AMEDDE-Net: A Deeply Supervised Multi-Scale Attention Network for Breast Cancer Lesion Segmentation in Digital Mammography

Dyah Titisari
Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, Indonesia
7022221019@student.its.ac.id

Andrew Prasetyo
Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, Indonesia
7022231006@student.its.ac.id

Eko Mulyanto Yuniarno
Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, Indonesia
ekomulyanto@ee.its.ac.id

I Ketut Eddy Purnama
Department of Computer Engineering, Sepuluh Nopember Institute of Technology, Indonesia
ketut@ee.its.ac.id

Mauridhi Hery Purnomo
Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, Indonesia
hery@ee.its.ac.id (corresponding author)

Received: 2 March 2026 | Revised: 8 April 2026 | Accepted: 21 April 2026

 DOI: https://doi.org/10.48084/etasr.18480

ABSTRACT

This study proposes the Attention, Multi-scale, Enhanced, Dilated, Dual-head, Ensemble Network (AMEDDE-Net), a deeply supervised segmentation framework that integrates multi-scale feature extraction, enhanced attention mechanisms, dilated convolutions, dual-head prediction, and adaptive ensemble fusion to improve lesion-focused learning and boundary delineation. The model was evaluated using a publicly available Contrast-Enhanced Spectral Mammography (CESM) dataset and five-fold cross-validation. Extensive hyperparameter tuning, ablation studies, and comparisons with state-of-the-art convolutional and transformer-based models were conducted. The experimental results indicated that AMEDDE-Net achieved superior performance, with 97.25% accuracy, a Dice coefficient of 65.10%, an F1-score of 65.10%, and a recall of 71.83%, outperforming conventional U-Net variants and the transformer-based TransU-Net. Ablation analyses confirmed the importance of multi-scale attention and deep supervision, whereas qualitative results demonstrated improved boundary accuracy and reduced False Negatives. These findings highlight the potential of AMEDDE-Net as a reliable tool for automated breast cancer screening and clinical decision support.

Keywords-cancer segmentation; digital mammography; deep supervision; multi-scale attention; deep learning; medical image analysis

I. INTRODUCTION

Breast cancer is the second leading cause of cancer-related mortality among women, imposing a substantial burden on the global healthcare system [1]. The early detection and precise segmentation of breast lesions are significant for improving clinical outcomes and reducing mortality rates [2]. Among the available imaging modalities, mammography is widely used for breast cancer screening due to its cost-effectiveness, broad accessibility, and well-established efficacy in facilitating early diagnosis [3]. However, manual lesion segmentation by radiologists remains time-consuming, prone to inter-observer variability, and expertise-dependent, highlighting the need for automated segmentation techniques [4]. Encoder–decoder

architectures, particularly the U-Net and its variants, have outperformed traditional approaches in tumor boundary delineation with computational efficiency [5-10]. Nevertheless, challenges persist, including the detection of subtle microcalcifications, handling diverse lesion morphologies and scales, and limited annotated mammography datasets [7].

Advances in Deep Learning (DL) have driven significant progress in breast cancer segmentation. For instance, U-Net 3+ with dense dilated convolution and MS-SSIM loss achieved a 98.47% Dice score on mammography segmentation [11], whereas a refined mask R-CNN combining mammography and ultrasound modalities achieved a classification accuracy of 99.20 % [12]. Transformer-based models have further enhanced global-contextual modeling in segmentation tasks. Trans U-Net and its variants improved segmentation accuracy for small and complex targets by combining transformer encoders with convolutional feature maps [13], whereas attention mechanisms integrating channel and spatial modules with CNN backbones enhanced lesion localization and interpretability [14]. Self-supervised transformer frameworks have also demonstrated performance gains of 5-7% over supervised CNN baselines across multiple breast imaging modalities [15, 16]. Multi-scale learning strategies and adversarial training frameworks have also been employed to address variability in lesion size, shape, and appearance [17, 18], supporting downstream clinical tasks such as automated classification and CAD systems [19-22]. Ensemble and hybrid approaches have achieved competitive results, with 99.76% accuracy reported using an integrated DBN-GCN-SSAE model [23].

Despite these advances, accurate lesion segmentation in digital mammography is challenging due to variations in lesion size, morphology, and contrast against complex background tissues [20]. Most existing approaches address these challenges in isolation: Attention U-Net lacks multi-scale extraction and lesion-specific prediction; U-Net++ lacks explicit attention and deep supervision; and TransU-Net lacks lesion-targeted learning and adaptive fusion.

To address these limitations, this study proposes the Attention, Multi-scale, Enhanced, Dilated, Dual-head, Ensemble Network (AMEDDE-Net), an enhanced U-Net-based segmentation framework validated on the Categorized Digital Database (CDD) for low-energy and subtracted Contrast-Enhanced Spectral Mammography (CESM) images [24], a publicly available benchmark hosted by The Cancer Imaging Archive (TCIA) comprising 2,006 high-resolution CESM images from 326 patients [25]. A rigorous comparative evaluation was conducted by incorporating hyperparameter optimization and benchmarking against established segmentation models, contributing to the development of clinically viable automated breast cancer segmentation systems [26]. Table I summarizes the limitations that the proposed model addresses compared to previous models.

TABLE I. ARCHITECTURAL COMPARISON OF AMEDDE-NET WITH BASELINE MODELS

Feature	**Attention U-Net**	**U-Net++**	**TransU-Net**	**AMEDDE-Net (proposed)**
Attention mechanism	Gate attention	None	Multi-head self-attention (ViT)	Gate + Lesion-enhancement attention
Multi-scale processing	No	Dense skip connections (nested)	No	Yes - 1×1, 3×3, 5×5, GMP branches
Dilated convolutions	No	No	No	Yes (d=2, effective RF=5×5)
Dual-head prediction	No	No	No	Yes - main + lesion-specific head
Deep supervision	No	Partial (multi-output)	No	Yes - 3 auxiliary supervision levels
Adaptive ensemble fusion	No	No	No	Yes - weighted α-fusion ($\alpha = 0.6$)
Global context	No	No	Yes (Transformer patch tokens)	Yes (Global Max Pooling branch)
Lesion-specific enhancement	No	No	No	Yes - $E_{small}(x)$ grouped conv module
Training paradigm	Supervised, single loss	Supervised, multi-scale loss	Supervised, patch-based	Supervised, multi-level deep supervision + ensemble

II. THE PROPOSED METHOD

Automated breast lesion segmentation in mammography is challenging because of feature suppression, insufficient multi-scale representation, and information loss during network propagation. To deal with these challenges, the AMEDDE-Net framework is proposed in this study, and its architecture is illustrated in Figure 1.

A. Preprocessing

Preprocessing is applied to standardize the input data, enhance image quality, and facilitate stable model training. The dataset consists of CESM images paired with corresponding binary tumor masks. All images and masks were resized to 256 × 256 pixels using bilinear interpolation to ensure a uniform spatial resolution. Pixel intensities and binary masks were normalized to [0, 1] by dividing by 255, and masks were expanded along the channel dimension for compatibility with the network input.

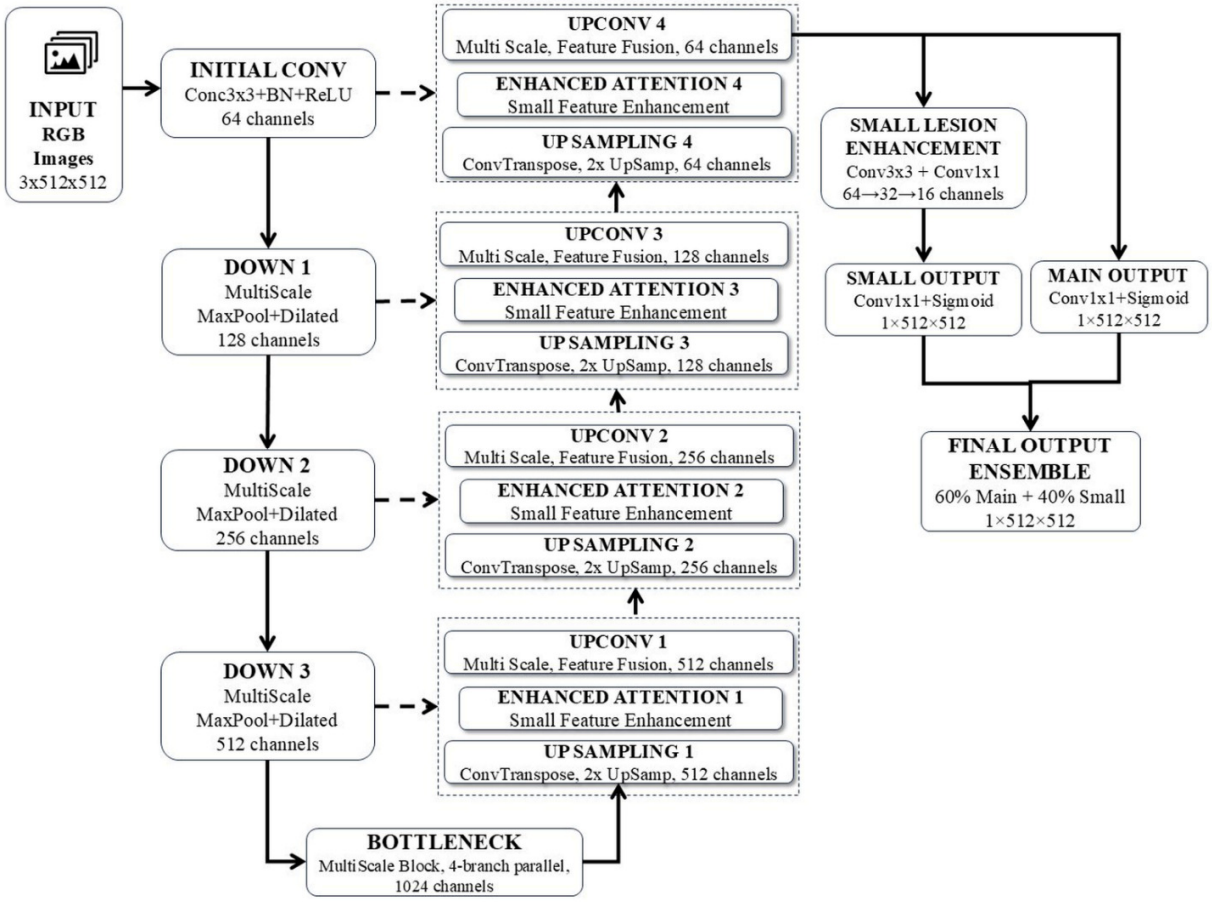

Fig. 1. AMEDDE-Net architecture.

After normalization, the dataset was split into training and testing sets at a 90:10 ratio with a fixed random seed for reproducibility, standardizing the data and reducing intensity variance for stable model convergence.

B. The AMEDDE-Net Architecture

The proposed architecture combines enhanced attention, multi-scale feature processing, dilated convolutions, dual-head prediction, deep supervision, and adaptive ensemble strategies to achieve accurate and robust breast lesion segmentation.

To improve mathematical clarity, all variables used in the equations throughout this section were defined. Let x denote the input feature map, g the gating signal, a the ensemble weight, k the kernel size, and d the dilation rate.

1) Enhanced Attention

The enhanced attention block emphasizes lesion-relevant features that are often suppressed in standard encoder-decoder architectures by combining gate signals from the decoder with skip connection features from the encoder, which are processed through 1×1 convolutions and batch normalization before additive fusion.

The transformed features are then fused additively, activated using ReLU, and projected into attention coefficients through a final convolution and sigmoid activation:

$$\psi = \sigma\left(\text{Conv}_{1\times1}\left(\text{ReLU}\left(W_g.\text{g} + W_x.x_l\right)\right)\right) \quad (1)$$

where g is the gate signal from the decoder, x_l is the skip-connection feature map from the encoder, and W_g and W_x are learnable 1×1 convolutional weight matrices.

As shown in Figure 2, the attention coefficients ψ, ranging between 0 and 1, generate a spatial importance map that highlights relevant regions in the feature space. To further enhance the detection of small and subtle lesions, a lesion-specific enhancement component $E_{small}(x)$ is incorporated. This module employs grouped convolutions to capture fine-grained spatial patterns:

$$E_{small}(x) = \sigma\left(\text{Conv}_{1\times1}\left(\text{ReLU}\left(\text{Conv}_{3\times3}^{(grouped=G)}(x)\right)\right)\right) \quad (2)$$

where G is the number of channel groups, set to $G = 4$ in this implementation, enabling fine-grained spatial pattern capture with reduced parameter overhead.

The final enhanced feature map is obtained through element-wise modulation of the input features with the lesion enhancement output:

$$\hat{x} = \left(x \odot E_{small}(x)\right) \quad (3)$$

where $\odot$ denotes element-wise multiplication, and $E_{small}(x) \in [0,1]$ is the lesion enhancement coefficient map.

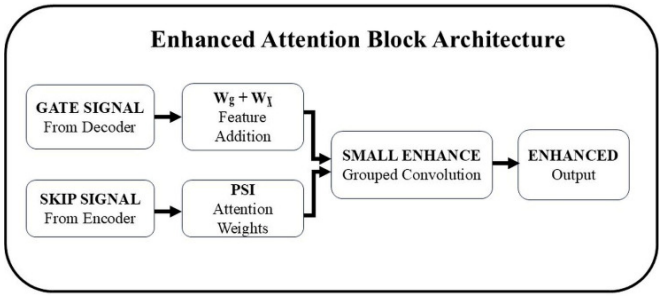

Fig. 2. Enhanced attention block architecture.

This dual-enhancement mechanism ensures that lesion features are both explicitly amplified and adaptively weighted, improving sensitivity to small lesions while maintaining computational efficiency.

2) Multi-Scale Feature Processing

To address scale variability in mammography lesions, the multi-scale block captures contextual information at multiple spatial scales using four parallel branches: a 1×1 convolution for channel projection, 3×3 and 5×5 convolutions for fine and medium-scale extraction, and global max pooling for broader contextual awareness, as depicted in Figure 3.

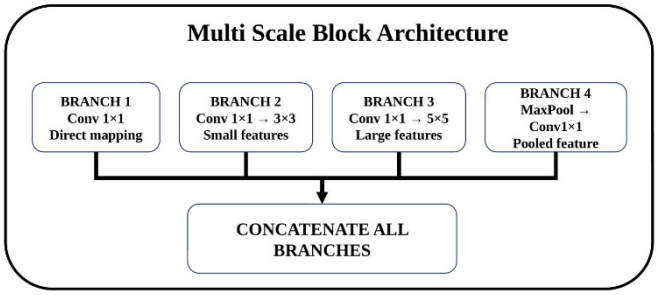

Fig. 3. Multi-scale block architecture.

The branch outputs are concatenated into a rich multi-scale representation, enabling simultaneous extraction of fine details and broader contextual cues without requiring image pyramids or multiple network passes:

$$F_{multi} = \text{Concat}(B_1, B_2, B_3, B_4) \quad (4)$$

3) Dilated Convolutions

Dilated convolutions expand the receptive field without increasing the parameter count or reducing the spatial resolution, thereby preserving lesion boundaries and capturing broader contextual information.

When $d = 1$, the operation reduces to standard convolution. Increasing d enlarges the receptive field while maintaining computational efficiency. The effective receptive field is given by:

$$RF_{effective} = k + (k-1)(d-1) \quad (5)$$

where k represents the kernel size, and d is the dilation rate. This linear scaling of the receptive field with the dilation rate efficiently captures multi-scale contextual information. In AMEDDE-Net, a dilation rate of $d = 2$ with 3×3 kernels results in an effective receptive field of 5×5, verified as $RF = 3 + (3-1)(2-1) = 5$ pixels, while maintaining parameter efficiency. The dilated convolution output is followed by batch normalization and ReLU activation:

$$F_{dilated} = \text{ReLU}\left(\text{BN}(F *_d K_d)\right) \quad (6)$$

This design enables the network to integrate a broader spatial context while preserving fine-grained lesion details and improving discrimination between malignant lesions and surrounding benign tissues.

4) Dual-Head Architecture

The dual-head architecture is designed to tackle the trade-off between general tissue segmentation and precise lesion detection, which often requires different feature representations. Instead of relying on a single output branch, the network employs two specialized prediction heads to optimize both tasks simultaneously.

The main head generates general segmentation predictions directly from the final decoder features using a 1×1 convolution, as defined in (7). In contrast, the lesion head incorporates additional spatial refinement through two sequential 3×3 convolutional layers with batch normalization and ReLU activation, as presented in (8), thereby enabling enhanced detection of small or subtle pathological structures.

$$H_{main} = \text{Conv}_{1\times1}(D_1) \quad (7)$$

$$H_{small} = \text{Conv}_{1\times1}\left(\text{ReLU}\left(\text{BN}\left(\text{Conv}_{3\times3}\left(\text{ReLU}\left(\text{BN}\left(\text{Conv}_{3\times3}(D_1)\right)\right)\right)\right)\right)\right) \quad (8)$$

Both heads produce probability maps via sigmoid activation, and their outputs are fused through a weighted ensemble formulation, where $\alpha = 0.6$ balances global segmentation accuracy and lesion sensitivity.

$$O_{final} = \alpha \cdot O_{main} + (1-\alpha) \cdot O_{small} \quad (9)$$

The main head emphasizes efficiency and general feature representation, whereas the lesion head focuses on fine-grained spatial detail. This complementary design improves the detection of small lesions while maintaining overall segmentation performance without requiring separate models or training stages.

5) Deep Supervision

Deep supervision is employed to improve gradient propagation and encourage intermediate decoder layers to learn discriminative representations. In the AMEDDE-Net, auxiliary

prediction branches are attached to multiple decoder stages, converting intermediate feature maps into segmentation outputs using 1×1 convolutions. These outputs are upsampled to the original image size and combined into the overall deep supervision loss:

$$\mathcal{L}_{ds} = w_1 \cdot \mathcal{L}\left(O^{up}_{deep1}, Y\right) + w_2 \cdot \mathcal{L}\left(O^{up}_{deep2}, Y\right) + w_3 \cdot \mathcal{L}\left(O^{up}_{deep3}, Y\right) \quad (10)$$

where the weights are set as $w_1 = 0.5$, $w_2 = 0.3$, and $w_3 = 0.2$, assigning higher importance to shallower decoder outputs that preserve finer spatial detail.

This multilevel supervision stabilizes training, accelerates convergence, and improves segmentation accuracy by enforcing meaningful feature learning across the decoder hierarchy.

III. RESULTS AND DISCUSSION

A. Dataset

This study employed the CDD-CESM dataset [24, 25], comprises of 2,006 high-resolution mammographic images (average resolution: 2,355×1,315 pixels) from 326 female patients aged 18-90 years, acquired at the National Cancer Institute, Cairo University, Egypt. It includes 310 masses, 48 architectural distortions, 222 asymmetries, 238 calcifications, 334 mass enhancements, 184 non-mass enhancements, and 751 normal images. Each case includes paired low-energy and subtracted CESM images in the CC and MLO views, with ACR BI-RADS 2013 annotations and expert-annotated binary segmentation masks as the ground truth.

The heterogeneous pathological patterns and multi-view paired imaging make CDD-CESM a suitable benchmark for evaluating DL models for breast cancer lesion segmentation. Its adoption is confirmed by recent studies. For example, authors in [27], who applied Swin Transformer and ConvNeXtV2 for ROI-stratified classification. Additionally, authors in [28] evaluated panoptic segmentation using a diffusion model, while in [29] explainable CNN architectures were assessed for binary lesion classification.

B. Experimental Setup

To address the limited availability of annotated data, an Albumentations-based augmentation pipeline incorporating geometric transformations, elastic deformations, intensity variations, and contrast enhancement was applied to improve model generalization. The proposed AMEDDE-Net was evaluated against established baselines, including U-Net++, R2U-Net, Attention U-Net, V-Net, and TransU-Net, all of which were trained under identical experimental conditions to ensure a fair comparison.

The evaluation was conducted at the pixel level on the test set (10% of the dataset, comprising 100 mammographic images resized to 512×512 pixels), resulting in a total of 26,214,400 pixels evaluated. Standardized metrics were utilized, namely the Dice coefficient, precision, recall, F1-score, and Intersection over Union (IoU), with IoU prioritized owing to its clinical relevance in measuring spatial agreement between predicted segmentations and expert-annotated lesion regions. This evaluation protocol allows assessing the proposed architecture and clarifies the strengths and limitations of commonly used segmentation models for CESM breast cancer imaging.

C. Parameter Tuning and Training Configuration

The optimal training configuration for the AMEDDE-Net was identified through hyperparameter tuning using random search across learning rates, batch sizes, and optimizers. As presented in Table II, Configuration 2 achieved optimum performance with an IoU of 38.01% and a Dice coefficient of 54.04%, using the AdamW optimizer with a Learning Rate (LR) of 0.0005 and batch size of 32, yielding high specificity (93.02%) and accuracy (89.91%).

D. AMEDDE-Net Performance Evaluation

As shown in Table III, the five-fold cross-validation results demonstrated the efficiency of the AMEDDE-Net, with a mean Dice coefficient ranging from 28.80% to 33.66% and IoU ranging from 19.46% to 23.65% across all folds. The sensitivity remained consistently high (89.91%-96.79%), confirming a strong capability to identify malignant regions, whereas the specificity varied between 57.28% and 71.92%, reflecting a deliberate emphasis on sensitivity to minimize False Negatives in clinical screening scenarios. The relatively stable mean performance across folds confirms that the model generalizes consistently, validating the effectiveness employed to prevent data leakage.

TABLE II. TOP-5 HYPERPARAMETER RESULT

No	LR	Batch	Opt	IoU	Dice	F1	Precision	Recall	Specificity	Accuracy
1	0.0001	32	AdamW	37.84	54.08	54.08	47.36	63.22	92.39	89.58
2	**0.0005**	**32**	**AdamW**	**38.01**	**54.04**	**54.04**	**48.91**	**60.99**	**93.02**	**89.91**
3	0.0001	32	AdamW	36.83	52.82	52.82	47.08	60.26	92.68	89.56
4	0.00001	8	Adam	33.27	48.51	48.51	38.89	68.54	88.27	86.32
5	0.00001	16	Adam	30.52	46.39	46.39	34.13	73.13	84.83	83.72

TABLE III. PERFORMANCE OF THE PROPOSED SEGMENTATION MODEL EVALUATED USING FIVE-FOLD CROSS-VALIDATION

Fold	Dice mean	Dice std dev	IoU mean	IoU std dev	Sensitivity mean	Sensitivity std dev	Specificity mean	Specificity std dev
1	30.303	26.180	21.077	21.061	93.757	13.275	69.476	12.243
2	33.662	26.601	23.658	21.694	89.906	17.175	71.924	12.986
3	28.808	24.084	19.457	18.865	92.335	15.439	65.659	11.856
4	29.120	25.757	20.117	20.659	96.786	7.381	57.280	15.158
5	31.855	26.257	22.211	21.143	95.003	11.562	63.239	12.116

E. Ablation Studies

Table IV summarizes the ablation results assessing the contribution of each component. The full enhanced model achieved the highest mean Dice of 50.4 (+26.47%, $p = 0.0012$, Cohen's $d = 3.341$), confirming the effectiveness of jointly integrating all components. Core architectural enhancements alone yielded significant gains (mean Dice = 46.3, +16.36%, $p = 0.0026$, Cohen's $d = 1.674$), whereas isolated modules presented limited or nonsignificant improvements - deep supervision only ($p = 0.8320$) and enhanced postprocessing ($p = 0.6292$) did not reach significance. The full enhanced model also exhibited the lowest coefficient of variation (3.820%), and multi-scale blocks in isolation yielded significant gains ($p = 0.0024$, Cohen's $d = 1.024$), collectively confirming that the performance gains of AMEDDE-Net arise from cohesive architectural integration rather than any single module.

TABLE IV. RESULTS OF ABLATION EXPERIMENTS

Model component	Mean_Dice	95%_CI	CV_%	vs_Baseline	*p*-value	Significant	Effect_size
Full enhanced model	50.4	[48.7, 51.7]	3.820	+26.47%	0.0012	Significant	3.341 (Large)
Full model - multi-scale	50.3	[48.4, 51.9]	4.400	+26.39%	0.0006	Significant	3.236 (Large)
Enhanced attention only	41.7	[39.3, 43.6]	6.400	+4.62%	0.0952	Not Significant	0.539 (Medium)
Architecture components	46.3	[43.4, 49.1]	8.100	+16.36%	0.0026	Significant	1.674 (Large)
Deep supervision only	39.5	[34.6, 43.2]	14.520	-0.94%	0.8320	Not Significant	-0.076 (Negligible)
Enhanced postprocessing	40.7	[35.6, 44.4]	14.530	+2.16%	0.6292	Not Significant	0.170 (Negligible)
Small lesion head only	40.8	[37.6, 43.8]	9.590	+2.48%	0.7215	Not Significant	0.249 (Small)
Enhanced preprocessing	45.3	[41.8, 47.4]	8.360	+13.71%	0.0625	Not Significant	1.398 (Large)
Multi-scale blocks only	44.1	[40.8, 47.4]	9.680	+10.66%	0.0024	Significant	1.024 (Large)

F. Performance Comparison

Table V presents a comparative evaluation of seven DL architectures for breast cancer lesion segmentation. AMEDDE-Net achieved an optimum overall performance, attaining 97.25% accuracy, a Dice of 65.10%, a recall of 71.83%, and an F1-score of 65.10%, indicating strong lesion localization capability. Conventional CNN-based models, such as U-Net and V-Net, demonstrated stable IoU scores of 53.51% and 52.75%, respectively, whereas TransU-Net achieved a Dice of 44.72%, a recall of 50.99%, and an F1-score of 40.41%. AMEDDE-Net consistently outperformed TransU-Net, with improvements of 20.38% in Dice, 20.84% in recall, 24.69% in F1-score, and 8.19% in accuracy, alongside a lower training loss (34.18 versus 31.77), indicating more stable convergence.

The mean IoU of 54.85% should be interpreted in the context of the CDD-CESM dataset's inherent difficulty, which includes subtle lesion types such as asymmetries and non-mass enhancements, alongside 751 normal images, evaluated under strict patient-level five-fold cross-validation to prevent data leakage. While moderate in absolute terms, AMEDDE-Net exhibited the highest IoU among all seven evaluated models and attained a recall of 71.83%, the clinically most critical metric for minimizing False Negatives in early breast cancer screening.

In contrast, U-Net++ and R2U-Net showed limited sensitivity with recall rates of 1.87% and 36.35%, and IoU of 45.58% and 14.77%, respectively, thereby reducing their clinical applicability. The superior performance of AMEDDE-Net reflects its integrated design, which combines multi-scale feature extraction, attention mechanisms, and deep supervision into a unified lesion-focused framework.

Figure 4 illustrates a qualitative comparison of four mammography cases. The AMEDDE-Net achieved the highest average IoU of 83.58% with consistent results (76.3%, 78.4%, 84.5%, 92.7%), whereas the other models presented notable limitations: U-Net under-segmented complex masses (53.73%), U-Net++ performed poorly on simple structures (48.53%), R2U-Net lacked robustness across lesion types, and V-Net failed in multi-region segmentation (50.55%).

The wide IoU range across models (48.53%-92.7%) highlights the strong influence of the architectural design on segmentation reliability. The consistently superior performance of AMEDDE-Net across lesion complexities confirms its suitability as a reliable clinical decision-support tool for accurate lesion delineation in mammography screening.

To further validate the reported metrics, a pixel-level confusion matrix was calculated for the 100-image test set, totaling 26,214,400 pixels, as depicted in Figure 5. The model achieved 785,629 True Positives and 24,708,129 True Negatives.

TABLE V. QUANTITATIVE COMPARISON OF SEGMENTATION PERFORMANCE ACROSS-BASELINE, CNN-BASED, TRANSFORMER-BASED, AND PROPOSED MODELS USING STANDARD EVALUATION METRICS

Method	Loss	Accuracy	Dice Coefficient	Mean IoU	Precision	Recall	F1-score
U-Net++	12.49	94.31	37.18	45.58	6.00	1.87	2.18
U-Net	12.40	94.24	40.34	53.51	50.30	21.50	23.01
R2U-Net	14.77	93.12	36.35	14.77	93.12	36.35	25.05
Attention U-Net	12.61	94.13	40.30	50.17	36.56	11.02	13.73
V-Net	10.67	94.94	44.00	52.75	41.46	25.37	24.17
TransU-Net	31.77	89.06	44.72	28.61	43.91	50.99	40.41
Proposed AMEDDE-Net	34.18	97.25	65.10	54.85	65.57	71.83	65.10

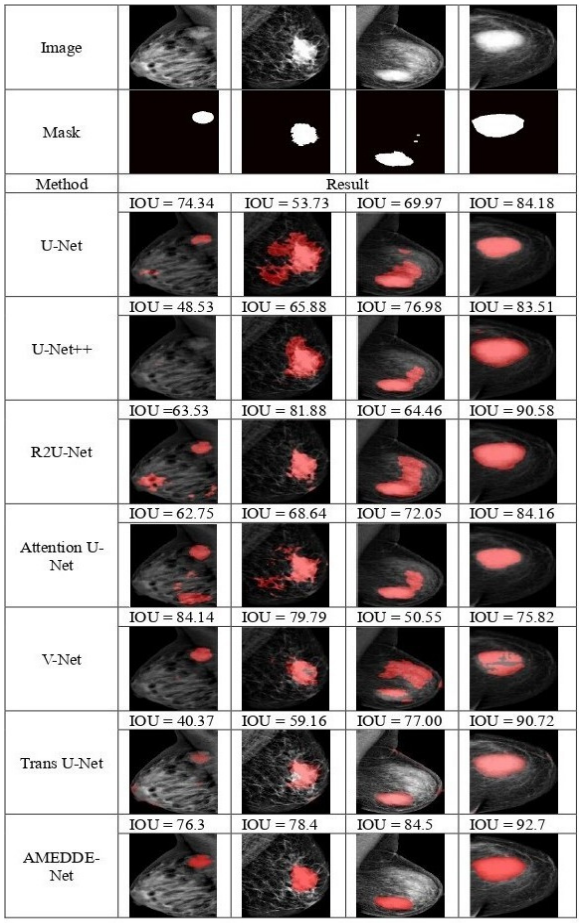

Fig. 4. Visual comparative analysis of segmentation performance across AMEDDE-Net and five other DL methods for four mammography cases with varying complexities.

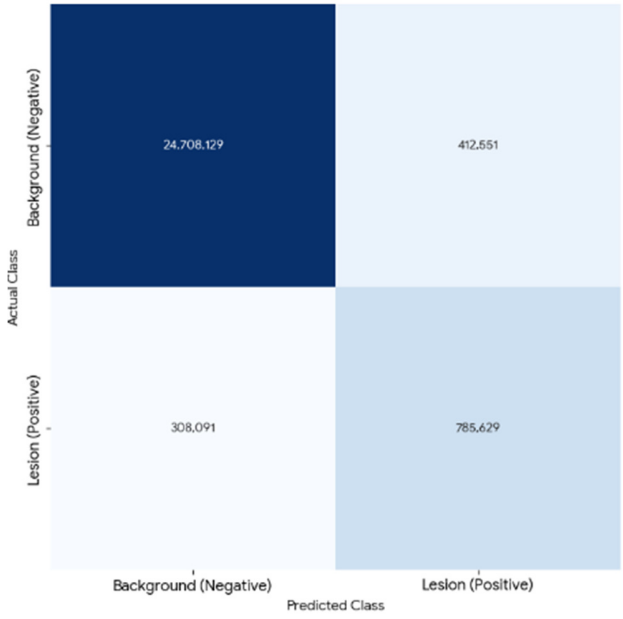

Fig. 5. Pixel-level confusion matrix for the AMEDDE-Net model.

IV. CONCLUSION

This study proposed the Attention, Multi-scale, Enhanced, Dilated, Dual-head, Ensemble Network (AMEDDE-Net), a deeply supervised multi-scale attention network for automated breast cancer lesion segmentation in digital mammography. The framework jointly integrated gate-guided attention, multi-scale feature extraction, dilated convolutions, dual-head prediction, deep supervision, and adaptive ensemble fusion. Among seven evaluated architectures, AMEDDE-Net achieved the best performance on the CDD-CESM dataset, attaining 97.25% accuracy, a Dice of 65.10%, and a recall of 71.83%, consistently outperforming both CNN-based and transformer-based models, including TransU-Net.

Ablation studies confirmed that performance gains arise from cohesive architectural integration ($p = 0.0012$, Cohen's $d = 3.341$) rather than any single module. The high recall of 71.83% was clinically significant, as minimizing False Negatives is critical in early breast cancer screening. The mean Intersection over Union (IoU) of 54.85%, while moderate in absolute terms, represented the highest among all evaluated models and reflects the inherent difficulty of the CDD-CESM dataset evaluated under strict patient-level cross-validation.

However, the model was validated on a single-institution dataset and operated on 2-D mammographic images without real-time deployment evaluation. Future work will pursue multi-institutional validation, extension to 3-D tomosynthesis, and lightweight architecture optimization for clinical deployment. Integration with explainability mechanisms and radiologist feedback loops will also be explored to support clinical adoption.

DECLARATION OF COMPETING INTERESTS

Not applicable.

ACKNOWLEDGMENT

This work was supported by a scholarship from the Ministry of Health of the Republic of Indonesia.

DATA AVAILABILITY

The dataset used in this study is the CDD-CESM [24].

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018, https://doi.org/10.3322/caac.21492.

[2] C. D. Lehman *et al.*, "National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium," *Radiology*, vol. 283, no. 1, pp. 49–58, Apr. 2017, https://doi.org/10.1148/radiol.2016161174.

[3] M. Woo *et al.*, "Subgroup evaluation to understand performance gaps in deep learning-based classification of regions of interest on mammography," *PLOS Digital Health*, vol. 4, no. 4, Apr. 2025, Art. no. e0000811, https://doi.org/10.1371/journal.pdig.0000811.

[4] Y. J. Suh, J. Jung, and B.-J. Cho, "Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning," *Journal of Personalized Medicine*, vol. 10, no. 4, Nov. 2020, Art. no. 211, https://doi.org/10.3390/jpm10040211.

[5] H. Zunair and A. Ben Hamza, "Sharp U-Net: Depthwise convolutional network for biomedical image segmentation," *Computers in Biology and Medicine*, vol. 136, Sept. 2021, Art. no. 104699, https://doi.org/10.1016/j.compbiomed.2021.104699.

[6] S. Li, M. Dong, G. Du, and X. Mu, "Attention Dense-U-Net for Automatic Breast Mass Segmentation in Digital Mammogram," *IEEE Access*, vol. 7, pp. 59037–59047, May 2019, https://doi.org/10.1109/ACCESS.2019.2914873.

[7] A. J. Bewersdorf, E. L. Bewersdorf, G. M. Fundaro, and G. Fundaro, "Mammographically Occult Breast Cancer in a Patient With Dense Breast Tissue," *Cureus*, vol. 17, no. 1, 2025, Art. no. e77789, https://doi.org/10.7759/cureus.77789.

[8] M. R. Islam *et al.*, "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images," *Machine Learning with Applications*, vol. 16, June 2024, Art. no. 100555, https://doi.org/10.1016/j.mlwa.2024.100555.

[9] S. Anari, S. Sadeghi, G. Sheikhi, R. Ranjbarzadeh, and M. Bendechache, "Explainable attention based breast tumor segmentation using a combination of UNet, ResNet, DenseNet, and EfficientNet models," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, Art. no. 1027, https://doi.org/10.1038/s41598-024-84504-y.

[10] H. Li, D. Chen, W. H. Nailon, M. E. Davies, and D. I. Laurenson, "Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography," *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 3–13, Jan. 2022, https://doi.org/10.1109/TMI.2021.3102622.

[11] S. M. Shaaban, M. Nawaz, Y. Said, and M. Barr, "An Efficient Breast Cancer Segmentation System based on Deep Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12415–12422, Dec. 2023, https://doi.org/10.48084/etasr.6518.

[12] P. Kumar *et al.*, "An AI-Based Method for Automated Breast Cancer Detection and Localization in Mammogram and Ultrasound Images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27266–27272, Oct. 2025, https://doi.org/10.48084/etasr.11951.

[13] J. Chen *et al.*, "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, vol. 97, Oct. 2024, Art. no. 103280, https://doi.org/10.1016/j.media.2024.103280.

[14] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, Apr. 2019, https://doi.org/10.1016/j.media.2019.01.012.

[15] S. Ahmad, E. S. Neal Joshua, N. T. Rao, R. M. Ghoniem, B. M. Taye, and S. Bharany, "A multi stage deep learning model for accurate segmentation and classification of breast lesions in mammography," *Scientific Reports*, vol. 15, no. 1, Oct. 2025, Art. no. 37103, https://doi.org/10.1038/s41598-025-21146-8.

[16] L. Wang, "Self-supervised learning and transformer-based technologies in breast cancer imaging," *Frontiers in Radiology*, vol. 5, Nov. 2025, https://doi.org/10.3389/fradi.2025.1684436.

[17] J. Chen, L. Chen, S. Wang, and P. Chen, "A novel multi-scale adversarial networks for precise segmentation of x-ray breast mass," *IEEE access*, vol. 8, pp. 103772–103781, Jun. 2020, https://doi.org/10.1109/ACCESS.2020.2999198.

[18] V. K. Singh *et al.*, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," *Expert Systems with Applications*, vol. 139, Jan. 2020, Art. no. 112855, https://doi.org/10.1016/j.eswa.2019.112855.

[19] Y. Gao, J. Lin, Y. Zhou, and R. Lin, "The application of traditional machine learning and deep learning techniques in mammography: a review," *Frontiers in Oncology*, vol. 13, Aug. 2023, https://doi.org/10.3389/fonc.2023.1213045.

[20] P. Buelens, S. Willems, L. Vandewinckele, W. Crijns, F. Maes, and C. G. Weltens, "Clinical evaluation of a deep learning model for segmentation of target volumes in breast cancer radiotherapy," *Radiotherapy and Oncology*, vol. 171, pp. 84–90, June 2022, https://doi.org/10.1016/j.radonc.2022.04.015.

[21] D. Dave *et al.*, "Diagnostic test accuracy of AI-assisted mammography for breast imaging: a narrative review," *PeerJ Computer Science*, vol. 11, Feb. 2025, Art. no. e2476, https://doi.org/10.7717/peerj-cs.2476.

[22] A. Jalalian, S. B. T. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clinical Imaging*, vol. 37, no. 3, pp. 420–426, May 2013, https://doi.org/10.1016/j.clinimag.2012.09.024.

[23] M. Sreevani and R. Latha, "An Advanced Ensemble of Deep Learning Models for Breast Cancer Segmentation and Classification with Two-Tier Optimization Algorithms," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27024–27029, Oct. 2025, https://doi.org/10.48084/etasr.12682.

[24] *Categorized Digital Database for Low energy and Subtracted Contrast Enhanced Spectral Mammography images*, The Cancer Imaging Archive, https://www.cancerimagingarchive.net/collection/cdd-cesm/.

[25] R. Khaled *et al.*, "Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research," *Scientific Data*, vol. 9, no. 1, Mar. 2022, Art. no. 122, https://doi.org/10.1038/s41597-022-01238-0.

[26] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. S. Elmaghraby, "Connected-UNets: a deep learning architecture for breast mass segmentation," *npj Breast Cancer*, vol. 7, no. 1, Dec. 2021, Art. no. 151, https://doi.org/10.1038/s41523-021-00358-x.

[27] J. H. Park, J. H. Lim, S. Kim, and J. Heo, "A Multi-label Artificial Intelligence Approach for Improving Breast Cancer Detection With Mammographic Image Analysis," *In Vivo*, vol. 38, no. 6, pp. 2864–2872, Nov. 2024, https://doi.org/10.21873/invivo.13767.

[28] K. Zhao, J. Prokop, J. Montalt-Tordera, and S. Mohammadi, "Panoptic Segmentation of Mammograms with Text-to-Image Diffusion Model," in *4th MICCAI Workshop, DGM4MICCAI 2024*, Marrakesh, Morocco, Oct. 2024, https://doi.org/10.1007/978-3-031-72744-3_10.

[29] S. Acosta-Jiménez, M. M. Mendoza-Mendoza, C. E. Galván-Tejada, J. M. Celaya-Padilla, J. I. Galván-Tejada, and M. A. Soto-Murillo, "Explainable Deep Learning for Breast Lesion Classification in Digital and Contrast-Enhanced Mammography," *Diagnostics*, vol. 15, no. 24, Dec. 2025, Art. no. 3143, https://doi.org/10.3390/diagnostics15243143.