International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 07 | Jul 2026 www.irjet.net p-ISSN: 2395-0072

Deep Learning and Vision Transformer-Based Automated Classification of Invasive Ductal Carcinoma in Breast Histopathology Images

Dropti[1], Prateek Gupta[2]

[1]*First Department of Computer Science and Engineering Shri Ram Institute of Sciences and Technology, Jabalpur, M.P., India*

[2]*Department of Computer Science and Engineering Shri Ram Institute of Sciences and Technology, Jabalpur, M.P., India*

Abstract – *Invasive ductal carcinoma (IDC) is the most common form of breast cancer, and its accurate identification from histopathological images is essential for reliable diagnosis and effective clinical decision-making. However, manual examination of breast tissue slides is time-consuming, subjective, and affected by inter-observer variability. Recent advances in deep learning and transfer learning have demonstrated significant potential for automated histopathological image analysis. This paper proposes an automated deep learning-based framework for the classification of IDC in breast histopathology images. The study utilizes the publicly available Breast Histopathology Images dataset, comprising 277,524 image patches categorized into non-IDC and IDC classes. To prevent patient-level data leakage a critical flaw in many existing medical AI models the framework employs a rigorous patient-wise data splitting strategy. Three deep learning approaches are implemented and compared: a baseline transfer learning DenseNet121 model, a fine-tuned DenseNet121 model, and a Vision Transformer (ViT) model. Results show that the fine-tuned DenseNet121 achieves the best performance with 86.60% accuracy and an ROC-AUC score of 0.9195 on unseen test patients. This study highlights that convolutional architectures remain highly effective for extracting local histopathological textures, outperforming transformer-based approaches which require significantly more data. The proposed framework demonstrates strong potential as a computer-aided diagnostic system to assist pathologists in real-world clinical environments.*

Key Words: Invasive Ductal Carcinoma (IDC), Breast Cancer, Histopathology Images, Deep Learning, DenseNet121, Fine-Tuning, Vision Transformer (ViT), Medical Image Classification.

1. INTRODUCTION

According to the World Health Organization (WHO), breast cancer poses the most significant risk to women, serving as the leading cause of cancer mortality in women worldwide [1]. With 2.3 million new cases annually, representing 11.7% of all new cancer cases in 2020 [2], early detection is critical for improving survival rates. The five-year survival rate for breast cancer is virtually 100% if discovered at an early stage, but it can drop drastically to as low as 15% if discovered later.

Invasive Ductal Carcinoma (IDC) is the most common and aggressive subtype of all breast cancers. Traditional diagnosis relies on the manual microscopic examination of tissue samples by expert pathologists [3], [4]. To assign an aggressiveness grade to a whole mount sample, pathologists typically focus on regions containing the IDC. However, this manual process is labor-intensive, time-consuming, and subject to diagnostic inconsistencies and inter-observer variability in large-scale screening environments [5]. Mammograms are currently the most utilized test, although they frequently return false positive results that show abnormal cells, leading to pointless, expensive, and unpleasant biopsies.

With the increasing availability of medical data and the rapid development of analytical methods, Artificial Intelligence (AI) and deep learning are revolutionizing the healthcare industry [6]. Convolutional Neural Networks (CNNs), such as DenseNet121, have shown strong capabilities in extracting discriminative local texture features from histopathological images [7]. Simultaneously, transformer-based architectures like Vision Transformers (ViT) have emerged as powerful alternatives for modeling global contextual relationships [8]. However, the integration of these models raises concerns about transparency, requiring Explainable Artificial Intelligence (XAI) methods like SHAP to interpret the inner workings of black-box models [9], [10].

This paper investigates whether transformer-based methods can outperform CNNs for IDC detection and proposes a robust, automated computer-aided diagnostic framework. By evaluating both CNN-based (DenseNet121) and transformer-based (ViT) architectures, this study aims to optimize automated disease detection while addressing critical dataset challenges such as patient-level data leakage [11], [12].

2. LITERATURE SURVEY

A. Machine Learning for Breast Cancer Detection

Machine learning has emerged as a central tool in breast cancer research. Prior works have extensively utilized Gradient Boosting, XGBoost, and LightGBM models. For instance, one prominent study explored the application of five supervised ML models on a real-world clinical dataset of 500 patients from Dhaka Medical College Hospital, where XGBoost emerged as the best performer with an impressive accuracy of 97% [1]. Another study by Arravalli et al. focused on identifying breast cancer using ML classifiers and incorporating XAI for model transparency [3]. A tailored approach for Indian breast types optimized machine learning algorithms for mammographic images, addressing population-specific characteristics (breast density, texture, lesion size) and achieving an Area Under the Curve (AUC) of 0.98 [4].

B. Deep Learning Across Imaging Modalities

Thermographic Data: Thermography serves as a non-invasive alternative. Dharani et al. designed an enhanced deep learning model using heatmaps to assess vascularity, achieving 96.8% accuracy [5]. Other frameworks have utilized hybrid heuristic and adopted deep learning [6]. Aidossov et al. coined the "black-box" issue in deep learning by combining CNN outputs with Bayesian Networks [7]. Similarly, Alshehri et al. improved CNN performance by integrating attention mechanisms (AMs), which allowed the network to focus on thermally significant regions [8].

Mammogram Data: Mammography remains the primary screening tool. Researchers like Seo et al. have used architectures like PMVnet to model relationships between standard mammographic views (CC and MLO) to reduce false positives [9]. Other frameworks, like those by Zeng et al., used ResNet18 CBAM to predict HER2, ER, and PR status directly from mammograms [10].

Ultrasound Data: Ultrasound imaging is highly accessible but suffers from operator dependency and variable image quality. Fatima et al. proposed architectures like InBnFUS and CNNDen-GRU, fusing Inception modules with inverted bottlenecks, achieving 99.0% accuracy on breast ultrasound images by successfully capturing both spatial and sequential features [11].

C. Deep Learning for Histopathology

Histopathological analysis remains the gold standard for a definitive diagnosis. Computational pathology utilizes CNN-based models to drastically reduce pathologist workload. Alzakari et al. proposed a Deep Sparse Wavelet Autoencoder (DSWAE) that preserved critical morphological features while achieving 100% precision for normal tissue, successfully minimizing false positives [12]. Furthermore, segmentation-classification pipelines, such as those combining Fuzzy C-Means clustering with U-Net-based CNNs by Potti et al., have achieved 99.23% accuracy [2].

3. DATASET: BREAST HISTOPATHOLOGY IMAGES

The study utilizes the publicly available Breast Histopathology Images dataset, a widely recognized benchmark for IDC detection. The original dataset consists of 162 whole mount slide images of breast cancer specimens scanned at 40x magnification.

From these slides, 277,524 image patches of size 50x50 pixels were extracted. The dataset is naturally imbalanced, containing 198,738 non-IDC (Class 0) patches and 78,786 IDC positive (Class 1) patches. Each patch's file name contains specific coordinate data and a unique Patient ID (e.g., 10253_idx5_x1351_y1101_class0.png), indicating exactly where the patch was cropped from the original whole mount slide. The hierarchical organization of these unique Patient IDs is critical for preventing data leakage during model evaluation.

4. PROPOSED WORK

To address the limitations of existing frameworks, a comprehensive deep learning pipeline was developed.

A. Patient-Wise Splitting Strategy

A critical flaw in existing medical AI research is performing a random image-level split, which inadvertently places patches from the same patient into both the training and testing sets. This causes data leakage, artificially inflating accuracy while failing to generalize in real-world scenarios. To guarantee the integrity of the evaluation, this study applies a strict 80:20 patient-wise train-test split. Unique patient IDs were extracted, and all images belonging to a specific patient were assigned to only one subset (either train or test). This ensures the models are evaluated on genuinely unseen patient tissue.

B. Data Preprocessing and Augmentation

A standardized preprocessing pipeline was developed using TensorFlow. Raw PNG images were decoded and resized from their original 50x50 dimensions to 224x224 pixels. Pixel intensities were normalized to a range of to ensure a uniform input format for the deep learning models.

To improve model generalization and reduce overfitting, a data augmentation pipeline was applied exclusively to the training set. Transformations included random horizontal

flipping, small rotations (0.1), and zoom transformations (0.1), simulating real-world variability in tissue orientation. The data pipeline was optimized using TensorFlow's tf.data API (shuffling, batching, prefetching) for efficient handling of the large-scale dataset.

C. Proposed Deep Learning Architectures

The framework investigates and compares three architectural approaches:

1. Baseline DenseNet121: A transfer learning approach utilizing DenseNet121 pretrained on ImageNet. The pretrained convolutional layers were frozen to preserve learned visual representations (edges, textures, structural patterns). A custom classification head was added, consisting of:

- Global Average Pooling (to reduce feature dimensions)
- Batch Normalization (to stabilize learning)
- Dense Layer with 128 units and ReLU activation (to learn IDC-specific features)
- Dropout at 0.5 (to reduce overfitting)
- Sigmoid output layer (for binary classification)

2. Proposed Fine-Tuned DenseNet121: While the frozen baseline model extracts generic visual features, it is limited in learning domain-specific histopathological tissue patterns. The proposed fine-tuning strategy addresses this by unfreezing the higher DenseNet layers (approximately the last 50 layers) while keeping the lower convolutional layers frozen. The model is retrained using a lower learning rate (1e-5) to ensure gradual, stable weight updates. This allows the network to learn breast tissue-specific patterns while avoiding catastrophic forgetting of pretrained knowledge.

3. Vision Transformer (ViT): A transformer-based approach was implemented to evaluate the efficacy of self-attention mechanisms in medical imaging. The ViT model divides images into patches, encodes positional information, and applies global self-attention mechanisms to model global contextual relationships across the tissue sample, followed by classification.

D. Proposed Algorithm Workflow

The systematic workflow for the proposed framework is executed as follows:

Step 1: Traverse patient folders to extract image file paths, corresponding labels, and Patient IDs to construct a structured DataFrame.

Step 2: Extract unique patient IDs and split the dataset 80:20 ensuring patient separation.

Step 3: Read, decode, resize to 224x224x3, and normalize image pixels.

Step 4: Apply data augmentation (flips, rotations, zooms) exclusively on training data.

Step 5: Construct the DenseNet121 baseline with a custom classification head and compile using the Adam Optimizer and Binary Crossentropy Loss.

Step 6: Fine-tune by unfreezing top layers and training at a 1e-5 learning rate.

Step 7: Train the parallel Vision Transformer (ViT) architecture.

Step 8: Generate predictions on the unseen testing dataset.

Step 9: Compute comparative performance metrics (Accuracy, ROC-AUC, Precision, Recall, F1-Score).

E. Evaluation Strategy

The evaluation protocol simulates real-world clinical deployment where the models, trained on tissue samples from a specific set of patients, are evaluated directly on entirely unseen patients without retraining. This strict 80:20 patient-wise evaluation strategy prevents the data leakage and overfitting commonly seen in standard random image-level splitting.

Model performance is comprehensively measured using Test Loss, Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Receiver Operating Characteristic - Area Under Curve (ROC-AUC) score. Because the Breast Histopathology dataset exhibits significant class imbalance between the majority non-IDC and minority IDC patches, the ROC-AUC score and F1-score are utilized as the primary evaluation metrics. These metrics provide a much more honest and reliable assessment of the model's ability to correctly identify the malignant class (IDC) compared to relying on overall accuracy alone.

5. RESULT

The three deep learning models were trained and evaluated on the same patient-wise split using identical preprocessing protocols. Performance was assessed using test loss, classification accuracy, and ROC-AUC scores.

A. Baseline Results

Table I summarizes the performance across the evaluated models. The baseline DenseNet121 achieved a test accuracy of 85.89%, a test loss of 0.3330, and an ROC-AUC score of 0.9177. These results demonstrate that the frozen DenseNet121 effectively captures discriminative histopathological features directly through transfer learning without any domain-specific weight updates.

TABLE I: Performance Comparison of Deep Learning Models

Model	Test Loss	Test Accuracy (%)	ROC-AUC
DenseNet121 (Baseline)	0.3330	85.89	0.9177
Vision Transformer (ViT)	0.3441	85.62	0.9165
Proposed Fine-Tuned DenseNet121	0.3170	86.60	0.9195

B. Vision Transformer Evaluation

The Vision Transformer (ViT) model obtained a test accuracy of 85.62%, a test loss of 0.3441, and an ROC-AUC of 0.9165. Although the transformer architecture successfully learned global contextual information from the image patches, its overall performance was slightly lower than the CNN-based DenseNet architectures.

C. Post-Tuning Analysis

The proposed Fine-Tuned DenseNet121 achieved the best overall performance, reducing test loss to 0.3170 and increasing test accuracy to 86.60%. Furthermore, it achieved the highest ROC-AUC score of 0.9195. The model's confusion matrix analysis demonstrated balanced classification performance, achieving an overall classification accuracy of approximately 87% and a strong recall of 76% for the cancerous class, indicating effective identification of malignant tissue patterns.

6. DISCUSSION

The experimental findings yield crucial practical insights into automated medical image classification. First, fine-tuning an existing CNN backbone provides the most reliable and robust performance for automated IDC breast cancer classification. Fine-tuning enabled the pretrained network to adapt specifically to the domain-specific characteristics of breast histopathology, minimizing classification error.

Second, the slightly lower performance of the Vision Transformer reveals a known constraint of self-attention mechanisms. Vision Transformers generally require substantially larger datasets to fully exploit their global attention capabilities.

In contrast, DenseNet121 benefits from dense feature reuse and efficient gradient propagation, making it highly effective for extracting the local texture and tissue morphology critical to identifying IDC on moderate-sized datasets.

Finally, the strict patient-wise split utilized in this study ensures that the reported accuracy (86.60%) reflects genuine real-world generalization on unseen patients. Standard random splits often create an optimistic bias (data leakage) that does not reflect actual clinical deployment conditions.

7. CONCLUSION

This paper presented an automated deep learning framework for the detection of Invasive Ductal Carcinoma (IDC) evaluated on 277,524 breast histopathology image patches. By employing a rigorous patient-wise data splitting strategy, the framework successfully prevented data leakage and ensured realistic model evaluation. An efficient preprocessing pipeline incorporating image resizing, normalization, and data augmentation was implemented to improve learning robustness.

The study performed a comparative analysis of CNN-based and transformer-based architectures. The proposed fine-tuned DenseNet121 model demonstrated superior generalization, achieving 86.60% accuracy and an ROC-AUC of 0.9195 on completely unseen patient data, outperforming both the frozen baseline and the Vision Transformer (ViT). The findings indicate that CNN architectures currently remain the most effective and computationally viable approach for texture-dependent histopathological analysis, as local tissue morphology plays a critical role in diagnosis.

8. FUTURE WORK

Although the proposed framework achieved promising performance, future studies should explore advanced deep learning architectures such as EfficientNet variants, Swing Transformers, and hybrid CNN-Transformer ensembles to further enhance feature extraction. Additionally, expanding the framework to encompass multi-class breast cancer classification (distinguishing multiple subtypes, benign vs. malignant categories, or tumor severity stages) will provide more comprehensive diagnostic support. Finally, validation on larger multi-center datasets and diverse patient populations is essential before real-world clinical deployment.

REFERENCES

[1] World Health Organization, "Breast Cancer: Prevention and Control." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer J. Clinicians, vol. 71, no. 3, pp. 209–249, May 2021.

[3] B. Jiang, et al., "Deep learning applications in breast cancer histopathological imaging: Diagnosis, treatment, and prognosis," Breast Cancer Res., 2024.

[4] S. K. Singh and K. S. Patnaik, "Patho-ai: A perceptive breast cancer identification and classification using deep learning methods integrated with explainable ai," SN Computer Science, vol. 6, p. 619, 2025.

[5] E. A. Rakha, J. S. Reis-Filho, F. Baehner, et al., "Breast cancer prognostic classification in the molecular era: The role of histological grade," Breast Cancer Res., vol. 12, no. 4, p. 207, Jul. 2010.

[6] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," Cold Spring Harbor Protocols, vol. 2008, no. 5, May 2008.

[7] T. Islam, M. A. Sheakh, M. S. Tahosin, et al., "Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable ai," Scientific Reports, vol. 14, no. 1, Apr. 2024.

[8] N. Sengodan, "Breast cancer histopathology classification using cbam-efficientnetv2 with transfer learning," 2025. [Online]. Available: https://arxiv.org/abs/2410.22392.

[9] D. Opoku, K. Owusu-Agyemang, et al., "SwinCNN+OE: A Swin Transformer and CNN Architecture for Breast Histopathology Classification With OOD and Grad-CAM Integration," IEEE Access, vol. 14, pp. 3897–3909, 2026.

[10] L. S. Potti and S. Maruthuperumal, "Breast cancer cell detection using fcm and prediction using unet based deep convolutional neural network," in 2024 5th IEEE Global Conference for Advancement in Technology (GCAT), 2024.

[11] M. Fatima, M. A. Khan, A. M. Mirza, et al., "An adaptive deep learning approach based on inbnfus and cnnden-gru networks for breast cancer and maternal fetal classification using ultrasound images," Scientific Reports, vol. 15, no. 1, p. 22027, 2025.

[12] D. Tellez, G. Litjens, P. Bándi, et al., "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," Med. Image Anal., vol. 58, Dec. 2019.

[13] M. S. Hosseini, "Computational pathology: A survey review and the way forward," J. Pathol. Informat., vol. 15, Dec. 2024.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[15] T. Zeng et al., "Bilateral Information-Guided Diagnosis of Breast Masses in Mammography Using Vision Transformer," IEEE Journal of Biomedical and Health Informatics, 2025.

[16] A. Thampi, Interpretable AI: Building Explainable Machine Learning Systems. New York, NY, USA: Simon and Schuster, 2022.