**International Journal of Electrical and Computer Engineering (IJECE)**
Vol. 16, No. 3, June 2026, pp. 1199~1212
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1199-1212 

# Optimized ResNet-50 framework for mammogram-based breast cancer classification: a comparative evaluation with EfficientNet-B0

**Muhammad Subali[1], Lulu Mawaddah Wisudawati[2], Teresa[3]**
[1]Department of Computer Science and Information Technology, Cendekia Abditama University, Tangerang, Indonesia
[2]Department of Informatics, Gunadarma University, Depok, Indonesia
[3]Department of Nursing, Cendekia Abditama University, Tangerang, Indonesia

**Article Info**

***Article history:***

Received Aug 26, 2025
Revised Jan 24, 2026
Accepted Mar 16, 2026

***Keywords:***

Breast cancer classification
Computer-aided diagnosis
Deep learning
EfficientNet-B0
Mammography
ResNet-50

**ABSTRACT**

Breast cancer remains one of the most prevalent malignancies worldwide, underscoring the need for accurate and reliable mammographic interpretation. Computer-aided diagnosis (CAD) based on deep learning has emerged as a promising approach to improve both screening performance and diagnostic consistency, yet fairness-driven comparisons between popular convolutional backbones on public mammogram benchmarks remain limited. This study provides a statistically validated, fairness-driven comparison of two widely used convolutional neural network architectures, ResNet-50 and EfficientNet-B0, for mammogram-based breast cancer classification under a rigorously controlled, clinically motivated protocol. The proposed "optimized ResNet-50" framework is defined by patient-level stratified undersampling, paired 5-fold cross-validation with identical partitions, harmonized augmentation and training configurations, and dual statistical testing (paired t-tests and Wilcoxon signed-rank tests), emphasizing methodological rigor rather than architectural novelty. Across MIAS and CBIS-DDSM benchmarks, the models demonstrated complementary strengths, with EfficientNet-B0 excelling in screening-oriented tasks (normal vs. abnormal) and ResNet-50 offering more robust performance for diagnostic-oriented tasks (benign vs. malignant). These findings highlight the value of fairness-driven evaluation protocols in CAD research and support the feasibility of integrating lightweight convolutional neural networks (CNNs) into tiered clinical workflows, where different backbones are strategically deployed for initial screening and confirmatory assessment.



***Corresponding Author:***

Muhammad Subali
Department of Computer Science and Information Technology, Cendekia Abditama University
Islamic Raya St., Kelapa Dua, Tangerang, Banten, Indonesia
Email: subali@uca.ac.id

## 1. INTRODUCTION

Breast cancer remains one of the most prevalent and life-threatening malignancies worldwide, imposing a substantial burden of incidence and mortality among women [1]. This burden calls for resource-stratified strategies to strengthen early detection and care pathways across diverse health-system contexts [2], with additional global-burden evidence further underscoring the need for timely intervention [3]. Mammography is the most widely adopted screening tool and is standardized through the BI-RADS® reporting framework; however, interpretation variability and the subtlety of certain lesion patterns can lead to

false positives and false negatives with downstream clinical and economic consequences [4]. Population-level estimates also confirm breast cancer's large contribution to female cancer incidence and deaths, reinforcing the urgency of effective screening programs [5]. Recent evaluations show that deep learning–based computer-aided diagnosis (CAD) can improve mammographic assessment and function as a complementary aid in clinical workflows [6], [7]. Broader deep-learning literature highlights the capacity of neural networks to learn multi-level representations and to support discovery across heterogeneous biomedical data [8], [9]. Transfer learning, in particular, facilitates robust performance in medical imaging despite limited annotated datasets [10].

Within this context, CAD leveraging machine learning (ML) and deep learning (DL) enables automated extraction of hierarchical, discriminative features that often surpass handcrafted descriptors in mammography and related modalities [11], while pretrained convolutional neural networks (CNNs) tailored to mammography continue to mature toward state-of-the-art practice [12]. EfficientNet variants have been explored for breast-imaging tasks with promising accuracy and scalability characteristics [13], [14]. In this study, we therefore focus on optimizing ResNet-50 for mammogram-based breast-cancer classification and provide a direct comparison to EfficientNet-B0.

At the backbone level, residual connections in ResNet stabilize deep optimization and enable effective feature reuse, and ResNet-50 has demonstrated robust performance in breast-cancer imaging applications [15]–[17]. Beyond architectures, feature engineering and baseline modelling remain relevant. Prior work spans handcrafted texture descriptors and CNN-based representations, each showing tangible gains in discriminability across datasets and tasks [18]–[20]. Dimensionality reduction and feature selection (*e.g.*, LASSO) have also been reported to improve performance and computational efficiency in breast-cancer classification pipelines [21]. Classical ML baselines—such as support vector machines and ensemble learners—have achieved strong results in several studies and datasets, although their reliance on engineered features can limit generalization across heterogeneous cohorts [22]–[25]. A broader review of ML/DL for medical image analysis further emphasizes the benefits of learned representations for end-to-end diagnostic modelling [26], and recent mammography applications on public datasets continue to show detection improvements under practical settings [27]. Comparative studies of ML algorithms also provide performance context across varied scenarios and evaluation protocols [28].

Despite these advances, class imbalance—where benign cases often outnumber malignant ones—remains a challenge that can shift decision boundaries and reduce sensitivity to minority classes in screening contexts. Recent pipelines have therefore incorporated different balancing strategies, such as stratified undersampling, class-weighted losses, and synthetic oversampling (*e.g.*, SMOTE), with demonstrated effectiveness in both ResNet-50-based configurations and contemporary deep-learning frameworks for breast-cancer detection and classification [29], [30]. In this study, we employ stratified undersampling at the patient level to mitigate majority-class bias while ensuring fair representation across training and validation folds. To reduce potential information loss, undersampling is combined with standardized augmentation and rigorous evaluation protocols.

In this study, we design and validate an optimized ResNet-50 framework for mammogram-based breast-cancer classification, with a direct comparison to EfficientNet-B0. Our methodology comprises: i) Rigorous preprocessing and augmentation tailored to mammographic variability; ii) Integration of patient-level stratified undersampling with a fixed random seed to balance classes reproducibly across folds; iii) Standardized transfer learning with frozen backbones and an identical minimal classifier head. Here, "optimized" highlights methodological improvements (balanced splitting, controlled classifier structure, and systematic hyperparameter tuning) designed to ensure reproducibility and clinical relevance, rather than proposing a novel network architecture; and iv) Comprehensive benchmarking against EfficientNet-B0 across accuracy, sensitivity, specificity, F1-score, and ROC–AUC. To ensure robustness, performance was reported as mean ± standard deviation (SD) across cross-validation folds, 95% confidence intervals (CI) for ROC–AUC estimated using the t-distribution across folds, and statistical validation using paired t-tests and Wilcoxon signed-rank tests.

In summary, this study makes three main contributions. First, it proposes a replicable, practice-minded optimization protocol for ResNet-50 tailored to mammography, reflecting both technical and clinical constraints. Second, it provides head-to-head comparative insights between ResNet-50 and EfficientNet-B0 with statistical validation across paired cross-validation folds under a fairness-driven experimental design. Third, it emphasizes clinically meaningful evaluation—including patient-level validation, sensitivity-focused operating points, and robust confidence-interval estimation—toward real-world CAD integration. Unlike prior incremental comparisons, the novelty of this work lies in this fairness-driven experimental design and clinically oriented evaluation protocol rather than in proposing a new neural architecture.

Through these contributions, the study aims to bridge the gap between algorithmic advances and clinical deployment, ensuring that CAD systems not only achieve high accuracy but also demonstrate

reliability, interpretability, and efficiency in breast cancer screening. The remainder of this paper is organized as follows. Section 2 presents the materials and methods, section 3 reports the experimental results and discusses the findings, and section 4 concludes the study.

## 2. METHOD

The overall workflow of the proposed framework is illustrated in Figure 1. The system begins with the acquisition of mammogram images obtained from two public benchmark datasets. Two experimental scenarios are considered: i) normal vs. abnormal classification using the MIAS dataset, and ii) benign vs. malignant classification using the CBIS-DDSM dataset. All images are subjected to preprocessing and region of interest (ROI) extraction. This step includes grayscale conversion, noise suppression with median filtering, breast boundary detection, removal of non-breast structures (*e.g.*, pectoral muscle), and cropping of relevant regions of interest. For the normal vs. abnormal scenario, the ROI corresponds to the whole breast area, while for the benign vs. malignant scenario, candidate mass regions are segmented and cropped.

Following ROI extraction, all images are standardized for CNN input. Specifically, each ROI is resized to 224×224 pixels, converted to three-channel RGB, and normalized to the [0, 1] range. Data augmentation is then applied during training to increase variability and reduce overfitting. To address the issue of class imbalance, we applied patient-level stratified undersampling with a fixed random seed to generate reproducible balanced splits. A paired 5-fold cross-validation design was then employed, ensuring that ResNet-50 and EfficientNet-B0 were trained and evaluated on identical dataset partitions for a fair comparison.

In the feature extraction stage, two deep CNN backbones were considered: ResNet-50 and EfficientNet-B0, both pretrained-on ImageNet. ResNet-50 employs residual connections to enable stable training of deeper architectures, while EfficientNet-B0 leverages compound scaling to balance depth, width, and resolution. For both backbones, the original fully connected classification layers were removed and replaced with a new dense layer followed by a SoftMax function to adapt to the binary classification tasks.

The classification stage outputs probability estimates for each class, aggregated across 5-fold cross-validation. Model performance was quantified using multiple metrics, including accuracy, sensitivity, specificity, precision, F1-score, and ROC–AUC. Results are reported as mean ± standard deviation across folds, with 95% confidence intervals for ROC–AUC estimated using the t-distribution across folds. To further ensure robustness, statistical significance of performance differences between ResNet-50 and EfficientNet-B0 was assessed using paired Student's t-tests and Wilcoxon signed-rank tests. This architecture enables a rigorous, reproducible, and statistically validated comparison between two representative CNN families across both screening-oriented (normal vs. abnormal) and diagnostic-oriented (benign vs. malignant) mammogram classification tasks. As summarized in Figure 1, this workflow explicitly combines patient-level splitting, paired cross-validation, and harmonized training configurations to control common sources of bias, which underpins the fairness-driven performance comparisons reported in the results.

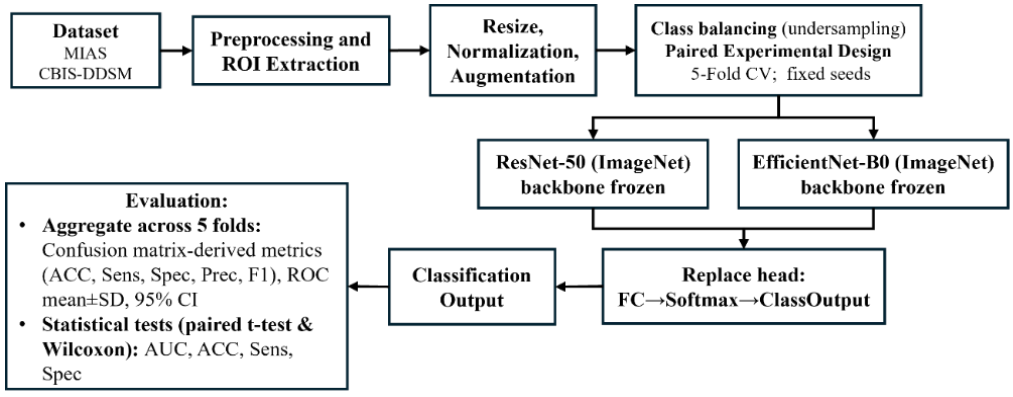


Figure 1. Workflow of the proposed mammogram classification framework using ResNet-50 and EfficientNet-B0

### 2.1. Dataset

This study employs two widely used public mammogram datasets. The MIAS database [31] provides low-resolution mammograms and was used for the normal vs. abnormal classification scenario,

comprising 107 normal and 89 abnormal cases (39 benign, 50 malignant). In contrast, the CBIS-DDSM dataset [32] offers high-resolution mammograms and was used for the benign vs. malignant classification scenario, containing 481 benign and 527 malignant cases. These datasets represent complementary benchmarks, supporting both screening-oriented and diagnostic-oriented evaluation tasks.

### 2.2. Preprocessing and ROI extraction

To ensure consistency and comparability across datasets, we adopted a standardized pre-processing protocol adapted from prior mammography CAD studies [18], [19]. As illustrated in Figure 2, all mammogram images were first converted to 8-bit grayscale, harmonizing intensity scales between formats. Noise suppression was performed using a 3×3 median filter, which effectively reduces salt-and-pepper noise without blurring lesion boundaries. For breast region segmentation, global thresholding (intensity level fixed at 0.08 after median filtering) was applied to separate foreground tissue from the background. The resulting binary mask was refined through hole filling, removal of small spurious components, and morphological opening with a disk structuring element. The largest connected component touching the image border was then identified, and its bounding box was used to automatically crop the breast region, producing the ROI–breast area. This procedure is critical for the normal vs. abnormal classification scenario, as it ensures that only relevant breast tissue is analyzed while excluding background artifacts.

Because non-breast structures, particularly the pectoral muscle, may interfere with lesion analysis, they were removed using an edge- and line-based detection strategy. Canny edge detection followed by a Hough transform constrained to oblique orientations (20°–70°) was applied in the upper corner region; the detected line defined a triangular mask covering the pectoral area, which was slightly dilated to ensure full suppression. This masking step was applied regardless of projection to minimize the chance of bright non-parenchymal structures being misinterpreted as suspicious masses. For mass segmentation within the cropped ROI, a histogram-guided thresholding approach was used to highlight dense regions. The threshold was derived adaptively from the dominant peak in the intensity distribution, followed by morphological refinement (dilation, erosion, hole filling, and area opening). Candidate blobs were further filtered using shape descriptors (circularity, solidity, extent) and location rules to reject elongated or corner-touching artifacts. The surviving blobs were enclosed by bounding boxes and cropped to produce ROI–Mass patches, which served as the input for the benign vs. malignant classification task.

In summary, the pre-processing pipeline combines classical image processing techniques—median filtering, thresholding, morphological analysis, and edge-based masking—with task-specific ROI extraction. Following prior works [18], [19], we tailored the ROI definition to the classification scenario: breast area ROIs for normal vs. abnormal and mass ROIs for benign vs. malignant. This tailored ROI strategy, summarized in Figure 2, concentrates learning on clinically relevant breast tissue and contributes to the stable ROC–AUC values reported in Table 1, particularly for the more challenging benign vs. malignant task.

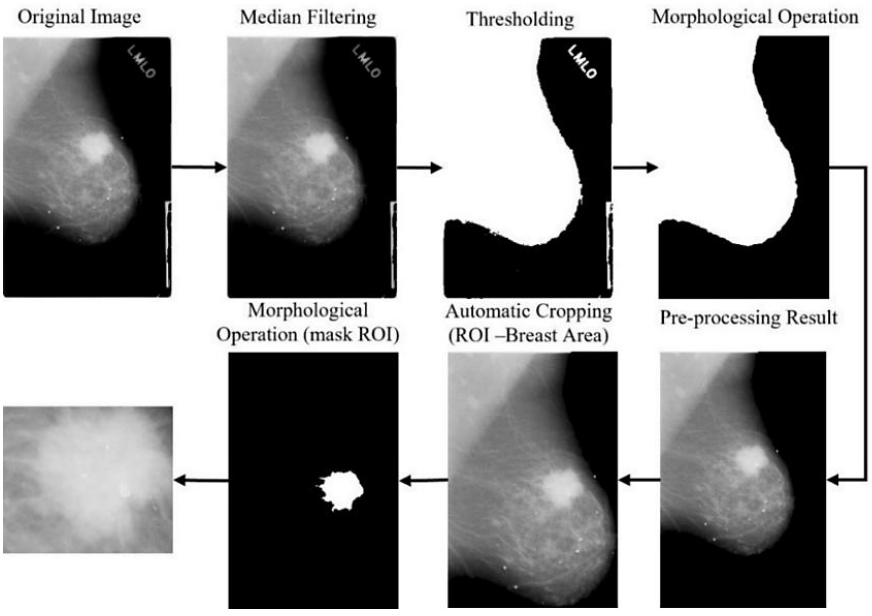


Figure 2. Pre-processing pipeline for mammogram ROI extraction

Table 1. Cross-validation performance (mean ± SD; 95% CI)

| Task | Model | LR | Ep | ACC | Sens | Spec | Prec | F1 | AUC (mean ± SD) | 95% CI (AUC) |
|---|---|---|---|---|---|---|---|---|---|---|
| Normal vs. Abnormal (MIAS) | ResNet-50 | 0.0001 | 30 | 0.893 ± 0.030 | 0.922 ± 0.049 | 0.865 ± 0.051 | 0.874 ± 0.042 | 0.896 ± 0.029 | 0.959 ± 0.027 | 0.925–0.993 |
| | **EfficientNet-B0** | **0.001** | **30** | **0.899 ± 0.031** | **0.888 ± 0.068** | **0.909 ± 0.098** | **0.918 ± 0.080** | **0.899 ± 0.029** | **0.981 ± 0.017** | **0.960–1.000** |
| Benign vs. Malignant (CBIS-DDSM) | **ResNet-50** | **0.001** | **15** | **0.833 ± 0.023** | **0.838 ± 0.068** | **0.827 ± 0.040** | **0.831 ± 0.025** | **0.833 ± 0.029** | **0.915 ± 0.013** | **0.899–0.931** |
| | EfficientNet-B0 | 0.0001 | 30 | 0.803 ± 0.019 | 0.838 ± 0.046 | 0.767 ± 0.052 | 0.784 ± 0.034 | 0.809 ± 0.019 | 0.902 ± 0.016 | 0.882–0.922 |

Note: 95% CIs via t-distribution across folds (paired 5-fold CV). LR = Learning Rate; Ep = Epochs; ACC = Accuracy; Sens = Sensitivity; Spec = Specificity; Prec = Precision; F1 = F1-score; AUC = ROC–AUC

### 2.3. Augmentation

After ROI extraction, the images were standardized for CNN input, as shown in Figure 3. All ROIs were resized to 224×224 pixels to match the input resolution required by ResNet-50 and EfficientNet-B0, both pretrained-on ImageNet. Pixel intensity normalization was applied by converting grayscale images to three-channel RGB and scaling pixel values to the [0, 1] range, thereby aligning the input distribution with that expected by the pretrained weights and facilitating effective transfer learning.

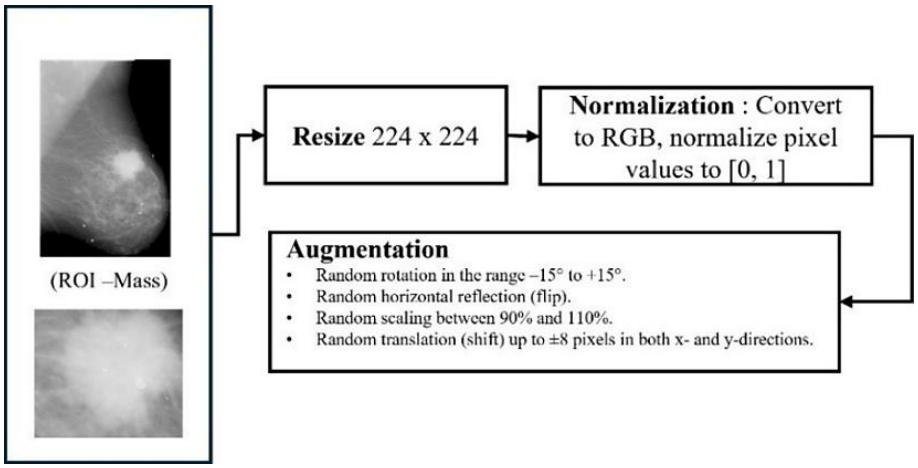


Figure 3. ROI standardization and augmentation workflow

To further improve robustness and reduce overfitting, data augmentation was incorporated during training. The applied transformations included random rotations within ±15°, random horizontal reflections (flips), random scaling between 90% and 110%, and random translations up to ±8 pixels in both horizontal and vertical directions. These augmentations simulate realistic variability in breast positioning and acquisition settings, thereby enriching the diversity of training samples without altering the underlying pathology. Similar augmentation strategies have been successfully employed in recent mammography CAD studies [6], [7].

This comprehensive augmentation pipeline ensured that the CNN models received consistent, high-quality inputs across both datasets and classification scenarios, enabling a fair and reproducible comparison between ResNet-50 and EfficientNet-B0. In practice, this standardized input and augmentation strategy supported the relatively narrow confidence intervals and consistent fold-wise performance observed in Table 1, especially on the smaller MIAS dataset.

### 2.4. Class balancing and experimental design

To address class imbalance in both MIAS (normal vs. abnormal) and CBIS-DDSM (benign vs. malignant), we applied patient-level stratified undersampling with a fixed random seed to generate reproducible, balanced splits. Model evaluation followed a paired 5-fold cross-validation protocol, in which ResNet-50 and EfficientNet-B0 shared identical train–test partitions for a fair, head-to-head comparison. Each fold comprised four partitions for training and one for testing, ensuring that every case was evaluated

exactly once in the test role. This paired design reduced sampling bias and allowed direct statistical comparison between the two backbones.

Model performance was reported as mean ± standard deviation across folds, with 95% confidence intervals for ROC–AUC estimated using the t-distribution. This experimental setup ensured reproducibility and provided statistically robust conclusions for both screening-oriented (normal vs. abnormal) and diagnostic-oriented (benign vs. malignant) classification tasks.

### 2.5. Model architectures

This study compares two widely used CNN backbones, ResNet-50 and EfficientNet-B0, pretrained on ImageNet and adapted for binary mammogram classification. To ensure a controlled and fair comparison, both backbones are used as fixed feature extractors: the pretrained weights are frozen, and a single, identical, minimal classifier head is attached to each backbone. This head maps extracted features to two output classes and is followed by a SoftMax to produce probability estimates. By holding the backbone parameters constant and using the same lightweight head, we isolate architectural differences: residual connections in ResNet-50 versus compound scaling in EfficientNet-B0—without confounding effects from classifier complexity or tuning depth.

From a representational standpoint, ResNet-50 leverages residual connections to stabilize information flow in deeper stacks, enabling robust feature extraction even with limited data [33]. EfficientNet-B0, in contrast, applies compound scaling across depth, width, and resolution using MBConv blocks, yielding an efficient accuracy–performance trade-off [34]. Optimization was intentionally conservative to prioritize reproducibility. We employed the Adam optimizer with early stopping and explored a compact hyperparameter grid (learning rates ∈ {1e-3, 1e-4}; epochs ∈ {15, 30}). Unless otherwise noted, section 3 reports the best-performing configuration for each backbone and dataset.

For clarity, Table 2 summarizes the key architectural and training settings, highlighting the symmetry between the two pipelines—identical classifier heads, frozen backbones, shared optimizer and stopping criteria, and a compact hyperparameter search space—thereby ensuring a clean architectural comparison.

Table 2. Architectural and training settings of ResNet-50 and EfficientNet-B0 for mammogram classification

| Aspect | ResNet-50 | EfficientNet-B0 |
|---|---|---|
| Pretraining | ImageNet | ImageNet |
| Backbone role | Frozen backbone (fixed feature extractor) | Frozen backbone (fixed feature extractor) |
| Classifier head | Minimal linear layer → SoftMax (2 classes) | Minimal linear layer → SoftMax (2 classes) |
| Key architectural trait | Residual connections for stable deep training | Compound scaling with MBConv blocks |
| Optimizer and control | Adam; early stopping (best validation loss retained) | Adam; early stopping (best validation loss retained) |
| Hyperparameters tested | Learning rates {1e-3, 1e-4}; epochs {15, 30} | Learning rates {1e-3, 1e-4}; epochs {15, 30} |

### 2.6. Evaluation metrics

Model performance was assessed using a comprehensive set of standard classification metrics widely adopted in medical imaging. These included overall accuracy (ACC), which measures the proportion of correctly classified samples, sensitivity (recall), which quantifies the ability to correctly identify positive cases such as abnormal or malignant lesions, specificity (Spec), which captures the ability to correctly identify negative cases (normal or benign), precision (Prec), which measures the proportion of true positives among all predicted positives, and the F1-score, defined as the harmonic mean of precision and recall.

In addition, the receiver operating characteristic area under the curve (ROC–AUC) was computed from predicted class probabilities to quantify ranking performance (*i.e.*, the probability that a randomly chosen positive sample is ranked higher than a negative one). ROC curves were generated, and mean ROC–AUC values are reported together with 95% confidence intervals (CI) across folds using the t-distribution. All other metrics (ACC, Sens, Spec, Prec, and F1) were derived from fold-wise confusion matrices and are summarized as mean ± standard deviation across folds. This combination of metrics ensured a comprehensive evaluation of both overall classification accuracy and clinically critical performance dimensions, with particular emphasis on sensitivity for screening scenarios and specificity for diagnostic contexts.

### 2.7. Statistical analysis

To verify whether the performance differences between ResNet-50 and EfficientNet-B0 were statistically meaningful, we performed paired hypothesis testing at the fold level, ensuring direct comparison across identical data partitions. For each learning-rate × epoch configuration in the compact grid, fold-wise

metric values (accuracy, F1-score, sensitivity, specificity, and ROC–AUC) from both models were compared using a paired Student's t-test; to accommodate potential non-normality, a Wilcoxon signed-rank test was also applied. Statistical significance was set at $p < 0.05$. In addition to reporting mean ± standard deviation, 95% confidence intervals for ROC–AUC were computed using the t-distribution across folds. This dual-testing framework ensures that performance differences are statistically validated rather than purely descriptive.

## 3. RESULTS AND DISCUSSION

### 3.1. Cross-validation performance

All results reported here were obtained under the standardized augmentation and patient-level stratified undersampling protocol described in section 2, with paired 5-fold cross-validation ensuring identical partitions across backbones. Among the tested configurations (LR ∈ {1e-3, 1e-4}, epochs ∈ {15, 30}), we report the best-performing setup for each backbone and dataset to provide a best-case comparison and ensure concise presentation of results. Table 1 presents the cross-validation results of ResNet-50 and EfficientNet-B0 across the two benchmark tasks: normal vs. abnormal (MIAS) and benign vs. malignant (CBIS-DDSM). On MIAS, EfficientNet-B0 achieved a mean ROC–AUC of $0.981 \pm 0.017$ (95% CI: 0.960−1.000), higher than $0.959 \pm 0.027$ (95% CI: 0.925−0.993) for ResNet-50. Beyond ROC–AUC, EfficientNet-B0 also outperformed ResNet-50 in accuracy (0.899 vs. 0.893), specificity (0.909 vs. 0.865), precision (0.918 vs. 0.874), and F1-score (0.899 vs. 0.896). ResNet-50 showed slightly higher mean sensitivity in its best configuration (0.922 vs. 0.888), although this difference was not statistically significant.

As shown in Figures 4(a) and 4(b), AUC varies systematically with the LR–epoch combination: EfficientNet-B0 benefits from longer training at a higher LR (1e-3, 30 epochs), reaching $0.981 \pm 0.017$, whereas ResNet-50 peaks at a lower LR with longer training (1e-4, 30 epochs) at $0.959 \pm 0.027$. The larger error bars for some EfficientNet-B0 settings indicate higher fold-to-fold variability, but its peak configuration still outperforms ResNet-50 on MIAS, aligning with Table 1. Figures 5(a) and 5(b) show the ROC curves for the best configurations.

For CBIS-DDSM, performance margins were narrower. ResNet-50 obtained a mean ROC–AUC of $0.915 \pm 0.013$ (95% CI: 0.899−0.931), slightly above EfficientNet-B0 at $0.902 \pm 0.016$ (95% CI: 0.882–0.922), with equal sensitivity (both 0.838). EfficientNet-B0 remained competitive on accuracy (0.803), precision (0.784), and F1-score (0.809), though trailing slightly. Figures 5(a) and 5(b) illustrate the impact of learning rate and training epochs on ROC–AUC for CBIS-DDSM: ResNet-50 is best at 1e-3, 15 epochs (longer training slightly degrades performance), while EfficientNet-B0 prefers 1e-4, 30 epochs. Figures 7(a) and 7(b) present the ROC curves for the best configurations, highlighting the relatively stable yet overlapping performance trends.

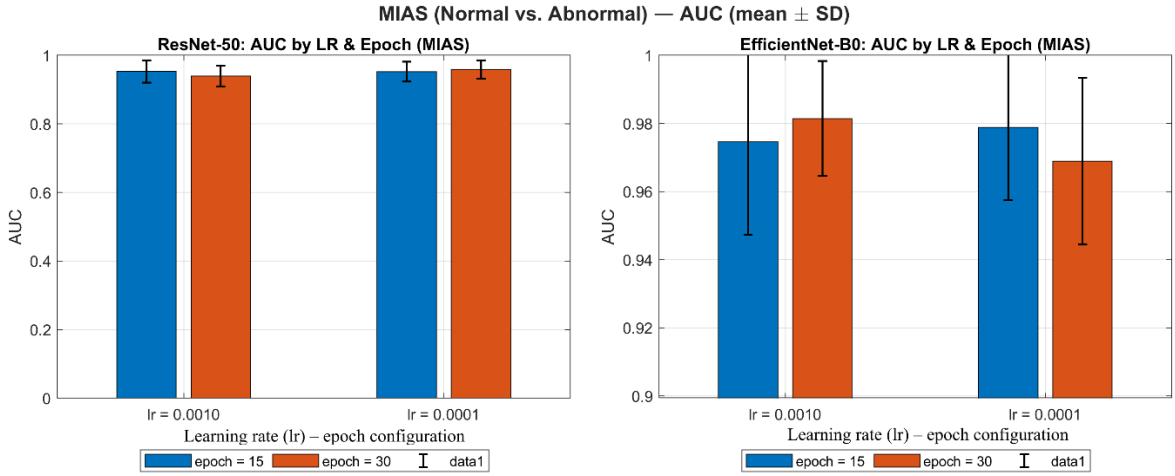


Figure 4. Comparison of mean ROC–AUC (± SD) over paired 5-fold cross-validation on MIAS (normal vs. abnormal) for ResNet-50 (left) and EfficientNet-B0 (right) under four hyperparameter settings: LR ∈ {1e-3, 1e-4} and epochs ∈ {15, 30}

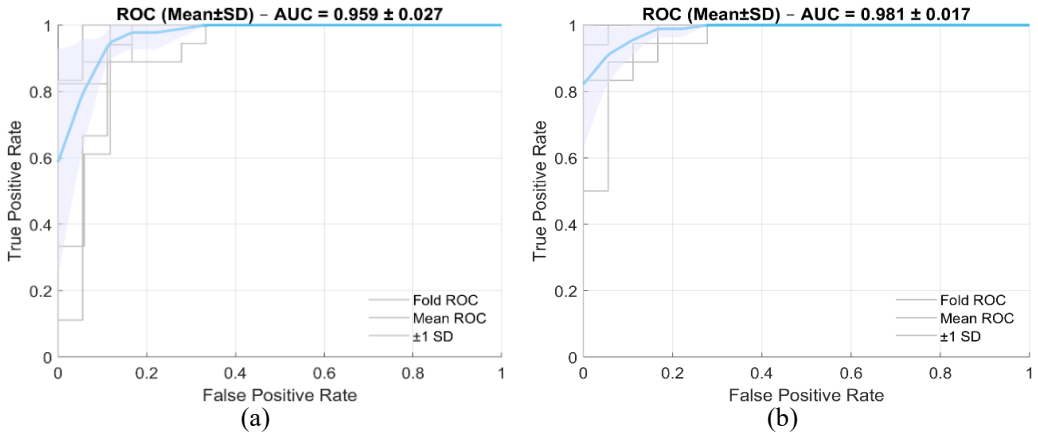


Figure 5. ROC curves (mean ± SD; paired, patient-level 5-fold CV) for the best-performing configurations on MIAS (normal vs. abnormal): (a) ResNet-50, LR = 1e-4, 30 epochs; (b) EfficientNet-B0, LR = 1e-3, 30 epochs. Solid lines show the mean; shaded bands denote ±1 SD. 95% CIs for AUC are reported in Table 1

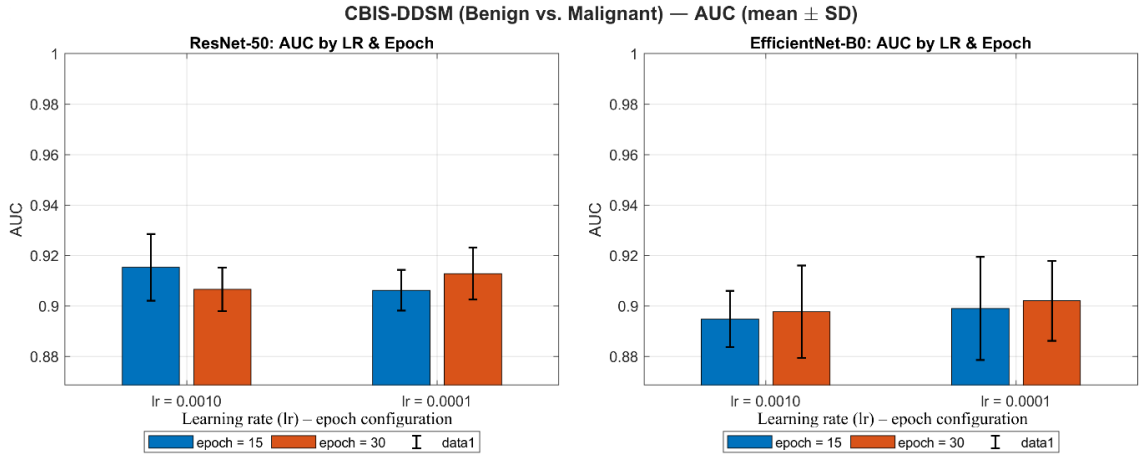


Figure 6. Comparison of mean ROC–AUC (± SD) over paired 5-fold cross-validation on CBIS-DDSM (benign vs. malignant) for ResNet-50 (left) and EfficientNet-B0 (right) under four hyperparameter settings: LR ∈ {1e-3, 1e-4} and epochs ∈ {15, 30}

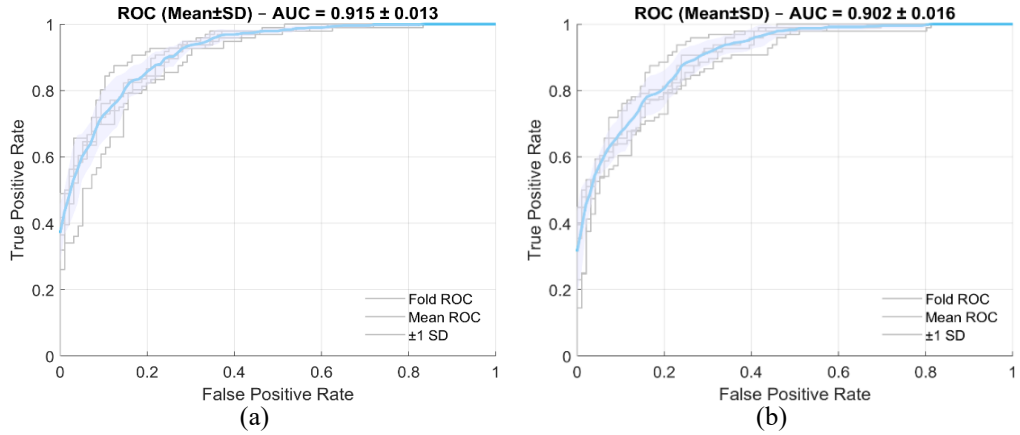


Figure 7. ROC curves (mean ± SD; paired, patient-level 5-fold CV) for the best-performing configurations on CBIS-DDSM (benign vs. malignant): (a) ResNet-50, LR = 1e-3, 15 epochs; (b) EfficientNet-B0, LR = 1e-4, 30 epochs. Solid lines show the mean; shaded bands denote ±1 SD. 95% CIs for AUC are reported in Table 1

### 3.2. Statistical significance analysis

To verify whether the performance differences between ResNet-50 and EfficientNet-B0 were statistically meaningful, we applied both paired Student's t-tests and Wilcoxon signed-rank tests across the 5-fold cross-validation. These complementary tests account for the dependency of paired folds and reduce the risk of bias caused by non-normal metric distributions. Statistical testing was performed under an identical LR–epoch setting for both backbones, selected as the best shared configuration per dataset from the hyperparameter grid in section 2.5 (see Table 2). While Table 1 reports the best-performing configuration for each backbone, Table 3 provides a controlled same-setting paired comparison (fold-wise, same-setting; thus, Δ Mean values are not derived from Table 1) to assess whether the observed differences are statistically meaningful under identical training conditions.

As summarized in Table 3, on the MIAS dataset (normal vs. abnormal), EfficientNet-B0 demonstrated small but statistically significant improvements over ResNet-50 in accuracy (Δ Mean = +0.045, p = 0.016), F1-score (Δ Mean = +0.048, p = 0.019), and ROC–AUC (Δ Mean = +0.027, p = 0.036) by paired t-test, while the Wilcoxon signed-rank test did not reach significance (p = 0.066). Sensitivity and specificity also showed positive gains (Δ Mean = +0.056 and Δ Mean = +0.033, respectively), but these differences were not statistically significant (p > 0.05). This indicates that EfficientNet-B0 provides more reliable overall discrimination on MIAS, particularly in terms of balanced performance (ACC, F1) and probability ranking (AUC).

Table 3. Statistical significance testing of ResNet-50 versus EfficientNet-B0 across cross-validation folds using paired t-test and Wilcoxon signed-rank test

| Task/Metric | Δ Mean (EffB0 – ResNet50) | t-test (p) | Wilcoxon (p) | Significance |
|---|---|---|---|---|
| Normal vs. Abnormal (MIAS) | | | | |
| Accuracy | +0.045 | 0.016 | 0.066 | p < 0.05 (t) |
| F1-score | +0.048 | 0.019 | 0.068 | p < 0.05 (t) |
| Sensitivity (Recall) | +0.056 | 0.226 | 0.144 | n.s. |
| Specificity | +0.033 | 0.426 | 0.285 | n.s. |
| ROC–AUC | +0.027 | 0.036 | 0.063 | p < 0.05 (t) |
| Benign vs. Malignant (CBIS-DDSM) | | | | |
| Accuracy | –0.033 | 0.054 | 0.063 | n.s. (borderline) |
| F1-score | –0.027 | 0.152 | 0.188 | n.s. |
| Sensitivity (Recall) | +0.002 | 0.960 | 1.000 | n.s. |
| Specificity | –0.069 | 0.073 | 0.125 | n.s. |
| ROC–AUC | –0.020 | 0.027 | 0.063 | p < 0.05 (t) |

Notes: Δ Mean = performance difference (EfficientNet-B0 – ResNet-50). Significance levels: $p < 0.05$ = significant; p < 0.01 = highly significant; n.s. = not significant. Same-setting comparison was used for statistical testing: MIAS (LR = 1e-4, Ep = 15) and CBIS-DDSM (LR = 1e-3, Ep = 15) for both backbones.

For the CBIS-DDSM dataset (benign vs. malignant), the trend reversed: ResNet-50 outperformed EfficientNet-B0 in terms of ROC–AUC (Δ Mean = –0.020, p = 0.027), indicating a statistically significant advantage by paired t-test, while the Wilcoxon signed-rank test did not reach significance (p = 0.066). Differences in accuracy (Δ Mean = –0.033, p = 0.054) and specificity (Δ Mean = –0.069, p = 0.073) favored ResNet-50 but did not reach statistical significance. Sensitivity differences were negligible (Δ = +0.002, p = 0.960). These results suggest that while EfficientNet-B0 generalizes better on the MIAS dataset, ResNet-50 remains more robust for the pathology-confirmed CBIS-DDSM task, particularly in discriminating benign from malignant cases.

Overall, the statistical analysis indicates that EfficientNet-B0 yields gains on MIAS, making it well-suited for screening-oriented classification where balanced metrics (ACC, F1) and probability ranking (AUC) are essential, whereas ResNet-50 maintains an advantage on CBIS-DDSM, supporting its suitability for diagnostic contexts where higher AUC reliability is required.

### 3.3. Comparison with previous studies

Table 4 presents a consolidated comparison of the proposed models with previous works on breast cancer classification across different datasets and modalities. Classical machine learning studies, such as those using the Wisconsin diagnostic breast cancer (WDBC) dataset [22], [24], [25], consistently reported high accuracies in the range of 96%–99% due to the relatively small size and homogeneous tabular features. Similarly, handcrafted texture-based approaches, including the use of 2D-DWT and GLCM combined with artificial neural networks (ANN), demonstrated competitive performance on mammography datasets such as MIAS and CBIS-DDSM, achieving accuracies above 90% [18], [19].

Table 4. Comparison of the proposed models with previous studies on breast cancer classification

| Dataset | Method | Accuracy/AUC | Notes |
|---|---|---|---|
| Mammograms (DDSM) [6] | CNN (CAD system) | CNN: 94.5%; ResNet-50: 95.8%; InceptionResNet-V2: 97.5% | DL outperformed conventional CAD; improved lesion detection. |
| CBIS-DDSM, INbreast [7] | End-to-end CNN classifier | AUC: 0.91 (CBIS-DDSM), 0.98 (INbreast) | Sensitivity: 86–87%, specificity: 80–96%. |
| Histopathology (1361 images) [14] | EfficientNet (CNN) | Accuracy: 94.26% | Precision: 94%, Recall: 94%. |
| MRI (1480 scans, Kaggle) [15] | ResNet-50 | Accuracy: 92.01% | MRI classification (healthy vs. malignant). |
| Mammograms [17] | Transfer learning (ResNet-50 v2) | Accuracy: 93.4% | Fine-tuned ResNet; improved sensitivity. |
| MIAS, UDIAT [18] | 2D-DWT + GLCM + ANN | Accuracy: 93.8% (3-class) | Classical handcrafted texture + ANN. |
| Mammograms (benign vs. malignant) [19] | Texture + ANN | Accuracy: 95.83% | Texture features combined with ANN. |
| MIAS [20] | Shearlet + GLCM/GLRLM + BiLSTM-CNN | Accuracy: 97.14% | Hybrid handcrafted + deep sequential features. |
| WDBC (tabular) [22] | SVM (best model) | Accuracy: 96.25%; AUC: 0.994 | Compared multiple ML classifiers. |
| WDBC [24] | ML classifiers (SVM, RF, NB) | Accuracy: 96.5% | RF and SVM showed strongest performance. |
| Kaggle (569 samples) [25] | SVC, RF, XGBoost | Accuracy: 98.7% (XGBoost) | ML baseline with feature-based learning. |
| CBIS-DDSM [27] | Deep CNN (tuned with augmentation) | Accuracy: 99.48% | Cloud-based training; hyperparameter optimization. |
| Mammograms [29] | ResNet-50 + SMOTE | Accuracy: 99% (balanced), 90% (imbalanced) | Addressed class imbalance explicitly. |
| BreakHis (histopath) [30] | FastLeakyResNet-CIR | Accuracy: 98.94% | Outperformed ResNet18/50, Inception, VGG. |
| **Mammograms (MIAS)** | **ResNet-50** | **Accuracy: 89.3% ± 3.0; AUC: 0.959 ± 0.027** | **Normal vs. abnormal classification.** |
| **Mammograms (CBIS-DDSM)** | **ResNet-50** | **Accuracy: 83.3% ± 2.3; AUC: 0.915 ± 0.013** | **Benign vs. malignant classification.** |
| **Mammograms (MIAS)** | **EfficientNet-B0** | **Accuracy: 89.9% ± 3.1; AUC: 0.981 ± 0.017** | **Normal vs. abnormal; best ROC–AUC.** |
| **Mammograms (CBIS-DDSM)** | **EfficientNet-B0** | **Accuracy: 80.3% ± 1.9; AUC: 0.902 ± 0.016** | **Benign vs. malignant classification.** |

Notes: Results are reported from heterogeneous datasets (mammography, histopathology, MRI, WDBC). Direct comparisons are not strictly valid due to dataset complexity, but the table illustrates performance trends across classical ML, handcrafted features, and deep CNNs. This study focuses specifically on mammography (MIAS and CBIS-DDSM).

With the advent of deep learning, CNN-based approaches have significantly advanced the state of breast cancer CAD. Al-antari *et al.* [6] achieved up to 97.5% accuracy using CNN variants on the DDSM dataset, while Shen *et al.* [7] demonstrated that end-to-end CNN classifiers trained on CBIS-DDSM and INbreast yielded AUC values of 0.91 and 0.98, respectively, surpassing many handcrafted approaches. Recent studies leveraging EfficientNet architectures also achieved strong results: Tasya *et al.* [14] reported 94.26% accuracy on histopathology images, and Bhuvaneswari *et al.* [13] showed near 99% performance on mammograms using EfficientNet-B3. Furthermore, advanced ResNet variants have also been effective in other imaging modalities such as MRI and histopathology, with Das *et al.* [15] and Zeng *et al.* [30] reporting accuracies above 92% and 98%, respectively.

Compared to these works, our study provides a direct evaluation of ResNet-50 and EfficientNet-B0 on public mammography datasets (MIAS and CBIS-DDSM). On the MIAS dataset, ResNet-50 achieved 89.3% accuracy with an ROC–AUC of 0.959, while EfficientNet-B0 slightly outperformed it with 89.9% accuracy and the highest ROC–AUC of 0.981. On the more challenging CBIS-DDSM dataset (benign vs. malignant), ResNet-50 obtained 83.3% accuracy and 0.915 ROC–AUC, outperforming EfficientNet-B0, which achieved 80.3% accuracy and 0.902 ROC–AUC. These results highlight that while EfficientNet-B0 achieved higher ROC–AUC on normal vs. abnormal classification, ResNet-50 remains more robust for discriminating benign and malignant cases.

It is important to note that Table 4 includes results obtained from heterogeneous datasets such as WDBC (tabular), histopathology, MRI, and mammography. Direct comparison across these modalities is not strictly valid, as dataset size, complexity, and imaging characteristics differ substantially. In general, higher accuracies reported on WDBC and histopathology datasets often reflect lower inter-patient variability compared to mammography, which is clinically more challenging due to subtle lesion appearances and imaging noise. Nevertheless, this consolidated comparison illustrates the performance landscape across classical ML, handcrafted texture methods, and modern CNN architectures. Within this landscape, our study demonstrates

that optimized ResNet-50 and EfficientNet-B0 achieve clinically meaningful performance on standardized mammogram benchmarks, with clear trade-offs between sensitivity and specificity. These comparisons confirm that although classical ML baselines such as ANN or SVM achieve high accuracies on curated datasets, their performance lags behind modern CNN architectures under rigorous, patient-level validation.

### 3.4. Discussion

This study provides a rigorous head-to-head comparison between two widely used convolutional neural network (CNN) backbones, ResNet-50 and EfficientNet-B0, for mammogram-based breast cancer classification under two binary tasks: normal vs. abnormal (MIAS) and benign vs. malignant (CBIS-DDSM). By employing paired cross-validation and patient-level undersampling, sampling bias was eliminated, and statistical testing (under controlled same-setting) confirmed that performance differences were configuration- and dataset-dependent. Specifically, EfficientNet-B0 achieved significantly higher accuracy, F1-score, and ROC–AUC on MIAS (by paired t-test), whereas ResNet-50 demonstrated a clear advantage on CBIS-DDSM, particularly in AUC reliability. These findings highlight the complementary strengths of compound scaling and residual learning in mammographic image analysis.

When contextualized against prior research, the proposed framework remains competitive with or superior to earlier approaches. Classical texture-based methods such as DWT–GLCM with ANN reported accuracies around 90% [18], [19], while CNN-based studies achieved higher but often less reproducible results [6], [7], [24], [25]. Our results, with ROC–AUC values of 0.981 (MIAS) and 0.915 (CBIS-DDSM), fall within the upper tier of reported performances. Importantly, the adoption of strict patient-level protocols, paired folds, and dual statistical testing improves reproducibility and credibility compared to many previous works.

From a clinical perspective, EfficientNet-B0, with stronger balanced performance on MIAS, appears more suited to screening contexts where balanced metrics (accuracy/F1) and ranking performance (AUC) are critical. ResNet-50, with more stable AUC on CBIS-DDSM, is better positioned for diagnostic confirmation, where specificity helps reduce unnecessary biopsies. Together, these results suggest a tiered workflow: EfficientNet-B0 as a first-pass screener and ResNet-50 as a confirmatory reader. Beyond direct application, the study establishes a transparent benchmark framework that future CAD research can adopt for fair and reproducible model evaluation.

Limitations. Despite the methodological rigor of the proposed protocol, three main limitations remain: i) relatively small curated datasets (MIAS and CBIS-DDSM) limit generalizability, ii) absence of external multi-institutional validation across scanners and acquisition protocols, and iii) evaluation restricted to two CNN backbones. Future work will address these by expanding datasets, testing broader architectures and ensembles, and incorporating ablation studies of preprocessing modules to quantify their contribution.

## 4. CONCLUSION

This study conducted a head-to-head comparison of ResNet-50 and EfficientNet-B0 for mammogram-based breast cancer classification under a fair and reproducible evaluation protocol. The primary purpose was to provide a fairness-driven, statistically validated comparison between these two widely used CNN backbones on public mammogram benchmarks. By integrating patient-level stratified undersampling, paired cross-validation, and dual statistical validation, the work establishes a transparent benchmark framework that addresses common methodological gaps in prior studies. Results indicate that EfficientNet-B0 is well suited for screening-oriented tasks (improving accuracy/F1 and ROC–AUC in normal vs. abnormal detection under the controlled same-setting comparison), while ResNet-50 demonstrates stronger reliability in diagnostic-oriented tasks (benign vs. malignant discrimination). These complementary strengths suggest potential for a tiered CAD workflow, where EfficientNet-B0 acts as a screener and ResNet-50 as a confirmatory model.

Importantly, the methodological contributions rather than architectural novelty constitute the primary advance of this study, emphasizing reproducibility, fairness, and clinical interpretability. The proposed protocol can also be reused to benchmark additional backbones, fusion strategies, and calibration methods under patient-level, statistically validated evaluation. Future extensions will include larger-scale external validation, integration of multimodal feature fusion, and ablation analyses to further quantify pipeline contributions and to narrow the gap between experimental CAD systems and routine clinical practice.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the valuable comments and suggestions provided by the reviewers, which have significantly improved the quality of this paper. The authors also thank the curators and maintainers of the CBIS-DDSM and MIAS mammographic datasets for making these resources publicly available to the research community.

**FUNDING INFORMATION**

This work was supported by the Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM), Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, under Contract No. 8018/LL4/PG/2025.

**AUTHOR CONTRIBUTIONS STATEMENT**

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, minimize authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Muhammad Subali | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Lulu Mawaddah Wisudawati | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | | |
| Teresa | | | | ✓ | | ✓ | ✓ | ✓ | | | | | ✓ | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

**CONFLICT OF INTEREST STATEMENT**

The authors declare that there is no conflict of interest regarding the publication of this paper.

**DATA AVAILABILITY**

The mammographic datasets utilized in this study are publicly accessible. The Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) can be obtained from *The Cancer Imaging Archive* at *https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM*. The mammographic image analysis society (MIAS) database is available at *http://peipa.essex.ac.uk/info/mias.html*.

**REFERENCES**

[1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] C. Duggan *et al.*, "The Breast Health Global Initiative 2018 Global Summit on improving breast healthcare through resource-stratified phased implementation: Methods and overview," *Cancer*, vol. 126, no. S10, pp. 2339–2352, 2020, doi: 10.1002/cncr.32891.

[3] WHO, "Global Health Estimates 2024: Disease burden by cause, age, sex, by country and region," *World Health Organization*. 2024.

[4] C. J. D'Orsi, E. A. Sickles, E. B. Mendelson, and E. A. M. et al., *ACR BI-RADS® Atlas: Breast imaging reporting and data system*, 5th ed. Reston, VA, USA: Amer. Coll. Radiol., 2013.

[5] GLOBOCAN, "Global cancer observatory: cancer today," *International Journal of Cancer*, vol. 149, no. 4. pp. 778–789, 2022.

[6] M. A. Al-antari, S. M. Han, and T. S. Kim, "Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 104084, 2020, doi: 10.1016/j.cmpb.2020.105584.

[7] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-48995-4.

[8] P. Chhabra and S. Goyal, "A thorough review on deep learning neural network," in *2023 International Conference on Artificial Intelligence and Smart Communication, AISC 2023*, 2023, pp. 220–226, doi: 10.1109/AISC56616.2023.10085166.

[9] V. B. Mathema, P. Sen, S. Lamichhane, M. Orešič, and S. Khoomrung, "Deep learning facilitates multi-data type analysis and predictive biomarker discovery in cancer precision medicine," *Computational and Structural Biotechnology Journal*, vol. 21, pp. 1372–1382, 2023, doi: 10.1016/j.csbj.2023.01.043.

[10] 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 Medical Imaging*, vol. 22, no. 1, 2022, doi: 10.1186/s12880-022-00793-7.

[11] F. A. Mohammed, K. K. Tune, B. G. Assefa, M. Jett, and S. Muhie, "Medical image classifications using convolutional neural networks: A survey of current methods and statistical modeling of the literature," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 699–735, 2024, doi: 10.3390/make6010033.

[12] M. Ben Ammar, F. L. Ayechi, R. Ksantini, and H. Mahjoubi, "Pre-trained deep convolutional neural network architectures for breast cancer diagnosis in mammography: Current state-of-the-art," in *17th International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2023 - Proceedings*, 2023, pp. 1–7, doi: 10.1109/INISTA59065.2023.10310612.

[13] S. Bhuvaneswari, G. Dhivya, J. V. P. Vimal, V. V. Anusuya, and V. Kalaivani, "Revolutionizing breast cancer detection: Leveraging EfficientNetB3 for accuracy and scalability," in *3rd International Conference on Electronics and Renewable Systems,*

*ICEARS 2025 - Proceedings*, 2025, pp. 1531–1537, doi: 10.1109/ICEARS64219.2025.10941143.
[14] W. Tasya, S. Sa'Idah, B. Hidayat, and F. Nurfajar, "Breast cancer detection using convolutional neural network with EfficientNet architecture," in *APWiMob 2022 - Proceedings: 2022 IEEE Asia Pacific Conference on Wireless and Mobile*, 2022, pp. 1–6, doi: 10.1109/APWiMob56856.2022.10014095.
[15] T. Das, D. S. K. Nayak, A. Kar, L. Jena, and T. Swarnkar, "ResNet-50: The deep networks for automated breast cancer classification using MR images," in *Proceedings of 2nd International Conference on Advancements in Smart, Secure and Intelligent Computing, ASSIC 2024*, 2024, pp. 1–6, doi: 10.1109/ASSIC60049.2024.10507980.
[16] A. Ashwini, M. A. M. Suji, B. P. Prathaban, D. F. D. Shahila, J. Jency Rubia, and N. V. Kumar, "Deep learning ResNet-50 framework for automatic early breast cancer diagnosis," in *2024 International Conference on Communication, Computing and Internet of Things, IC3IoT 2024 - Proceedings*, 2024, pp. 1–5, doi: 10.1109/IC3IoT60841.2024.10550264.
[17] S. Deshnande, R. K. Patel, S. S. Chouhan, and H. Vishwakarma, "Transfer learning with ResNet-50 for enhanced mammographic breast cancer identification," in *5th International Conference on Circuits, Control, Communication and Computing, I4C 2024*, 2024, pp. 58–63, doi: 10.1109/I4C62240.2024.10748454.
[18] L. M. Wisudawati, S. Madenda, E. P. Wibowo, and A. A. Abdullah, "Feature extraction optimization with combination 2D-discrete wavelet transform and gray level co-occurrence matrix for classifying normal and abnormal breast tumors," *Modern Applied Science*, vol. 14, no. 5, p. 51, 2020, doi: 10.5539/mas.v14n5p51.
[19] L. M. Wisudawati, S. Madenda, E. P. Wibowo, and A. A. Abdullah, "Benign and malignant breast tumors classification based on texture analysis and backpropagation neural network," *Computer Optics*, vol. 45, no. 2, pp. 227–234, 2021, doi: 10.18287/2412-6179-CO-769.
[20] S. Vijayalakshmi, B. K. Pandey, D. Pandey, and M. E. Lelisho, "Innovative deep learning classifiers for breast cancer detection through hybrid feature extraction techniques," *Scientific Reports*, vol. 15, no. 1, p. 22212, 2025, doi: 10.1038/s41598-025-06669-4.
[21] M. M. Hassan *et al.*, "A comparative assessment of machine learning algorithms with the least absolute shrinkage and selection operator for breast cancer detection and prediction," *Decision Analytics Journal*, vol. 7, 2023, doi: 10.1016/j.dajour.2023.100245.
[22] V. A. Telsang and K. Hegde, "Breast cancer prediction analysis using machine learning algorithms," in *Proceedings of the 2020 IEEE International Conference on Communication, Computing and Industry 4.0, C2I4 2020*, 2020, pp. 1–4, doi: 10.1109/C2I451079.2020.9368911.
[23] A. Bharat, N. Pooja, and R. A. Reddy, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," in *2018 IEEE 3rd International Conference on Circuits, Control, Communication and Computing, I4C 2018*, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739696.
[24] S. Ara, A. Das, and A. Dey, "Malignant and benign breast cancer classification using machine learning algorithms," in *2021 International Conference on Artificial Intelligence, ICAI 2021*, 2021, pp. 97–101, doi: 10.1109/ICAI52203.2021.9445249.
[25] S. K. Mohapatra, A. Jain, Anshika, and P. Sahu, "Comparative approaches by using machine learning algorithms in breast cancer prediction," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, 2022, pp. 1874–1878, doi: 10.1109/ICACITE53722.2022.9823470.
[26] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 26731–26769, 2023, doi: 10.1007/s11042-022-14305-w.
[27] J. K. Sandhu, C. Sharma, A. Kaur, P. V. Singh Gogna, and V. Sharma, "Improving breast cancer detection with deep learning techniques: A study using CBIS-DDSM dataset," in *Proceedings - 4th International Conference on Technological Advancements in Computational Sciences, ICTACS 2024*, 2024, pp. 1703–1707, doi: 10.1109/ICTACS62700.2024.10841056.
[28] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative study of machine learning algorithms for breast cancer prediction," in *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, 2020, pp. 796–801, doi: 10.1109/ICSSIT48917.2020.9214267.
[29] 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.
[30] R. Zeng *et al.*, "FastLeakyResNet-CIR: A novel deep learning framework for breast cancer detection and classification," *IEEE Access*, vol. 12, pp. 70825–70832, 2024, doi: 10.1109/ACCESS.2024.3401729.
[31] J. Suckling *et al.*, *Mammographic image analysis society (mias) database v1. 21*. University of Cambridge Repository, 2015.
[32] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography dataset for CAD (CBIS-DDSM)," *Scientific Data*, vol. 4, 2017, doi: 10.1038/sdata.2017.177.
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[34] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning, ICML 2019*, 2019, vol. 2019-June, pp. 10691–10700.

## BIOGRAPHIES OF AUTHORS

**Muhammad Subali** received his bachelor's degree in physics from Universitas Indonesia in 1990, his master's degree in electrical engineering from Universitas Trisakti in 1997, and his doctorate from Universitas Gunadarma in 2007. He is a Head Lecturer in the Faculty of Informatics Engineering at Cendekia Abditama University, Tangerang, Banten, Indonesia. His primary research interests include signal and image processing, machine learning, artificial intelligence, Internet of Things (IoT), and deep learning. He can be contacted at: subali@uca.ac.id.

**Lulu Mawaddah Wisudawati** received her Bachelor's degree in informatics engineering from Gunadarma University, Indonesia, in 2008. She obtained a double master's degree in information systems management from Gunadarma University, Indonesia, and in computer vision from Université de Bourgogne, France, in 2013. She earned her doctoral degree in industrial technology from Gunadarma University in 2020. She is currently a lecturer and researcher in the Department of Informatics Engineering at Gunadarma University. Her research interests include medical imaging, machine learning, artificial intelligence, Internet of Things (IoT), image processing, and deep learning. She can be contacted at: lulu_mawadah@staff.gunadarma.ac.id

**Teresa** received her bachelor's degree in nursing science (PSIK) from the Faculty of Medicine, Universitas Indonesia (FKUI) in 1992 and her master's degree in leadership and nursing management from the Faculty of Nursing, Universitas Indonesia in 2020. She is currently a lecturer at the Faculty of Nursing, Universitas Cendekia Abditama (FIK-UCA), Tangerang, Banten, Indonesia, where she also serves as the Head of the Research and Community Service Unit (UPPM). Her main research interests focus on nursing management, particularly in the relationship between information technology and the improvement of healthcare service quality in hospitals. She can be contacted at: teresa@uca.ac.id.