



A Two-Stage Deep Learning Approach with DoubleU-Net for Enhanced Early Skin Cancer Detection in Smart Healthcare

Bandi Lakshmi Prasanna^{1,2*}, Ravi Boda¹, Raja Murali Prasad³

¹ Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad 500075, India

² Department of Electronics and Communication Engineering, Institute of Aeronautical Engineering, Hyderabad 500043, India

³ Department of Electronics and Communication Engineering, Marri Laxman Reddy Institute of Technology and Management, Hyderabad 500043, India

Corresponding Author Email: lakshmiprasanna447@gmail.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.120706>

ABSTRACT

Received: 8 April 2025
Revised: 20 June 2025
Accepted: 27 June 2025
Available online: 31 July 2025

Keywords:

DoubleU-Net, skin cancer detection, lesion segmentation, convolutional neural networks, International Skin Imaging Collaboration (ISIC) datasets, sensitivity, precision, dice coefficient

Precise lesion segmentation is necessary for the identification of skin cancer using dermoscopic pictures, but this is still difficult since lesion color, texture, and border abnormalities vary. In order to overcome this, we suggest a brand-new design named DoubleU-Net, which is very different from other multi-stage variations like U-Net++ and layered U-Nets. DoubleU-Net uses two different U-Net models in a sequential pipeline, in contrast to U-Net++, which combines dense skip connections within a single U-Net framework. In our method, the first U-Net creates an initial segmentation mask by acting as a coarse lesion extractor. The original input image is then concatenated with this mask and sent into a second U-Net, which concentrates on improving boundary regions and correcting early segmentation problems. Better feature learning and border delineation are made possible by this explicit two-stage refinement method, which is not included in the U-Net++ architectural philosophy, where refinement is carried out implicitly through layered connections. We assess DoubleU-Net using the International Skin Imaging Collaboration (ISIC) 2017, 2018, and 2019 benchmark public datasets. Intersection over Union (IoU), Dice Coefficient, Sensitivity, and Precision are used to evaluate performance. The proposed model performs better than the most advanced techniques on every metric. We obtain 98.89% in IoU, 91.04% in Dice Coefficient, 92.77% in Sensitivity, and 97.52% in Precision for ISIC 2017. The corresponding values for ISIC 2018 are 93.50%, 97.79%, 91.52%, and 97.78%. We achieve 92.86%, 98.55%, 92.64%, and 96.89% for ISIC 2019. The findings show that DoubleU-Net is useful for improving lesion boundary visual clarity and quantitative accuracy.

1. INTRODUCTION

Particularly in body parts that are often exposed to sunlight or artificial ultraviolet (UV) sources like tanning beds, prolonged exposure to UV radiation can cause inappropriate skin cell growth, which can result in skin cancer [1]. Melanoma, which starts in the melanocyte cells that produce melanin, is the most aggressive and deadly type of skin cancer [2] among the others [3-5]. When melanoma is discovered early, the 5-year survival rate is above 99%, but it drastically decreases if the disease spreads to other organs, according to the American disease Society. This emphasizes the vital role that early diagnosis and treatment play in enhancing patient outcomes [6].

However, the clinical diagnosis of skin cancer can be subjective and open to human error, which makes it expensive and time-consuming. In an effort to improve accuracy and decrease diagnostic variability, researchers have been investigating computer-aided diagnosis (CAD) systems that

use dermoscopic pictures for analysis more and more [7]. One of the core tasks in these systems is skin lesion segmentation, which involves identifying each pixel as either healthy or diseased [8]. Segmentation is the first and most important phase in the image analysis pipeline, and it has a big influence on how well later classification or diagnostic stages work [9].

Low contrast, erratic borders, and aberrations like hair, pen markings, or gel reflections are some of the reasons why traditional techniques like thresholding (bi-level and multi-level) frequently fail [10]. Dermoscopic pictures of skin lesions impacted by these problems are shown in Figure 1. Additionally, the segmentation method becomes more complicated and less precise due to changes in lesion color, size, and texture.

In order to get around these issues, current research has used pre-processing [11-13] and post-processing [14-16] methods in addition to color space conversion (e.g., Red Green Blue (RGB) to Hue Saturation and Value (HSV), Luminance and Chrominance (YUV), or Cyan, Magenta and Yellow (CMY))

to improve lesion visibility [17]. Most significantly, the field of medical picture segmentation has changed as a result of deep learning (DL). Multi-level features are automatically extracted by Convolutional Neural Networks (CNNs), and in pixel-wise segmentation tasks, designs such as Fully Convolutional Networks (FCNs) and U-Net-based models have shown higher performance [7, 8, 10, 12].

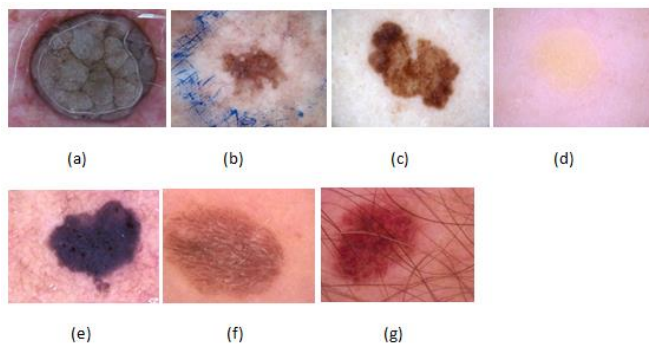


Figure 1. Image samples of skin lesions with (a) gel, (b) pen marks, (c) irregular boundaries, (d) unclear lesion, (e) dark color lesion area, (f) light color lesion area, (g) lesion with hair

In sophisticated medical image segmentation, U-Net continues to show limits despite its efficacy. Ineffective feature fusion, semantic gaps between encoder and decoder routes, and information loss from recurrent downsampling are some of the issues that U-Net may face, as noted in references [6, 8, 13, 16]. In difficult circumstances, these issues may lead to either under or over segmentation. Improvements to address these problems are suggested by studies like MAGRes-UNet [8] and IDUNet++ [16], although segmenting lesions with fuzzy borders or poor contrast is still a difficulty.

Furthermore, the generalizability of conventional deep models is further hampered by the fact that medical datasets are frequently small and need domain expertise for annotation. The creation of improved architectures specifically suited for medical segmentation tasks has been prompted by these constraints.

The U-Net architecture, which minimizes information loss by using skip links between the encoding and decoding channels, is one of the most well-known designs to handle these issues. Nonetheless, a persistent obstacle is the semantic disparity between encoder and decoder properties [15, 18]. We suggest the DoubleU-Net, which combines two U-Net modules in a cascaded manner, to address these drawbacks. In order to solve edge misclassifications and enable multi scale feature augmentation, this architecture enables the first U-Net to produce a coarse mask that is concatenated with the original input and sent to the second U-Net for refinement.

The following briefly describes the contributions of our work:

- **Creation of U-Net for Skin Lesion Segmentation:** A basic U-Net model was developed in order to effectively detect lesion boundaries.
- **DoubleU-Net Architecture Advancement:** To increase segmentation accuracy across lesion types, we suggest an improved architecture that combines twin U-Nets with staged learning and fine-tuned skip connections.
- **Comprehensive Dataset Evaluation:** We assess our approach using the International Skin Imaging Collaboration (ISIC) 2017, 2018, and 2019 datasets, which serve as

benchmarks for the segmentation of skin lesions.

- **Robust Evaluation Metrics:** To give a thorough performance evaluation, our analysis uses the Dice Coefficient, Intersection over Union (IoU), Sensitivity, and Precision.

- **Visualization Techniques:** To demonstrate the efficacy of the model, segmentation outputs are compared with ground truth.

- **Clinical Potential and Research Impact:** In addition to providing dermatologists with useful insights, the suggested model points to exciting paths for practical applications and AI-assisted diagnostics.

- **Future Extensions:** For data augmentation, real-time deployment, and usability enhancements for clinical settings, we recommend integration with generative models.

This paper's remaining sections are arranged as follows: In Section 2, relevant literature is reviewed; in Section 3, the suggested architecture is explained; in Section 4, the experimental setup, results, and comparisons are covered; and in Section 5, conclusions and future directions are presented.

2. RELATED WORK

Many methods have been developed for automatic skin lesion segmentation in the last ten years [19-21]. Performance across a range of segmentation difficulties has significantly improved with the change from conventional hand-crafted feature engineering to deep learning-based architectures, particularly CNNs [22, 23].

2.1 Segmentation using traditional techniques

Early segmentation attempts were based on low-level pixel features, such as histogram thresholding [24-26], unsupervised color clustering [27], and region-merging techniques [28, 29]. Although morphological procedures, edge detectors, and active contour methods [30] were also investigated [31], these methods were not able to generalize to different lesion sizes, shapes, and colors. As a result, they were less appropriate for practical clinical applications due to their limited performance, particularly in low contrast or noisy image conditions.

2.2 Segmentation using deep learning techniques

By utilizing CNNs' feature learning capabilities, deep learning has significantly increased segmentation accuracy by enabling the automatic extraction of complex patterns. To improve lesion border precision and lower segmentation errors, a number of enhanced CNN-based models have been suggested.

Many researchers have investigated cutting-edge methods to enhance the categorization and segmentation of skin lesions. Using Mask-CN and Coarse-SN, Xie et al. [29] combined segmentation and classification in a multitask framework, enabling lesion delineation and disease prediction at the same time. In order to improve resilience under a range of lighting circumstances, Kumar et al. [30] used shading-attenuated and grayscale picture representations to overcome illumination variance.

Hasan et al. [31] developed lightweight structures using depth-wise separable convolutions to simplify the model while preserving pixel-level accuracy, which made their method appropriate for real-time or resource-constrained applications.

A hybrid JAEO-LeNet model with TransUNet-based preprocessing was presented by Babu and Philip [32], with an emphasis on improved feature extraction for early skin cancer detection. To enhance the acquisition of fine lesion boundaries, Xie et al. [33] created spatial and channel attention modules with high-resolution input handling. In order to guarantee structural consistency and contextual comprehension, Khan et al. [34] used DenseNet for deep feature extraction after first segmenting using MASK-RCNN. Similar to this, Qamar et al. [35] improved boundary localization by combining multi-scale feature aggregation and edge prediction, which enabled the model to retain intricate textures and subtle lesion edges. When taken as a whole, these papers show the variety of methodological developments meant to increase automated skin lesion analysis's precision, resilience, and effectiveness.

Considering these developments, a large number of current models still mostly rely on single-stage encoder-decoder architectures, which have significant drawbacks. One of these is insufficient feature fusion across spatial scales, which results in less-than-ideal segmentation of lesions with varied sizes and textures. Furthermore, these models frequently have trouble generalizing to lesions with hazy or ambiguous borders, which leads to inadequate lesion margin delineation. Moreover, their capacity to rectify inaccurate predictions is restricted due to the lack of a specific refinement step following initial segmentation. The DoubleU-Net architecture, a two-stage cascaded segmentation framework that improves multi-scale feature extraction, lesion border localization, and offers an extra refinement step for higher segmentation accuracy, is proposed in this work to address these issues.

2.3 Segmentation using color-based techniques

To improve segmentation performance, a number of recent methods have combined color space transformations with attention mechanisms. Pour and Seker [36] used feature map concatenation and the International Commission on Illumination (IELAB) color space to enhance lesion visibility, particularly under different lighting circumstances. By combining Squeeze and Excitation Network (SENet) with multiscale cross attention (MSC) modules and Cross-Scale Feature Fusion (CSFF) blocks, Liu et al. [37] were able to increase contextual comprehension and spatial precision. In a similar vein, Tran and Pham [38] used fuzzy logic in conjunction with additive attention mechanisms, adding fuzzy energy-based shape distances to the loss function to improve shape conformance. Despite improving saliency detection and lesion localization, these attention-based methods are still mostly focused on single-stage segmentation frameworks. Because of this, they frequently struggle to handle complex, multi-scale lesion patterns or high noise conditions, which reduces their generalizability and resilience. The need for multi-stage architectures, such as the DoubleU-Net proposal, which specifically addresses these problems through structured refinement and deeper contextual learning, is further highlighted by this.

2.4 Gap and motivation for DoubleU-Net

It is clear from the reviewed literature that single-stage U-Net models are the backbone of current segmentation techniques, which frequently fail to capture fine-grained lesion borders, particularly when there is low contrast or unclear

edges. While multi-branch topologies and attention methods have been developed to improve performance, they often result in a more complex model without adequately refining coarse segmentation results. Remarkably, relatively few studies use a real two-stage segmentation pipeline, in which a second network explicitly refines the output of an initial segmentation network. A cascaded dual-stage design is introduced in the proposed DoubleU-Net architecture to overcome these drawbacks. A typical U-Net generates a coarse segmentation mask in the first step. After concatenating this mask with the original input image, it is fed into a second U-Net for more precise refining. By using this technique, the model may reduce the semantic gap between intermediate and final outputs, enhance border precision, and fix early segmentation errors. DoubleU-Net successfully fills the architectural gap in the existing literature by ignoring traditional single-pass models and concentrating on structured multi-stage refinement, providing an accurate method for precise lesion segmentation in intricate dermoscopic pictures.

3. PROPOSED METHOD

We developed an effective skin lesion segmentation algorithm to aid in the early detection of skin cancer. Because skin lesions vary widely in size, shape, color, and physical appearance, it can be complicated to differentiate between healthy skin and the affected regions. Therefore, appropriate skin lesion segmentation is essential. Sometimes these variations are too large for conventional segmentation methods to handle, which results in inaccurate or insufficient lesion borders. To overcome these challenges, we developed a detailed method utilizing the Double U-Net architecture, which has advanced feature extraction capabilities and accurate border identification. The technical architecture and execution of the suggested segmentation pipeline are addressed in this section.

3.1 Design flow

A simplified and effective method for segmenting skin lesions utilizing the Double U-Net architecture is shown in Figure 2. The pipeline employs deep learning for accurate and automatic lesion border detection while preserving the structural integrity of dermoscopic images.

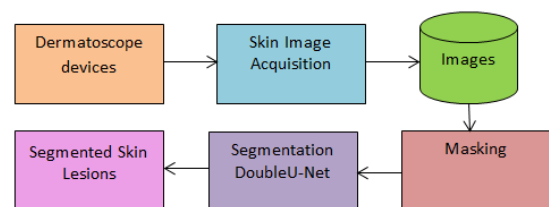


Figure 2. Design procedure for DoubleU-Net skin lesion segmentation

Dermatoscope devices, which are professional instruments used by dermatologists to obtain high-resolution pictures of skin lesions, are the initial step in the procedure. By offering vital details like lesion color, texture, and border features that are significant in the diagnosis of skin malignancies like melanoma, these devices improve the imaging of both surface and underlying skin structures. After being taken, the photos are gathered during the Skin Image Acquisition stage, when

they are digitally processed and systematically kept in the Images database. This makes up the segmentation system's input dataset. This pipeline retains full-color (RGB) pictures throughout, when compared to conventional image processing methods that transform pictures to grayscale. This choice is consistent with standard dermoscopic procedures, in which the ability to differentiate between various kinds of lesions depends on color information. Color asymmetry and pigmentation variations serve as essential diagnostic indicators for melanoma identification. Thus, the ability of the model to learn sensitive visual characteristics is enhanced by the retention of RGB information.

The pictures pass through a masking stage, where additional preprocessing techniques will be used to highlight regions of interest (e.g., erasing marks, hair, or non-lesion artifacts). This ensure that the model preserves crucial context while processing just the necessary lesion locations. The Double U-Net Segmentation module then obtains the preprocessed picture. Double U-Net utilizes a cascaded dual U-Net structure in which the second U-Net enhances the first's segmentation output. With the assistance of this architecture, the model can recognize lesion boundaries with high accuracy by learning complex spatial hierarchies and fixing any mistakes in the initial mask prediction.

The segmented skin lesions, which identify the lesion region with pixel-level precision, are the model's final result. Lesion size, symmetry, border irregularity, and other crucial clinical characteristics might be measured using these binary masks in additional diagnostic analysis.

3.2 DoubleU-Net architecture

The recommended DoubleU-Net design uses a cascaded dual U-Net architecture, in which two identical U-Net networks repeatedly process the input picture, to increase segmentation accuracy. In Figure 3, the complete architecture is shown. Four encoder blocks, each with two convolutional layers and a 3×3 kernel size, analyze the input picture in Network 1 first. Batch normalization and ReLU activation are then performed. Following each encoder block, the feature maps are down sampled and dominating spatial characteristics are preserved using max-pooling with a 2×2 window and stride of 2. Deeper semantic characteristics are captured by passing these encoder outputs via a bottleneck block with 1024 filters.

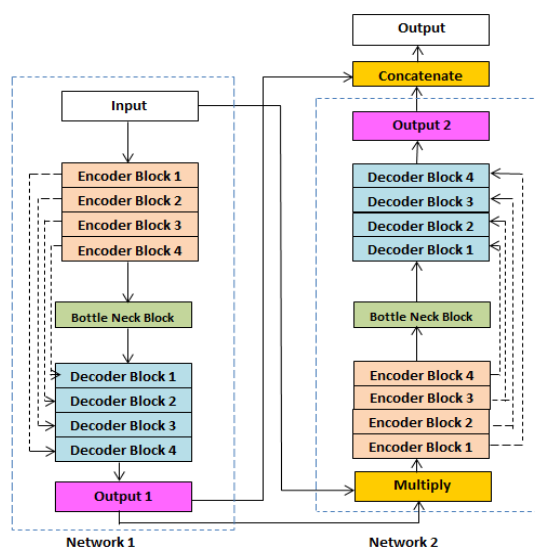


Figure 3. DoubleU-Net architecture

Similar to the encoder structure, Network 1's decoder path upsamples feature maps using transposed convolutions (2×2). In order to restore spatial information lost during downsampling, matching encoder features at each level are concatenated to their corresponding decoder blocks via skip connections. The decoder generates the first segmentation mask, Output1, by performing a sigmoid activation after a 1×1 convolution. In contrast to conventional U-Nets, DoubleU-Net incorporates Network 2 to add a refining method. The original input picture and the segmentation mask Output1 generated by Network 1 are multiplied element by element to provide the input to Network 2. The second U-Net can concentrate on specific characteristics since this multiplication improves areas that are probably related to lesions. With four encoder blocks, a bottleneck, and four decoder blocks, Network 2 is structurally identical to Network 1. Each decoder block in Network 2 only gets features from its associated encoder block in Network 2 since skip connections in Network 2 are internal. The impact of Network 1 is exclusively transferred through the improved input picture via multiplication; there are no common skip connections between the two U-Nets.

A sigmoid activation function is used after a 1×1 convolution to create Network 2's output, or Output2. The revised segmentation mask is then produced by concatenating Output1 and Output2 and running them through a final 1×1 convolution with sigmoid activation. A two-step refinement is made possible by this cascaded architecture: Network 1 produces a coarse segmentation, while Network 2 improves accuracy and specificity by sharpening the borders and correcting misclassified areas.

i) Encoder Explanation

The encoders in the first U-Net, Encoder1, and the second, Encoder2, have the same configurations. To add non-linearity and stabilize training, each encoder block consists of two 3×3 convolutional layers, followed by batch normalization and ReLU activation. A 2×2 max-pooling operation with a stride of 2 is used to produce downsampling, which reduces dimensionality while maintaining important spatial characteristics. For high-level abstract feature representation, a bottleneck layer uses 1024 filters, whereas the number of filters gradually rises via encoder levels, usually utilizing 64, 128, 256, and 512 filters. After processing the original input picture, Encoder1 creates a coarse segmentation mask (Output1) by sending encoded features to Decoder1. The original image and Output1 are multiplied element-wise to provide an upgraded input that Encoder2 may use to extract more fine-grained characteristics and focus more accurately on possible lesion locations.

ii) Decoder Explanation

Every decoder reflects this structure backwards, matching its corresponding encoder. In order to preserve spatial and contextual features, the decoder blocks first upsample using 2×2 transposed convolutions, then concatenate using encoder-derived skip connections. Skip connections from Encoder 1 are used by Decoder1, and Decoder2 incorporates Encoder2's skip connections. Batch normalization, ReLU activation, and two 3×3 convolutional layers are also included in each decoder block. Each decoder ends with a pixel-wise probability map (Output1 from the first stage and Output2 from the second) produced by a 1×1 convolution with sigmoid activation. To create the final refined segmentation mask, these two outputs are concatenated and then subjected to a further 1×1

convolution and sigmoid activation. This cascaded dual decoder method improves resilience against complicated and low-contrast skin lesion pictures, improves border accuracy, and corrects early prediction mistakes to enable comprehensive segmentation.

4. EXPERIMENTAL RESULTS

This section contains details on the training process, evaluation criteria, and description of each dataset we used to judge our methodology.

4.1 Training process

The proposed method used the Keras framework, with TensorFlow as the backend. In this research NVIDIA RTX 3090 GPU device is used, which offers the computing power required for deep learning applications. The proposed DoubleU-Net model consists of approximately 64 million trainable parameters, which is nearly double that of the traditional U-Net architecture, 31 million parameters. Despite the increase in complexity, the model achieves significantly better segmentation accuracy across multiple datasets. The model was trained with 100 epochs, batch size 16, learning rate 1e-4, and binary cross-entropy is used as a loss function to optimize memory utilization and improve the stability of gradient updates. Adam optimizer was chosen for training, as it adapts the learning rate for each parameter and has shown effective performance in a variety of deep learning tasks. Binary cross-entropy was the loss function that was employed. Since it reduces the divergence between the predicted and actual pixel values in the segmentation output, this option is frequently used for segmentation tasks using binary masks. The model was trained using training and validation datasets. When run on the designated hardware setup, the suggested DoubleU-Net model took about 20 minutes per epoch to train on the ISIC 2019 dataset. The overall training time was around 34 hours approximately, for a full training cycle of 100 epochs. This demonstrates the higher computational cost of the dual-stage design, which requires more training time than traditional single-stage models even if it offers better segmentation performance.

Figure 4 illustrates the training and validation accuracy graph of the DoubleU-Net for 100 epochs. Initially, during the first 20 epochs, the training accuracy shows a steady increase from a low starting point, while the validation accuracy exhibits some fluctuations. This indicates that the model is in its early learning phase and adapting to the data. As training progresses into the middle phase (epochs 20–50), both accuracies improve significantly, with the gap between the two curves narrowing. This demonstrates that the model is learning effectively and generalizing better to the validation data.

In the later phase (epochs 50–100), the training and validation accuracies converge and stabilize at high values, close to 1.0, showing that the model has achieved near perfect performance. The near overlap of the two curves indicates strong generalization with minimal overfitting. Early fluctuations in validation accuracy around epochs 10–20 suggest minor overfitting initially, but this resolves as the model continues training. The graph shows that the model performed extremely well overall, attaining high accuracy and indicating efficient training with balanced generalization.

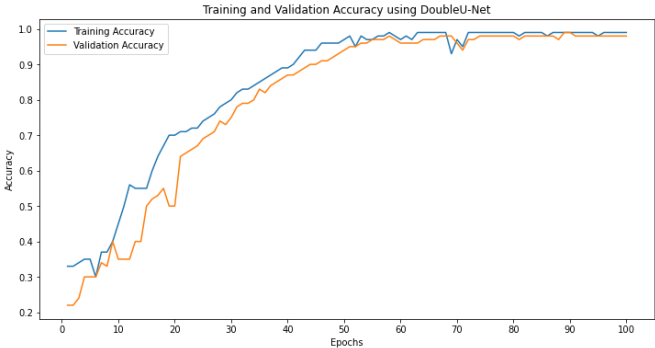


Figure 4. DoubleU-Net's training and validation accuracy graph

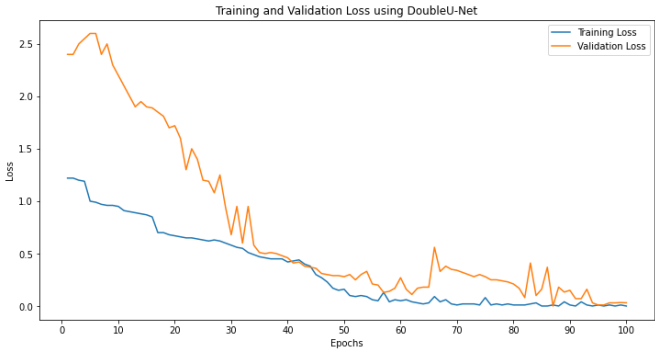


Figure 5. DoubleU-Net's training and validation loss graph

The training and validation loss for 100 epochs is shown in Figure 5. DoubleU-Net successfully learned from the training data is first demonstrated by the training loss fast drop. On the other hand, the validation loss begins higher and fluctuates throughout the early epochs (up to epoch 30), indicating that the model is adjusting to the validation data while controlling for random overfitting tendencies.

The model's performance significantly improves as training goes on, as indicated by the steady decrease in both training and validation losses. The validation loss stabilizes and follows the training loss's declining trend around the middle epochs (30–60), suggesting that the model is improving its performance in identifying new data.

Both curves converge close to zero in the latter stage (epochs 60–100), indicating little loss and highlighting the model's remarkable capacity for precise data segmentation. A well-trained model with good generalization and little overfitting is shown by the training and validation loss curves' near overlap over the latter epochs.

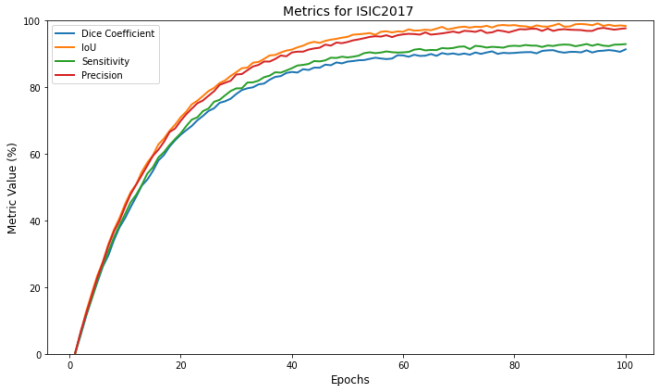


Figure 6. Graph representing key metrics on ISIC 2017

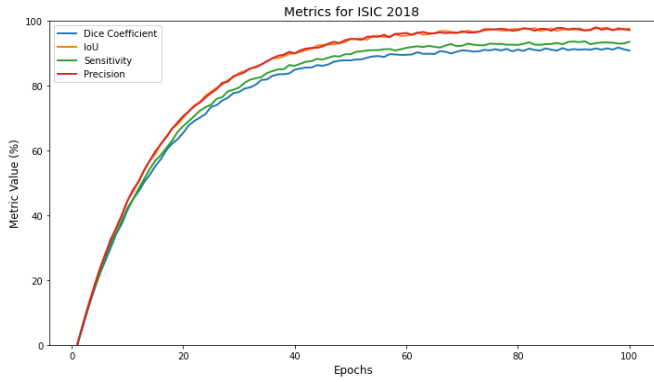


Figure 7. Graph representing key metrics on ISIC 2018

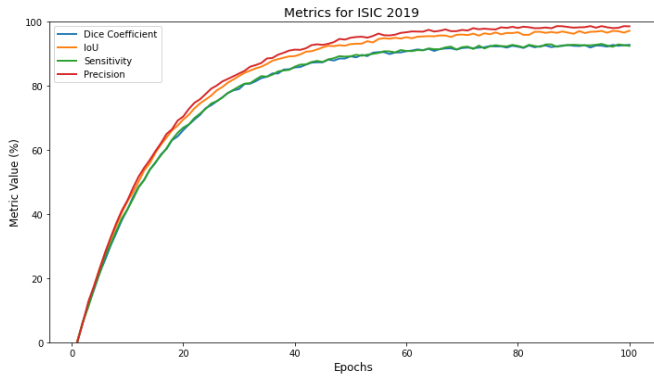


Figure 8. Graph representing key metrics on ISIC 2019

For the ISIC 2017, ISIC 2018, and ISIC 2019 datasets, the training curves for measures like Dice Coefficient, IoU, Sensitivity, and Precision are plotted and shown in Figures 6-8. Throughout the training epochs, they consistently improved. The model's capacity for learning was demonstrated by the initial sharp rise in all metrics, which was followed by stabilization as training went on. The model's ability to precisely detect lesion boundaries is demonstrated by precision, which commonly exceeded the other requirements. The model's robustness and suitability for skin lesion segmentation tasks are demonstrated by the consistent patterns observed in all datasets.

4.2 Evaluation metrics

The effectiveness of the suggested model is assessed using different performance metrics, each of which provides details on the model's capability to accurately identify skin lesions and distinguish them from healthy tissue. Table 1 indicates the metrics for segmentation.

4.3 Datasets

The DoubleU-Net was evaluated using ISIC 2019 [39], ISIC 2018 [40], and ISIC 2017 [41]. These datasets are part of the ISIC and are widely used in research on the diagnosis of skin cancer. They are publicly available datasets. Statistics related to datasets are indicated in Table 2. Additionally, we reduced the input images spatial dimension to 256X256 pixels using the resize method.

Table 1. Metrics for evaluating skin lesion segmentation

Performance Metrics	Metric Description	Formula
Dice Coefficient (DC)	Dice Coefficient measures the overlap between the predicted segmentation and the ground truth.	$DC = \frac{2 \cdot A \cap B }{ A + B }$
		A: ground truth (the actual skin lesion segmentation). B denotes the predicted segmentation. A ∩ B : number of pixels that are common between the ground truth and the predicted segmentation. A indicates the number of pixels in the ground truth. B indicates the number of pixels in the predicted segmentation.
Intersection over Union (IoU)	IoU quantifies the ratio of the overlap area to the total combined area of the predicted and ground truth masks.	$IoU = \frac{ A \cap B }{ A \cup B }$ A ∩ B is the region of overlap between the ground truth and predicted masks A ∪ B is the total area covered by both the ground truth and predicted masks.
Sensitivity	It measures the proportion of predicted TP that are correctly identified by model.	$Sensitivity = \frac{TP}{TP + FN}$ TP: It describes the number of pixels that are accurately predicted to be a part of the skin lesion (the model correctly detects the lesion). FN: It describes the quantity of pixels that are incorrectly recognized as background while, in fact, they are part of the skin lesion (the model fails to detect a lesion).
Precision	It measures the proportion of predicted positive cases TP and FP that are actually TP.	$Precision = \frac{TP}{TP + FP}$ TP: It describes the number of pixels that are accurately predicted to be a part of the skin lesion (the model correctly detects the lesion). FP: It describes the quantity of pixels that are incorrectly identified as belonging to a skin lesion when, in fact, they are part of the background (the model inaccurately labels healthy skin as a lesion).

Table 2. Dataset statistics before augmentation

Citation	Dataset	Total Number of Images	Training Dataset (70%)	Validation Dataset (10%)	Testing Dataset (20%)
[39]	ISIC 2019	25,331	17732	2533	5066
[40]	ISIC 2018	2594	1815	259	520
[41]	ISIC 2017	2000	1400	200	400

Table 3. Dataset statistics after augmentation

Dataset	Total Number of Images	Training Dataset (70%)	Validation Dataset (10%)	Testing Dataset (20%)
ISIC 2019	103000	72100	10300	20600
ISIC 2018	46935	32855	4693	9387
ISIC 2017	4116	2881	412	823

Oversampling-based data augmentation was used in all datasets to address the problem of class imbalance that arises in skin lesion datasets. To maintain class distribution and reduce bias, we used a Random Over sampler with 70% training, 10% validation, and 20% testing in a stratified split due to computational resource limitations. In order to match the majority class, counts for each dataset, other classes were enhanced using transformations such as rotation, flipping, zooming, and contrast changes. With 1,372 photos, the nevus class made up the majority of the three classes in the ISIC 2017 collection. After augmentation, the overall dataset size was 4,116 pictures, with the remaining two classes being enhanced to this amount. There are seven classifications in the ISIC 2018 dataset, with melanocytic nevi being the most prevalent with 6,705 photos. The dataset was expanded from 10,019 to 46,935 images by upsampling each of the other classes to 6,705 images.

Similarly, melanocytic nevi once again emerged as the majority class with 12,875 photos in the ISIC 2019 dataset, which has 8 classes. Following augmentation, all classes were equal to this number, yielding a 103,000-image enhanced dataset.

Table 3 shows the number of images after augmentation. Class-wise balancing not only makes the datasets uniform but it also improves the model's generalization and reduces overfitting. This is particularly important for tiny or skewed class distributions, which are typical in medical picture

datasets.

4.4 Results

The DoubleU-Net model's segmentation results are evaluated on three popular datasets: ISIC2017, ISIC2018, and ISIC2019. DoubleU-Net's dual-network architecture resulted in a more prolonged training period, but the smooth segmentation outputs and accuracy gains compensate the computational burden. Both quantitative and qualitative factors are taken consideration while analyzing the effectiveness of the recommended method. The suggested method improves the SOTA approaches, according on the quantitative findings of the proposed approach on ISIC 2017, which are shown in Table 4. A few visualization outputs of the suggested method on the ISIC 2017 are displayed in Figure 9. The DoubleU-Net is capable of producing high-quality segmentation results, as shown by its efficacy on the ISIC2017 dataset. The model correctly represents the borders of the lesions while retaining uniformity among lesions of various sizes and shapes. Even in the presence of objects or background noise, and the outcomes demonstrate reliable and flawless lesion area segmentation. The ability of the model to generalize and adapt to the unique challenges of the ISIC2017 dataset illustrates its reliability for skin lesion segmentation challenges in the real world.

Table 4. Comparison of suggested segmentation model with SOTA segmentation models

Citation	Method	Dataset	Sensitivity	Precision	DC	IoU
[2]	LSCS-Net	ISIC2017	91.95	95.14	91.17	85.88
[6]	L-UNet	PH2	91.15	97.22	91.15	90.98
[8]	MAGRes-UNet	HAM10000	91.06	95.47	94.68	95.89
[18]	UCM-Net	ISIC2017	-	-	88.45	79.29
		ISIC2018			89.35	80.85
[21]	GA-UNet	ISIC2018	90.41	91.99	90.18	82.13
[37]	M-VAN Unet,	ISIC2018	92.58	96.69	91.27	84.17
[42]	SICU-Net	ISIC2017	86.37	94.65	85.15	74.13
[42]	DICU-Net	ISIC2017	85.38	94.79	84.77	73.56
[42]	TICU-Net	ISIC2017	83.64	96.21	85.63	74.88
[35]	FAT-Net	ISIC2017	83.92	97.25	85.00	76.53
[43]	FRCU-Net	ISIC2017	91.50	98.61	-	97.27
[42]	SICU-Net	ISIC2018	89.39	96.95	90.94	83.39
[42]	DICU-Net	ISIC2018	92.02	94.61	90.44	82.55
[42]	TICU-Net	ISIC2018	91.57	95.85	90.96	83.42
[44]	CKD-Net	ISIC2018	90.55	97.01	87.79	80.41
[45]	FAT-Net	ISIC2018	91.00	96.99	89.03	82.02
[46]	AFLN-DGCL	ISIC2018	-	-	90.00	83.50
[47]	FRCU-Net	ISIC2018	90.40	97.90	-	96.30
Proposed Method	Double U-Net	ISIC2017	92.77	97.52	91.04	98.89
		ISIC2018	93.50	97.79	91.52	97.78
		ISIC2019	92.86	98.55	92.64	96.89

In order to compare the outcomes with SOTA methods, we also evaluated our approach on ISIC 2018. Table 4 shows that this approach improves performance significantly when compared to the other approaches. Figure 10 exhibits a qualitative assessment of the suggested approach efficacy. The segmentation outcomes further illustrate the model's ability to determine lesion boundaries for the ISIC 2018 dataset. In

challenging circumstances with overlapping textures and varying lesion intensities, the model works excellently. The extraordinary commitment of the segmented outputs to the ground truth masks highlights the ability of the models to generalize across multiple lesion patterns. The model's ability to recognize aberrant regions and understand complex patterns is shown by the precision with which lesions are distinguished

from the background.

The ISIC 2019 dataset was used in assessing our methods later, and Table 4 indicates the proposed approach outperformed the current methodologies by a wide margin. In addition, we also present some qualitative segmentation results of the suggested method on the ISIC2019 dataset in Figure 11, indicating that the model can generate continuous and accurate segmentation outputs that precisely match the offered ground truth masks. The smoothness around the lesion boundaries shows how well the model preserves the morphological features of the lesions. Because DoubleU-Net can correctly learn the features of the lesion, these results suggest that it is very efficient for clinical applications in skin cancer analysis.

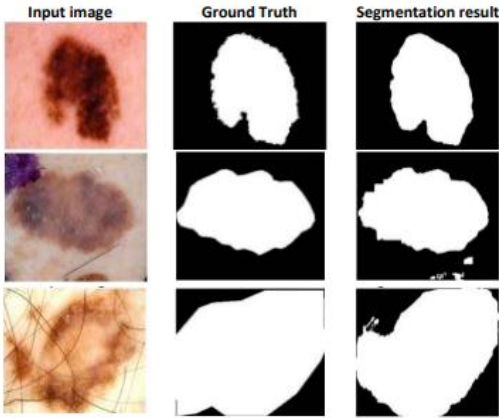


Figure 9. DoubleU-Net segmentation results on ISIC2017

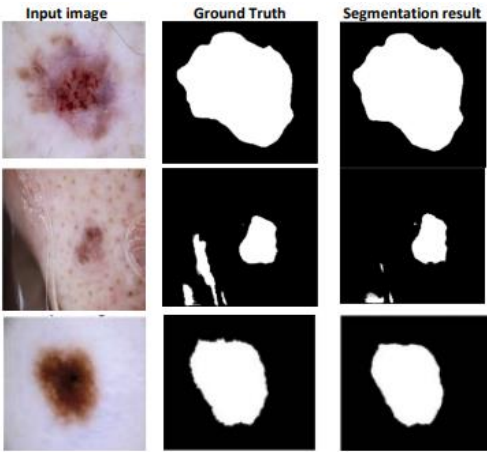


Figure 10. DoubleU-Net segmentation results on ISIC2018

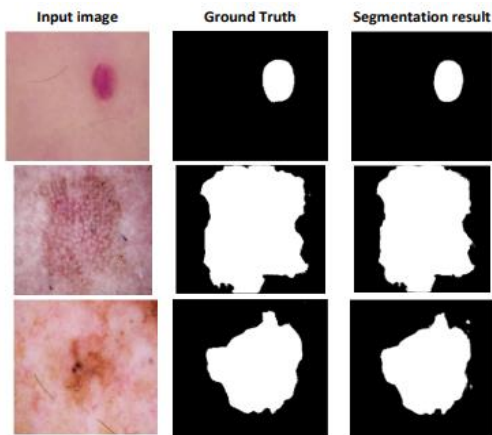


Figure 11. DoubleU-Net segmentation results on ISIC2019

5. CONCLUSION AND FUTURE DIRECTIONS

Using the ISIC 2017, ISIC 2018, and ISIC 2019 benchmark datasets, this study validated the efficacy of its DoubleU-Net architecture for skin lesion segmentation. The model outperformed the traditional U-Net by using a cascaded dual U-Net structure, especially in fine-grained segmentation and border localization. High performance was achieved by the DoubleU-Net in terms of Precision with 97.52% on ISIC 2017, 97.79% on ISIC 2018, and 98.55% on ISIC 2019, respectively, sensitivity 92.77% on ISIC 2017, 93.50% on ISIC 2018, and 92.86% on ISIC 2019, dice coefficient 91.04% on ISIC 2017, 91.52% ON 2018, and 92.64% on ISIC 2019, and intersection over union 98.89% with ISIC 2017, 97.78% with ISIC 2018, and 96.89% on ISIC 2019. These findings indicate the model's ability to aid in early identification of skin cancer.

It is essential to realize a few limitations though. The datasets used during the current research are carefully selected. Variations in illumination, resolution, and background noise can cause the model's performance to decrease in real world clinical scenarios, particularly when using low-quality or smartphone acquired photos. Moreover, DoubleU-Net's two-stage architecture greatly increases the memory and computational demands, restricting its use on edge devices often observed in smart healthcare applications.

Future research should concentrate on improving the architecture for lightweight implementation in order to overcome such challenges. To minimize the model's size and training time without reducing accuracy, methods including knowledge distillation, quantization, and model pruning might be considered. Additionally, cross modal fusion techniques, which include medical information such as age, lesion location, and patient history with image data, can improve diagnostic precision and enhance the system's clinical utility.

Model comprehension must be given the highest priority when applying AI systems in healthcare from an ethical point of view, making sure that predictions can be publicly communicated to doctors. Adherence to data protection laws like the GDPR is also essential, especially when managing private patient information. In conclusion, this suggested DoubleU-Net architecture has potential in developing automated dermatological evaluation and provides an excellent framework for accurate skin lesion segmentation. Its effective deployment in actual smart healthcare systems will depend on continuing developments in effectiveness, generalizability, and ethical integration.

REFERENCES

[1] Ichim, L., Popescu, D. (2020). Melanoma detection using an objective system based on multiple connected neural networks. *IEEE Access*, 8: 179189-179202. <https://doi.org/10.1109/ACCESS.2020.3028248>

[2] Din, S., Mourad, O., Serpedin, E. (2024). LSCS-Net: A lightweight skin cancer segmentation network with densely connected multi-rate atrous convolution. *Computers in Biology and Medicine*, 173: 108303. <https://doi.org/10.1016/j.compbimed.2024.108303>

[3] Adegun, A., Viriri, S. (2021). Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artificial Intelligence Review*, 54(2): 811-841. <https://doi.org/10.1007/s10462-020-09865-y>

[4] Ashraf, R., Afzal, S., Rehman, A.U., Gul, S., et al (2020).

- Region-of-interest based transfer learning assisted framework for skin cancer detection. *IEEE Access*, 8: 147858-147871. <https://doi.org/10.1109/ACCESS.2020.3014701>
- [5] Adegun, A.A., Viriri, S. (2019). Deep learning-based system for automatic melanoma detection. *IEEE Access*, 8: 7160-7172. <https://doi.org/10.1109/ACCESS.2019.2962812>
 - [6] Alafer, F., Siddiqi, M.H., Khan, M.S., Ahmad, I., Alhujaili, S., Alrowaili, Z., Alshabibi, A.S. (2024). A comprehensive exploration of L-UNet approach: Revolutionizing medical image segmentation. *IEEE Access*, 12: 140769-140791. <https://doi.org/10.1109/ACCESS.2024.3413038>
 - [7] Kumar, K.A., Vanmathi, C. (2024). Segmentation and detection of skin cancer using fuzzy cognitive map and deep Seg Net. *Soft Computing*, 28(5): 4575-4592. <https://doi.org/10.1007/s00500-024-09644-9>
 - [8] Hussain, T., Shouno, H. (2024). MAGRes-UNet: Improved medical image segmentation through a deep learning paradigm of multi-attention gated residual U-Net. *IEEE Access*, 12: 40290-40310. <https://doi.org/10.1109/ACCESS.2024.3374108>
 - [9] Khan, S., Khan, M.A., Noor, A., Fareed, K. (2024). SASAN: Ground truth for the effective segmentation and classification of skin cancer using biopsy images. *Diagnosis*, 11(3): 283-294. <https://doi.org/10.1515/dx-2024-0012>
 - [10] Talavera-Martinez, L., Bibiloni, P., Gonzalez-Hidalgo, M. (2020). Hair segmentation and removal in dermoscopic images using deep learning. *IEEE Access*, 9: 2694-2704. <https://doi.org/10.1109/ACCESS.2020.3047258>
 - [11] Chen, P., Huang, S., Yue, Q. (2022). Skin lesion segmentation using recurrent attentional convolutional networks. *IEEE Access*, 10: 94007-94018. <https://doi.org/10.1109/ACCESS.2022.3204280>
 - [12] Alahmadi, M.D. (2022). Multiscale attention U-Net for skin lesion segmentation. *IEEE Access*, 10: 59145-59154. <https://doi.org/10.1109/ACCESS.2022.3179390>
 - [13] Bindhu, A., Thanammal, K.K. (2023). Segmentation of skin cancer using Fuzzy U-network via deep learning. *Measurement: Sensors*, 26: 100677. <https://doi.org/10.1016/j.measen.2023.100677>
 - [14] Tang, P., Liang, Q., Yan, X., Xiang, S., Sun, W., Zhang, D., Coppola, G. (2019). Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging. *Computer Methods and Programs in Biomedicine*, 178: 289-301. <https://doi.org/10.1016/j.cmpb.2019.07.005>
 - [15] Zhou, Z., Quan, X., Niu, Y. (2023). IDUNet++: An improved convolutional neural network for melanoma skin lesion segmentation based on UNet++. In *Second International Conference on Electronic Information Engineering, Big Data, and Computer Technology (EIBDCT 2023)*, Xishuangbanna, China, pp. 568-573. <https://doi.org/10.1117/12.2674749>
 - [16] Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K. (2018). Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*. <https://doi.org/10.48550/arXiv.1802.06955>
 - [17] Hajabdollahi, M., Esfandiarpour, R., Khadivi, P., Soroushmehr, S.M.R., Karimi, N., Samavi, S. (2020). Simplification of neural networks for skin lesion image segmentation using color channel pruning. *Computerized Medical Imaging and Graphics*, 82: 101729. <https://doi.org/10.1016/j.compmedimag.2020.101729>
 - [18] Yuan, C., Zhao, D., Agaian, S.S. (2024). UCM-Net: A lightweight and efficient solution for skin lesion segmentation using MLP and CNN. *Biomedical Signal Processing and Control*, 96: 106573. <https://doi.org/10.1016/j.bspc.2024.106573>
 - [19] Sharen, H., Jawahar, M., Anbarasi, L.J., Ravi, V., Alghamdi, N.S., Suliman, W. (2024). FDUM-Net: An enhanced FPN and U-Net architecture for skin lesion segmentation. *Biomedical Signal Processing and Control*, 91: 106037. <https://doi.org/10.1016/j.bspc.2024.106037>
 - [20] Liu, X., Song, L., Liu, S., Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3): 1224. <https://doi.org/10.3390/su13031224>
 - [21] Khouy, M., Jabrane, Y., Ameer, M., Hajjam El Hassani, A. (2023). Medical image segmentation using automatic optimized U-Net architecture based on genetic algorithm. *Journal of Personalized Medicine*, 13(9): 1298. <https://doi.org/10.3390/jpm13091298>
 - [22] Ünver, H.M., Ayan, E. (2019). Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics*, 9(3): 72. <https://doi.org/10.3390/diagnostics9030072>
 - [23] Yuan, Y., Chao, M., Lo, Y.C. (2017). Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. *IEEE Transactions on Medical Imaging*, 36(9): 1876-1886. <https://doi.org/10.1109/TMI.2017.2695227>
 - [24] Abraham, N., Khan, N.M. (2019). A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, pp. 683-687. <https://doi.org/10.1109/ISBI.2019.8759329>
 - [25] Cao, Y., Liu, S., Peng, Y., Li, J. (2020). DenseUNet: densely connected U-Net for electron microscopy image segmentation. *IET Image Processing*, 14(12): 2682-2689. <https://doi.org/10.1049/iet-ipr.2019.1527>
 - [26] Ibtehaz, N., Rahman, M.S. (2020). MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121: 74-87. <https://doi.org/10.1016/j.neunet.2019.08.025>
 - [27] Ganesan, P., Sathish, B.S., Leo Joseph, L.M.I. (2020). HSL color space based skin lesion segmentation using fuzzy-based techniques. In *Advances in Electrical Control and Signal Systems: Select Proceedings of AECSS 2019*, pp. 903-911. https://doi.org/10.1007/978-981-15-5262-5_69
 - [28] Chi, W., Ma, L., Wu, J., Chen, M., Lu, W., Gu, X. (2020). Deep learning-based medical image segmentation with limited labels. *Physics in Medicine & Biology*, 65(23): 235001. <https://doi.org/10.1088/1361-6560/abc363>
 - [29] Xie, Y., Zhang, J., Xia, Y., Shen, C. (2020). A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Transactions on Medical Imaging*, 39(7): 2482-2493. <https://doi.org/10.1109/TMI.2020.2972964>
 - [30] Kumar, A., Hamarneh, G., Drew, M.S. (2020).

- Illumination-based transformations improve skin lesion segmentation in dermoscopic images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 3132-3141. <https://doi.org/10.1109/CVPRW50498.2020.00372>
- [31] Hasan, M.K., Dahal, L., Samarakoon, P.N., Tushar, F.I., Martí, R. (2020). DSNet: Automatic dermoscopic skin lesion segmentation. *Computers in Biology and Medicine*, 120: 103738. <https://doi.org/10.1016/j.compbiomed.2020.103738>
- [32] Babu, R.R., Philip, F.M. (2024). Optimized deep learning for skin lesion segmentation and skin cancer detection. *Biomedical Signal Processing and Control*, 95: 106292. <https://doi.org/10.1016/j.bspc.2024.106292>
- [33] Xie, F., Yang, J., Liu, J., Jiang, Z., Zheng, Y., Wang, Y. (2020). Skin lesion segmentation using high-resolution convolutional neural network. *Computer Methods and Programs in Biomedicine*, 186: 105241. <https://doi.org/10.1016/j.cmpb.2019.10524>
- [34] Khan, M.A., Akram, T., Zhang, Y.D., Sharif, M. (2021). Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework. *Pattern Recognition Letters*, 143: 58-66. <https://doi.org/10.1016/j.patrec.2020.12.015>
- [35] Qamar, S., Ahmad, P., Shen, L. (2021). Dense encoder-decoder-based architecture for skin lesion segmentation. *Cognitive Computation*, 13(2): 583-594. <https://doi.org/10.1007/s12559-020-09805-6>
- [36] Pour, M.P., Seker, H. (2020). Transform domain representation-driven convolutional neural networks for skin lesion segmentation. *Expert Systems with Applications*, 144: 113129. <https://doi.org/10.1016/j.eswa.2019.113129>
- [37] Liu, S., Zhuang, Z., Zheng, Y., Kolmaniè, S. (2023). A VAN-based multi-scale cross-attention mechanism for skin lesion segmentation network. *IEEE Access*, 11: 81953-81964. <https://doi.org/10.1109/ACCESS.2023.3298826>
- [38] Tran, T.T., Pham, V.T. (2022). Fully convolutional neural network with attention gate and fuzzy active contour model for skin lesion segmentation. *Multimedia Tools and Applications*, 81(10): 13979-13999. <https://doi.org/10.1007/s11042-022-12413-1>
- [39] ISIC Challenge Datasets 2019. <https://challenge.isic-archive.com/data/#2019>.
- [40] ISIC Challenge Datasets 2018. <https://challenge.isic-archive.com/data/#2018>.
- [41] ISIC Challenge Datasets 2017. <https://challenge.isic-archive.com/data/#2017>.
- [42] Ramadan, R., Aly, S. (2022). CU-Net: A new improved multi-input color U-Net model for skin lesion semantic segmentation. *IEEE Access*, 10: 15539-15564. <https://doi.org/10.1109/ACCESS.2022.3148402>
- [43] Xie, Y., Zhang, J., Lu, H., Shen, C., Xia, Y. (2020). SESV: Accurate medical image segmentation by predicting and correcting errors. *IEEE Transactions on Medical Imaging*, 40(1): 286-296. <https://doi.org/10.1109/TMI.2020.3025308>
- [44] Shorfuzzaman, M. (2022). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. *Multimedia Systems*, 28(4): 1309-1323. <https://doi.org/10.1007/s00530-021-00787-5>
- [45] Arora, R., Raman, B., Nayyar, K., Awasthi, R. (2021). Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomedical Signal Processing and Control*, 65: 102358. <https://doi.org/10.1016/j.bspc.2020.102358>
- [46] Wang, X., Jiang, X., Ding, H., Zhao, Y., Liu, J. (2021). Knowledge-aware deep framework for collaborative skin lesion segmentation and melanoma recognition. *Pattern Recognition*, 120: 108075. <https://doi.org/10.1016/j.patcog.2021.108075>
- [47] Sarker, M.M.K., Rashwan, H.A., Akram, F., Singh, V.K., et al. (2021). SLSNet: Skin lesion segmentation using a lightweight generative adversarial network. *Expert Systems with Applications*, 183: 115433. <https://doi.org/10.1016/j.eswa.2021.115433>