Segmentation of Skin Lesions in Dermatoscopic Images Using ResU-Net with Enhanced Attention and Grad-CAM

Torang Four Yones Manullang[1*], Ernawati[2], Desi Andreswari[3]

[1,2,3]Informatics Engineering, Faculty of Engineering, Universitas Bengkulu, Bengkulu, Indonesia
E-mail: [1*]torangfym04@gmail.com, [2]ernawati@unib.ac.id, [3]desi.andreswari@unib.ac.id

(Received: 28 Jul 2026, revised: 18 Aug 2026, accepted: 26 Aug 2026)

Abstract

Skin lesions exhibit substantial variations in shape, size, color, texture, and boundary clarity, which make automatic segmentation of dermoscopic images challenging. Although previous studies have achieved improved segmentation performance through various attention-based architectures, most of them primarily focus on enhancing segmentation metrics, with limited attention to model interpretability and pixel-level feature refinement. This study developed a ResU-Net model with Enhanced Attention (ResUNetEA) for skin lesion segmentation and employed Grad-CAM to support model interpretability. The ISIC 2018 Task 1 dataset containing 2,594 paired dermoscopic images and segmentation masks was used. Images and masks were resized to 256 × 256 pixels, normalized or binarized, and divided into training, validation, and test sets using an 80:10:10 ratio. Training data were augmented using horizontal and vertical flips, rotation, brightness and contrast adjustment, CLAHE, elastic transformation, random resized crop, and Gaussian blur. The proposed architecture integrates residual blocks, channel attention through an Efficient Attention Block, pixel attention, and multi-scale feature fusion. The model was trained for 40 epochs using Adam with a learning rate of 0.0001 and a combination of BCEWithLogitsLoss and Dice Loss. The test results achieved a loss of 0.1264, a Dice coefficient of 0.8914, and an IoU of 0.8080. Predicted masks generally followed the location and shape of the lesions, while Grad-CAM highlighted the lesion regions and their boundary-transition areas. These findings indicate that ResUNetEA provides good segmentation performance and useful visual information about the model focus.

Keywords: Skin Lesion Segmentation, ResU-Net, Enhanced Attention, Grad-CAM, ISIC 2018.

I. INTRODUCTION

Skin lesions are abnormal changes in skin tissue that can indicate various dermatological diseases, including melanoma [1]. Skin lesions are typically examined using dermatoscopic images because they provide a more detailed view of the skin's surface structure and pigmentation patterns and facilitate systematic analysis of lesion characteristics [1][2]. However, the process of manually identifying and delineating lesion borders remains highly dependent on the examiner's experience and can be subjective, particularly when lesions have irregular shapes, low contrast, or colors that resemble the surrounding skin tissue. Additionally, the presence of artifacts such as hair, shadows, and light reflections can complicate the determination of a lesion's location and borders [3]. These conditions make manual segmentation highly dependent on the examiner's experience and potentially lead to subjective assessments [1]. Therefore, an automated segmentation method capable of consistently separating the lesion area from normal skin is needed.

Automated approaches based on deep learning have been widely applied in dermoscopic image analysis because they are capable of learning visual patterns directly from the data [4]. In medical images, deep learning can extract hierarchical feature representations to handle variations in texture, intensity, and the complex shapes of objects [5]. One of its key components is the Convolutional Neural Network (CNN), which learns image features ranging from simple levels to semantic features [6]. CNNs also remain a key component in various developments of medical image segmentation models [7].

In segmentation tasks, CNNs are generally implemented using an encoder–decoder structure. The encoder extracts features and reduces spatial resolution, while the decoder reconstructs the features into a segmentation map at the pixel level [8]. U-Net is one of the most widely used encoder–decoder architectures because it incorporates skip connections to preserve the spatial information necessary for determining the location and boundaries of objects [9]. However, the baseline U-Net still has limitations in capturing the complex characteristics of lesions. Previous research reported that the baseline U-Net achieved a Dice score of 85.40% and an

Intersection over Union (IoU) of 77.09%, indicating that there is still room to improve segmentation performance [10].

ResU-Net extends U-Net by adding residual connections to preserve the flow of information and gradients, thereby enabling the network to be trained more stably [11]. However, not all generated features carry the same level of importance for the lesion area. Attention mechanisms can help amplify relevant features and suppress less important information in skin lesion segmentation [12]. In medical image segmentation in general, these mechanisms are also used to increase the model's focus on target object regions [13].

Several studies have developed the U-Net architecture and its variants to improve segmentation performance. Jafari et al. developed DRU-Net by combining residual and dense connections, but did not implement attention mechanisms or interpretability analysis [5]. Pennisi et al. developed Attention Squeeze U-Net for skin lesion segmentation with relatively low computational complexity [14]. Yuan et al. developed the Improved Dense Recurrent Residual U-Net on the ISIC 2017 dataset by integrating Dense Recurrent Residual Connections, Squeeze-and-Excitation Blocks, and a Feature Adaptation Module, achieving a Dice score of 86.92% and an IoU of 78.86% [15]. Rehman et al. proposed Attention Res-UNet for skin lesion segmentation on the ISIC 2018 dataset by integrating residual connections with Squeeze-and-Excitation, Atrous Spatial Pyramid Pooling, and attention gates, together with Focal Tversky Loss, and reported a Dice score of 89.14% and an IoU of 81.16% [16]. Although Rehman et al. and the present study both use attention-enhanced ResU-Net, their feature enhancement strategies differ: Rehman et al. use SE, ASPP, and attention gates, whereas this study uses Efficient Attention, Pixel Attention, and Multi-Scale Feature Fusion. Their experimental configurations also differ in terms of the number of epochs, batch size, optimizer, and loss function. Furthermore, this study evaluates the contribution of each component through an ablation study and includes Grad-CAM-based interpretability analysis, which was not performed by Rehman et al.

In addition to quantitative performance, interpretability is necessary to determine whether a model's predictions are based on relevant regions of the image. In medical image segmentation, interpretability methods can help reveal the spatial relationships between the features learned by the model and the target objects [9]. Gradient-weighted Class Activation Mapping (Grad-CAM) uses gradient information from convolutional layers to generate activation maps of the regions contributing to the prediction [17]. Grad-CAM has also been applied in skin cancer classification to enhance the transparency of model decisions [18]. However, its application as a tool for interpretability in ResU-Net-based skin lesion segmentation remains limited.

Based on these issues, this study develops ResUNetEA, a ResU-Net architecture that integrates channel attention via the Efficient Attention Block, pixel attention, and multi-scale feature fusion. The model was applied to the ISIC 2018 Task 1 dataset and evaluated using the Dice coefficient and IoU. Grad-CAM was used as a supporting analysis to demonstrate the model's focus on lesion regions and their boundary transition areas. The contributions of this study lie in the integration of feature enhancement mechanisms into ResU-Net and the use of Grad-CAM to support the interpretability of segmentation results.

II. RESEARCH METHODOLOGY

This study employs an experimental approach to develop a skin lesion segmentation model for dermatoscopic images using the ResUNetEA architecture. The research stages include dataset collection, data preparation and preprocessing, dataset splitting, data augmentation, model architecture design, training, evaluation, and interpretability analysis using Grad-CAM. The dataset used is ISIC 2018 Task 1, which provides dermatoscopic images along with ground-truth masks. The research workflow is shown in Figure 1.

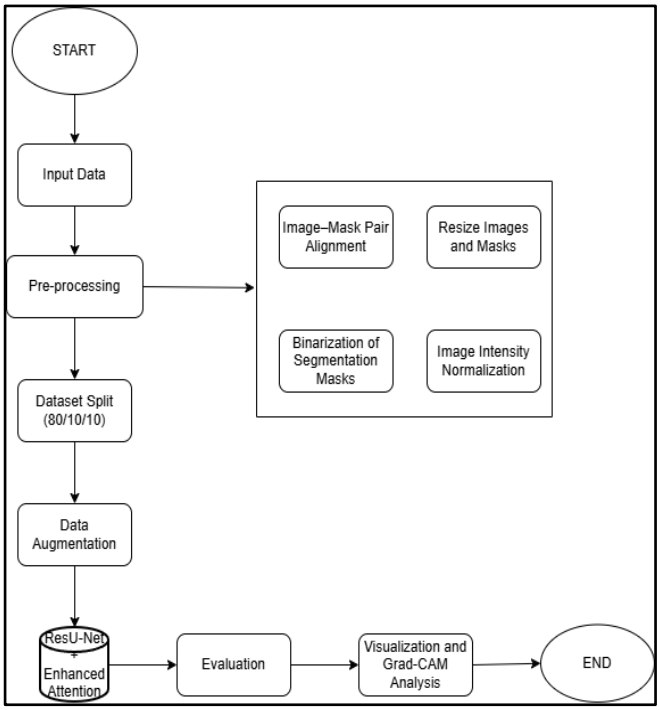

Figure 1. Research Workflow

A. Input Data

The experiments employed the ISIC 2018 Challenge Task 1: Lesion Segmentation dataset accessed through Kaggle . The dataset on that page is a copy of the official ISIC 2018 Task 1 dataset, which is intended for the binary segmentation of skin lesion areas in dermatoscopic images. The research dataset consists of 2,594 dermatoscopic images in RGB format and 2,594 segmentation masks as ground truth. The dermatoscopic images are used as model input, while the masks serve as targets during the training and evaluation processes.

The original images have nonuniform resolutions and exhibit considerable variation in lesion morphology, dimensions, pigmentation, texture, and boundary clarity. Several samples also contain visual interference, including

DOI: 10.34148/teknika.v15i3.1534
TEKNIKA, Volume 15(3), November 2026, pp. 412-420
ISSN 2549-8037, EISSN 2549-8045

hair, shadowing, specular reflection, and nonuniform illumination. Such diversity provides challenging conditions for assessing whether the model can separate lesion tissue from the surrounding normal skin.

B. Pre-processing Data

The pre-processing stage is performed to standardize the data format and ensure that each image has a corresponding mask. Alignment is based on filenames, taking into account the _segmentation suffix in the mask filenames. This stage is necessary to prevent mismatches between images and ground truth.

Next, all images and masks are resized to 256 × 256 pixels. Images are resized using bilinear interpolation, while masks use nearest-neighbor interpolation to ensure that label values remain unchanged. Image intensity values are then normalized from the range of 0–255 to 0–1 to help maintain the stability of the optimization process. Segmentation masks are binarized to 0 for the background and 1 for lesion areas.

C. Dataset Split

The 2,594 image–mask pairs obtained from the ISIC 2018 Task 1 dataset were independently divided into training, validation, and testing subsets using a random seed of 42 and an 80:10:10 proportion (Table 1). This division was not based on the official ISIC 2018 test set split. The resulting subsets consisted of 2,075 image–mask pairs for training, 259 for validation, and 260 for testing. The training data was used to update the model parameters, the validation data was used to monitor performance during training, and the test data was used for the final evaluation.

Table 1. Dataset Distribution Per Subset

Dataset Subset	Number of Images	Proportion
Training	2,075	80%
Validation	259	10%
Testing	260	10%

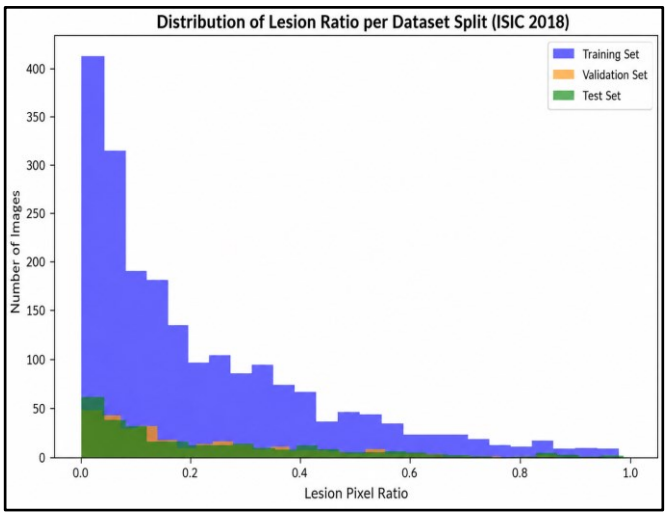

Figure 2. Distribution of Lesion Pixel Ratios Per Subset

In Figure 2, the distribution of lesion-to-pixel ratios shows that most images have relatively small lesion areas compared to the entire image field, while others have lesions that cover large areas. This pattern appears in the training, validation, and test datasets, ensuring that all three sets contain a variety of difficulty levels. Images with a low lesion-to-pixel ratio are more prone to pixel imbalance because the background is far more dominant. Conversely, images with a high lesion ratio require the model to preserve shapes and boundaries that are often irregular. Table 2 shows statistical values of the lesion area ratio for each dataset split.

Table 2. Statistical Values of the Lesion Area Ratio for Each Dataset Split

Dataset Subset	Mean	Std	Min	Max
Training	0.2165	0.2091	0.0030	0.9783
Validation	0.1889	0.1802	0.0032	0.8679
Testing	0.2193	0.2248	0.0038	0.9877

D. Data Augmentation

Augmentation was restricted to the training subset to broaden sample variability and reduce the risk of overfitting. The transformations used include horizontal flip, vertical flip, rotation, brightness and contrast adjustments, CLAHE, elastic transform, random resized cropping, and Gaussian blur (Table 3). Geometric transformations are applied simultaneously to both the image and the mask to maintain spatial correspondence. Figure 3 is a sample of the augmented data.

Table 3. Data Augmentation Parameters

Augmentation Method	Parameter Settings
HorizontalFlip	p = 0.5
VerticalFlip	p = 0.5
Rotate	limit = ±15°, p = 0.5
ElasticTransform	alpha = 10, sigma = 5, p = 0.2
RandomResized Crop	size = (256,256), scale = (0.8–1.0), ratio = (0.9–1.1), p = 0.3
RandomBrightness Contrast	brightness_limit = 0.15, contrast_limit = 0.15, p = 0.3
CLAHE	clip_limit = 2.0, tile_grid_size = (8,8), p = 0.3
GaussianBlur	blur_limit = 3, p = 0.2

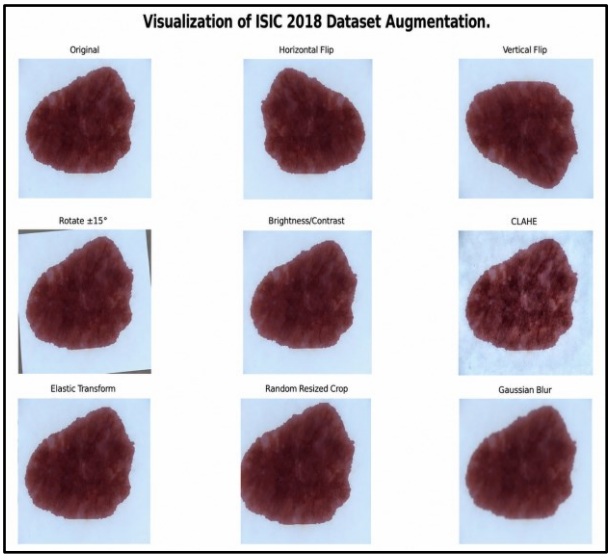

Figure 3. Examples of Augmentation Applied to ISIC 2018 Training Data

E. Model Architecture

ResUNetEA retains the main U-Net structure consisting of an encoder, bottleneck, decoder, and skip connections. The encoder extracts features using 64, 128, 256, and 512 channels, while max-pooling reduces the spatial resolution. The bottleneck processes the deepest features using 1,024 channels. The encoder extracts features and reduces spatial resolution, whereas the decoder reconstructs the segmentation map [8]. Skip connections preserve spatial information required to reconstruct object locations and boundaries [9]. Meanwhile, the ResNet concept is implemented through residual connections within each ConvBlock to maintain information flow and improve training stability [11].

Each ConvBlock contains two 3×3 convolutions in the main path and a 1×1 convolution in the residual path to adjust the channel dimensions. The outputs of both paths are summed to form the residual connection. This differs from the U-Net skip connection, which concatenates encoder and decoder features at the same spatial resolution.

After each ConvBlock in the encoder and decoder, an Efficient Attention Block is applied as a channel-attention mechanism. Attention strengthens relevant features and suppresses less important information in skin lesion and medical image segmentation [12][13]. In the decoder, transposed convolution restores the spatial resolution, followed by concatenation with the corresponding encoder features. Pixel Attention is applied only to the final decoder output to enhance the model's spatial focus on lesion regions and boundary-transition areas.

The outputs of the four decoder stages, d1, d2, d3, and d4, are combined through multi-scale feature fusion. Each feature is projected to 64 channels, while lower-resolution features are upsampled to match the resolution of d1. The fused features are processed using a 3×3 convolution, followed by a 1×1 convolution to produce a single-channel binary segmentation mask of 256×256 pixels. The overall architecture is illustrated in Figure 4.

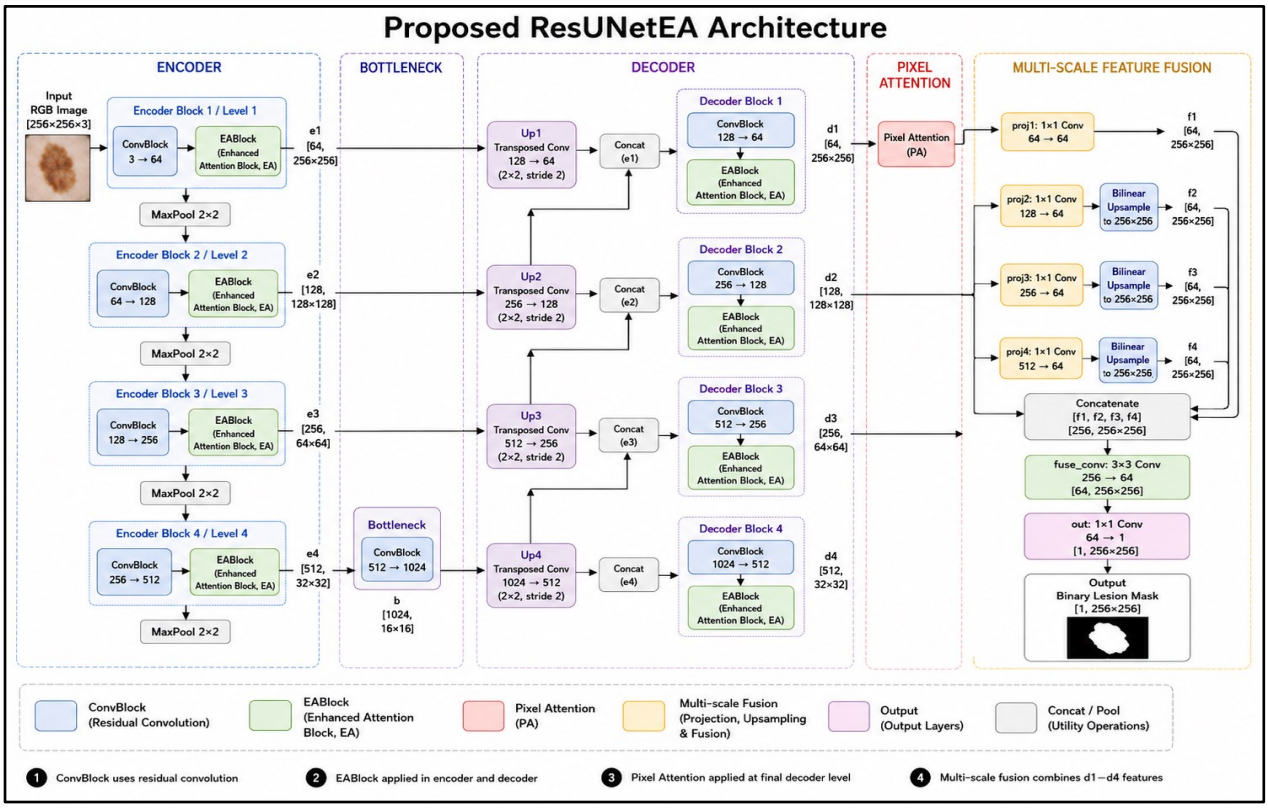

Figure 4. Proposed ResUNetEA Architecture

DOI: 10.34148/teknika.v15i3.1534
TEKNIKA, Volume 15(3), November 2026, pp. 412-420
ISSN 2549-8037, EISSN 2549-8045

CC BY SA

F. Training Strategy

Model optimization was carried out for a maximum of 40 epochs with eight image–mask pairs processed per batch. Adam was selected as the optimizer and initialized with a learning rate of 0.0001. The training configuration is summarized in Table 4. Dermoscopic images were used as model inputs, while binary segmentation masks served as the ground-truth targets. During each training iteration, the predicted segmentation logits were compared with the ground-truth masks, and the model parameters were updated through backpropagation.

Table 4. Model Training Configuration

Parameter	Value
Input Size	256 x 256 piksel
Batch Size	8
Max Epoch	40 epoch
Optimizer	Adam
Learning Rate	1e-4
Loss Function	Combined Loss
Komponen *Loss*	0,5 BCEWithLogitsLoss + 0,5 Dice Loss
Scheduler	ReduceLROnPlateau
Callback	Early stopping(7patience), model checkpoint
Best Model Selection Criterion	Validation Dice
Mixed Precision	GradScaler/AMP
GPU	NVIDIA Tesla T4

The optimization objective assigned equal contributions to BCEWithLogitsLoss and Dice Loss. The first component evaluates binary prediction errors at the individual-pixel level, whereas the second penalizes insufficient spatial agreement between predicted lesion regions and their annotations. Validation Dice was monitored by ReduceLROnPlateau; when the score failed to improve for three successive epochs, the learning rate was multiplied by 0.5. Automatic Mixed Precision was enabled to reduce memory consumption and improve computational efficiency. In addition, manual early stopping was configured with a patience of seven epochs, while the network weights associated with the highest validation Dice were retained for final testing.

G. Evaluation Method

Model performance was evaluated using the Dice coefficient (Formula 1) and Intersection over Union/IoU (Formula 2), which measure the overlap between the predicted segmentation mask and the ground-truth mask [19]. The model output was converted into pixel-wise probabilities using the sigmoid function and then binarized with a threshold of 0.5 to obtain the final segmentation mask for Dice and IoU calculation.

In addition, Grad-CAM was applied to the bottleneck layer to generate activation heatmaps as a supporting interpretability analysis. Feature maps and gradients from the bottleneck layer were captured using forward and backward hooks. The gradients were averaged to obtain channel-wise weights, which were combined with the corresponding feature maps to produce the activation map. After applying ReLU, the activation map was resized to the input-image resolution and overlaid on the original image. The resulting heatmaps were compared with the original images, ground-truth masks, and predicted masks to identify the regions contributing to the segmentation output. Grad-CAM was used as a supporting analysis of the model's spatial focus rather than as a quantitative evaluation metric, but to examine the spatial regions that contributed to the segmentation output.

$$Dice = \frac{2|A \cap B|}{|A|+|B|} \quad (1)$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

where A is the pixel area of the model's segmentation result, and B is the pixel area of the ground-truth mask.

III. RESULT AND DISCUSSION

A. Training Performance

The training process was evaluated using the changes in loss and Dice coefficient over 40 epochs. Figure 5 shows that the training loss fell from 0.3102 at the beginning of training to 0.1025 in the final epoch, while the validation loss decreased from 0.2565 to 0.1340. This downward trend indicates that the model progressively reduced the difference between the predicted segmentation masks and the ground-truth masks during training.

The Dice coefficient also increased throughout the training process, as presented in Figure 6. The training Dice improved from 0.7448 to 0.9111, whereas the validation Dice increased from 0.7601 to 0.8803. The improvement in both training and validation Dice indicates that the model increasingly learned to identify the location, shape, and coverage of lesion regions. Minor fluctuations in the validation values occurred during several epochs, but the overall trend remained upward.

The simultaneous decrease in loss and increase in Dice demonstrate that the optimization process proceeded consistently. The combined use of BCEWithLogitsLoss and Dice Loss allowed the model to learn both pixel-wise classification and spatial overlap between the predicted and ground-truth masks. In addition, the Adam optimizer and ReduceLROnPlateau scheduler supported parameter updates when the validation Dice began to stagnate.

Although the final training Dice was higher than the validation Dice, the difference remained moderate and the validation performance continued to improve. This result indicates that the model did not only memorize the training data but was also able to maintain its segmentation performance on unseen validation data. The model with the highest validation Dice was subsequently selected for final evaluation on the test set.

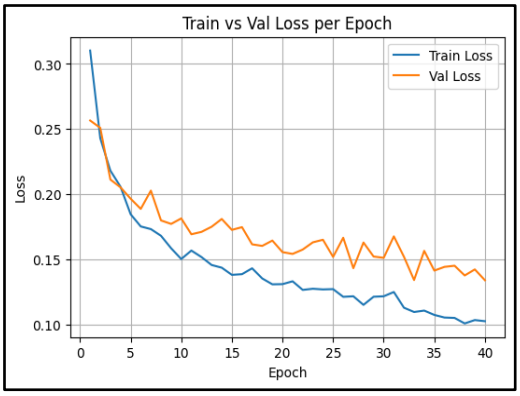

Figure 5. Training and Validation Loss Over 40 Epochs

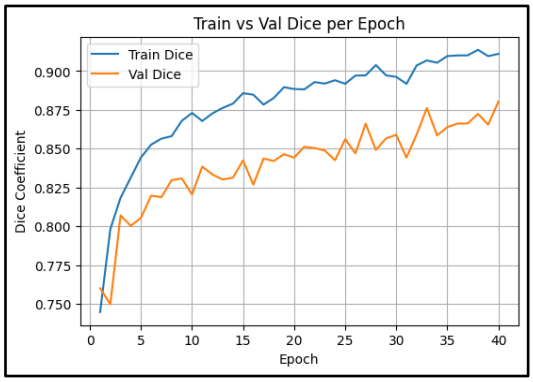

Figure 6. Training and Validation Dice Over 40 Epochs

B. Test Performance

The model selected based on the highest validation Dice score was evaluated on the test set to measure its ability to generalize to previously unseen data. The evaluation results are summarized in Table 5. The model achieved a test loss of 0.1264, a Dice coefficient of 0.8914, and an Intersection over Union (IoU) of 0.8080. These results indicate a high degree of overlap between the predicted segmentation masks and the corresponding ground-truth masks

Table 5. Model Evaluation Results

Dataset	Loss	Dice Score	IoU
Train	0.0946	0.9111	0.8530
Validation	0.1340	0.8802	0.7951
Test	0.1264	0.8914	0.8080

The training set produced the highest Dice and IoU values because these data were directly involved in updating the model parameters. However, the validation and test results remained relatively close to the training results. The test Dice was slightly higher than the validation Dice, while the test loss was lower than the validation loss. This variation can occur because the validation and test sets contain different lesion characteristics and difficulty levels.

The Dice value was higher than the IoU because the two metrics use different formulations. Dice gives greater weight to the intersection between the predicted and ground-truth regions, whereas IoU compares the intersection directly with the total combined area. Nevertheless, both metrics consistently indicate that the predicted masks substantially overlapped with the lesion regions in the ground truth.

Overall, the relatively small differences among the training, validation, and test results indicate that ResUNetEA maintained good generalization performance. The model was able to segment lesion regions on unseen test images without showing a substantial decline in performance compared with the training data. These findings support the effectiveness of the combined residual blocks, attention mechanisms, and multi-scale feature fusion used in the proposed architecture.

C. Segmentation and Grad-CAM Analysis

Figure 7 presents representative examples consisting of the original dermoscopic image, the ground-truth mask, the predicted mask, and the Grad-CAM visualization. In general, the predicted masks were able to follow the location and overall shape of the lesions. Several predictions closely matched the ground truth, although minor differences remained along some lesion boundaries.

In some cases, the predicted masks showed slight over-segmentation or under-segmentation. Over-segmentation occurred when the predicted lesion area extended beyond the ground-truth boundary, whereas under-segmentation occurred when part of the lesion area was not included in the prediction. These differences were generally observed in images with low contrast, irregular lesion shapes, unclear boundaries, or artifacts such as hair, shadows, and uneven illumination.

Despite these boundary differences, the overall similarity between the predicted masks and ground truth was consistent with the test Dice coefficient of 0.8914 and IoU of 0.8080. These values indicate that the predicted lesion regions substantially overlapped with the corresponding ground-truth masks.

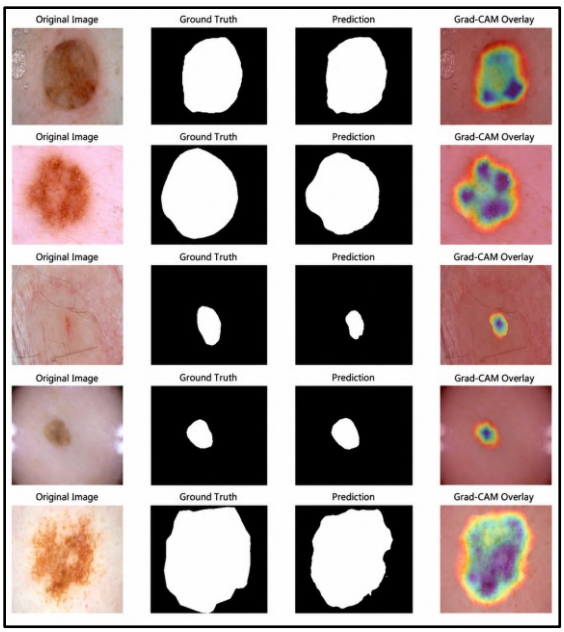

Figure 7. Visualization of Original Images, Ground-Truth Masks, Predicted Masks, and Grad-CAM Heatmaps

The Grad-CAM visualizations showed that high-activation regions were generally concentrated on the lesion area and around the lesion boundaries. Red and yellow regions indicate areas that contributed strongly to the model output, whereas blue regions represent lower contributions. In several examples, stronger activation appeared near the boundary-transition area, indicating that edge information played an important role in distinguishing lesion regions from normal skin.

Grad-CAM was used only as a supporting interpretability analysis and not as a quantitative measure of segmentation accuracy. Therefore, its results should be interpreted together with the predicted masks, ground-truth masks, Dice coefficient, and IoU. Overall, the visualization results support the quantitative evaluation by showing that ResUNetEA generally focused on spatial regions relevant to lesion segmentation.

D. Ablation Study

An ablation study was conducted to analyze the contribution of each component in the ResU-NetEA architecture. Training and validation curves were analyzed to assess model performance, training stability, and generalization across epochs for each model configuration.

a.) ResU-Net: The baseline ResU-Net was trained under the same experimental settings as the other ablation configurations, with a maximum of 40 epochs and an early stopping patience of seven epochs (Figure 8a). The training curves show a gradual decrease in loss and an increase in Dice, with training loss decreasing from approximately 0.337 to 0.116 and training Dice increasing from approximately 0.719 to 0.902. The best validation performance was achieved at epoch 23, with a validation Dice of 0.8590 and a training Dice of 0.8929. Although the validation performance generally improved during training, the validation Dice exhibited noticeable fluctuations across several epochs. After reaching its best value, the validation Dice showed no further improvement, while the training Dice continued to increase, indicating an increasing generalization gap. As no improvement was observed for seven consecutive epochs, the early stopping criterion was triggered.

b.) ResU-Net + Efficient Attention: The addition of Efficient Attention (EA) did not provide an improvement in validation performance compared with the baseline ResU-Net under the current experimental setting (Figure 8b). Based on the model selected using the highest validation Dice, this configuration achieved a validation Dice of 0.8549 at epoch 12, with a training Dice of 0.8783. This result was slightly lower than that of the baseline ResU-Net (0.8590). After reaching its best performance, the validation Dice fluctuated without showing further improvement, whereas the training Dice continued to increase. This behavior suggests that the model increasingly adapted to the training data without achieving better validation performance. Furthermore, training stopped earlier at epoch 19 due to the early stopping criterion.

c.) ResU-Net + Efficient Attention + Pixel Attention: The addition of Pixel Attention to ResU-Net with Efficient Attention resulted in improved validation performance and more controlled fluctuations in the validation Dice (Figure 8c). Based on the model selected according to the highest validation Dice, this configuration achieved a best validation Dice of 0.8675 at epoch 29, with a training Dice of 0.9003 at the same epoch. This result was higher than both the baseline ResU-Net (0.8590) and ResU-Net + Efficient Attention (0.8549), representing an improvement of approximately 0.0126 over the ResU-Net + Efficient Attention configuration. Although fluctuations were still observed across several epochs and a gap between training and validation performance remained, the validation Dice consistently maintained a higher performance level than the previous configuration. These results suggest that the addition of Pixel Attention complements Efficient Attention by enabling more effective feature refinement, resulting in improved validation performance for skin lesion segmentation.

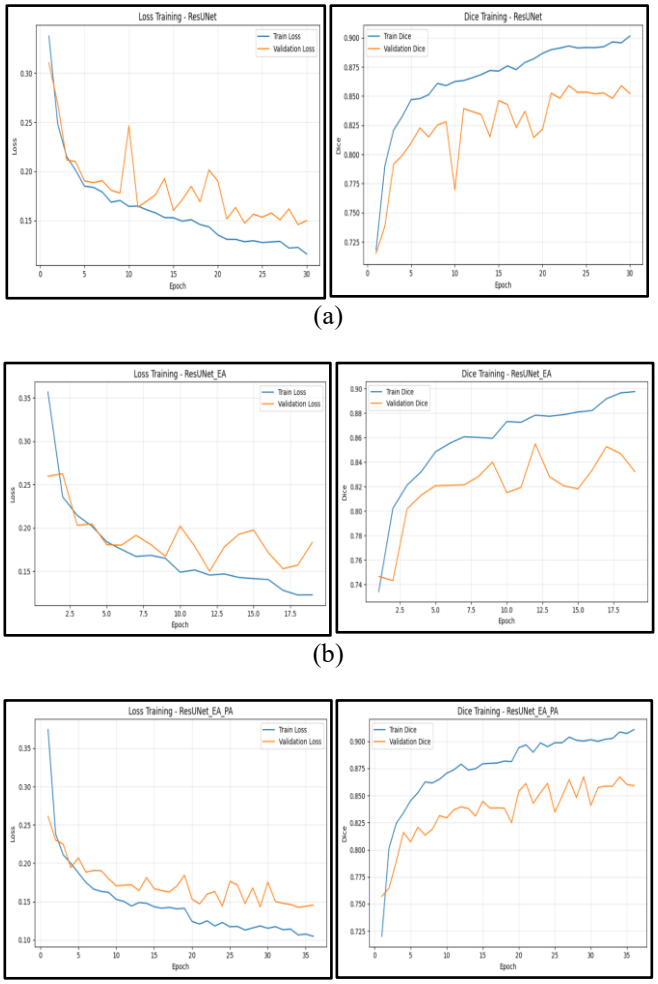

Figure 8. Comparison of training and validation loss and Dice curves across the ablation configurations: (a) ResUNet, (b) ResUNet + Efficient Attention, and (c) ResUNet + Efficient Attention + Pixel Attention.

E. Comparison with Prior Work

Table 6 presents a cross-study comparison between the proposed ResUNetEA and several skin lesion segmentation models reported by Yuan et al. [15]. The proposed method was evaluated using the ISIC dataset, while Yuan et al. [15] also used the ISIC dataset. ResUNetEA achieved a Dice coefficient of 0.8914 and an IoU of 0.8080, which were numerically higher than those reported for U-Net, R2U-Net, DA-Net, and IDR2U-Net.

However, the results should be interpreted as a general performance reference, as the two studies differ in dataset version, data partitioning, preprocessing, augmentation, and training configurations. Therefore, the numerical comparison does not represent a controlled direct comparison under identical experimental settings.

Table 6. Cross-Study Comparison with Results Reported by Yuan et al. [15]

Method	Dice	Jaccard/IoU
U-Net [15]	0.7898	0.6893
R2U-Net [15]	0.8247	0.7409
DA-Net [15]	0.8437	0.7537
IDR2U-Net [15]	0.8692	0.7886
ResUNetEA (Proposed)	**0.8914**	**0.8080**

IV. CONCLUSION

This study developed ResUNetEA for skin lesion segmentation in dermoscopic images by integrating residual blocks, Efficient Attention Blocks, Pixel Attention, and multi-scale feature fusion into the U-Net architecture. The model was trained and evaluated using the ISIC skin lesion dataset with Dice coefficient and Intersection over Union as the primary evaluation metrics.

The proposed model achieved a test loss of 0.1264, a Dice coefficient of 0.8914, and an IoU of 0.8080. These results indicate that the predicted masks generally achieved a high degree of overlap with the ground-truth lesion regions. Visual analysis also showed that the predicted masks were able to follow the location and overall shape of the lesions, although minor over-segmentation and under-segmentation remained in some boundary regions.

Grad-CAM visualizations indicated that the model primarily focused on lesion regions and boundary-transition areas. Therefore, Grad-CAM provided supporting visual information regarding the spatial regions that contributed to the segmentation output, although it was not used as a quantitative measure of segmentation accuracy.

Overall, ResUNetEA demonstrated good segmentation performance and maintained consistent results on unseen test data. Future work may evaluate the model on external clinical datasets, improve its robustness against artifacts and unclear lesion boundaries.

REFERENCES

[1] G. Çetinel, B. M. Aydın, S. Gül, D. Akgün, and R. Öztaş Kara, "An evaluation of skin lesion segmentation using deep learning architectures," *Sakarya University Journal of Computer and Information Sciences*, vol. 7, no. 3, pp. 449–459, 2024, doi: 10.35377/saucis.1543993

[2] S. Rachmawati, A. E. Minarno, and Y. Azhar, "Klasifikasi Lesi Kulit Menggunakan Convolutional Neural Network," *Jurnal Repositor*, vol. 4, no. 3, pp. 363–372, 2024, doi: 10.22219/repositor.v4i3.32327.

[3] R. A. S. Lubis, "Evaluasi tingkat akurasi penggunaan metode deep learning dengan algoritma convolutional neural network pada citra dermatoskopi untuk efisiensi diagnosis kutaneus melanoma: Sebuah telaah literatur sistematis," *Holistik Jurnal Kesehatan*, vol. 18, no. 3, pp. 289–299, May 2024, doi: 10.33024/hjk.v18i3.167.

[4] Z. Mirikharaji, K. Abhishek, A. Bissoto, C. Barata, S. Avila, E. Valle, M. E. Celebi, and G. Hamarneh, "A survey on deep learning for skin lesion segmentation," *Medical Image Analysis*, vol. 88, p. 102863, 2023, doi: 10.1016/j.media.2023.102863.

[5] M. Jafari, D. Auer, S. Francis, J. Garibaldi, and X. Chen, "DRU-Net: An Efficient Deep Convolutional Neural Network for Medical Image Segmentation," *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, 2020, pp. 1144–1148, doi: 10.1109/ISBI45749.2020.9098391.

[6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022, doi: 10.1109/TPAMI.2021.3059968.

[7] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Processing*, vol. 16, no. 5, pp. 1243–1267, 2022, doi: 10.1049/ipr2.12419.

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, 2021, doi: 10.1038/s41592-020-01008-z.

[9] C. Seibold, J. Künzel, A. Hilsmann, and P. Eisert, "From explanations to segmentation: Using explainable AI for image segmentation," in *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 616–626, 2022, doi: 10.5220/0010893600003124.

[10] T.-D.-T. Phan, S.-H. Kim, H.-J. Yang, and G.-S. Lee, "Skin lesion segmentation by U-Net with adaptive skip connection and structural awareness," *Applied Sciences*, vol. 11, no. 10, p. 4528, 2021, doi: 10.3390/app11104528.

[11] L. Huang, A. Miron, K. Hone, and Y. Li, "Segmenting medical images: From U-Net to Res-UNet and nnU-

Net," *arXiv preprint arXiv:2407.04353*, 2024, doi: 10.48550/arXiv.2407.04353.

[12] P. T. Le, B.-T. Pham, C.-C. Chang, Y.-C. Hsu, T.-C. Tai, Y.-H. Li, and J.-C. Wang, "Anti-Aliasing Attention U-Net Model for Skin Lesion Segmentation," *Diagnostics*, vol. 13, no. 8, p. 1460, 2023, doi: 10.3390/diagnostics13081460.

[13] A. Qurri and M. Almekkawy, "Improved U-Net with attention for medical image segmentation," *Sensors*, vol. 23, no. 20, Art. no. 8589, 2023, doi: 10.3390/s23208589.

[14] Pennisi, D. D. Bloisi, V. Suriani, D. Nardi, A. Facchiano, and A. R. Giampetruzzi, "Skin lesion area segmentation using Attention Squeeze U-Net for embedded devices," *Journal of Digital Imaging*, vol. 35, pp. 1217–1230, 2022, doi: 10.1007/s10278-022-00634-7.

[15] Y. Yuan, D. Zhao, and Z. Luo, "Improved dense recurrent residual U-Net for skin lesion segmentation," *Journal of Electronics and Information Science*, vol. 9, no. 1, pp. 27–37, 2024, doi: 10.23977/jeis.2024.090105.

[16] Rehman, M. A. Butt, and M. Zaman, "Attention Res-UNet: Attention residual U-Net with focal Tversky loss for skin lesion segmentation," *International Journal of Data Science and Smart Technology*, vol. 15, no. 1, pp. 1–17, 2023, doi: 10.4018/IJDSST.315756.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017, doi: 10.1109/ICCV.2017.74.

[18] M. E. Ramadhan and J. Zeniarja, "Implementation of Deep Transfer Learning and Explainable AI in Skin Cancer Classification," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 5, pp. 2266–2279, Sep. 2025, doi: 10.32520/stmsi.v14i5.5425.

[19] S. Sonia and Y. Yohannes, "Penerapan model U-Net untuk segmentasi gigi pada citra radiografi panoramik orang dewasa," *Jurnal Algoritme*, vol. 5, no. 2, pp. 188–198, 2025, doi: 10.35957/algoritme.v5i2.10965.