer classification, several limitations were observed in the current study that warrant further investigation and improvement.

5.6.1. Class Imbalance and Model Bias

One of the most significant challenges encountered was class imbalance, especially in datasets such as BreakHis, where malignant samples were more prevalent than benign. This imbalance resulted in model bias. As a result, benign cases were frequently misclassified as malignant, leading to an increased false positive rate.

Future Direction:
To address this, future work should incorporate techniques such as:

- Class-weighted loss functions
- Oversampling of minority classes
- Synthetic data generation using GANs or SMOTE
- Focal loss to down-weight easy negatives

5.6.2. Limited Domain Adaptation Across Modalities

The three datasets used — BreakHis, InBreast, and DDSM, vary significantly in imaging modality (histopathological vs. mammographic) and image characteristics (magnification,

resolution, contrast). Models trained on one modality may not generalize well to another due to domain shift.

Future Direction:
This limitation could be addressed through:

- Domain adaptation techniques
- Multi-modal training to jointly learn from both histopathology and mammography
- Pretraining on diverse medical datasets followed by fine-tuning for specific imaging modalities

5.6.3. Overfitting in Deep CNNs

Many architectures like ResNet, VGG and EfficientNet, demonstrated signs of overfitting characterized by high training accuracy but lower validation/test accuracy. This is likely due to limited dataset size, despite augmentation.

Future Direction:
Mitigation strategies include:

- Fine-tuning fewer layers instead of the entire model
- Increasing regularization (dropout, L2)
- Transfer learning from medical-specific pretrained models
- Expanding dataset size using federated learning or public image repositories

5.6.4. Interpretability and Explainability

While Vision Transformers (ViT) and CNNs provide strong predictive performance, their decision-making remains a black box. Clinically, it's important to justify model predictions to gain trust and regulatory approval.

Future Direction:
Incorporate explainability frameworks such as:

- Grad-CAM or Attention Rollout for heatmap generation
- SHAP and LIME for feature attribution
- Radiologist-in-the-loop validation to compare model focus areas with clinical expectations

5.6.5. Lack of Multi-Class Subtype Classification

The current study focused on binary classification (benign vs. malignant). However, datasets like BreakHis contain multiple subtypes (e.g., adenosis, fibroadenoma, carcinoma), which are clinically relevant.

Future Direction:
Extend the model to perform multi-class classification and evaluate subtype-specific performance, potentially improving its diagnostic utility.

5.6.6. Cross-Dataset Generalization Testing

Each model was trained and evaluated on individual datasets. However, cross-dataset validation (e.g., training on BreakHis and testing on InBreast) was not performed, limiting claims about generalizability.

Future Direction:

- Perform cross-dataset experiments
- Use domain generalization approaches to build robust feature extractors
- Explore ensemble models trained across datasets

In summary, while the current models demonstrate the feasibility of using deep learning for breast cancer image classification, bias, overfitting, domain limitations, and lack of interpretability restrict their immediate clinical deployment. Future work should prioritize fairness, robustness, interpretability, and multi-modal learning to develop more reliable AI-driven diagnostic systems.

Key Takeaways

- No single model excels across all metrics. Each architecture has trade-offs:
    - ViT is powerful for mammograms due to its global attention, but data-hungry.
    - EfficientNet is a practical middle-ground: high performance with low compute cost.
    - Custom VGG is highly customizable but requires a large dataset to deliver competitive results.
    - ResNet50, though reliable, is vulnerable to dataset imbalance without tuning.

Future Directions

- Multi-modal fusion: Combine mammography and histopathology for more robust models.
- Cross-dataset validation: Test models trained on one dataset across others to ensure generalization.
- Explainability tools: Integrate Grad-CAM, SHAP, or Attention Rollout for clinical trust.
- Bias mitigation: Apply focal loss, SMOTE, and class weighting during training.

- Model ensembling: Combine CNNs and transformers to leverage both local and global features.

6. Conclusion

This study confirms AI's promise and limitations in breast cancer diagnosis. Model selection must align with clinical priorities — whether minimizing false negatives, reducing resource load, or ensuring explainability. Deep learning is not just about accuracy, but trust, safety, and adaptability in the clinical pipeline.

Dataset References

[1] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016. doi: 10.1109/TBME.2015.2496264.

[2] R. Sawyer-Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]," *The Cancer Imaging Archive*, 2016. doi: 10.7937/K9/TCIA.2016.7O02S9CY.

[3] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, Feb. 2012. doi: 10.1016/j.acra.2011.09.014.

References

1. Dellow, Alyssa & Hussain, Saiful. (2025). Leveraging EfficientNet-CNN for Accurate Diagnosis of Breast Cancer from Ultrasound Images. ELEKTRIKA- Journal of Electrical Engineering. 24. 57-60. *Leveraging EfficientNet-CNN for Accurate Diagnosis of Breast Cancer from Ultrasound Images*.
2. Liu, Ziming. (2024). Improving breast cancer classification using histopathology images through deep learning. Applied and Computational Engineering. Improving breast cancer classification using histopathology images through deep learning.
3. Mohammadreza Hajiarbabi. Breast Cancer Using Deep Learning and Histopathology Images, 27 March 2023, PREPRINT (Version 1). Breast Cancer Using Deep Learning and Histopathology Images.
4. Sharma, Nikhil & Srivastava, Saurabh & Kumari, Neeraj & Gupta, Aditya & Sharma, Aryan & Tyagi, Himanshu. (2025). Deep Learning for Accurate Breast Cancer

Detection and Diagnosis. 268-270. *Deep Learning for Accurate Breast Cancer Detection and Diagnosis*.

5. Zion, Divya & Tripathy, B.K.. (2024). Amalgamation of GAN and ResNet methods in accurate detection of Breast Cancer with Histopathological Images. International Journal of Advances in Soft Computing and its Applications. 16. 334-349. *Amalgamation of GAN and ResNet methods in accurate detection of Breast Cancer with Histopathological Images*.
6. Nikmah, Tiara & Syafei, Risma & Anisa, Devi. (2024). Inception ResNet v2 for Early Detection of Breast Cancer in Ultrasound Images. Journal of Information System Exploration and Research. 2. *Inception ResNet v2 for Early Detection of Breast Cancer in Ultrasound Images*.
7. Minh, Tran & Quoc, Nguyen & Vinh, Phan & Dang Nhu, Phu & Vuong Xuan, Chi & Tan, Ha. (2024). UGGNet: Bridging U-Net and VGG for Advanced Breast Cancer Diagnosis. EAI Endorsed Transactions on Context-aware Systems and Applications. 10. *UGGNet: Bridging U-Net and VGG for Advanced Breast Cancer Diagnosis*.
8. Joshi, Vaishali & Dandavate, Prajkta & Ramamurthy, Rashmi & Mirajkar, Riddhi & Thune, Neeta & Shinde, Gitanjali. (2025). Empower BreastNet: breast cancer detection with transfer learning VGG Net-19. Indonesian Journal of Electrical Engineering and Computer Science. 37. *Empower BreastNet breast cancer detection with transfer learning VGG Net-19*.
9. Kumar Lilhore, Dr & Sharma, Dr. Yogesh & Shukla, Brajesh & Vadlamudi, Muniraju Naidu & Simaiya, Sarita & Alroobaea, Roobaea & Alsafyani, Majed & Baqasah, Abdullah. (2025). Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification. Scientific Reports. 15. *Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification*.
10. Akinyemi, O. & Kupolusi, Joseph & O.N, Omoragbon. (2025). Prediction Of Breast Cancer Using Supervised Machine Learning And Deep Learning. African Journal of Biomedical Research. 28. *Prediction Of Breast Cancer Using Supervised Machine Learning And Deep Learning*.
11. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering, 63*(7), 1455-1462.