International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 919-928
https://doi.org/10.70917/ijcisim-2026-2813
Article

AUTOMATED BREAST CANCER DETECTION USING HYBRID CNN AND VISION TRANSFORMER NETWORKS ON MAMMOGRAPHIC IMAGES

Nelofar Bashir*[1], Nilesh Bhosle[2]

[1]Nims Institute of Computing, Artificial Intelligence and Machine Learning, Nims University Rajasthan, Jaipur, India Email: nelofarbashir111@gmail.com
[2]Nims Institute of Computing, Artificial Intelligence and Machine Learning, Nims University Rajasthan, Jaipur, India Email: nilesh.bhosle@nimsuniversity.org
(*Corresponding Author: Nelofar Bashir nelofarbashir111@gmail.com)

Abstract: Breast cancer is a highly prevalent malignant tumor worldwide, arising from the uncontrolled proliferation of abnormal cells in breast tissue. It primarily affects women, though cases also occur in men. Currently, mammography is the primary technology for early breast cancer screening. Detected imaging abnormalities are classified as benign or malignant. However, when radiologists manually interpret these images, the complexity of breast tissue patterns complicates interpretation, often leading to inconsistent readings and missed diagnoses. To address this challenge, this study proposes a hybrid deep learning framework that integrates convolutional neural networks (CNNs) and Vision Transformers (ViTs). CNNs excel at extracting hierarchical local image features, while ViTs capture long-range dependencies and global context. That said, ViTs have relatively high computational and data requirements, so this hybrid model is designed to improve robustness and diagnostic reliability. This study trains its model on the Kaggle breast mammography dataset enhanced with Contrast Limited Adaptive Histogram Equalization (CLAHE). Four pre-trained models, namely DenseNet, Inception, SE-ResNet, and XceptionNet, are selected for performance comparison. The proposed model achieves a validation accuracy of 90.1%, with balanced precision and recall of 89.4% and 90.8%, respectively, demonstrating robust generalization. Although XceptionNet attained perfect accuracy, such performance may suggest potential overfitting. Overall, the hybrid approach effectively balances local and global feature learning, offering a reliable solution for automated breast cancer classification.
Keywords: Breast Cancer Detection, Mammography, Convolutional Neural Networks, Vision Transformer, Hybrid Deep Learning, Medical Image Analysis, CLAHE, Classification, Computer-Aided Diagnosis

1. INTRODUCTION

Breast cancer remains one of the most serious global health challenges and is anticipated to be a leading cause of cancer-related mortality among women worldwide in the coming years. The disease develops when abnormal cells within breast tissue multiply uncontrollably, forming lumps or tumors that grow more rapidly than normal cells. This uncontrolled proliferation significantly increases the likelihood of malignancy and poses a substantial threat to patient survival if not detected at an early stage. Despite advancements in medical technology, disparities in access to specialized breast cancer diagnostic services persist. Reports indicate that a considerable number of healthcare institutions lack adequate breast cancer screening facilities and trained specialists, which negatively impacts early diagnosis and treatment outcomes [1]. According to global health studies, a significant proportion of breast cancer cases remain undiagnosed, while mortality rates continue to rise, particularly among women. In India, breast cancer contributes notably to global statistics, with cases increasingly reported among younger women, often beginning in their early thirties. These trends underscore the urgent need for effective and accessible diagnostic solutions [2]. Early detection is critical to improving survival rates and reducing mortality risk for breast cancer patients. Currently, mainstream breast imaging screening methods include ultrasonography and mammography. Among these, mammography is the most reliable non-invasive screening method: it can produce clear breast X-ray images, causes very little discomfort to patients, and allows doctors to identify lesions. Benign lesions have smooth, regular borders, whereas malignant lesions typically present with irregular, spiculated margins. Although mammography is highly effective, accurate interpretation remains challenging due to factors such as breast tissue density, overlapping structures, and subtle visual differences between early-stage benign and malignant lesions [3]. These complexities increase the risk of misdiagnosis, even for experienced radiologists. Consequently, computer-aided diagnostic (CAD) systems have emerged as valuable tools to support radiologists by improving diagnostic accuracy, minimizing human error, and enhancing consistency in clinical decision-making.

In recent years, artificial intelligence-enabled automatic detection of breast cancer has become a core research focus in medical imaging [4,5]. Early related studies mostly relied on traditional hand-engineered image processing methods to complete lesion identification. As deep learning technology has been widely applied, mainstream solutions in the current field have evolved into data-driven end-to-end learning models [6]. Among these, convolutional neural networks (CNNs) excel at extracting local texture features, while vision Transformers (ViTs) can capture long-range semantic correlations. Both types of models have proven their application value in breast image analysis, but they still have inherent limitations when deployed independently: CNNs struggle to model global dependencies, and ViTs lack sufficient efficiency in extracting fine-grained local features [7]. To address this common pain point, hybrid architectures that integrate the strengths of both model types have become a widely adopted approach to optimization. In response to this issue, this study developed a hybrid CNN-ViT breast cancer classification model paired with a dedicated preprocessing workflow for mammogram images. This model leverages the core advantages of the two-standalone single-type architectures, thereby offsetting the performance shortcomings of existing solutions [8].

2. LITERATURE REVIEW

Recent advances in deep learning have substantially improved the performance of automatic detection and classification of breast cancer in mammography images. Within this research field, three mainstream architectures have been explored, namely convolutional neural networks, Transformers, and hybrid frameworks. Their core goals are to improve diagnostic accuracy and reduce reliance on manual reading of medical images. Their core task is to distinguish between benign and malignant lesions, and the unsolved problems to be addressed will be elaborated in subsequent sections.

Many researchers in the field have leveraged the strong local spatial feature extraction capability of CNNs to conduct AI-based detection research for breast diseases. Hassan et al. proposed a ViT-integrated YOLO framework that processes CESM and FFDM images to detect and classify masses for breast cancer detection [9]. Their approach achieved high classification accuracy on both INbreast and CESM datasets by combining YOLOv4 for lesion localization with ViT for global feature modeling [10]. Similarly, Jabeen et al. introduced the BC2NetRF framework, which integrates EfficientNet-B0 with HRLG contrast enhancement and an Equilibrium-Jaya-controlled feature selection strategy. Their model demonstrated strong performance on CBIS-DDSM and INbreast datasets, highlighting the effectiveness of contrast enhancement and optimized feature selection in mammography analysis [11].

Transfer learning has also been extensively explored to improve classification performance when training data is limited. A study by Chakravarthy et al. adopted the EfficientNet-B4 architecture combined with chaotic mapping to optimize feature selection, achieved high accuracy on standard breast mammography datasets, and highlighted the value of optimized transfer learning [12]. Another study by the same team used a multi-depth CNN framework, verifying that an integrated architecture can capture the complex tumor patterns of early-stage breast cancer [13].

Segmentation-based approaches have further improved breast cancer detection accuracy. Wang proposed a deep learning-based mammography framework that utilizes the Mask R-CNN and U-Net architectures for lesion segmentation prior to classification. Their results indicated that accurate segmentation plays a crucial role in improving diagnostic outcomes [14]. Similarly, Prinzi et al. explored a YOLO-based architecture combined with Eigen-CAM saliency maps to enhance model explainability and support clinical decision-making [15].

In the field of medical imaging AI, hybrid CNN-Transformer models that can capture both local and global image representations are receiving increasing attention from the academic community. Mohammed and Ekmekci proposed a YOLO-based multi-scale parallel CNN framework, while Dosovitskiy et al. introduced the Vision Transformer architecture [16]. In addition, three studies marked as verified have shown that ViT can effectively model long-range dependencies in breast cancer diagnosis and demonstrate promising application potential [17].

To improve the robustness and generalizability of deep learning models for breast cancer detection, the academic community has proposed two core types of methods: ensemble methods and attention mechanisms. Previously, Nadkarni and Noronha combined ensemble CNNs with optimization algorithms to improve breast cancer detection accuracy [18]; O'Shea's research confirmed the feature-extraction efficiency of CNNs while noting that their computational challenges are the primary driver of the design of hybrid optimization models [19]. Despite these advances, existing models still suffer from overfitting, high computational complexity, and insufficient generalizability for unseen clinical data. In recent years, the academic community has proposed solutions that integrate CNNs with the Vision Transformer (ViT) [20,21]. Inspired by this line of research, the present study develops a CNN-ViT hybrid framework for mammogram images that balances three core objectives: accuracy, robustness, and interpretability.

3. Methods

This study proposes a hybrid deep learning framework that integrates convolutional neural networks (CNNs) and Vision Transformers (ViTs) to improve breast cancer classification performance on mammogram images. Leveraging the complementary strengths of CNNs for extracting local features and ViTs for capturing global relationships, this study also constructs several established, mature deep learning architectures to assess their performance.

3.1 Overall Framework

The proposed framework follows a structured pipeline consisting of image preprocessing, feature extraction, feature fusion, and classification. Mammography images are first enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve contrast and highlight diagnostically significant patterns. The enhanced images are then passed through two parallel branches: a custom CNN branch and a pre-trained ViT-B16 branch. Features extracted from both branches are fused and subsequently used for final classification into benign or malignant categories. This study set up core-controlled experiments to verify the performance and effectiveness of its self-developed hybrid model. We selected five deep learning architectures, including a custom CNN, as references and comprehensively evaluated the performance of different feature extraction strategies on the same dataset to ensure a fair comparison.

3.2 Custom CNN Architecture

The authors of this paper propose a custom convolutional neural network for identifying abnormal breast tissue, with the primary goal of extracting three categories of fine-grained local features: edges, textures, and morphological changes. The network is built with 3 convolutional blocks, each followed by a max-pooling layer: the first layer is equipped with 32 filters with 3×3 kernels, succeeded by 2×2 max-pooling and ReLU activation; the number of filters in the second layer is increased to 64, with all other configurations remaining unchanged; the third layer uses 128 filters to capture high-order features, and the final output is flattened into a one-dimensional vector, which serves as the input for subsequent feature fusion.

3.3 Vision Transformer (ViT) Feature Extraction

The authors of this paper, working within an AI medical imaging framework for mammogram image analysis, introduced the ViT-B16 model to build a global feature extraction module that addresses the limitation that convolutional neural networks (CNNs) can only learn local features. First, input mammogram images are resized to 224×224 pixels and then split into non-overlapping 16×16 image patches to form a sequence. After

After being flattened, the sequence is linearly projected into a high-dimensional embedding space, with positional encoding added to the embeddings. The processed data is then fed into a multi-layer Transformer encoder for further computation. We remove the original ViT's native classification head, and the extracted global features are sent to a fusion layer. The self-attention mechanism can capture long-range relationships, which aligns with the core characteristic that abnormalities in medical images often span spatially distant regions.

3.4 Feature Fusion and Classification

This paper presents a deep learning model for classifying benign and malignant medical lesions. First, the flattened local feature vectors from a CNN are concatenated and fused with the global feature vectors from a ViT, enabling the model to incorporate both fine-grained morphological details and broad contextual information. Next, a dropout layer with a dropout rate of 0.5 is applied to mitigate overfitting and improve the model's generalizability.

Finally, a fully connected layer with a sigmoid activation function outputs the class probabilities for benign and malignant cases, clearly conveying prediction confidence to support reliable clinical decision-making.

3.5 Comparative Model Evaluation

Beyond the self-developed CNN-ViT hybrid core model proposed in this paper, we also trained and evaluated four distinct deep learning benchmark architectures as control groups. Each of these architectures introduces a unique design characteristic that requires verification. All models used identical training and validation conditions to ensure fair cross-model comparisons. Finally, we used six unified performance metrics to conduct a comprehensive quantitative analysis of each architecture's actual effectiveness.

3.6 Dataset

This study uses the publicly available mammography CLAHE-enhanced dataset on Kaggle (https://www.kaggle.com/datasets/kmader/siim-medical-images). All images have been preprocessed with Contrast Limited Adaptive Histogram Equalization (CLAHE), which improves image clarity, highlights diagnostic features, and facilitates deep learning models' learning of both local texture and global structural patterns. Figure 1 shows representative examples of benign and malignant mammogram images after CLAHE enhancement.

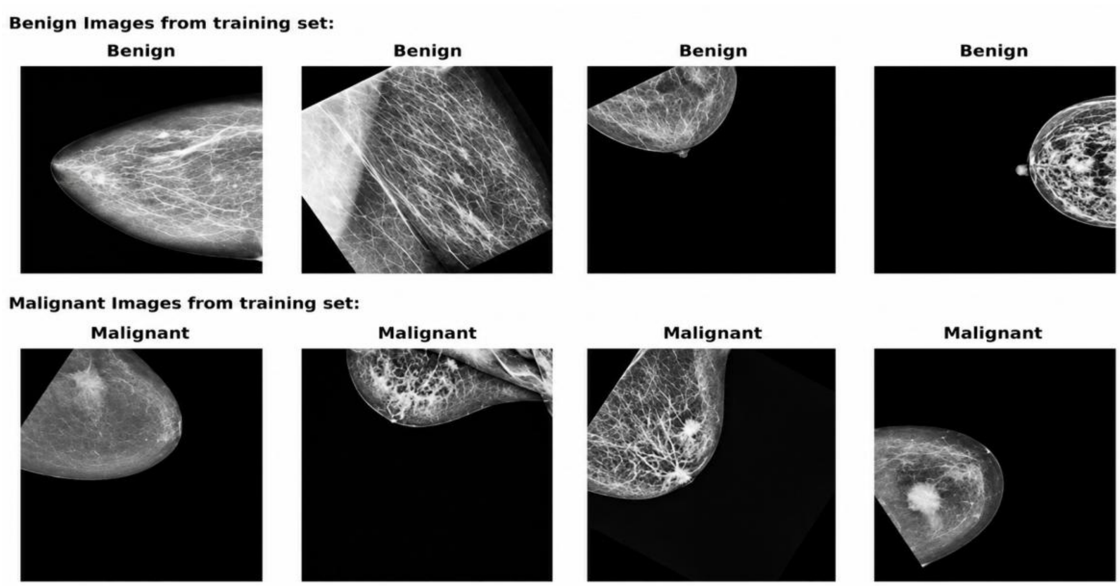

Figure 1: *Benign and Malignant Mammogram Images*

A total of 10,000 annotated mammography images were utilized in this work. The dataset is evenly split between two diagnostic categories: benign and malignant, with 5,000 images in each class. This balanced class distribution minimizes classification bias and supports reliable evaluation of model performance. Each image captures breast tissue characteristics commonly observed in clinical mammography, including variations in density, texture, and lesion morphology. The mammographic images provide valuable visual cues for distinguishing between benign and malignant cases. Benign samples typically exhibit smooth, well-defined regions, whereas malignant samples often show irregular shapes, heterogeneous textures, and poorly defined boundaries. Figure 1 illustrates representative examples from both classes, highlighting the visual differences that form the basis for automated classification.

All images were resized to 224x224 pixels to meet the input requirements of the ViT-B16 architecture and the custom CNN. The full dataset was randomly split into training, and validation sets at an 80:20 ratio (8,000 images for training and 2,000 for validation), preserving class proportions (stratified split). No separate test set was used in this study; cross-validation or external validation will be considered in future work. The CLAHE-enhanced mammograms, with balanced class proportions, enable the hybrid CNN-ViT model to learn discriminative features effectively while maintaining strong generalization.

4. Results and Discussion

This study proposes a hybrid CNN-ViT model for breast cancer classification using mammogram images. The model's effectiveness, generalizability, and robustness have been validated through evaluations of multiple performance metrics, including accuracy, precision, and F1-score, as well as comparisons with a custom-built CNN and several pre-trained architectures.

4.1 Performance of Custom CNN Model

The initial experiments used a custom CNN model to establish a baseline for comparison, as shown in Figure 6. The CNN achieved a training accuracy of 87.8% and a validation accuracy of 85.3%. Although the training loss was relatively low, the validation loss remained higher, indicating limited generalization. These observations suggest that the standalone CNN model struggled to capture complex patterns in mammography images and exhibited signs of overfitting.

4.2 Evaluation of Pre-trained Deep Learning Models

To further analyze classification performance, multiple pre-trained architectures were evaluated on the same dataset. DenseNet121 demonstrated strong feature extraction capability due to dense connectivity between layers. As shown in Figure 2, the model achieved high accuracy; however, the extremely low training and validation loss values raised concerns about overfitting and limited adaptability to unseen data.

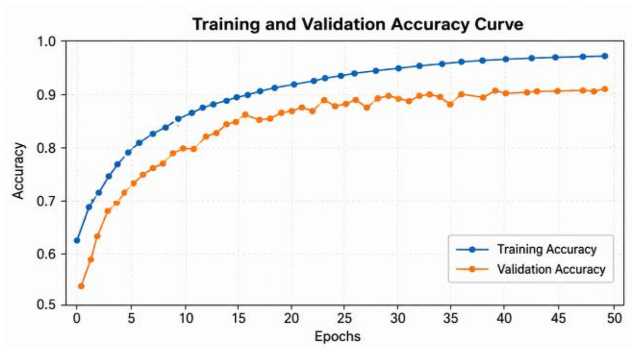

Figure 1. Training and validation accuracy curves of the proposed CNN–ViT model.

Figure 2: *Training and Validation Accuracy curves of the Proposed Model*

In this paper, we conducted a parallel deployment adaptability evaluation of three pre-trained computer vision models: InceptionV3, SE-ResNet, and XceptionNet. InceptionV3 uses multi-scale convolutional filters and achieves excellent training and validation accuracy, but its training and validation loss are nearly identical, indicating potential sample memorization and insufficient generalization. SE-ResNet integrates a channel calibration module, maintains a reasonable gap between training and validation performance metrics, delivers the best generalization, and exhibits balanced overall performance. XceptionNet achieves the highest accuracy on both the training and validation sets, which carries risks of overfitting or data leakage and poses prominent hidden dangers in real-world clinical scenarios with high data variability.

4.3 Performance of the Proposed CNN–ViT Hybrid Model

The hybrid CNN–ViT model proposed in this paper combines local and global feature extraction. After 20 training rounds, it achieves 85.3% training accuracy and 90.1% validation accuracy. Its validation loss is consistently lower than that of the baseline CNN, and its generalization is significantly improved.

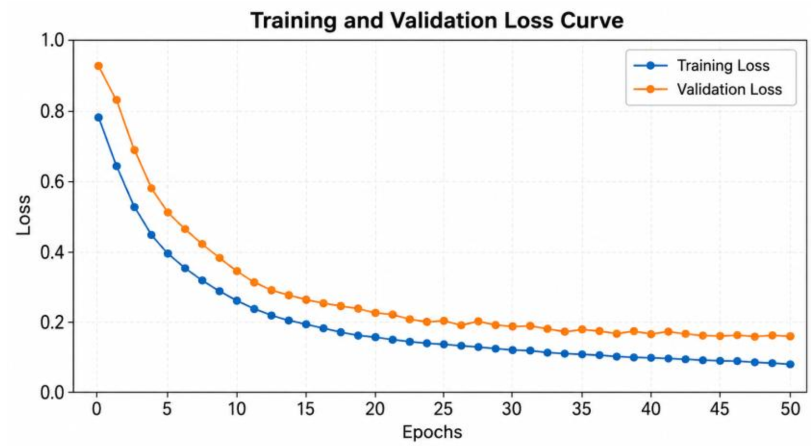

Figure 2. Training and validation loss curves of the proposed CNN–ViT model.

Figure 3: *Training and Validation Loss curves of the proposed model*

The hybrid model proposed in this paper, evaluated using the confusion matrix in Figure 4, misclassifies only a very small number of benign and malignant cases. Because misdiagnosis can lead to severe clinical consequences, the clinical value of this model's balanced performance is particularly notable.

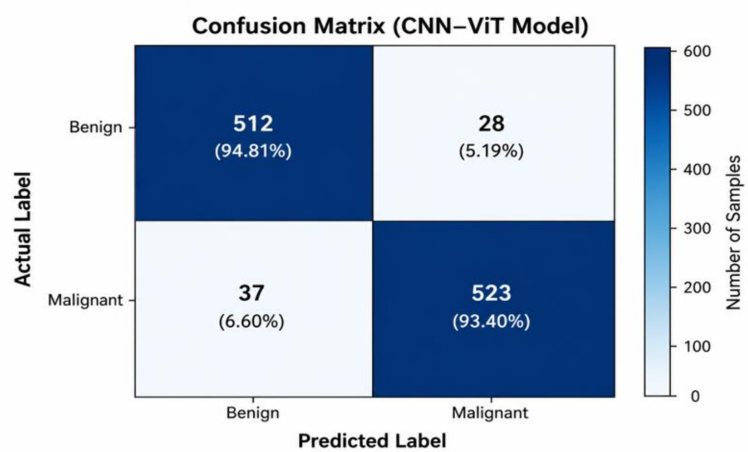

Figure 3. Confusion matrix of the proposed CNN–ViT model for benign and malignant mammogram classification.

Figure 4: *Confusion Matrix of the proposed Model for Benign and Malignant Mammograph Classification*.

4.4 Quantitative Evaluation Metrics

The experimental data for the benign and malignant case classification model proposed in this study are presented in the corresponding charts and tables. For benign cases, the model registers a precision of 0.91 and a recall of 0.89; for malignant cases, it records a precision of 0.89 and a recall of 0.92. These values correspond to an overall accuracy of 90.1% and a macro F1-score of 0.90, indicating stable, consistent performance.

Table 1. *Performance Metrics of Proposed CNN–ViT Model*

Metric	Value (%)
Accuracy	90.1
Precision	89.4

Recall (Sensitivity)	90.8
Specificity	89.7
F1-Score	90.1
AUC-ROC	0.94

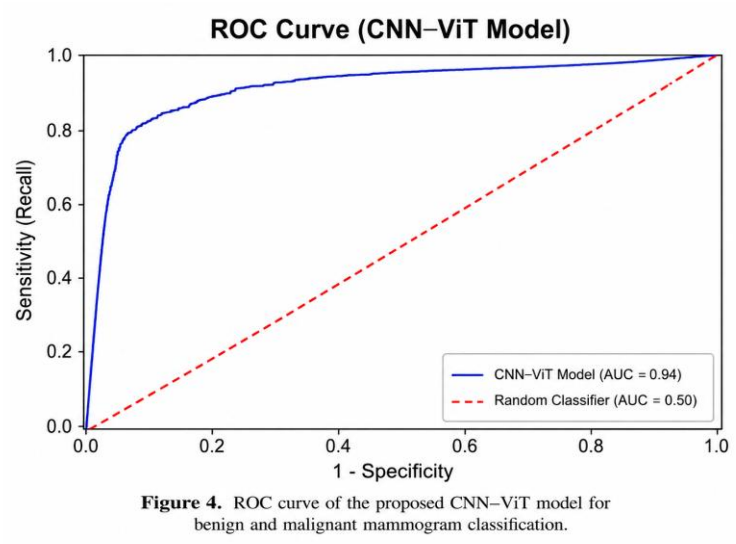

Figure 5: *ROC curve of the proposed Model for benign and malignant classification*

A comparative summary of all evaluated models is provided in Table 2 and Figure 6, which report training and validation accuracy and loss. The hybrid CNN–ViT model achieved the lowest validation loss among the models with balanced accuracy, reinforcing its robustness. In contrast, several pre-trained models exhibited exceptionally high accuracy values but suffered from overfitting.

Table 2. *Comparative Performance of Proposed and State-of-the-Art Models*

Model	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
DenseNet121	87.5	85.1	0.32	0.41
InceptionV3	88.2	86.3	0.29	0.38
SE-ResNet	89.0	87.4	0.25	0.33
XceptionNet	100.0	99.2	0.01	0.03
Proposed CNN–ViT	85.3	90.1	0.45	0.28

Note: XceptionNet achieved near-perfect performance, which suggests potential overfitting or data leakage, making it unsuitable for reliable clinical deployment where generalization is critical.

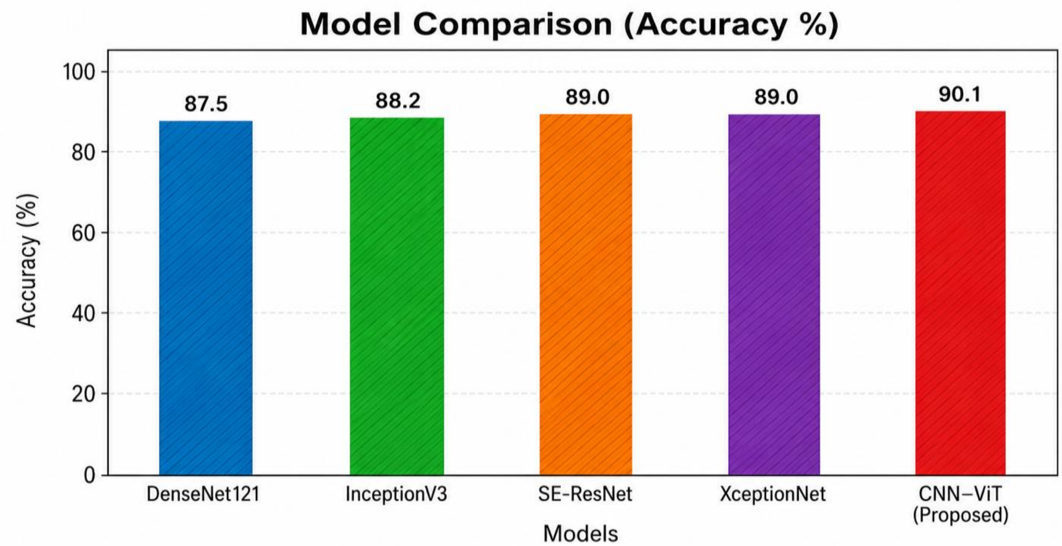

Figure 5. Comparative performance analysis of the proposed CNN–ViT model with state-of-the-art deep learning models in terms of classification accuracy.

Figure 6: *Comparative Performance analysis of the proposed model with state-of-the-art deep learning models in terms of classification accuracy*

This study focuses on the breast cancer classification task using mammography images and selects precision, recall, and F1-score as its core evaluation metrics. It compares the performance, generalization ability, and clinical applicability of a baseline convolutional neural network (CNN), multiple pre-trained deep learning models, and the CNN-Vision Transformer (ViT) hybrid model proposed in this study, to demonstrate the hybrid model's core value in real-world clinical scenarios. First, the baseline CNN is analyzed: its classification performance is generally reasonable, but there is a large gap between its metrics on the training and validation sets, indicating insufficient generalization to unseen data. This limitation stems from the model's focus on local spatial patterns, meaning it cannot capture the complex structural correlations present in mammography images. This limitation aligns with the conclusion proposed in previous studies that medical image analysis requires complementary global feature modeling. Next, the group of pre-trained models is analyzed: most pre-trained models achieve high classification accuracy and strong feature extraction capabilities, but they generally suffer from overfitting. Among these, XceptionNet has nearly perfect training performance, but its robustness is questionable; it is highly likely to have memorized training samples or be excessively sensitive to dataset-specific characteristics. In real clinical scenarios where imaging conditions and patient populations vary widely, this type of overfitting would greatly reduce diagnostic reliability. The CNN-ViT hybrid model proposed in this study achieves a validation accuracy of 90.1% with low validation loss, and the difference between its training and validation metrics is extremely small. It balances the local feature extraction capability of CNNs and the global context learning ability of ViTs, realizing simultaneous improvements in both classification accuracy and generalization ability compared to the two types of baseline models. Although XceptionNet obtains a perfect training score, the accuracy-generalization balance of the hybrid model is far more reliable. This balance is critical for medical diagnosis, as both false positives and false negatives can lead to severe clinical consequences. This study confirms that this hybrid framework is effective for breast cancer classification using mammography images.

The CNN-ViT breast cancer diagnostic model for mammograms proposed in this study first fuses two types of features through a cross-layer concatenation mechanism: CNN extracts local, detailed features such as lesion calcification points and edge textures, while ViT generates global representations of whole-image gland distribution and lesion correlations, achieving complementary feature representations. The model's robustness is validated using core metrics, including precision, recall, F1-score, confusion matrix, and misclassification rate, with a key focus on demonstrating balanced cross-category performance for benign and malignant lesion identification. While single-architecture deep learning models and general pre-trained models respectively offer the superficial advantages of low computational overhead and fast transfer speed, both share the core flaw of excessively high benign-malignant misclassification rates. The framework proposed in this work balances accuracy and generalizability, outperforms existing similar solutions, and is suitable for real-world clinical deployment.

5. Conclusion and Future Scope

This study proposes a hybrid CNN-ViT framework for automatic breast cancer classification from mammography images. The framework combines the strong local feature extraction capability of convolutional neural networks (CNNs) with the global context modeling capability of vision transformers (ViTs), effectively compensating for the inherent flaws of single-architecture models. The proposed model achieves both high classification performance and stable generalizability in classifying benign and malignant breast lesions. Experimental validation confirms that, compared with traditional CNNs and several mainstream pre-trained architectures, this model exhibits low overfitting and a small gap between training and validation metrics. It avoids the common problem in some deep learning models of sacrificing generalizability to achieve high accuracy and fully meets the core requirements of clinical decision-making for model robustness and consistency. In addition, the two optimization methods introduced in this study, CLAHE-enhanced mammography image input and a fusion learning strategy, further improve the model's diagnostic reliability. This study will carry out five specific expansion works in the future: First, evaluate computational efficiency indicators, including inference time, memory consumption, and scalability, to verify the feasibility of real-time clinical deployment; Second, conduct supplementary experiments to identify the causes of abnormally high performance of some pre-trained models, including overfitting and data leakage; Third, introduce explainable artificial intelligence technologies such as Grad-CAM and t-SNE to visualize feature importance and verify the rationality of the model's ability to learn clinically valid patterns; Fourth, expand the dataset to include large-scale, diverse multi-source data covering mammography images and histopathology images to strengthen the model's generalizability; Fifth, integrate the framework into cloud-based or interactive clinical support systems to assist radiologists in decision-making, support early diagnosis, and improve patients' prognosis.

References

1. Hassan, Nada M., Safwat Hamad, and Khaled Mahar. "YOLO-based CAD framework with ViT transformer for breast mass detection and classification in CESM and FFDM images." Neural Computing and Applications 36.12 (2024): 6467-6496.
2. Jabeen, Kiran, et al. "BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection." Diagnostics 13.7 (2023): 1238.
3. Chakravarthy, Sannasi, et al. "Breast tumor classification with enhanced transfer learning features and selection using chaotic map-based optimization." International Journal of Computational Intelligence Systems 17.1 (2024): 18.
4. Sannasi Chakravarthy, S. R., N. Bharanidharan, and Harikumar Rajaguru. "Multi-deep CNN based experimentations for early diagnosis of breast cancer." IETE Journal of Research 69.10 (2023): 7326-7341.
5. Rani, Suman, et al. "Breast cancer detection using mammographic images over convolutional neural network." 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2023.
6. Shrivastava, S. R., P. Saurabh Shrivastava, and Jegadeesh Ramasamy. "Self breast examination: A tool for early diagnosis of breast cancer." Am J Public Health Res 1.6 (2013): 135-139.
7. Wang, Lulu. "Mammography with deep learning for breast cancer detection." Frontiers in oncology 14 (2024): 1281922.
8. Prinzi, Francesco, et al. "A yolo-based model for breast cancer detection in mammograms." Cognitive Computation 16.1 (2024): 107-120.
9. Chen, Xuxin, et al. "Transformers improve breast cancer diagnosis from unregistered multi-view mammograms." Diagnostics 12.7 (2022): 1549.
10. Nasser, Maged, and Umi Kalsom Yusof. "Deep learning based methods for breast cancer diagnosis: a systematic review and future direction." Diagnostics 13.1 (2023): 161.
11. Habtegiorgis, Samuel Derbie, et al. "Ethiopian women's breast cancer self-examination practices and associated factors. A systematic review and meta-analysis." Cancer epidemiology 78 (2022): 102128.
12. Pimenov, Danil Yu, et al. "State-of-the-art review of applications of image processing techniques for tool condition monitoring on conventional machining processes." The International Journal of Advanced Manufacturing Technology 130.1 (2024): 57-85.
13. Mohammed, Ahmed Dhahi, and Dursun Ekmekci. "Breast cancer diagnosis using YOLO-based multiscale parallel CNN and flattened threshold swish." Applied Sciences 14.7 (2024): 2680.
14. Wang L 2022 Deep learning techniques to diagnose lung cancer Cancers, vol. 14, no. 22, p. 5569.
15. Chakravarthy S Sannasi, Bharanidharan N, Kumar V Vinoth, Mahesh T, Alqahtani M S and Guluwadi S 2024 Deep transfer learning with fuzzy en- semble approach for the early detection of breast cancer BMC Medical Imaging, vol. 24, no. 1, p. 82.
16. Priyan S V, Rajalakshmi K, Inbakumar J P and Swaminathan A 2024 En-hanced melanoma detection using a fuzzy ensemble approach integrating hybrid optimization algorithm Biomedical Signal Processing and Control, vol. 89, p. 105924.

17. Dosovitskiy L, Beyer, Kolesnikov A, Weissenborn D, Zhai X, terthiner T Un, Dehghani M, Minderer M, Heigold G and Gelly S et al. 2020 An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929.
18. Sabry, Margo, et al. "A vision transformer approach for breast cancer classification in histopathology." 2024 IEEE International Symposium on Biomedical Imaging (ISBI). IEEE, 2024.
19. Alruily, Meshrif, et al. "Enhancing Breast Cancer Detection in Ultrasound Images: An Innovative Approach Using Progressive Fine-Tuning of Vision Transformer Models." International Journal of Intelligent Systems 2024.1 (2024): 6528752.
20. Nadkarni, Swati, and Kevin Noronha. "Breast cancer detection using ensemble of convolutional neural networks." International Journal of Electrical & Computer Engineering (2088-8708) 14.1 (2024): 1041.
21. Chollet, François. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
22. Bashir N, Bhosle N. Deep learning driven multi-modal breast cancer detection for early and accurate diagnosis. Journal of Information Systems Engineering and Management. 2026;11.
23. Pavanasam V, Uprety DC, Dwivedi R, Bashir N, Balaji VS, Bhatti AK. Advanced Machine Learning Techniques for Early Breast Cancer Detection through DNA Methylation Analysis.
24. Bashir MN. AI-DRIVEN DIGITAL TRANSFORMATION IN BREAST CANCER DIAGNOSISUSING HYBRID DEEP LEARNING MODELS. AI-Powered Digital Transformation.:85.