Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

DOI: 10.37943/23OUMR1748

Beibit Abdikenov
PhD, Director of Science and Innovation Center "Artificial Intelligence"
beibit.abdikenov@astanait.edu.kz; orcid.org/0000-0002-0284-0949
Astana IT University, Kazakhstan

Tomiris Zhaksylyk
Master of Science, Researcher at Science and Innovation Center "Artificial Intelligence"
Zhaksylyk.tomiris@astanait.edu.kz; orcid.org/0009-0002-8749-1967
Astana IT University, Kazakhstan

Aruzhan Imasheva
Master's student, Junior Researcher at Science and Innovation Center "Artificial Intelligence"
aruzhan.imasheva@astanait.edu.kz; orcid.org/0009-0004-3790-2339
Astana IT University, Kazakhstan

Yerzhan Orazayev
Master of Science, Researcher at Science and Innovation Center "Artificial Intelligence"
y.orazayev@astanait.edu.kz; orcid.org/0009-0004-9826-625X
Astana IT University, Kazakhstan

Danara Suleimenova
Master of Medicine, Researcher at Science and Innovation Center "Artificial Intelligence"
Astana IT University, Kazakhstan
danara.suleimenova@gmail.com; orcid.org/0000-0003-0396-5249

FUSION VIEW-NET: DUAL-VIEW DEEP LEARNING FOR ROBUST MAMMOGRAPHIC BREAST CANCER CLASSIFICATION

Abstract: Breast cancer is still one of the top causes of cancer-related death for women globally, and better patient outcomes depend on early identification. Although mammography is the main imaging modality used for screening, the delicate nature of early clinical symptoms and inter-reader variability sometimes compromise diagnostic accuracy. We examine the application of deep convolutional neural networks (CNNs) to automated classification of mammogram images in this work. FusionView-Net (FV-Net) is also presented, a novel dual-view integration framework that combines data from mediolateral oblique (MLO) and craniocaudal (CC) views to improve diagnostic precision. To produce a more comprehensive depiction of the breast tissue than conventional single-view methods, FV-Net combines contextual and spatial data from both standard perspectives. Two publicly available mammography datasets, which have been properly divided to allow for both seen-unseen data configurations and cross-dataset generalization testing, are used to assess the approach. A variety of CNN architectures are evaluated on separate and combined datasets, including ResNet18 and a specially created CNN. Findings indicate that FV-Net significantly increases model robustness and classification accuracy, as evidenced by consistently better F1 scores and ROC AUC values, especially when combined with ResNet18 and the custom CNN. The necessity for flexible models in actual clinical settings is shown by generalization studies, which further highlight the significance of

Received: 26.05.2025 Accepted: 30.06.2025 Published: 30.09.2025

DOI: 10.37943/23OUMR1748

dataset diversity by showing a noticeable drop in performance when domain shifts are present. Our results demonstrate how well multi-view fusion works for CNN-based mammography classification and provide useful guidance for choosing architectures and training methods. The development of trustworthy, broadly applicable AI technologies to assist radiologists in the early diagnosis of breast cancer is made possible by FV-Net.

Keywords: Deep Learning, Mammography, Breast Cancer, Computer-Aided Diagnosis (CADx), Medical Image Analysis, Classification.

Introduction

According to the World Health Organization, breast cancer accounts for over 2.3 million new cases and roughly 685,000 deaths per year, making it the most common cancer among women and a major cause of cancer-related deaths globally [1]. The best way to lower the death rate from breast cancer is still early detection, and mammography-based screening programs significantly increase survival rates. Despite its proven benefits, mammography is not without challenges. Breast density, image quality, and the subtlety of early-stage tumor features are some of the factors that frequently impair diagnostic accuracy. Furthermore, inconsistent interpretations and missed or false-positive results can arise from inter-reader variability among radiologists, which is frequently brought on by fatigue, differences in experience, or high case volumes.

In the analysis of medical images, including mammograms, convolutional neural networks (CNNs) have demonstrated great promise. Deep learning has been used in a number of studies for tasks like malignancy classification [4], breast density assessment [3], and lesion detection [2]. Notable models have reached or even exceeded radiologist-level accuracy in large-scale screening contexts [5], [6]. Performance has also been enhanced by transfer learning techniques that use pretrained architectures like ResNet and DenseNet, particularly in situations with little labeled data [7], [8].

Despite these developments, the majority of CNN-based mammography models process the mediolateral oblique (MLO) and craniocaudal (CC) standard views separately. This differs from clinical workflows, where radiologists use both perspectives to more accurately describe abnormalities. Although some recent attempts have investigated multi-view learning through the use of parallel pipelines [9], view concatenation [10], or attention-based fusion [11], many of these approaches are not very good at accurately simulating the contextual and spatial relationships between views.

The ability of dual-stream architectures to process multiple views or modalities in parallel has made them popular in medical image analysis, especially in breast cancer screening, where standard mammography exams include four views (L-CC, L-MLO, R-CC, and R-MLO). Before combining their individual feature embeddings, these architectures frequently have two (or more) branches intended to process CC and MLO views separately [12], [13]. Better modeling of view-specific anatomical and pathological cues is made possible by this design, which makes it easier to learn specialized features from each view. For example, Lotter et al. [15] and Ribli et al. [14] showed that integrating bilateral and ipsilateral views via different processing paths improves diagnostic accuracy and more accurately represents the diagnostic workflow of radiologists.

This paradigm has been further adopted by recent transformer-based or graph-based architectures. In their evaluation of multi-view transformer and graph models, Manigrasso et al. [12] demonstrate that explicitly encoding view relationships improves performance over conventional CNNs, even with relatively small amounts of data. In a similar vein, Yala et al. [16] reinforced the advantages of separate-stream learning by proposing Mirai, a transformer-based model that independently processes each mammography view before aggregating

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

representations. Dual-stream designs have a higher computational burden despite the performance improvements. Particularly in high-resolution mammography, where full-resolution images can surpass 3000×5000 pixels, each parallel stream results in additional memory and computational demands. Scalability and real-time applicability are limited by this overhead, which is further increased when heavy-weight backbones such as ResNet-50 or ViT variants are used [13], [17]. Additionally, unless pretraining techniques or large datasets are used to prevent overfitting, training stability may be jeopardized [18]. In clinical settings with limited deployment infrastructure, this resource-performance trade-off is especially crucial. In order to preserve view-specific modeling capability without having to pay for duplicate backbones, a number of studies have started looking into more effective alternatives, such as attention-based fusion [19] or lightweight dynamic routing strategies, even though the dual-stream approach is still a good choice for multi-view medical image analysis.

In order to close this gap, we present FusionView-Net (FV-Net), a novel dual-view integration framework that creates a more comprehensive and richer representation of mammographic data by processing CC and MLO views simultaneously. FV-Net seeks to increase classification accuracy, robustness, and clinical utility by matching the architecture design to radiologists' diagnostic reasoning.

To test the generalizability of our models, we apply both intra- and cross-dataset evaluations to two publicly accessible mammography datasets. To understand how view integration and data diversity affect performance, we benchmark a number of CNN architectures under single-view and dual-view configurations, including ResNet18 and a custom-designed CNN. Our results show that the proposed dual-view approach significantly improves F1 scores and ROC AUC across architectures, with the custom CNN showing particularly strong generalization performance.

This paper makes several contributions:

Introduces a novel, clinically informed dual-view fusion architecture for mammogram classification.

Provides comprehensive benchmarking of CNN models across individual and fused-view configurations.

Offers new insights into the effects of dataset diversity and domain shift on model robustness in real-world scenarios.

The remainder of the paper is structured as follows: Section 2 details our proposed method, FusionView-Net. Section 3 describes the experimental setup, including datasets, evaluation metrics, and model training. Section 4 presents the results and performance analysis. Finally, Section 6 concludes with key findings and directions for future research.

Methods and Materials

In this section, we describe the detailed methodology of FusionView-Net (FV-Net), a deep learning framework designed for mammographic lesion classification. We discuss the workflow, including raw image acquisition, fusion strategies, preprocessing pipelines, and the CNN-based classification process. Additionally, we outline the evaluation protocols used to assess the model's robustness and generalizability across the VinDr-Mammo and CMMD datasets.

Datasets

VinDr-Mammo is a large-scale full-field digital mammography (FFDM) dataset that provides standard views (CC and MLO) with expert-annotated BI-RADS assessments. For the purpose of this study, BI-RADS categories were converted into binary labels—benign or malignant—based on standard BI-RADS interpretation guidelines [20]. Specifically, BI-RADS 2 to 4 were categorized as benign, while BI-RADS 5 was treated as malignant.

DOI: 10.37943/23OUMR1748

CMMD (Chinese Mammography Database) contains full-field digital mammography (FFDM) images with clinical labels and detailed annotations of tumor presence and type. Like Vin-Dr-Mammo, both CC and MLO views are available for each breast.

An overview of the number of benign and malignant images in each dataset is provided in Table 1.

Table 1. Overview of utilized datasets

Dataset	Number of Benign Images	Number of Malignant Images	Label Source
VinDr-Mammo	6368	226	BI-RADS
CMMD	1108	2626	Biopsy-confirmed

Proposed Methodology

We propose FusionView-Net (FV-Net), a novel deep learning framework that uses complementary information from the two common mammography views MLO and CC, to classify mammogram images into benign or malignant categories. In order to predict lesion malignancy, the suggested approach combines these perspectives into a single fused representation, which is then processed by a convolutional neural network (CNN). Raw image acquisition, view fusion, preprocessing, and classification are all included in the overall workflow, which is shown in Fig. 1.

Architecture Overview

The custom CNN model is presented in this study. *Fig. 2* depicts proposed architecture. The data acquisition procedure, fusion tactics, preprocessing pipelines, CNN-based classification, and evaluation protocols are all covered in detail in this section.

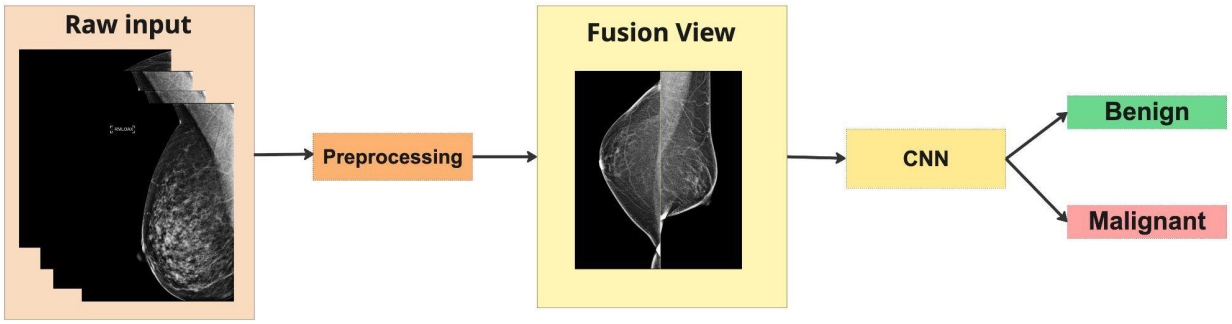

Figure 1. Workflow of proposed FusionView-Net (FV-Net) methodology.

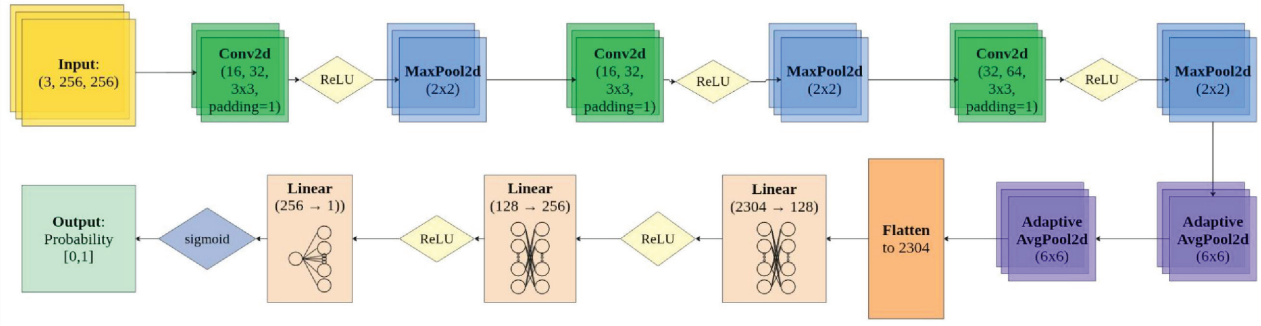

Figure 2. Custom CNN architecture.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

The CNN architecture can be described mathematically as a function $f(x; \theta)$, which maps the input image $x \in \mathbb{R}^{\{H \times W \times C\}}$ to an output prediction (e.g., class probability or segmentation mask), where θ represents the learnable parameters (weights and biases).

- Convolutional Layers

Each convolutional layer applies a set of learnable kernels to the input tensor. The output feature map x^{l+1} at layer $l+1$ is given by:

$$x_{i,j,k}^{l+1} = \sum_{m+1}^{C} \sum_{u+1}^{K} \sum_{v+1}^{K} W_{u,v,m,k}^{l} \cdot x_{i+u,j+v,m}^{l} + b_{k}^{l} \quad (1)$$

where K is the kernel size, C is the number of input channels, and b^{l}_{k} is the bias term.

- Activation Function

After each convolution, a nonlinear activation function $\sigma(\cdot)$ is applied. For ReLU:

$$\mathrm{ReLU}(x) = \max(0, x) \quad (2)$$

This produces the activated feature maps $z^{l+1} = \sigma(x^{l+1})$.

- Pooling Layers

Pooling layers reduce the spatial dimensions of feature maps. For instance, 2×2 max pooling is defined as:

$$x_{i,j,k}{}^{l+1} = max\, z_{2i+p,2j+q,k}{}^{l} \quad (3)$$

- Fully Connected Layer

After flattening, fully connected layers map the extracted features to the output space. For classification, a softmax function is applied:

$$y = Softmax\,(W^{fc}) \cdot z^{L} + b^{fc} \quad (4)$$

$$Softmax(y_i) = \frac{e^{y_i}}{\sum_{j}^{i} e^{y_i}} \quad (5)$$

- Complete Model

The complete CNN is a composition of the above layers:

$$f(x;\, \theta) = x^{L} = f^{L} \circ f^{L-1} \circ \ldots \circ f^{1}(x) \quad (6)$$

Workflow Overview

The FV-Net workflow consists of four main stages, seamlessly integrated to process mammographic images for lesion classification. First, raw input acquisition involves collecting both craniocaudal (CC) and mediolateral oblique (MLO) views of each breast from mammographic imaging, capturing distinct anatomical perspectives to provide complementary diagnostic information. Next, the fusion pipeline combines these CC and MLO views into a single image through one of two fusion strategies – direct fusion or cropped fusion, as described in the fusion strategies section – creating a unified representation that encapsulates critical diagnostic details from both views. Following this, the preprocessing stage normalizes the fused image to a pixel intensity range of [0, 1] and resizes it to a standardized resolution of 256×256 pixels, ensuring compatibility with the convolutional neural network (CNN) architecture while enhancing computational efficiency. Finally, the classification stage feeds the preprocessed fused image into a CNN-based model, which leverages transfer learning where necessary to predict whether the lesion is benign or malignant.

DOI: 10.37943/23OUMR1748

Preprocessing Pipelines

To investigate the impact of preprocessing on model performance, two distinct preprocessing pipelines are implemented, corresponding to the fusion strategies described above:

- *Preprocessing Pipeline 1 (Direct Fusion Pipeline):* The CC and MLO images are concatenated without any cropping. The concatenated image is normalized to a pixel intensity range of [0, 1] using min-max normalization and resized to 256 × 256 pixels using bilinear interpolation. This pipeline prioritizes simplicity and retains all image information, including non-breast regions. The pipeline is shown on *Fig. 3*.
- *Preprocessing Pipeline 2 (Cropped Fusion Pipeline):* Prior to concatenation, each CC and MLO image undergoes a breast isolation step. This involves applying an automated segmentation algorithm to detect and extract the breast tissue, removing the background. The cropped images are then concatenated, normalized to [0, 1], and resized to 256 × 256 pixels using bilinear interpolation. This pipeline aims to reduce noise and focus the model on clinically relevant features. The pipeline is shown on *Fig. 4*.

The preprocessing pipelines are designed to ensure consistency in input dimensions and intensity ranges while allowing for a comparative analysis of the impact of background removal on classification accuracy.

The fused images are processed by a CNN-based classifier to predict whether a lesion is benign or malignant. The CNN architecture is based on a pre-trained model (e.g., ResNet 18) fine-tuned via transfer learning to adapt to the mammography domain. Transfer learning is employed to leverage features learned from large-scale datasets (e.g., ImageNet) while tailoring the model to the specific characteristics of mammographic images. The final fully connected layer of the CNN is modified to output two classes (benign or malignant), and the model is trained using a binary cross-entropy loss function. The training process involves optimization with the Adam optimizer and a learning rate scheduler to ensure convergence.

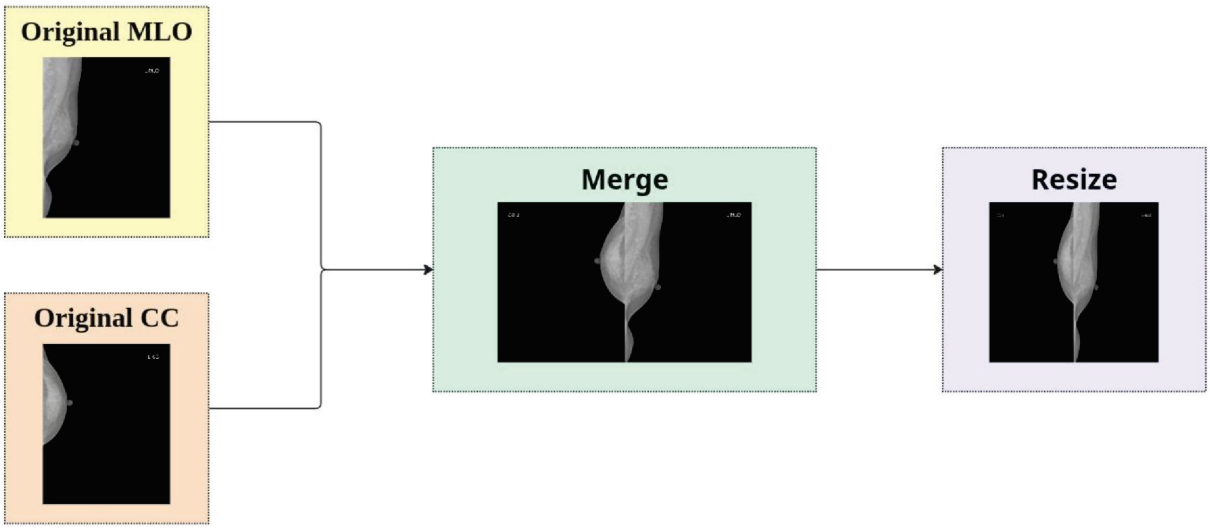

Figure 3. Image preprocessing pipeline 1: Direct Fusion Pipeline.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

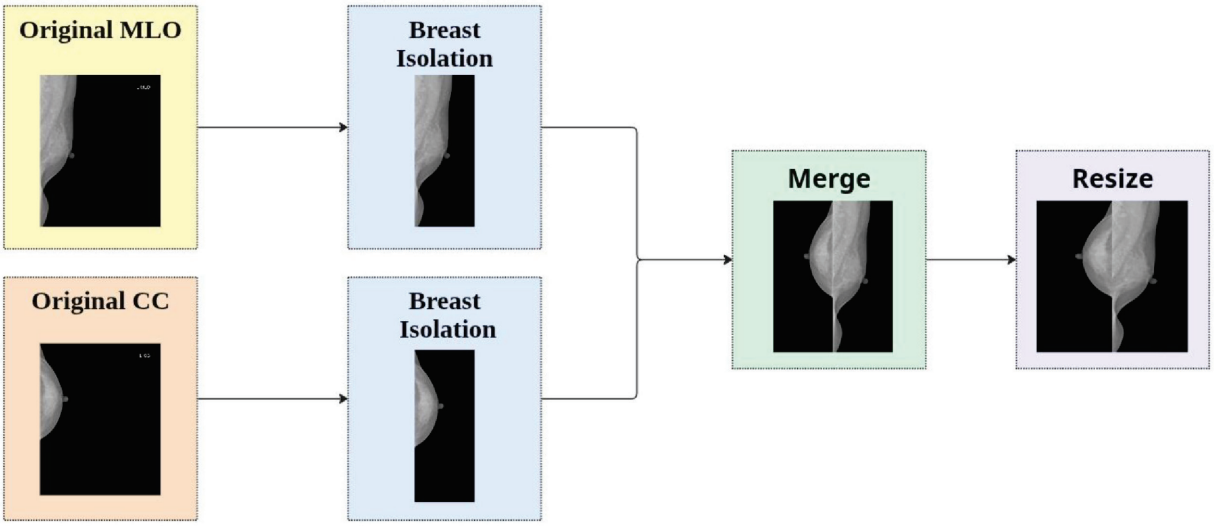

Figure 4. Image preprocessing Pipeline 2: Cropped Fusion Pipeline.

Evaluation Protocols

To assess the robustness and generalizability of FusionView-Net (FV-Net), the model is evaluated using two distinct datasets: the VinDr-Mammo dataset, and the CMMD. The evaluation is conducted under two protocols designed to test the model's performance under varying conditions of data familiarity and generalization. Both protocols utilize standard performance metrics, including accuracy, sensitivity, specificity, and F1.

The first setting, termed Seen-Patient Evaluation, involves splitting each dataset (VinDr-Mammo and CMMD) such that images from the same patients may appear in both training and testing sets, though no identical images are shared. This setup allows the model to leverage patient-specific distributional patterns, potentially enhancing performance on familiar data. The dataset is partitioned using a stratified split to maintain class balance, and the model is tested separately on each dataset to evaluate its performance when trained and tested within the same dataset.

The second and third settings, Unseen-Patient Evaluation, are designed to rigorously test the model's generalizability by ensuring that all patients in the test set are excluded from the training set. Two configurations are employed: in the first, the model is trained on one dataset (e.g., VinDr-Mammo) and tested on the other (e.g., CMMD), and vice versa, to assess cross-dataset generalization. In the second configuration, both datasets are combined, but patient-level partitioning ensures that no patient's images appear in both training and testing sets. This setup simulates real-world clinical scenarios where the model encounters entirely new patients, providing a stringent test of its ability to generalize across diverse populations and imaging conditions.

By evaluating FV-Net on the VinDr-Mammo and CMMD datasets under these protocols, the study aims to comprehensively assess its robustness and generalizability, leveraging both within-dataset and cross-dataset scenarios to ensure applicability in diverse clinical settings. The visual representation of the protocols is given in *Fig. 5*.

DOI: 10.37943/23OUMR1748

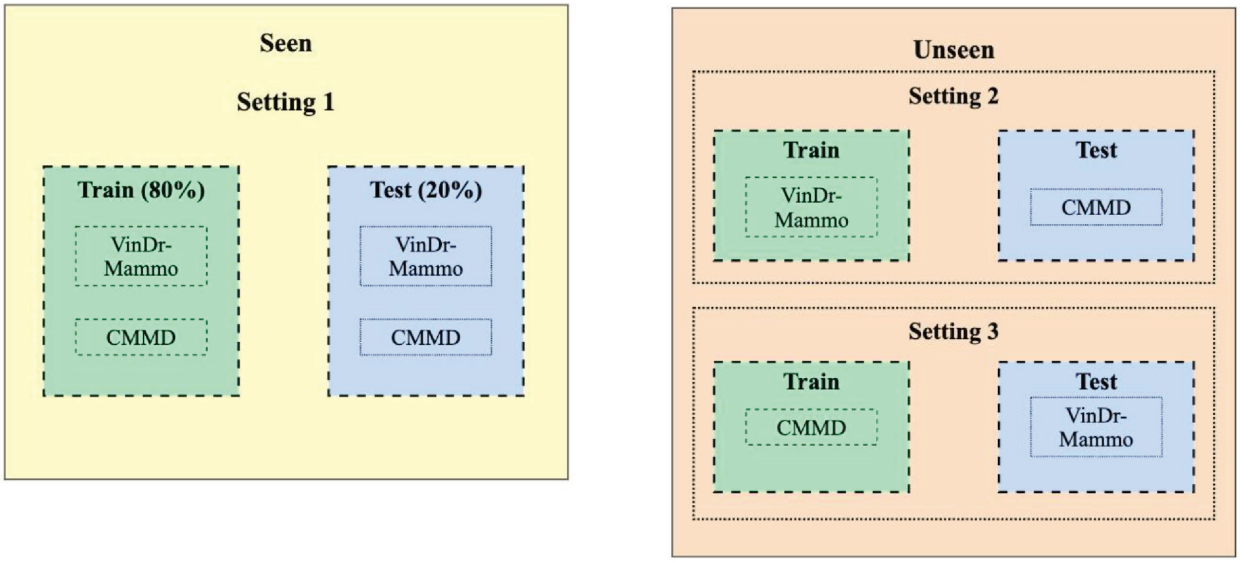

Figure 5. Evaluation Protocols.

Implementation Details

The experiments were conducted on a dedicated workstation with two 11 GB video memory-equipped NVIDIA GeForce RTX 2080 Ti GPUs. The system used CUDA and cuDNN for hardware acceleration and was set up with Ubuntu 20.04. PyTorch was used as the main framework for model construction and training in Python 3.9, where the deep learning operations were developed. Every training and inference activity was carried out within the PyTorch ecosystem, making use of its adaptable model-building tools and effective data management features. Training was split over both GPUs using the DataParallel module to optimize GPU consumption.

Below is a list of the primary training parameters that were employed in the experiments:

- Batch size: 32 (adjusted based on available GPU memory)
- Optimizer: Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$

- Update biased first moment estimate:

$$m_t = \beta 1 * m_{\{t-1\}} + (1 - \beta 1) * g_t \tag{7}$$

- Update biased second raw moment estimate:

$$V_t = \beta 2 * v_{\{t-1\}} + (1 - \beta 2) * g_t^2 \tag{8}$$

- Compute bias-corrected moment estimates:

$$\hat{M}_t = \frac{m_t}{(1 - \beta 1^t)} \tag{9}$$

$$\hat{v}_t = \frac{v_t}{(1 - \beta 2^t)} \tag{10}$$

- Update parameters:

$$\Theta_t = \theta_{\{t-1\}} - \alpha * \frac{\hat{m}_t}{(\sqrt{\hat{v}_t} + \varepsilon)} \tag{11}$$

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

- Learning rate: Initialized at 1×10^{-4} with a cosine annealing schedule
  - Loss function: Binary Cross-Entropy Loss

$$L_{BCE} = -[y * \log(\hat{y}) + (1 - y) * \log(1 - \hat{y})] \quad (12)$$

  - For a batch of N samples, the mean BCE loss is:

$$L_{BCE} = -\left(\frac{1}{N}\right) * \Sigma [y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)] \quad (13)$$

- Training duration: Between 30 and 50 epochs, with early stopping based on validation performance

To maintain experimental consistency, the same preprocessing routine and training configuration were used across all dataset variations and model runs. Additionally, random seeds were fixed to ensure reproducibility in data splits and weight initialization. This environment and pipeline allowed for the efficient execution of extensive experimentation, including comparisons of neural networks, cross-dataset generalization tests, and evaluation of the two fusion strategies described earlier.

Results

ResNet18 and a custom convolutional neural network are trained and tested across three evaluation settings (depicted in *Fig. 5*) and two preprocessing pipelines (depicted in *Fig. 3* and *Fig. 4*).

Results under Pipeline 1 (No Cropping)

In Setting 1 with the same distribution of training and testing datasets, ResNet18 achieved the highest overall performance with an accuracy of 0.8780, F1 score of 0.7934, a strong recall of 0.8521, which shows its ability to capture true positive cases effectively. On the other hand, custom CNN model performed competitively with an accuracy of 0.8497 and F1 score of 0.7370. However, it was more prone to false positives with lower precision of 0.7103, compared to ResNet18.

In setting 2 with domain generalization, both models performed poorly. The accuracy of ResNet18 dropped to 0.4146, and its F1 score fell to 0.2852. Similarly, custom CNN showed poor performance with an accuracy of 0.4109 and an F1 score of 0.2784.

In setting 3 with a different unseen domain, ResNet18 performed better than in Setting 2. It achieved an accuracy of 0.7779 and an F1 score of 0.8734—its highest F1 score across all settings—driven by an exceptionally high precision of 0.9719. The model was confident when predicting positive cases, although recall was more modest at 0.7930. In contrast, custom CNN showed substantially lower performance, with an accuracy of 0.6152 and an F1 score of 0.6577. While the recall was decent (0.7201), the precision was much lower at 0.6052.

Results under Pipeline 2 (With Cropping)

Incorporation of cropping step into preprocessing pipeline showed slightly better overall performance of models.

In Setting 1, ResNet18's performance improved slightly, reaching an accuracy of 0.8819. Although the F1 score of 0.7904 was similar to that in Pipeline 1, its precision increased to 0.7986, suggesting more reliable predictions. Interestingly, custom CNN also benefited from cropping, improving its accuracy to 0.8722 and, notably, surpassing ResNet18 in F1 score with a value of 0.7937. This was largely driven by its substantially higher recall (0.8944), demonstrating that cropping enabled the model to detect more true positives at the expense of slightly lower precision (0.7135).

DOI: 10.37943/23OUMR1748

Setting 2 revealed a more pronounced advantage of cropping. ResNet18's accuracy increased from 0.4146 in Pipeline 1 to 0.6277 in Pipeline 2, and its F1 score more than doubled, reaching 0.6899. This improvement was largely due to enhanced precision (0.8330). Custom CNN also showed better performance, with an increase in F1 score from 0.2784 to 0.5666. However, it remained behind ResNet18 in all metrics, suggesting that while cropping helped, it was not sufficient for the custom CNN to match the robustness of ResNet18 in this setting.

In Setting 3, cropping slightly reduced ResNet18's performance compared to Pipeline 1, with accuracy dropping from 0.7779 to 0.7262 and F1 score from 0.8734 to 0.7666. However, it maintained high and balanced precision and recall (0.7700 and 0.764), indicating consistent performance. The custom CNN showed a slight improvement in this setting compared to the no-cropping pipeline, increasing its F1 score to 0.6958, with a recall of 0.7542 and a precision of 0.6452.

Table 2. Overview of Experimental Results

Preprocessing Pipeline	Evaluation settings	Model	Accuracy	Precision	Recall	F1
Pipeline 1 (no cropping)	Setting 1 (Seen)	Resnet18	0.8780	0.7423	0.8521	0.7934
		Custom CNN	0.8497	0.7103	0.7716	0.7370
	Setting 2	Resnet18	0.4146	0.2236	0.3935	0.2852
		Custom CNN	0.4109	0.2123	0.4146	0.2784
	Setting 3	Resnet18	0.7779	0.9719	0.7930	0.8734
		Custom CNN	0.6152	0.6052	0.7201	0.6577
Pipeline 2 (with cropping)	Setting 1 (Seen)	Resnet18	0.8819	0.7986	0.7823	0.7904
		Custom CNN	0.8722	0.7135	0.8944	0.7937
	Setting 2	Resnet18	0.6277	0.8330	0.5887	0.6899
		Custom CNN	0.5972	0.5591	0.5745	0.5666
	Setting 3	Resnet18	0.7262	0.7700	0.7641	0.7666
		Custom CNN	0.6952	0.6452	0.7542	0.6958

Discussion

The results from our experiments offer a nuanced understanding of the comparative strengths and limitations of ResNet18 and the custom convolutional neural network under varying domain conditions and preprocessing pipelines. The experimental results offer several key insights that are critical for interpreting model behavior and guiding future work in robust image classification.

Model Performance and Generalization

ResNet18 consistently showed superior performance across different settings and pipelines, particularly excelling in precision. In setting 1, its performance was notably strong across both pipelines, showing that it is more suitable for cases where the training and testing distributions are aligned similarly. The real distinction across models is apparent from the other 2 settings with domain generalization. The severe drop in performance under Pipeline 1 for Setting 2 highlights the sensitivity of both models to domain shifts, a common challenge in real-world applications. ResNet18's recovery in Setting 3 and substantial improvement under cropping conditions (Pipeline 2) in Setting 2 underscore its capacity to adapt more effectively when aided by appropriate preprocessing.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

In contrast, custom CNN showed promising performance with its high recall rates, which is an important metric, especially when it comes to high-stakes domains such as medical imaging, where it is crucial not to miss positive cases. In some cases, it trailed behind ResNet18, but it achieved notable improvements under Pipeline 2, particularly in Setting 1, where it outperformed ResNet18 in F1 score due to a remarkably high recall. This suggests that the model is adept at identifying true positives when focused on relevant image regions, albeit at the cost of higher false positive rates (lower precision). This high sensitivity suggests that custom CNN may be well-suited for applications where detecting every relevant instance is prioritized over minimizing false positives. In domain generalization scenarios (Settings 2 and 3), custom CNN showed consistent improvements under the cropping pipeline. While it did not fully close the gap with ResNet18 in these unseen domains, its gains in recall indicate meaningful enhancement in generalization when provided with a more targeted input representation. These results suggest that custom CNN, despite its simpler architecture, possesses strong detection capability and, when supported with effective preprocessing, can serve as a competitive and practical model, especially in recall-critical contexts.

Impact of Cropping as a Preprocessing Strategy

The results indicate that it can be critical to use the right preprocessing to enhance robustness and generalization of models. Cropping, which is designed to focus on the breast tissue rather than the background, improved performance across domain generalization settings for both models.

The benefit was apparent in Setting 2, a challenging unseen dataset scenario, where cropping led to substantial gains in accuracy and F1 score for both models. This improvement suggests that reducing input noise and directing attention to semantically relevant regions helps mitigate the effects of domain shift. While ResNet18 saw a slight decrease in performance in Setting 3 under the cropping pipeline, it still maintained balanced precision and recall. For custom CNN, cropping yielded notable and consistent improvements, especially in Settings 1 and 3. These relative improvements suggest that targeted preprocessing can significantly boost the effectiveness of more compact or domain-specific architectures.

Trade-offs Between Precision and Recall

A key takeaway from the results is the complementary trade-off between precision and recall observed in the two models. ResNet18 consistently favored precision, leading to fewer false positives, which can be advantageous in contexts where incorrect positive predictions have higher costs. In contrast, custom CNN tended to favor recall, especially when cropping was applied. This behavior means that the model was more aggressive in identifying potential positive cases, which is particularly valuable in fields like medical imaging, where missing a positive case (false negative) can have more serious consequences than flagging a false positive. High recall, as seen in multiple settings, especially the remarkable 0.8944 recall in Setting 1 with cropping, demonstrates the model's ability to effectively detect relevant patterns when provided with clear and focused input. Tasks requiring high certainty may favor ResNet18, while sensitivity-critical applications may benefit more from custom CNN.

Conclusion

This study evaluated the performance of ResNet18 and a custom CNN across multiple domain conditions and preprocessing strategies, offering insights into their respective strengths and use cases. While ResNet18 consistently showed strong generalization and precision, the custom CNN model demonstrated compelling performance, particularly in terms of recall and sensitivity, and especially when paired with cropping. Importantly, cropping significantly improved both models' robustness under domain shift, with particularly strong relative gains

for custom CNN. These findings illustrate that preprocessing techniques are not merely enhancements, they can be transformative, especially for models with lighter architectures or domain-specific optimizations.

Looking forward, future research should explore adaptive or learned preprocessing mechanisms, such as attention-based cropping or dynamic region selection, to further amplify the strengths of each architecture. Evaluating model behavior under a broader range of domain shifts and tasks will also help deepen our understanding of how to build robust, flexible, and context-aware deep learning systems for real-world applications.

Acknowledgement

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number BR24993145.

References

[1] World Health Organization. (2021). Breast cancer. Geneva, Switzerland: World Health Organization. Retrieved from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). *Deep learning to improve breast cancer detection on screening mammography*. Radiology, 292(3), 535–540.

[3] Lehman, C.D., Wellman, R.D., Buist, D.S.M., Kerlikowske, K., Tosteson, A.N.A., & Miglioretti, D.L. (2019). *Mammographic breast density assessment using deep learning: Clinical implementation*. Radiology, 290(1), 52–58.

[4] McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., et al. (2020). *International evaluation of an AI system for breast cancer screening*. Nature, 577(7788), 89–94.

[5] Karaca Aydemir, B.K., Telatar, Z., Güney, S. *et al.* (2025). *Detecting and classifying breast masses via YOLO-based deep learning*. Neural Comput & Applic 37, 11555–11582. https://doi.org/10.1007/s00521-025-11153-1

[6] Lotter, W., Sorensen, G., Ding, J., Kim, H., Ghassemi, M., Haider, Z., et al. (2021). *Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach*. Nature Medicine, 27(2), 244–249.

[7] Carriero, A., Groenhoff, L., Vologina, E., Basile, P., & Albera, M. (2024). *Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024*. Diagnostics, 14(8), 848. https://doi.org/10.3390/diagnostics14080848.

[8] Songsaeng, Chatsuda & Pradaranon, Varanatjaa & Chaichulee, Sitthichok. (2021). *Multi-Scale Convolutional Neural Networks for Classification of Digital Mammograms With Breast Calcifications*. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3104627.

[9] Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J.L., & Guevara López, M.A. (2020). *Representation learning for mammography classification using multi-view information*. Computer Methods and Programs in Biomedicine, 190, 105361.

[10] Manigrasso, F., Milazzo, R., Russo, A. S., Lamberti, F., Strand, F., Pagnani, A., & Morra, L. (2025). *Mammography classification with multi-view deep learning techniques: Investigating graph and transformer-based architectures.* Medical Image Analysis*, 99*, 103320. https://doi.org/10.1016/j.media.2024.103320.

[11] Nasir, I.M., Alrasheedi, M.A., & Alreshidi, N.A. (2024). *MFAN: Multi-Feature Attention Network for Breast Cancer Classification*. Mathematics, *12*(23), 3639. https://doi.org/10.3390/math12233639.

[12] F. Manigrasso, R. Milazzo, A.S. Russo, F. Lamberti, F. Strand, A. Pagnani, & L. Morra. (2025). *Mammography classification with multi-view deep learning techniques: Investigating graph and transformer-based architectures*. Medical Image Analysis, vol. 99, Art. no. 103320. https://doi.org/10.1016/j.media.2024.103320

[13] Yang, B., Peng, H., Luo, X., & Wang, J. (2024). *Multi-stages attention breast cancer classification based on nonlinear spiking neural P neurons with autapses*. arXiv. https://arxiv.org/abs/2312.12804

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 23, SEPTEMBER 2025

[14] Ribli D., Horváth A., Unger Z., Pollner P. (2021). *Detecting and classifying lesions in mammograms with deep learning*. Scientific Reports, vol. 8, no. 1, p. 4165. https://doi.org/10.1038/s41598-018-22437-z.

[15] Lotter W., Sorensen K., Golan T., Barzilay R. (2021). Breast cancer detection with a transformer-based model for high-resolution mammograms. *Nature Communications*, vol. 12, p. 518. https://doi.org/10.1038/s41467-020-20407-z.

[16] Yala A., Mikhael P., Strand F., Lin G., Smith K., Barzilay R.. (2022). *Toward robust mammography-based models for breast cancer risk. Science Translational Medicine*, vol. 14, no. 629, eabj5325. https://doi.org/10.1126/scitranslmed.abj5325.

[17] Trivizakism E., Tsiknakis S., Vamvakas G., Marias K. (2020). *A deep learning approach for automatic classification of breast lesions on mammography.* Journal of Healthcare Engineering, vol. 2019, Art. no. 4180212. https://doi.org/10.1155/2019/4180212.

[18] Wu N., Phang J., Park J., Shen Y.,. Huang Z, Zorin M. (2023). *Self-supervised learning for mammography*. Medical Image Analysis, vol. 87, 102787. https://doi.org/10.1016/j.media.2023.102787.

[19] Raghu M., Zhang C., Kleinberg J., Bengio S. (2021). *Do vision transformers see like convolutional neural networks?* Advances in Neural Information Processing Systems, vol. 34, pp. 12116–12128, 2021.

[20] D'Orsi, C.J., Sickles, E.A., Mendelson, E.B., & Morris, E.A. (2014). *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. American College of Radiology.