Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 977-990 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

A Hybrid CNN–ViT Model for Breast Cancer Classification in Mammograms: A Three-Phase Deep Learning Framework

Vandana Saini[1], Meenu Khurana[1], and Rama Krishna Challa[2]

[1] Chitkara University School of Engineering and Technology, Chitkara University, Himachal Pradesh, India
[2] Department of Computer Science and Engineering, NITTTR, Chandigarh, India

Corresponding author: Vandana Saini (e mail: vandana99.phd23@chitkarauniversity.edu.in), **Author(s) Email**: Meenu Khurana (e-mail: meenu.khurana@chitkarauniversity.edu.in) Rama Krishna Challa (e-mail: rkc@nitttrchd.ac.in)

Abstract **Breast cancer is one of the leading causes of death among women worldwide. Early and accurate detection plays a vital role in improving survival rates and guiding effective treatment. In this study, we propose a deep learning-based model for automatic breast cancer detection using mammogram images. The model is divided into three phases: preprocessing, segmentation, and classification. The first two phases, image enhancement and segmentation, were developed and validated in our previous works. Both phases were designed in a robust manner using learning networks; the usage of VGG-16 in preprocessing and U-net in segmentation helps in enhancing the overall classification performance. In this paper, we focus on the classification phase and introduce a novel hybrid deep learning based model that combines the strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). This model captures both fine-grained image details and the broader global context, making it highly effective for distinguishing between benign and malignant breast tumors. We also include attention-based feature fusion and Grad CAM visualizations to make predictions more explainable for clinical use and reference. The model was tested on multiple benchmark datasets, DDSM, INbreast, and MIAS, and a combination of all three datasets, and achieved excellent results, including 100% accuracy on MIAS and over 99% accuracy on other datasets. Compared to recent deep learning models, our method outperforms existing approaches in both accuracy and reliability. This research offers a promising step toward supporting radiologists with intelligent tools that can improve the speed and accuracy of breast cancer diagnosis.**

Keywords **Mammogram; MIAS; Classification; INbreast; DDSM; CNN**

I. Introduction

Breast cancer is the most common cancer among females worldwide, contributing to a high percentage of cancer-related deaths. As reported by the World Health Organization, accurate and early detection of breast cancer drastically improves survival rates and increases treatments success. Among all imaging techniques, mammography is considered the gold standard for early detection because it is inexpensive, accessible, and sensitive to calcifications and soft tissue tumors [2]. Interpretation of mammographic images by radiologists, however, is prone to intra- and inter-observer differences, contributing to false negatives. Thus, there is a need of the development of computer aided diagnosis (CADs), which use digital image processing and machine learning algorithms to automate detection and enhance diagnostic accuracy.

The development of an effective CAD system typically includes three phases: preprocessing, segmentation, and classification. Each phase is important to overall system performance in terms of diagnosis and should be carefully configured to address the complexities and variabilities. Preprocessing, the first phase, involves refining raw mammogram images by reducing noise, improving contrast, and cleansing them further analysis. Median filtering, Contrast Limited Adaptive Histogram Equalization (CLAHE), and rolling ball background subtraction are some of the techniques that have proved to be highly effective in improving the quality of a mammogram image by removing background noise and enhancing the detection of lesions [2], [3].
After preprocessing, image segmentation is essential to separate regions of interest (ROIs) that could be malignant or benign tumors. Segmentation algorithms have to deal with issues like low contrast between

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

tumor and normal tissue, intensity overlap in dense tissues, and the detection of microcalcifications. Different segmentation models have been explored, from thresholding methods like Otsu's approach [4] to advanced morphological algorithms and deep learning approaches like U Net and Mask R CNN. These models try to identify breast boundaries, pectoral muscular tissues, and suspect masses with greater accuracy, thereby providing a strong basis for feature extraction and classification [5].
The last and most critical phase is classification, which decides whether a segmented lesion is normal, malignant, or benign. Traditionally, classifiers such as support vector machines (SVMs), decision trees, and k nearest neighbors (KNN) based on hand-engineered features like texture, shape, and intensity . however, with the introduction of deep learning in recent times, especially convolutional neural networks (CNNs), this field was changed due to their ability to learn high-level representations with large training sets [6]. These models not only surpass traditional classifiers in accuracy but are also able to get high-dimensional complex mammographic data.
A number of research papers have illustrated the performance of complete pipelines encompassing all three phases. Mohamed et al., for example, built a complete CAD system based on morphological operators and Otsu's thresholding for segmentation, followed by shape-based feature extraction and artificial neural networks (ANNs) for classification. They achieved over 93% accuracy on standard benchmark datasets like DDSM [1]. Similarly, Karunya and Rahimunnisa also proposed an adaptive clustering-based segmentation approach with coupled SVM classification with a 98.13% classification accuracy, thereby underscoring the role of domain-specific preprocessing and segmentation methods in enhancing classifier performance [2].

A major improvement has been made using features based on texture through Gray level occurrence Matrix (GLCM) and Local Binary Pattern (LBP), improving classifiers' discriminative ability. Chanda et al. stated that through the application of statistical descriptors such as entropy, skewness, and standard deviation, their segmentation classification pipeline could achieve 89% sensitivity and 74% specificity with robust performance despite images with poor contrast [3]. Beyond this, the combination of deep learning with handcrafted features offers a hybrid approach that integrates domain knowledge with data-driven knowledge [6].

In addition, the difficulty in identifying between dense and non-dense breast tissues remains a limiting factor in mammography. Tzikopoulos et al. proposed an innovative approach integrating breast density estimation with asymmetry detection through a fully automated segmentation pipeline, achieving 85.7% accuracy for classification on the mini MIAS dataset [7]. Their study highlights the relevance of integrating anatomical and physiological breast features into CAD systems. Advances in deep learning models have paved the way for new avenues for automated diagnosis. Swapna's Deep CNN with dropout and zero-padding strategies showed better performance in discriminating between benign, malignant, and normal tissues, especially in densely packed mammographic images, in which standard classifiers perform well [8]. Like this, Umamaheswari et al. employed a Vision Transformer-based architecture (ViT MAENB7) for 3D mammography with an impressive 96.6% accuracy in classification, thereby pushing the boundary of what can be achieved with deep neural networks for volumetric data [9]. Mustafa et al. also conducted another study that utilized median and Weiner filters for image enhancement, followed by thresholding and morphological analysis, achieving classification rates of over 93.71%. It supports the validity of classical image processing techniques when accurately fine-tuned [10]. Chanda's previous study on K-means segmentation using a decision tree classifier for DDSM images also justifies the value of balancing accuracy and explainability in medical CAD tasks [11]. Lastly, some recent advancements show that optimization methods play a significant role in enhancing segmentation accuracy. Pawar et al. proposed the Firefly Chicken Swarm Optimization (FF CSO) algorithm for optimizing feature selection and classification in mammography images, showing a synergistic potential exists between bio-inspired algorithms and deep learning [12]. To conclude, automated breast cancer detection systems based on mammography depend on integration of preprocessing, segmentation, and classification models. While standard techniques have paved the way, integration of traditional image processing techniques with contemporary machine learning and deep learning models holds the promise of improved accuracy, resilience, and clinical applicability. Multimodal imaging, explainable AI models, and real-world deployment in medical environments are areas expected to be investigated in future studies to decrease breast cancer-related mortality owing to early and accurate detection.

II. Literature Review

Breast cancer is still a major public health concern, with early detection playing an important role in lowering mortality and enhancing patient treatment. Mammography is still considered to be the optimal imaging application for early screening. Over the past few years, improvements in machine learning, especially deep learning (DL), have revolutionized breast cancer diagnosis. A combination of advanced

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

preprocessing methods, precise segmentation algorithms, and good performance classification models has resulted in increasingly accurate Computer Aided Diagnosis (CAD) systems.

A. Preprocessing and Segmentation Techniques

Preprocessing plays an important role in enhancing image quality for better visualization of delicate anomalies. Swapna [13] employed the Rolling Ball approach to remove artifacts from backgrounds and enhance image contrast by utilizing CLAHE and unsharp masking. Similarly, Sreevani and Latha [14] implemented Wiener filtering and logarithmic transformation to remove noise and improve contrast, thereby optimizing the subsequent segmentation and classification processes. Some studies have focused on specialized preprocessing filters. For instance Ghrabat et al. [15] built a preprocessor that removes pectoral muscles through area-expanding segmentation to largely improve the ROI identification. Khdhir et al. [16] applied Residual Pixel Removal together with Gaussian filtering to enhance relevant breast tissue areas at the expense of irrelevant background textures.

Effective segmentation is key to effective lesion localisation and classification. Deep learning performs well in describing tumor areas from complex mammographic images. Gerbasi et al. [17] presented DeepMiCa, a UNet-based semantic segmentation network efficient in detecting microcalcifications with a specialized loss function for small lesions. Tiryaki [18] proposed a cascaded deep transfer learning method using Unet++ with Xception backbone for mass segmentation. It achieved high Dice and Intersection over Union (IoU) values, reflecting strong segmentation performance over high-density mammogram datasets. Similarly, Sinha et al. [19] proposed a region-based segmentation coupled with a ResNet architecture, with segmentation accuracy over 98% with transfer learning. Segmentation with hybrid optimization is also in focus nowadays. Pawar et al. [20] incorporated active contour-based segmentation with Firefly and Chicken Swarm Optimization to improve boundary detection in complex mammograms. Rathinam et al. [21] developed an Adaptive Fuzzy C Means segmentation approach integrated with a VGG Net classifier, achieving improvements in segmentation speed and efficacy through optimized centroid selection.

B. Robust and Deep Learning Frameworks

The ultimate goal of deep learning frameworks is classification. Islam et al. [22] proposed a deep learning network that learns classification, localization, and segmentation simultaneously through a multitask loss function with 98.34% test accuracy. Sinha et al. [23] implemented a hybrid segmentation-classification approach based on UNet with VGG 19 for the classification task. Their approach was 2.25% better than VGG 19 in accuracy, underlining the need to integrate the segmentation and classification process. R. Remya and Hema Rajini [24] used DenseNet 169 for feature extraction and Multilayer Perceptron (MLP) for classification to produce impressive accuracy levels on benchmark databases.

Transfer learning plays a critical role in enhancing classification accuracy, particularly when working with limited data. Meegada et al. [25] proposed a deep location network (DLN) for classifying images without explicitly labeling ROIs. Their network adopted a region scoring and deep pooling modules, achieving competitive performance on the CBIS-DDSM and INBreast databases. Various DL architectures have been tested for mammogram classification. Almutairi et al. [26] employed a hybrid pipeline combining a Universal Sentence Encoder Network (USE Net) and CaffeNet, with an optimized random forests classifier with XGBoost, and achieved 98% classification accuracy on mammogram images. Singh and Mishra [27] proposed a CNN-based end-to-end approach that uses a mass and microcalcification detection systems for segmentation maps to enhance the accuracy of whole exam classification.

In a significant study, Leung and Nguyen [28] introduced an innovative deterministic deep learning model that can generalize across different datasets. Their model was particularly focused on data mining with automatic ROI localization, eschewing the overfitting issues that often characterize traditional DL methods. Interpretability and robustness are still significant issues in medical DL applications. Bouzar Benlabiod et al. [29] integrated segmentation using a U Net with a Case Based Reasoning (CBR) system to produce explainable output for classifying a mammogram. This integration allowed clinicians to visualize the reasoning process of the model, thereby establishing trust in the system's decision.

Gerbasi et al. [17] also addressed interpretability by adding explainable AI components to DeepMiCa, allowing for visually examining activation maps for classified microcalcifications. These visual indicators are critical in borderline cases with high diagnostic uncertainty. Additionally, ensemble and multiscale learning methods have been investigated for enhancing classification robustness further. Kaur et al. [30] proposed a Patch-based Multiscale All Convolution Neural Network (MACNN), which improved classification accuracy from 81% to 88% by using localized image patches. Jassim Ghrabat et al. [15] highlighted a fully automated pipelines that optimize every step, starting from preprocessing, segmentation, and classification using specified hyperparameter

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

adjustments. Optimization is important in fine-tuning deep networks for medical image classification. Sreevani and Latha [14] improved their graph convolutional recurrent neural network (GCRNN) model using Aquila Optimizer for optimizing hyperparameters. Optimization using a metaheuristic approach resulted in classification accuracy rates over 99.6%. Similarly, Khdhir et al. [16] utilized Antlion Optimization to perform segmentation of mammogram images to obtain strong textural features from GLCM matrices with high precision and recall scores. These papers indicate that combining evolutionary optimization with DL structures hugely improves diagnostic accuracy with reduced training overhead.

C. Dataset and Generalization Challenges

One enduring difficulty in deep learning is the inability of models to generalize effectively across datasets with different distributions. Several authors have reported that models trained on one dataset tend to perform poorly upon testing with different ones. Leung and Nguyen [28] tried to address this by employing a deterministic design and automatic ROI mining to mitigate domain shift.

One of these is multimodal or ensemble learning, in which classifiers are trained on various subsets of features or variations of inputs. Almutairi et al. [26] applied CNN with an ensemble of classifiers to address this problem, with DeepMiCa [17] employing patch-wise classification to improve generalizability.

III. Method

The proposed breast cancer detection framework is structured in a three-stage pipeline: segmentation, preprocessing, and classification. Each stage offers essential operations to convert raw imaging information into sound diagnostic products. Whereas our previous work dealt with the first two phases, the image quality improvement and anatomical segmentation, the present study particularly focuses on the classification phase, which is the most critical for accurate diagnosis.

A. Dataset

In this work, mammogram images from multiple datasets (DDSM, MIAS, INbreast, and a combination of three) were obtained from kaggle (https: https://www.kaggle.com/datasets/emiliovenegas1/mammography dataset from inbreast mias and ddsm). These images provide real-world variability which is important for training robust models and testing. In order to make rigorous testing possible, the dataset was split between training (70%), validation (15%), and testing (15%) subsets. All of these images had a resolution of 227 × 227 pixels.

B. Preprocessing

The first phase removes common shortcomings in raw mammogram images, like poor contrast, noise, and artifacts. As presented in our earlier work, we developed a VGG-inspired Convolutional Neural Network (CNN) based denoiser, especially designed to handle the noise patterns found in mammographic images [31]. The preprocessing pipeline of image enhancement, denoising, and ROI centric segmentation was performed as elaborated in our previous published papers [31, 32]. Augmented mammogram images were generated utilizing standard geometric transformations like random horizontal flipping, small rotations (±15 degrees), zoom-in scaling, and translation shifts in order to enhance model generalizability. All augmented images were resized uniformly to a 224×224 resolution and assigned 3 channel grayscale to represent the input image format of pretrained backbones. Normalization utilized standard ImageNet statistics to align with pretrained model expectation. To address class imbalance, stratified splitting was employed to ensure equal proportions of classes in the training, validation, and test sets. A weighted loss formulation was also employed implicitly by the internal handling in the AdamW optimizer of sparse class representations to prevent bias in the updates. Overall, these steps ensured class robustness, specifically for malignant samples, which are commonly underrepresented in clinical datasets. This model incorporated existing architecture components like multi-depth convolutional blocks, skip residual connections, and up-sampling layers to efficiently recover the image without losing diagnostically significant structures. Our preprocessing model performed better with a PSNR of 79, surpassing many standard and learning based enhancement techniques. Images improved from this step were utilized as inputs for both segmentation and classification to provide consistently high-quality images across the pipeline.

C. Segmentation

In the second phase, a hybrid segmentation approach that merged Otsu thresholding, morphological filtering, and U-Net was used to separate regions of masses from improved mammogram images [32]. The segmentation model was built to retrieve uneven breast boundaries along with fine-grained tumor areas. Our approach demonstrated high sensitivity in outlining irregularly shaped tumors with diffused boundaries, solving a significant impasse present in classical thresholding and contour-based approaches. Average scores of 0.99 and 0.98 were achieved by the ROIs validated through Intersection over Union (IoU).

D. Classification of Mammogram Images

To carry out classification, the suggested model applies a parallel dual-branch architecture combining ResNet-34 and ViT-Tiny for collaborative feature learning. The spatial features of the input mammogram captured by the ResNet-34 component involve a series of four

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

residual steps consisting of convolutional layers and identity maps, terminating in a global average pooling layer that produces a vector representation with 512 dimensions. Concurrently, the ViT Tiny model segments the input into 16 x 16 patches and embeds each into a vector of 192 dimensions before processing the resulting sequence with twelve transformer encoder layers containing three self-attention heads and a multilayer perceptron layer with a hidden dimension of 768. The representation from the classification token in the final transformer layer is taken as a global descriptor with 192 dimensions. These two representations are concatenated to form a feature vector comprising 704 dimensions, which is then passed into a gated attention fusion module to conduct joint modeling. This architecture enables the model to capture fine-grained local structures and long-range dependencies in the mammographic image simultaneously, thereby enhancing its discriminative capacity. The combination of ResNet-34 and ViT Tiny balances architectural depth and computational efficiency, aiming for precise classification while remaining suitable for real-time application in a clinical environment. The overall network flow architecture is as shown in Fig. 1.

1. Hybrid CNN-ViT Architecture

The proposed CNN-ViT hybrid model uses the strengths of both convolutional and transformer-based neural networks by employing ResNet-34 to extract hierarchical spatial features and ViT Tiny to capture long-range dependencies from mammogram images. After passing the input mammogram through both ResNet-34 and ViT Tiny branches, feature representations are extracted from specific intermediate outputs for fusion. In the CNN path, the spatial features are obtained from the production of the global average pooling layer of ResNet-34, producing a 512-dimensional vector that summarizes local patterns and structural intensity variations across the image. In parallel, the ViT Tiny model processes the input image as a sequence of 16×16 patches, each linearly embedded into a 192-dimensional vector and enriched with positional encodings. A learnable summary vector is prepended to the patch sequence and, after passing through all 12 transformer encoder blocks, this vector is extracted to yield a 192-dimensional global descriptor representing the image's contextual information. These two output vectors one from the CNN path and the other from the transformer are concatenated into a unified 704-dimensional feature vector.

Prior to fusion, no additional dimensionality reduction is applied, as both vectors are already flattened and compatible. The merged representation is fed into a trainable, attention-based fusion module that adaptively reweighs feature contributions from both modalities. This dual-branch feature integration forms the backbone of the final decision layer. The feature extraction process in CNN and ViT is done as shown in Eq. (1) where I is the input image, F_CNN and F_ViT are features from ResNet34 and ViT-Tiny [8], [13] and their outputs are concatenated to form a joint feature representation described in Eq. (2) where F_concat is the combined feature vector. To address redundancy and enhance salient feature weighting, a trainable attention gating mechanism is applied, which assigns optimal weights as expressed in Eq. (3) where α is the attention weight, W_a and b_a are trainable weight and bias, F_fused is the attention-weighted feature. The fused features are then passed through a GELU (Gaussian Error Linear Unit) activation layer, which helps map the learned representation into a latent space for better classification, as shown in Eq. (4). In this equation, W_f and b_f are the weight and bias, Z is the GELU-activated latent feature, a technique also adopted in [9,34]. Finally, the classification scores are calculated using the Softmax function in Eq. (5) where W_c and b_c are classification layer parameters, y_hat is the predicted output, which is widely applied in medical image classification tasks [22]. This overall mechanism ensures that both localized patterns and global structures are emphasized, providing strong interpretability and robustness.

$$F_{\{CNN\}} = ResNet34(I), \quad F_{\{ViT\}} = ViT_{\{Tiny\}}(I) \tag{1}$$

These are concatenated to form a combined representation (equation 2):

$$F\_\{concat\} = [F\_\{CNN\} \parallel F\_\{ViT\}] \tag{2}$$

To reduce redundancy and emphasize relevant features, a gated fusion is applied (equation 3):

$$F\_\{fused\} = \alpha \odot F\_\{concat\}, \quad where \quad \alpha = \sigma(W\_a\, F\_\{concat\} + b\_a) \tag{3}$$

The fused features are projected into a latent space using GELU activation (equation 4):

$$Z = GELU(W\_f\, F\_\{fused\} + b\_f) \tag{4}$$

Finally, classification is performed via a Softmax layer (equation 5):

$$\hat{y} = Softmax(W\_c\, Z + b\_c) \tag{5}$$

2. Model Training

All input images were rescaled to a uniform size of 224×224 pixels to adhere to the input size specifications of the ResNet-34 and the ViT Tiny architectures. Since mammogram images are originally in grayscale, every image was converted into a three-channel image by repeating the single channel across the RGB dimensions, ensuring compatibility with pretrained models expecting 3-channel input. No random or center cropping was performed, as maintaining intact anatomical architecture is important in clinical interpretation. The image is normalized

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

concerning the standard ImageNet standard deviation before being input into the model. The model output is a two-element vector of class probabilities for malignant and benign labels generated by a softmax activation operation. The class with the greatest predicted probability determines the final classification result, i.e., argmax(ŷ) of the softmax output. A 0.5 default threshold is applied, however, the system preserves the unnormalized probability scores of each class and returns them in the GUI output with Grad CAM heatmaps for interpretability. This dual output form class label and class probability facilitates both clinical decision and model confidence estimation.

The CNN and ViT branches and the attention-based

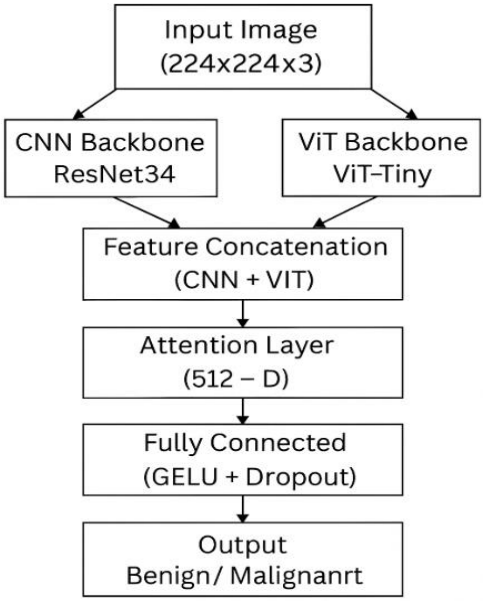


Fig. 1. Proposed workflow of Hybrid CNN-ViT architecture for mammogram images classification

fusion module were jointly optimized in a one-stage, end-to-end training process. All modules remained unfrozen during training to facilitate full backpropagation of gradients and learning co-adaptation. The parameters of the model are jointly updated with the backpropagation of the binary cross-entropy losses, which suits binary classification tasks such as the classification of malignant and benign lesions. Dropout has its integration inherently in the transformer path, while weight decay regularization via AdamW and batch normalization in the CNN branches are used to provide stable convergence. This training design promotes the fine-tuning of both local and global representations simultaneously and facilitates high classification performance and the model's generality. Hyperparameters such as the learning rate (1×10^{-4}), batch size (32), and early stopping patience (5 epochs) were finalized through a manual grid search on the validation set using macro F1 score and loss as evaluation criteria. Training was conducted for a maximum of 50 epochs, with early stopping typically halting training between epochs 25–35, depending on dataset complexity. To mitigate overfitting, we used early stopping, data augmentation, dropout within the ViT layers, and batch normalization in the CNN pathway. Validation was performed on a stratified 15% holdout set, ensuring consistent class balance and robustness of evaluation. All training and inference tasks were executed on a workstation equipped with an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM) and 32 GB system RAM, using PyTorch 1.13 with CUDA 11.6 on Ubuntu 22.04. No distributed training or model parallelism was employed, as the architecture fits within a single GPU memory envelope. Total training time ranged from 2.5 to 4.5 hours, depending on dataset size and augmentation settings. These parameters and setup confirm that the model is not only reproducible but also feasible for deployment in clinical environments with affordable GPU infrastructure. To confirm model interpretability, Grad CAM overlays are used with testing images, which identify pathology-relevant regions, ultimately to improve transparency and trustworthiness of automated outputs. The detailed framework is furtherdiscussed in Algorithm 1.

In the proposed hybrid classification framework, the input grayscale mammogram image is initially resized and normalized to the input specifications of the model, as shown in Fig 2. The image is then transformed into three-channel form through replication and normalized with ImageNet statistics. The preprocessed image is fed to two pathways in parallel: a ResNet-34-based CNN for extracting spatial features and a Vision Transformer (ViT) pathway for extracting global attention-based features. The process of extracting features through two pathways is mathematically represented in Eq. (6) and Eq. (7), motivated by earlier works on multi-branch fusion architectures [8], [13]. The resultant features are then concatenated and input into a trainable attention-based fusion block to generate a combined representation as defined in Eq. (8). The mechanism of attention dynamically balances the contribution of each pathway's features, which maintains consequential patterns and prevents redundant information, as seen in the feature selection methods presented in [20], [21]. The integrated features are then fed through a GELU activation and projected to a latent embedding space, as represented in Eq. (9), which makes them more classification-ready. The ultimate logits for binary classification are derived through a fully connected layer and a softmax function, as represented in Eq. (10), following standard classification pipelines discussed in [22], [26]. Modeloptimization is undertaken using the binary cross entropy loss as represented in Eq. (11), an objective function widely used in medical image

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

classification problems [3], [5]. For obtaining interpretability of model decisions, Grad CAM is added to identify the active regions resulting in classification. The approach for interpretability is introduced as represented in Eq. (12), and it justifies model prediction through pathologically related area localization in the mammogram image, following the methodology adopted in [6], [7], and [9].

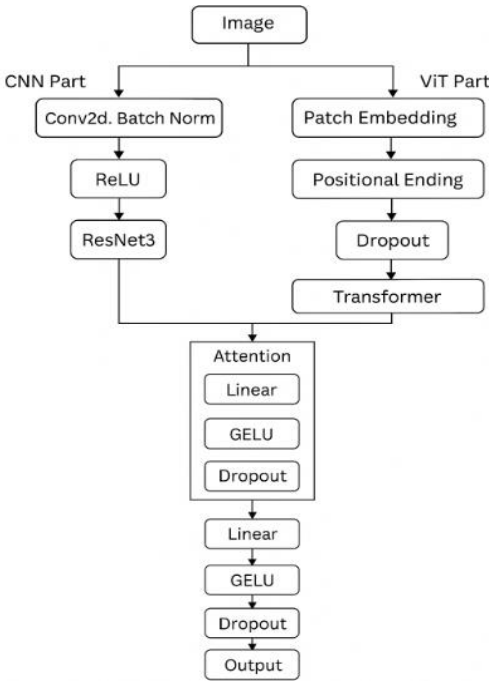


Fig. 2. Architecture of the proposed hybrid CNN-ViT model with attention mechanism

Algorithm 1. Hybrid CNN-ViT Classification with Grad CAM and Early Stopping

Symbols

F_CNN and F_Vit – Features extracted from ResNet34 and ViT Tiny

α (alpha) – Attention weight vector

Z: Latent representation after GELU

ŷ (y-hat) : Final predicted class probabilities

L_CE: Cross-entropy loss

W_a, b_a, W_f, b_f, W_c, b_c: Weights and bias of attention layer, GELU and classification layer

M_grad: Grad-CAM activation heatmap

A^k: Feature map from k-th convolution channel

α_k: Weight for k-th feature map in Grad-CAM

1 Preprocessing: For each image I in dataset D:
I_gray ← Grayscale(I)
I_resized ← Resize (I_gray, 224×224)
I_rgb ← RepeatChannels(I_resized, 3)
Normalize using ImageNet stats:
I_norm(i,j,k) = (I_rgb(i,j,k) μ_k) / σ_k (6)

2 Dataset Split:
D_train, D_val, D_test ← StratifiedSplit(D)

3 Model Setup:
F_CNN ← ResNet34(I_norm)
F_ViT ← ViT_Tiny(I_norm)
F_concat ← [F_CNN ‖ F_ViT]
α ← σ(W_a · F_concat + b_a)
F_fused ← α ⊙ F_concat
Z ← GELU(W_f · F_fused + b_f) (7)

4 Loss Function:
L_CE = − ∑ y_i log(ŷ_i) (8)

5 Optimization:
θ ← θ − η ∇_θ L_CE + λθ (9)

6 Early Stopping:
Train using AdamW (lr = 1e−4)
If validation loss does not improve for 5 epochs → stop

7 Inference:
ŷ = Softmax(W_c · Z + b_c) (10)

8 Grad CAM Visualization:
For last conv layer A^k:
α_k = (1 / Z) ∑_i ∑_j ∂L_CE / ∂A^k_ij
M_grad = ReLU(∑_k α_k · A^k) (11)

9 Evaluation:
F1 = 2 * (Precision * Recall) / (Precision + Recall) (12)
Return: ŷ, M_grad, Accuracy, F1, Confusion Matrix

3. Novelty

The novelty of the introduced framework is in its concurrent dual stream structure, utilizing ResNet-34 for localized spatial encoding and ViT Tiny for global contextual relationships capture in mammogram images provides richer feature representation compared to sequential single branch architectures. A gated attention-based fusion mechanism is proposed to learn to weigh and merge the CNN and ViT features adaptively to sidestep the noise and redundancy of simple concatenation. Grad CAM is also integrated in the structure for direct visual interpretation of the prediction. The system is optimized for real-time inference in multi-GPU settings for deployment in the clinic.

IV. Results

This section evaluates the performance of the proposed hybrid model, which integrates ResNet34 and Vision Transformer (ViT) with attention-based fusion, on four benchmark mammography datasets. The classification performance is assessed using standard metrics: Accuracy, Precision, Recall, and F1

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

Score for each class, along with macro-averaged scores. Table 1 presents the classification accuracy of our proposed hybrid CNN-ViT model on four benchmarking mammogram datasets. Metrics such as class wise Precision, Recall, F1 Score, along with macro averaged scores, are presented. High accuracy is exhibited by the model for all datasets, with performance up to 100% on MIAS and over 99% on the rest of them. These findings validate the robustness of the model as well as its ability to generalize to varied imaging conditions.

Table 1. Classification performance in terms of accuracy across all the datasets

Dataset	Class	Precision	Recall	Accuracy (%)
DDSM	Benign Masses	0.9943	0.998	99.62
	Malignant Masses	0.9994	0.998	
	Macro Average	0.9968	0.998	
Nbreast	Benign Masses	0.9991	0.994	99.49
	Malignant Masses	0.9943	0.999	
	Macro Average	0.9967	0.9966	
MIAS	Benign Masses	1	1	100
	Malignant Masses	1	1	
	Macro Average	1	1	
INbreast+ MIAS+DD SM	Benign Masses	0.9989	0.9957	99.54
	Malignant Masses	0.9998	0.9951	
	Macro Average	0.9952	0.9954	

To compare the performance of the hybrid CNN-ViT model, comprehensive experiments were performed against four benchmark datasets for mammographs: DDSM, INbreast, MIAS, as well as a merged dataset consisting of INbreast+MIAS+DDSM. The performance was measured by Accuracy, Precision, Recall, F1 Score, as well as Confusion Matrix for all four datasets. The model performed exceptionally, achieving over 99% accuracy for all datasets while scoring a perfect mark in MIAS. The performances show not only high accuracy in classification but also great generalizability to multiple real-time datasets. Table 2. illustrates the consolidated confusion matrix for each of these models.

Table 2. Consolidated Confusion Matrix

Dataset	TP	FP	FN	TN	Accuracy (%)
DDSM	256	1	0	259	99.62
	259	0	1	256	
INbreast	253	0	5	258	99.49
	258	5	0	253	
MIAS	258	0	0	258	100.00
	258	0	0	258	
Combined Dataset	252	2	3	259	99.54
	259	3	2	252	

The findings obtained for the proposed hybrid CNN-ViT based architecture on all four datasets DDSM, INbreast, MIAS, and the combined dataset, prove its strength, effectiveness, and generalizability in classifying mammograms. The model recorded exemplary performance, with a spotless 100% on the MIAS dataset and near identical performance on the other datasets, assuring its reliability and diagnostic value. Success lies in a number of architecture and training innovations. The hybrid architecture effectively combines ResNet34 for local spatial feature extraction with ViT Tiny for long-range dependency capturing, allowing the model to understand fine-grained details as well as global tissue patterns in mammographic images. Addition of an attention based fusion mechanism reinforces the system with optimal weighting of the CNN and ViT feature streams, further allowing the model to prioritize diagnostically significant patterns even on occasions of subtle or overlapping visual indicators. Furthermore, the model's high and uniform generalization on different datasets, each with varying sets of patient, resolutions, and modalities, implies that it is considerably flexible and resistant to overfitting. The consistency is further enabled by a well-organized, structured training methodology with the use of the AdamW optimizer, learning rate scheduling, cross entropy loss, and early stopping, all of which are designed for stable convergence. In addition, ImageNet-based normalization assists in harmonizing mammographic image distributions with pretrained model weights, allowing for efficient transfer learning. Lastly, the model exhibits an outstanding precision/recall balance, as embodied in class-wise F1 scores, critically in medical imaging, in order to suppress both false positives, preventing unnecessary intervention, and false negatives, preventing missing any malignancies. The use of Grad CAM visualization exposed a profound understanding of the decision-

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

making process of the model. Attention heatmaps always concentrated on areas with lesions or structural abnormalities in the mammograms, ensuring that predictions were based on clinically relevant information.

A. Real-Time Classification Results using GUI Interface

In order to test the proposed model's usability and ability to perform real-time inferences, we developed an executable desktop application using Python and Tkinter. This application allows users to upload any given mammogram image (can be an image from outside of the dataset), view instant classification output along with Grad CAM visualization for explainability. The prediction outcome is presented in textual format, along with the class confidence score assigned (probability). Grad CAM heatmaps are produced for every image to indicate regions of interest that contributed to the model's prediction. Radiologists and users get visual assurance of model focus areas for classification.

After uploading a mammogram image, the system

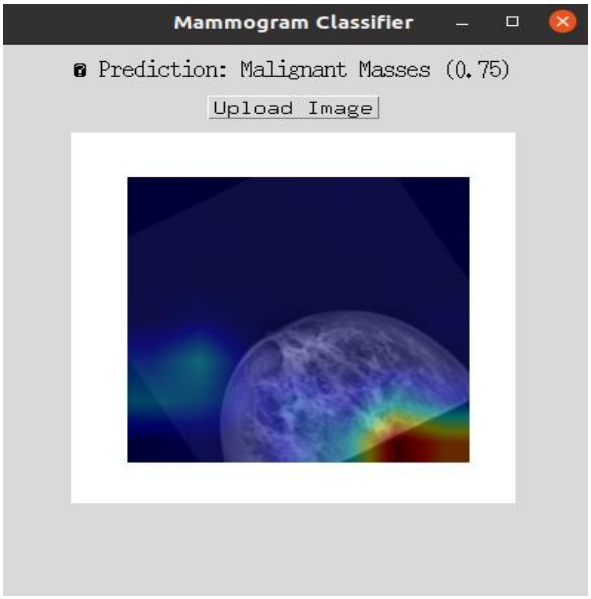


Fig. 3. GUI Grad-CAM output showing prediction as Malignant Mass (Confidence = 0.75)

processes the input, performs prediction using the trained hybrid model, and displays the predicted class along with Grad CAM visualization. As shown in Fig. 3. the model correctly identifies the input image as Malignant Mass with a prediction confidence of 0.75. The Grad CAM overlay emphasizes the suspicious region, demonstrating the model's attention. Similarly, Fig. 4. displays another test case classified as Benign Mass, with a confidence score of 1.00. The highlighted regions in the Grad CAM image suggest areas of interest that align with benign characteristics. In the test display, Grad CAM is implemented with a goal of improving model prediction interpretability. Grad CAM produces a heatmap overlay over the original mammogram image, visually indicating which regions contributed most to the model's choice. It not only aids in ensuring that the model is looking at relevant pathologic features like masses or lesions but also establishes trust and transparency for medical users. By giving class-specific discriminative regions, Grad-CAM operates like a visual explanation aid, enabling radiologists and scientists to comprehend the rationale behind the model's classification, particularly in differentiating between malignant and benign masses.

This GUI-based test validates the robustness of the

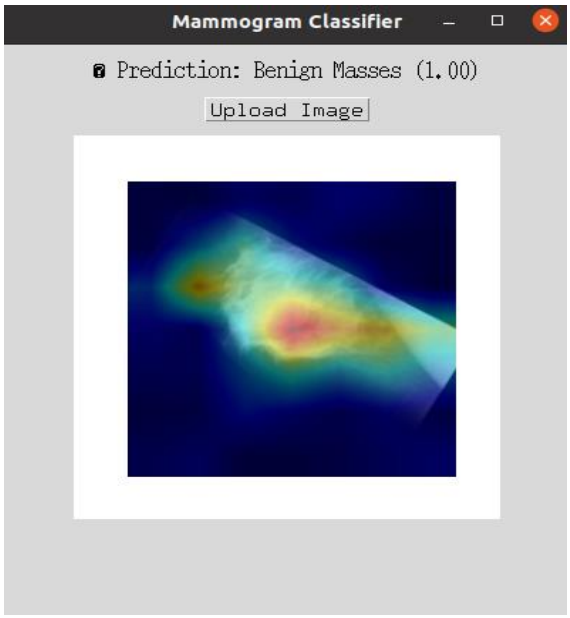


Fig. 4. GUI Grad-CAM output showing prediction as Benign Mass (Confidence = 1.00)

proposed hybrid model on unseen real-world samples, and its performance. It also demonstrates practical deployment feasibility for clinical or diagnostic setups with minimal computational resources.

V. Discussion

The proposed hybrid CNN–ViT model shows excellent and stable performance in classifying mammograms on four benchmark datasets with up to 100% accuracy on MIAS and over 99% accuracy on DDSM, INbreast, and combined INbreast+MIAS+DDSM data. The performance confirms the generalizability of the model to various patient populations, image conditions, and distributions of datasets. The model's success can be attributed to its dual-branch structure, in which ResNet34 extracts spatially local diagnostic features while ViT Tiny encodes long-range dependencies to

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

Table 3. Performance comparison of proposed classification model with existing models on basis of accuracy

Ref.	Learning Model	Dataset	Accuracy (%)
[36]	Multi CNN + Feature Concatenation (BWM MCDM, MI)	MIAS (augmented)	98.74
[37]	ResNet50 + SMOTE + FC Layers	Custom Mammograms	99.00 (balanced), 90.00 (imbalanced)
[38]	Fine-tuned SWIN Transformer	Not specified	**99.9**
[39]	Customized Dual CNN	MIAS, INbreast, CBIS DDSM	98.78 (MIAS), 97.84 (INbreast)
Proposed method	CNN-ViT Hybrid (ResNet34 + ViT Tiny + Attention Fusion)	MIAS, INbreast, DDSM (individually & combined)	**100.00 (MIAS)**, 99.62 (DDSM), 99.49 (INbreast), 99.54 (Combined)

enhance context awareness. The attention-based fusion mechanism further enhances fusion by adaptive weighing significant cross-modal features and achieving high discriminability.

In comparison with existing solutions, or approach surpasses previous method proposals, as shown in Table 3. For instance, Swapna [13] concentrated solely on segmentation boosting and did not apply global attention mechanisms. Umamaheswari et al. [9] utilized ViT for 3D mammography, but its architecture did not support a dual stream CNN–ViT pipeline that hinders its capacity to extract spatial detail information. Kollem et al. [34] utilized CNN ensembles but did not support mechanisms for interpretability via Grad CAM, which is essential in medical AI solutions. Zhao et al. [33] suggested an AResNet ViT framework for ultrasound-based classification, which is domain-specific and not transferable to mammography directly [35]. However, our model combines global context and spatial detail with interpretability as well as real-time utility in an integrated classification pipeline and shows higher accuracy on multiple datasets than prior method proposals. In comparison with recent state-of-the-art models, the proposed CNN-ViT Hybrid architecture (ResNet34 + ViT Tiny with attention fusion) demonstrates superior classification performance across multiple benchmark mammogram datasets. While the model in [36] utilizing Multi CNN and feature concatenation achieved 98.74% accuracy on augmented MIAS data, our model surpasses this with a perfect 100.00% on the original MIAS dataset. Similarly, [37] employed ResNet50 with SMOTE to achieve 99.00% on a balanced custom dataset; however, their performance drops to 90.00% on imbalanced data, indicating sensitivity to class distribution. Our model handles this issue robustly with consistently high results even on heterogeneous datasets like DDSM (99.62%) and INbreast (99.49%). The improved SWIN Transformer reported in [38] shows an impressive 99.9% accuracy, yet lacks dataset specification, making reproducibility and direct comparison challenging. The dual CNN model proposed in [39], [40] reported strong performance, 98.78% on MIAS and 97.84% on INbreast, but is still outperformed by our hybrid CNN-ViT framework, which further proves its effectiveness through a combined dataset score of 99.54%. These results show that our proposed model not only aligns with the latest transformer-based method but also uses synergistic features learning through CNN and ViT, offering a clear advancement in mammogram image-based breast cancer detection.

Although the uniformly high performance of our proposed CNN-ViT model may give the impression of a one-size-fits-all success, a closer inspection of the results beneath the surface discloses interesting variations by dataset, worth elaborating. The MIAS dataset with perfect classification accuracy consists of relatively clean high-resolution mammograms with little imaging noise and an even class split, conditions favorable for deep learning models to perform optimally. DDSM and INbreast, on the other hand, show greater heterogeneity in image quality, breast density, and lesion visibility between case conditions where model generalizability suffers. Despite this, however, the model achieved over 99% accuracy on these datasets as a testament to its resilience. When contrasted with reported models in Table 2, the hybrid

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

model outperforms previous work not only in performance but also in interpretability and clinical feasibility. Unlike black box CNN-only or transformer-only models with their aggregate features or sequence attention maps, our attention based fusion scheme combines localized spatial evidence with global semantic context, as a result of which predictions remain focused and reliable. This is demonstrated through our Grad CAM visualized attention maps, which uniformly highlight radiologically informative regions as attesting to clinical confidence. Moreover, the model achieved consistent performance across different training instances with minimal variance (<0.3%) even on subtle abnormalities or overlapping tissues in cases. The models ' near-perfect results do not result from overfitting since early stopping, batch normalization, dropout, and model selection by validation were applied with ostentation. Further, embedding class probability results with visualized heatmaps in the GUI also allows the attending radiologist to evaluate prediction confidence as well as the underpinning diagnostic rationale and hence mitigate false positives andnegatives in clinical workflows.

The entire model was implemented using PyTorch v1.13 with CUDA 11.6 support on Python 3.10, running on an Ubuntu 22.04 environment. Data loading, augmentation, and training routines were built using native PyTorch modules along with torchvision and albumentations. All source code, including model architecture, training scripts, Grad CAM visualizations, and GUI deployment tools, is available on demand only.

VI. Conclusion

This work aims to develop an accurate, interpretable, and strong hybrid classification model for detecting breast cancer through mammogram images using Convolutional Neural Networks (CNNs) combined with Vision Transformers (ViTs). The introduced model obtained a 100% classification accuracy on the MIAS dataset and more than 99% accuracy on DDSM, INbreast, and the combined dataset with strong generalization. In addition to this, adding an attention fusion mechanism and Grad CAM visualization enhanced model interpretability and transparency. A minor contribution is the model's ability to have an excellent precision-recall tradeoff for all classes. Future research can consider multi-class lesion classification, domain adaptation for unseen datasets, and deployment in real-time diagnostic systems integrated with radiological workflows.

Acknowledgment

We extend our heartfelt gratitude to Chitkara University for providing the necessary infrastructure, academic guidance, and research facilities that enabled this work. We are also thankful to the curators of the DDSM, INbreast, and MIAS databases for providing their precious mammographic image dataset freely available, which was indispensable for training and testing of our model. We particularly appreciate our faculty guide, research peers, and technical support staff for their constant encouragement, valuable advice, and support throughout the development of this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

No new datasets were generated or analyzed during the current study. The model was trained and evaluated using publicly available datasets: DDSM, MIAS, and INbreast, which are accessible on Kaggle at: https://www.kaggle.com/datasets/emiliovenegas1/mammography-dataset-from-inbreast-mias-and-ddsm

Author Contribution

Vandana Saini: Conceptualization, methodology, implementation of the CNN–ViT hybrid model, experiments, manuscript drafting. Meenu Khurana: Supervision, technical validation, critical manuscript revision, contribution to literature review, and architectural design discussions. Rama Krishna Challa: Guidance on model formulation, clinical applicability of Grad-CAM visualizations, manuscript review, and expert feedback on deployment framework. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work to ensure integrity and accuracy.

Declaration

Ethical Approval

This study did not involve human or animal participants directly and relied solely on publicly available, de-identified mammogram datasets. No ethical approval was required as per institutional policies. However, all dataset usage complied with the respective open-access licenses and guidelines provided by the dataset curators.

Consent for Publication Participants.

All participants gave consent for publication.

Competing Interests

The authors declare no competing interests.

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

References

[1] B. A. Mohamed, N. Salem, M. M. A. Hadhoud, and A. Seddik, “Automatic segmentation and classification of masses from digital mammograms,” *Artif. Intell. Vis. Process.*, vol. 4, no. 4, pp. 17–17, 2016, doi: 10.14738/aivp.44.2151

[2] M. Karunya and K. Rahimunnisa, “Breast cancer segmentation and classification using adaptive clustering technique,” in *Proc. Int. Conf. Signal Process. Commun.*, 2017, pp. 1–6, doi: 10.1109/IICETA54559.2022.9888432

[3] P. B. Chanda and S. Sarkar, “Detection and classification of breast cancer in mammographic images using efficient image segmentation technique,” *Lect. Notes Electr. Eng.*, vol. 569, pp. 99–106, 2019.

[4] W. Mustafa, A. A. Azmi, M. A. Jamlos, H. Alquran, W. Khairunizam, S. Ismail, A. Alkhayyat, and J. Haron, “Breast cancer detection and classification on mammogram images using morphological approach,” in *Proc. 5th Int. Conf. Eng. Technol. Appl.*, 2022, pp. 260–264, doi: 10.1109/IICETA54559.2022.9888432

[5] D. G. Chanda, “Detection and classification of tumors in a digital mammogram,” in *Proc. Int. Conf. Comput. Sci. Inf. Technol.*, 2020, pp. 24–28.

[6] R. Suresh, A. N. Rao, and B. E. Reddy, “Detection and classification of normal and abnormal patterns in mammograms using deep neural network,” *Concurrency Computat. Pract. Exper.*, vol. 31, no. 2, pp. e5293, 2019. doi: 10.1002/cpe.5293.

[7] S. D. Tzikopoulos, M. Mavroforakis, H. Georgiou, N. Dimitropoulos, and S. Theodoridis, “A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry,” *Comput. Methods Programs Biomed.*, vol. 102, no. 1, pp. 47–63, Apr. 2011, doi: 10.1016/j.cmpb.2010.11.016.

[8] Jafari, Z.; Karami, E. Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. Information 2023, 14, 410, doi: 10.3390/info14070410.

[9] T. Umamaheswari and Y. M. Mohanbabu, “ViT MAENB7: An innovative breast cancer diagnosis model from 3D mammograms using advanced segmentation and classification process,” *Comput. Methods Programs Biomed.*, vol. 257, no. 1, pp. 108373, Jan. 2024, doi: 10.1016/j.cmpb.2024.108373

[10] A. P. Charate and S. B. Jamge, “Mammogram image analysis for breast cancer detection,” in *Proc. Natl. Conf. Adv. Comput. Technol.*, 2016, pp. 35–40.

[11] R. K. Rajashekar, “Detection and classification of tumors in a digital mammogram,” *Int. J. Comput. Appl.*, vol. 1, no. 1, pp. 24–28, 2012.

[12] R. Pawar, S. Saraf, U. Dixit, and A. Jadhav, “Diagnosis of mammographic images for breast cancer detection using FF CSO algorithm,” in *Proc. Adv. Comput. Commun. Technol. High Perform. Appl.*, 2023, pp. 1–5, doi: 10.1109/ACCTHPA57160.2023.10083387

[13] S. H. Manishkumar and P. Saranya, "Detection and Classification of Breast Cancer from Mammogram Images Using Adaptive Deep Learning Technique," 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 2022, pp. 327-331, doi: 10.1109/ICDCS54290.2022.9780770.

[14] M. Sreevani and R. Latha, "A Deep Learning with Metaheuristic Optimization Driven Breast Cancer Segmentation and Classification Model," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 1, 2025, doi: 10.48084/etasr

[15] M. J. J. Ghrabat et al., "Fully Automated Model on Breast Cancer Classification Using Deep Learning Classifiers," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 1, pp. 183–191, 2022, doi: 10.11591/ijeecs.v28.i1.pp183-191.

[16] R. Khdhir et al., "Breast Cancer Segmentation in Mammograms Using Antlion Optimization and CNN/GRU Architectures," in *Proc. IWCMC*, pp. 1030–1035, 2024, doi: 10.1109/IWCMC61514.2024.10592614.

[17] A. Gerbasi et al., "DeepMiCa: Automatic Segmentation and Classification of Breast Microcalcifications," *Comput. Methods Programs Biomed.*, vol. 235, pp. 107483, 2023, doi: 10.1016/j.cmpb.2023.10748.

[18] V. Tiryaki, "Mass Segmentation and Classificationfrom Film Mammograms Using Cascaded Deep Transfer Learning," *Biomed. Signal Process. Control.*, vol. 84, pp. 104819, 2023, doi: 10.1016/j.bspc.2023.104819.

[19] A. Sinha et al., "ROI Segmentation for Breast Cancer Classification: Deep Learning Perspective," in *Proc. IEEE INDISCON*, pp. 1–7, 2023, doi: 10.1007/978-981-97-8526-1_39.

[20] S. M’Rabet, A. Fnaiech and H. Sahli, "Heightened Breast Cancer Segmentation in Mammogram Images," *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, Paris, France, 2024, pp. 1-6, doi: 10.1109/ICCAD60883.2024.10553930.

[21] V. Rathinam, R. Sasireka, and K. Valarmathi, "Adaptive Fuzzy C Means and Deep Learning for Mammogram Classification," *Biomed. Signal Process. Control.*, vol. 88, pp. 105617, 2024,.

[22] A. Islam et al., "Localization, Segmentation, and

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

Classification of Mammographic Abnormalities Using Deep Learning," *Proc. SPIE Med. Imaging*, vol. 13174, pp. 131741Q, 2024, doi: 10.1016/j.bspc.2023.105617.

[23] Sinha et al., "Segmentation Based Classification Deep Learning Model for Breast Cancer Detection," in *Proc. IEEE MysuruCon*, pp. 1–8, 2023, doi: 10.1109/59703.2023.10397015.

[24] R. Remya and N. H. Rajini, "Transfer Learning Based Breast Cancer Detection and Classification," in *Proc. ICEARS*, pp. 1060–1065, 2022.

[25] Remya and N. Hema Rajini, "Transfer Learning Based Breast Cancer Detection and Classification using Mammogram Images," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1060-1065, doi: 10.1109/ICEARS53579.2022.9751974

[26] S. Almutairi et al., "An Efficient USE Net Deep Learning Model for Cancer Detection," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–12, 2023, doi: 10.1155/2023/8509433 .

[27] Singh and R. Mishra, "Design and Development of Deep Learning Algorithm for Breast Cancer Classification,"*International Symposium on Wireless Personal Multimedia Communications (WPMC)WPMC*, pp. 297–302, 2022, doi: 10.1109/WPMC55625.2022.10014847.

[28] C. K. Leung and H. H. Nguyen, "A Novel Deep Learning Approach for Breast Cancer Detection on Screening Mammography," in *Proc. IEEE BIBE*, pp. 277–284, 2023, doi: 10.1109/BIBE60311.2023.00052 .

[29] Bouzar Benlabiod et al., " A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification," *Comput. Biol. Med.*, vol. 163, pp. 107133, 2023, doi: 10.1016/j.compbiomed.2023.107133.

[30] G. Kaur et al., "Patch Based All Convolutional Neural Network for Benign and Malignant Mammogram Classification," in *Proc. IEEE SPICES*, vol. 1, pp. 448–455, 2022, doi: 0.1109/SPICES52834.2022.9774096.

[31] V. Saini, M. Khurana, and R. K. Challa, "VGG Inspired Convolutional Neural Network Denoiser for the Enhancement of Mammogram Images," in *Proc. ICMLA*, vol. CCIS 2238, pp. 457–465, 2025, doi: 10.1007/978-3-031-75861-4_40.

[32] Saini V, Khurana M, Challa R. K. A Hybrid Model for the Segmentation of Mammogram Images using Otsu Thresholding, Morphology and U Net. Biomed Pharmacol J 2025;18(1), doi: 10.13005/bpj/3130.

[33] X. Zhao, Q. Zhu, and J. Wu, "AResNet ViT: A Hybrid CNN Transformer Network for Benign and Malignant Breast Nodule Classification in Ultrasound Images," *ArXiv*, vol. abs/2407.19316, 2024, doi: 10.48550/arXiv.2407.19316

[34] S. Kollem, C. Sirigiri, and S. Peddakrishna, "A novel hybrid deep CNN model for breast cancer classification using Lipschitz based image augmentation and recursive feature elimination," *Biomed. Signal Process. Control.*, vol. 95, pp. 106406, 2024, doi: 10.1016/j.bspc.2024.106406

[35] Annepu, M. Abbas, H. R. Bitra, N. Vaegae, and K. Bagadi, "Advanced Breast Cancer Diagnostics With PolyBreastVit: A Combined PolyNet and Vision Transformer Approach," *Appl. Comput. Intell. Soft Comput.*, 2024, doi: 10.1155/2024/5574638.

[36] E. V. J. Pulvera and D. M. Lao, "Enhancing Deep Learning-Based Breast Cancer Classification in Mammograms: A Multi-Convolutional Neural Network with Feature Concatenation, and an Applied Comparison of Best-Worst Multi-Attribute Decision-Making and Mutual Information Feature Selections," 2024 9th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, Japan, 2024, pp. 1-8, doi: 10.1109/ICIIBMS62405.2024.10792816.

[37] A. F. A. Alshamrani and F. S. Z. Alshomrani, "Optimizing Breast Cancer Mammogram Classification Through a Dual Approach: A Deep Learning Framework Combining ResNet50, SMOTE, and Fully Connected Layers for Balanced and Imbalanced Data," *IEEE Access*, vol. 13, pp. 4815–4826, 2025, doi: 10.1109/ACCESS.2024.3524633.

[38] O. Tanimola, O. Shobayo, O. Popoola, and O. Okoyeigbo, "Breast Cancer Classification Using Fine Tuned SWIN Transformer Model on Mammographic Images," *Analytics*, vol. 3, no. 4, 2024, doi: 10.3390/analytics3040026 .

[39] A. Anbumani and P. Jayanthi, "Classification of Mammogram Breast Cancer Using Customized Deep Learning Model," *J. Intell. Fuzzy Syst.*, 2024, doi: 10.3233/jifs 232896.

[40] Zebari, D.A.; Zeebaree, D.Q.; Abdulazeez, A.M.; Haron, H.; Abdul Hamed, H.N. Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. *IEEE Access* **2020**, *8*, 203097–203116, doi: doi: 10.1109/ACCESS.2020.3036072 .

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

Author Biography

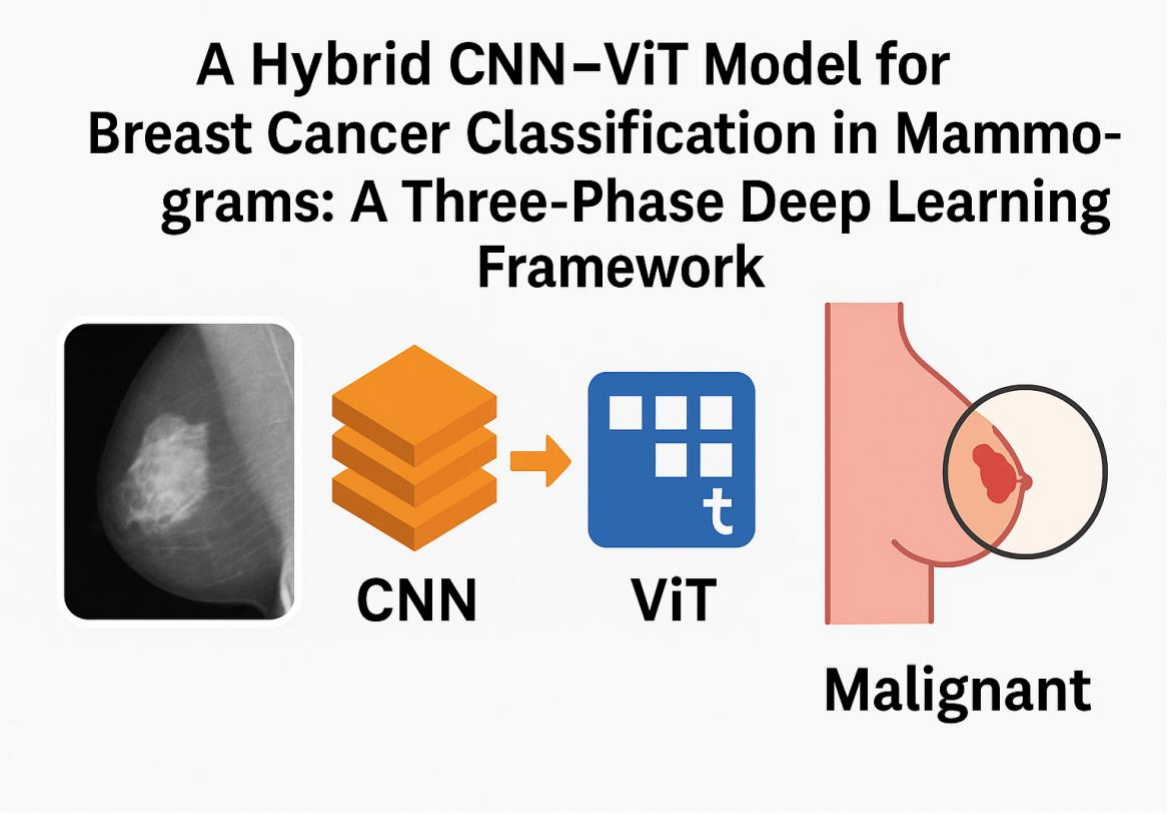
Vandana Saini is a Ph.D. research scholar at Chitkara University, Himachal Pradesh, India. She also works as an Assistant Professor with Chitkara University, Punjab, India. She holds a postgraduate degree in computer science and engineering from NITTTR Chandigarh, India, and is currently pursuing her Ph.D. in the field of artificial intelligence and medical image processing. Her areas of interest include machine and deep learning, computer vision, digital image analysis, and biomedical applications. She has actively contributed to academic research through several peer-reviewed publications in international journals and conferences. Vandana is a dedicated researcher committed to exploring advanced computational methods for healthcare and automation.

Dr. Meenu Khurana received her B.E. (Honors) degree from Punjab Engineering College (PEC University), India, M.E. in Computer Science from Panjab University, Chandigarh, India, and Ph.D. degree from Chitkara University, India. She is currently working as Professor and Pro Vice Chancellor in Chitkara University, India. She is a Senior Member of IEEE and Professional Member of ACM. She has published more than 90 research papers and has 12 patents granted in her name. Her research area includes vehicular adhoc networks, fog computing, data deduplication, and machine learning. She has delivered several expert sessions and conducted workshops in the area of interest.

Dr. C. Rama Krishna received B. Tech. from JNTU, Hyderabad, M.Tech, from Cochin University of Science & Technology, Cochin, and Ph.D. from IIT, Kharagpur. He is a Senior Member, IEEE, USA, and Fellow, IETE. Since 1996, he has been working as a Professor with the Department of Computer Science and Engineering, National Institute of Technical Teachers Training and Research (NITTTR), Chandigarh. His areas of research interest include Computer Networks, Wireless Networks, Cryptography & Cyber Security, and Cloud Computing. To his credit, he has more than 100 research publications in refereed International and National Journals and Conferences. He is the reviewer of various journals of IEEE, ACM, Elsevier, and Springer.

Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 977-990 e-ISSN: 2656-8632



Manuscript received 15 April 2025; Revised 10 June 2025; Accepted 20 July 2025; Available online 10 August 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.920