OPEN ACCESS
Vietnam Journal of Computer Science
(2026) 1–30

DOI: 10.1142/S2196888826500041

World Scientific
www.worldscientific.com

Tri-ResNet: A Parallel Multi-Branch and Transfer Learning Triple-Input Model for Breast Cancer Detection and Classification

Abdelhafidh Kacher* and Medjeded Merati†

Laboratoire Informatique et Mathématiques (LIM)
Faculty of Mathematics and Computer Science
University of Tiaret, BP 78, Zaâroura, Tiaret 14000, Algeria
**abdelhafidh.kacher@univ-tiaret.dz*
†*medjeded.merati@univ-tiaret.dz*

Saïd Mahmoudi

ILIA Department, Faculty of Engineering, UMONS
20 Place du Parc, Mons 7000, Belgium
said.mahmoudi@umons.ac.be

Accepted 20 February 2026
Published 17 March 2026

Breast cancer remains a leading cause of cancer-related deaths among women, highlighting the need for accurate computer-aided diagnosis systems (CADs). Convolutional neural networks (CNNs) have demonstrated substantial progress in medical image analysis, significantly improving diagnostic accuracy. This paper introduces Tri-ResNet, a triple-input model composed of parallel fine-tuned ResNet-based branches using transfer learning (TL) for efficient breast cancer classification. The model simultaneously processes full mammogram images (FMs), regions of interest images (ROIs), and contrast-enhanced ROI images (CLAHE-enhanced ROIs) using Contrast-Limited Adaptive Histogram Equalization (CLAHE). Extensive experiments were conducted using multiple pre-trained models across single-input and multi-input architectures. Tri-ResNet achieved outstanding results on the Mini-DDSM, MIAS, and INbreast datasets, with peak performance on MIAS reaching 99.62% accuracy for normal–abnormal classification and 99.14% for benign–malignant classification, while maintaining competitive results on Mini-DDSM and INbreast. The model consistently outperformed single-input models and state-of-the-art approaches, demonstrating the effectiveness of multi-input CNNs for enhancing automated breast cancer diagnosis.

Keywords: Breast cancer; CNN; parallel network; multi-input; computer-aided diagnosis; transfer learning.

*Corresponding author.

1. Introduction

Breast cancer is the most prevalent cancer in the world and one of the primary cause of cancer-related deaths among women, particularly in developing countries.[1] In Algeria, breast cancer is the leading cancer among women, as reported by the National Cancer Registry, with cases increasing annually and an age-standardized incidence rate of 82.8.[2,3] Early detection significantly improves survival rates, and relies on accurate mammographic interpretation followed by biopsy for malignancy confirmation. Radiologists manually inspect mammograms to detect and identify abnormalities such as masses, calcifications, or architectural distortions, using the Breast Imaging Reporting and Data System (BI-RADS) reporting system for classification.[4]

Despite advances in medical imaging technology, manual interpretation remains challenging due to overlapping breast tissue and variations in size and shape of early-stage cancerous lesions. This process requires the extreme caution by highly qualified radiologists.[5] Misdiagnosis can lead to delayed treatment or fatal outcomes in the case of false negatives (FNs), where real cancerous lesions are missed, or to unnecessary interventions and patient anxiety in the case of false positives (FPs), where non-cancerous findings are incorrectly identified as malignant. In many developing countries, the shortage of experienced radiologists further amplifies these challenges.[6] Therefore, improving mammogram interpretation efficiency is crucial through automated diagnostic systems.

Recently, computer-aided diagnosis systems (CAD) systems have increasingly improved diagnostic precision while reducing the time required for diagnosis compared to manual examination.[7] Convolutional neural networks (CNNs), in particular, have shown strong performance in medical applications, particularly, in the analysis of mammograms, improving breast cancer detection, localization, segmentation and the classification of breast tumors.[4,8]

However, CNNs require extensive labeled images for training, which are difficult to obtain in medical imaging due to cost, ethical constraints, and limited expert availability.[9] Transfer learning (TL) addresses this limitation by reusing pre-trained models trained on large datasets such as ImageNet,[10] for domain adaptation such as medical image classification, by transferring the learned weights instead of training the model from scratch.[11,12]

In this work, we evaluate several widely used pre-trained CNN architectures, including ResNet50V2,[13] ResNet152V2,[14] InceptionResNetV2,[15] DenseNet121,[16] NASNetMobile,[17] InceptionV3,[18] VGG16,[19] and VGG19.[19] These models were evaluated to determine their suitability for breast cancer classification when applied to different image types derived from the same case.

The experiments were structured into four main components:

- Full mammogram image (FM) classification: We evaluated the performance of the selected pre-trained CNN models on FMs across binary and multi-class classification (normal–abnormal, benign–malignant, and normal–benign–malignant).

- Regions of interest image (ROI) classification: The same models were tested on ROIs extracted from the original mammograms. Focusing on these localized regions helps the models concentrate specifically on suspicious tumor areas, enabling more targeted classification.
- Contrast-limited adaptive histogram equalization (CLAHE)-enhanced ROI classification: We investigated the impact of applying CLAHE[20] to the ROIs. CLAHE enhances local contrast and improves the visibility of subtle edges and textures that may be indistinguishable in the original images. This enhancement helps the CNN better differentiate between normal and abnormal tissues.
- Triple-input CNN model: We constructed a triple-input architecture consisting of three parallel branches, each using one of the top-performing pre-trained models identified in the previous experiments. This model simultaneously processes FMs and corresponding ROIs and CLAHE-enhanced ROIs. Combining these complementary inputs from the same case ensures that no relevant features are missed, as each image contributes unique diagnostic information. This multi-input fusion significantly improved classification accuracy.

All experiments were conducted using preprocessed, balanced, and augmented images from the Mini-DDSM dataset.[21] Through these experiments, we developed Tri-ResNet, a parallel multi-branch architecture integrating pre-trained CNN models. The model's novelty lies in processing three distinct image inputs, each handled by a dedicated branch. This design enables the extraction of complementary macroscopic and microscopic tumor characteristics. By combining ResNet50V2 and ResNet152V2 architectures, Tri-ResNet achieved outstanding performance on the Mini-DDSM dataset, surpassing single-input CNN models and several state-of-the-art approaches. The model also demonstrated strong generalization on the MIAS[22] and INbreast[23] datasets. In addition, the models are evaluated using accuracy, precision, recall, $F1$-score, specificity, area under the ROC curve (AUC), and error rate metrics.

The remainder of this paper is organized in the following way. Section 2 reviews related work in the field of breast cancer detection and classification. Section 3 describes the datasets and methodology. Section 4 reports experimental results, an ablation study, and comparisons with state-of-the-art models. Section 5 discusses the findings, limitations and future work. Section 6 concludes this paper.

2. Related Work

This section provides a comprehensive review of recent research studies on CNN-based breast cancer classification using TL on public mammography datasets. Numerous studies have explored fine-tuned CNNs, multi-channel architectures, and multi-input strategies to enhance diagnostic accuracy. In addition, the section highlights various preprocessing and data augmentation techniques employed in these studies.

Multi-modal studies, such as Muduli *et al.*[24] combined mammography images obtained from the MIAS, DDSM,[25] and INbreast datasets with ultrasound images. Augmentation involved geometric and photometric transformations, such as Gaussian noise injection and gamma correction. Their multi-modal CNN achieved 96.55% and 90.68% accuracies on MIAS and DDSM images, respectively. Meanwhile, Mohapatra *et al.*[26] fine-tuned AlexNet,[27] VGG16, and ResNet50[13] for multi-class classification using FMs. The fine-tuned AlexNet and VGG16 models only achieved a maximum accuracy of 65% on Mini-DDSM dataset.

Recent works have explored hybrid models and multi-channel fusion of features from multiple pre-trained models. Several studies also introduced multi-input architectures where enhanced images serve as independent inputs, demonstrating performance gains.

Salama *et al.*[28] fine-tuned ResNet50 and VGG16 by replacing the final fully connected (FC) layer with a Support Vector Machine (SVM) classifier[29] for benign–malignant classification. Using MIAS, DDSM, and CBIS-DDSM[30] datasets, the fine-tuned ResNet50 combined with SVM achieved top performance on DDSM with an accuracy of 97.98%. Aslan[31] combined CNNs with Bidirectional Long Short-Term Memories (BiLSTM) layers for multi-class classification. Preprocessing involved the removal of defects and noise, while augmentation incorporated Gaussian noise, sharpening filters, salt-and-pepper noise, and Gaussian filters. The model achieved accuracies of 98.56% and 92.26% on MIAS and INbreast datasets, respectively.

Khamparia *et al.*[32] proposed a modified VGG16 for normal–abnormal classification on ROIs from the DDSM dataset. Their model achieved an accuracy of 94.3%. Al-Nawashi *et al.*[33] performed a multi-class classification of extracted features from mammogram images using a custom CNN model. Using CLAHE-enhanced images from a private dataset, the proposed model achieved an accuracy of 98.71%, while minimizing computational time. Eltoukhy *et al.*[34] proposed a CAD system for classifying breast masses and normal tissues on ROIs. Using Exact Gaussian–Hermite Moments (EGHMs), the extracted feature vector is fed into the AdaBoost classifier for prediction. Using images from the MIAS and Image Retrieval in Medical Applications (IRMA) datasets, the classifier achieved accuracies of 90.56% and 93.27% on MIAS and IRMA, respectively.

Alternatively, Jafari and Karami[35] investigated the fusion of five pre-trained models for breast cancer detection. By combining extracted features from FMs using AlexNet, ResNet50, MobileNet-Small,[36] ConvNeXtSmall,[37] and EfficientNet,[38] their multi-channel model achieved accuracies of 94.5% and 96% on the MIAS and DDSM datasets, respectively. Similarly, Maqsood *et al.*[39] introduced the Transferable Texture CNN (TTCNN), which preserves texture information for benign–malignant classification. They applied a modified CLAHE technique for preprocessing images from the DDSM, INbreast, and MIAS datasets. By integrating extracted features from InceptionResNetV2, InceptionV3, VGG16, VGG19, GoogLeNet,[40] ResNet18[13] ResNet50, and ResNet101,[13] and using entropy-based feature selection, the model achieved 99.08% accuracy on DDSM images.

Alshayeji and Al-Buloushi[41] developed a triple-channel model for binary and multi-class classifications by integrating a custom CNN, InceptionResNetV2, and Xception.[42] Using ROIs from the DDSM dataset, the triple-channel model achieved 97.26% accuracy for binary classification and 99.13% for multi-class classification. Mohammed and Ekmekci[43] developed a CAD system for multi-class classification by integrating the You Only Look Once (YOLO) object detector[44] with a multi-scale CNN architecture, combining DenseNet and InceptionNet.[18] Applying CLAHE and noise injection for preprocessing, their system achieved an accuracy of 98.08% on the INbreast dataset.

Investigating multi-input models, Jabeen *et al.*[45] used FMs and contrast-enhanced FMs as inputs of a dual-branch EfficientNet-B0 model for benign–malignant classification. The dual-input model achieved accuracies of 95.4% on the CBIS-DDSM and 99.7% on the INbreast datasets.

Finally, Sahu *et al.*[46] proposed a multi-modal CNN framework integrating ShuffleNet[47] and ResNet18 for breast cancer classification. Evaluated on images from different dataset modalities, including Mini-DDSM and ultrasound datasets, their framework achieved 99.17% accuracy for normal–abnormal classification and 98% for benign–malignant classification on the Mini-DDSM dataset.

While CNN models have achieved remarkable success, the emergence of Vision Transformers (ViTs) has introduced a new paradigm. Numerous studies have explored transformer-based architectures to enhance classification performance. In parallel, Sarker *et al.*[48] proposed a multi-view transformer built on Swin Transformer blocks. The model fuses multiple views of mammogram images from the CBIS-DDSM and VinDr-Mammo datasets. The proposed model achieved an accuracy of 95.50% on the VinDr-Mammo images and 68.63% for mass lesion classification on CBIS-DDSM images.

Proposing a complete CAD system, Hassan *et al.*[49] combined YOLOv4 and ViTs for breast mass detection, localization, and classification. They used contrast-enhanced spectral mammography images from the INbreast dataset. The ROIs are first detected and extracted using YOLOv4 and subsequently classified as benign or malignant by a ViT-based classifier, achieving a classification accuracy of 95.65%.

Most previous studies have focused on developing fine-tuned architectures for breast cancer classification across multiple categories. However, these studies present two main limitations: (1) the majority rely on a single input image per model, even when proposing multi-channel or multi-scale architectures; (2) while other studies have introduced multi-input models, most approaches analyze either FMs or ROIs, meaning multiple inputs are derived from the same original image rather than incorporating diverse input types.

Moreover, several studies employed geometric transformations for augmentation, including image scaling, translation, shifting, cropping, and zooming for augmentation. However, these methods may alter the critical characteristics of breast cancer tumors, especially in ROIs.

To address these limitations, we introduce a novel multi-input model that integrates multiple pre-trained CNNs for breast cancer detection and classification. This approach extracts complementary features from different image types including FMs, ROIs, and CLAHE-enhanced ROIs to enrich feature diversity and improve classification performance. Additionally, this study explores photometric transformations for augmentation, including noise addition, to improve the model's robustness.

3. Materials and Methods

This section describes the dataset and methodology used in this study, including the preprocessing pipeline, data balancing and augmentation procedures, and the selection of pre-trained base models through a series of experiments conducted to construct the proposed architecture. Figure 1 illustrates the overall workflow of the methodology.

The experiments were conducted using images from the public Mini-DDSM dataset, a lightweight version of the widely used DDSM dataset. This dataset has been adopted in numerous recent studies (Refs. 26, 46, and 50), demonstrating its relevance and utility in breast cancer research. Mini-DDSM contains mammogram images in 16-bit PNG and 8-bit JPEG formats, labeled as normal, benign, or malignant. In addition to FMs, it provides rich metadata, including patient age,

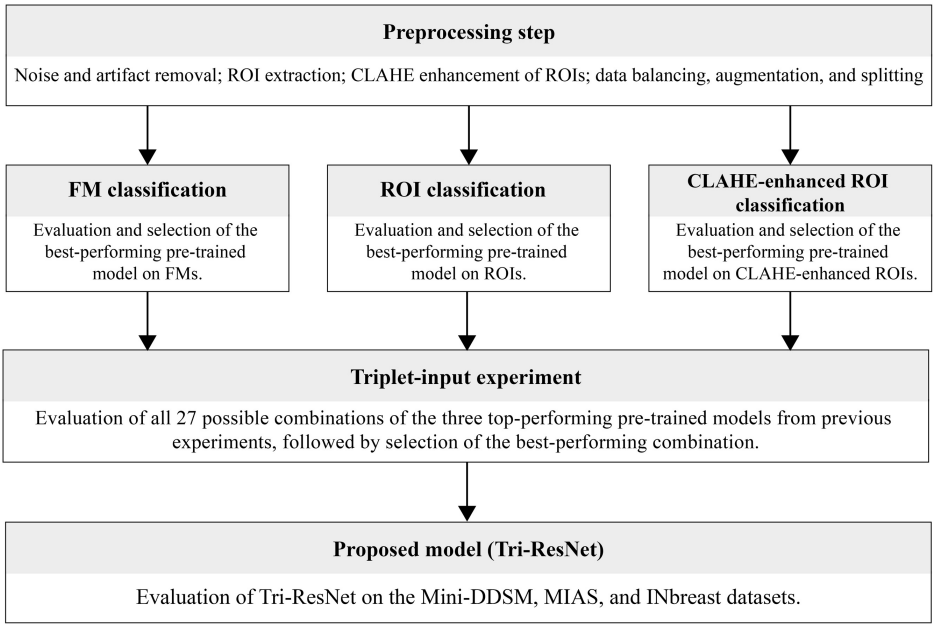


Fig. 1. Overall architecture of the proposed methodology.

BI-RADS category, disease type, and radiologist-annotated tumor binary masks indicating lesion location and shape.

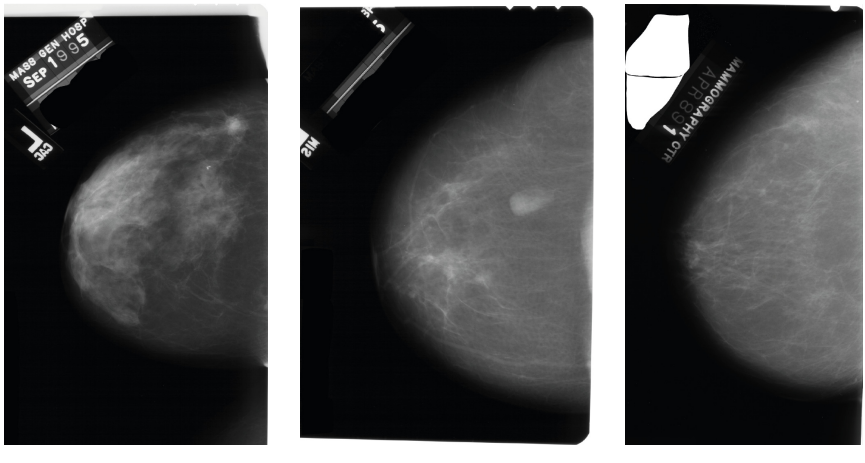
Since this study relies on complementary image types as model inputs, we filtered the dataset to retain only cases containing both FMs and corresponding tumor binary masks. These masks were essential for extracting ROIs used in our multi-input architecture.

3.1. *Image preprocessing*

Image preprocessing is an essential step to enhance CNN performance, as it helps normalize images by reducing variability, removing irrelevant features, and enhancing relevant patterns. This enables the model to learn more effectively and produce accurate predictions. The preprocessing pipeline included ROI extraction using tumor masks, and contrast enhancement using CLAHE. Many Mini-DDSM images contain noise, labels, or artifacts (Fig. 2), requiring multiple cleaning steps:

(a) Initially, morphological operations (erosion and dilation),[51] combined with binary thresholding,[52] were applied to generate a binary mask of the original FM. Contour detection was then used to identify and isolate the largest connected component (the breast). Smaller objects such as labels and background noise were removed, and the image was cropped to retain only the breast region, eliminating blank space (Fig. 3(a)).

(b) The second step focuses on extracting ROIs. Tumor masks were used to determine bounding boxes around the lesions. ROIs were extracted by cropping these bounding boxes from the original FM (Fig. 3(b)). For normal cases, ROIs were randomly extracted since no suspicious lesion is present in the breast tissue.



(a) (b) (c)

Fig. 2. Sample images from the Mini-DDSM dataset: (a) malignant case; (b) benign case; (c) normal case.

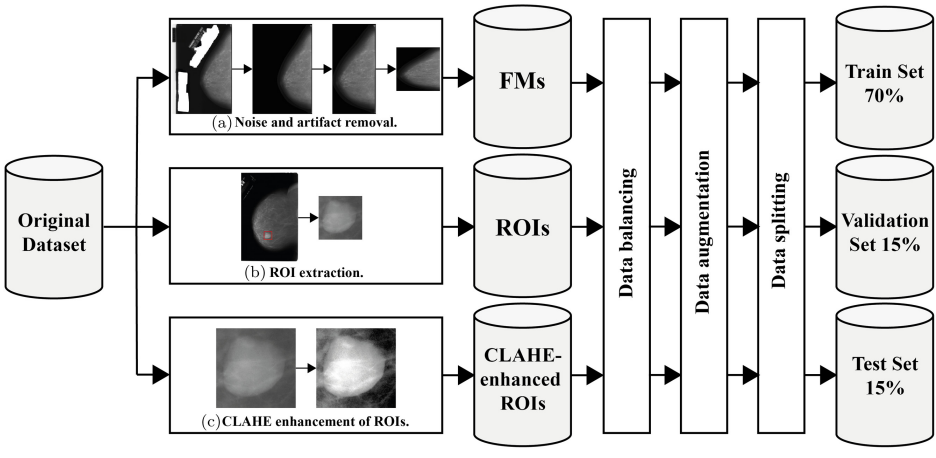


Fig. 3. Overview of the preprocessing steps.

(c) In the third step, ROIs were enhanced using CLAHE (Fig. 3(c)), which improves local contrast and highlights low-contrast structural patterns without excessively amplifying noise or adjacent tissues. This technique has been widely adopted in breast cancer imaging studies (Refs. 39 and 43), achieving notable results.

(d) Finally, all processed images were resized to 224×224 to match the input requirements of pre-trained CNN models and subsequently normalized to improve training stability.

Following preprocessing, the final dataset included 13,854 images, divided across three distinct sets (FMs, ROIs, and CLAHE-enhanced ROIs) each containing 4,618 images (1,984 normal, 1,317 benign, 1,317 malignant) as summarized in Table 1.

3.2. *Data balancing, augmentation, and splitting*

In this study, we perform three types of classifications: normal–abnormal, benign–malignant, and normal–benign–malignant classifications. Since Mini-DDSM is imbalanced, typically containing more normal cases, each class was balanced separately for each classification task to avoid biased model performance.

Balancing was performed as follows:

- For normal–abnormal classification, the benign and malignant classes were downsampled to 992 cases each, producing a balanced abnormal set of 1,984 cases matching the number of normal cases.
- For benign–malignant classification, both classes were already balanced at 1,317 cases each.
- For normal–benign–malignant classification, the normal class was downsampled to 1,317 cases to match the benign and malignant classes.

Each case is represented by the set of three related image types: FM and corresponding ROI, and CLAHE-enhanced ROI obtained from the same patient sample.

To improve generalization, image transformation techniques are frequently used to expand the dataset.[12] Although geometric transformations such as rotating, zooming, and cropping, are commonly applied for augmentation, they may distort or alter the appearance of breast cancer tumor.

Recent studies (Refs. 24, 31, 43, and 53) demonstrate that using photometric transformations, such as Gaussian blur and Gaussian noise is more appropriate for improving medical image classification. Based on these findings, Gaussian noise and Gaussian blur were used for augmentation. Table 1 provides detailed counts before and after augmentation.

Following the data balancing and augmentation steps, each case (FM with its corresponding ROI and CLAHE-enhanced ROI) was assigned a unique numeric identifier to maintain consistency across image type sets. To prevent data leakage, the dataset was split strictly at the case level rather than at the individual image level. Specifically, the training, validation, and testing split was first performed on the FMs using the split-folders library with a fixed random seed, generating a 70%, 15%, and 15% split for training, validation, and testing sets, respectively. This initial split produced distinct sets of image filenames corresponding to each case. Once the FM split was established, the ROIs and CLAHE-enhanced ROIs were assigned to the train, validation, and test sets by matching filenames with the FM split. This ensured that all images originating from the same case were consistently assigned to the same subset, eliminating any possibility of cross-set leakage. The process preserved the correspondence between the original FMs and the derived ROIs and CLAHE-enhanced ROIs, maintaining case-level integrity across all image types.

To confirm alignment across all image types, we implemented a verification procedure after each split, comparing filenames across all image type sets. No mismatches were observed, and all filenames in the ROI and CLAHE-enhanced ROI sets perfectly matched the FM filenames in the corresponding subsets. This rigorous verification guarantees that each case is represented in only one dataset split, ensuring the validity of subsequent model training and evaluation.

Table 1. Mini-DDSM images used in the experiments.

			Abnormal					
	Normal		Benign		Malignant		Total	
Set/class	Bef. aug.	Aft. aug.	Bef. aug.	Aft. aug.	Bef. aug.	Aft. aug.	Bef. aug.	Aft. aug.
FMs	1,984	7,936	1,317	5,268	1,317	5,268	4,618	18,472
ROIs	1,984	7,936	1,317	5,268	1,317	5,268	4,618	18,472
CLAHE-enhanced ROIs	1,984	7,936	1,317	5,268	1,317	5,268	4,618	18,472

By performing case-level splitting on mammograms and propagating this split to all derived image types, the dataset preparation prevents data leakage and ensures reliable evaluation of the model's performance.

3.3. *Training, validation, and testing processes*

Throughout the experiments, in each evaluated architecture, the features extracted from each base model are flattened and passed through a fine-tuning block (which varies depending on the experiment architecture), and finalized by a final prediction layer with a softmax activation function. During training, the model's predictions are compared to ground-truth labels using Categorical Crossentropy loss function, which measures the discrepancy between predicted and actual labels. The Adam optimizer iteratively updates the model's parameters to minimize the loss. After each training epoch, the model is evaluated on the validation data, and a ModelCheckpoint callback saves the model's weights whenever validation accuracy improves, ensuring that only the best-performing weights are retained for testing. Once the training process is complete, the saved weights with the highest validation accuracy are reloaded for testing. The model is then evaluated on the test dataset, and the results are presented in the experimental results section.

3.4. *Experimental design*

This study introduces a triple-input model that integrates the strengths of multiple pre-trained base models and complementary image inputs from the same case to achieve accurate breast cancer classification. The proposed architecture consists of three branches, each dedicated to processing a specific type of image input:

- Branch-A processes FMs, capturing global structural features.
- Branch-B processes ROIs, focusing on localized regions likely to contain abnormalities.
- Branch-C processes CLAHE-enhanced ROIs, highlighting subtle features that may be less visible in standard ROIs.

This multi-input design enables the model to capture a wide range of features from different input images. Each branch incorporates a pre-trained base model selected based on experimental evaluations.

To develop the proposed architecture, two main experiments were conducted:

3.4.1. *Single-input model experiment*

In the first experiment, we evaluated the performance of several pre-trained base models for single-image classification. The models included ResNet50V2, InceptionV3, InceptionResNetV2, VGG16, VGG19, DenseNet121, and NASNetMobile. These models were selected based on recent studies presented in the related work section, where they were frequently fine-tuned for breast cancer classification.

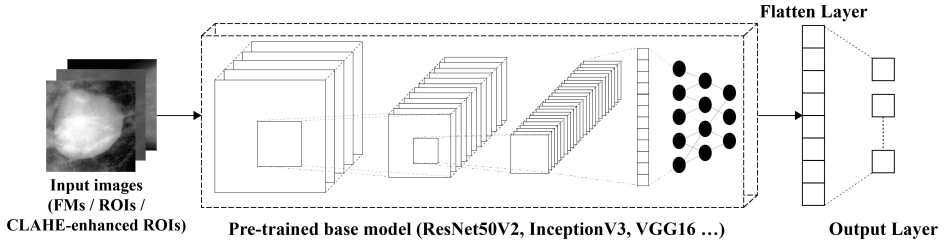


Fig. 4. Architecture of the single-input model.

Each pre-trained base model was initialized with ImageNet weights (excluding the top layers). Extracted features were flattened and passed through a softmax-activated prediction layer. Figure 4 shows the architecture used in the single-input experiments for both binary and multi-class classification.

To comprehensively evaluate the impact of different input types, the experiment was repeated using the same set of pre-trained models and fine-tuning strategy with: (1) FMs, to establish the baseline performance, (2) ROIs, to evaluate region-focused classification, and (3) CLAHE-enhanced ROIs, to investigate the effect of contrast enhancement on classification accuracy.

3.4.2. *Triple-input model experiment*

In the second experiment, we combined the strengths of the top-performing pre-trained base models from the previous experiments into a triple-input architecture. The model consists of three parallel branches, each corresponding to a pre-trained base model. The model captures diverse and complementary features from each input type, and the extracted features from the three branches were flattened, concatenated, and passed through a softmax-activated prediction layer for final classification (Fig. 5). This experiment aimed to assess the impact of combining multiple image types and multiple pre-trained base models on overall classification performance.

The pre-trained base models assigned to each branch were selected based on their performance in the corresponding single-input experiments:

- Branch-A: top-performing model with FMs,
- Branch-B: top-performing model with ROIs,
- Branch-C: top-performing model with CLAHE-enhanced ROIs.

3.5. *Proposed model construction*

The final proposed architecture, namely Tri-ResNet, builds upon the triple-input architecture with additional fine-tuning enhancements. The output of each pre-trained base model is first passed through a dedicated FC layer. The concatenated features from all branches are then refined using a fine-tuning block composed of: (1)

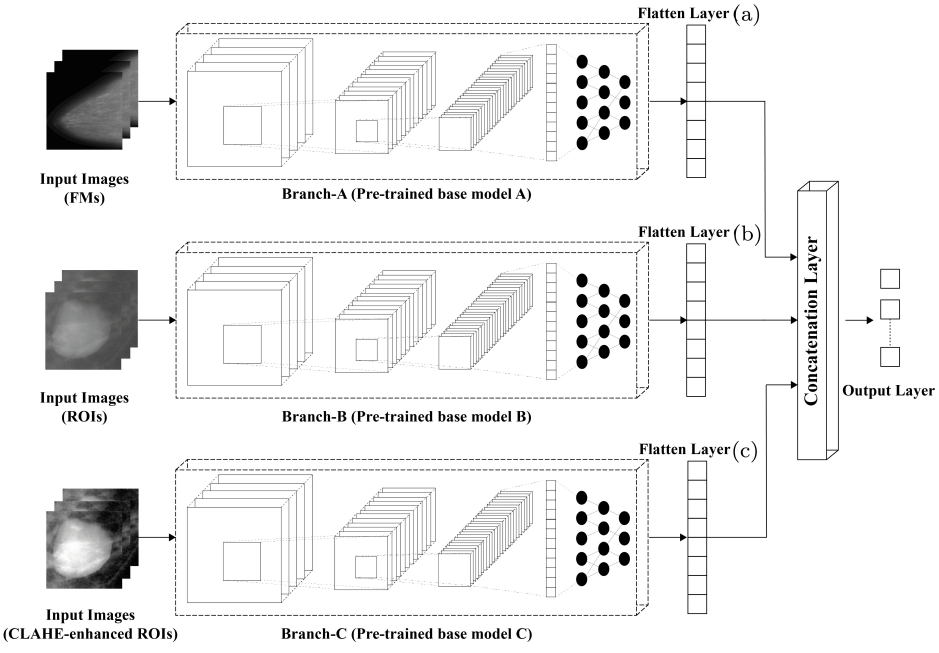


Fig. 5. Architecture of the triple-input model.

batch normalization for training stability, (2) FC layer with 512 neurons, (3) a dropout layer to reduce overfitting, and (4) another FC layer with 64 neurons.

Finally, the output of this block is fed into a softmax-activated prediction layer. The complete architecture of the final model is illustrated in Fig. 6.

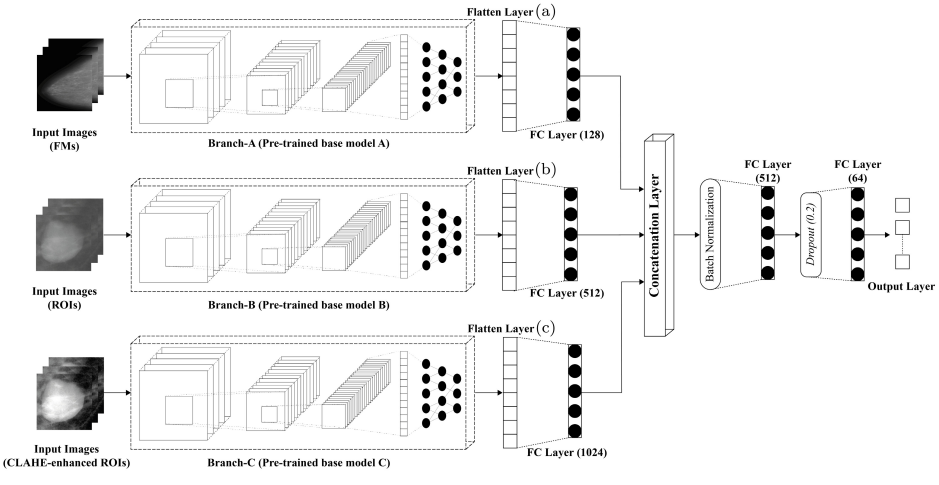


Fig. 6. Architecture of the proposed Tri-ResNet model.

4. Experimental Results

This section presents the experimental results using multiple evaluation metrics. We compare the performance of all conducted experiments on both binary and multi-class classifications. Additionally, we discuss the results of the Tri-ResNet model on the Mini-DDSM, MIAS, and INbreast datasets and compare its performance with several state-of-the-art models.

4.1. *Evaluation metrics*

The performance of all models is evaluated using commonly adopted classification metrics,[54] including accuracy[9] "(1)", precision[9,55] "(2)", recall (sensitivity)[9,55] "(3)", $F1$-score[9,55,56] "(4)", specificity[9,55] "(5)", AUC,[9,57] and error rate[54] "(6)".

These metrics are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{5}$$

$$\text{Error Rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{6}$$

where

- True Positive (TP): The number of instances correctly classified as positive class.
- FP: The number of instances incorrectly classified as positive class when they are actually negative class.
- True Negative (TN): The number of instances correctly classified as negative class.
- FN: The number of instances incorrectly classified as negative class when they are actually positive class.

4.2. *Single-input model results*

In this experiment, we first evaluated single-input fine-tuned models using FMs. ResNet50V2 achieved 93.16% across accuracy, recall, precision, $F1$-score, and specificity, with an AUC of 0.9778 and an error rate of 0.0684 for normal–abnormal classification. It achieved 92.10% across all five metrics, with an AUC 0.9705 and an error rate of 0.0790 for benign–malignant classification. For multi-class classification,

it achieved 91.91% accuracy, 92.20% precision, 91.66% recall, 91.93% $F1$-score, and 96.12% specificity, 0.9823 AUC and an error rate of 0.0537.

Using ROIs, ResNet50V2 achieved 98.49% across accuracy, recall, precision, $F1$-score, and specificity, with an AUC of 0.9991 and an error rate of 0.0151 for normal–abnormal classification. It achieved 92.92% across all five metrics, with an AUC of 0.9786 and error rate of 0.0708 for benign–malignant classification. For multi-class classification, it achieved 94.23% accuracy, 94.25% precision, 93.97% recall, 94.11% $F1$-score, 97.13% specificity, 0.9920 AUC and an error rate of 0.0392. These results demonstrate accuracy improvements of +5.33%, +0.82%, and +2.32% over FM-based models for the corresponding binary and multi-class classifications.

Applying CLAHE to ROIs further improved performance. ResNet50V2 achieved 98.74% across accuracy, recall, precision, $F1$-score, and specificity, with an AUC of 0.9986 and error rate of 0.0126 for normal–abnormal classification. It achieved 94.94% across the same five metrics, with an AUC of 0.9852 and error rate of 0.0506 for benign–malignant classification. For multi-class classification, it achieved 94.77% accuracy, 94.88% precision, 94.44% recall, 94.66% $F1$-score, 97.45% specificity, 0.9931 AUC, and an error rate of 0.0355. Compared to ROI-based models, accuracy increased by +0.25%, +2.02%, and +0.54% for the three classification tasks.

Tables 2–4 summarize the detailed results obtained in this experiment.

Table 2. Results of the single-input models using FMs.

Base model	Accuracy (%)	Recall (%)	Precision (%)	$F1$-score (%)	Specificity (%)	AUC	Error rate
Normal–Abnormal Classification							
ResNet50V2	**93.16**	**93.16**	**93.16**	**93.16**	**93.16**	**0.9778**	**0.0684**
DenseNet121	90.09	90.09	90.09	90.09	90.09	0.9659	0.0991
InceptionResNetV2	87.57	87.57	87.57	87.57	87.57	0.9489	0.1243
NASNetMobile	86.15	86.15	86.15	86.15	86.15	0.9346	0.1385
VGG16	85.64	85.64	85.64	85.64	85.64	0.9271	0.1436
InceptionV3	83.84	83.84	83.84	83.84	83.84	0.9182	0.1616
VGG19	82.12	82.12	82.12	82.12	82.12	0.8946	0.1788
Benign–Malignant Classification							
ResNet50V2	**92.10**	**92.10**	**92.10**	**92.10**	**92.10**	**0.9705**	**0.0790**
DenseNet121	87.67	87.67	87.67	87.67	87.67	0.9475	0.1233
InceptionResNetV2	86.09	86.09	86.09	86.09	86.09	0.9302	0.1391
NASNetMobile	79.96	79.96	79.96	79.96	79.96	0.8765	0.2004
InceptionV3	76.80	76.80	76.80	76.80	76.80	0.8466	0.2320
VGG16	76.36	76.36	76.36	76.36	76.36	0.8526	0.2364
VGG19	69.09	69.09	69.09	69.09	69.09	0.7574	0.3091
Normal–Benign–Malignant Classification							
ResNet50V2	**91.91**	**91.66**	**92.20**	**91.93**	**96.12**	**0.9823**	**0.0537**
DenseNet121	85.12	84.28	86.36	85.31	93.34	0.9640	0.0968
InceptionResNetV2	83.44	81.84	84.88	83.33	92.71	0.9533	0.1091
NASNetMobile	78.76	77.75	79.84	78.78	90.18	0.9309	0.1396
InceptionV3	76.74	75.18	77.94	76.53	89.36	0.9188	0.1537
VGG16	74.67	63.84	81.10	71.45	92.56	0.9067	0.1701
VGG19	66.50	49.22	75.16	59.49	91.87	0.8497	0.2235

Table 3. Results of the single-input models using ROIs.

Base model	Accuracy (%)	Recall (%)	Precision (%)	$F1$-score (%)	Specificity (%)	AUC	Error rate
Normal–Abnormal Classification							
ResNet50V2	**98.49**	**98.49**	**98.49**	**98.49**	**98.49**	**0.9991**	**0.0151**
InceptionResNetV2	97.61	97.61	97.61	97.61	97.61	0.9969	0.0239
DenseNet121	96.89	96.89	96.89	96.89	96.89	0.9952	0.0311
InceptionV3	96.85	96.85	96.85	96.85	96.85	0.9930	0.0315
NASNetMobile	96.39	96.39	96.39	96.39	96.39	0.9936	0.0361
VGG16	92.40	92.40	92.40	92.40	92.40	0.9791	0.0760
VGG19	91.02	91.02	91.02	91.02	91.02	0.9688	0.0898
Benign–Malignant Classification							
ResNet50V2	**92.92**	**92.92**	**92.92**	**92.92**	**92.92**	**0.9786**	**0.0708**
InceptionResNetV2	87.86	87.86	87.86	87.86	87.86	0.9552	0.1214
DenseNet121	84.01	84.01	84.01	84.01	84.01	0.9239	0.1599
InceptionV3	83.12	83.12	83.12	83.12	83.12	0.9048	0.1688
NASNetMobile	81.80	81.80	81.80	81.80	81.80	0.8987	0.1820
VGG16	73.39	73.39	73.39	73.39	73.39	0.8033	0.2661
VGG19	66.88	66.88	66.88	66.88	66.88	0.7473	0.3312
Normal–Benign–Malignant Classification							
ResNet50V2	**94.23**	**93.97**	**94.25**	**94.11**	**97.13**	**0.9920**	**0.0392**
InceptionResNetV2	90.69	89.80	91.58	90.68	95.87	0.9831	0.0615
DenseNet121	87.95	87.10	88.90	87.99	94.56	0.9748	0.0792
NASNetMobile	86.94	86.14	87.73	86.92	93.97	0.9672	0.0864
InceptionV3	85.42	84.79	85.98	85.38	93.09	0.9626	0.0968
VGG16	75.35	68.98	80.21	74.17	91.49	0.9148	0.1601
VGG19	72.48	68.39	75.28	71.67	88.77	0.8978	0.1802

Overall, most pre-trained models performed better with ROIs than with FMs, likely due to the removal of irrelevant structures such as pectoral muscles, non-cancerous tissues, and background artifacts. Performance improved further with CLAHE-enhanced ROIs, highlighting the importance of preprocessing and contrast enhancement in CNN-based medical image analysis.

4.3. *Triple-input model results*

This experiment combined the best-performing pre-trained models from the single-input experiments. After comparative analysis, ResNet50V2 consistently outperformed all other models across all classification tasks. Therefore, ResNet50V2 was always part of the optimal configuration of the triple-branch model.

To expand the experimental scope, deeper ResNet architectures (ResNet101V2 and ResNet152V2) were evaluated alongside ResNet50V2. All 27 possible model combinations, accounting for branch order and duplicates were tested. The optimal configuration consisted of ResNet50V2 for Branch-A (FMs), ResNet152V2 for Branch-B (ROIs), and ResNet50V2 for Branch-C (CLAHE-enhanced ROIs).

This configuration achieved excellent results (Table 5). For normal–abnormal classification, it achieved 99.20% across accuracy, precision, recall, $F1$-score, and specificity with an AUC of 0.9991 and an error rate of 0.0080. For benign–malignant

Table 4. Results of the single-input models using CLAHE-enhanced ROIs.

Base model	Accuracy (%)	Recall (%)	Precision (%)	$F1$-score (%)	Specificity (%)	AUC	Error rate
Normal–Abnormal Classification							
ResNet50V2	**98.74**	**98.74**	**98.74**	**98.74**	**98.74**	**0.9986**	**0.0126**
DenseNet121	98.53	98.53	98.53	98.53	98.53	0.9973	0.0147
InceptionResNetV2	97.61	97.61	97.61	97.61	97.61	0.9963	0.0239
InceptionV3	97.10	97.10	97.10	97.10	97.10	0.9950	0.0290
VGG16	96.35	96.35	96.35	96.35	96.35	0.9927	0.0365
NASNetMobile	96.14	96.14	96.14	96.14	96.14	0.9943	0.0386
VGG19	94.12	94.12	94.12	94.12	94.12	0.9863	0.0588
Benign–Malignant Classification							
ResNet50V2	**94.94**	**94.94**	**94.94**	**94.94**	**94.94**	**0.9852**	**0.0506**
InceptionResNetV2	89.57	89.57	89.57	89.57	89.57	0.9630	0.1043
DenseNet121	88.75	88.75	88.75	88.75	88.75	0.9496	0.1125
NASNetMobile	87.48	87.48	87.48	87.48	87.48	0.9442	0.1252
InceptionV3	86.03	86.03	86.03	86.03	86.03	0.9324	0.1397
VGG16	84.32	84.32	84.32	84.32	84.32	0.9213	0.1568
VGG19	79.39	79.39	79.39	79.39	79.39	0.8752	0.2061
Normal–Benign–Malignant Classification							
ResNet50V2	**94.77**	**94.44**	**94.88**	**94.66**	**97.45**	**0.9931**	**0.0355**
DenseNet121	91.82	91.07	92.31	91.68	96.21	0.9835	0.0551
InceptionV3	89.72	89.21	90.24	89.72	95.17	0.9765	0.0681
InceptionResNetV2	89.63	89.13	90.15	89.64	95.13	0.9833	0.0687
NASNetMobile	87.10	86.30	87.75	87.02	93.97	0.9699	0.0858
VGG16	86.43	84.28	88.18	86.19	94.35	0.9700	0.0900
VGG19	82.55	80.15	84.38	82.21	92.58	0.9516	0.1156

classification, it achieved 96.46% across all five metrics with an AUC of 0.9907 and an error rate of 0.0354. For multi-class classification, it achieved 96.59% accuracy, 96.63% precision, 96.59% recall, 96.61% $F1$-score, and 98.31% specificity, 0.9902 AUC and an error rate of 0.0226.

These results demonstrate that combining complementary image types and variant deeper ResNet architectures significantly improves performance over the best single-input models, confirming the effectiveness of multi-input fusion and model diversity for breast cancer classification.

4.4. *Ablation study*

We first evaluated multiple pre-trained models individually using FMs, ROIs, and CLAHE-enhanced ROIs. The results showed that models trained on CLAHE-enhanced ROIs consistently achieved the best results, highlighting the importance of region-focused inputs and contrast enhancement.

Next, the three image types were integrated into a triple-input model, and the configuration using ResNet50V2 for Branch-A, ResNet152V2 for Branch-B, and ResNet50V2 for Branch-C achieved the highest accuracy. Therefore, it was selected as the backbone of Tri-ResNet.

Table 5. Summary of the results of the triple-input models.

#	Base model A, Base model B, Base model C	Precision (%)	Recall (%)	Accuracy (%)	$F1$-score (%)	Specificity (%)	AUC	Error rate
Normal–Abnormal Classification								
1	ResNet50V2, ResNet50V2, ResNet50V2	98.99	98.99	98.99	98.99	98.99	0.9956	0.0101
2	ResNet50V2, ResNet50V2, ResNet101V2	98.61	98.61	98.61	98.61	98.61	0.9951	0.0139
3	ResNet50V2, ResNet50V2, ResNet152V2	98.95	98.95	98.95	98.95	98.95	0.9964	0.0105
		...						
7	**ResNet50V2, ResNet152V2, ResNet50V2**	**99.20**	**99.20**	**99.20**	**99.20**	**99.20**	**0.9985**	**0.0080**
		...						
25	ResNet152V2, ResNet152V2, ResNet50V2	98.91	98.91	98.91	98.91	98.91	0.9943	0.0109
26	ResNet152V2, ResNet152V2, ResNet101V2	98.91	98.91	98.91	98.91	98.91	0.9967	0.0109
27	ResNet152V2, ResNet152V2, ResNet152V2	98.78	98.78	98.78	98.78	98.78	0.9985	0.0122
Benign–Malignant Classification								
1	ResNet50V2, ResNet50V2, ResNet50V2	96.21	96.21	96.21	96.21	96.21	0.9864	0.0379
2	ResNet50V2, ResNet50V2, ResNet101V2	95.32	95.32	95.32	95.32	95.32	0.9836	0.0468
3	ResNet50V2, ResNet50V2, ResNet152V2	95.70	95.70	95.70	95.70	95.70	0.9877	0.0430
		...						
7	**ResNet50V2, ResNet152V2, ResNet50V2**	**96.46**	**96.46**	**96.46**	**96.46**	**96.46**	**0.9907**	**0.0354**
		...						
25	ResNet152V2, ResNet152V2, ResNet50V2	95.32	95.32	95.32	95.32	95.32	0.9874	0.0468
26	ResNet152V2, ResNet152V2, ResNet101V2	94.82	94.82	94.82	94.82	94.82	0.9772	0.0518
27	ResNet152V2, ResNet152V2, ResNet152V2	94.69	94.69	94.69	94.69	94.69	0.9842	0.0531
Normal–Benign–Malignant Classification								
1	ResNet50V2, ResNet50V2, ResNet50V2	95.03	95.03	95.03	95.03	97.51	0.9874	0.0332
2	ResNet50V2, ResNet50V2, ResNet101V2	94.99	94.99	94.99	94.99	97.49	0.9805	0.0334
3	ResNet50V2, ResNet50V2, ResNet152V2	95.62	95.58	95.58	95.60	97.81	0.9847	0.0294
		...						
7	**ResNet50V2, ResNet152V2, ResNet50V2**	**96.63**	**96.59**	**96.59**	**96.61**	**98.31**	**0.9902**	**0.0226**
		...						
25	ResNet152V2, ResNet152V2, ResNet50V2	96.17	96.17	96.17	96.17	98.08	0.9897	0.0256
26	ResNet152V2, ResNet152V2, ResNet101V2	94.27	94.23	94.27	94.25	97.13	0.9791	0.0383
27	ResNet152V2, ResNet152V2, ResNet152V2	96.37	96.25	96.29	96.31	98.19	0.9900	0.0246

Focusing on the challenging benign–malignant classification on the Mini-DDSM dataset, we iteratively refined the triple-input model through extensive experimentation and optimization, ultimately developing the final proposed architecture. The refinement process began by adding a uniform FC layer with 128 neurons to all branches, which slightly reduced accuracy from 96.46% to 95.58%. We then evaluated heterogeneous FC layers (128 neurons for Branch-A, 512 for Branch-B, and 1024 for Branch-C), which improved accuracy to 97.41%. As a final fine-tuning step, we added a 512-neuron FC layer after the concatenation layer, followed by a 128-neuron FC layer. Dropout layers were inserted between FC layers to mitigate overfitting, and batch normalization was applied before them. This configuration achieved an accuracy of 98.42%, using Categorical Crossentropy as the loss function and the Adam optimizer with a learning rate of 0.001.

The final stage of the ablation study focused on hyperparameter optimization. We evaluated alternative loss functions, including Cosine Similarity and Mean Squared

Error, but none matched the performance of Categorical Crossentropy, which consistently delivered the best results. For the optimizer, we tested SGD, RMSProp, Adagrad, and Nadam; although Adagrad achieved comparable performance, Adam was retained due to its stability across earlier experiments. We also tested learning rates of 0.01, 0.001, 0.0001, and 0.00001, with 0.0001 yielding the highest accuracy of 98.48% and thus selected as the optimal value. Finally, extending training to 50 epochs showed that performance stabilized after the 30th epoch, indicating no further improvement beyond that point.

The fine-tuning steps and their corresponding accuracy improvements are summarized in Tables 6 and 7.

4.5. *Tri-ResNet model results*

Building on the findings from previous experiments, the Tri-ResNet architecture was fine-tuned using the optimal hyperparameters and achieved its highest performance on the Mini-DDSM dataset. The model reached 99.54% across accuracy, precision, recall, $F1$-score, and specificity, with an AUC of 0.9995 and an error rate of 0.0046 for normal–abnormal classification. For benign–malignant classification, Tri-ResNet achieved 98.61% across all metrics, with an AUC of 0.9973 and an error rate of 0.0139. In the multi-class classification, it attained 98.78% accuracy, 98.86% precision, 98.74% recall, 98.80% $F1$-score, and 99.43% specificity, with an AUC of 0.9995 and an error rate of 0.0080.

To assess generalization, Tri-ResNet was further evaluated on two additional public mammography datasets, MIAS and INbreast. These datasets were preprocessed, balanced, and augmented using the same pipeline applied to Mini-DDSM. On the MIAS dataset, the model achieved 99.62% across accuracy, precision, recall, $F1$-score, and specificity, with an AUC of 1.0 and an error rate of 0.0038 for normal–abnormal classification. For benign–malignant classification, it achieved 99.14% across all metrics, with an AUC of 0.9998 and an error rate of 0.0086. For multi-class

Table 6. Impact of architectural modifications on model accuracy.

Experiment	Accuracy (%)
Triple-input model (ResNet50V2–ResNet50V2–ResNet50V2)	96.21
Triple-input model (ResNet50V2–ResNet152V2–ResNet50V2)	96.46
Adding a Dense (256) layer after each branch output	95.58
Adding Dense (64) for Branch-A, Dense (128) for Branch-B, and Dense (256) for Branch-C	97.41
Adding Dense (128) for Branch-A, Dense (512) for Branch-B, and Dense (1024) for Branch-C	97.60
Adding Dense (512) after the concatenation layer	97.79
Adding Dense (64) before the output layer	98.36
Adding a dropout layer before the output layer	95.95
Our proposed model (Tri-ResNet) architecture	**98.42**

Table 7. Impact of hyperparameter modifications on model accuracy.

Experiment		Accuracy (%)
Loss function:	Cosine Similarity	94.75
	Mean Squared Error	94.44
	Categorical Crossentropy	**98.42**
Optimizer:	SGD	97.35
	RMSProp	96.27
	Adagrad	98.42
	Nadam	97.22
	Adam	**98.42**
Learning rate:	0.01	95.45
	0.001	98.42
	0.0001	**98.48**
	0.00001	96.59
Epochs:	10	98.48
	30	**98.61**

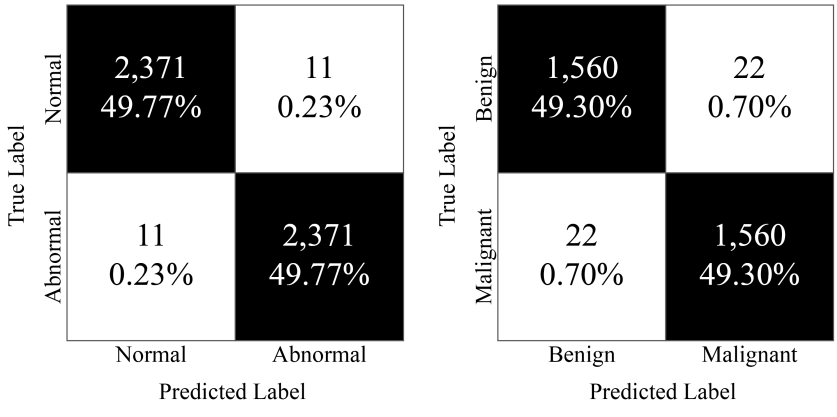
classification, it attained 98.85% accuracy, 100.00% precision, 98.85% recall, 99.42% $F1$-score, and 100.00% specificity, with an AUC of 0.9999 and an error rate of 0.0038.

On the INbreast dataset, Tri-ResNet achieved 98.75% across accuracy, precision, recall, $F1$-score, and specificity, with an AUC of 0.9869 and an error rate of 0.0125 for normal–abnormal classification. For benign–malignant classification, it reached 98.53% across all metrics, with an AUC of 0.9990 and an error rate of 0.0147. For multi-class classification, it attained 98.53% accuracy, 99.01% precision, 98.53% recall, 98.77% $F1$-score, and 99.51% specificity, with an AUC of 0.9996 and an error rate of 0.0081.

Table 8 presents the results of Tri-ResNet on the Mini-DDSM, MIAS, and INbreast datasets.

Table 8. Performance of the Tri-ResNet model on the Mini-DDSM, MIAS, and INbreast datasets.

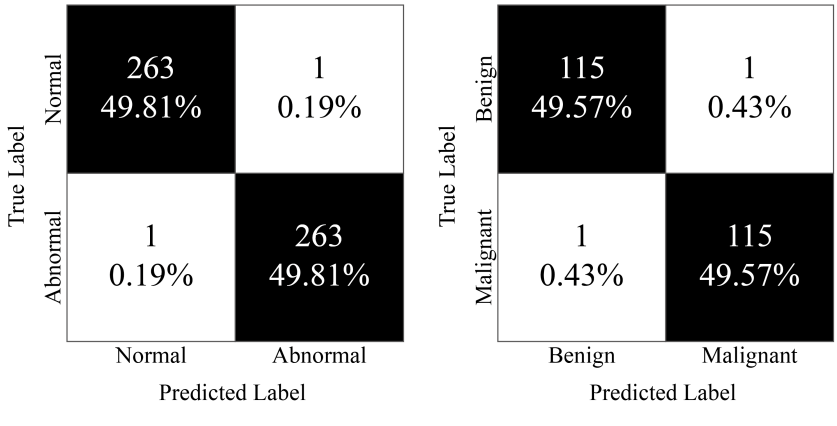
Dataset	Accuracy (%)	Recall (%)	Precision (%)	$F1$-score (%)	Specificity (%)	AUC	Error rate
Normal–Abnormal Classification							
Mini-DDSM	99.54	99.54	99.54	99.54	99.54	0.9995	0.0046
MIAS	99.62	99.62	99.62	99.62	99.62	1.0000	0.0038
INbreast	98.75	98.75	98.75	98.75	98.75	0.9869	0.0125
Benign–Malignant Classification							
Mini-DDSM	98.61	98.61	98.61	98.61	98.61	0.9973	0.0139
MIAS	99.14	99.14	99.14	99.14	99.14	0.9998	0.0086
INbreast	98.53	98.53	98.53	98.53	98.53	0.9990	0.0147
Normal–Benign–Malignant Classification							
Mini-DDSM	98.78	98.74	98.86	98.80	99.43	0.9995	0.0080
MIAS	98.85	98.85	100.00	99.42	100.00	0.9999	0.0038
INbreast	98.53	98.53	99.01	98.77	99.51	0.9996	0.0081



(a) Normal–Abnormal Classification (b) Benign–Malignant Classification

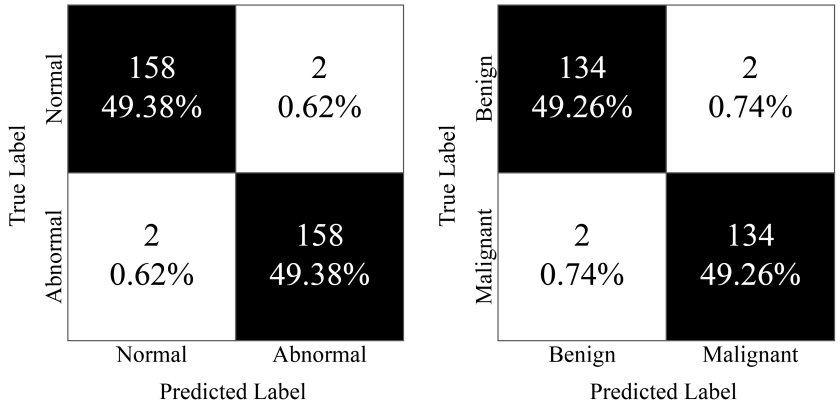
Fig. 7. Confusion matrices of the Tri-ResNet model on the Mini-DDSM dataset.

Figures 7–9 present the confusion matrices of the Tri-ResNet model across all datasets. The matrices display both the predicted case counts and their corresponding percentages for the binary classifications. High TP values indicate the model's effectiveness in correctly identifying abnormal cases in the first classification and malignant cases in the second. Whereas high TN values reflect its ability to accurately recognize normal and benign cases in both binary classifications. Low FP values are important, as they reduce the number of normal and benign cases incorrectly classified as abnormal and malignant. FN values correspond to abnormal and malignant cases misclassified as normal and benign. Minimizing these values is critical as missed abnormal and malignant cases can lead to delayed diagnosis and serious clinical consequences.



(a) Normal–Abnormal Classification (b) Benign–Malignant Classification

Fig. 8. Confusion matrices of the Tri-ResNet model on the MIAS dataset.



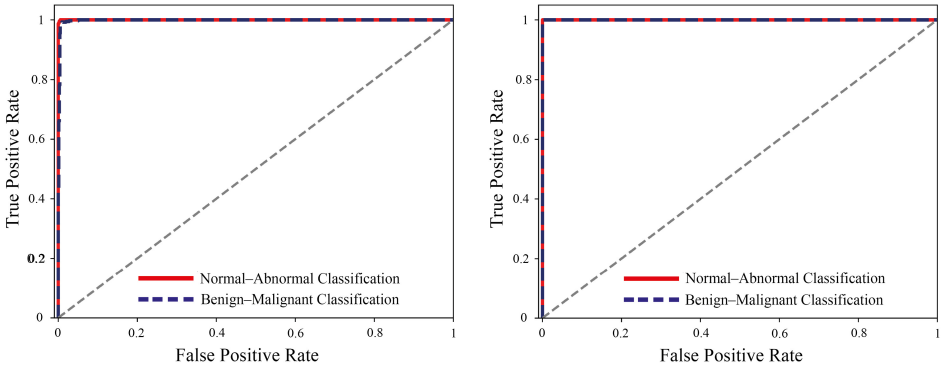
(a) Normal–Abnormal Classification (b) Benign–Malignant Classification

Fig. 9. Confusion matrices of the Tri-ResNet model on the INbreast dataset.

Figure 10 shows the ROC curves for the binary classifications, offering additional insight into the model's performance. The consistently high AUC values across all datasets demonstrate the model's strong capability. The ROC curves rise near the origin, reflecting high TP rates with minimal FP rates, and remain close to the upper-left corner of the ROC space, reflecting high TP rates with minimal FP rates. Overall, these ROC curves provide a comprehensive visual representation of the robustness and reliability of the Tri-ResNet model in breast cancer classification.

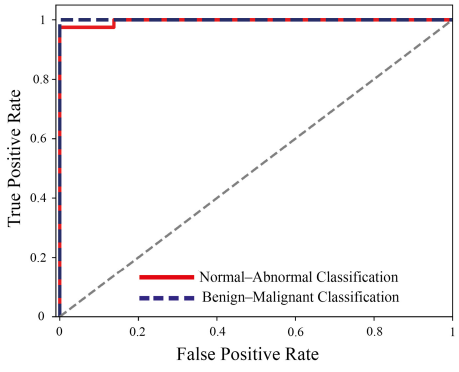
4.6. *Experimental environment and settings*

We conducted our experiments using Google Colab's cloud-based environment with T4 GPU acceleration. Model development and fine-tuning were performed in Python



(a) Mini-DDSM dataset (b) MIAS dataset

Fig. 10. ROC curves of the Tri-ResNet model.



(c) INbreast dataset

Fig. 10. (*Continued*)

3.10.12 using TensorFlow 2.15.0, while OpenCV library was used for image pre-processing. Table 9 summarizes the hyperparameter settings of the proposed model.

4.7. *Comparison with state-of-the-art models*

Table 10 compares the performance of the proposed Tri-ResNet model with state-of-the-art approaches on the CBIS-DDSM, Mini-DDSM, and DDSM datasets for both binary and multi-class classifications. Tri-ResNet consistently outperformed existing models across all evaluation metrics and all classifications. The only exception occurs in the benign–malignant classification, where the model by Maqsood *et al.*[39] achieved higher accuracy, recall, and specificity. However, Tri-ResNet achieved comparable results and was further evaluated using precision, $F1$-score, and AUC metrics, suggesting stronger overall performance. Additionally, the model was further

Table 9. Hyperparameter values used for the proposed model.

Parameter	Value
Batch size	128
Loss function	Categorical Crossentropy
Optimization function	Adam
Learning rate	0.0001
Momentum	0.9
Epochs	30
Callback	ModelCheckpoint (saving weights based on best validation accuracy)
TensorFlow version	2.15.0
Initial pre-trained weights	ImageNet
Dropout	0.2
Activation function	Softmax
Input shape	(224, 224, 3)
Training split	0.7

Table 10. Comparison between the Tri-ResNet model and state-of-the-art models using Mini-DDSM, CBIS-DDSM, and DDSM datasets.

Study	Dataset	Accuracy (%)	Recall (%)	Precision (%)	*F*1-score (%)	Specificity (%)	AUC
Normal–Abnormal Classification							
Ref. 46	Mini-DDSM	99.17	99.25	99.50	99.37	99.00	0.9917
Ref. 41	DDSM	97.26	92.37	95.31	93.77	92.37	—
Ref. 32	DDSM	94.30	93.70	93.50	93.60	—	0.9330
Tri-ResNet	Mini-DDSM	**99.54**	**99.54**	**99.54**	**99.54**	**99.54**	**0.9995**
Benign–Malignant Classification							
Ref. 39	DDSM	**99.08**	**99.19**	—	—	**98.96**	—
Ref. 32	DDSM	98.02	—	—	98.15	—	0.9827
Ref. 46	Mini-DDSM	98.25	98.00	98.00	—	98.49	0.9800
Ref. 28	DDSM	96.98	97.63	97.51	95.97	—	0.9846
Ref. 45	CBIS-DDSM	95.40	95.40	95.35	—	—	0.9800
Ref. 24	DDSM	90.68	92.72	—	—	88.21	—
Ref. 48	CBIS-DDSM	68.63	—	—	—	—	0.7137
Tri-ResNet	Mini-DDSM	98.61	98.61	**98.61**	**98.61**	98.61	**0.9973**
Normal–Benign–Malignant Classification							
Ref. 58	DDSM	96.47	—	—	—	—	0.9600
Ref. 35	DDSM	96.00	94.70	97.00	—	—	—
Ref. 26	Mini-DDSM	65.00	—	—	—	—	—
Tri-ResNet	Mini-DDSM	**98.78**	**98.74**	**98.86**	**98.80**	**99.43**	**0.9995**

evaluated on the MIAS and INbreast to assess its generalization across different mammography datasets.

Similarly, Table 11 presents a comparison with state-of-the-art methods on the MIAS dataset. Tri-ResNet achieved the highest performance across all metrics and classification tasks.

Finally, Table 12 shows the comparison on the INbreast dataset. While the dual-input model reported by Jabeen *et al.*[45] achieved higher performance for benign–malignant classification. However, it performed less effectively on the CBIS-DDSM

Table 11. Comparison between the Tri-ResNet model and state-of-the-art models using MIAS dataset.

Study	Dataset	Accuracy (%)	Recall (%)	Precision (%)	*F*1-score (%)	Specificity (%)	AUC
Normal–Abnormal Classification							
Ref. 34	MIAS	90.56	—	—	—	—	0.8850
Tri-ResNet	MIAS	**99.62**	**99.62**	**99.62**	**99.62**	**99.62**	**1.0000**
Benign–Malignant Classification							
Ref. 28	MIAS	96.99	96.88	96.99	96.45	—	0.9781
Ref. 39	MIAS	96.57	96.11	97.06	96.58	97.03	0.9657
Ref. 24	MIAS	96.55	97.28	—	—	95.92	—
Tri-ResNet	MIAS	**99.14**	**99.14**	**99.14**	**99.14**	**99.14**	**0.9998**
Normal–Benign–Malignant Classification							
Ref. 31	MIAS	98.56	98.82	98.36	98.56	99.33	—
Ref. 35	MIAS	94.50	96.32	91.80	—	—	—
Ref. 58	MIAS	92.54	—	—	—	—	0.8500
Tri-ResNet	MIAS	**98.85**	**98.85**	**100.00**	**99.42**	**100.00**	**0.9999**

Table 12. Comparison between the Tri-ResNet model and state-of-the-art models using the INbreast dataset.

Study	Dataset	Accuracy (%)	Recall (%)	Precision (%)	$F1$-score (%)	Specificity (%)	AUC
Benign–Malignant Classification							
Ref. 45	INbreast	**99.40**	99.40	**99.40**	**99.40**	—	**1.0000**
Ref. 39	INbreast	96.82	95.99	97.73	96.85	97.68	0.9684
Ref. 49	INbreast	95.65	**100.00**	95.00	97.43	75.00	0.8800
Ref. 24	INbreast	91.28	99.43	—	—	83.13	—
Tri-ResNet	INbreast	98.53	98.53	98.53	98.53	**98.53**	0.9990
Normal–Benign–Malignant Classification							
Ref. 43	INbreast	98.08	97.93	97.13	97.51	—	—
Ref. 31	INbreast	92.26	92.95	91.22	88.53	92.95	0.9751
Tri-ResNet	INbreast	**98.53**	**98.53**	**99.01**	**98.77**	**99.51**	**0.9996**

dataset. In contrast, the proposed Tri-ResNet model demonstrated consistently strong performance across all three datasets, confirming its robustness and generalizability.

5. Discussion of Limitations and Future Work

The results of this study highlight the effectiveness of the proposed Tri-ResNet model for the classification of breast cancer. Its strengths include consistently high performance across three independent datasets and its ability to integrate complementary image features through a triple-input architecture for more precise predictions. The use of TL and fine-tuned models helps overcome the challenges of limited dataset size faced in the MIAS and INbreast datasets. Additionally, the modular design of Tri-ResNet allows flexible adaptation to other imaging modalities by adjusting the number of inputs and branches and reconfiguring the backbone networks with lightweight pre-trained alternatives.

The Tri-ResNet model was trained and evaluated in a controlled environment on Google Colab using an NVIDIA Tesla T4 GPU (16 GB). Training required approximately 1.5 h for the normal–abnormal classification (30 epochs), 1.1 h for the benign–malignant classification, and 1.4 h for the multi-class classification, with resource usage reaching 13.7 GB of VRAM and 6.7 GB of RAM. The model includes 167.8 million trainable parameters out of a total of 273.3 million.

During inference, the average processing time was 93 ms per image for both binary classifications and 94 ms for the multi-class classification, with memory consumption of 4.1 GB VRAM and 2.8 GB RAM. Table 13 compares the training and inference times of Tri-ResNet with single-input architectures (ResNet50V2 and ResNet152V2) under the same experimental conditions. Although the multi-branch architecture significantly improves performance across all metrics, it also increases computational complexity. Training time increased by approximately +497.4% and +112.1% relative to ResNet50V2 and ResNet152V2, respectively, and inference time increased by +122.21% and +51.33%. Despite these increases, the Tri-ResNet model still

Table 13. Training and inference time comparison.

Model architecture	Training time (s/30 epochs)	Average inference time (ms/input)
Normal–Abnormal Classification		
Single-input model (ResNet50V2)	925	40
Single-input model (ResNet152V2)	2,556	59
Tri-ResNet	5,321	93
Benign–Malignant Classification		
Single-input model (ResNet50V2)	599	44
Single-input model (ResNet152V2)	1,712	61
Tri-ResNet	4,001	93
Normal–Benign–Malignant Classification		
Single-input model (ResNet50V2)	906	42
Single-input model (ResNet152V2)	2,580	65
Tri-ResNet	5,199	94

satisfies the real-time prediction requirement (below 100 ms), making it suitable for clinical decision-support systems.

Despite its strong classification performance, the proposed Tri-ResNet model presents certain operational limitations that must be addressed for effective real-world deployment. The model introduces an operational dependency on the availability of ROIs corresponding to suspicious regions in the mammogram. In this study, ROI extraction was based on the tumor masks provided in the used datasets. However, this is not the case in real-world screening scenarios, meaning that ROI extraction must occur prior to the classification stage. This dependency can be addressed through one of the following strategies:

- Automated detection or segmentation: A detection or segmentation model, such as a U-Net-based architecture or a YOLO-based mass detector, can be integrated upstream of the classifier. This module would automatically identify suspicious regions, extract the corresponding ROIs, and feed them into the triple-input classifier. Although this enables a fully automated end-to-end workflow, the performance of the detection stage directly influences the final classification accuracy, and errors in ROI localization may propagate to the classifier.
- Manual annotation by radiologists: Radiologists using annotation tools may manually describe ROIs. While this ensures accurate localization of suspicious areas, it is time-consuming and unsuitable for large-scale screening applications. Although unsuitable for real-time deployment, manual annotation remains valuable for ensuring high-quality ground truth.

Beyond its strong performance on Mini-DDSM, the model also demonstrated good generalization on the MIAS and INbreast datasets. Nonetheless, validation on large, multi-institutional datasets remains essential to fully establish robustness and real-world applicability.

To address these limitations, future work will focus on distributed training and inference, pruning redundant layers, knowledge distillation, and quantization to

reduce computational cost and memory usage ensuring that the model remains efficient and scalable for clinical deployment. In parallel, automated ROI extraction must be explicitly incorporated into the workflow, resulting in a fully operational CAD system that follows a two-stage pipeline: (1) automated lesion detection and segmentation for ROI extraction, and (2) multi-input CNN classification using Tri-ResNet.

This complete two-stage pipeline will be implemented and evaluated on real-world mammograms by integrating an automated lesion detection module with the proposed triple-input classifier, ultimately enabling a fully automated clinical decision-support system for breast cancer detection and classification.

6. Conclusion

In this paper, we proposed a multi-branch architecture for breast cancer classification. The model, named Tri-ResNet, is a triple-input CNN that integrates the strengths of ResNet50V2 and ResNet152V2. These pre-trained models were selected through a systematic evaluation of nine state-of-the-art architectures (ResNet50V2, ResNet101V2, ResNet152V2, InceptionV3, InceptionResNetV2, VGG16, VGG19, DenseNet121, and NASNetMobile), which were rigorously assessed for their effectiveness in classifying three distinct input types (FMs, ROIs, and CLAHE-enhanced ROIs). The top-performing models were incorporated into the final Tri-ResNet architecture, comprising three parallel branches: Branch-A (ResNet50V2) for FMs, Branch-B (ResNet152V2) for ROIs, and Branch-C (ResNet50V2) for CLAHE-enhanced ROIs. This configuration consistently outperformed all single-input models across both binary and multi-class classification tasks, demonstrating the advantage of integrating complementary image representations within a unified framework.

Although Tri-ResNet achieved superior performance on the Mini-DDSM, MIAS, and INbreast datasets, its multi-branch design increases computational complexity, memory usage, and inference time compared to single-branch models. Nevertheless, the model showed strong generalization when evaluated on external datasets, achieving comparable results across independent imaging sources. An additional limitation is the model's reliance on pre-annotated lesion masks for ROI extraction.

Future work will focus on (1) developing a fully automated end-to-end CAD system for breast cancer detection, segmentation, and classification, (2) validating the Tri-ResNet model on large and private mammography datasets (currently under construction) to further enhance robustness and generalization, and (3) addressing computational constraints through model optimization to maintain diagnostic performance while improving efficiency.

Data Availability

The datasets used in this study are publicly available from their original sources: Mini-DDSM (https://www.kaggle.com/datasets/cheddad/miniddsm2).

MIAS https://www.repository.cam.ac.uk/handle/1810/250394).
INbreast (https://www.kaggle.com/datasets/martholi/inbreast).

ORCID

Abdelhafidh Kacher https://orcid.org/0000-0001-8548-6385
Medjeded Merati https://orcid.org/0000-0003-2525-6814
Saïd Mahmoudi https://orcid.org/0000-0001-8272-9425

References

1. C. E. DeSantis, F. Bray, J. Ferlay, J. Lortet-Tieulent, B. O. Anderson and A. Jemal, International variation in female breast cancer incidence and mortality rates, *Cancer Epidemiol. Biomark. Prev.* **24**(10) (2015) 1495–1506, doi: 10.1158/1055-9965.EPI-15-0535.
2. National Institute of Public Health, Ministry of Health, Algeria, Registre des tumeurs d'Alger (2023), https://insp.dz/images/PDF/Les%20registres/TumeursAlger/Registre%20des%20Tumeurs%20d'Alger%20bulletin%202023%20final.pdf.
3. J. Ferlay, M. Ervik, F. Lam, M. Laversanne, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram and F. Bray, Global cancer observatory: Cancer today (2024), https://gco.iarc.who.int/today.
4. S. Gao, J. Liu, L. Li, D. Yang, Y. Miao, X. Zhang, Q. Han, Y. Shi, J. Wu and K. Zhang, Application of deep learning technology in breast cancer: A systematic review of segmentation, detection, and classification approaches, *BioMed. Eng. OnLine* **25** (2026) 19, doi: 10.1186/s12938-025-01502-5.
5. O. Díaz, A. Rodríguez-Ruíz and I. Sechopoulos, Artificial intelligence for breast cancer detection: Technology, challenges, and prospects, *Eur. J. Radiol.* **175** (2024) 111457, doi: 10.1016/j.ejrad.2024.111457.
6. E. B. Friedberg, D. Corn, J. D. Prologo, H. Fleishon, R. Pyatt, R. Duszak and P. Cook, Access to interventional radiology services in small hospitals and rural communities: An ACR membership intercommission survey, *J. Am. Coll. Radiol.* **16**(2) (2019) 185–193, doi: 10.1016/j.jacr.2018.10.002.
7. H. Fujita, AI-based computer-aided diagnosis (AI-CAD): The latest review to read first, *Radiol. Phys. Technol.* **13**(1) (2020) 6–19, doi: 10.1007/S12194-019-00552-4.
8. K. Kansal, S. Kumar and K. Kansal, Advances in deep learning techniques for breast cancer classification: A comprehensive review, *Arch. Comput. Methods Eng.* **33** (2026) 187–222, doi: 10.1007/s11831-025-10306-8.
9. J. Majidpour and H. Beitollahi, A comprehensive examination of machine learning and deep learning approaches for breast cancer detection, classification, segmentation, augmentation, and feature selection, *Arch. Comput. Methods Eng.* **33** (2025) 1–32, https://doi.org/10.1007/s11831-025-10359-9.
10. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.* **115**(3) (2015) 211–252, doi: 10.1007/s11263-015-0816-y.
11. H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros and T. Ganslandt, Transfer learning for medical image classification: A literature review, *BMC Med. Imaging* **22**(1) (2022) 69, doi: 10.1186/s12880-022-00793-7.
12. L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A. S. Albahri, B. S. N. Al-dabbagh, M. A. Fadhel, M. Manoufali, J. Zhang, A. H. Al-Timemy, Y. Duan, A. Abdullah,

L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh and Y. Gu, A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications, *J. Big Data* **10**(1) (2023) 1–82, doi: 10.1186/S40537-023-00727-2.

13. K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2016)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
14. K. He, X. Zhang, S. Ren and J. Sun, Identity mappings in deep residual networks, *Proc. 14th European Conf. Computer Vision (ECCV 2016)*, Amsterdam, Netherlands, 2016, pp. 630–645, doi: 10.1007/978-3-319-46493-0_38.
15. C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, Inception-v4, Inception-ResNet and the impact of residual connections on learning, *Proc. Conf. Artificial Intelligence (AAAI 2017)*, San Francisco, CA, USA, 2017, pp. 770–778, doi: 10.1609/aaai.v31i1.11231.
16. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, HI, USA, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
17. B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, Learning transferable architectures for scalable image recognition, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, USA, 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.
18. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the Inception architecture for computer vision, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2016)*, Las Vegas, NV, USA, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
19. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *Proc. IEEE 3rd Int. Conf. Learning Representations (ICLR 2015)*, San Diego, CA, USA, 2015, pp. 1–14, doi: 10.48550/arXiv.1409.1556.
20. K. Zuiderveld, Contrast limited adaptive histogram equalization, in *Graphics Gems*, ed. P. S. Heckbert (Academic Press, 1994), pp. 474–485, doi: 10.1016/B978-0-12-336156-1.50061-6.
21. C. D. Lekamlage, F. Afzal, E. Westerberg and A. Cheddad, Mini-ddsm: Mammography-based automatic age estimation, *Proc. 3rd Int. Conf. Digital Medicine and Image Processing (DMIP 2020)*, Kyoto, Japan, 2021, pp. 1–6, doi: 10.1145/3441369.3441370.
22. J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal and J. Savage, Mammographic image analysis society (MIAS) database v1.21 (2015), doi:10.17863/CAM.105113.
23. I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso and J. S. Cardoso, INbreast: Toward a full-field digital mammographic database, *Acad. Radiol.* **19**(2) (2012) 236–248, doi: 10.1016/J.ACRA.2011.09.014.
24. D. Muduli, R. Dash and B. Majhi, Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach, *Biomed. Signal Process. Control* **71** (2022) 102825, doi: 10.1016/j.bspc.2021.102825.
25. M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang and S. Munishkumaran, Current status of the digital database for screening mammography, in *Digital Mammography: Nijmegen, 1998*, eds. N. Karssemeijer, M. Thijssen, J. Hendriks and L. van Erning (Springer, Dordrecht, 1998), pp. 457–460, doi: 10.1007/978-94-011-5318-8_75.
26. S. Mohapatra, S. Muduly, S. Mohanty, J. Ravindra and S. N. Mohanty, Evaluation of deep learning models for detecting breast cancer using histopathological mammograms images, *Sustain. Oper. Comput.* **3** (2022) 296–302, doi: 10.1016/j.susoc.2022.06.001.
27. A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM* **60**(6) (2017) 84–90, doi: 10.1145/3065386.

28. W. M. Salama, A. M. Elbagoury and M. H. Aly, Novel breast cancer classification framework based on deep learning, *IET Image Process.* **14**(13) (2020) 3254–3259, doi: 10.1049/iet-ipr.2020.0122.
29. C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.* **20**(3) (1995) 273–297, doi: 10.1007/BF00994018.
30. R. Sawyer-Lee, F. Gimenez, A. Hoogi and D. Rubin, Curated breast imaging subset of digital database for screening mammography (CBIS-DDSM) (2016), doi:10.7937/K9/TCIA.2016.7O02S9CY.
31. M. F. Aslan, A hybrid end-to-end learning approach for breast cancer diagnosis: Convolutional recurrent network, *Comput. Electr. Eng.* **105** (2023) 108562, doi: 10.1016/j.compeleceng.2022.108562.
32. A. Khamparia, S. Bharati, P. Podder, D. Gupta, A. Khanna, T. K. Phung and D. N. Thanh, Diagnosis of breast cancer based on modern mammography using hybrid transfer learning, *Multidimens. Syst. Signal Process.* **32**(2) (2021) 747–765, doi: 10.1007/S11045-020-00756-7.
33. M. M. Al-Nawashi, O. M. Al-Hazaimeh and M. K. Khazaaleh, A new approach for breast cancer detection-based machine learning technique, *Appl. Comput. Sci.* **20**(1) (2024) 1–16, doi: 10.35784/acs-2024-01.
34. M. M. Eltoukhy, M. Elhoseny, K. M. Hosny and A. K. Singh, Computer aided detection of mammographic mass using exact Gaussian–Hermite moments, *J. Ambient Intell. Humaniz. Comput.* **15**(1) (2024) 1139–1147, doi: 10.1007/s12652-018-0905-1.
35. Z. Jafari and E. Karami, Breast cancer detection in mammography images: A CNN-based approach with feature selection, *Information* **14**(7) (2023) 410, doi: 10.3390/info14070410.
36. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, USA, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
37. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell and S. Xie, A ConvNet for the 2020s, *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 11966–11976, doi: 10.1109/CVPR52688.2022.01167.
38. M. Tan and Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, *Proc. 36th Int. Conf. Machine Learning*, Long Beach, CA, USA, 2019, pp. 6105–6114, doi: 10.48550/arXiv.1905.11946.
39. S. Maqsood, R. Damaševičius and R. Maskeliūnas, TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages, *Appl. Sci.* **12**(7) (2022) 3273, doi: 10.3390/app12073273.
40. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2015)*, Boston, MA, USA, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
41. M. H. Alshayeji and J. Al-Buloushi, Breast cancer classification using concatenated triple convolutional neural networks model, *Big Data Cogn. Comput.* **7**(3) (2023) 142, doi: 10.3390/bdcc7030142.
42. F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, HI, USA, 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
43. A. D. Mohammed and D. Ekmekci, Breast cancer diagnosis using YOLO-based multiscale parallel CNN and flattened threshold swish, *Appl. Sci.* **14**(7) (2024) 2680, doi: 10.3390/app14072680.

44. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You only look once: Unified, real-time object detection, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2016)*, Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
45. K. Jabeen, M. A. Khan, J. Balili, M. Alhaisoni, N. A. Almujally, H. Alrashidi, U. Tariq and J.-H. Cha, 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, doi: 10.3390/diagnostics13071238.
46. A. Sahu, P. K. Das and S. Meher, High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets, *Biomed. Signal Process. Control* **80** (2023) 104292, doi: 10.1016/j.bspc.2022.104292.
47. X. Zhang, X. Zhou, M. Lin and J. Sun, ShuffleNet: An extremely efficient convolutional neural network for mobile devices, *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, USA, 2018, pp. 6848–6856, doi: 10.1109/CVPR.2018.00716.
48. S. Sarker, P. Sarker, G. Bebis and A. Tavakkoli, MV-Swin-T: Mammogram classification with multi-view Swin transformer, *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI 2024)*, Athens, Greece, 2024, pp. 1–5, doi: 10.1109/ISBI56570.2024.10635578.
49. N. M. Hassan, S. Hamad and K. Mahar, YOLO-based CAD framework with ViT transformer for breast mass detection and classification in CESM and FFDM images, *Neural Comput. Appl.* **36**(12) (2024) 6467–6496, doi: 10.1007/s00521-023-09364-5.
50. S. R. Kebede, F. G. Waldamichael, T. G. Debelee, M. Aleme, W. Bedane, B. Mezgebu and Z. C. Merga, Dual view deep learning for enhanced breast cancer screening using mammography, *Sci. Rep.* **14**(1) (2024) 1–15, doi: 10.1038/s41598-023-50797-8.
51. P. Soille, *Morphological Image Analysis: Principles and Applications* (Springer, Berlin, 2004), doi: 10.1007/978-3-662-05088-0_3.
52. X. Sun, Y. Gao, T. Xin and N. Song, Binary restrictive threshold method for item exposure control in cognitive diagnostic computerized adaptive testing, *Front. Psychol.* **12** (2021) 517155, doi: 10.3389/fpsyg.2021.517155.
53. Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz and S.-H. Wang, Improved breast cancer classification through combining graph convolutional network and convolutional neural network, *Inf. Process. Manage.* **58**(2) (2021) 102439, doi: 10.1016/j.ipm.2020.102439.
54. G. Naidu, T. Zuva and E. M. Sibanda, A review of evaluation metrics in machine learning algorithms, in *Proc. Artificial Intelligence Application in Networks and Systems (CSOC 2023)*, eds. R. Silhavy and P. Silhavy (Springer International Publishing, Cham, 2023), pp. 15–25, doi: 10.1007/978-3-031-35314-7_2.
55. A. A. Taha and A. Hanbury, Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool, *BMC Med. Imaging* **15** (2015) 29, doi: 10.1186/s12880-015-0068-x.
56. C. J. V. Rijsbergen, A theoretical basis for the use of co-occurrence data in information retrieval, *J. Doc.* **33**(2) (1977) 106–119, doi: 10.1108/eb026637.
57. J. A. Hanley and B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology* **143** (1982) 29–36, doi: 10.1148/radiology.143.1.7063747.
58. V. S. Gnanasekaran, S. Joypaul, P. Meenakshi Sundaram and D. D. Chairman, Deep learning algorithm for breast masses classification in mammograms, *IET Image Process.* **14**(12) (2020) 2860–2868, doi: 10.1049/iet-ipr.2020.0070.