SJER Scientific Journal of Engineering Research

futuristech PT. Teknologi Futuristik Indonesia

Date of publication May 6, 2026, date of current version May 6, 2026.
e-ISSN: 3109-1725
Digital Object Identifier 10.64539/sjer.v2i3.2026.489

Article

BReMS-Net: Prediction-Guided Coarse-to-Fine Refinement with Boundary-Aware Multi-Scale Dilated Fusion for Robust Breast Mass Segmentation

Tayyba Sarfraz[1,*], Tan Ling[1], Ahmad Ijaz[1]

[1] Department of Computer Science and Technology, School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, 210044, China; e-mail: tayybasarfraz2001@gmail.com (T. Sarfraz), cillatan0@nuist.edu.cn (T. Ling), officialahmed000@gmail.com (A. Ijaz).
* Correspondence Author

The authors received no financial support for the research, authorship, and/or publication of this article.

Abstract: Breast masses in mammograms are important to segment for computer-aided diagnosis (CAD) to enhance early detection and treatment decisions. Current approaches face challenges in segmenting lesions with low lesion-to-tissue contrast and diverse textures, resulting in misclassification or poor segmentation accuracy. To overcome this challenge, this paper introduces BReMS-Net, a multi-stage segmentation network to improve contextual learning and refined boundaries. We used an MBA-Net backbone with two major components: a Multi-scale Hybrid Dilated Convolution (MHD) module to extract multi-scale contextual features, and a Boundary Feature Auxiliary (BFA) module to strengthen boundary representations via coarse-to-fine feature fusion. Furthermore, a lightweight Prediction-Guided Refinement Module (PRM) uses initial predictions to produce attention maps, remove background clutter, and progressively refine boundary areas. The model has been evaluated on a cross-dataset basis, trained on the CBIS-DDSM dataset and tested on the INbreast dataset, and the results show that the BReMS-Net produces a Dice coefficient of 93.12% and an HD95 of 0.9826, which demonstrate competitive performance compared to several state-of-the-art deep learning methods. These results underline its generalization and robustness. Overall, the framework provides a robust and efficient approach to breast mass segmentation and has important implications for the performance and clinical relevance of automatic breast cancer diagnosis systems.

Keywords: Breast mass segmentation; Mammography; Multi-stage deep learning; Hybrid dilated convolution; Computer-aided diagnosis.

1. Introduction

Breast cancer is one of the most frequent cancers in women, and it is a major cause of cancer related deaths in the world. Population-based screening programs are important to reduce mortality, as the early detection and timely interventions significantly impact patient outcomes [1]. Mammography is the most common screening modality since it has the potential to expose suspicious findings before clinical manifestations [2], [3]. Radiologists determine masses in clinical interpretation based on their size, internal density patterns, shape, and margin [4]. Mass boundary delineation is highly needed, as visual features have a high correlation with malignancy risk and directly affect the diagnostic confidence and subsequent treatment planning [5]. Nevertheless, manual annotation is time consuming, labor intensive, and prone to inter-observer variability. Thus, the design of strong automated breast mass segmentation algorithms is encouraged [6]. With these advancements, mammography mass segmentation is still very difficult due to several factors: (i) lesions may be blurred, irregular, or spiculated edges (ii) lesion textures may be similar to normal tissue, especially in dense breasts, and (iii) the foreground mass region usually occupies only a small portion of the image, causing a very large class imbalance [7], [8]. Therefore, this can result in over smoothing contours, spiked edges, and loss of detections with single-stage models, particularly in low contrast or small lesion cases [9]. Despite recent advances, breast mass

segmentation remains a significant challenge due to the combined impact of low contrast, complex tissue structures, and severe class imbalance. These challenges prevent current methods from providing accurate and clinically meaningful boundary segmentations.

Existing deep learning frameworks, such as single-stage encoder-decoder networks, struggle to effectively combine multi-scale contextual information with fine-grained boundary predictions. In particular, the lack of multi-scale contextual representation results in poor feature understanding in complex mammograms, and weak boundary modeling leads to inaccurate and inconsistent lesion boundaries. Moreover, the post-processing steps in existing methods often fail to refine the estimated segmentation result, especially in areas of low contrast. This results in discontinuities and classification errors in complex mammograms.

To address these constraints, we suggest an advanced breast mass segmentation system based on MBA-Net [10] and a Prediction-Guided Refinement Module (PRM). Firstly, the backbone network learns hierarchical features of mammograms by residual encoding with Multi-scale Hybrid Dilated Convolution (MHD) modules to expand the receptive field to obtain contextual information at different scales. To enhance the boundary representation, the Boundary Feature Auxiliary (BFA) module is proposed to provide the decoding process with multi-granularity edge cues. Nevertheless, with these complicated textures and low contrast of mammograms, the fused prediction can still have rough or jagged edges. Thus, a lightweight Prediction-Guided Refinement Module is proposed to introduce refinement to the fused segmentation prediction, creating an attention map based on the coarse output and steering the network to concentrate on the regions with lesions. The refinement mechanism enhances the continuity of contours and limits the number of segmentation errors that result from unclear bounding pixels.

In order to further reduce foreground-background imbalance and enhance stability of optimization, we use an Adaptive Weight Optimization loss (AWloss), which dynamically focuses on hard samples when training. Experiments using the publicly available CBIS-DDSM dataset show that the proposed framework improved boundary delineation and competitive segmentation compared to the representative baseline and more recent algorithms in the segmentation task. Unlike existing encoder-decoder extensions that rely solely on architecture depth or attention, the proposed framework explicitly comprises the coarse-to-fine refinement, boundary-aware feature modulation, and adaptive weight loss in a single optimization problem. The innovation is not just in architectural elements but in the integrated multi-scale context continuity, explicit supervision of the boundaries, and the stages of the error correction specific to mammography qualities.

To address these challenges, this work presents a unified framework with the following contributions:

- A multi-stage breast mass segmentation framework, BReMS-Net, is proposed based on an MBA-Net backbone, integrating boundary-aware learning and coarse-to-fine refinement for robust mammographic lesion delineation.
- A Multi-scale Hybrid Dilated Convolution (MHD) module is designed to systematically enlarge the receptive field, enabling rich multi-scale contextual feature extraction without increasing computational overhead.
- A Boundary Feature Auxiliary (BFA) module is introduced to strengthen lesion boundary representations through hierarchical multi-granularity edge feature interaction between coarse and fine-grained feature maps.
- A lightweight Prediction-Guided Refinement Module (PRM) is proposed, which generates an attention map from the coarse segmentation output to iteratively suppress background interference and correct misclassified boundary pixels, yielding precise contour delineation.

The remainder of this paper is organized as follows. Section 2 presents the related work. Section 3 describes the BReMS-Net architecture and approach. Section 4 describes the results and comparisons. Section 5 reports the ablation study. Section 6 describes the discussion. Section 7 describes the limitations and future work. Finally, Section 8 concludes the paper.

2. Related Work

The segmentation of the mass in mammography is a critical issue in computer-aided diagnosis, as it determines accurate boundaries that allow clinical assessment of malignancy related symptoms, including speculation, margin irregularity, and shape distortion [11], [12]. Nonetheless, the large amounts of glandular tissue, low lesion to background contrast, noise during acquisition, and a large range of mass size and appearance degrade the performance of segmentation [13], [14]. These reasons have led to the research not only seeing through the normal thresholding and region growing pipelines but also deep learning methods, which develop hierarchical representations, and more recently hybrid architectures that combine convolutional feature learning with attention mechanisms [15], [16]. Moreover, due to the high cost of pixel level labeling and the need to be reliable during screening procedures, annotation efficient learning and clinically oriented research have attracted interest [17], [18].

DVF-YOLO-Seg, which is proposed by Abudukelim et al. [19], is a two-stage mammography pipeline that is aimed at enhancing breast mass segmentation, especially of small, low contrast lesions commonly associated with overcoming single-stage methods. YOLOv10 is improved

with the help of DualConv feature extraction and Varifocal Loss. The boxes of the detector are calculated to perform VRP-SAM to mask refinement, report Dice 0.802, F1 0.806 on the CBIS-DDSM mass subgroup, but results on lesions smaller than 5 mm fall (Dice 0.750), indicating that the toughest clinical cases remain tough. Moreover, there are performance constraints due to two-stage dependency and single dataset 2D mammogram evaluation, and the authors note that there may be constraints to performance in terms of deploying costs of the system and the residual failure modes, resulting in sub validation of generalizability and clinical implications in practice.

Wang et al. [20] introduce Bi-CBMSegNet, which is a lightweight encoder-decoder architecture that can be used in challenging mammography cases like dense tissue and low contrast. The method gives especially high Dice scores on publicly available datasets with desirably good computing performance by fusing global and local enhancement modules to reason about long range context and narrow boundary regions, respectively. The reported findings could be enriched by more intensive studies of ablation and external validation in order to prove the clinical soundness, and generalization to different methods of acquisition, social groups, or severe cases of dense breasts have not been examined completely.

Shamrat *et al*. [21] introduce MammoSegNet to overcome a widespread limitation of mammography CAD research. Accurate separation of lesions is challenging, especially in dense tissue, low contrast, and edge tumors, and many deep baselines are expensive to implement. In order to increase the number of multi-scale features and localize the border, they append PFIT augmentation to an Inception ResNet based CNN, which uses Inception ResNet, squeeze and excitation attention, and dilated convolutions. They show high segmentation with BCDR-D01 to train/stratify 5-fold validation, as well as INbreast to test across the dataset. Nevertheless, the work is constrained by dataset accuracy, resizing, computing limits, false boundary errors, and ineffective external baselines, which partially affect the more global SOTA comparison.

In general, the current literature has revealed that hybrid CNN-Transformer architectures, lightweight global-local augmentation models, and more focused learning and clinical preparedness are all leading to more precise segmentation outcomes [22], [23]. Among the challenges that have remained in mammography segmentation are boundary over smoothing, partial contour reconstruction due to dilated convolution designs, poor modeling of smooth intensity changes at lesion edges, and extreme class imbalance, which leads to false detections and biased training [24], [25].

The current segmentation approaches show significant advancements in their attention mechanisms, hybrid CNN-Transformer frameworks, and multi-resolution feature learning [26], [27]. Numerous methods continue to have issues of over smoothing of boundaries, discontinuities due to rough dilation, and sensitivity to class imbalance and poor cross-dataset validation [28]. These limitations drive the proposed model that integrates a two-stage refinement plan and multi-scale context modeling, boundary sensitive feature direction, and hard sample based optimum to generate more precise and clinically consistent breast mass delineation in mammography [29], [30].

3. Methods

3.1. Overall Architecture of the Proposed Framework

The proposed BReMS-Net framework is based on an improved MBA-Net architecture to achieve correct segmentation of breast mass in mammography. As shown in Figure 1, the network has an encoder-decoder architecture with a modified version of the ResNet-50 backbone to generate hierarchical features of the input mammogram. Multi-scale Hybrid-Dilated Convolution (MHD) in the encoder is also added to supplement the representation of contextual features and increase the size of the receptive fields, only to keep its spatial continuity. Boundary Feature Auxiliary (BFA) modules are included in skip connections to deliver multi-granularity edge guidance to the decoder and enhance the reconstruction of contours with upsampling. Coarse segmentation prediction is obtained after feature fusion at multi-level. To enhance the accuracy of boundaries and eliminate possible misclassifications, the Prediction-Guided Refinement Module (PRM) is used on the fused prediction.

The PRM applies the coarse prediction to create an attention map, which actively identifies the lesion-related areas and prevents the irrelevant background responses so that the final segmentation mask produced by the network can be more accurate. The proposed architecture is based on the coarse-to-fine segmentation strategy, where the initial prediction produced by the backbone network is optimized by the PRM to enhance the boundary accuracy and reduce the segmentation errors. The PRM uses the coarse segmentation prediction obtained from the multi-level output fusion stage to guide feature refinement.

First, a coarse prediction map is generated by a 1x1 convolution and a sigmoid activation function. This coarse prediction is then used to generate an attention map by performing a 3x3 convolution operation. The generated attention map indicates possible areas of lesions and inhibits unnecessary background reactions. Multiplication of the fused feature representation with the attention map is used to enable the network to focus on lesion related regions. The refined features are further passed through convolution layers to give the resulting final segmentation output. Let F represent the fused feature representation that was derived by using the multi-level output integration module. The rough forecast P_c is obtained as:

$$P_c = \sigma\left(\mathrm{Conv}_{1\times1}(F)\right) \quad (1)$$

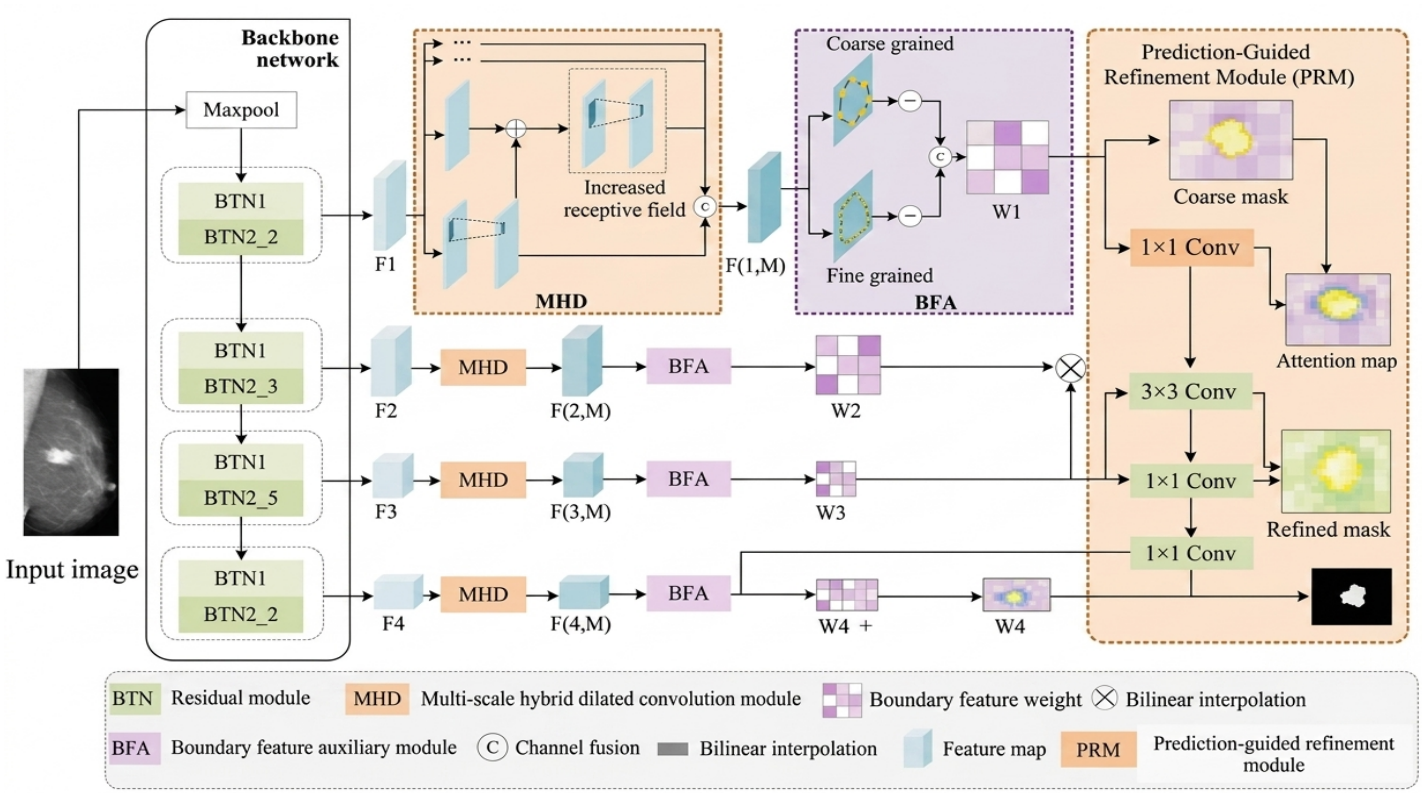

Figure 1. Overview of the Proposed Framework Architecture.

where σ denotes the sigmoid activation function. The attention map A is obtained as:

$$A = \sigma\left(\mathrm{Conv}_{3\times3}(P_c)\right) \tag{2}$$

The refined feature representation F_r is computed using element-wise multiplication.

$$F_r = F \otimes A \tag{3}$$

Finally, the refined segmentation output P_f is produced as:

$$P_f = \sigma\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Conv}_{3\times3}(F_r)\right)\right) \tag{4}$$

3.2. Multi-Scale Hybrid Dilated Convolution (MHD) Module

We use the MHD module in the proposed model to gain multi-scale context with the preservation of spatial relationships. The module augments receptive fields with multi-scale parallel convolution with dilation rates in a sawtooth format and cross-path fusion of features between neighboring branches in order to decrease discontinuities and avoid duplicate processing. The input feature map F_s is categorized into four groups based on the number of channels, with a 1x1 convolutional layer to expand the nonlinear features to enhance the network's expression ability, and the four sub-feature maps x_s^j would be formed. The sub-features of each group can be defined as:

$$x_s^j = Conv_{1\times1}\left(Divide(F_s)\right), j = 1,2,3,4 \tag{5}$$

where F_s refers to the initial feature map of the sth layer Stage, x_s^j refers to the jth sub-feature map, Conv 1x1 refers to the convolution operation 1x1, Divide refers to dividing the feature map by the number of channels.

The former is a convolution (dilation 1) of 3 by 3:

$$y_s^1 = Conv_{3\times3}, (x_s^1) \tag{6}$$

In order to maintain local continuity but expand the receptive field, both the second and third branches take the preceding branch output and carry out dilated convolution.

$$y_s^2 = dConv_{3\times3}^{r=2}, (x_s^2 + y_s^1) \tag{7}$$

$$y_s^3 = dConv_{3\times3}^{r=3}, (x_s^3 + y_s^2) \tag{8}$$

where, y_s^i represents the ith sub-output feature maps, $Conv_{3\times3}$ represents a 3x3 convolution, $dConv_{3\times3}^{r=2}$ and $dConv_{3\times3}^{r=3}$ represents the dilation convolution operation using expansion coefficients 2 and 3, respectively. The fourth branch maintains fine details because it does not transform the sub-feature map further, but only retains the initial sub-feature maps x_s^4 the better feature map is produced as a concatenation and fusion of the results of multiple receptive fields through 1x1 convolution:

$$F_{(s,M)} = Conv_{1\times1}, \left(Concat(y_s^1, y_s^2, y_s^3, x_s^4)\right) \quad (9)$$

$F_{(s,M)}$ represents the feature map of the sth layer with the MHD module, and Concat represents the operation of channel splicing. The design provides a compromise between local and contextual information, enhancing the network recognition of lesions at different scales and reducing the negative impact of high levels of dilation.

3.3. Boundary Feature Auxiliary (BFA) Module

The BFA module based on the proposed model is used to capture and transfer multi-granularity edge cues with the aim of enhancing the representation of boundaries. The enhanced feature map $F_{(s,M)}$ derived from the MHD module is passed to the BFA module, which employs maximum pooling as a different granularity to model mass areas at different sizes, and boundary cues are acquired by deriving pooled responses of the original feature map.

$$f_{s,border}^{j} = F_{(s,M)} - MAX_j, \left(F_{(s,M)}\right), j = 1,2 \quad (10)$$

$F_{(s,M)}$ represents an input feature map of the BFA module at the sth layer, $f_{s,border}^{j}$ as an edge feature map of the output of $F_{(s,M)}$, after the jth maximum pooling operation, and MAX_j is a maximum pooling operation having varying granularities. The resulting features of edges are concatenated to make boundary weights, which are compressed to the channel dimension of the decoder features, followed by a sigmoid activation:

$$W_s = \sigma, \left(Conv_{1\times1}, \left(Concat\left(f_{s,border}^{1}, f_{s,border}^{2}\right)\right)\right) \quad (11)$$

where W_s is the weight of the sth layer on the boundary feature and σ denotes the sigmoid activation function. These boundary weights are added to the decoder restoration process to emphasize areas of contours and compensate for the detail lost in upsampling. In BReMS-Net, BFA is used at both levels to offer boundary guidance when making Coarse predictions and correcting refinements.

3.4. Adaptive Weight Optimization Loss (AW-loss)

In order to address these issues, we adopt the Adaptive Weight Optimization loss (AW-loss) in the proposed model. AW-loss is an implementation of a multi-output supervisory loss, a loss that is based on focal loss to prioritize hard pixels. Assuming a pixel has the predicted probability p and label n ∈ {0,1} the probability of correct prediction is:

$$p_{\text{true}} = \begin{cases} p, & n = 1 \\ 1 - p, & n = 0 \end{cases} \quad (12)$$

and the focal loss is:

$$FL(p_{true}) = -(1 - p_{true})^{\gamma} \log(p_{true}) \quad (13)$$

where γ is the parameter of focusing $\gamma = 2$. When larger contributions to the loss on misclassified pixels are used, the model tries to focus on the harder regions, especially towards boundaries. Deep supervision is implemented by attaching output heads to multiple stages of decoding. Let FL_q denote the focal loss of the qth output head, and AW-loss gives an adaptive weight to every output in proportion to its relative loss.

$$w_q = \frac{FL_q}{\sum_{q=1}^{N} FL_q} \quad (14)$$

and the final loss is calculated as:

$$AWloss = \sum_{q=1}^{N} w_q \; FL_q \quad (15)$$

where N is the number of supervised outputs.

3.5. Dataset and Preprocessing

The experiment was performed on the public CBIS-DDSM mammography dataset [31], which is a subset of DDSM of 2,478 cases and 1,696 annotated breast masses on craniocaudal (CC) and mediolateral oblique (MLO) views and pixel-wise binary masks. This dataset is also popular to be taken as a reference for evaluating mass segmentation performance under clinically relevant variations of tissue appearance and lesion morphology. All mammograms were initially in the DICOM format, but they were converted to a single image representation. Histogram equalization was done to increase lesion to background contrast. Due to different dimensions of the images, mammograms and respective masks were zero-padded to square dimensions by centering the breast part, which gave equal input on both phases of the proposed framework.

The dataset was divided into training and testing groups, with 1,357 and 339 images, respectively. Figure 2 illustrates the proposed training and evaluation pipeline.

In order to provide an early estimate of the cross-dataset generalization, the trained model was also tested on a subset of the INbreast dataset [32] without fine-tuning. Only preprocessing alignment was done due to the dissimilarity in the acquisition protocol and annotation format. Moreover, a statistical analysis of the dataset shows a large class imbalance between the lesion and background. The average percentage of pixels corresponding to lesion regions in the dataset is less than 5-10% of the total image pixels, with the remaining area representing background tissue. Also, the dataset displays significant variability in lesion size, shape, and boundary features, from small areas of low contrast to large, irregular, and spiculated lesions.

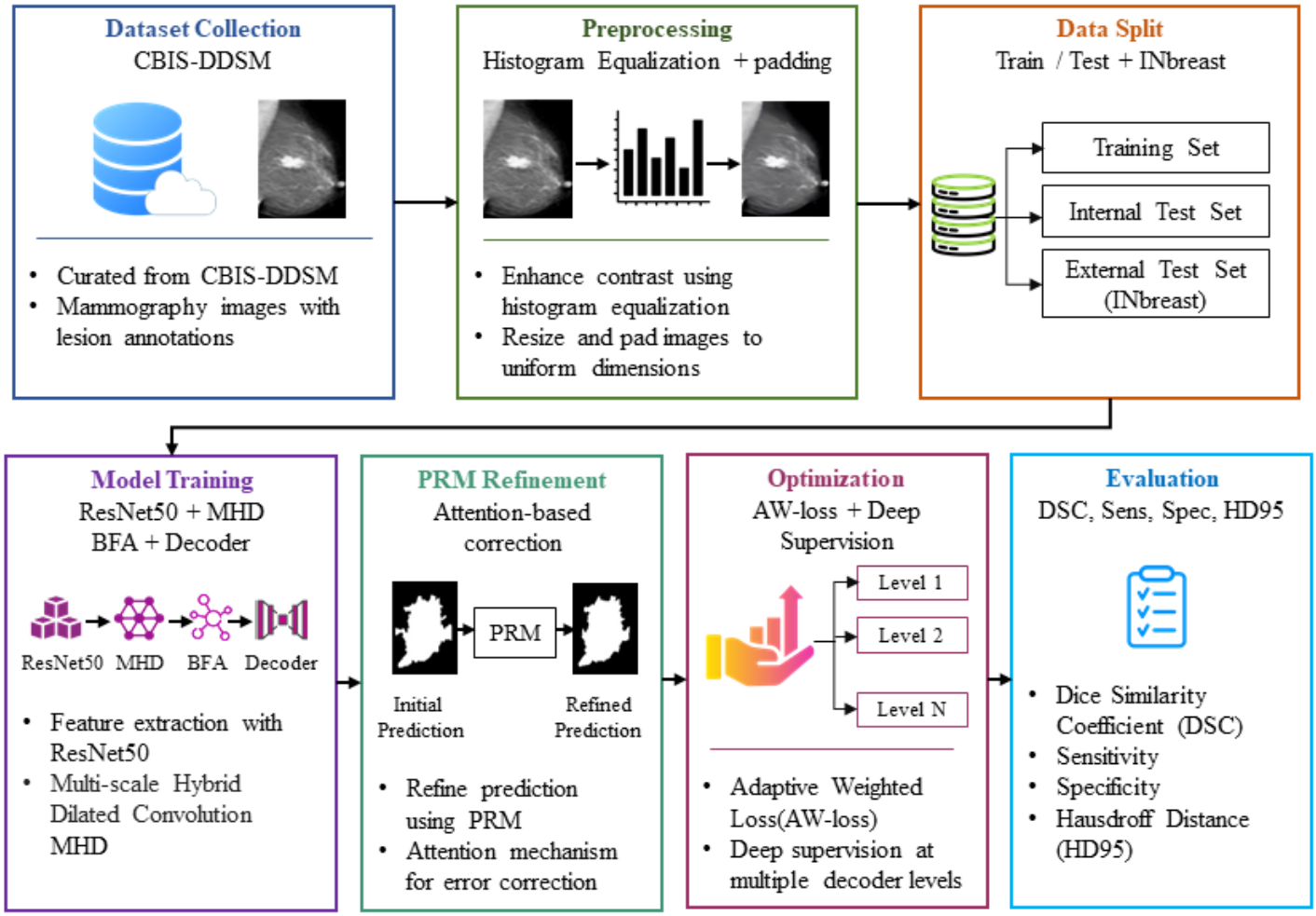

Figure 2. Training and Evaluation Pipeline.

Table 1. Training Hyperparameters.

Hyperparameter	Value
Hardware	NVIDIA GeForce RTX 2080 Ti
Initial Learning Rate	1×10^{-4}
Learning Rate Decay	Exponential (0.9)
Optimizer	Adam
Loss Function	AW-loss (Focal loss, $\gamma = 2$)
Deep Supervision	5 output heads
Train/Test Split	1357 / 339 (8:2)
Preprocessing	Histogram equalization, zero-padding

This variability underlines the difficulty of the segmentation task and motivates the use of deep multi-scale feature extraction and boundary refinement methods.

3.6. Evaluation Metrics

We report region-based and boundary-based metrics to comprehensively analyze the segmentation performance of the proposed framework on CBIS-DDSM. Besides sensitivity and specificity as the measures of lesion detection consistency and the background discrimination, we also use the Dice Similarity Coefficient (DSC) to measure overlap between predicted masks and ground truth, and the 95% Hausdorff Distance (HD95) to measure the accuracy of borders. The sensitivity is the percentage of true lesion pixels that are correctly identified, where Specificity is used to measure the percentage of true negatives and represents the ability to reduce false positives in normal tissue. The Dice Similarity Coefficient (DSC) measures the overlap between predicted and actual positive samples by comparing their twice intersection to the total number of positive elements. These measures are determined as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

$$Specificity = \frac{TN}{TN + FP} \tag{17}$$

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{18}$$

3.7. Details of Training

The BReMS-Net model has been trained on an NVIDIA GeForce RTX 2080 Ti with an initial learning rate 1×10^{-4} with exponential decay (0.9). The model was trained with the Adam optimizer with a batch size of 8 for 100 epochs. The training hyperparameters are listed in

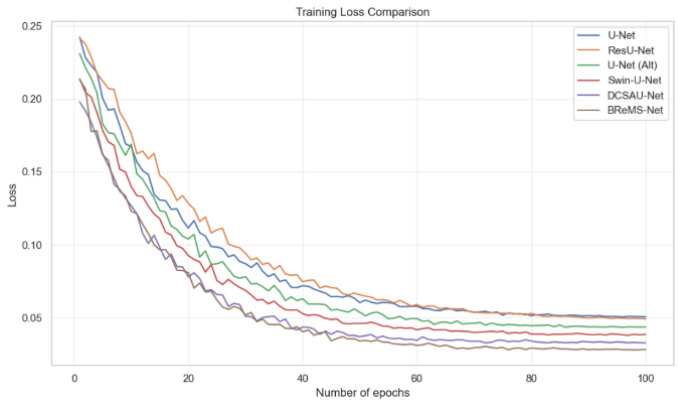

Figure 3. Convergence Loss Comparison across Segmentation Models.

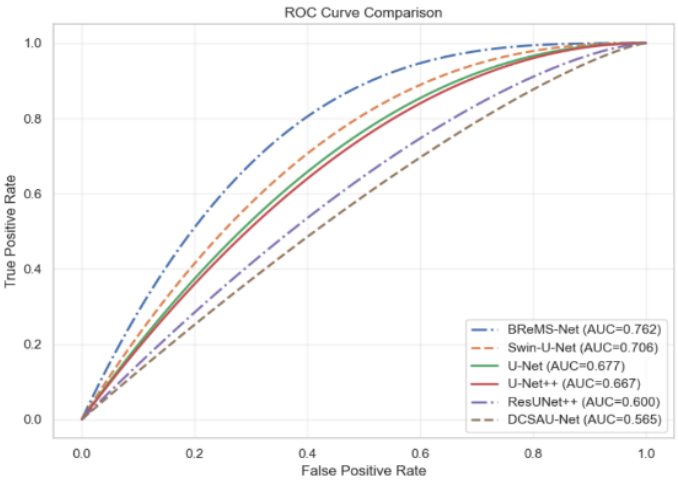

Figure 4. Comparison of Area under the ROC Curve.

Table 2. Comparison of Segmentation Methods on CBIS-DDSM.

Model	Sens. (%)	Spec. (%)	DSC (%)	HD95
U-Net	82.10	79.50	80.30	2.4591
U-Net++	85.40	84.20	84.60	1.9573
ResUNet++	84.10	83.00	82.90	1.6516
Swin-UNet	87.20	85.80	86.50	1.3042
DCSAU-Net	88.50	90.20	89.10	1.4597
MBA-Net	89.20	88.70	88.90	1.2538
BReMS-Net	**91.40**	**90.85**	**93.12**	**0.9826**

Table 1. It used AW-loss with a focal term $\boldsymbol{\gamma = 2}$ and deep supervision over five outputs to deal with class imbalance and focus on hard pixels. It was trained using the CBIS-DDSM dataset with an 8:2 split (1357/339), applying histogram equalization and zero-padding for preprocessing.

4. Experimental Results and Analysis

4.1. Comparative Experimental Evaluation

In order to prove the effectiveness and competitiveness of the proposed model, comparison experiments were conducted with CBIS-DDSM with the help of the following dataset preparation. We compared the proposed approach on the popular medical image segmentation baselines, including DCSAU-Net, U-Net, U-Net++, ResUNet++, and Swin-UNet. In addition, we also provide results of the single-stage MBA-Net backbone to isolate the gain of the proposed multi-stage refining technique. Table 2, shows the quantitative analysis between the proposed framework and some state-of-the-art segmentation models on the CBIS-DDSM dataset. The proposed multi-stage integration has more advantages in comparison to the single-stage MBA-Net. Although the current MBA-Net is already presented with multi-scale context extraction (MHD), boundary-aware decoding (BFA), and imbalance sensitive optimization (AW-loss), the optimization stage of BReMS-Net reduces residual segmentation error by directly conditioning on coarse prediction.

This method of correction restores missed boundary regions, which are usually ambiguous after a single forward pass, improves the consistency of contours, and reduces the number of false positives in dense tissue. Figure 3 illustrates a convergence loss comparison across different segmentation models. Although the proposed modules allow architecture to reach more uniform values of losses than the underlying ResNet baseline, with an improved behavior in the class imbalanced optimization, all models exhibit a steep loss reduction during the initial epochs. Specifically, the proposed BReMS-Net structure ensures the minimal loss during the convergence, meaning that the multi-scale context modeling and the boundary-guided decoding synergize their efforts to enhance the feature learning and reduce training instability.

In terms of model design, U-Net and ResUNet largely rely on local processing and may not capture enough global information in whole mammograms, resulting in over-smoothed edges and poor lesion/background contrast. Despite the fact that U-Net+ enhances the multi-scale fusion with the assistance of layered skip connections, the presence of weakly contrasted lesions may lead to ambiguous outlines. Transformer-based Swin-UNet can improve global reasoning, although when fine-grained cues of a contour are not explicitly maintained, it can lead to boundary effects. DCSAU-Net is very effective in integrating low and high-level semantics, but in the case of lesion texture that is extremely near the background tissue, it could perform poorly, particularly in challenging dense breast cases. However, BReMS-Net incorporates multi-scale contextual continuity and explicit boundary weighting as well as refinement-based error correction, leading to increased robustness in the mammography's particular problems.

Based on the ROC curves in Figure 4, BReMS-Net has the most competitive performance of the compared methods with an AUC of 0.762, which means that it maintains the most desirable sensitivity-specificity trade-off at all the thresholds. However, AUC values of U-Net, U-Net++, ResUNet++, Swin-UNet, and DCSAU-Net are lower, which indicates decreased separability of mass tissue and

Figure 5. Segmentation results of different methods.

Table 3. Effect of Different Modules on Mass Segmentation.

Model	Sens. (%)	Spec. (%)	DSC (%)	HD95
ResNet	75.52	82.32	79.12	2.4308
ResNet + MHD	81.25	87.28	83.73	2.0546
ResNet + BFA	86.13	88.27	87.02	1.3858
ResNet + MHD + BFA	89.42	91.30	90.24	1.2232
Proposed Model	**91.40**	**90.85**	**93.12**	**0.9826**

background tissue using the same evaluation setting. The ROC performance of BReMS-Net is explained by its better feature representation, with multi-scale context extraction and boundary-guided decoding, which reinforce the characterization of the lesions, and AW-loss, which further stabilizes the optimization under extreme class imbalance conditions by focusing on the challenging pixels. To qualitatively compare the segmentation behavior with the clinical conditions in practice, CBIS-DDSM cases were chosen and plotted. This instance indicates common failure modes of mammography. In this case, we provide the cropped mammogram area, ground truth mask, and predicted masks of all the methods.

Case (jagged / spiculated boundary): As shown in Figure 5, the cropped mammogram and its reference annotation are shown in (a) and (b), respectively. The prediction of the U-Net in (c) has severe boundary over-smoothing and does not retain spiculated margin features. In (d), the U-Net++ still recovers a more realistic outline, but still sharp protrusions get smoothed, and in (e), ResUNet++ still emphasizes a smoother and more regularized contour that eliminates all minor irregularities. The output of Swin-UNet in (f) preserves the morphology of the dominant lesion but slightly absorbs the fine serrations into the interior of the lesion, whereas in (g) the DCSAU-Net preserves more local structure but has remaining contour discontinuities. Conversely, in the (h), the proposed BReMS-Net model produces the more consistent delineation, with spiculated margin properties and better contour coherence, thus exhibiting the competitive agreement with the ground truth of all the considered methods in this tricky situation.

5. Ablation Analysis

To check the contribution of each component in the proposed BReMS-Net, we performed an ablation experiment on CBIS-DDSM. ResNet-50 is just an encoder that is used to extract features and needs to be coupled with a decoder to recover spatial resolution in order to make pixel-wise predictions. Consequently, every ablation condition utilized the identical decoder design that had prediction heads and bilinear upsampling, but only the desired modules were switched on or off so that all variants remained consistent. First, we used the ResNet-50+ decoder to construct a one-stage baseline. The MHD, BFA, and AW-loss were subsequently introduced with the original MBA-Net design to estimate their individual and combined impact on segmentation performance. Finally, the incorporation of the Prediction-Guided Refinement Module (PRM) enhances the performance of segmentation towards better refinement of the coarse prediction maps through attention-guided feature enhancement to obtain better lesion delineation and minimize segmentation errors.

Table 3 summarizes the effect of different modules on mass segmentation performance. On the whole, the default ResNet model is not good, especially in terms of sensitivity, proving that a fixed receptive field and the lack of context modeling may lead to the absence of mass regions. MHD improves the detection and overlap through the provision of contextual information at multiple scales and the maintenance of spatial continuity. Boundary accuracy (HD95) is also evident to be enhanced by BFA addition, and this illustrates the power of multi-granularity edge cues in the recovery of contour information. The AW-loss uses hard samples, so the extreme foreground-background ratio typical of mammography is greatly reduced, and the performance is greatly improved. The PRM refinement module is intended to further enhance HD95 and Dice metrics, based on a strong single-stage framework, followed by the mitigation of the residual segmentation errors and the enhancement of boundary continuity based on the attention-guided refinement, indicating that the second-stage conditioning on the coarse prediction effectively removes residual error as well as refines the boundary structure.

The findings suggest that every component makes complementary improvements. MHD mainly increases sensitivity and overlap by increasing the spatial context; BFA greatly decreases boundary deviation as measured by lower HD95; and AW-loss enhances stability in extreme class imbalance. The combination of these modules in the proposed model provides a robust single-stage base, and the PRM refinement phase provides extra boundary accuracy, which proves explicit coarse-to-fine correction.

6. Discussion

In terms of mechanisms, the enhanced performance of the proposed approach can be explained by the interaction of its main components. The Multi-scale Hybrid Dilated Convolution (MHD) module boosts contextual information by capturing multiscale features while maintaining spatial connections, which is particularly useful in detecting lesions at different scales. The Boundary Feature Auxiliary (BFA) module enhances the boundary cues explicitly by incorporating multiscale boundary features, and thus reduces the phenomenon of edge blurring and facilitates accurate boundary detection. In addition, the Prediction-Guided Refinement Module (PRM) further refines the coarse segmentation prediction by generating attention maps that focus on lesion areas and suppress irrelevant background responses, resulting in improved boundary smoothness and fewer boundary errors.

Computational complexity is another aspect to consider. Nevertheless, the added complexity is justified as the model exhibits good inference speed on contemporary GPUs. In experiments, the model takes about 0.08 seconds to process a mammogram on an NVIDIA RTX 2080 Ti, confirming its potential for deployment in clinical workflows. Compared to prior studies, the proposed BReMS-Net shows competitive performance in region-based and improved performance in boundary-based evaluation. Traditional encoder-decoder frameworks such as U-Net and ResUNet++ typically exhibit smooth contours and lack sufficient contextual information, especially in low-contrast mammographic regions. Likewise, hybrid models such as DCSAU-Net and MBA-Net improve multi-scale feature representation, but do not specifically address residual boundary uncertainties after initial mass predictions. Transformer-based methods, such as Swin-UNet, focus on global feature learning but might not generate accurate boundaries. In contrast, the proposed BReMS-Net combines multi-scale contextual learning, boundary-aware feature modulation, and a prediction-guided refinement approach, which allows for more precise boundary extraction and reduces segmentation errors, particularly along lesion contours. These strengths demonstrate that the proposed method is a reliable and competitive solution compared with recent breast mass segmentation methods.

7. Limitations and Future Work

Although the proposed framework achieves strong performance on CBIS-DDSM, several factors restrict its clinical readiness. This study mainly performs its evaluation on the CBIS-DDSM dataset, which may limit generalizability across institutions, scanner manufacturers, and population distributions. While this study includes preliminary tests with an external dataset, no quantitative cross-dataset evaluation is provided. Thus, the generalizability of the proposed model needs to be further tested in future multi-center studies. Moreover, the refinement process introduces more computational operations compared to a single-pass segmentation network, although it provides significantly higher precision in the accuracy of boundaries and better segmentation consistency, inspiring further efficiency work to enable screening scale deployment. It is also based on pixel-wise masks, which highlights the importance of learning annotations efficiently, and clinical use could also be enhanced by interpretability and uncertainty estimations to enable effective human supervision.

Moreover, no multicenter or prospective clinical validation is provided in this study. The differences in imaging devices, acquisition protocols, and demographics of patients across hospitals can affect the strength of segmentation. Future research will be done with clinical centers to test the proposed framework using multi-institutional data and clinical procedures.

8. Conclusion

This paper presents BReMS-Net, a multi-stage breast mass segmentation framework that addresses the persistent challenges of blurred lesion boundaries, boundary over-smoothing, and severe foreground-background class imbalance in mammographic imaging. Evaluated on CBIS-DDSM, BReMS-Net performs well in region-based metrics with a Dice coefficient of 93.12%, and HD95 of 0.9826, with improved boundary segmentation. These findings demonstrate the effectiveness of the proposed architecture for delineating breast masses, but quantitative evaluation on external data is needed to fully evaluate its generalizability. The proposed framework shows potential for supporting CAD systems in breast cancer screening.

9. Declarations

9.1. Author Contributions

Tayyba Sarfraz: Conceptualization, Methodology, Writing Original Draft; **Tan Ling:** Supervisor, Conceptual Guidance, Critical Review, Formal Analysis; **Ahmad Ijaz:** Data Visualization, Formal Analysis, Investigation and Review.

9.2. Institutional Review Board Statement

Not applicable.

9.3. Informed Consent Statement

Not applicable.

9.4. Data Availability Statement

The datasets used in this study are publicly available at:

- https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset.
- https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset.

9.5. Acknowledgment

Not applicable.

9.6. Conflicts of Interest

The authors declare no conflict of interest.

10. References

[1] Q. Guo *et al.*, "Context-aware feature complementary screening network for mass segmentation in whole mammograms," *Multimedia Systems*, vol. 32, no. 1, p. 10, 2026. https://doi.org/10.1007/s00530-025-02072-1.

[2] D. E. M. Jaincy, P. Venkatasubbu, "An empirical study for breast cancer detection using MRI images," *Biomedical Signal Processing*, vol. 118, p. 109640, 2026. https://doi.org/10.1016/j.bspc.2026.109640.

[3] A. Poonia, M. Meena, A. S. Yadav, S. Maheshwari, and D. Songara, "Efficient Breast Cancer Detection and Classification Model by Analyzing Mammogram Images Using ViT-Aided MobileNet With LSTM Network Based on Adaptive Segmentation," *Computational Intelligence*, vol. 42, no. 1, p. e70176, 2026. https://doi.org/10.1111/coin.70176.

[4] N. Brancati and M. Frucci, "USE-MiT: Attention-based model for breast ultrasound images segmentation," *Computer Methods and Programs in Biomedicine Update*, p. 100226, 2026. https://doi.org/10.1016/j.cmpbup.2025.100226.

[5] L. Yadav, G. Chandra, and D. Yadav, "Breast cancer stage detection by differentiating benign and malignant tumor using L0H-CWSNN and FZB-IS," *Expert Systems with Applications*, vol. 307, p. 131089, 2026. https://doi.org/10.1016/j.eswa.2026.131089.

[6] D. Mathew, K. V. Grace, and M. M. S. J. Preetha, "Breast cancer detection and classification using optimisation enabled deep learning model," *International Journal of Bioinformatics Research and Applications*, vol. 22, no. 2, pp. 107-125, 2026. https://doi.org/10.1504/IJBRA.2026.152611.

[7] R. Meegada and H. K. Bhuyan, "Segmentation and Feature Extraction Based Breast Cancer Detection Analysis," in *2026 Sixth International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT)*, 2026, pp. 1-6: IEEE. https://doi.org/10.1109/ICAECT68478.2026.11426107.

[8] V. Sreelekshmi, K. Pavithran, and J. J. Nair, "An integrated model for early breast cancer prediction using microcalcifications and patient risk factors," *Discover Artificial Intelligence*, vol. 6, no. 1, p. 30, 2026. https://doi.org/10.1007/s44163-025-00775-y.

[9] A. Wahiba and R. El Mostafa, "Pixel Intensity in Mammography: A Factor of Error in Breast Cancer Detection," in *EPJ Web of Conferences*, 2026, vol. 350, p. 03006: EDP Sciences. https://doi.org/10.1051/epjconf/202635003006.

[10] K. Mo et al., "Deep Electrical Impedance Spectroscopic Tomography for the Characterization of Tissue Architecture and Composition," *IEEE Transactions on Instrumentation and Measurement*, vol. 75, 2026. https://doi.org/10.1109/TIM.2026.3667225.

[11] D. M. Jaincy and V. Pattabiraman, "An empirical study for breast cancer detection using MRI images," *Biomedical Signal Processing and Control*, vol. 118, p. 109640, 2026. https://doi.org/10.1016/j.bspc.2026.109640.

[12] M. Basith, P. Praveen, and P. C. S. Reddy, "Adaptive deep Q-GAN framework for enhanced breast cancer detection in medical imaging," *Biomedical Signal Processing and Control*, vol. 112, p. 108638, 2026. https://doi.org/10.1016/j.bspc.2025.108638.

[13] B. Khati, S. Mukherjee, A. Sinitca, D. Kaplun, and R. Sarkar, "Reciprocal cooperative gating fusion of SqueezeNet and ShuffleNetV2 for breast cancer detection in histopathology images," *Scientific Reports*, vol. 16, art. No. 5904, 2026. https://doi.org/10.1038/s41598-026-36375-8.

[14] T. Lehnen, D. Polenske, B. D. Wichtmann, and N. C. Lehnen, "AI software as a third reader in breast cancer screening—a prospective diagnostic observational study," *European Radiology*, pp. 1-11, 2026. https://doi.org/10.1007/s00330-026-12359-0.

[15] B. N. Chua, D. K. H. Thng, T. B. Toh, and D. Ho, "Artificial intelligence for breast cancer management," *Communications Medicine*, vol. 6, 2026. https://doi.org/10.1038/s43856-025-01342-3.

[16] E. Elías-Cabot, S. Romero-Martín, J. L. Raya-Povedano, A. Rodríguez-Ruiz, and M. Álvarez-Benito, "AI-based triage and decision support in mammography and digital tomosynthesis for breast cancer screening: a paired, noninferiority trial," *Nature Medicine*, vol. 32, pp. 1296–1305, 2026. https://doi.org/10.1038/s41591-026-04277-x.

[17] C. Ma, H. Zhang, and L. Guo, "DFMFI: Ultrasound Breast Cancer Detection Method Based on Dynamic Fusion Multi-Scale Feature Interaction Model," *Computerized Medical Imaging and Graphics*, vol. 128, p. 102710, 2026. https://doi.org/10.1016/j.compmedimag.2026.102710.

[18] M.-J. Lee et al., "Single-tube total analysis system for ratiometric detection of exosomal miRNAs in breast cancer diagnosis," *Chemical Engineering Journal*, vol. 529, p. 173140, 2026. https://doi.org/10.1016/j.cej.2026.173140.

[19] H. Abudukelimu *et al.*, "DVF-YOLO-Seg: A two-stage breast mass segmentation model with enhanced feature extraction and small lesion detection," Digital Health, vol. 11, 2025. https://doi.org/10.1177/20552076251374192.

[20] Y. Wang, M. Ali, T. Mahmood, A. Rehman, and T. Saba, "Robust Bi-CBMSegNet framework for advancing breast mass segmentation in mammography with a dual module encoder-decoder approach," *Scientific Reports*, vol. 15, no. 1, p. 24434, 2025. https://doi.org/10.1038/s41598-025-09775-5.

[21] F. J. M. Shamrat *et al.*, "MammoSegNet: a convolutional network analysis for segmenting tumor tissue masses in digital mammograms of breast cancer patients," *Neural Computing and Applications*, vol. 37, no. 32, pp. 26437-26484, 2025. https://doi.org/10.1007/s00521-025-11631-6.

[22] T. Fatma, P. K. Sahu, S. Choudhury, and A. Wunnava, "Magnification-independent breast cancer diagnosis using a GWO-enhanced vision transformer with multi-stage stain normalization," *Scientific Reports*, vol. 16, 2026. https://doi.org/10.1038/s41598-026-42490-3.

[23] Q. Chen, Y. Zhao, X. Luo, W. He, and G. Shi, "Hybrid Deep Learning and Classification Framework for Automatic Traffic Inspection Classification Based on Image Detection," *Transactions on Emerging Telecommunications Technologies*, vol. 37, no. 2, p. e70325, 2026. https://doi.org/10.1002/ett.70325.

[24] A. Jain, R. K. Rupani, K. P. Arunachalam, and D. Veeraswamy, "Multi-scale feature fusion for breast cancer detection using circular dilated convolutional transformer optimized by enhanced wombat algorithm," *Computers and Electrical Engineering*, vol. 134, p. 111087, 2026. https://doi.org/10.1016/j.compeleceng.2026.111087.

[25] S. Mohammadi and M. A. Livani, "A two-stage self-supervised learning framework for breast cancer detection with multi-scale vision transformers," *Information Sciences*, vol. 735, p. 123061, 2026. https://doi.org/10.1016/j.ins.2025.123061.

[26] G. Shruthi and P. Ravikumar, "A Hybrid CNN-Transformer Model for Tumor-Infiltrating Lymphocyte Score Prediction in Breast Cancer Histopathological Image," *Engineering, Technology & Applied Science Research*, vol. 16, no. 2, pp. 32893-32898, 2026. https://doi.org/10.48084/etasr.15757.

[27] P. J. Ho *et al.*, "Breast Cancer Screening Knowledge and Sentiments in Singaporean Women: Mixed Methods Study Using Topic Modeling, Sentiment Analysis, and Structured Questionnaire Data," *Journal of Medical Internet Research,* vol. 28, p. e78439, 2026. https://doi.org/10.2196/78439.

[28] R. Varsha and S. Veni, "Hybrid Deep Learning: Parallel CNN and Swin Transformer Fusion Network for Breast Cancer Diagnosis," in *2026 International Conference on Electric Power and Renewable Energy (EPREC)*, 2026. https://doi.org/10.1109/EPREC66546.2026.11412005.

[29] H. Acikgoz, A. Aytekin, and S. Gezici, "BreasTransNeXt: An Enhanced Multi-Module Vision Transformer For Early Breast Cancer Diagnosis," *Journal of Imaging Informatics in Medicine*, pp. 1-20, 2026. https://doi.org/10.1007/s10278-026-01863-w.

[30] R. M. Al-Tam, A. M. Al-Hejri, F. A. Hashim, S. M. Narangale, M. A. Al-Antari, and S. A. Alzakari, "An Interpretable Ensemble Transformer Framework for Breast Cancer Detection in Ultrasound Images," *Diagnostics*, vol. 16, no. 4, p. 622, 2026. https://doi.org/10.3390/diagnostics16040622.

[31] L. G. Falconi, M. Perez, W. G. Aguilar, and A. Conci, "Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 5, no. 2, pp. 154-165, 2020. https://doi.org/10.25046/aj050220.

[32] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236-248, 2012. https://doi.org/10.1016/j.acra.2011.09.014.