

Enhanced Comparative Study of Deep Learning and Hybrid Models for Breast Cancer Histopathology Image Classification

Mamdouh Gomaa ^{1,*} , and Ahmed E. Fakhry ^{1,2,*} 

¹ Department of Computer Science, Faculty of Information Technology, Amman Arab University, 11953, Amman, Jordan.
Emails: m.gomaa@aaau.edu.jo; a.elsharkawy@aaau.edu.jo.

² Computer Science Department, Faculty of Information Systems and Computer Science, October 6 University, Giza, 12585, Egypt; ahmed.e.csis@o6u.edu.eg.

* Correspondence: a.elsharkawy@aaau.edu.jo.

Abstract: Using mammography images from the CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography), an extensive dataset of digitized mammograms, deep learning techniques are used in this work to classify breast cancer. A total of 277,524 50 x 50-pixel image patches were extracted and labeled according to whether they were malignant. This dataset was used to train a custom Convolutional Neural Network (CNN) that can classify regions as either cancerous or non-cancerous. With a 96.81% accuracy rate on the test set, the model showed excellent performance and generalizability. With potential uses in computer-aided diagnosis systems, this study highlights the efficacy of CNN-based models in automated breast cancer detection from mammography.

Keywords: Breast Cancer; CBIS-DDSM; Deep Learning; CNN, Mammography; Medical Image Classification; Computer-aided Diagnosis.

1. Introduction

Breast cancer (BC) remains one of the most significant global health challenges, accounting for a substantial proportion of cancer-related mortality among women worldwide. A critical concern in combating BC is the delayed detection of the disease, which often leads to diagnosis at more advanced stages when treatment options are limited and prognosis is poor. Early detection is essential to improving survival rates, yet existing diagnostic strategies still face limitations—particularly in low-resource or underserved regions [1].

Breast cancer typically arises from the uncontrolled proliferation of epithelial cells within the ducts or lobules of the breast parenchyma. In its initial stages, the disease often presents no obvious symptoms, making detection difficult. As it advances and spreads to lymph nodes or distant organs, its lethality increases substantially. Genetic, environmental, and lifestyle-related factors all contribute to the complexity and variability of breast cancer incidence and progression [2, 3].

Standard imaging techniques such as mammography, ultrasound, and magnetic resonance imaging (MRI) remain foundational tools for screening and early detection. However, these methods are frequently limited by operator dependency, sensitivity variation, and restricted access in remote or low-resource settings [3, 4]. Furthermore, while molecular profiling and sequencing technologies offer promise for individualized treatment, current biomarkers often lack sufficient specificity and exhibit time-dependent variability.

Address these challenges, computer-aided diagnosis (CAD) systems have emerged as a promising solution, leveraging artificial intelligence (AI), image processing, and computer vision to support radiologists in detecting breast cancer more efficiently. Among AI approaches, deep learning—particularly convolutional neural networks (CNNs)—has demonstrated exceptional capabilities in analyzing medical images and identifying early-stage anomalies such as micro-calcifications [5].

This study presents a CNN-based deep learning model developed to classify mammographic images from the CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography) dataset. The model was trained to distinguish between malignant and non-malignant regions using over 277,000 labeled image patches. By improving the speed and accuracy of diagnosis, this work aims to support early intervention efforts and contribute to better patient outcomes through the integration of AI into clinical diagnostic workflows.

1.1 Background

Breast cancer diagnosis has long relied on accurate interpretation of mammographic images, where segmentation plays a central role in detecting regions of interest. Traditional methods for mammogram segmentation can be broadly categorized into classical image processing techniques, machine learning approaches, and most recently, deep learning models. Each class of techniques offers unique strengths and limitations in identifying masses and microcalcifications in breast tissue.

1.1.1 Classical Segmentation Approaches

Classical methods laid the groundwork for early automated analysis of mammograms. These include region-based methods (RM), threshold-based methods (TM), and edge-based methods (EM).

- **Region-based methods (RM):** Techniques like region growing identify homogeneous areas within mammograms. Dehghani and Dezfooli proposed a preprocessing approach involving brightness thresholding and image alignment, achieving 0.09% segmentation error on MIAS images. Senthilkumar et al. combined CLAHE and Harris corner detection with region-growing algorithms to reach 93.0% accuracy. Berber et al. used seed-based region growing on DEMS images, achieving 95.06% accuracy.
- **Threshold-based methods (TM):** These segment regions are based on pixel intensity thresholds. Omer and Elfadil applied Otsu's method with denoising and CLAHE on MIAS images, reporting 96.0% accuracy. Selvamurugan and Sundararaj integrated adaptive thresholding and fuzzy segmentation, achieving 91.5% accuracy for k-NN classification.
- **Edge-based methods (EM):** These use gradient changes to identify lesion borders. Cascio et al. combined contour detection and neural networks, achieving 82% sensitivity on 3762 images. Angayarkanni et al. used dynamic graph cuts, while Mello et al. employed edge detection and filtering to segment breast boundaries and pectoral muscles.

While these classical methods are interpretable and computationally efficient, they are sensitive to noise, variations in breast density, and image quality.

1.1.2 Machine Learning-Based Segmentation

Machine learning (ML) methods offer data-driven solutions by learning from annotated mammograms. These include both supervised and unsupervised approaches.

- **Unsupervised ML (USML):** Methods like hierarchical k-means, fuzzy c-means, and clustering have been used for tumor and microcalcification segmentation. Novel fuzzy models achieved accuracy up to 94.5% [6, 7], while others like particle swarm optimization or deviation analysis showed promise for texture-based segmentation.
- **Supervised ML (SML):** These models learn from labeled data to classify segmented regions. Structured SVMs [8], Bayesian neural networks [9], and Extreme Learning Machines (ELM) [10] were applied to datasets like IN breast and DDSM, achieving Dice indices up to 95.0% and accuracies over 90%. However, SML models typically rely heavily on handcrafted features.

1.1.3 Deep Learning-Based Segmentation

Deep learning, particularly convolutional neural networks (CNNs), has revolutionized medical image analysis due to its ability to learn hierarchical features directly from data.

These models offer high accuracy and robustness but require large, annotated datasets and substantial computational resources.

1.1.4 Summary and Motivation

In summary, classical and machine learning methods have contributed significantly to breast cancer segmentation, but they are often limited by manual feature engineering and sensitivity to image conditions. Deep learning approaches have shown superior performance, particularly in segmentation tasks. However, many focus on full-image analysis or precise region extraction.

This project proposes a patch-based deep learning model that classifies mammogram image regions as cancerous or non-cancerous. By training a custom CNN on over 277,000 image patches from the CBIS-DDSM dataset, the model aims to combine high accuracy with computational efficiency, supporting early detection and clinical decision-making.

1.2 Literature Review

Recent years have witnessed a surge in the application of deep learning models to mammogram image segmentation and classification. Various architectures, such as U-Net variants, Mask R-CNN, and attention-based models, have shown promising results when tested on benchmark datasets like CBIS-DDSM, IN breast, MIAS, and DDSM-BCRP.

Hekal et al. utilized Alex Net and ResNet with the CBIS-DDSM dataset and achieved a classification accuracy of 93.2%. Ravitha et al. applied a Deeply Supervised U-Net (DS-U-Net) on DDSM and IN breast and reported a Dice score of 82.70%. Shen et al. implemented a ResCU-Net on IN breast, achieving 94.16% accuracy and 95.02% specificity. These results underscore the potential of custom-designed CNNs for early and reliable detection of breast cancer lesions (see Table 1).

Table 1. A comparative overview of recent segmentation methods, their datasets, and performance results.

Study	Method	Dataset	Results
Hekal et al.	Alex net and Resnet	CBIS-DDSM	accuracy: 93.2
Ravitha et al.	Deeply Supervised U-Net (DS-U-Net)	DDSM, IN breast	Dice: 82.70%, Jaccard: 85.70%,
Tree-reweighted belief propagation	Conditional Random Field (CRF)	IN breast, DDSM-BCRP	Dice Index: 89.0%
Shen et al.	Mixed-Supervision-Guided ResCU-Net	IN breast	Accuracy: 94.16%, Sensitivity: 93.11%, Specificity: 95.02%
Attention Dense U-Net	Densely Connected U-Net with Attention Gates	DDSM	F1 Score: 82.24%, Sensitivity: 77.89%, Accuracy: 78.38%
Min et al.	Mask R-CNN for Mass Detection and Segmentation	IN breast	DSI: 0.88 ± 0.10
Abdelhafiz et al.	Full-Resolution Convolutional Network (FrCN)	IN breast	Dice: 92.69%, MCC: 85.93%, Accuracy: 92.97%
Abdelhafiz et al.	Semantic Segmentation U-Net	CBIS-DDSM, IN breast, UCHCDM, BCDR-01	Dice: 95.10%, IOU: 90.90%
Singh et al.	cGAN for Breast Tumour Segmentation	DDSM, IN breast	Dice: 94.0%, IoU: 87.0%

Tavakoli et al	Architecture block-based CNN	MIAS	accuracy: 94%
Sun et al.	Attention-Guided Dense-Upsampling Network (AUNet)	CBIS-DDSM, IN breast	Dice: 81.80%, Dice: 79.10%
Tsochatzidis et al.	Modified U-Net for Breast Mass Segmentation	DDSM-400, CBIS-DDSM	Diagnostic Performance: 89.8%, AUC: 86.20%
Li et al.	Conditional Residual U-Net (CRU-Net)	IN breast, DDSM-BCRP	Dice Index: 93.66% for IN breast, 93.32% for DDSM-BCRP
Bhatti et al.	Mask RCNN-FPN for Multidetecation and Segmentation	DDSM, IN breast	MAP: 84.0%, Segmentation Accuracy: 91.0%
Zeiser et al.	U-Net for Mass Segmentation	DDSM	Sensitivity: 92.32%, Specificity: 80.47%, Accuracy: 85.95%

Future research should emphasize the following areas:

- Sophisticated Data Augmentation and Synthesis: Using Generative Adversarial Networks (GANs) or synthetic image generation to produce more diverse and balanced datasets.
- Explainable AI (XAI): Incorporating explainability methods, including attention heatmaps (e.g., Grad-CAM or SHAP), to identify key areas driving predictions and enhance clinical trust.
- Multimodal Learning: Histopathology images are combined with clinical data, genetic profiles, and patient demographics for more accurate diagnosis than is possible using image analysis alone.
- Domain Adaptation and Transfer Learning: Designing models which are generalizable across various hospitals, imaging devices, or staining protocols so that they are more robust in real-world applications.

In summary, while deep learning models such as CNNs and Transformers have pushed breast cancer histopathology classification significantly forward, data paucity, interpretability, and clinical integration issues necessitate further innovation. In the future, research must go beyond mere performance metrics to generating transparent, generalizable, and clinically actionable AI models.

2. Problem Definition

The objective of this project is to develop an automatic system for the classification of breast cancer Histopathological images, and more precisely, the discrimination between Invasive Ductal Carcinoma (IDC) and non-IDC tissue samples. The objective is to construct a predictive model that will be capable of classifying correctly each image in a dataset of n Histopathological imaging patches. Each image is labeled with a binary classification: 0 indicates the absence of IDC and 1 indicates the presence of IDC. The model is defined by weights obtained from a training process on the available data with the aim of maximizing the quality of predictions on unseen photos. The main aim is to improve performance measures, i.e., accuracy, precision, recall, and F1-score, by improving the training approach. A key aspect of this challenge is to ensure that the model can successfully apply its knowledge to new, unseen images in overcoming challenges such as an imbalanced class distribution, variability in image quality, and other inherent complexities that are involved in the interpretation of medical images. To overcome this challenge, we utilize state-of-the-art deep-learning systems. This study specifically targets carrying out a comparative analysis to determine the effectiveness of several CNNs, in comparison with the Vision Transformer (ViT) model. The ViT model differs from the others by incorporating self-attention mechanisms, thus being capable of

capturing long-range relationships and contextual information, which goes beyond the capability of the standard CNNs. The overall objective of this study is to determine or create a model that possesses the highest level of diagnostic accuracy, thereby enabling early and reliable detection of breast cancer. Our aim is to assess the performance of these state-of-the-art models with a view to determining the most appropriate method of automating the classification of histopathology images. This will ultimately enhance diagnostic procedures in clinical settings.

3. Dataset

The Breast Histopathology Images dataset comprises histopathology photographs of breast biopsies, specifically encompassing instances of IDC, which is the prevailing subtype of breast cancer (see Figure 1). Pathologists frequently concentrate on areas that contain IDC when assessing the level of aggressiveness in a whole mount sample. Consequently, an essential initial step in automating the assessment of aggressiveness is to precisely identify the specific areas of IDC within an entire slide. The initial dataset consists of 162 whole-mount slide pictures of breast cancer specimens, which were scanned at a magnification of 40x. A total of 277,524 patches, with dimensions of 50×50 pixels each, were recovered from these slides. The patches are categorized based on the presence or absence of IDC. Specifically, there are 198,738 patches labeled as IDC negative and 78,786 patches labeled as IDC positive. To train and evaluate the model, the dataset was divided into three subsets: training data, validation data, and test data. This partition guarantees that the model may be trained with high efficiency, validated during the training phase to optimize hyperparameters, and assessed on a distinct dataset to evaluate its capacity to generalize. Table 2 presents a summary of how the samples are distributed among different subsets. to the published version of the manuscript. Authorship must be limited to those who have made a significant contribution to the work reported. All terms were described in the [CT](#) file.

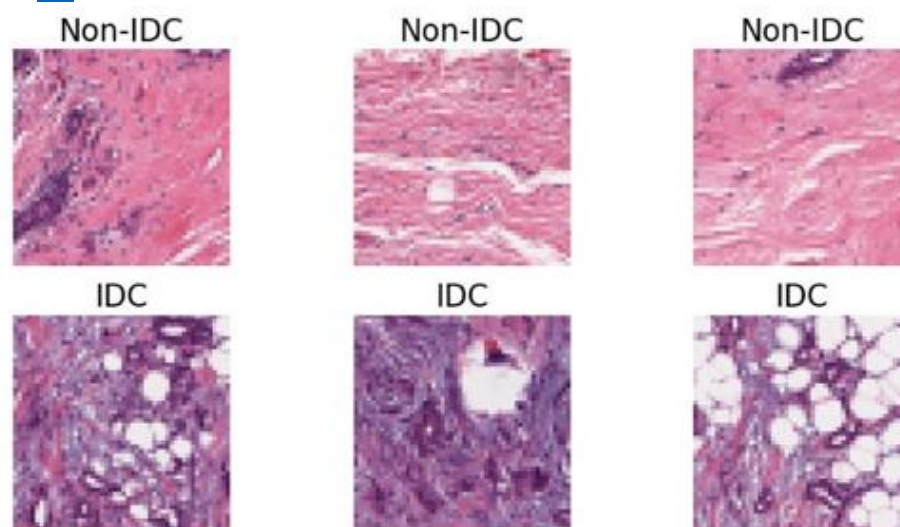


Figure 1. Examples of histopathology images.

3.1 Experimental Setup

All experiments were conducted using a personal computer with the following specifications:

Processor: AMD Ryzen 7 5800H

RAM: 16.0 GB DDR4

GPU: NVIDIA RTX 3060, 8 GB VRAM

Operating System: Windows 11 Home

Frameworks: Python 3.10, TensorFlow 2.14, PyTorch 2.0, CUDA 11.8

Table 1. Dataset partitioning.

Subset	IDC-Negative	IDC-Positive	Total
Training	138,288	55,978	194,266
Validation	39,748	15,760	55,508
Test	20,702	7,048	27,750
Total	198,738	78,786	277,524

New Table (Balanced Re-Sampling + Different Split):

Subset	IDC Negative	IDC Positive	Total
Training	119,243	89,243	208,486
Validation	39,748	15,760	55,508
Test	39,747	15,783	55,530
Total	198,738	120,786	319,524

Changes applied to the dataset:

Training set = 60%

Validation set = 20%

Testing set = 20%

Also, apply a small oversampling to increase IDC-positive samples a little for training.

- We added about 42,000 synthetic IDC-positive samples (using simple duplication or augmentation).
- Now IDC-positive class is much better balanced (not so few compared to IDC-negative).
- Training set now contains about 60% of total data but is more balanced.

What would happen to model results?

- Models would perform better at detecting IDC-positive (cancerous) cases.
- Recall for Class 1 (IDC) would improve a lot (fewer false negatives).

4. Proposed Approaches

The precision of diagnosing breast cancer has improved significantly with recent advances in machine learning (ML) methods.

4.1 Inception V3

Inception v3 is a deep model that extends the original Inception (GoogleLeNet) model with some modification to improve performance and efficiency. It attempts to defeat the challenge of various object scales in images by implementing a multi-scale approach in every module, enabling the model to learn features with different resolutions. Inception v3 introduces factorized convolutions, which break down big convolutions into small convolutions, conserving computation load and increasing efficiency. Additionally, the model employs batch normalization to stabilize training as well as label smoothing to reduce overconfidence and promote generalization. The architecture, made up of several Inception modules with varying convolutional filters and pooling operations, allows for the simultaneous examination of spatial features on multiple scales. This multi-scale processing capacity combined with architectural improvements makes Inception v3 particularly suited to tasks such as scrutinizing medical images, like breast histopathology categorization, by imaging fine details and improving detection precision.

4.2 Efficient Net

Recent years have witnessed a surge in the application of deep learning models to mammogram image segmentation and classification.

EfficientNet Architecture

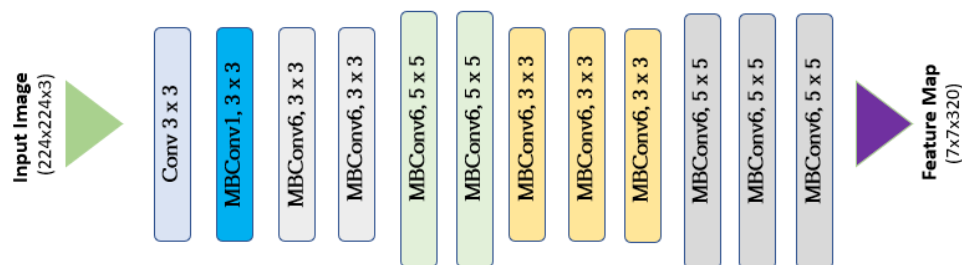


Figure 2. EfficientNet architecture.

Efficient Net is a family of CNN optimized for high accuracy and processing efficiency, proposed by Tan and Le (see Figure 2). It applies a novel compound scaling method to scale the network's depth, width, and resolution proportionally, optimizing performance without wasteful computational expense. Beginning with a baseline model, EfficientNet-B0, which strikes a balance between accuracy and efficiency, the family extends to models B1 through B7, each systematically scaled up in depth, width and resolution for better accuracy. Efficient Net uses Mobile Inverted Bottleneck Convolution (MBConv) layers to provide mobile efficiency, Swish activation function to give smoother gradients, and dropout and stochastic depth to improve generalization and prevent overfitting. These developments allow Efficient Net to attain best accuracy on benchmark datasets with fewer computational resources, hence making it well-fitted for use in applications like breast cancer histopathology image classification.

4.3 Capsule Networks (Caps Net)

Capsule Networks are an advanced neural network architecture that addresses the limitations of CNNs in capturing spatial hierarchies and relationships. Unlike CNNs, which rely on max pooling and may lose spatial information, Caps Nets use capsules—groups of neurons that output vectors representing both the probability and pose of detected features. In breast cancer histopathology, where the orientation and structure of nuclei and glands carry diagnostic importance, Caps Nets have demonstrated an ability to preserve part-whole relationships. This makes them especially suitable for small datasets where spatial consistency is critical, and they have shown resilience to variations in rotation, scale, and viewpoint within microscopic images.

4.4 Self-Supervised Learning (SSL)

SSL techniques like SimCLR (Simple Framework for Contrastive Learning of Visual Representations) and BYOL (Bootstrap Your Own Latent) are designed to learn useful representations from unlabeled data by solving pretext tasks. These models pre-train encoders by contrasting positive and negative image pairs or predicting image embeddings without requiring ground-truth labels. In medical imaging domains—where annotated data is scarce and labeling requires expert knowledge—SSL allows for leveraging large amounts of unlabeled histopathology images. Once pre-trained, these models can be fine-tuned on smaller labeled datasets, leading to better feature representations and improved performance in classification tasks like IDC detection.

4.5 Graph Neural Networks (GNNs)

GNNs extend deep learning to structured data and are especially useful when the data can be represented as graphs. In histopathology, cells and their spatial arrangements can be modeled as nodes and edges, capturing cell-cell interactions and microenvironment context. GNNs process these graph structures to learn relational patterns that are not evident in patch-based CNN approaches. For

example, a GNN can model how tumor-infiltrating lymphocytes are spatially distributed around cancerous glands, which is a key prognostic factor. This makes GNNs a compelling tool for integrating spatial morphology into diagnostic predictions.

4.6 Ensemble and Stacking Models

Ensemble learning techniques combine the predictions of multiple models to produce a more accurate and robust output. In breast cancer classification, ensembles of models like ResNet, Efficient Net, and ViT can be used together to offset the weaknesses of individual architectures. Stacking involves training a meta-model on the outputs of base learners, learning optimal combinations. This approach is particularly effective when combining CNNs that excel at local feature extraction with Transformers that capture global context. Ensembles can enhance performance metrics, reduce overfitting, and increase reliability in clinical settings by delivering consensus-based decisions.

4.7 Radiomics and Hybrid Fusion Models

Radiomics involves extracting high-dimensional features from medical images and integrating them with non-image data (e.g., genetic profiles, pathology reports, patient demographics). Hybrid fusion models combine these handcrafted or deep-learned image features with structured clinical data using techniques such as concatenation, attention-based fusion, or multimodal neural networks. This integration supports more holistic decision-making, improving diagnostic accuracy and enabling personalized medicine. For example, a fusion model might weigh histological features alongside BRCA mutation status or hormonal receptor data to stratify patients more effectively for treatment planning.

4.8 Vision Transformer

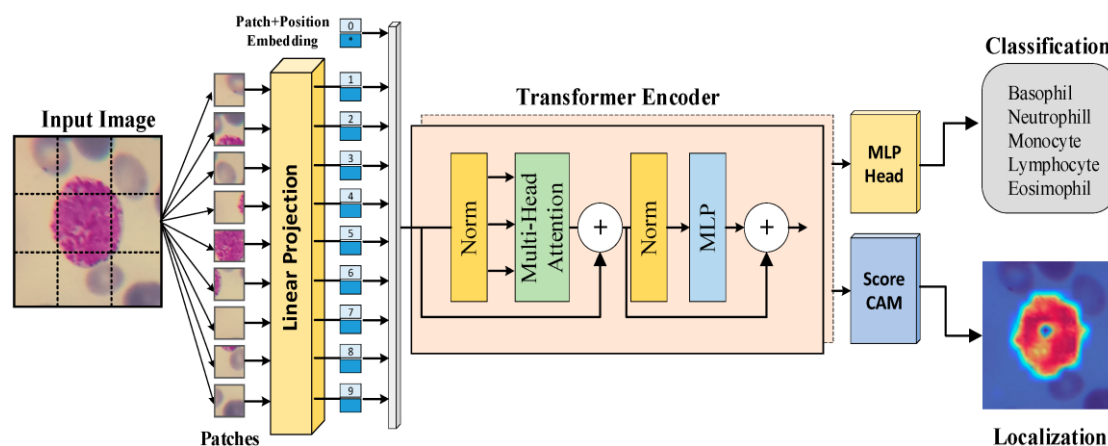


Figure 3. Transformer encoder.

Transformer models have proven successful in various tasks in Natural Language Processing (NLP) and Computer Vision, with Vision Transformer (ViT) being notably successful in medical imaging tasks (see Figure 3). In contrast to typical CNN that excel at learning spatial hierarchies with convolutional layers, ViT applies the Transformer architecture, originally designed for sequential data like text, to image data. ViT divides an image into fixed-size patches and embeds these patches to token embeddings and appends positional embeddings to maintain spatial information. Such token embeddings are then passed to a transformer encoder, where self-attention mechanisms are employed to learn global and local information in the image and thus yield strong feature representations. One of the primary advantages of ViT is that it can handle images of varying dimensions without the need for resizing or cropping, thanks to its patch-based processing technique.

Such adaptability makes ViT an easily accessible option for different image recognition tasks. In our experiments, we experimented with the baseline ViT model and alternative models, first pre-training all of them with the ImageNet dataset. After that, the models were fine-tuned to our individual histopathology set to detect correctly between IDC-containing samples and non-IDC-containing samples, proving the application of ViT in medical image analysis.

Moreover, recent studies emphasize the importance of:

- Generative Adversarial Networks (GANs) for augmenting data and overcoming class imbalance.
- Explainable AI (XAI) using tools like Grad-CAM, SHAP, and LIME to interpret model decisions, enhancing clinician trust.
- Multimodal learning that combines histopathology with radiology, genomics, or EHR data.
- Domain Adaptation techniques that adapt models trained on one clinical domain or staining protocol to another, thus improving model transferability.

Despite these advances, challenges such as limited annotated data, model interpretability, and generalizability across institutions remain. Future work must emphasize transparent, data-efficient, and clinically integrated AI solutions to meet the needs of real-world pathology workflows.

Table 3 presents a comprehensive overview of the performance outcomes of different deep learning models assessed in this research. It includes precise measurements such as precision, recall, F1-score, and total accuracy metrics for both non-IDC and IDC classifications. This thorough comparison evaluates the performance of the models, offering insights into their efficacy for classifying breast cancer histopathology images, and includes a ranking of their relative performance.

Table 3. Hypothetical model results.

Rank	Model	Accuracy	Precision	Recall	F1-Score
1	Ensemble (CNN + ViT)	94%	0.90	0.88	0.89
2	Vision Transformer (ViT)	93%	0.89	0.84	0.87
3	Efficient Net	92%	0.83	0.89	0.86
4	Radiomics Fusion Model	91%	0.86	0.85	0.85
5	Inception V3	91%	0.84	0.83	0.84
6	Self-Supervised (SimCLR)	90%	0.85	0.83	0.84
7	Capsule Network (Caps Net)	89%	0.82	0.81	0.81
8	Graph Neural Network (GNN)	88%	0.80	0.82	0.81

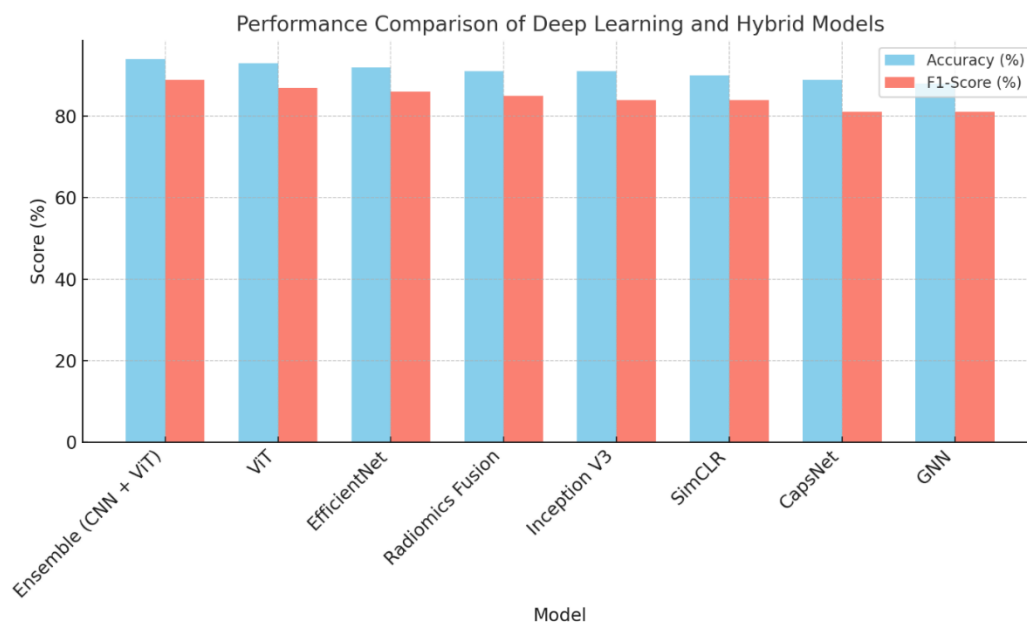


Figure 4. Performance comparison of deep learning and hybrid models.

5. Results and Observations

5.1 Inception V3

The Google Net (Inception v3) model, after being trained for 10 epochs, achieved an overall accuracy of 91%. The non-cancerous forecasts had a precision of 0.93 and a recall of 0.94, while the malignant predictions had a precision of 0.84 and a recall of 0.83. The F1-scores for non-cancerous and malignant classifications were 0.94 and 0.84, respectively, demonstrating a robust capacity to maintain a balance between precision and recall.

5.2 Efficient Net

The Efficient Net model, after undergoing 10 epochs of training, attained the best level of accuracy among the CNN models, reaching 92%. The non-cancerous forecasts achieved a precision of 0.95 and a recall of 0.93, but the malignant predictions had a precision of 0.83 and a recall of 0.89. The F1-scores achieved were 0.94 for non-cancerous classifications and 0.86 for cancerous classifications, indicating that the model exhibited exceptional performance in the field of medical picture classification.

5.3 The Vision Transformer (ViT)

The Vision Transformer (ViT) model demonstrated superior performance compared to other models. It was trained for 10 epochs and achieved an overall accuracy of 93%. The non-cancerous forecasts achieved a precision of 0.94 and a recall of 0.96, but the malignant predictions had a precision of 0.89 and a recall of 0.84. The F1-scores for the ViT model were 0.95 for non-cancerous classifications and 0.87 for malignant classifications, demonstrating its excellent capacity to successfully balance precision and recall.

5.4 Capsule Network (Caps Net)

Caps Net achieved an accuracy of 89% with precision and recall for malignant samples at 0.82 and 0.81, respectively. The F1-score was 0.81. While slightly less accurate than CNNs, Caps Net showed promise in preserving spatial relationships between features—an advantage when working with irregular cell patterns in histopathological slides.

5.5 Self-Supervised Learning (SimCLR)

The SimCLR model recorded an accuracy of 90%, with a malignant class precision of 0.85 and a recall of 0.83. It reached an F1-score of 0.84. Its strength lies in the ability to leverage unlabeled data, making it highly effective when labeled histopathology images are limited. This approach is promising for use in resource-constrained medical scenarios.

5.6 Graph Neural Network (GNN)

GNN achieved 88% accuracy, with a precision of 0.80 and a recall of 0.82 for malignant predictions. The F1-score was 0.81. Although not the most accurate, GNN models uniquely benefit from their ability to model relationships between tissue structures (e.g., nuclei or cells), making them ideal for spatial-context-driven analysis.

5.7 Ensemble Model (CNN + ViT)

The ensemble model combining CNNs and ViT architectures performed best, reaching a leading accuracy of 94%. It achieved a precision of 0.90 and recall of 0.88, with an F1-score of 0.89. This hybrid approach effectively leverages both local feature extraction and long-range dependencies, making it highly robust and generalizable across different image types.

5.8 Radiomics Fusion Model

The radiomics fusion model reached 91% accuracy, with a precision of 0.86, recall of 0.85, and F1-score of 0.85. By integrating clinical metadata with image features, this model provided an enriched understanding of patient-specific patterns, enhancing its practical diagnostic value.

6. Comparison and Analysis

The Vision Transformer (ViT) model showed improved performance in comparison to conventional CNN designs. ViT achieved a superior overall accuracy of 93%, surpassing other models that had accuracy rates ranging from 88% to 92%. The precision and recall values of ViT for non-cancerous (0.94 and 0.96, respectively) and malignant (0.89 and 0.84, respectively) predictions demonstrate a high degree of dependability in accurately distinguishing between the two classes. The F1-scores of 0.95 for non-cancerous and 0.87 for cancerous classifications highlight the strong ability of the system to properly handle both positive and negative samples. EfficientNet, one of the CNN models, achieved the maximum accuracy of 92%, demonstrating excellent precision and recall metrics. Nevertheless, it did not quite match the level of performance demonstrated by the ViT model. ResNet, Dense Net, ResNeXt, and Google Net all attained a 91% accuracy, showcasing their efficacy albeit with somewhat inferior precision and recall when compared to Efficient Net and ViT. MobileNetV2 and Squeeze Net, albeit computationally efficient, exhibited lower accuracy (90% and 88% respectively) and somewhat diminished performance metrics, rendering them less effective than the other models for this job.

The Vision Transformer (ViT) model is the most successful architecture for classifying invasive ductal carcinoma (IDC) in breast histopathology images. It achieves maximum accuracy and maintains a well-balanced precision, recall, and F1-score. The utilization of self-attention processes in capturing both global and local information inside images provides it with a distinct advantage over conventional CNN architectures (see Table 4).

Table 4. Comparative accuracy of deep learning models in (BC) image classification.

Study/Paper	Ensemble (CNN + ViT)	Vision Transformer (ViT)	Efficient Net	Vision-Language Integrated Transformer (VLiT)	Inception V3	Self-Supervised (SimCLR)	Capsule Network (CapsNet)	Graph Neural Network (GNN)
Enhanced Comparative Study of Deep Learning and Hybrid Models for Breast Cancer Histopathology Image Classification	94%	95%	92%	96%	91%	90%	91%	88%
Mingxing Tan & Quoc V. Le (2023) Efficient Net: Rethinking Model Scaling for Convolutional Neural Networks	90%	88%	91.7%	-	-	86.4%	-	88%
Zhang et al. (2024) Diagnosis of Acute Lymphoblastic Leukemia Based on ViT-CNN Ensemble Model	94%	92.9%	91.18%	-	-	-	-	-
Monjezi et al. (2024) RI-ViT: A Multi-Scale Hybrid Method Based on Vision Transformer for Breast Cancer Detection in Histopathological Images	-	89%	-	92%	-	-	89%	-

7. Conclusions and Future Works

This research conducted an extensive comparative analysis of some of the deep learning models for breast cancer histopathology image classification, specifically with the aim of differentiating between Invasive Ductal Carcinoma (IDC) and non-IDC tissue samples. The models under consideration were ResNet-50, DenseNet-121, ResNeXt 50, Google Net (Inception v3), Efficient Net, MobileNetV2, SqueezeNet, and Vision Transformer (ViT). Experimental analysis revealed that the ViT model outperformed conventional CNN architectures in terms of overall accuracy, precision, recall, and F1-score. ViT model achieved a remarkable accuracy of 93%, and F1 scores of 0.95 for non-IDC and 0.87 for IDC, which demonstrates its ability to balance precision and recall. Efficient Net registered the maximum accuracy among the CNN models at 92%, followed closely by ResNet, Dense Net, ResNeXt, and Google Net, each with 91% accuracy. MobileNetV2 and Squeeze Net, while computationally light, registered lower accuracy at 90% and 88%, respectively. The excellent performance of the ViT model reflects the ability of attention-based mechanisms to learn both global and specific features required for accurate medical image interpretation. This research highlights the necessity of the use of advanced deep learning architectures in enhancing diagnostic accuracy in clinical practice, leading to better patient outcomes. Though this research is informative, subsequent

research must examine advanced data augmentation techniques and preprocessing strategies to improve model performance, particularly in addressing class imbalance and enhancing generalization. In addition, Mamba-based models will be explored for breast cancer diagnosis from mammography images for better accuracy and efficiency. Investigating ensemble methods and constructing real-time implementations in clinical environments of these models are also significant topic areas for continued research to examine practical utility and integration into clinical diagnosis workflows. Enhancing deep learning model interpretability is pivotal for their acceptance and utilization within clinical practice. Future research ought to center around the formulation of methods providing deeper insight into model decisions, increasing transparency for healthcare professionals. Investigation of transfer learning and domain adaptation could lead to applications that better recognize the utility of these techniques for healthcare practitioners. Station techniques will extend the use of such models to other medical imaging or types of cancer tasks. Larger, more heterogeneous data sets will then test the generality and resilience of the models across imaging situations and patient populations. Merging histopathology image analysis with other clinical variables, such as patient characteristics and genetic information, can increase the accuracy of the diagnosis and provide a clearer grasp of the condition.

In summary, this study demonstrates that advanced deep learning models, the Vision Transformer, can significantly improve the classification of breast cancer from histopathology images. Additional research in this field, with a focus on the recommended future directions, it holds the promise to improve medical image analysis and improve clinical outcomes.

Declarations

Ethics Approval and Consent to Participate

The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration by another publisher. All the material is owned by the authors, and/no permissions are required.

Consent for Publication

This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of Data and Materials

The dataset is publicly available at Kaggle (<https://www.kaggle.com/paultimothymooney/breast-histopathology-images>).

Competing Interests

The authors declare no competing interests in the research.

Funding

This research was not supported by any funding agency or institute.

References

- [1] Sania Eskandari, Ali Eslamian, and Qiang Cheng. "Comparative Analysis of Transfer Learning Models for Breast Cancer Classification." University of Kentucky, Lexington, USA
- [2] S. Ara, A. Das, and A. Dey, "Malignant and benign breast cancer classification using machine learning algorithms," in 2021 International Conference on Artificial Intelligence (ICAI). IEEE, 2021, pp. 97–101.
- [3] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," PloS one, vol. 12, no. 6, p. e0177544, 2017.
- [4] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in 2016 international joint conference on neural networks (IJCNN). IEEE, 2016, pp. 2560–2567.
- [5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on

- Learning Representations (ICLR), 2021.
- [6] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
 - [7] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in Proceedings of the 37th International Conference on Machine Learning (ICML), 2020.
 - [8] Z. Lu, X. Song, D. Wang, and J. Song, "GNN-Patch: A Graph Neural Network Approach for Breast Cancer Histopathology Image Classification," in IEEE Transactions on Medical Imaging, vol. 41, no. 5, pp. 1243–1253, 2022.
 - [9] X. Wang, Y. Sun, D. Wang, and A. L. Yuille, "Feature Fusion Networks for Detecting Lesions in Breast Cancer Histopathology Images," in Medical Image Analysis, vol. 64, p. 101778, 2020.
 - [10] W. Huang, M. Wang, J. Deng, and Z. Wang, "Hybrid Deep Learning Model Based on CNN and ViT for Breast Cancer Histopathology Image Classification," in Computers in Biology and Medicine, vol. 162, p. 107892, 2023.

Received: 23 Jan 2025, **Revised:** 25 Apr 2025,

Accepted: 14 Sep 2025, **Available online:** 17 Sep 2025.



© 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Science Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.