bioengineering

MDPI

Article

Engineering the Image Representation for Deep Learning in Contrast-Enhanced Mammography: A Systematic Analysis of Preprocessing and Anatomical Masking

Roberta Fusco [1], Vincenza Granata [1,*], Paolo Vallone [1], Teresa Petrosino [1], Maria Daniela Iasevoli [1], Mauro Mattace Raso [1], Davide Pupo [1], Piero Trovato [1], Igino Simonetti [1], Paolo Pariante [1], Vincenzo Cerciello [2], Gerardo Ferrara [3], Modesta Longobucco [4], Giulia Capuano [4], Roberto Morcavallo [4], Caterina Todisco [4], Fabiana Antenucci [5], Mario Sansone [5], Daniele La Forgia [4] and Antonella Petrillo [1]

[1] Radiology Division, Istituto Nazionale Tumori-IRCCS-Fondazione G. Pascale, 80131 Naples, Italy; r.fusco@istitutotumori.na.it (R.F.); p.vallone@istitutotumori.na.it (P.V.); t.petrosino@istitutotumori.na.it (T.P.); m.iasevoli@istitutotumori.na.it (M.D.I.); m.mattaceraso@istitutotumori.na.it (M.M.R.); davide.pupo@istitutotumori.na.it (D.P.); piero.trovato@istitutotumori.na.it (P.T.); igino.simonetti@istitutotumori.na.it (I.S.); p.pariante@istitutotumori.na.it (P.P.); a.petrillo@istitutotumori.na.it (A.P.)

[2] Division of Health Physics, Istituto Nazionale Tumori-IRCCS-Fondazione G. Pascale, IRCCS di Napoli, 80131 Naples, Italy; v.cerciello@istitutotumori.na.it

[3] Pathology Division, Istituto Nazionale Tumori-IRCCS-Fondazione G. Pascale, 80131 Naples, Italy; gerardo.ferrara@istitutotumori.na.it

[4] Struttura Semplice Dipartimentale di Radiodiagnostica Senologica, IRCCS Istituto Tumori Giovanni Paolo II, Via Orazio Flacco 65, 70124 Bari, Italy; m.longobucco@oncologico.bari.it (M.L.); g.capuano@oncologico.bari.it (G.C.); r.morcavallo@oncologico.bari.it (R.M.); todisco.katia@gmail.com (C.T.); d.laforgia@oncologico.bari.it (D.L.F.)

[5] Biomedical Engineering Faculty, Università degli Studi di Napoli Federico II, 80125 Naples, Italy; f.antenucci@studenti.unina.it (F.A.); msansone@unina.it (M.S.)

* Correspondence: v.granata@istitutotumori.na.it

Abstract

Deep-learning models applied to contrast-enhanced mammography (CEM) are known to be highly sensitive to the input image representation. However, preprocessing is often treated as a secondary step and rarely analyzed as an independent design variable. In this work, we present a systematic engineering analysis of a deterministic, label-independent preprocessing pipeline for CEM images. The pipeline integrates intensity normalization, global histogram matching, local contrast enhancement, denoising, and anatomically constrained breast masking. Using a controlled experimental design, identical deep-learning architectures were trained under different input representations to isolate the impact of preprocessing on classification performance and stability. Across convolutional neural network architectures, anatomically constrained preprocessing consistently improves discrimination performance, reduces variability across cross-validation folds, and enhances training stability. Breast mask-based representations demonstrate substantial gains in AUROC and AUPRC compared to raw DICOM inputs. These findings highlight image preprocessing as a first-class engineering component in medical AI pipelines. Breast masking significantly improves robustness and generalization, independently of network architecture complexity. From a clinical perspective, improving model robustness and sensitivity to malignant lesions may contribute to more reliable AI-assisted decision support in contrast-enhanced mammography, particularly in settings characterized by acquisition variability and heterogeneous patient populations.

Check for updates

Academic Editor: John A. St. Cyr

Received: 2 February 2026
Revised: 24 February 2026
Accepted: 5 March 2026
Published: 11 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: oncology; breast; deep learning; convolutional neural networks

1. Introduction

Contrast-enhanced mammography (CEM) has emerged as a valuable imaging modality for breast cancer assessment, combining high-resolution mammographic imaging with functional information derived from iodinated contrast uptake. By enhancing lesion conspicuity and suppressing background parenchymal tissue, CEM improves sensitivity compared to conventional mammography, particularly in dense breasts [1–11].

Despite these advantages, CEM images remain affected by substantial variability related to acquisition protocols, scanner vendors, patient anatomy, and contrast dynamics. This variability poses significant challenges for automated image analysis systems, especially deep-learning (DL)-based methods, which are known to be highly sensitive to input data distribution and representation [12–28].

Recent advances in convolutional neural networks (CNNs) and hybrid CNN–Transformer architectures have demonstrated promising results in mammographic image classification and lesion detection [29–42]. However, most published studies focus on architectural innovations or performance optimization while treating image preprocessing as a fixed or auxiliary step. As a result, the role of preprocessing as an engineering design variable remains underexplored.

From an engineering perspective, preprocessing defines the representation space in which learning takes place. In medical imaging, this representation directly affects signal-to-noise ratio, anatomical consistency, and the balance between diagnostically relevant and irrelevant information. Prior studies have shown that intensity normalization, histogram-based enhancement, and contrast equalization can improve lesion visibility and downstream learning stability in mammography [32–34,43]. Nevertheless, these techniques are often evaluated in isolation or embedded within complex pipelines without systematic comparison.

Another critical yet underinvestigated aspect is the explicit incorporation of anatomical constraints. Breast masking techniques aim to suppress background regions and acquisition artifacts, allowing models to focus exclusively on anatomically meaningful content. Automated breast region segmentation tools, such as the LIBRA framework [7,8], have been widely adopted for quantitative breast imaging and density assessment, but their role as a preprocessing strategy for deep learning remains insufficiently characterized.

The objective of this study is to perform a controlled, engineering-oriented evaluation of image preprocessing strategies for deep learning applied to contrast-enhanced mammography. Rather than proposing a novel classification architecture, we focus on isolating the impact of preprocessing and anatomical masking on model performance, robustness, and stability. By training identical deep-learning models under different input representations, we aim to quantify how preprocessing choices influence downstream learning independently of network design.

The main contributions of this work are as follows:

1. The definition of a deterministic, label-independent preprocessing pipeline for CEM images, designed to harmonize intensity distributions and enhance diagnostically relevant structures.
2. A systematic comparison of raw, enhanced, and anatomically constrained image representations under identical training conditions.
3. An engineering analysis of the effects of breast masking on classification performance, training stability, and generalization.

2. Materials and Methods

This study used exclusively the publicly available CDD-CESM dataset hosted by The Cancer Imaging Archive (TCIA) [31,33]. The dataset includes 1003 recombined contrast-enhanced spectral mammography (CESM) images acquired using dual-energy techniques. For each examination, cranio-caudal (CC) and mediolateral oblique (MLO) views were obtained approximately two minutes after intravenous administration of iodinated contrast material (1.5 mL/kg). Each acquisition consists of low-energy (26–31 kVp) and high-energy (45–49 kVp) exposures, which are digitally recombined through subtraction to generate contrast-enhanced images highlighting areas of contrast uptake. The dataset provides annotated images and associated clinical information, enabling supervised learning experiments without the need for additional institutional data. All image preprocessing, anatomical breast masking, and convolutional neural network (CNN) analyses were performed centrally to ensure methodological consistency.

2.1. Study Design Overview

This study was designed as a controlled experimental analysis aimed at isolating the effect of image preprocessing on deep-learning-based classification of contrast-enhanced mammography images. All experiments were conducted using identical network architectures, training procedures, data splits, and evaluation metrics, while varying only the input image representation. The methodological workflow followed a structured pipeline consisting of image preprocessing and representation generation, deep-learning model training, and performance evaluation.

Image preprocessing included intensity normalization, contrast harmonization, local enhancement, denoising, and anatomical breast masking. All preprocessing operations were applied globally and independently of class labels to prevent data leakage and preserve the validity of the experimental comparison. All images were first converted to a uniform format and rescaled to a fixed spatial resolution. Pixel intensity normalization was performed using min–max scaling to standardize the dynamic range. Global histogram matching was applied to reduce inter-image contrast variability within the dataset. Local contrast enhancement and denoising were subsequently performed using fixed parameter settings across all images.

All processed images were resized to a unified input dimension compatible with the CNN architectures. Anatomical breast masking was applied using an automated segmentation procedure implemented consistently across the entire dataset. Preprocessing was executed centrally using identical code and parameter configurations for all images. Data splitting into training, validation, and test sets was performed after preprocessing to ensure strict separation between model development and evaluation phases.

A detailed description of each preprocessing step is provided in the following sections.

2.2. Image Preprocessing Pipeline

The image preprocessing pipeline was designed to standardize contrast-enhanced mammography images and to explicitly control the input representation provided to deep-learning models. All preprocessing steps were deterministic, label-independent, and applied uniformly to the entire dataset prior to data splitting, ensuring reproducibility and preventing information leakage.

The pipeline consists of five sequential stages: intensity normalization, global histogram matching, local contrast enhancement, denoising, and anatomical breast masking. Each component was selected based on established image processing principles and prior evidence in mammographic analysis [32–34,43].

2.2.1. Intensity Normalization and Percentile-Based Windowing

Raw CEM DICOM images exhibit wide intensity variability due to differences in the acquisition machine, breast composition, and contrast agent uptake. To mitigate the influence of extreme intensity outliers and harmonize the dynamic range across images, a percentile-based windowing strategy was adopted.

Specifically, pixel intensities were clipped between the 5th and 99.5th percentiles of the image intensity distribution, following robust normalization strategies previously applied in mammographic image processing [32]. This approach suppresses acquisition artifacts and isolated high-intensity pixels while preserving diagnostically relevant contrast-enhanced structures.

After windowing, intensities were linearly rescaled to the [0, 1] interval and converted to an 8-bit representation. All images were then resized to a fixed spatial resolution to ensure homogeneity of input dimensions across the dataset.

2.2.2. Global Histogram Matching

To further reduce inter-patient and inter-device variability, a global histogram matching (GHM) procedure was applied. Histogram-based normalization has been widely used in mammography to improve contrast consistency and reduce domain shifts between acquisitions [43].

A global reference cumulative distribution function (CDF) was computed from the entire dataset after background exclusion. Each individual image histogram was then transformed to match this reference distribution. Unlike adaptive or local normalization strategies, this global approach enforces consistent intensity statistics across all images while preserving relative contrast relationships within each image.

By standardizing intensity distributions at the dataset level, GHM reduces spurious correlations related to scanner-specific characteristics and facilitates more stable learning in downstream deep-learning models.

2.2.3. Local Contrast Enhancement

Following global normalization, local contrast enhancement was applied to improve the visibility of fine structural details and lesion boundaries. A Local Contrast Mapping (LCM) operator was employed, implemented by subtracting a Gaussian-smoothed version of the image from the normalized image, thereby enhancing high-frequency components.

This operation is conceptually similar to unsharp masking techniques and has been previously reported to improve lesion detectability in mammographic images [33]. By emphasizing local intensity variations, LCM enhances edges and subtle textural patterns that may be informative for classification.

Subsequently, Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied to further enhance micro-contrast while controlling noise amplification. CLAHE has been extensively adopted in computer-aided detection systems for mammography due to its ability to improve contrast in low-visibility regions without introducing excessive artifacts [33].

2.2.4. Denoising

To reduce noise amplification introduced by local contrast enhancement, a wavelet-based denoising strategy was applied using the denoise_wavelet function from the Scikit-image library (version 0.25.2). Denoising was performed using multilevel discrete wavelet decomposition with a Daubechies wavelet (db1), two decomposition levels, and soft thresholding. The BayesShrink method was employed for threshold estimation, with sigma rescaling enabled. The denoising step was applied after local contrast enhancement and

prior to anatomical breast masking, ensuring consistent noise reduction while preserving structural image content relevant for downstream convolutional neural network training.

2.2.5. Anatomical Breast Masking

In addition to intensity-based enhancement, anatomically constrained representation was generated by isolating the breast region from background structures and acquisition artifacts. Automated breast segmentation was performed using the LIBRA framework, a validated tool for breast region extraction released as part of CaPTk [7,8].

LIBRA employs intensity-based filtering, morphological refinement, and anatomical priors to separate breast tissue from the background, pectoral muscle, and non-anatomical elements. The algorithm was applied directly to the original CEM DICOM images to generate a binary breast mask for each mammographic view.

The resulting masks were used to create an alternative input representation, referred to as breast-mask-based images, in which all non-breast pixels were suppressed. This representation explicitly encodes anatomical constraints while discarding background information that is unlikely to contribute to lesion characterization (Figure 1).

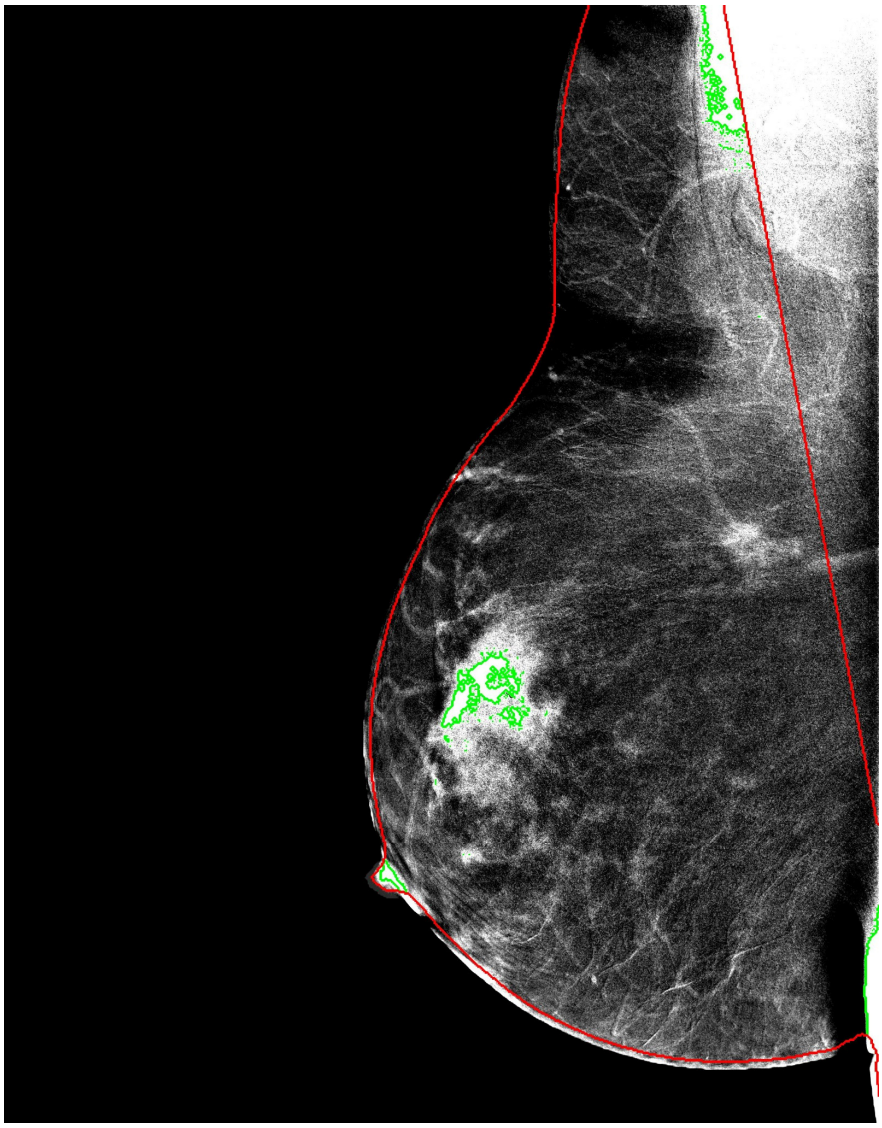


Figure 1. Example of breast segmentation using the LIBRA framework. The red contour outlines the breast boundary after background and pectoral muscle removal, while green areas indicate detected lesions within the anatomically constrained breast region.

Unlike lesion-level segmentation, breast masking does not incorporate diagnostic labels and therefore preserves independence between preprocessing and classification. By introducing an explicit anatomical prior at the input level, this strategy reduces the dimensionality of the learning problem and encourages models to focus on anatomically meaningful structures. The Figure 2 reports representative examples of the multi-stage enhancement pipeline.

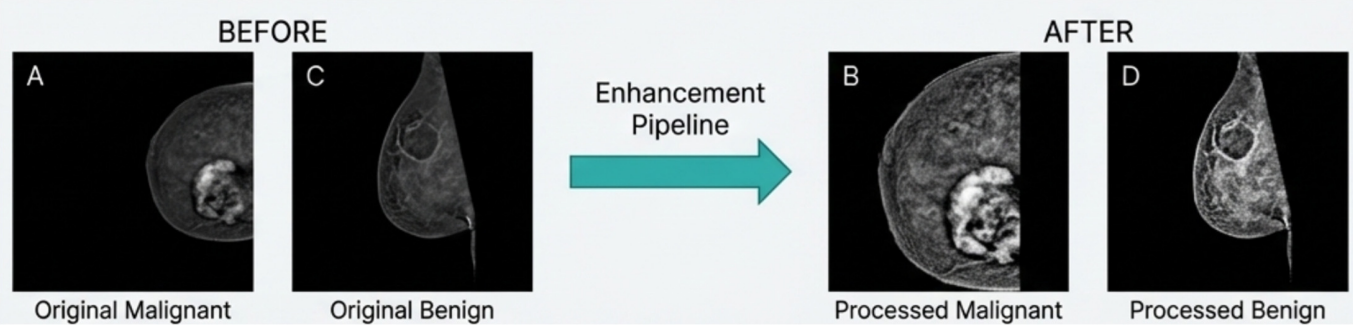


Figure 2. Representative examples of the multi-stage enhancement pipeline. Panels (**A**,**C**) show original malignant and benign CEM images, while panels (**B**,**D**) display the corresponding processed images after normalization and contrast enhancement. The pipeline enhances tissue contrast and structural visibility while preserving lesion morphology.

2.2.6. Preprocessing Integrity and Data Leakage Considerations

All preprocessing operations were implemented as deterministic and label-independent transformations. No class-conditioned statistics, outcome information, or partition-specific parameters were used at any stage of the pipeline.

Percentile-based intensity clipping was applied independently to each image based solely on its internal intensity distribution. Therefore, this step does not introduce cross-sample statistical coupling.

Global histogram matching relied on a single reference cumulative distribution function (CDF) computed across the dataset to standardize intensity statistics. Although this reference distribution was estimated prior to data partitioning, it represents a deterministic normalization transform and does not encode class-specific or patient-specific information. No preprocessing parameter was optimized based on model performance, validation results, or test outcomes.

Data partitioning into training, validation, and test subsets was performed strictly at the patient level after preprocessing, ensuring that no patient contributed images to multiple partitions. While fold-specific estimation of statistical normalization parameters could theoretically reinforce partition independence, the applied transformations are fully label-independent and do not incorporate predictive information. Consequently, meaningful performance inflation due to preprocessing-induced leakage is not expected.

The global histogram reference acts as a fixed-intensity standardization mapping and does not encode class-dependent or predictive information. Therefore, although estimated on the full dataset, it does not introduce outcome-driven leakage nor provide privileged information to the training process.

2.3. Deep-Learning Models

To avoid confounding effects related to architectural complexity, this study intentionally focused on a limited set of well-established and widely adopted convolutional neural network (CNN) architectures. The objective was not to maximize classification performance, but to evaluate how different input representations affect learning behavior and robustness when using standard and interpretable deep-learning models.

Two representative CNN architectures were selected: VGG16 and MobileNetV2. These models were chosen to span different computational regimes—high-capacity and lightweight—while remaining conceptually simple and widely used as baseline architectures in medical imaging studies [28,33,37,39]. This selection enables assessing whether the impact of anatomical breast masking persists independently of model capacity and computational complexity.

2.3.1. VGG16

VGG16 is a deep convolutional neural network characterized by a sequential architecture composed of stacked 3×3 convolutional layers followed by max-pooling operations and fully connected layers [28]. Despite its relatively large number of parameters, VGG16 remains conceptually simple and provides a strong baseline for image classification tasks.

Its homogeneous design makes VGG16 particularly suitable for studying the effects of input preprocessing, as changes in performance can be more directly attributed to differences in image representation rather than to complex architectural mechanisms.

2.3.2. MobileNetV2

MobileNetV2 is a lightweight convolutional neural network architecture designed for computational efficiency, based on depthwise separable convolutions and inverted residual blocks [39]. This design significantly reduces the number of parameters and computational cost while preserving representational capacity.

Compared to conventional deep CNNs, MobileNetV2 is particularly well suited for deployment-oriented and resource-constrained scenarios, which are increasingly relevant in medical imaging applications. Its inclusion in this study allows evaluating whether the benefits of anatomical breast masking extend to efficient architectures, thereby assessing the generality of representation-driven performance gains beyond high-capacity models.

2.4. Training Strategy

The models were trained using an identical supervised learning framework to ensure comparability across experiments. The classification task was formulated as a binary problem, distinguishing malignant from benign/negative cases. Data partitioning was performed strictly at the patient level to prevent information leakage, ensuring that all images belonging to a given patient were assigned exclusively to a single subset. No patient contributed data to more than one fold or to both training and test sets.

Training was performed using cross-entropy loss optimized with the Adam optimizer. A fixed batch size of 32 was employed across all experiments. To mitigate class imbalance, a controlled undersampling strategy was applied to the majority class, ensuring balanced mini-batches during training.

A stratified five-fold cross-validation scheme was adopted at the patient level to assess model robustness and reduce variance related to data partitioning. In each fold, approximately 80% of patients were used for training and internal validation, while 20% were held out for fold-based validation. After cross-validation, model performance was further evaluated on an independent hold-out test set defined prior to training, comprising 15% of malignant patients with balanced sampling of benign/negative cases. The test set remained completely unseen during model development and hyperparameter selection.

Model weights were initialized using ImageNet pretraining, followed by fine-tuning of all layers. Since contrast-enhanced mammography images are inherently grayscale, each preprocessed image was converted to a three-channel format by duplicating the single intensity channel across the RGB dimensions. This step ensured compatibility with convolutional neural networks pretrained on ImageNet, which expect three-channel inputs. Standard ImageNet normalization (mean = [0.485, 0.456, 0.406]; standard deviation = [0.229, 0.224, 0.225])

was subsequently applied without modification. Because all three channels contain identical intensity values, this procedure preserves the intrinsic grayscale structure while aligning the input distribution with pretrained backbone expectations.

To isolate the impact of image representation, data augmentation was intentionally restricted to mild geometric transformations, specifically horizontal flipping and small rotations. More aggressive augmentation strategies were avoided, as they could introduce artificial variability in image geometry or intensity distributions, potentially confounding the representation-level effects under investigation. By maintaining minimal augmentation, the experimental design preserved anatomical realism and ensured that observed performance differences were primarily attributable to input representation rather than synthetic variability.

2.5. Experimental Setup

Each CNN architecture was trained under two different input representations, both derived from the same preprocessing pipeline described in Section 2.2. The first representation consists of preprocessed full-field contrast-enhanced mammography images. The second representation is obtained by applying an automatically generated breast mask to the same preprocessed images, thereby retaining only breast tissue and suppressing background regions.

All preprocessing steps, including intensity normalization, contrast enhancement, and denoising, were identical in both conditions.

All other experimental parameters—including network architecture, optimizer, learning rate, batch size, and training schedule—were kept constant. This design ensures that observed performance differences can be attributed exclusively to the effect of anatomical constraint in the input representation.

CNN Training Configuration and Computational Environment

All experiments were implemented in Python 3.x using PyTorch (version 2.5.1+cu121) and Torchvision (version 0.20.1+cu121) libraries. Pretrained ImageNet weights were used for all backbone architectures (VGG16 and MobileNetV2). Transfer learning was implemented by initially freezing all convolutional backbone layers and setting their parameters to non-trainable during a five-epoch warm-up phase, during which only the newly initialized final fully connected classification layer was optimized. This strategy allowed the classification head to adapt to the CEM domain while preserving pretrained feature representations. After the warm-up phase, all backbone layers were unfrozen by enabling gradient computation across the entire network, and fine-tuning was performed jointly on all layers using differential learning rates to allow gradual adaptation of pretrained filters.

Optimization was performed using the AdamW optimizer with separate learning rates for backbone and classification layers. The learning rate for the classification head was set to 1×10^{-4}, while the backbone learning rate was set to 3×10^{-5}, with a weight decay coefficient of 1×10^{-4}. Learning rate scheduling was implemented using a Cosine Annealing scheduler with T_max = 20 epochs. Training was conducted for a maximum of 30 epochs per fold, with early stopping applied based on validation performance using a patience of 6 epochs.

The loss function was Cross-Entropy Loss without class weighting. Class imbalance at the case level was addressed through moderate undersampling of the majority class with a ratio of 2.0. All input images were resized to 224×224 pixels. During training, light data augmentation was applied, including random rotations within ± 10 degrees, random affine transformations with translation up to $\pm 5\%$, scaling between 0.9 and 1.1, shear up to ± 8 degrees, and horizontal flipping with probability 0.5.

Model evaluation was performed using five-fold stratified cross-validation at the patient level, followed by assessment on an independent hold-out test set comprising 15% of malignant patients with balanced benign/negative sampling.

The optimal classification threshold was determined using the Youden index computed on validation predictions within each fold of the five-fold cross-validation. To obtain a single operating point for final evaluation, validation probabilities from all folds were aggregated, and a global threshold was estimated from the combined validation outputs. This threshold was then applied unchanged to the independent hold-out test set. The test data were not used at any stage for threshold optimization.

All numerical operations were performed using NumPy (2.2.6), data handling with Pandas (2.3.3), performance metrics with Scikit-learn (1.7.2), image preprocessing with Scikit-image (0.25.2) and OpenCV (4.12.0.88), DICOM handling with pydicom (3.0.1), EfficientNet implementation via efficientnet_pytorch (0.7.1), visualization using Matplotlib (3.10.7), and progress monitoring with tqdm (4.67.1). GPU acceleration was enabled using CUDA 12.1.

Experiments were conducted on a 64-bit Windows system (build 10.0.22631) equipped with an Intel® Core™ i9-12900 CPU (2.40 GHz), 64 GB RAM, and 6 GB GPU memory. To ensure reproducibility, a fixed random seed (42) was applied to Python, NumPy, and PyTorch operations, and deterministic cuDNN settings were enabled across all runs.

2.6. Evaluation Metrics

Model performance was evaluated using a combination of threshold-dependent and threshold-independent metrics commonly recommended for imbalanced binary classification tasks in medical imaging [31,33]. The primary evaluation metric was the Area Under the Receiver Operating Characteristic Curve (AUROC), which quantifies the model's ability to discriminate between classes independently of the decision threshold. Additionally, the Area Under the Precision–Recall Curve (AUPRC) was reported to provide a more informative assessment under class imbalance.

During cross-validation, AUROC and AUPRC were computed independently for each fold using the predicted probabilities of the validation subset and subsequently averaged across the five folds to estimate robustness and variance. For the independent hold-out test set, AUROC and AUPRC were computed on merged predictions obtained from the final trained model across the entire test cohort, without fold averaging.

To evaluate classification performance at a fixed operating point, Balanced Accuracy—defined as the average of sensitivity and specificity—was computed. Sensitivity (recall for malignant cases) and specificity were also reported separately to characterize error asymmetry.

3. Results

When trained on original images, VGG16 achieved an AUROC of 0.815 on the independent test set (Table 1). The confusion matrix analysis shows a specificity of 0.812 and a sensitivity for malignant cases of 0.608, resulting in an overall accuracy and balanced accuracy of 0.705. Training histories across the five cross-validation folds exhibit a consistent pattern: training loss decreases monotonically, whereas validation loss shows oscillations and late increases. Similarly, training accuracy steadily increases above 0.90, while validation accuracy stabilizes around 0.70–0.75, widening the gap between the two curves. This behavior indicates progressive overfitting and limited generalization when learning directly from original images.

Table 1. Performance metrics obtained on the independent hold-out test set for VGG16 and MobileNetV2 under original and breast-mask-based input representations. Accuracy represents the overall proportion of correctly classified cases. AUROC (Area Under the Receiver Operating Characteristic Curve) quantifies threshold-independent discrimination capability. AUPRC (Area Under the Precision–Recall Curve) provides a complementary evaluation under class imbalance conditions. Recall (Malignant) corresponds to sensitivity for malignant cases, while specificity reflects the correct identification of benign/negative cases. Balanced Accuracy is defined as the arithmetic mean of sensitivity and specificity and was used to account for class imbalance.

Model	Input Representation	Accuracy (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	Recall–Malignant (95% CI)	Specificity (95% CI)	Balanced Accuracy (95% CI)
VGG16	Original images	0.705 (0.612–0.789)	0.815 (0.742–0.881)	0.782 (0.704–0.852)	0.608 (0.487–0.719)	0.812 (0.706–0.892)	0.705 (0.622–0.782)
VGG16	Breast mask images	0.774 (0.689–0.848)	0.872 (0.812–0.924)	0.891 (0.834–0.939)	0.726 (0.612–0.823)	0.821 (0.719–0.899)	0.774 (0.694–0.847)
MobileNetV2	Original images	0.651 (0.561–0.733)	0.732 (0.658–0.806)	0.718 (0.639–0.794)	0.602 (0.488–0.712)	0.739 (0.631–0.829)	0.651 (0.569–0.728)
MobileNetV2	Breast mask images	0.712 (0.624–0.791)	0.788 (0.716–0.857)	0.771 (0.694–0.842)	0.641 (0.528–0.748)	0.793 (0.689–0.874)	0.712 (0.631–0.787)

For MobileNetV2 trained on original images, ROC analysis on the test set shows an AUROC of 0.732. The corresponding confusion matrix yields a sensitivity of 0.602 for malignant cases and a specificity of 0.739, resulting in an accuracy and balanced accuracy of 0.651. These results indicate moderate discriminative performance, comparable to VGG16, with relatively higher specificity than sensitivity. This pattern suggests that background structures and acquisition-related variability in original images negatively affect generalization, particularly for lightweight architectures.

When the same architectures were trained on breast-mask-based images, a systematic improvement in performance and training stability was observed for both networks (Table 1). For VGG16, breast masking led to a clear increase in discriminative performance. The model achieved an AUROC of 0.872 and a balanced accuracy of 0.774 on the test set. Sensitivity for malignant cases increased to 0.726, while specificity remained high at 0.821. Training curves across all five cross-validation folds show more stable convergence, with reduced divergence between training and validation loss and a smaller gap between training and validation accuracy. This behavior indicates improved generalization compared to training on original images.

Similarly, MobileNetV2 benefited from breast masking. When trained on breast-mask-based images, the model achieved an AUROC of 0.788 and a balanced accuracy of 0.712, with sensitivity for malignant cases increasing to 0.641 while specificity improved to 0.793.

To assess whether these improvements were statistically significant, paired comparisons were conducted between the original and breast-mask-based conditions on the independent test set. AUROC values were compared using DeLong's test for correlated ROC curves, and classification outcomes were evaluated using McNemar's test on matched predictions. In addition, fold-level AUROC and balanced accuracy were compared using a paired Wilcoxon signed-rank test across the five cross-validation folds. These analyses confirmed that breast masking significantly improved discrimination performance and reduced misclassification rates ($p < 0.05$ for both architectures).

Overall, these findings confirm that anatomical constraint improves learning behavior for both high-capacity and lightweight models, reducing the negative impact of background regions and enhancing generalization performance when training exclusively on the public CDD-CESM dataset.

Across both architectures, breast-mask-based images resulted in more consistent training dynamics across cross-validation folds, as illustrated in Figures 3 and 4. For models trained on original images, validation loss and accuracy curves exhibit larger inter-fold variability and clear signs of overfitting, with an increasing divergence between training and validation performance (Figures 3A and 4A).

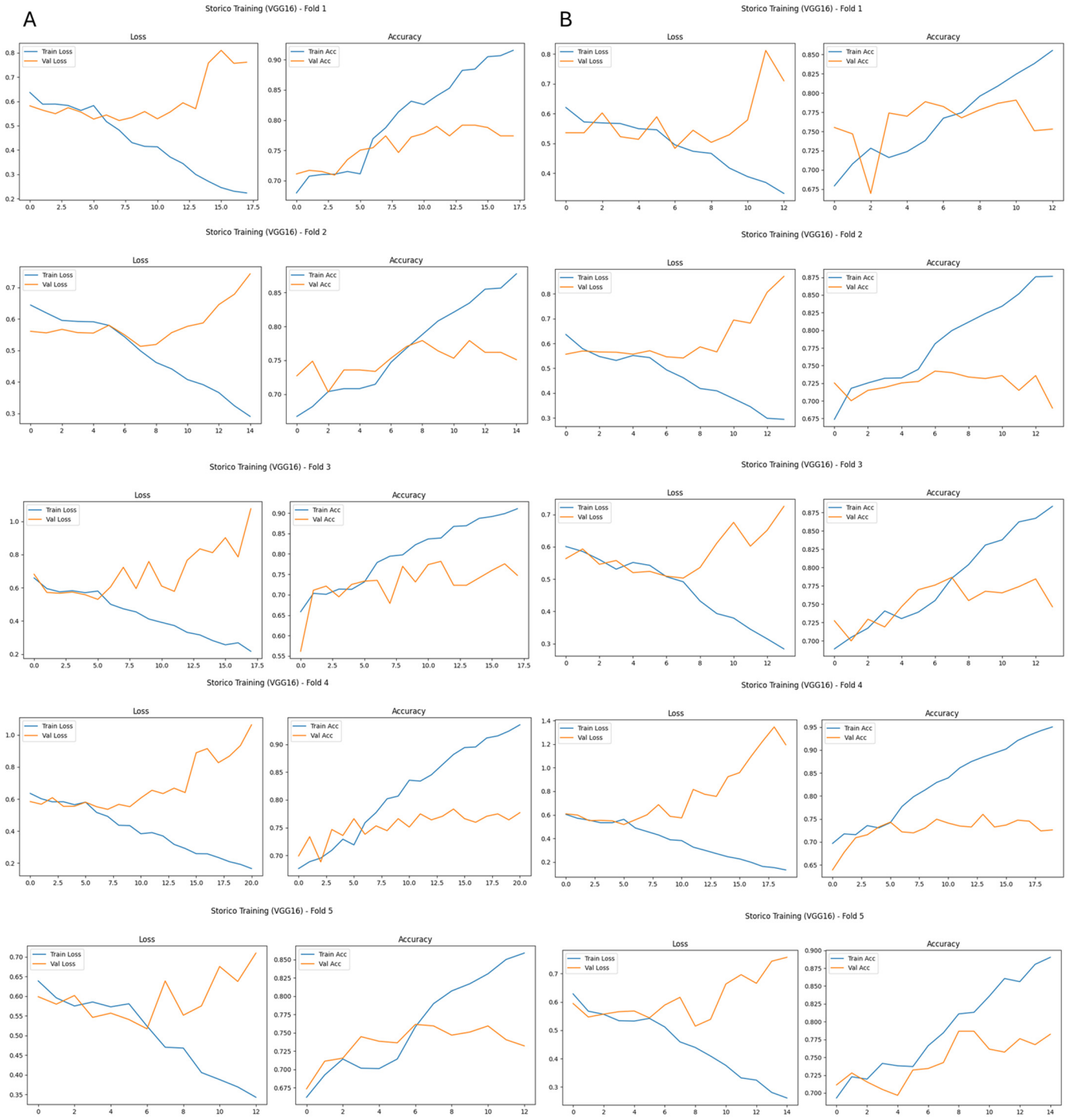


Figure 3. Training history of the VGG16 model across five cross-validation folds. (**A**) Loss and accuracy curves using original images, showing progressive divergence between training and validation performance. (**B**) Loss and accuracy curves using breast-mask-based images, demonstrating more stable convergence and reduced overfitting.

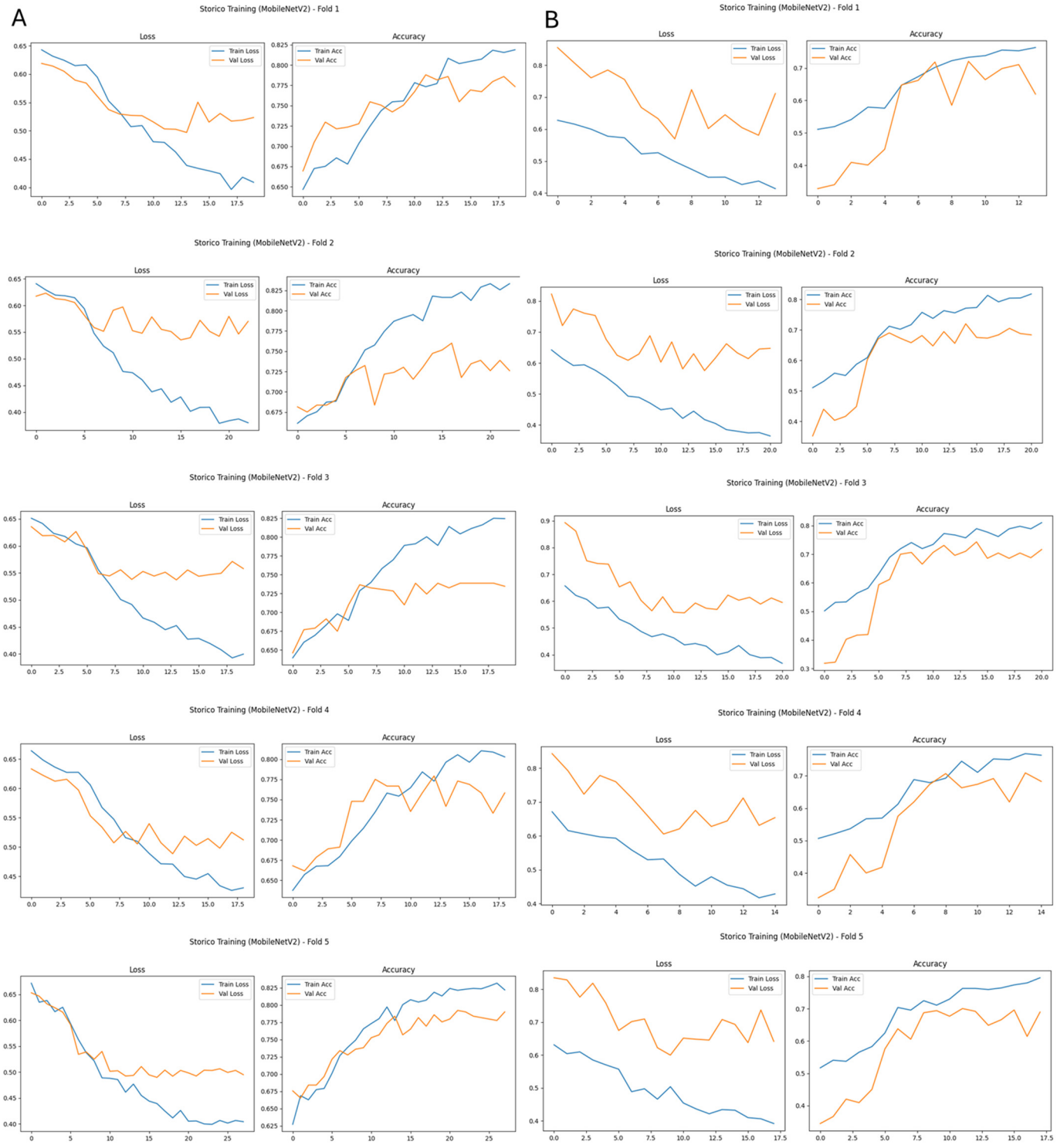


Figure 4. Training history of the MobileNetV2 model across five cross-validation folds. (**A**) Loss and accuracy curves using original images, showing variability across folds and limited generalization. (**B**) Loss and accuracy curves using breast-mask-based images, demonstrating improved stability and more consistent validation performance.

In contrast, when breast-mask-based images are used, both VGG16 and MobileNetV2 show smoother convergence behavior, reduced variance across folds, and more stable validation performance (Figures 3B and 4B). These trends indicate improved generalization and reduced sensitivity to data partitioning, highlighting the stabilizing effect of anatomical constraint on the learning process.

Overall, the results demonstrate the following:

- Training on original images leads to moderate performance and limited generalization for both VGG16 and MobileNetV2.
- Breast-mask-based representations systematically improve discrimination, sensitivity to malignant cases, and cross-fold training stability.
- The observed improvements are consistent across both a high-capacity model (VGG16) and a lightweight architecture (MobileNetV2), indicating that the effect of anatomical masking is largely independent of model complexity.

The robustness of the preprocessing framework is further supported by the cross-validation design and independent test evaluation. Since all preprocessing transformations were deterministic, label-independent, and applied uniformly across experimental conditions, the relative performance comparison between the original and breast-mask-based representations does not depend on partition-specific statistical estimation. The consistent trends observed across cross-validation folds (Figures 3 and 4) indicate that the reported improvements are not driven by partition-specific statistical artifacts.

4. Discussion

In recent years, research in medical imaging has largely focused on architectural innovation, including increasingly complex convolutional and hybrid deep-learning models [13,17,30]. While such approaches have improved benchmark performance, comparatively less attention has been devoted to deterministic preprocessing and input representation design, despite their recognized influence on learning behavior in medical imaging pipelines [14–16]. In breast imaging, background variability, acquisition-dependent artifacts, and heterogeneous tissue composition can introduce confounding information that neural networks may inadvertently exploit [7,10,23,24,44–50]. When preprocessing is not explicitly controlled, models may learn dataset-specific correlations that reduce generalization across folds or institutions, an issue particularly critical in medical imaging, where datasets are smaller and less standardized than natural image benchmarks [15].

Several studies have reported that increasing architectural complexity does not consistently translate into improved robustness or generalization in mammographic classification tasks [26,30]. In contrast, representation-level interventions—such as anatomical masking or region selection strategies—have demonstrated the ability to stabilize training and enhance discrimination performance without altering network topology [18,23,25]. Moreover, preprocessing pipelines are often underreported or insufficiently standardized, limiting reproducibility and cross-study comparability [18,23]. In many CEM studies, lesion-level cropping or ROI extraction guided by radiologist annotations is employed [18,21,24]; although effective, such strategies introduce annotation dependency and may inadvertently couple preprocessing with diagnostic labels. Additionally, preprocessing variations are frequently combined with architectural modifications or attention mechanisms [26,30], making it difficult to isolate the specific contribution of input representation. Limited emphasis has also been placed on training stability and cross-fold variability, despite the known susceptibility of medical imaging datasets to overfitting and domain variability [15].

The present study addresses these methodological gaps through a controlled, deterministic, and label-independent preprocessing framework. By maintaining identical training conditions and network architectures while modifying only the input representation, we explicitly isolate the impact of anatomical breast masking on discrimination performance and training stability.

Models trained on original contrast-enhanced mammography images achieved moderate discriminative performance but exhibited reduced sensitivity to malignant lesions and signs of overfitting, as reflected by a divergence between the training and validation curves.

This behavior aligns with previous observations showing that background structures and acquisition variability can introduce confounding signals during model training [7,9,23,24], particularly in relatively small datasets [15,20].

The introduction of anatomical breast masking produced consistent improvements across all evaluated metrics and stabilized training dynamics. Both VGG16 and MobileNetV2 demonstrated higher AUROC and AUPRC values, increased sensitivity to malignant cases, and improved balanced accuracy. Training convergence became smoother, with reduced inter-fold variability. From an engineering perspective, breast masking can be interpreted as the incorporation of an explicit anatomical prior at the input level, constraining learning to clinically meaningful regions without requiring lesion-level annotations. The consistency of performance improvements across cross-validation folds and on the independent test set further supports that the observed gains are attributable to anatomical masking rather than to partition-specific statistical effects. An alternative interpretation of the observed performance gains is that breast masking reduces confounding background signals, such as image borders, acquisition artifacts, textual markers, or peripheral noise, rather than exclusively enhancing anatomical focus. This explanation is plausible and does not contradict the proposed mechanism. From a machine-learning perspective, removing irrelevant background structures reduces spurious correlations and limits the possibility that the model exploits non-diagnostic cues. Therefore, the observed improvements may result from a combined effect of anatomical constraint and suppression of acquisition-related confounders. Both mechanisms contribute to improved representation quality and training stability.

Similar strategies have been shown to enhance robustness and reduce overfitting in breast imaging tasks [7,15,23], and suppression of irrelevant background information has previously been associated with improved classification performance in CEM and MRI studies [10,24,25,44–46].

Importantly, these improvements were observed across architectures with substantially different computational profiles. VGG16, a high-capacity convolutional network [28], and MobileNetV2, a lightweight, efficiency-oriented architecture [39], both exhibited comparable relative gains when trained on anatomically constrained images. This consistency suggests that the impact of anatomical masking is largely independent of model complexity. Prior comparative analyses have indicated that architectural sophistication alone does not guarantee improved generalization when input variability remains uncontrolled [26,30]; our findings reinforce the central role of representation design in shaping learning behavior in data-limited medical imaging settings.

Preprocessing should therefore be regarded as a core engineering component of medical AI pipelines rather than a secondary implementation detail. By improving input quality instead of increasing model complexity, it is possible to enhance robustness and reproducibility without incurring additional computational cost. This is particularly relevant for deployment-oriented models, where lightweight architectures such as MobileNetV2 are preferred for scalability and efficiency [33,39].

The deliberate restriction of this study to two well-established convolutional architectures was intended to isolate representation-level effects from architectural confounders. VGG16 provides a canonical high-capacity baseline [28], while MobileNetV2 represents an efficient, deployment-oriented design [39]. Excluding deeper residual or compound-scaled networks minimized the influence of advanced architectural mechanisms, enabling clearer attribution of observed improvements to preprocessing choices.

Several limitations warrant consideration. The analysis was confined to two CNN architectures and a binary classification task; future work should examine whether similar trends persist in lesion localization or temporal CEM analysis frameworks [40,41]. More

refined anatomical or hybrid representations integrating radiomic or functional parametric features may further enhance performance [6,12,18,21,27]. Clinical implementation will also require continued validation of automated segmentation reliability across heterogeneous acquisition settings. Finally, external validation on multi-center datasets will be essential to confirm generalizability across diverse protocols and populations [19,42].

In summary, this study demonstrates that anatomical breast masking exerts a measurable and architecture-agnostic influence on deep-learning performance in contrast-enhanced mammography. By constraining the input to anatomically relevant regions, it is possible to achieve more stable training, improved discrimination, and enhanced generalization. Although the present analysis focused on binary classification, representation-level stabilization is not inherently task-specific and may extend to multi-class or localization frameworks, where structured input constraints can facilitate feature learning and spatial focus. These findings reinforce the importance of treating preprocessing and input representation as foundational engineering decisions in medical artificial intelligence systems.

Although the results demonstrate consistent improvements across cross-validation folds and an independent hold-out test set, all experiments were conducted on a single publicly available dataset (CDD-CESM, TCIA). Therefore, conclusions regarding robustness across different acquisition protocols, scanners, or patient populations should be interpreted as preliminary. External validation on independent multi-center datasets will be necessary to confirm the generalizability of the proposed preprocessing strategy in heterogeneous clinical environments.

5. Conclusions

This study evaluated the impact of anatomical breast masking on deep-learning performance in contrast-enhanced mammography using two standard convolutional neural network architectures, VGG16 and MobileNetV2. The results show that models trained on original images achieve moderate performance and limited generalization, whereas breast-mask-based representations consistently improve discrimination, sensitivity to malignant cases, and training stability across cross-validation folds. These improvements are observed for both high-capacity and lightweight architectures, indicating that the benefit of anatomical masking is largely independent of model complexity.

Overall, the findings highlight image representation design, particularly anatomical constraint, as a key engineering component in medical AI pipelines, offering a simple, computationally neutral, and architecture-agnostic strategy to improve robustness and generalization in contrast-enhanced mammography.

Author Contributions: Conceptualization, R.F.; Methodology, R.F., V.G., P.V., T.P., M.D.I., M.M.R., D.P., P.T., I.S., P.P., V.C., G.F., M.L., G.C., R.M., C.T., F.A., M.S., D.L.F. and A.P.; Validation, V.G.; Formal analysis, R.F. and F.A.; Investigation, R.F., V.G., P.V., T.P., M.D.I., M.M.R., D.P., P.T., I.S., P.P., V.C., G.F., M.L., G.C., R.M., C.T., F.A., M.S., D.L.F. and A.P.; Writing—original draft, R.F., V.G. and A.P.; Writing—review & editing, R.F., V.G., P.V., T.P., M.D.I., M.M.R., D.P., P.T., I.S., P.P., V.C., G.F., M.L., G.C., R.M., C.T., F.A., M.S., D.L.F. and A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian Ministry of Health through the Ricerca Corrente funds.

Institutional Review Board Statement: Not applicable because the study performed the analysis on publicly available CDD-CESM dataset hosted by The Cancer Imaging Archive (TCIA) https://www.cancerimagingarchive.net/collection/cdd-cesm/ (accessed on 1 February 2026).

Informed Consent Statement: Not applicable because the study performed the analysis on publicly available CDD-CESM dataset hosted by The Cancer Imaging Archive (TCIA) https://www.cancerimagingarchive.net/collection/cdd-cesm/ (accessed on 1 February 2026).

Data Availability Statement: The raw data for data analysis was uploaded by https://www.cancerimagingarchive.net/collection/cdd-cesm/ (accessed on 1 February 2026). Results findings are reported in the manuscript.

Acknowledgments: The authors are grateful to Alessandra Trocino, librarian of the National Cancer Institute of Naples, Italy. Moreover, the authors are grateful to Paola Gargiulo and Giuditta Giannotti for their research support. Rosa Briante, Serena Santucci, Martina Totaro, and Martina Tarallo are scholarship holders of the Radiology Division at the National Cancer Institute of Naples, Italy.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. AIOM. *I Numeri del Cancro in Italia*; AIOM: Milan, Italy, 2020.
2. American Cancer Society. *Breast Cancer Facts Figures 2019–2020*; American Cancer Society Inc.: Atlanta, GA, USA, 2019. Available online: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf (accessed on 1 February 2026).
3. Schwab, F.D.; Huang, D.J.; Schmid, S.M.; Schötzau, A.; Güth, U. Self-detection and clinical breast examination: Comparison of the two "classical" physical examination methods for the diagnosis of breast cancer. *Breast* **2015**, *24*, 90–92. [CrossRef]
4. Heywang-Köbrunner, S.; Viehweg, P.; Heinig, A.; Küchler, C. Contrast-enhanced MRI of the breast: Accuracy, value, controversies, solutions. *Eur. J. Radiol.* **1997**, *24*, 94–108. [CrossRef]
5. Dessouky, B.; Elsaid, N.; Shaaban, Y. Role of contrast-enhanced digital mammography in evaluation of breast lesions. *Menoufia Med. J.* **2017**, *30*, 861.
6. Fusco, R.; Sansone, M.; Filice, S.; Granata, V.; Catalano, O.; Amato, D.M.; Di Bonito, M.; D'Aiuto, M.; Capasso, I.; Rinaldo, M.; et al. Integration of DCE-MRI and DW-MRI Quantitative Parameters for Breast Lesion Classification. *Biomed. Res. Int.* **2015**, *2015*, 237863. [CrossRef] [PubMed] [PubMed Central]
7. Keller, B.M.; Nathan, D.L.; Wang, Y.; Zheng, Y.; Gee, J.C.; Conant, E.F.; Kontos, D. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Med. Phys.* **2012**, *39*, 4903–4917. [CrossRef]
8. Saslow, D.; Boetes, C.; Burke, W.; Harms, S.; Leach, M.O.; Lehman, C.D.; Morris, E.; Pisano, E.; Schnall, M.; Sener, S.; et al. American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography. *CA A Cancer J. Clin.* **2007**, *57*, 75–89.
9. Maglogiannis, I.; Zafiropoulos, E.; Anagnostopoulos, I. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Appl. Intell.* **2007**, *30*, 24–36. [CrossRef]
10. Petrillo, A.; Fusco, R.; Di Bernardo, E.; Petrosino, T.; Barretta, M.L.; Porto, A.; Granata, V.; Di Bonito, M.; Fanizzi, A.; Massafra, R.; et al. Prediction of Breast Cancer Histological Outcome by Radiomics and Artificial Intelligence Analysis in Contrast-Enhanced Mammography. *Cancers* **2022**, *14*, 2132. [CrossRef]
11. Kim, J.J.; Kim, J.Y.; Hwangbo, L.; Suh, H.B.; Son, Y.; Nickel, M.D.; Grimm, R. Ultrafast Dynamic Contrast-Enhanced MRI Using Compressed Sensing: Associations of Early Kinetic Parameters with Prognostic Factors of Breast Cancer. *AJR Am. J. Roentgenol.* **2021**, *217*, 56–63. [CrossRef]
12. Feng, S.; Yin, J. Radiomics of dynamic contrast-enhanced magnetic resonance imaging parametric maps and apparent diffusion coefficient maps to predict Ki-67 status in breast cancer. *Front Oncol.* **2022**, *12*, 847880. [CrossRef]
13. Nicosia, L.; Pesapane, F.; Bozzini, A.C.; Latronico, A.; Rotili, A.; Ferrari, F.; Signorelli, G.; Raimondi, S.; Vignati, S.; Gaeta, A.; et al. Prediction of the Malignancy of a Breast Lesion Detected on Breast Ultrasound: Radiomics Applied to Clinical Practice. *Cancers* **2023**, *15*, 964. [CrossRef]
14. Cui, H.; Sun, Y.; Zhao, D.; Zhang, X.; Kong, H.; Hu, N.; Wang, P.; Zuo, X.; Fan, W.; Yao, Y.; et al. Radiogenomic analysis of prediction HER2 status in breast cancer by linking ultrasound radiomic feature module with biological functions. *J. Transl. Med.* **2023**, *21*, 44. [CrossRef] [PubMed]
15. Sinha, S.; Lucas-Quesada, F.A.; DeBruhl, N.D.; Sayre, J.; Farria, D.; Gorczyca, D.P.; Bassett, L.W. Multifeature analysis of Gd-enhanced MR images of breast lesions. *J. Magn. Reson. Imaging* **1997**, *7*, 1016–1026. [CrossRef] [PubMed]
16. Vomweg, T.W.; Buscema, P.M.; Kauczor, H.U.; Teifke, A.; Intraligi, M.; Terzi, S.; Heussel, C.P.; Achenbach, T.; Rieker, O.; Mayer, D.; et al. Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Med. Phys.* **2003**, *30*, 2350–2359. [CrossRef] [PubMed]
17. Sathya, D.J.; Geetha, K. Mass classification in breast DCE-MR images using an artificial neural network trained via a bee colony optimization algorithm. *Science* **2013**, *39*, 294. [CrossRef]

18. Petrillo, A.; Fusco, R.; Barretta, M.L.; Granata, V.; Mattace Raso, M.; Porto, A.; Sorgente, E.; Fanizzi, A.; Massafra, R.; Lafranceschina, M.; et al. Radiomics and artificial intelligence analysis by T2-weighted imaging and dynamic contrast-enhanced magnetic resonance imaging to predict Breast Cancer Histological Outcome. *Radiol. Med.* **2023**, *128*, 1347–1371. [CrossRef] [PubMed]
19. Petrillo, A.; Fusco, R.; Petrosino, T.; Vallone, P.; Granata, V.; Rubulotta, M.R.; Pariante, P.; Raiano, N.; Scognamiglio, G.; Fanizzi, A.; et al. A multicentric study of radiomics and artificial intelligence analysis on contrast enhanced mammography to identify different histotypes of breast cancer. *Radiol Med.* **2024**, *129*, 864–878. [CrossRef] [PubMed]
20. Degenhard, A.; Tanner, C.; Hayes, C.; Hawkes, D.J.O.; Leach, M.; The UK MRI Breast Screening Study. Comparison between radiological and artificial neural network diagnosis in clinical screening. *Physiol. Meas.* **2002**, *23*, 727–739. [CrossRef]
21. Li, C.; Song, L.; Yin, J. Intratumoral and Peritumoral Radiomics Based on Functional Parametric Maps from Breast DCE-MRI for Prediction of HER-2 and Ki-67 Status. *J. Magn. Reson. Imaging* **2021**, *54*, 703–714. [CrossRef]
22. Shin, J.K.; Kim, J.Y. Dynamic contrast-enhanced and diffusion-weighted MRI of estrogen receptor-positive invasive breast cancers: Associations between quantitative MR parameters and Ki-67 proliferation status. *J. Magn. Reson. Imaging* **2017**, *45*, 94–102. [CrossRef]
23. Fusco, R.; Piccirillo, A.; Sansone, M.; Granata, V.; Vallone, P.; Barretta, M.L.; Petrosino, T.; Siani, C.; Di Giacomo, R.; Petrillo, A.; et al. Radiomic and Artificial Intelligence Analysis with Textural Metrics, Morphological and Dynamic Perfusion Features Extracted by Dynamic Contrast-Enhanced Magnetic Resonance Imaging in the Classification of Breast Lesions. *Appl. Sci.* **2021**, *11*, 1880. [CrossRef]
24. La Forgia, D.; Fanizzi, A.; Campobasso, F.; Bellotti, R.; Didonna, V.; Lorusso, V.; Moschetta, M.; Massafra, R.; Tamborra, P.; Tangaro, S.; et al. Radiomic Analysis in Contrast-Enhanced Spectral Mammography for Predicting Breast Cancer Histological Outcome. *Diagnostics* **2020**, *10*, 708. [CrossRef] [PubMed]
25. Marino, M.A.; Leithner, D.; Sung, J.; Avendano, D.; Morris, E.A.; Pinker, K.; Jochelson, M.S. Radiomics for Tumor Characterization in Breast Cancer Patients: A Feasibility Study Comparing Contrast-Enhanced Mammography and Magnetic Resonance Imaging. *Diagnostics* **2020**, *10*, 492. [CrossRef]
26. Stefano, A.; Bini, F.; Giovagnoli, E.; Dimarco, M.; Lauciello, N.; Narbonese, D.; Pasini, G.; Marinozzi, F.; Russo, G.; D'Angelo, I. Comparative Evaluation of Machine Learning-Based Radiomics and Deep Learning for Breast Lesion Classification in Mammography. *Diagnostics* **2025**, *15*, 953. [CrossRef]
27. Leithner, D.; Horvat, J.V.; Marino, M.A.; Bernard-Davila, B.; Jochelson, M.S.; Ochoa-Albiztegui, R.E.; Martinez, D.F.; Morris, E.A.; Thakur, S.; Pinker, K. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: Initial results. *Breast Cancer Res.* **2019**, *21*, 106. [CrossRef]
28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
29. Makeev, A.; Rodal, G.; Ghammraoui, B.; Badal, A.; Glick, S.J. Exploring CNN potential in discriminating benign and malignant calcifications in conventional and dual-energy FFDM: Simulations and experimental observations. *J. Med. Imaging* **2021**, *8*, 033501. [CrossRef]
30. Available online: https://www.cancerimagingarchive.net/collection/cdd-cesm/ (accessed on 20 September 2025).
31. Khaled, R.; Helal, M.; Alfarghaly, O.; Mokhtar, O.; Elkorany, A.; El Kassas, H.; Fahmy, A. Categorized Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography images. *Cancer Imaging Arch.* **2021**, *16*. [CrossRef]
32. Rofena, A.; Guarrasi, V.; Sarli, M.; Piccolo, C.L.; Sammarra, M.; Zobel, B.B.; Soda, P. A deep learning approach for virtual contrast enhancement in Contrast Enhanced Spectral Mammography. *Comput. Med. Imaging Graph.* **2024**, *116*, 102398. [CrossRef] [PubMed]
33. Vaish, R.; Shukla, P. Deep-CNN for Disease Classification using Enhanced Mammographic Images. *Front. Biomed. Technol.* **2025**, *12*, 435–445.
34. Brahmareddy, A.; Selvan, M.P. TransBreastNet a CNN transformer hybrid deep learning framework for breast cancer subtype classification and temporal lesion progression analysis. *Sci. Rep.* **2025**, *15*, 35106. [CrossRef] [PubMed]
35. Kaddes, M.; Ayid, Y.M.; Elshewey, A.M.; Fouad, Y. Breast cancer classification based on hybrid CNN with LSTM model. *Sci. Rep.* **2025**, *15*, 4409. [CrossRef]
36. Lilhore, U.K.; Sharma, Y.K.; Shukla, B.K.; Vadlamudi, M.N.; Simaiya, S.; Alroobaea, R.; Alsafyani, M.; Baqasah, A.M. Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification. *Sci. Rep.* **2025**, *15*, 12082. [CrossRef] [PubMed] [PubMed Central]
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
38. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
39. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.

40. Ma, L.; Zhang, Y.; Liu, Y.; Wu, J. Transm: Transformer-based multiview network for breast cancer detection from mammograms. *Biomed. Signal Process. Control* **2023**, *83*, 104710.
41. Li, W.; Wang, M.; Zhao, J. Glam-net: A global-local attention-based multiview network for mammogram classification. *J. Biomed. Inform.* **2021**, *124*, 103939. [CrossRef]
42. Lyu, P.-F.; Wang, Y.; Meng, Q.-X.; Fan, P.-M.; Ma, K.; Xiao, S.; Cao, X.C.; Lin, G.-X.; Dong, S.-Y. Mapping the intellectual structure and research hotspots of artificial intelligence in cancer research: A bibliometric analysis. *Front. Oncol.* **2022**, *12*, 854742. [CrossRef] [PubMed]
43. Alshamrani, K.; Alshamrani, H.A.; Alqahtani, F.F.; Almutairi, B.S. Enhancement of Mammographic Images Using Histogram-Based Techniques for Their Classification Using CNN. *Sensors* **2022**, *23*, 235. [CrossRef] [PubMed] [PubMed Central]
44. Magni, V.; Cozzi, A.; Muscogiuri, G.; Benedek, A.; Rossini, G.; Fanizza, M.; Di Giulio, G.; Sardanelli, F. Background parenchymal enhancement on contrast-enhanced mammography: Associations with breast density and patient's characteristics. *Radiol. Med.* **2024**, *129*, 1303–1312. [CrossRef] [PubMed]
45. Yang, W.; Yang, Y.; Zhang, N.; Yin, Q.; Zhang, C.; Han, J.; Zhou, X.; Liu, K. The features associated with mammography-occult MRI-detected newly diagnosed breast cancer analysed by comparing machine learning models with a logistic regression model. *Radiol. Med.* **2024**, *129*, 751–766. [CrossRef] [PubMed]
46. Kim, J.Y.; Kim, J.J.; Lee, H.J.; Hwangbo, L.; Song, Y.S.; Lee, J.W.; Lee, N.K.; Hong, S.B.; Kim, S. Added value of diffusion-weighted imaging in detecting breast cancer missed by artificial intelligence-based mammography. *Radiol. Med.* **2025**, *130*. [CrossRef] [PubMed]
47. Fusco, R.; Granata, V.; Sansone, M.; Rega, D.; Delrio, P.; Tatangelo, F.; Romano, C.; Avallone, A.; Pupo, D.; Giordano, M.; et al. Validation of the standardized index of shape tool to analyze DCE-MRI data in the assessment of neo-adjuvant therapy in locally advanced rectal cancer. *Radiol. Med.* **2021**, *126*, 1044–1054. [CrossRef]
48. Granata, V.; Fusco, R.; Catalano, O.; Avallone, A.; Palaia, R.; Botti, G.; Tatangelo, F.; Granata, F.; Cascella, M.; Izzo, F.; et al. Diagnostic accuracy of magnetic resonance, computed tomography and contrast enhanced ultrasound in radiological multimodality assessment of peribiliary liver metastases. *PLoS ONE* **2017**, *12*, e0179951. [CrossRef]
49. Petrillo, A.; Fusco, R.; Petrillo, M.; Granata, V.; Delrio, P.; Bianco, F.; Pecori, B.; Botti, G.; Tatangelo, F.; Caracò, C.; et al. Standardized Index of Shape (DCE-MRI) and Standardized Uptake Value (PET/CT): Two quantitative approaches to discriminate chemo-radiotherapy locally advanced rectal cancer responders under a functional profile. *Oncotarget* **2016**, *8*, 8143–8153. [CrossRef]
50. Fusco, R.; Petrillo, M.; Granata, V.; Filice, S.; Sansone, M.; Catalano, O.; Petrillo, A. Magnetic Resonance Imaging Evaluation in Neoadjuvant Therapy of Locally Advanced Rectal Cancer: A Systematic Review. *Radiol. Oncol.* **2017**, *51*, 252–262. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.