Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.075537

ARTICLE

LANET: A Deep Lightweight Attention Network for Skin Cancer Segmentation

Abdulrahman Dira Khalaf[1,2,*], Hazlina Hamdan[1,*], Alfian Abdul Halin[1] and Noridayu Manshor[1]

[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), Serdang, Malaysia
[2]Department of Computer Center, University of Fallujah, Anbar, Iraq

*Corresponding Authors: Hazlina Hamdan. Email: hazlina@upm.edu.my; Abdulrahman Dira Khalaf. Email: adk1973@uofallujah.edu.iq

Received: 03 November 2025; Accepted: 30 December 2025; Published: 27 May 2026

ABSTRACT: Current automated lesion segmentation methods have limited success, particularly for segmenting small, irregular, or heterogeneous lesions. Moreover, such models require significant computational power, which restricts their scalability and clinical application. To overcome these limitations, a lightweight LANET, which is a layer-attention network based on an encoder–decoder deep-learning architecture, has the explicit goal of increasing the segmentation performance and computational efficiency. The LANET is coupled with three new modules: (i) an attention module that includes a depthwise separable convolution operator to reduce the number of parameters, (ii) a custom attention mechanism, and (iii) an atrous spatial pyramid pooling (ASPP) module designed to model substantial features at multiple scales under ideal conditions. Through experiments on benchmark datasets, LANET demonstrated robustness, resulting in accuracies of 96.44%, 96.8%, 96.3%, and 97.9% for HAM10000, ISIC 2017, ISIC 2018, and PH2, respectively. These results exceed those of classical architectures, such as U-Net, UNet++, and DeepLabv3+, as well as more recent state-of-the-art approaches. Simultaneously, it integrates only 846,786 parameters of the LANET, which leads to a minimum number of overall parameters, and thus, lower computational costs in terms of inference. Furthermore, techniques such as Grad-CAM and activation-map visualizations help explain model decisions and highlight clinically relevant regions. The results show that the LANET provides a robust, scalable, and interpretable real-time segmentation system. This design specifically improves the segmentation of small- or low-contrast lesions. This approach offers a practical path for integrating efficient segmentation models into clinical workflows for skin disease analyses.

KEYWORDS: Attention; deep learning; lightweight models; segmentation; skin cancer; U-Net

1 Introduction

Skin cancer is an enormous issue worldwide and is becoming more common every year. By 2025, the USA is expected to have 104,960 instances [1]. This statistic shows that the disease is becoming more common and that there is a clear need for precise tests. Traditional machine learning (ML) techniques for skin lesion analysis typically rely on manually crafted features, such as color, texture, and shape descriptors [2,3]. Although these approaches offer interpretability, the significant heterogeneity in lesion appearance makes interpretation particularly difficult, particularly in low-contrast and uneven-margin cases. In contrast, deep learning (DL) methods learn hierarchical representations directly from data and operate on raw images, resulting in more robust and accurate segmentation performance [4].

The performance of segmentation has improved in recent years; however, existing methods exhibit inherent trade-offs. Fully convolutional neural network (FCNN) architectures such as U-Net [5] and its variants, including UNet++ [6], reduce the semantic gap between the encoder and decoder layers and

achieve strong segmentation accuracy. Despite their success, these models often struggle with small or subtle lesions and require significant computational resources, which affects their reproducibility and real-time usability [7]. Models such as Dermo-Seg [8] and MSREA-Net [9] enhance feature aggregation through attention mechanisms and improve lesion identification; however, their complex architectures lead to high computational costs. Transformer-based approaches such as DermoSegDiff [10], CTH-Net [11], and UNETR [12] improve boundary modelling by capturing long-range dependencies. However, they require considerable computational resources and struggle with narrow or poorly defined margins.

Many existing segmentation models fail to capture subtle or poorly defined lesion edges and often require high computational power [9]. The LANET is designed to address these issues by improving the boundary sensitivity while maintaining a low computational overhead. Artifacts such as hair, ruler marks, and lighting variations further degrade the segmentation accuracy [13]. To address these challenges, we introduce LANET, a lightweight attention-based network that combines depthwise convolutions, attention modules, and atrous spatial pyramid pooling (ASPP) [14], which enables effective multiscale context aggregation without increasing computational complexity. The design specifically targeted small and low-contrast lesion boundaries. The LANET balances high performance with computational efficiency, making it both correct and clinically practical. The main contributions of this study are as follows:

1. LANET reduces the number of parameters while improving boundary detection, particularly for small- or low-contrast lesions.
2. The architecture integrates lightweight ASPP and attention modules to capture the lesion structure without increasing computational cost.
3. LANET is evaluated on four public datasets with diverse imaging conditions and lesion variations.
4. The results show that the LANET maintains a strong accuracy while preserving the boundary detail and operating with a compact parameter count.

This paper presents LANET, a lightweight attention-based segmentation framework for pixel-level skin lesion delineation. The LANET combines depthwise separable convolutions, a customized attention mechanism, and an ASPP module within an encoder–decoder structure to enhance the boundary sensitivity at a low computational cost. The model was evaluated using four benchmark datasets: HAM10000, ISIC 2017, ISIC 2018, and PH2, covering diverse imaging conditions and lesion types. The LANET achieves superior Dice, IoU, and sensitivity scores compared with the baseline models. These results demonstrate the effectiveness of the compact architecture for accurate lesion boundary extraction and support its potential clinical use. Section 2 summarizes the related work. Section 3 describes the datasets and LANET method. Section 4 presents experimental results and interpretability analysis. Section 5 reports ablation studies, followed by a discussion in Section 6. Section 7 outlines the limitations, and Section 8 provides future research directions.

2 Related Work

Accurate diagnosis of skin cancer depends on reliable lesion segmentation. Recent research has focused on deep learning (DL) and hybrid models [12,15], particularly on publicly available dermoscopic datasets, such as ISIC [16]. This section summarizes key studies across four methodological categories: (i) CNN-based, (ii) transformer-based, (iii) attention-based, and (iv) lightweight segmentation. Each category offers unique strengths and limitations that motivate the design of LANET.

2.1 CNN-Based Model

Convolutional neural networks (CNNs) remain the foundation of most methods used to segment skin lesions. Classic architectures such as U-Net [5] and its derivatives have been widely extended to improve pixel-level accuracy. For example, UNet++ [6] and DFF-UNet [7] integrated deep feature fusion, whereas MRP-UNet [17] incorporated multiscale input fusion and pyramid dilated convolutions to capture lesions of varying sizes better. These fully convolutional networks (FCNs) highlight the strength of CNNs in local feature learning; however, they often struggle to effectively capture the global context. To address this, several models have begun to embed attention- or transformer-based modules in CNN frameworks.

2.2 Transformer-Based Model

Transformer-based methods address the limitations of CNN by capturing long-range dependencies and local features. Mas-TransUNet [18] incorporates CNN encoding and transformer modules for simultaneous global and local feature extractions. DuaSkinSeg is a two-encoder-based method using MobileNetV2 and Vision Transformer to utilize convolutional features with contextual representation [19]. BDFormer employs a boundary-aware dual-decoder transformer to improve edge accuracy [20], and EM-Net incorporates a morphology-aware design to better represent the lesion structure [21]. These approaches enhance segmentation performance, but often introduce greater model complexity, limiting their suitability for real-time or resource-constrained environments. However, their computational demands hinder their automatic clinical use in real-time applications, which require speed and hardware efficiency.

2.3 Attention-Based Model

One way to improve the precision and performance is to introduce an attention mechanism. Attention Squeeze U-Net focuses on embedded devices [22]. This is not due to the dataset size, but to the deployment environment, as large datasets are mostly required only for offline model training. Attention networks are intrinsically lightweight, which improves their performance, while making them more compact. While alternative efficient methods, such as boundary-aware attention and a hybrid CNN-transformer encoder-decoder, enhance the fusion of region features, they also introduce additional computational overhead [11]. They identified a trade-off between segmentation accuracy and clinical applicability.

2.4 Lightweight Models

Lightweight versions are popular for real-time applications. DSNet is a lightweight CNN with depthwise convolutions and boundary-aware loss, achieving strong segmentation performance with fewer parameters [23]. In addition to architectural design, a model explores compression techniques such as pruning, quantization, and knowledge distillation to reduce inference time and memory usage. As such, “lightweight” in clinical settings involves a smaller parameter count and faster runtime on mobile or embedded systems, where GPUs are not available. Several strategies based on swine transformers with parameter sharing and anti-aliasing upsampling have been investigated; however, challenges remain in the ambiguous segmentation of lesions [24]. The use of lightweight designs allows point-of-care implementation to be integrated into research models.

This categorization emphasizes incremental developments in CNN-based, transformer-based, and lightweight methods, thereby providing high segmentation accuracy and clinical utility trade-offs. Despite notable progress in lesion segmentation, existing approaches face three significant challenges: (i) high computational costs that limit real-time use, (ii) reduced accuracy for small- or low-contrast lesions, and (iii) limited adaptability in resource-constrained clinical environments. These limitations indicate the importance of a lightweight yet accurate model. In the following section, LANET explain that lightweight designs

help bridge the gap between research models and point-of-care deployment. These gaps were addressed by combining depthwise convolutions, a customized attention block, and an optimized ASPP module.

3 Materials and Methods

LANET was implemented in MATLAB 2024a and trained on an Ubuntu server with an NVIDIA Tesla V100 GPU (16 GB). The model was trained on four datasets with random allocation and an 80:10:10 split for training, validation, and testing. The images were resized to 224 × 224 pixels for training, and a batch size of 32 pixels was used. The optimization was based on stochastic gradient descent with momentum (SGDM), and 20 epochs.

3.1 Dataset

Table 1 lists the human-to-machine ratio (HAM10000), which consists of images and the ground truth [25]. The dataset consists of 10,015 images organized into seven categories, with a resolution of 600 × 450 pixels in the JPEG format. The ground-truth dataset is the same size, but it is in binary images in the PNG format. Table 2 lists other public datasets used in this study. The ISIC 2017 and 2018 datasets are widely used as benchmarks for skin cancer detection and consist of dermoscopic images. The ISIC 2017 comprises 2750 images and supports three tasks: lesion segmentation, dermoscopic feature identification, and lesion categorization into melanoma, seborrheic keratosis, and benign nevi [16]. The ISIC 2018 dataset comprises 3694 images. It is used for multiclass classification across seven diagnostic categories [26]. Both datasets offer expert annotations and have established themselves as benchmarks for assessing DL models in automated skin-lesion analyses. The PH2 dataset consists of 200 skin lesion images, each with its corresponding label, at a resolution of 768 × 560 [27]. It is commonly used to assess the generality and efficacy of a model. Fig. 1 shows the sample images and corresponding ground-truth masks from all four datasets.

Table 1: Seven classes of HAM10000.

Class	Abbreviation	Number of Images
Actinic keratoses	AKIEC	327
Basal cell carcinoma	BCC	514
Benign keratoses	BKL	1099
Dermatofibroma	DF	115
Melanoma	MEL	1113
Melanocytic nevi	NV	6705
Vascular lesions	VASC	142

Table 2: Skin cancer public datasets.

Dataset	No. of Images for the 80:10:10 Split	Image Size	Classes
HAM10000	10,015	600 × 450	AKIEC, BKL, DF, VASC, MEL, NV, BCC
ISIC 2018	3694	Different sizes	AKIEC, BKL, DF, VASC, MEL, NV, BCC
ISIC 2017	2750	Different sizes	Melanoma, Nevus, Seborrheic Keratosis
PH2	200	768 × 560	Melanoma, Common Nevi, Atypical Nevi

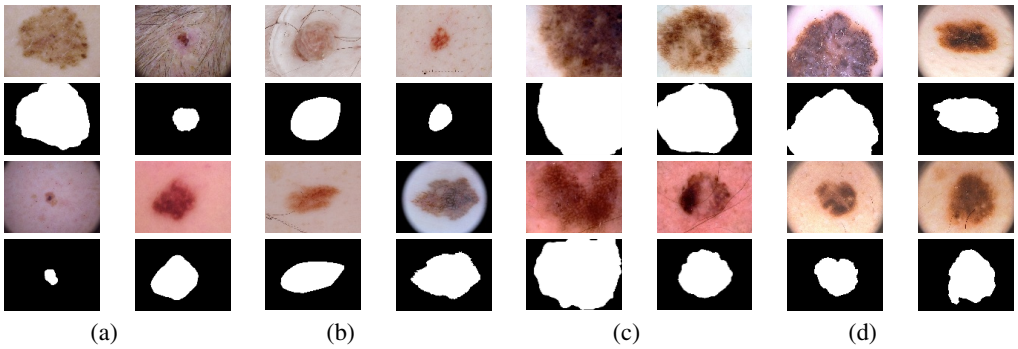


Figure 1: Examples of typical dermoscopic images from each dataset; each row shows the source image (color) with its corresponding ground truth mask (black and wight) for: (**a**) HAM10000; (**b**) ISIC 2018; (**c**) ISIC 2017; and (**d**) PH2.

3.2 Proposed Method

LANET is based on U-Net [5]. The encoder-decoder architecture was designed for efficiency and segmentation accuracy (Fig. 2). It includes three main components: (i) an encoder that incorporates attention mechanisms and grouped convolutions; (ii) a decoder; and (iii) an ASPP module for multiscale contextual understanding. The model accepts three channels (R, G, and B) as input for an image of size 224 × 224 × 3, and outputs a pixel-wise segmentation map with two classes (lesion and background).

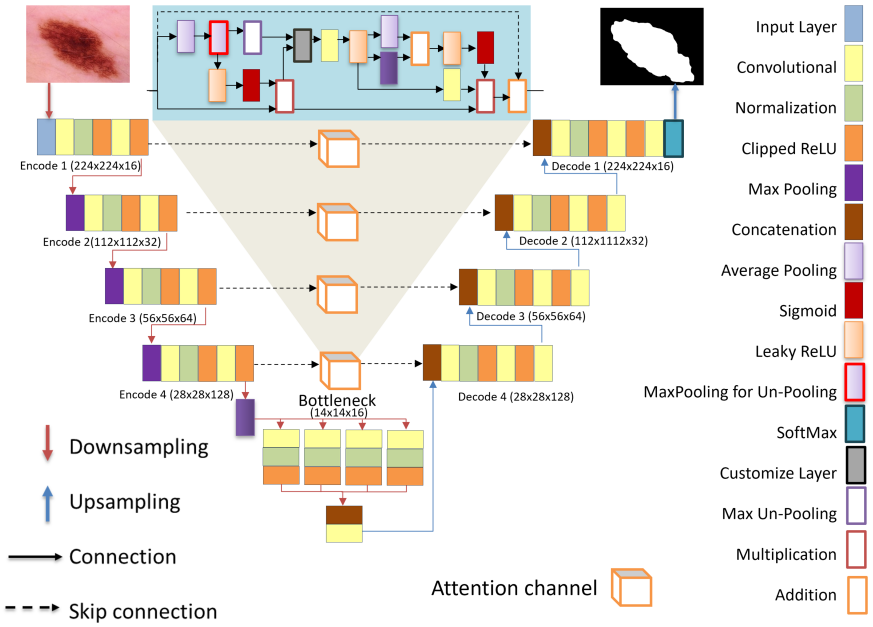


Figure 2: The main framework of the LANET.

3.2.1 Encoder

The encoder forms four blocks that extract the features at diverse levels. Each block had a 2D convolutional layer, batch normalization, and three activation functions. The Rectified Linear Unit (ReLU), leaky ReLU, and clipped ReLU activation functions were computed using Eqs. (1)–(3). Depthwise separable convolutions were employed to improve computational efficiency. The scale factor was 0.01 and the ceiling was 6. Multiscale pooling was applied through average and max-pooling operations, enabling feature

representations at varying spatial resolutions. Moreover, channel attention was incorporated into each block via sigmoid gating from each source with a residual connection to increase feature selectivity. The pattern used 16, 32, 64, and 128 feature maps to show a gradual increase in feature depth across the encoder stages. Fig. 3 shows the feature map.

$$f(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

$$f(x) = \begin{cases} x, & x \geq 0 \\ scale \times x, & x < 0 \end{cases} \tag{2}$$

$$f(x) = min\,(ceiling, max\ (0, x)) = \begin{cases} x, & x < 0 \\ x, & 0 \leq x \leq ceiling \\ ceiling, & x \geq ceiling \end{cases} \tag{3}$$

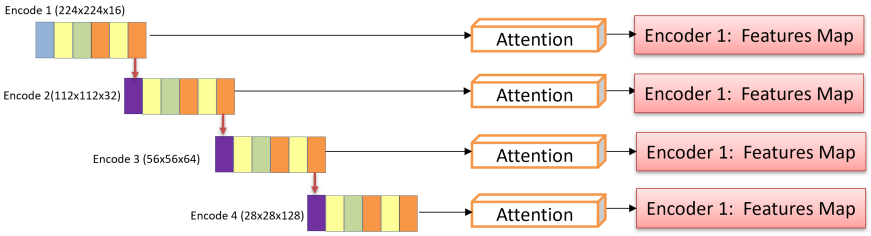


Figure 3: Extracting feature maps by encoding layers and attention.

3.2.2 Atrous Spatial Pyramid Pooling (ASPP) Module

The atrous spatial pyramid pooling (ASPP) module [14] improves multiscale feature representation by employing parallel atrous convolutions with various rates of dilation. This design enlarges the receptive field and does not increase the number of model parameters; thus, a fine lesion boundary and coarser long-range information can be captured by the network. As illustrated in Fig. 4, LANET and ASPP use dilation rates of 1, 6, 12, and 18 to combine local boundary details with the global semantic context sufficiently, owing to the large variation in lesion size, shape, and margin complexity.

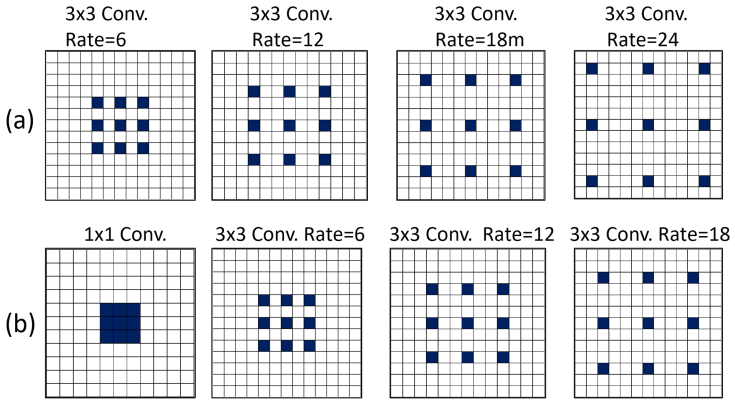


Figure 4: ASPP architecture: (**a**) original ASPP module; (**b**) ASPP module with dilated convolutions at different rates for land cover types.

Unlike typical U-Nets, which are based on repeated pooling and therefore tend to lose spatial information, the integration of ASPP enables LANET to retain detailed structures and acquire rich contextual cues. This provides the model with more immunity to irregular, diffuse or uncertain regions. ASPP integration helps the model capture multiscale contextual information without significantly increasing the computational cost of DeepLab-style models. The design seeks to improve the accuracy while maintaining suitable efficiency for real-time clinical use.

3.2.3 Block Attention

The attention mechanism between the encoder and decoder stages captured the lesion features more effectively. A LANET uses an attention mechanism inspired by a convolutional block attention module (CBAM) [28]. It combines channel and spatial attention and refines feature representations through two consecutive steps of attention map inference. This process helps the network to focus on more informative channels and spatial regions. Algorithm 1 contains the pseudocode for the pooled-unpooled attention block obtained by merging the original and additional information in its attention block. The exact process is applied to the other encoders within the LANET, as indicated by the arrow lines in Fig. 5. Channel attention detects key features in an image. By contrast, spatial attention identifies specific locations with valuable information.

Algorithm 1: Pseudocode for the pooled–unpooled attention block

Input: $X \in \mathbb{R}^{H\times W\times C}$
Output: $X_{\text{att}} \in \mathbb{R}^{H\times W\times C}$
Pooling and Unpooling Residual
$(P, idx) \leftarrow \text{MaxPool}_{2\times 2}(X)$
$U \leftarrow \text{MaxUnpool}_{2\times 2}(P, idx)$
$R \leftarrow X \leftarrow U$
Channel Attention
$\text{r} \leftarrow \text{GAP}(R) \in \mathbb{R}^{C}$
$w \leftarrow \sigma(MLP_{LeakyReLU}) \in \mathbb{R}^{C}$
$X_{\text{ch}} \leftarrow X.w$
Spatial Attention
$M_{\text{avg}} \leftarrow mean(X_{\text{att}}) \in \mathbb{R}^{H\times W}$
$M_{\max} \leftarrow max(X_{\text{att}}) \in \mathbb{R}^{H\times W}$
$M \leftarrow stack\left(M_{\text{avg}}, M_{\max}\right) \in \mathbb{R}^{H\times W\times 2}$
$S = \sigma\left(\text{Conv}_{7\times 7}(M)\right) \in \mathbb{R}^{H\times W}$
Return
$X_{\text{att}} \leftarrow X_{\text{ch}}.S$

The LANET introduces a pooled–unpooled attention mechanism that restores the information removed during pooling as shown in Fig. 5. The module first performs max pooling and unpooling to recover the spatial layout and then computes the pixel-wise difference between the original and unpooled features. This subtraction reveals subtle activations suppressed by pooling. It retains edge-level variations that are critical for medical segmentation. The spatial attention module enhances clinically important locations by emphasizing reconstructed border cues and reducing the influence of background artifacts. The proposed mechanism explicitly recovers lost information and incorporates leaky ReLU to improve the gradient flow for

low-contrast lesions. The following simplified example shows how **pool → unpool → subtraction** recovers suppressed features.

Input Feature Map (4 × 4): $X = \begin{bmatrix} 16 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{bmatrix}$

Step 1: Max-Pooling (2 × 2): Pooling outputs the maximum from each block. $P = \begin{bmatrix} 16 & 13 \\ 14 & 15 \end{bmatrix}$

Step 2: Max-Unpooling with Stored Indices: $U = \begin{bmatrix} 16 & 0 & 13 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 14 & 15 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Step 3: Residual (X − U) Recovers Suppressed Values: $R = \begin{bmatrix} 0 & 2 & 3 & 0 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 0 & 0 & 1 \end{bmatrix}$

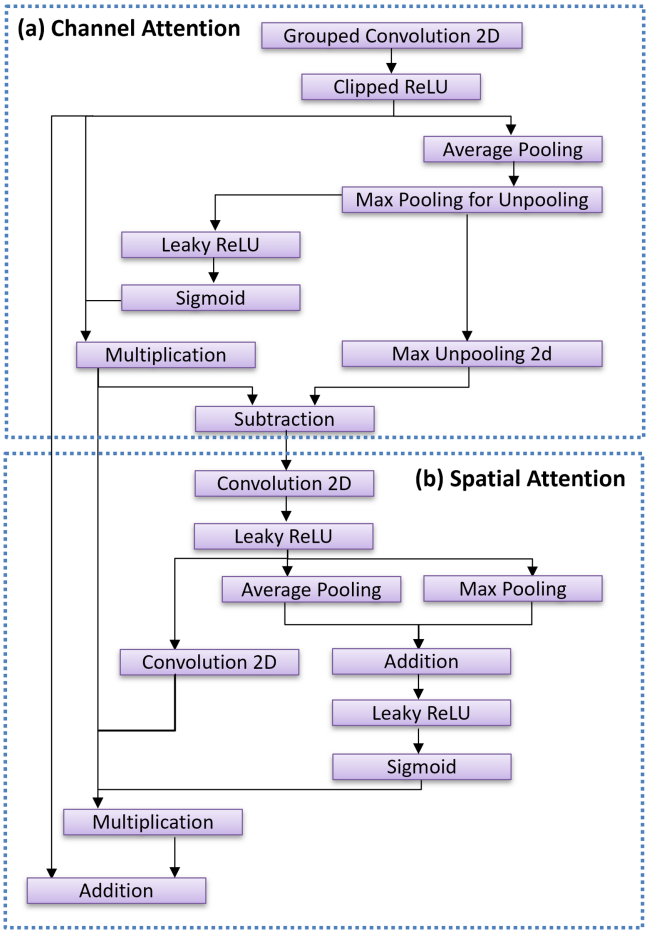


Figure 5: Block attention modules: (**a**) Channel Attention; (**b**) Spatial Attention.

3.2.4 Decoder

The decoder reconstructs the segmentation mask from the encoded feature maps using transposed convolutions for upsampling. Skip connections are introduced by the corresponding encoder layers to preserve spatial information. Each upsampled feature map was refined using convolutional and activation layers to ensure accurate localization and boundary recovery. The final prediction layer consisted of a one-by-one convolution to reduce the feature dimensions to two classes: lesion and background. The activation function applies softmax to generate pixel-wise class probabilities, thereby producing the final segmentation output.

At each encoder level, the output features are generated by customization layers that compute the difference between the maximum unpooling of the max pooling and the input. The first type is unpooling, which uses the pooling concept to start the decoding operations, as illustrated in Fig. 6. The learnable parameters of the LANET model, which are the outputs, were initialized from the first convolutional encoding layer. Three activation functions were used in this study: ReLU, leaky ReLU, and clipped ReLU, Eqs. (1)–(3), respectively: The last convolutional decoding layer was followed by a softmax activation layer, resulting in 846,786 parameters, as listed in Table 3, which provides details on the distribution of the parameters of the proposed architecture. For more details, refer to Appendix A.

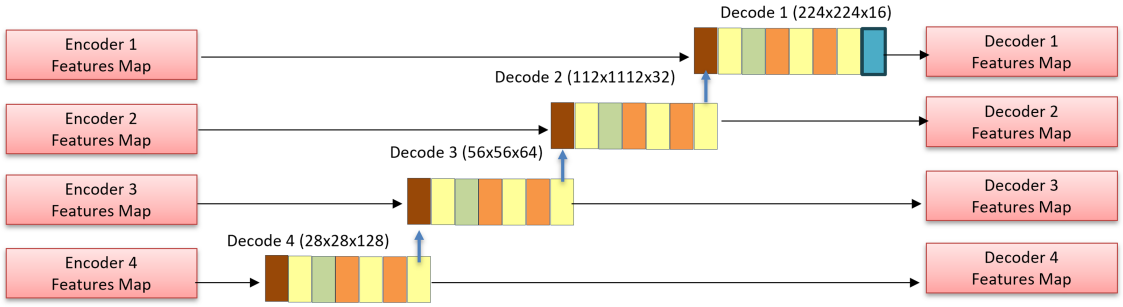


Figure 6: Extracting feature maps by Decoding layers.

Table 3: Distribution of the learning parameters of the LANET model.

Block	Layers	Learnable Sizes	Learnable
Encoder_Block 1	Conv, BN, GroupedConv	224 × 224 × 16	3232
Encoder_Block 2	Conv, BN, GroupedConv	112 × 112 × 32	15,328
Encoder_Block 3	Conv, BN, GroupedConv	56 × 56 × 64	60,352
Encoder_Block 4	Conv, BN, GroupedConv	28 × 28 × 128	239,488
Bottleneck	Conv, BN	14 × 14 × 16	57,536
Decoder_Block 1	Transposed_Conv, Conv, BN	28 × 28 × 128	329,472
Decoder_Block 2	Transposed_Conv, Conv, BN	56 × 56 × 64	107,392
Decoder_Block 3	Transposed_Conv, Conv, BN	112 × 112 × 32	27,072
Decoder_Block 4	Transposed_Conv, Conv, BN	224 × 224 × 16	6914
Total learnable parameters		846,786	

U-Net, U-Net++, and U-Net+3 were implemented to determine the number of parameters, and their results were compared with those of the LANET. It contains 846,786 parameters, resulting in a small memory footprint suitable for the deployment of hardware with limited resources. Table 3 compares these criteria

using a state-of-the-art (SOTA) model. The memory size is computed by [29] using Eq. (4). For example, the LANET is equal to 3.2 megabytes.

3.3 Cross-Validation (Improves Robustness)

To ensure the robustness and reliability of the proposed model, the LANET was evaluated on the HAM10000 dataset using 5-fold cross-validation. Data were randomly subdivided into five subsets. In each cycle, four subsets were trained in, and one was validated such that each image was validated exactly once. Hence, the final performance measures reported the average and standard deviation across all folds, providing a better estimate of how well the model can be generalized in practice.

To better illustrate the learning behavior of the model, we tracked the accuracy and loss of both the training and validation sets across all the iterations. As shown in Fig. 7, the curves demonstrate stable learning, with consistent trends between the two sets. The validation accuracy closely followed the training accuracy, and the loss decreased smoothly for both, indicating an effective convergence without overfitting. The final validation accuracy reached 96.12%, which was consistent with the quantitative results reported in the main experiments. This comparison confirms that LANET generalizes well to unseen data and maintains reliable performance throughout the training process.

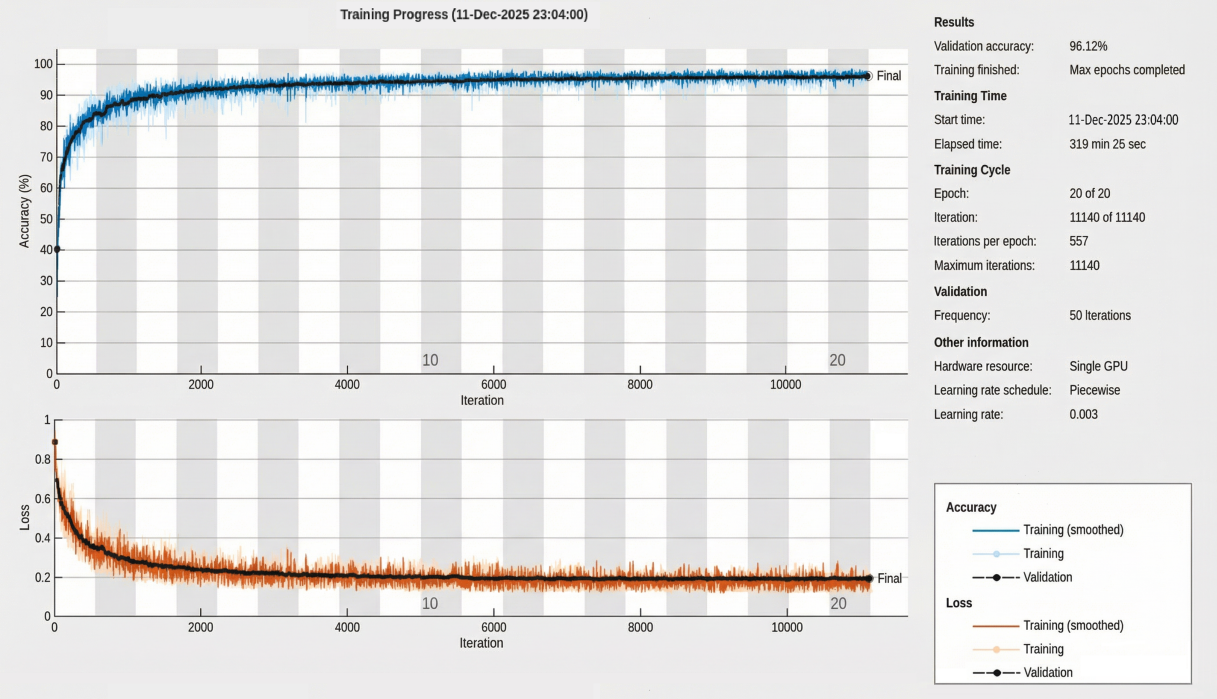


Figure 7: Progress training model.

Table 4 presents the comparison criteria for testing these models using the LANET. All GFLOPs are computed for an input of 224 × 224 using the convention that one multiply–accumulate (MAC) equals two floating-point operations. Compared to U-Net (7.42 GFLOPs) and U-Net++ (8.45 GFLOPs), the proposed LANET achieves higher segmentation accuracy (as mentioned in the results section) while requiring only 4.22 GFLOPs. This highlights superior computational efficiency of LANET.

$$Size\,of\,memory\ (in\,MegaBytes) = \frac{N_{parametres} \times 4\,bytes}{1024^2} \tag{4}$$

$Size\ of\ memory = \frac{846786 \times 4}{1024^2} = \frac{3387144}{1048576} = 3.2$ MB. Assuming 32-bit float precision (4 bytes per parameter), a LANET with 846,786 parameters requires approximately 3.2 MB of memory.

Table 4: Model complexity comparison.

Model	Number of Parameters (Million)	Number of Layers	Memory Size (MB)	GFLOPs
U-Net [5]	31	**70**	118.3	7.42
U-Net++ [6]	36.2	116	138.1	8.45
U-Net+3 [30]	26.9	92	102.6	6.80
LANET	**0.846**	136	**3.2**	**4.22**

Note: Bold values indicate the best result in each column.

3.4 Evaluation Metrics

Five primary metrics were used to evaluate the performance of the LANET. Evaluation metrics are beneficial for the overall analysis and offer helpful information regarding many characteristics of the segmentation results. Eqs. (5)–(9) show the subsequent metrics. Where: "*TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative*"

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Dice\,Coefficient\,(Dice) = \frac{2TP}{2TP + FP + FN} \tag{8}$$

$$Intersection\,over\,Union\,(IoU) = \frac{TP}{TP + FP + FN} \tag{9}$$

4 Results

The LANET performed well in terms of segmentation on all four datasets. The results are presented as mean ± standard deviation in Tables 5–8. The model demonstrated stable performance in accuracy, Dice, IoU, sensitivity, and specificity compared with the baseline and SOTA methods. All the results reported here are for segmentation performance, that is, the accuracy of segmentation, Dice, IoU, sensitivity, and specificity. Across the datasets, LANET consistently outperformed the baseline and contemporary architecture in most evaluation metrics. To evaluate whether the improvements of LANET over the SOTA models were statistically significant, we applied a paired t-test.

Table 5: Quantitative comparison of LANET and baseline segmentation models based on Accuracy, Dice, IoU, Sensitivity, and Specificity.

Citation	Accuracy	Dice	IoU	Sensitivity	Specificity
U-Net [5]	90.87 ± 0.46	77.24 ± 0.52	56.69 ± 0.49	60.87 ± 0.47	93.98 ± 0.41
U-Net++ [6]	95.65 ± 0.32	89.09 ± 0.37	81.31 ± 0.33	86.94 ± 0.31	**96.35 ± 0.29**
U-Net+3 [30]	95.41 ± 0.30	88.24 ± 0.35	77.67 ± 0.31	83.70 ± 0.29	95.83 ± 0.27
DeepLabv3+ [31]	95.84 ± 0.28	89.71 ± 0.33	74.21 ± 0.30	84.34 ± 0.27	96.87 ± 0.25
MRP-UNet [17]	94.64 ± 0.26	92.95 ± 0.29	90.18 ± 0.26	89.85 ± 0.25	90.18 ± 0.23
LANET	**96.44 ± 0.22**	**94.53 ± 0.25**	**92.61 ± 0.21**	**91.98 ± 0.20**	92.48 ± 0.19

Note: Bold values indicate the best result in each column.

Table 6: Evaluation of SOTA performance in comparison to LANET for the ISIC 2017 dataset.

Citation	Accuracy	Dice	IoU	Sensitivity	Specificity
U-Net [5]	96.46 ± 0.36	86.17 ± 0.42	64.79 ± 0.40	70.65 ± 0.39	**97.12 ± 0.31**
U-Net++ [6]	96.36 ± 0.34	87.72 ± 0.39	82.98 ± 0.33	88.65 ± 0.31	96.43 ± 0.29
U-Net+3 [30]	96.35 ± 0.33	87.27 ± 0.38	73.71 ± 0.32	79.04 ± 0.30	96.87 ± 0.27
DeepLabv3+ [31]	94.91 ± 0.29	85.71 ± 0.35	**83.64 ± 0.28**	**90.74 ± 0.27**	94.65 ± 0.25
MRP-UNet [17]	93.51 ± 0.26	**93.14 ± 0.31**	92.41 ± 0.25	90.42 ± 0.23	92.41 ± 0.22
EM-Net [21]	93.97 ± 0.24	86.42 ± 0.29	78.70 ± 0.24	85.33 ± 0.22	92.17 ± 0.21
IDSNet [32]	94.45 ± 0.23	86.53 ± 0.27	78.68 ± 0.23	84.41 ± 0.21	93.73 ± 0.20
LANET	**96.78 ± 0.22**	92.90 ± 0.24	81.90 ± 0.21	90.30 ± 0.20	92.83 ± 0.19

Note: Bold values indicate the best result in each column.

Table 7: Evaluation of SOTA performance vs. the LANET model on the ISIC 2018 dataset.

Citation	Accuracy	Dice	IoU	Sensitivity	Specificity
U-Net [5]	85.96 ± 0.48	75.11 ± 0.57	64.58 ± 0.54	75.55 ± 0.49	85.65 ± 0.45
U-Net++ [6]	89.05 ± 0.39	80.26 ± 0.44	68.67 ± 0.42	80.34 ± 0.37	88.36 ± 0.33
U-Net+3 [30]	89.71 ± 0.37	81.32 ± 0.40	71.01 ± 0.36	81.87 ± 0.35	89.09 ± 0.31
DeepLabv3+ [31]	90.41 ± 0.34	82.46 ± 0.41	67.63 ± 0.38	83.98 ± 0.33	90.54 ± 0.29
DSU-Net [33]	94.31 ± 0.26	90.04 ± 0.28	83.43 ± 0.27	92.22 ± 0.25	96.14 ± 0.23
WFC_AS_KL [34]	94.59 ± 0.25	90.27 ± 0.29	**84.06 ± 0.25**	86.08 ± 0.24	**98.95 ± 0.20**
EM-Net [21]	94.70 ± 0.23	90.30 ± 0.27	83.60 ± 0.23	**92.40 ± 0.22**	93.90 ± 0.21
MSHV-Net [35]	94.79 ± 0.21	89.16 ± 0.26	80.43 ± 0.22	87.91 ± 0.20	97.01 ± 0.18
LANET	**96.32 ± 0.19**	**92.21 ± 0.20**	83.37 ± 0.18	90.84 ± 0.19	98.22 ± 0.17

Note: Bold values indicate the best result in each column.

Table 8: Comparison of SOTA performance and the LANET model on the PH2 dataset.

Citation	Accuracy	Dice	IoU	Sensitivity	Specificity
U-Net [5]	92.00 ± 0.50	91.10 ± 0.40	82.00 ± 0.35	91.40 ± 0.42	90.20 ± 0.38
U-Net++ [6]	93.30 ± 0.46	91.90 ± 0.38	84.40 ± 0.33	92.30 ± 0.41	91.10 ± 0.36
U-Net+3 [30]	93.20 ± 0.44	92.10 ± 0.37	85.30 ± 0.32	93.30 ± 0.40	91.60 ± 0.35
DeepLabv3+ [31]	94.70 ± 0.39	93.60 ± 0.34	86.10 ± 0.29	93.20 ± 0.38	92.00 ± 0.34
WFC_AS_KL [35]	88.98 ± 0.47	83.22 ± 0.51	74.67 ± 0.46	87.73 ± 0.44	93.39 ± 0.38
LiteMamba-Bound [36]	94.79 ± 0.29	**95.62 ± 0.33**	**91.70 ± 0.25**	**96.20 ± 0.27**	92.67 ± 0.32
BDFormer [20]	95.14 ± 0.26	92.66 ± 0.35	86.32 ± 0.28	95.15 ± 0.31	**95.20 ± 0.25**
EM-Net [21]	96.34 ± 0.22	94.03 ± 0.30	88.92 ± 0.24	94.57 ± 0.26	94.06 ± 0.22
LANET	**97.89 ± 0.18**	93.70 ± 0.21	91.60 ± 0.19	95.80 ± 0.20	92.90 ± 0.21

Note: Bold values indicate the best result in each column.

4.1 HAM10000 Dataset

Table 5 evaluation of our method on the HAM10000 dataset LANET achieved accuracies of 96.44%, Dice = 94.53% and IoU = 92.61% over the HAM10000 dataset (shown in Table 5). All these results are higher than U-Net, U-Net++, DeepLabv3+, and MRP-UNet. The model shows a 17.3-point gain in Dice over U-Net and a 5.4-point gain over U-Net++. Compared with MRP-UNet, LANET improves Dice by 1.58 points, confirming its stronger boundary detection ability. A five-fold cross validation was conducted. LANET achieved an average Dice of 96.4% ± 0.6% and IoU of 89.1% ± 0.7%, showing stable performance. The results in Fig. 8 compare the five SOTA metrics. Fig. A1 in Appendix B shows the predictions in green color and the ground truth in red color.

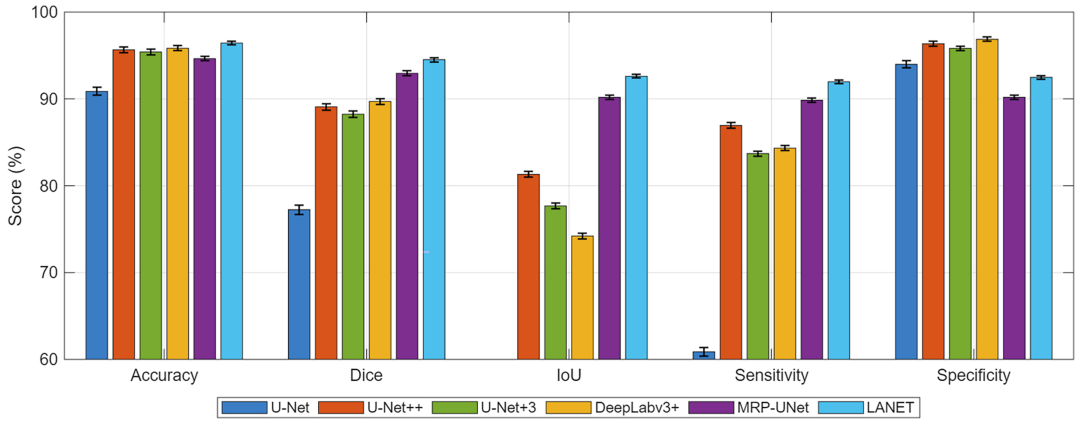


Figure 8: Comparison of accuracy, Dice, IoU, sensitivity, and specificity for LANET and baseline models on HAM10000.

The LANET returned a Dice score of 94.53% and accuracy of 96.44%, revealing competitive in-domain performance. The results transferred to ISIC 2017, ISIC 2018, and PH2 for LANET were the best with Dice scores from 90.61% to 91.09% and accuracies over the threshold, and all datasets were higher than 93%, showing a stable performance in segmentation over polling positions including different datasets.

4.2 ISIC 2017 Dataset

LANET obtained the highest accuracy (96.78%) and one of the best Dice scores (92.90%) among all models. Its Dice score exceeds those of U-Net, U-Net++, and U-Net+3 by 6.7, 5.2, and 5.6 points, respectively. The sensitivity (90.30%) and specificity (92.83%) showed that the model accurately detected lesion areas, while limiting false positives. The results were statistically significant ($p < 0.05$). The summary statistics of the performance of the models are shown in Fig. 9. Qualitative examples demonstrate the superior performance of LANET in tracing lesion borders compared to other competitors. Fig. A2 in Appendix B shows some segmented image samples. Qualitative examples support the quantitative findings, showing accurate segmentation across various lesion shapes and sizes.

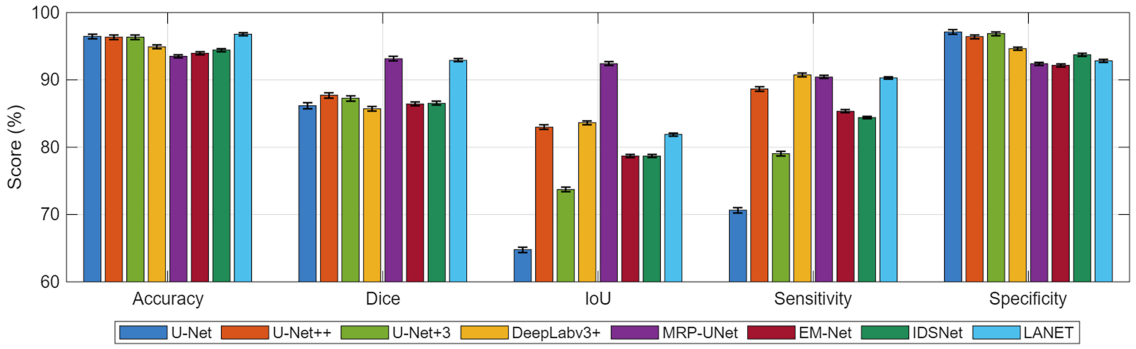


Figure 9: Metric comparison for LANET and SOTA models on ISIC 2017.

4.3 ISIC 2018 Dataset

Table 7 presents the comparative results of various DL models against the LANET. LANET achieved the highest overall accuracy (96.32%) and Dice score (92.21%) for the ISIC 2018. Its Dice score surpasses U-Net by 17.1 points and exceeds DSU-Net and EM-Net by roughly 2 points. The sensitivity (90.84%) and specificity (98.22%) demonstrated a balanced discrimination between the lesion and background regions. A statistical comparison of the performances is shown in Fig. 10. Across all SOTA models, LANET produced the second-highest specificity, while maintaining competitive sensitivity. Examples of the representative segmentations are shown in Fig. A3 in Appendix B. The qualitative outputs showed an improved boundary precision and reduced noise sensitivity.

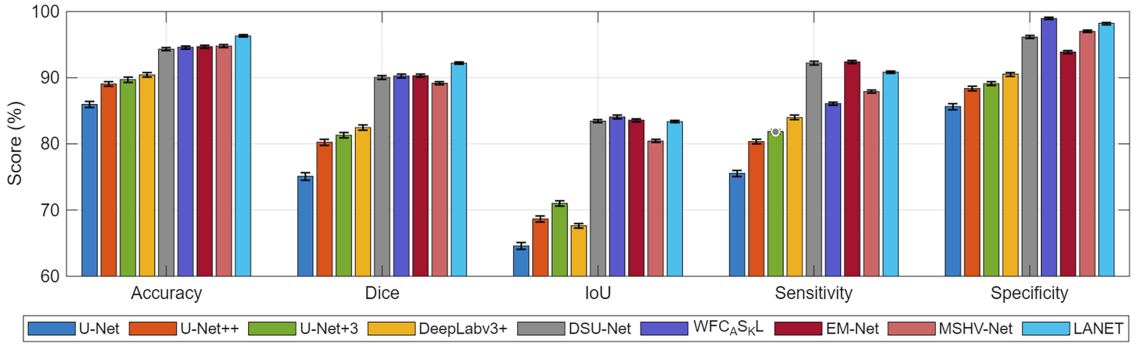


Figure 10: Statistical performance of LANET and competing models on the ISIC 2018 dataset.

4.4 PH2 Dataset

Table 8 lists the performance of the PH2 dataset. LANET achieved the highest accuracy (97.89%), followed by strong Dice (93.70%) and IoU (91.60%). It improved Dice over U-Net by 18.4 points, U-Net++ by 5.3 points, and DeepLabv3+ by 2.7 points. The LANET also achieved a high sensitivity (95.80%) and specificity (92.90%), indicating reliable lesion recognition with few false alarms. The model consistently outperformed the classical and transformer-based methods ($p < 0.05$). The qualitative samples showed accurate structural preservation even in challenging low-contrast lesions. The qualitative examples in Fig. A4 in Appendix B shows that the LANET is more accurate at boundary rendering and lesion localization than the other models.

4.5 Cross-Dataset Evaluation

A cross-dataset experiment was performed using LANET to evaluate its generalizability to other imaging datasets, as shown in Table 9. The model was trained on the dataset HAM10000 (refer to Table 5) and tested on three non-introduced datasets (ISIC 2017, ISIC 2018, and PH2) without any retraining or fine-tuning. These datasets are diverse with respect to illumination, lesion type, and color normalization. Cross-dataset testing shows that the LANET maintains a strong performance when applied to datasets with different imaging characteristics, indicating good generalization.

Table 9: Quantitative evaluation of LANET on cross-dataset testing.

Test Dataset	Accuracy	Dice	IoU	Sensitivity	Specificity
ISIC 2017	93.81	90.79	82.00	88.89	91.71
ISIC 2018	94.02	91.09	83.07	89.02	92.12
PH2	93.46	90.61	82.11	88.50	91.51

4.6 Model Visualization for Interpretability

The LANET model uses interpretability tools (i.e., convolutional activation maps and Grad-CAM) to illustrate the image regions that drive its predictions. These visualizations are concentrated in the lesion regions rather than irrelevant artifacts to mitigate clinical trust hesitation and provide transparency. Because DL models are considered "black boxes," these interpretability techniques shed light on what LANET has thought of. Through a performance assessment using the Dice coefficient, we demonstrated that the LANET ensures a good overlap between the predictions and ground-truth masks for multiple datasets. From Table 10, the high Dice scores show reliable and accurate segmentation and generalization, even when considering the imbalance between classes and slight boundary changes.

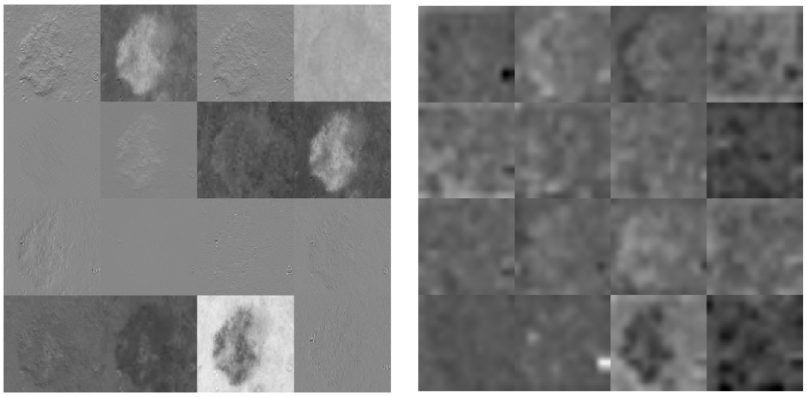
Table 10: Dice model among four public datasets.

Model	HAM10000	ISIC 2017	ISIC 2018	PH2
U-Net [5]	77.24	86.17	75.11	91.10
U-Net++ [6]	89.09	87.72	80.26	91.90
U-Net+3 [30]	88.24	87.27	81.32	92.10
DeepLabv3+ [31]	89.71	85.71	82.46	93.60
LANET	**94.53**	**92.90**	**92.21**	**93.70**

Note: Bold values indicate the best result in each column.

4.6.1 Interpretability Based on Convolutional Layers

Interpretability can be described as a clear technical explanation of a model. Visualization of these activations helps us decipher how the model transforms the input lesion images and discover the handcrafted features that establish class decisions. Interpretability techniques highlight the image regions that influence the model's predictions and help verify that the LANET focuses on clinically relevant structures during segmentation. This explicability is particularly important in medical imaging, where (even if correct) predictive but inscrutable models do not play a useful clinical role. Fig. 11 shows the feature activation maps generated by the LANET for various layers. These maps demonstrate how the network gradually refines the lesion-relevant information. The shallow layers are sensitive to the overall texture variations, whereas the latter capture fine edges and structures. The focus of the LANET during segmentation can be seen in the highlighted regions, demonstrating that the lesion contours and high-contrast regions were effectively addressed by our model.



(a) **(b)**

Figure 11: Activation maps from LANET's convolutional layers; (**a**) early layers emphasize low-level cues such as textures and simple edges; and (**b**) deeper layers capture higher-level semantic patterns, including lesion shape and boundary complexity.

4.6.2 Interpretability with Grad-CAM

To interpret how the LANET localizes lesion regions, the output scores and underlying convolutional features were linked using Grad-CAM. The strong-gradient locations correspond to the image regions that significantly affect the model decision. To explore the decision making of LANET, Grad-CAM was performed on several decoder layers. As shown in Fig. 12, shallow layers focus on fine-grained details of the lesion by emphasizing the edges and irregular margins, whereas deeper layers highlight the overall structure of the lesions. This evolution shows that the LANET uses both local border cues and the global semantic context during segmentation. The class-specific activation map obtained is given by Eq. (10):

$$\mathrm{CAM}^c = \mathrm{ReLU}\left(\sum_k a_k^c A^k\right) \tag{10}$$

A^k is the k^{th} feature map of the final convolutional layer and a_k^c is the importance weight for feature map k for class c.

The weights a_k^c represent how strongly feature map k influences the output score for class c. They are computed as:

$$a_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{11}$$

y^c is the score for class c. A_{ij}^k is the activation at the spatial location (i, j) in feature map k.Z is the total number of spatial locations in A^k.

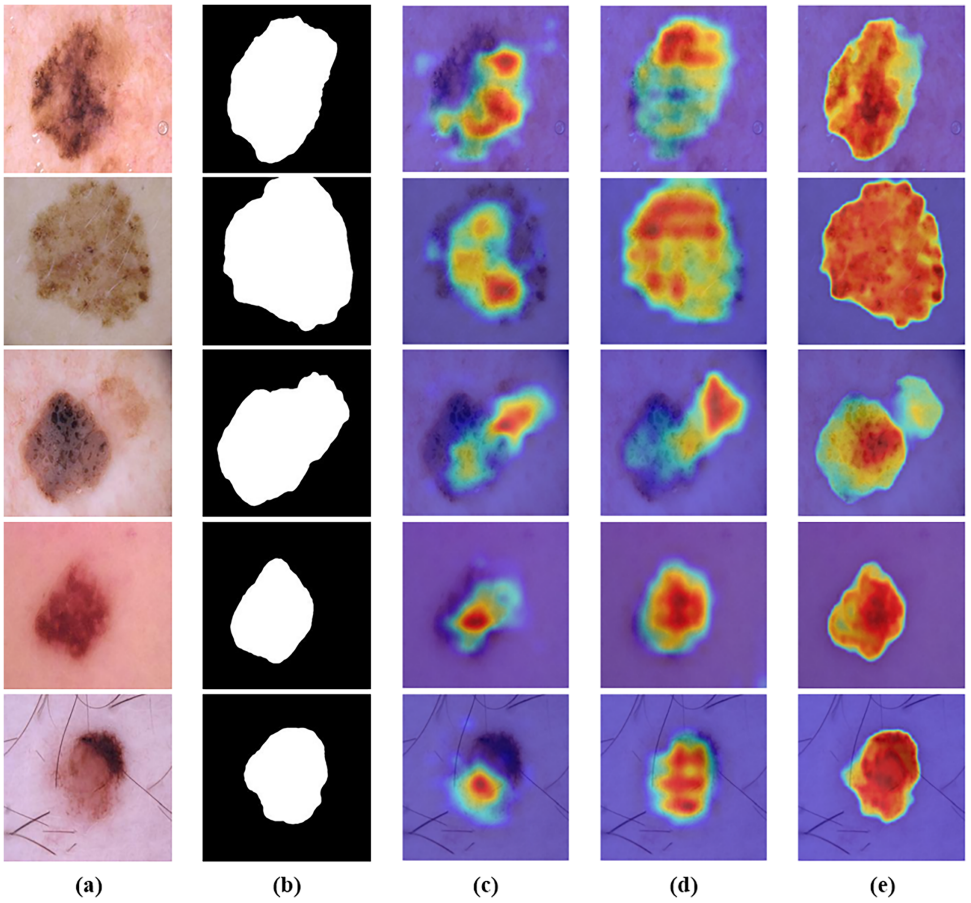


Figure 12: LANET Grad-CAM overlays at ISIC 2018 (**a**) Input image; (**b**) Predicted mask; (**c**); (**d**); and (**e**) Grad-CAM overlay from shallow, intermediate, and deep layers, respectively.

A more comprehensive Grad-CAM analysis over the entire ISIC 2018 dataset found that in 88% of the cases, the highlighted regions coincided with relevant areas (e.g., borders of lesions and pigment structures), as shown in Fig. 12. Reduced overlap was observed primarily for low-contrast or visual occlusions. These results indicate the ability of LANET to reliably zoom in diagnostically meaningful regions and corroborate the interpretability and clinical reliability of our model.

5 Ablation Experiment

Ablation studies have demonstrated the contribution of each module to a given performance level. Specifically, the Dice and IoU scores decreased significantly without the ASPP module, confirming that multiscale context aggregation is required to distinguish the lesion topological boundary across varying sizes. Sensitivity decreased without an attention mechanism, indicating that a model with one has an improved focus on subtle or low-contrast lesions. This step not only reduces the computational cost but also preserves the accuracy, thereby confirming the relevance of depthwise separable convolutions for lightweight implementations. Together, these findings justify the inclusion of each architectural element and confirm the design choices made for the LANET.

This study analyzed the roles of each constituent of the LANET architecture using the ISIC 2018 dataset. To evaluate the contribution of individual components to the LANET, an extensive ablation study was performed. The configurations tested are listed in Table 11, which summarizes the results of this study, where all variants were trained and evaluated using the same experimental setup for a fair comparison. Furthermore, the baseline U-Net model reached an Intersection over Union (IoU) of 75.11%. The incorporation of depthwise separable convolutions into U-Net increased IoU to 78.20%. The attention method further improved performance by increasing the IoU to 81.10%. The combination of the proposed improvements, including the ASPP, bespoke attention mechanisms, and depthwise convolutions, resulted in the best performance metrics, achieving an IoU of 81.90%. The total LANET architecture increased the IoU by approximately 6.79% compared with the basic U-Net. Fig. 13 shows the effect of each part on the LANET.

Table 11: Ablation study results for LANET using the ISIC 2018 dataset.

Citation	Accuracy	Dice	IoU	Sensitivity
U-Net (Baseline)	85.96 ± 0.42	75.11 ± 0.38	64.58 ± 0.55	75.55 ± 0.47
U-Net + Depthwise Conv.	91.20 ± 0.33	83.15 ± 0.29	78.20 ± 0.26	84.30 ± 0.35
U-Net + Depthwise Conv. + Attention	94.80 ± 0.27	87.50 ± 0.31	81.10 ± 0.24	87.20 ± 0.28
U-Net + Depthwise Conv.+ Attention + ASPP (LANET)	**96.78 ± 0.25**	**92.90 ± 0.22**	**81.90 ± 0.20**	**90.30 ± 0.23**

Note: Bold values indicate the best result in each column.

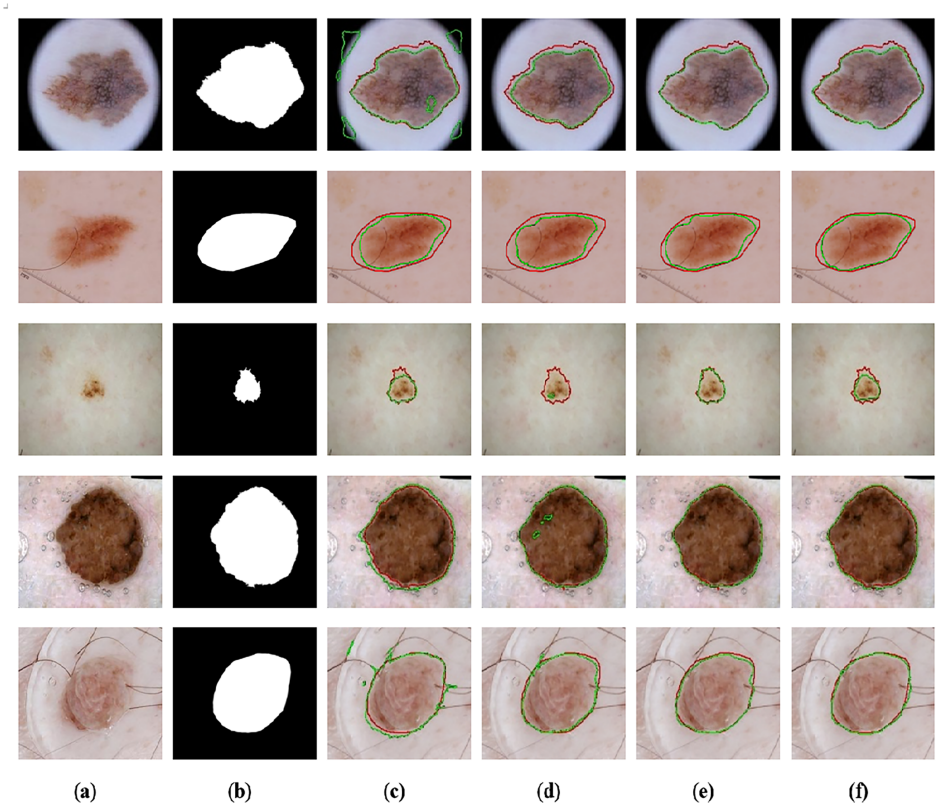


Figure 13: Effect of each part by the LANET model, where prediction in green color and the ground truth in red color: (**a**) Original images; (**b**) Ground truth; (**c**) Baseline U-Net segmentation; (**d**) U-Net with depthwise convolutions; (**e**) U-Net with depthwise convolutions and attention mechanism; and (**f**) LANET segmentation.

6 Discussion

LANET offers compact architecture that maintains a strong segmentation of accuracy while requiring minimal computational resources. With a combination of depthwise separable convolutions, custom attention mechanism and ASPP module in a compact encoder–decoder architecture, LANET achieves high seg-mentation performance with only 0.85 million parameters and 3.2 MB memory footprint. The results across the HAM10000, ISIC 2017, ISIC 2018 and PH2 datasets indicate that LANET achieves state-of-the-art performance in comparison with U-Net, U-Net++, and DeepLabv3+., indicating a strong boundary preservation ability and robust lesion localization.

LANET's low computational demand supports deployment in clinical settings, where hardware resources may be limited. Its low latency and small memory footprint make it well suited for real-time applications in clinical workstations as well as mobile or embedded systems. These properties are relevant because many large-capacity models require hardware that is not attainable in resource-constrained environments. It is encouraging that, with proper model design, lightweight models can achieve clinically relevant accuracy without compromising interpretability or usability.

Extensions, such as model compression, uncertainty estimation, and multimodal integration, may further improve clinical use. Model compression or knowledge distillation approaches can further reduce memory consumption. Taking uncertainty estimation into account might bring to the fore cases that are difficult or ambiguous for the model and may support clinician trust in towards human–AI collaboration. Further cross-validation and enlargement of the multi-dataset evaluation would also serve to test the stability across populations and imaging parameters. Next, by applying the LANET to other imaging modalities, such as MRI or CT, one can test the generality of its feature representations and attention mechanisms.

The Waterloo skin-tone dataset [37] provides complementary information. LANET consistently achieved high accuracy and sensitivity across samples, demonstrating robustness in lesion detection under diverse skin tones. Table 12 shows high Dice and IoU scores manifested significant overlaps in some cases although over- or under-segmentation was observed in others, particularly in lesions with low contrast or irregular boundaries.

Table 12: Quantitative results of evaluating a sample of the Waterloo dataset.

Images	Accuracy	Dice	IoU	Sensitivity	Specificity
1	0.98854	0.95268	0.90963	0.99776	0.98734
2	0.95105	0.91762	0.84777	0.96719	0.94472
3	0.92951	0.86534	0.76265	0.98732	0.91230
4	0.87074	0.21742	0.12197	1.00000	0.86837
5	0.99081	0.85643	0.74891	0.98708	0.99092
6	0.89346	0.75332	0.60426	0.60426	1.00000

These results correspond to the qualitative examination in which most examples had good structural agreement; however, some were subject to either false positives or boundary drift. These observations illustrate that improvements in texture modelling and contrast normalization may further improve robustness. Fig. 14 shows the qualitative results of the LANET segmentation on six example images from the Waterloo dataset. These results indicate that lightweight and interpretable architectures can achieve high performance and practical clinical value.

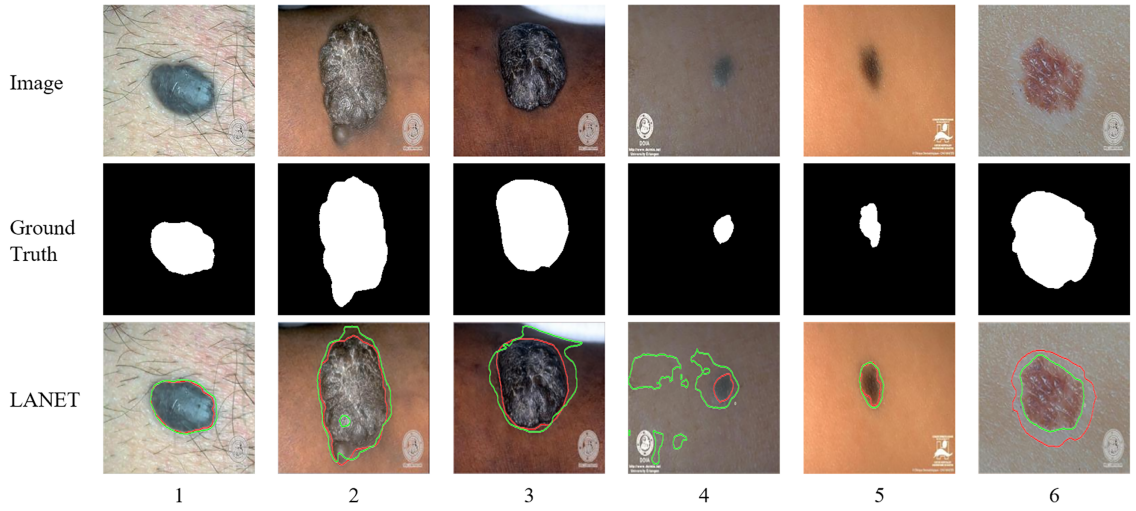


Figure 14: Qualitative results of testing six images from the Waterloo dataset, where the prediction in green color and the ground truth is in red.

7 Limitations and Future Work

The datasets used may not have covered the full range of clinical imaging variations. Broader evaluations across institutions and devices would improve our understanding of LANET's robustness. Furthermore, the datasets they were trained on may not truly represent the variety of clinical imaging devices, acquisition techniques, and patient populations, highlighting the importance of having diverse large-scale datasets to validate the robustness and generalization. More importantly, the model relies solely on image features and cannot leverage patient metadata or clinical knowledge when the diagnosis is ambiguous owing to a lack of clinical background. Future research should address these limitations in the future. Although LANET perform well on public datasets, their applicability to wider clinical populations remains uncertain.

Future studies will continue to investigate the robustness across acquisition technologies, lighting conditions, and lesion properties and include more diverse datasets collected from multiple sources under various conditions. This evaluation aimed to provide a comprehensive view of how well LANET can transfer learned representations to new, unseen clinical data distributions.

8 Conclusion

This study introduced the LANET, a lightweight attention-based network for skin lesion segmentation. LANET integrates depthwise convolutions, attention, and multiscale contexts to achieve accurate segmentation with a low computational cost. The compact design makes it suitable for real-time clinical use. A LANET consists of only 0.85 million parameters and costs approximately 3.2 MB of memory. It performs better than the widely used architectures on the four benchmark datasets. The results demonstrate that the LANET provides precise, robust, and reusable segmentation. Its small size and low latency make it appropriate for real-time implementation of mobile clinical hardware. These features limit the adoption of DL in dermatology by providing models that work well in both high-resource and low-resource contexts.

Future work will examine performance across wider datasets and clinical settings. This study provides a robust basis for the development of future diagnostic tools that can help clinicians in routine clinical practice. The use of LANET in other medical imaging applications can further demonstrate this versatility. Thus, LANETs offer a feasible and effective stepping stone for the resource-efficient, interpretable, and clinically deployable segmentation of skin lesions.

Acknowledgement: The authors gratefully acknowledge the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), for the institutional support provided for this research. The authors also sincerely thank the reviewers for their careful evaluation and constructive comments, which contributed to improving the quality of the manuscript.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Abdulrahman Dira Khalaf, Hazlina Hamdan; data collection: Abdulrahman Dira Khalaf; analysis and interpretation of results: Abdulrahman Dira Khalaf, Hazlina Hamdan, Alfian Abdul Halin, Noridayu Manshor; draft manuscript preparation: Abdulrahman Dira Khalaf. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets analyzed in this study are publicly available. HAM10000, ISIC 2017, and ISIC 2018 datasets are available from the International Skin Imaging Collaboration (ISIC) archive. The PH2 dataset is available from the Dermatology Service of Hospital Pedro Hispano. No new data was generated or analyzed in this study. All datasets are available from their respective public repositories without restriction.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A Detailed Layer-Wise Configuration and Learnable Parameter Distribution of the Proposed Lanet Architecture

Layer Type	Activation Size	Learnable Parameters	Count
2-D Convolution_1	224 × 224 × 16 × 1	Weights 3 × 3 × 3 × 16, Bias 1 × 1 × 16	448
Batch Normalization_1	224 × 224 × 16 × 1	Offset 1 × 1 × 16, Scale 1 × 1 × 16	32
Grouped Convolution_1	224 × 224 × 16 × 1	Weights 3 × 3 × 1 × 1 × 16, Bias 1 × 1 × 1 × 16	160
2-D Convolution_2	224 × 224 × 16 × 1	Weights 3 × 3 × 16 × 16, Bias 1 × 1 × 16	2320
2-D Convolution_3	224 × 224 × 16 × 1	Weights 1 × 1 × 16 × 16, Bias 1 × 1 × 16	272
2-D Convolution_4	112 × 112 × 32 × 1	Weights 3 × 3 × 16 × 32, Bias 1 × 1 × 32	4640
Batch Normalization_2	112 × 112 × 32 × 1	Offset 1 × 1 × 32, Scale 1 × 1 × 32	64
Grouped Convolution_2	112 × 112 × 32 × 1	Weights 3 × 3 × 1 × 1 × 32, Bias 1 × 1 × 1 × 32	320
2-D Convolution_5	112 × 112 × 32 × 1	Weights 3 × 3 × 32 × 32, Bias 1 × 1 × 32	9248
2-D Convolution_6	112 × 112 × 32 × 1	Weights 1 × 1 × 32 × 32, Bias 1 × 1 × 32	1056
2-D Convolution_7	56 × 56 × 64 × 1	Weights 3 × 3 × 32 × 64, Bias 1 × 1 × 64	18,496
Batch Normalization_3	56 × 56 × 64 × 1	Offset 1 × 1 × 64, Scale 1 × 1 × 64	128
Grouped Convolution_3	56 × 56 × 64 × 1	Weights 3 × 3 × 1 × 1 × 64, Bias 1 × 1 × 1 × 64	640
2-D Convolution_8	56 × 56 × 64 × 1	Weights 3 × 3 × 64 × 64, Bias 1 × 1 × 64	36,928
2-D Convolution_9	56 × 56 × 64 × 1	Weights 1 × 1 × 64 × 64, Bias 1 × 1 × 64	4160
2-D Convolution_10	28 × 28 × 128 × 1	Weights 3 × 3 × 64 × 128, Bias 1 × 1 × 128	73,856
Batch Normalization_4	28 × 28 × 128 × 1	Offset 1 × 1 × 128, Scale 1 × 1 × 128	256
Grouped Convolution_4	28 × 28 × 128 × 1	Weights 3 × 3 × 1 × 1 × 128, Bias 1 × 1 × 1 × 128	1280
2-D Convolution_11	28 × 28 × 128 × 1	Weights 3 × 3 × 128 × 128, Bias 1 × 1 × 128	147,584
2-D Convolution_12	28 × 28 × 128 × 1	Weights 1 × 1 × 128 × 128, Bias 1 × 1 × 128	16,512
2-D Convolution_13	14 × 14 × 16 × 1	Weights 1 × 1 × 128 × 16, Bias 1 × 1 × 16	2064
Batch Normalization_5	14 × 14 × 16 × 1	Offset 1 × 1 × 16, Scale 1 × 1 × 16	32
2-D Convolution_14	14 × 14 × 16 × 1	Weights 3 × 3 × 128 × 16, Bias 1 × 1 × 16	18,448
Batch Normalization_6	14 × 14 × 16 × 1	Offset 1 × 1 × 16, Scale 1 × 1 × 16	32
2-D Convolution_15	14 × 14 × 16 × 1	Weights 3 × 3 × 128 × 16, Bias 1 × 1 × 16	18,448

(Continued)

(continued)

Layer Type	Activation Size	Learnable Parameters	Count
Batch Normalization_7	14 × 14 × 16 × 1	Offset 1 × 1 × 16, Scale 1 × 1 × 16	32
2-D Convolution_16	14 × 14 × 16 × 1	Weights 3 × 3 × 128 × 16, Bias 1 × 1 × 16	18,448
Batch Normalization_8	14 × 14 × 16 × 1	Offset 1 × 1 × 16, Scale 1 × 1 × 16	32
Transposed Convolution_1	28 × 28 × 128 × 1	Weights 2 × 2 × 128 × 64, Bias 1 × 1 × 128	32,896
2-D Convolution_17	28 × 28 × 128 × 1	Weights 3 × 3 × 256 × 128, Bias 1 × 1 × 128	295,040
Batch Normalization_9	28 × 28 × 128 × 1	Offset 1 × 1 × 64, Scale 1 × 1 × 64	256
Grouped Convolution_5	28 × 28 × 128 × 1	Weights 3 × 3 × 1 × 1 × 128, Bias 1 × 1 × 1 × 128	1280
Transposed Convolution_2	56 × 56 × 64 × 1	Weights 2 × 2 × 64 × 128, Bias 1 × 1 × 64	32,832
2-D Convolution_18	56 × 56 × 64 × 1	Weights 3 × 3 × 128 × 64, Bias 1 × 1 × 64	73,792
Batch Normalization_10	56 × 56 × 64 × 1	Offset 1 × 1 × 64, Scale 1 × 1 × 64	128
Grouped Convolution_6	56 × 56 × 64 × 1	Weights 3 × 3 × 1 × 1 × 64, Bias 1 × 1 × 1 × 64	640
Transposed Convolution_3	112 × 112 × 32 × 1	Weights 2 × 2 × 32 × 64, Bias 1 × 1 × 32	8224
2-D Convolution_19	112 × 112 × 32 × 1	Weights 3 × 3 × 64 × 32, Bias 1 × 1 × 32	18,464
Batch Normalization_11	112 × 112 × 32 × 1	Offset 1 × 1 × 32, Scale 1 × 1 × 32	64
Grouped Convolution_7	112 × 112 × 32 × 1	Weights 3 × 3 × 1 × 1 × 32, Bias 1 × 1 × 1 × 32	320
Transposed Convolution_4	224 × 224 × 16 × 1	Weights 2 × 2 × 16 × 32, Bias 1 × 1 × 16	2064
2-D Convolution_20	224 × 224 × 16 × 1	Weights 3 × 3 × 32 × 16, Bias 1 × 1 × 16	4624
Batch Normalization_12	224 × 224 × 16 × 1	Offset 1 × 1 × 16, Scale 1 × 1 × 16	32
Grouped Convolution_8	224 × 224 × 16 × 1	Weights 3 × 3 × 1 × 1 × 16, Bias 1 × 1 × 1 × 16	160
2-D Convolution_21	224 × 224 × 2 × 1	Weights 1 × 1 × 16 × 2, Bias 1 × 1 × 2	34

Appendix B Qualitative Segmentation Results on Public Skin Lesion Datasets

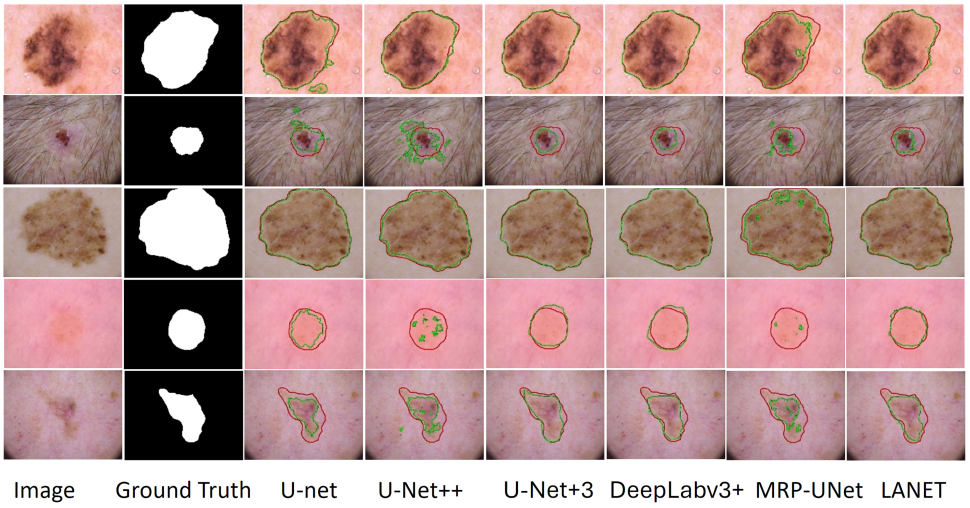


Figure A1: Examples of HAM10000 segmentations. Green indicates LANET predictions; red indicates ground truth.

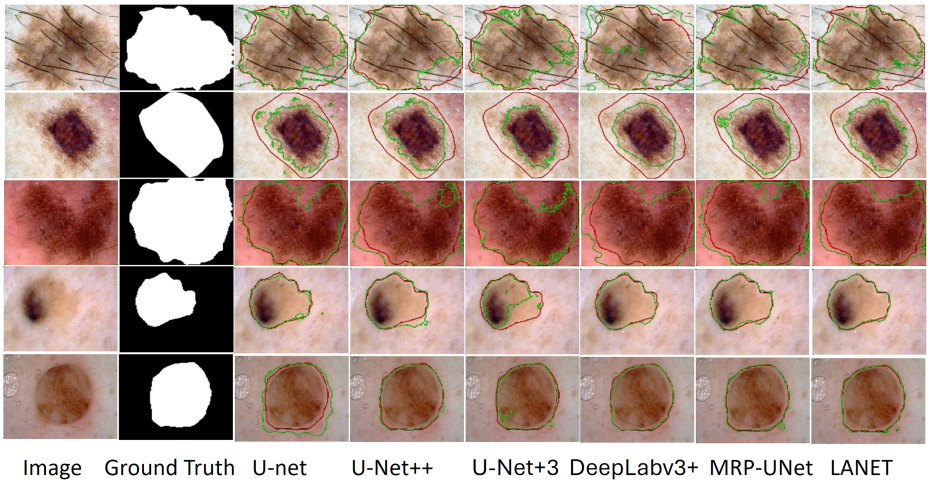


Figure A2: Examples of ISIC 2017 segmentations with predictions in green and ground truth in red.

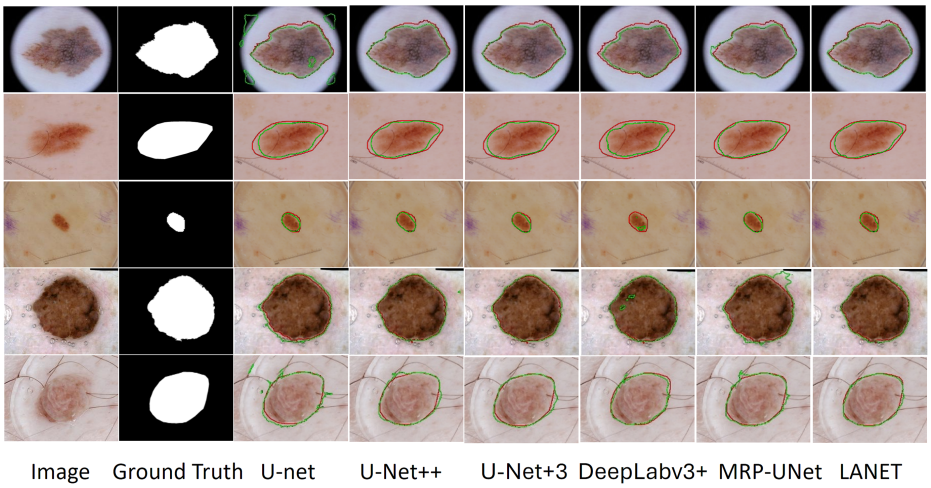


Figure A3: Segmentation examples from ISIC 2018 showing predictions (green) and ground truth (red).

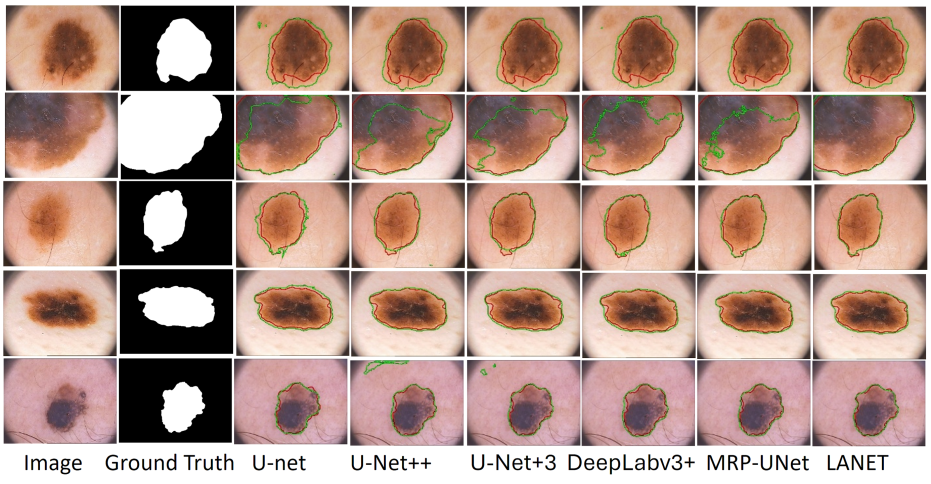


Figure A4: Segmentation results from the PH2 dataset with predictions shown in green and ground truth in red.

References

1. Mph RLS, Sung H, Mph TBK, Msph ANG. Cancer statistics 2025. A Cancer J Clin. 2025:10–45. doi:10.3322/caac.21871.
2. Hameed M, Zameer A, Raja MAZ. A comprehensive systematic review: advancements in skin cancer classification and segmentation using the ISIC dataset. Comput Model Eng Sci. 2024;140(3):2131–64. doi:10.32604/cmes.2024.050124.
3. Kassem MA, Hosny KM, Damaševičius R, Eltoukhy MM. Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. Diagnostics. 2021;11(8):1390. doi:10.3390/diagnostics11081390.
4. Yang G, Luo S, Greer P. Advancements in skin cancer classification: a review of machine learning techniques in clinical image analysis. Multimed Tools Appl. 2025;84(11):9837–64. doi:10.1007/s11042-024-19298-2.
5. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention—MICCAI 2015. Cham, Switzerland: Springer International Publishing; 2015. p. 234–41. doi:10.1007/978-3-319-24574-4_28.
6. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: a nested U-Net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham, Switzerland: Springer International Publishing; 2018. p. 3–11. doi:10.1007/978-3-030-00889-5_1.
7. Khan S, Khan A, Teng Y. DFF-UNet: a lightweight deep feature fusion U-Net model for skin lesion segmentation. IEEE Trans Instrum Meas. 2025;74:5030214. doi:10.1109/TIM.2025.3565715.
8. Arshad S, Amjad T, Hussain A, Qureshi I, Abbas Q. Dermo-seg: resNet-UNet architecture and hybrid loss function for detection of differential patterns to diagnose pigmented skin lesions. Diagnostics. 2023;13(18):2924. doi:10.3390/diagnostics13182924.
9. Yang G, Nie Z, Wang J, Yang H, Yu S. MSREA-Net: an efficient skin disease segmentation method based on multi-level resolution receptive field. Appl Sci. 2023;13(18):10315. doi:10.3390/app131810315.
10. Bozorgpour A, Sadegheih Y, Kazerouni A, Azad R, Merhof D. DermoSegDiff: a boundary-aware segmentation diffusion model for skin lesion delineation. In: Predictive intelligence in medicine. Cham, Switzerland: Springer Nature; 2023. p. 146–58. doi:10.1007/978-3-031-46005-0_13.
11. Ding Y, Yi Z, Xiao J, Hu M, Guo Y, Liao Z, et al. CTH-Net: a CNN and transformer hybrid network for skin lesion segmentation. iScience. 2024;27(4):109442. doi:10.1016/j.isci.2024.109442.
12. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. UNETR: transformers for 3D medical image segmentation. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2022 Jan 3–8; Waikoloa, HI, USA. p. 1748–58. doi:10.1109/WACV51458.2022.00181.
13. Khalaf AD, Hamdan H, Abdul Halin A, Manshor N. Segmentation and classification of skin cancer diseases based on deep learning: challenges and future directions. IEEE Access. 2025;13(1):90163–84. doi:10.1109/ACCESS.2025.3569170.
14. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018;40(4):834–48. doi:10.1109/TPAMI.2017.2699184.
15. Wu H, Chen S, Chen G, Wang W, Lei B, Wen Z. FAT-Net: feature adaptive transformers for automated skin lesion segmentation. Med Image Anal. 2022;76(9):102327. doi:10.1016/j.media.2021.102327.
16. Codella NCF, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018 Apr 4–7; Washington, DC, USA. p. 168–72. doi:10.1109/ISBI.2018.8363547.
17. Liu Z, Hu J, Gong X, Li F. Skin lesion segmentation with a multiscale input fusion U-Net incorporating Res2-SE and pyramid dilated convolution. Sci Rep. 2025;15(1):7975. doi:10.1038/s41598-025-92447-1.
18. Upadhyay AK, Bhandari AK. MaS-TransUNet: a multiattention swin transformer U-Net for medical image segmentation. IEEE Trans Radiat Plasma Med Sci. 2025;9(5):613–26. doi:10.1109/TRPMS.2024.3477528.

19. Ahmed A, Sun G, Bilal A, Li Y, Ebad SA. Precision and efficiency in skin cancer segmentation through a dual encoder deep learning model. Sci Rep. 2025;15(1):4815. doi:10.1038/s41598-025-88753-3.
20. Ji Z, Ye Y, Ma X. BDFormer: boundary-aware dual-decoder transformer for skin lesion segmentation. Artif Intell Med. 2025;162(6):103079. doi:10.1016/j.artmed.2025.103079.
21. Zhu K, Yang Y, Chen Y, Feng R, Chen D, Fan B, et al. EM-Net: effective and morphology-aware network for skin lesion segmentation. Expert Syst Appl. 2025;285(1):127668. doi:10.1016/j.eswa.2025.127668.
22. Pennisi A, Bloisi DD, Suriani V, Nardi D, Facchiano A, Giampetruzzi AR. Skin lesion area segmentation using attention squeeze U-Net for embedded devices. J Digit Imag. 2022;35(5):1217–30. doi:10.1007/s10278-022-00634-7.
23. Chen Y, Yang G, Dong X, Zeng J, Qin C. DSNET: a lightweight segmentation model for segmentation of skin cancer lesion regions. IEEE Access. 2025;13:31095–104. doi:10.1109/ACCESS.2025.3539521.
24. Naveed A, Naqvi SS, Khan TM, Iqbal S, Wani MY, Khan HA. AD-Net: attention-based dilated convolutional residual network with guided decoder for robust skin lesion segmentation. Neural Comput Appl. 2024;36(35):22277–99. doi:10.1007/s00521-024-10362-4.
25. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018;5(1):180161. doi:10.1038/sdata.2018.161.
26. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin lesion analysis toward mela-noma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). arXiv:1902.03368. 2019.
27. Mendonça T, Ferreira PM, Marques JS, Marcal ARS, Rozeira J. PH2—a dermoscopic image database for research and benchmarking. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013 Jul 3–7; Osaka, Japan. p. 5437–40. doi:10.1109/EMBC.2013.6610779.
28. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: Computer vision—ECCV 2018. Cham, Switzerland: Springer International Publishing; 2018. p. 3–19. doi:10.1007/978-3-030-01234-2_1.
29. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Vol. 1. Cambridge, MA, USA: MIT press Cambridge; 2016.
30. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, et al. UNet 3+: a full-scale connected UNet for medical image segmentation. In: ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2020 May 4–8; Barcelona, Spain. p. 1055–9. doi:10.1109/icassp40776.2020.9053405.
31. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Computer vision—ECCV 2018. Cham, Switzerland: Springer International Publishing; 2018. p. 833–51. doi:10.1007/978-3-030-01234-2_49.
32. Liu S, Wang P, Lin Y, Zhou B. IDSNet: unifying local and global context for skin lesion image segmentation. Biomed Signal Process Control. 2025;108(1):107961. doi:10.1016/j.bspc.2025.107961.
33. Zhong L, Li T, Cui M, Cui S, Wang H, Yu L. DSU-Net: dual-stage U-Net based on CNN and transformer for skin lesion segmentation. Biomed Signal Process Control. 2025;100(1):107090. doi:10.1016/j.bspc.2024.107090.
34. Kumari P, Agrawal RK, Priya A. Weighted fuzzy clustering approach with adaptive spatial information and Kullback-Leibler divergence for skin lesion segmentation. Int J Mach Learn Cybern. 2025;16(7):5317–37. doi:10.1007/s13042-025-02575-3.
35. Qu H, Gao Y, Jiang Q, Wang Y. MSHV-Net: a multi-scale hybrid vision network for skin image segmentation. Digit Signal Process. 2025;162(2):105166. doi:10.1016/j.dsp.2025.105166.
36. Ho QH, Nguyen TNQ, Tran TT, Pham VT. LiteMamba-Bound: a lightweight Mamba-based model with boundary-aware and normalized active contour loss for skin lesion segmentation. Methods. 2025;235(5):10–25. doi:10.1016/j.ymeth.2025.01.008.
37. Mahavar F. Skin cancer detection [Internet]. San Francisco, CA, USA: Kaggle; 2023 [cited 2026 Jan 1]. Available from: https://www.kaggle.com/datasets/fatemehmehrparvar/skin-cancer-detection.