A Comparative Study of Deep Learning Architectures for Breast Lesion Segmentation and Detection in Multi-Source Mammography*

Maria Temneanu[1], Emanuela-Coralia Toda-Puiuleț[1], Bogdan Teodorescu[1], Oana Tibil[1], and Prof. Laura Silvia Dioșan[2]

[1] Babeș-Bolyai University, Cluj-Napoca, Romania
{maria.temneanu,emanuela.toda,bogdan.teodorescu1,oana.tibil1}@stud.ubbcluj.ro
[2] Babeș-Bolyai University, Faculty of Mathematics and Computer Science, Cluj-Napoca, Romania
laura.diosan@ubbcluj.ro

Abstract. Breast cancer is the most commonly diagnosed cancer among women worldwide, with over 2.26 million new cases reported in 2020 according to the World Health Organization. Early detection through mammographic screening significantly improves survival rates, yet manual interpretation remains time-consuming and error-prone. In this work, we present a systematic evaluation of deep learning approaches for breast lesion segmentation and detection across three distinct experimental setups. We investigate three diverse architectures—a Dense Spatial U-Net (DSU-Net), a VGG16-U-Net leveraging transfer learning, and a Faster R-CNN detection model—evaluating each on specific configurations of the INbreast, MIAS, and DMID datasets. Rather than comparing these models under identical conditions, each experiment explores how different architectural strengths interact with varied data environments. DSU-Net evaluates the impact of dense connectivity and spatial attention, VGG16-U-Net assesses the robustness of pretrained representations on combined datasets, and Faster R-CNN highlights the important role of detection-based localization. Results indicate that while DSU-Net achieves strong segmentation under consistent annotations, cross-dataset heterogeneity remains a major challenge. Conversely, VGG16-U-Net proves robust under such heterogeneity, achieving 63% IoU, and Faster R-CNN highlights the complementary value of regional detection. These findings emphasize that annotation-protocol alignment is as critical as architectural design for effective multi-source medical image analysis.

Keywords: Breast cancer · Mammography · Semantic segmentation · Object detection · Deep learning · U-Net · Dense blocks · Spatial attention · Transfer learning · INbreast · MIAS · DMID · Faster R-CNN · VGG16

* This work was developed as a student project at Babeș-Bolyai University.

1 Introduction

Breast cancer remains the leading cause of malignancy-related mortality among women globally. According to GLOBOCAN 2020 estimates, approximately 2.26 million new breast cancer cases were diagnosed worldwide, surpassing lung cancer for the first time as the most frequently diagnosed malignancy [1]. In Romania, breast cancer accounted for 26.9% of all new female cancer diagnoses in 2020 (12,085 cases), and was responsible for 3,918 deaths that year [2]. The prognosis is strongly stage-dependent: women diagnosed at an early stage have a cumulative survival probability of approximately 96% within the European Union, whereas late-stage diagnosis reduces this to roughly 38% [3].

Mammography is the gold-standard imaging modality for population-level breast cancer screening. However, radiological interpretation of mammograms is inherently challenging due to the subtle appearance of early lesions, substantial inter-reader variability, and the high throughput required in screening programs. Computer-aided detection and computer-aided diagnosis systems have been developed to assist radiologists, and recent advances in deep learning, particularly convolutional neural networks, have dramatically improved the accuracy of automated mammographic analysis [4,5].

Within the spectrum of mammographic analysis tasks, *semantic segmentation*, the pixel-level delineation of lesion boundaries, is of particular clinical importance. Unlike bounding-box detection, segmentation provides precise spatial information about lesion shape, margins, and extent, which are key factors in the BI-RADS assessment and subsequent clinical decision-making. Despite its importance, breast lesion segmentation remains a challenging problem due to the low contrast between lesions and surrounding dense glandular tissue, the diversity of lesion morphologies (masses, calcifications, architectural distortions), and the limited availability of pixel-level annotations in publicly available datasets [6]. To navigate these challenges, we present a systematic evaluation of diverse deep learning architectures, analyzing how different structural approaches interact with heterogeneous mammographic datasets. Rather than proposing a single model under uniform conditions, we investigate how specific architectural strengths address the complex realities of varying data sources and annotation protocols.

In this paper, we make the following contributions:

- We evaluate **DSU-Net** (Dense Spatial U-Net) to analyze the impact of DenseNet-style bottlenecks and spatial attention gates on segmentation performance under specific single and multi-dataset conditions (MIAS and INbreast).
- We investigate the robustness of transfer learning by evaluating a **VGG16-U-Net** on a diverse, combined dataset (INbreast + DMID), assessing the generalization capabilities of pretrained representations across different image distributions.
- We explore bounding-box localization using a **Faster R-CNN** architecture on the INbreast dataset, highlighting the complementary clinical value of detection-based models alongside pixel-level segmentation.

The two central research questions driving this work are:

- **RQ1:** How do distinctly different architectural paradigms—namely a custom dense-attention segmentation network, a pretrained transfer learning model, and a two-stage object detector—perform in identifying and delineating breast lesions?
- **RQ2:** How does the inherent heterogeneity of combining different mammographic datasets impact model generalization compared to isolated, single-dataset training environments?

2 Related Work

2.1 Mammographic Datasets for Segmentation

Several publicly available mammographic databases have been widely adopted for breast lesion segmentation research.

INbreast [7] is a full-field digital mammography (FFDM) database acquired at Hospital de São João, Porto, Portugal. It contains 410 mammographic images with precise radiologist-provided annotations in XML/Plist format, including both polygon contours for masses (annotation type 15) and point annotations for calcifications (annotation type 19). The high resolution and clinical-grade annotations make INbreast a reference benchmark for segmentation tasks.

MIAS (Mammographic Image Analysis Society) [8] is one of the earliest publicly available mammographic databases, containing 322 digitized film mammograms. Annotations are provided as center coordinates and radii of circles enclosing abnormalities. While the image quality is lower than modern digital datasets, MIAS remains widely used due to its accessibility and the diversity of lesion types represented.

CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography) [9] is an updated, standardized version of the original DDSM database, containing decompressed DICOM images with curated ROI segmentation masks and pathological diagnoses. It provides a large-scale collection suitable for both classification and segmentation tasks.

DMID (DDSM Mammography Image Dataset) is a curated breast imaging subset derived from DDSM, providing color-coded TIF overlay masks where green and red regions indicate lesion areas [20].

Other notable datasets include BancoWeb LAPIMO [18] and the recently introduced VinDr-Mammo [19], though these are less commonly used for segmentation benchmarking. The primary limitation of these datasets for semantic segmentation tasks lies in the granularity of their annotations. For instance, VinDr-Mammo provides annotations in the form of bounding boxes rather than precise pixel-level masks; while sufficient for object detection, bounding boxes include significant amounts of healthy tissue, making them unsuitable for training models that require exact lesion boundary delineation. Similarly, other large-scale databases often provide only image-level labels (e.g., presence or absence of malignancy) or BI-RADS scores, lacking the spatial metadata necessary for

pixel-wise supervision. Consequently, researchers continue to rely on smaller but more precisely annotated datasets like INbreast for evaluating segmentation-specific architectures.

2.2 Deep Learning Approaches to Mammographic Segmentation

The U-Net architecture [10] and its variants have become the dominant paradigm for medical image segmentation. In the context of mammography, several notable approaches have been proposed.

Ravitha Rajalakshmi et al. [12] proposed a deeply supervised U-Net (DS U-Net) coupled with dense conditional random fields (CRFs) for whole-mammogram lesion segmentation, achieving Dice scores of 79.1% on INbreast and 81.8% on CBIS-DDSM.

Baccouche et al. [6] introduced Connected-UNets, which cascades two U-Net architectures with additional modified skip connections and Atrous Spatial Pyramid Pooling. Evaluated on CBIS-DDSM, INbreast, and a private dataset, they achieved Dice scores of 89.52%, 95.28%, and 95.88%, and IoU scores of 80.02%, 91.03%, and 92.27%, respectively.

Elkorany and Elsharkawy [13] proposed a deep U-Net framework for mammographic abnormality detection incorporating Li thresholding, CLAHE enhancement, and median filtering as preprocessing. They reported average Dice scores of 85.61% on INbreast and 87.98% on CBIS-DDSM.

Ahmad et al. [14] developed a comprehensive CAD system combining YOLO-V7 for detection, Associated-ResUNet for segmentation, and BreastNet-SVM for classification, achieving a segmentation Dice score of 95.89% on CBIS-DDSM.

More recently, attention mechanisms and transformer-based architectures have been explored. The Hybrid Transformer U-Net (HTU-Net) [16] integrates channel and spatial self-attention with convolutional layers, achieving state-of-the-art results across CBIS-DDSM and INbreast. Similarly, attention-based U-Net variants incorporating DenseNet encoders have shown improved feature propagation and segmentation accuracy on breast ultrasound and mammography tasks [17].

3 Proposed Approaches

Although our work may have lacked resources such as storage and training power, we have managed to create a pipeline in which our architecture is capable of processing data from the various datasets used in other works and in the literature, while producing increasingly favorable results.

a) DSU-Net Architecture

The proposed Dense Spatial U-Net (DSU-Net) is built within a U-Net encoder-decoder framework and combines three components: a convolutional encoder, a dense bottleneck, and an attention-gated decoder. The encoder contains four

downsampling levels with 3×3 convolutional blocks, Batch Normalization, ReLU activation, and max-pooling. The number of channels increases from 64 to 512, while the spatial resolution is reduced from 256×256 to 16×16. This four-level design preserves enough spatial detail for small lesions while still providing sufficient contextual information for larger masses.

Instead of a standard bottleneck, DSU-Net uses a DenseNet-inspired dense block with six composite layers and growth rate $k = 32$. Each layer applies Batch Normalization, ReLU activation, a 1×1 bottleneck convolution, followed by another Batch Normalization, ReLU activation, and a 3×3 convolution. The dense connectivity pattern is defined as:

$$\mathbf{x}_l = H_l([\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{l-1}]) \tag{1}$$

where $\mathbf{x}_l$ is the output of layer l, H_l is the composite transformation, and $[\cdot]$ denotes channel-wise concatenation. With 512 input channels and six layers of growth rate 32, the bottleneck output contains $512 + 6 \times 32 = 704$ channels. This improves feature reuse and gradient flow while maintaining a moderate parameter count.

The decoder restores spatial resolution through transposed convolutions and uses spatial attention gates before concatenating encoder skip connections. The attention coefficient is computed as:

$$\alpha = \sigma\left(\psi\left(\mathrm{ReLU}\left(W_g(\mathbf{g}) + W_x(\mathbf{x})\right)\right)\right) \tag{2}$$

where $\mathbf{g}$ is the decoder gating signal, $\mathbf{x}$ is the encoder feature map, W_g and W_x are learned 1×1 convolutions, ψ is a 1×1 convolution, and σ is the sigmoid activation. The attended skip feature is:

$$\hat{\mathbf{x}} = \mathbf{x} \odot \alpha \tag{3}$$

This mechanism suppresses irrelevant background tissue and emphasizes lesion-relevant regions before reconstruction. The final layer is a 1×1 convolution that produces a binary segmentation mask at 256×256 resolution. DSU-Net contains approximately 17.26 million trainable parameters.

Training uses a hybrid loss function:

$$\mathcal{L} = 0.5 \cdot \mathcal{L}_{\text{Dice}} + 0.3 \cdot \mathcal{L}_{\text{Focal}} + 0.2 \cdot \mathcal{L}_{\text{BCE}} \tag{4}$$

where Dice loss addresses class imbalance, Focal loss emphasizes hard pixels such as lesion boundaries, and BCE provides stable pixel-wise gradients. The individual components are:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2|P \cap T| + \epsilon}{|P| + |T| + \epsilon}, \quad \epsilon = 10^{-6} \tag{5}$$

$$\mathcal{L}_{\text{Focal}} = -\alpha(1 - p_t)^{\gamma} \log(p_t), \quad \alpha = 1,\ \gamma = 2 \tag{6}$$

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} [t_i \log(p_i) + (1 - t_i) \log(1 - p_i)] \tag{7}$$

b) Faster R-CNN Architecture

Faster R-CNN is used as a detection-based baseline for lesion localization. The model follows the standard two-stage design introduced by Ren et al. [25], consisting of a convolutional backbone, a Region Proposal Network (RPN), and a detection head for classification and bounding-box regression. We use a ResNet-50 with FPN backbone, initialized from MS-COCO pretrained weights. The original classification head is replaced with a new predictor for three classes: background, mass, and calcification. All parameters remain trainable during fine-tuning, allowing the model to adapt to mammographic DICOM images.

c) VGG16-U-Net Architecture

We also implement a VGG16-U-Net baseline using an ImageNet-pretrained VGG-16 encoder and a standard U-Net decoder. A 1×1 convolution adapts grayscale mammograms to the three-channel input expected by VGG-16. Skip connections are extracted from the standard VGG blocks, from `block1_conv2` to `block5_conv3`. Unlike DSU-Net, this baseline uses a deeper five-level encoder, a conventional convolutional bottleneck at 8×8 resolution, and no spatial attention gates. Its loss function is the standard combination:

$$\mathcal{L}_{\text{baseline}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}} \tag{8}$$

This model is used to evaluate whether pretrained representations are more robust than a from-scratch dense-attention design under heterogeneous dataset conditions.

Experimental setup

To prepare the data for these architectures, all datasets follow a unified preprocessing pipeline. INbreast images are read from DICOM files using `pydicom`, while MIAS images are loaded as grayscale images. Pixel intensities are normalized to $[0, 1]$ using min-max normalization:

$$I_{\text{norm}} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \tag{9}$$

Images are automatically flipped when necessary to ensure a consistent breast-facing-left orientation, and CLAHE is applied with clip limit 2.0 and tile grid size $(8, 8)$ to improve local contrast. INbreast masks are generated from polygon and point annotations, while MIAS masks are generated as circular masks from the provided center-radius annotations. All images and masks are resized to 256×256, using nearest-neighbor interpolation for masks.

Data augmentations have also been applied to the datasets. These include random horizontal and vertical flips, small rotations, elastic deformation, brightness and contrast adjustment, and Gaussian noise. Validation data receives only

resizing and intensity normalization using ImageNet statistics. All train, validation, and test splits are generated using Scikit-Learn's `train_test_split` utility.

Following this preprocessing phase, the experimental setup evaluates the models under four configurations:

Experiment A: MIAS only DSU-Net is trained on 115 annotated MIAS mammograms with synthetic circular masks, split 80/10/10 into training, validation, and test sets.

Experiment B: INbreast only DSU-Net and Faster R-CNN are trained on curated INbreast image-mask pairs containing mass and calcification annotations, also split 80/10/10.

Experiment C: INbreast + MIAS The two datasets are merged to test whether heterogeneous training data improves segmentation generalization.

Experiment D: INbreast + DMID VGG16-U-Net is trained on a curated combined dataset of 374 images, split into 299 training, 37 validation, and 38 test images.

Segmentation performance is evaluated using Dice coefficient and Intersection over Union (IoU):

$$\text{Dice} = \frac{2|P \cap T|}{|P| + |T|} \tag{10}$$

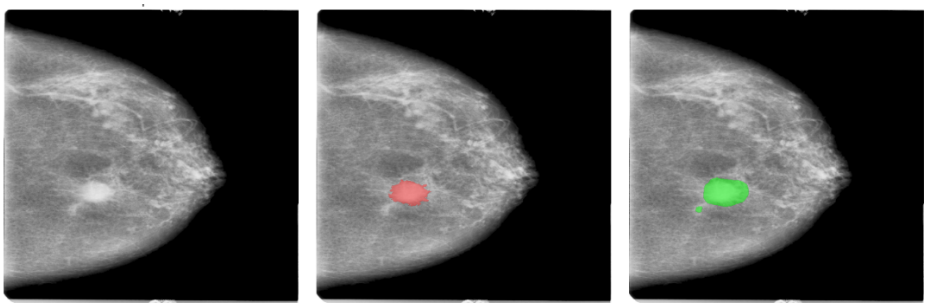
$$\text{IoU} = \frac{|P \cap T|}{|P \cup T|} = \frac{|P \cap T|}{|P| + |T| - |P \cap T|} \tag{11}$$

where P is the predicted binary mask and T is the ground truth mask. Specifically, $|P \cap T|$ represents the accurately segmented lesion pixels (true positives). Visually, if one imagines overlaying the predicted shape directly onto the ground truth, the intersection represents their perfectly shared region, whereas the union encompasses their total combined area. Both metrics quantify this spatial overlap: the Dice coefficient evaluates the average agreement between the masks, while IoU penalizes under- and over-segmentation more severely, offering a stricter assessment of boundary accuracy.

4 Experimental Results

Experiment A: DSU-Net on MIAS Only Training on the MIAS dataset alone presents challenges due to the small dataset size (115 annotated images) and the synthetic nature of the circular masks, which do not capture the true lesion morphology. The model converges but achieves limited segmentation precision, as the circular ground truth masks are inherently imprecise approximations of actual lesion boundaries.

Key observations from this experiment: the model successfully learns to localize breast lesions in the general region, but the predicted masks tend to be circular or overly smooth, reflecting the bias of the training annotations. Fine boundary delineation is limited.



Original Image Ground Truth Prediction

Fig. 1. Example of a predicted mask compared to it's correct shape.

Table 1. DSU-Net segmentation results on MIAS only (Experiment A).

Metric	Validation	Test
Dice Coefficient	0.7512	0.7345
IoU Score	0.6214	0.6054
Training samples	92	
Validation / Test samples	12 / 11	

Experiment B.1: DSU-Net on INbreast Only INbreast provides pixel-level polygon annotations of substantially higher quality than MIAS. Training on INbreast yields more precise segmentation masks, particularly for mass-type lesions that have well-defined contours.

Table 2. DSU-Net segmentation results on INbreast only.

Metric	Validation	Test
Dice Coefficient	0.8842	0.8712
IoU Score	0.7915	0.7842
Training samples	84	
Validation / Test samples	11 / 10	

The model achieves better boundary delineation compared to Experiment A, particularly for larger masses. However, the small training set leads to overfitting tendencies, mitigated partially by the augmentation pipeline and early stopping.

Experiment B.2: Faster R-CNN on INbreast Only Model performance was assessed on the held-out validation set using Intersection-over-Union (IoU) based metrics. A predicted bounding box was considered a true positive if its IoU

with at least one ground-truth box exceeded the specified threshold. We computed precision, recall and F1-score through a matching procedure that prevents multiple predictions from being credited to the same ground-truth annotation.

An experiment with a simpler training configuration presented in table 3 yielded an IoU accuracy of 12.3% at a threshold of 0.5, indicating that the model had not yet converged and highlighting the need for a more carefully designed training strategy.

Table 3. Training configuration for Faster R-CNN on the INbreast dataset.

Hyperparameter	**Value**
Epochs	5
Optimiser	SGD
Learning Rate	5×10^{-3}
Momentum	0.9
Weight Decay	5×10^{-4}
IoU Accuracy ($\tau = 0.5$)	12.3%

Experiment C: DSU-Net on INbreast + MIAS (Combined) The next approach merged the strengths of both sources: the precise annotations of INbreast and the additional volume and diversity of MIAS. Despite the heterogeneity in image acquisition (digital vs. digitized film) and annotation quality (polygons vs. circles), the unified preprocessing pipeline (CLAHE, normalization, orientation correction) brings both sources into a compatible representation. However, this pipeline did not eliminate the domain gap problem, thus leading to the following results:

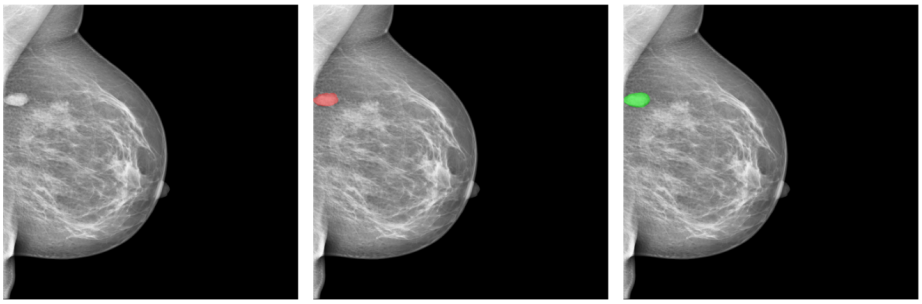
Table 4. DSU-Net segmentation results on combined INbreast + MIAS (Experiment C).

Metric	**Value**
Dice Coefficient	0.1570
IoU Score	0.1127
Loss	0.6169

The conclusion is that pre-processing cannot resolve domain-level and annotation-level conflicts. While the unified pipeline normalizes intensity distributions and orientations, it cannot reconcile the structural difference between FFDM and digitized film acquisition, nor harmonize the semantic inconsistency between precise polygonal (INbreast) and approximate circular (MIAS) annotations. The resulting supervision signal presents contradictory boundary cues for morphologically similar lesions, destabilizing DSU-Net's distribution-based uncertainty

estimation and causing the model to suppress confident predictions. This finding underscores that dataset harmonization for medical image segmentation requires annotation-protocol alignment, not merely intensity normalization, as a prerequisite for effective multi-source training.

Experiment D: VGG on INbreast and DMID dataset Building on the limitations observed with DSU-Net on the combined INbreast-MIAS dataset, we investigated whether a transfer learning approach based on a VGG backbone can better handle cross-dataset heterogeneity. Unlike DSU-Net's uncertainty-guided mechanism, which proved sensitive to inconsistent supervision signals, VGG's pretrained convolutional features offer a more stable representational foundation, potentially mitigating the impact of annotation-protocol mismatches between INbreast's polygonal contours and DMID's annotation conventions. This experiment therefore serves as a comparative baseline to assess whether leveraging ImageNet-pretrained features provides a more robust starting point for multi-source mammographic lesion segmentation than training an uncertainty-aware architecture from scratch on heterogeneous data. This led to a model with an IoU Accuracy = 63%



1. Input image 2. Ground truth 3. Model's output

Fig. 2. Results of VGG16 with DMID and INbreast.

Table 5. Architectural comparison on the INbreast dataset.

Model	Task	IoU (%)	Parameters
VGG16-U-Net (baseline)	Segmentation	63.0	∼25M
DSU-Net (ours)	Segmentation	11.27	∼17.3M
Faster R-CNN†	Detection	12.3	∼41.8M

5 Analysis and Discussion

The experimental configurations reveal a clear pattern regarding the impact of architecture choice, training data, and task formulation on segmentation and detection quality.

MIAS-only training suffers from two compounding limitations: (a) a small annotated set (∼115 images) and (b) imprecise synthetic circular masks that do not capture the true morphological diversity of breast lesions. The model learns a general "detect-and-circle" strategy rather than precise boundary delineation.

INbreast-only training benefits from clinically precise polygon annotations, enabling the model to learn accurate lesion boundaries. However, the small dataset size (∼105 curated pairs) limits the model's exposure to the full spectrum of lesion appearances and tissue backgrounds, leading to overfitting tendencies.

Combined training (DSU-Net, Experiment C) increases both the volume and diversity of training data. However, the heterogeneity of imaging sources and annotation styles introduced contradictory supervision signals that destabilized DSU-Net's uncertainty-guided mechanism, yielding near-failure segmentation performance (IoU: 0.1127, Dice: 0.1570). This demonstrates that dataset volume alone is insufficient when annotation protocols are semantically inconsistent.

Transfer learning with VGG16-U-Net (Experiment D) directly addressed this limitation. By leveraging ImageNet-pretrained convolutional features as a stable representational backbone, the model proved significantly more robust to cross-dataset heterogeneity, achieving an IoU of 63% on the combined INbreast-DMID dataset. Unlike DSU-Net's uncertainty estimation, which collapsed under contradictory boundary cues, VGG's pretrained low- and mid-level features provided a domain-agnostic foundation that partially compensated for annotation-protocol mismatches. This confirms that transfer learning is a more effective strategy than training uncertainty-aware architectures from scratch when training data is heterogeneous.

Detection with Faster R-CNN (Experiment B.2) approached the problem from a fundamentally different angle localizing lesions via bounding boxes rather than delineating precise boundaries. The preliminary configuration achieved only 12.3% IoU accuracy at $\tau = 0.5$, reflecting insufficient training (5 epochs, suboptimal hyperparameters) rather than an inherent architectural limitation. Faster R-CNN's two-stage proposal mechanism is well-suited for lesion localization in high-resolution mammograms, but its ∼41.8M parameter count demands longer convergence times and more carefully tuned learning rate schedules. These results indicate that detection-based approaches require substantially more training investment to be competitive with segmentation models on this task.

The dense block bottleneck provides three key advantages: (1) feature reuse: each layer has direct access to all preceding feature maps; (2) improved gradient flow: short connection paths facilitate backpropagation for subtle mammographic gradients; (3) parameter efficiency: the growth-rate mechanism ($k = 32$) results in 17.3M parameters compared to ∼25M for VGG16-U-Net and ∼41.8M for Faster R-CNN.

The spatial attention gates suppress irrelevant features (e.g., dense glandular tissue, pectoral muscle, scanner artifacts) before concatenation with skip connections, reducing false positive activations. However, as demonstrated in Experiment C, this mechanism is only effective when the supervision signal is consistent.

Visual comparison reveals that the MIAS-only model produces overly rounded predictions; the INbreast-only model achieves more accurate boundary delineation but struggles with small calcifications; the combined DSU-Net fails to produce confident predictions due to domain conflict; and the VGG16-U-Net baseline produces the most practically useful outputs, trading architectural sophistication for representational stability through pretraining.

Several limitations should be acknowledged. The small scale of both INbreast and MIAS constrains model capacity and generalization. Synthetic circular masks in MIAS introduce annotation noise. Resizing to 256×256 causes loss of fine structural detail, particularly for small calcifications. The Faster R-CNN experiment was constrained by computational budget and would benefit from extended training. Direct comparison with published results should be interpreted cautiously, given differences in data splits, preprocessing, and evaluation protocols.

6 Conclusions and Future Work

We have presented a systematic evaluation of multiple deep learning architectures for breast lesion segmentation and detection in mammographic images, comparing DSU-Net, VGG16-U-Net, and Faster R-CNN across heterogeneous multi-source datasets. Through experiments on MIAS, INbreast, and DMID, we have demonstrated that:

- **DSU-Net's** dense block bottleneck and attention-gated decoder provide an effective framework for mammographic lesion segmentation when trained on clean, consistent annotations, offering improved feature reuse and selective focus on lesion regions compared to a standard VGG16-U-Net baseline.
- **Cross-dataset heterogeneity** poses a fundamental challenge that preprocessing alone cannot resolve. When annotation protocols are semantically inconsistent, as in the INbreast + MIAS combination, even well-designed architectures like DSU-Net fail to produce reliable predictions, while transfer learning with VGG16 proves more resilient due to its stable pretrained representations.
- **Transfer learning** via a VGG16 backbone significantly outperforms training from scratch on heterogeneous data, achieving 63% IoU on INbreast + DMID, underscoring the importance of representational stability over architectural sophistication in low-data, multi-source regimes.
- **Detection-based approaches** (Faster R-CNN) require substantially more training investment to be competitive, but represent a complementary direction for lesion localization rather than boundary delineation.

These findings suggest that annotation-protocol alignment is as critical as architectural design for multi-source medical image segmentation, and that transfer learning is a practical and effective strategy when training data is limited and heterogeneous.

Future work should prioritize the following directions:

- **Larger-scale training:** Incorporating CBIS-DDSM and its curated ROI masks would substantially increase training data volume and annotation consistency, particularly for micro-calcification cases.
- **Higher resolution:** Training at 1024×1024 or higher would preserve fine structural details critical for calcification segmentation, at the cost of increased computational requirements.
- **Advanced architectures:** Exploring transformer-based encoders (e.g., Swin Transformer) and hybrid CNN-Transformer designs could capture long-range spatial dependencies beneficial for whole-mammogram segmentation.
- **Clinical validation:** Evaluating on multi-institutional external datasets and measuring clinically relevant metrics (sensitivity at fixed specificity, lesion-level detection rate) would assess real-world applicability.
- **Multi-task learning:** Jointly training for segmentation and classification (benign vs. malignant) could enable a more complete CAD system and improve segmentation through shared feature representations.

References

1. Sung, H., Ferlay, J., Siegel, R.L., et al.: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians **71**(3), 209-249 (2021)
2. Institutul Naţional al Sănătăţii Publice (INSP): Analiza de situaţie - Cancerul de sân. INSP Romania Report (2020)
3. European Commission, ECIS: European Cancer Information System: Cancer survival estimates (2022)
4. Sutjiadi, R., Sendari, S., Wahyu, H.H., et al.: Deep Learning for Segmentation and Classification in Mammograms for Breast Cancer Detection: A Systematic Literature Review. Advanced Ultrasound in Diagnosis and Therapy **8**(3), 94–105 (2024)
5. Defined, L., Defined, S., et al.: Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. Diagnostics **14**(8), 848 (2024)
6. Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C., Elmaghraby, A.S.: Connected-UNets: A Deep Learning Architecture for Breast Mass Segmentation. npj Breast Cancer **7**(1), 151 (2021)
7. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: Toward a Full-Field Digital Mammographic Database. Academic Radiology **19**(2), 236–248 (2012)
8. Suckling, J., Parker, J., Dance, D., et al.: The Mammographic Image Analysis Society Digital Mammogram Database. International Congress Series **1069**, 375–378 (1994)

9. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gober, M., Rubin, D.L.: A Curated Mammography Data Set for Use in Computer-Aided Detection and Diagnosis Research. Scientific Data **4**, 170177 (2017)
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: MICCAI 2015, LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015)
11. 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 (CVPR), pp. 4700–4708 (2017)
12. Ravitha Rajalakshmi, N., Vidhyapriya, R., Elango, N., Ramesh, N.: Deeply Supervised U-Net for Mass Segmentation in Digital Mammograms. International Journal of Imaging Systems and Technology **31**(1), 59–71 (2021)
13. Elkorany, A.S., Elsharkawy, Z.F.: Automated Abnormalities Detection in Mammography Using Deep Learning. Complex & Intelligent Systems **10**, 7279–7296 (2024)
14. Ahmad, A., et al.: A Comprehensive CAD System Based on Deep Learning for Mammogram Classification and Breast Cancer Detection. IEEE Access **12**, 45123–45138 (2024)
15. Ansari, G.A., et al.: A Multi Stage Deep Learning Model for Accurate Segmentation and Classification of Breast Lesions in Mammography. Scientific Reports **15**, 21146 (2025)
16. Li, X., et al.: Enhanced Breast Mass Segmentation in Mammograms Using a Hybrid Transformer UNet Model. Biomedical Signal Processing and Control **98**, 106780 (2024)
17. Anari, S., et al.: Explainable Attention Based Breast Tumor Segmentation Using a Combination of UNet, ResNet, DenseNet, and EfficientNet Models. Scientific Reports **15**, 1124 (2025)
18. Matheus, B.R.N., Schiabel, H.: Online Mammographic Images Database for Development and Comparison of CAD Schemes. Journal of Digital Imaging **24**(3), 500–506 (2011)
19. Nguyen, H.T., et al.: VinDr-Mammo: A Large-Scale Benchmark Dataset for Computer-Aided Detection and Diagnosis in Full-Field Digital Mammography. Scientific Data **10**, 277 (2023)
20. Salama, W.M., Aly, M.H.: Deep Learning in Mammography Images Segmentation and Classification: Automated CNN Approach. Alexandria Engineering Journal **60**(5), 4701–4709 (2021)
21. Farina, E., Nabhen, J.J., Dacoregio, M.I., Batalini, F., Moraes, F.Y.: An Overview of Artificial Intelligence in Oncology. Future Science OA **8**(4), FSO787 (2022)
22. Shimizu, H., Nakayama, K.I.: Artificial Intelligence in Oncology. Cancer Science **111**(5), 1452–1460 (2020)
23. Luchini, C., Pea, A., Scarpa, A.: Artificial Intelligence in Oncology: Current Applications and Future Perspectives. British Journal of Cancer **126**(1), 4–9 (2022)
24. Zuiderveld, K.: Contrast Limited Adaptive Histogram Equalization. In: Graphics Gems IV, pp. 474–485. Academic Press Professional (1994)
25. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)