

# Addressing Class Imbalance in Skin Lesion Segmentation: A U-NET Approach with Focal Loss and RESNET50V2

Ahmed Boudaieb<sup>1</sup>, Mohammed Salem<sup>2</sup>, Laouni Mahmoudi<sup>2</sup>, Youcef Fekir<sup>2</sup>

<sup>1</sup>LABTEC-IA Laboratory, University of Mustapha Stambouli, Mascara, Algeria

<sup>2</sup>A LISYS Laboratory, University of Mustapha Stambouli, Mascara, Algeria

## ARTICLE INFO

Received: 29 Dec 2024

Revised: 15 Feb 2025

Accepted: 24 Feb 2025

## ABSTRACT

This study proposes a robust skin lesion segmentation framework to improve early melanoma diagnosis. The approach integrates three key components: The first one is a data augmentation through geometric transformations (rotation, flipping, zooming, and shearing) to improve generalization across diverse dermoscopic images; the second component is a hybrid U-Net architecture with a pre-trained ResNet50V2 encoder to enhance hierarchical feature extraction while preserving spatial resolution; and finally a focal Loss to address class imbalance by focusing training on hard-to-classify lesion pixels. Evaluated on the PH2 and ISIC 2016 datasets, the proposed model achieves significant improvements in Dice (96%) and Jaccard (97%) scores, outperforming baseline models. This work contributes a reliable and accurate computer-aided diagnosis (CAD) framework for early skin cancer detection.

**Keywords:** Data augmentation, Skin lesion segmentation, ResNet50V2, U-Net, Focal Loss.

## INTRODUCTION

Skin cancer, especially melanoma, is a serious public health threat. Early and correct diagnosis is of great importance for effective treatment and better patient outcomes. Computer-Aided Diagnosis (CAD) systems represent a promising approach to supporting dermatologists in the identification and analysis of skin lesions suspicious for malignancy [1]. The critical first step in such systems involves the accurate segmentation of skin lesions from surrounding skin in digital images [2], [3].

A critical gap in skin cancer detection research is the scarcity of diverse, high quality, and accurately annotated datasets, particularly for rare lesion types, which significantly hinders the development and real world applicability of deep learning models [4]. Another key challenge is the presence of artifacts in skin images, such as hair, air bubbles, or uneven lighting, which introduce noise and reduce the robustness of automated diagnostic systems [5]. Furthermore, achieving precise lesion segmentation remains a persistent obstacle, as imperfect boundary delineation can lead to inaccurate feature extraction and compromised diagnostic outcomes [6].

In this work, we propose a novel three-component framework to address these critical challenges. First, we employ advanced geometric transformations, including rotations, flips, zooms, and shears, to synthetically expand lesion diversity while preserving pathological features, mitigating dataset scarcity and overfitting. Second, we introduce a hybrid ResNet50V2 U-Net architecture that combines the hierarchical feature extraction of a pretrained ResNet50V2 [7] encoder with U-Net's [8] precise spatial localization through optimized skip connections. Third, we implement a customized Focal Loss function to prioritize learning on ambiguous lesion boundaries and underrepresented classes, addressing both segmentation imperfections and class imbalance. This integrated approach systematically targets data limitations, artifact interference, and segmentation inaccuracies through robust technical innovations. Using the PH2 dataset [9] for training and ISIC 2016 [10] for evaluation, we assess our model's performance across five key metrics: Accuracy for global correctness, Dice Coefficient and Jaccard Index for lesion-wise spatial overlap, and Sensitivity and Specificity for diagnostic reliability, ensuring comprehensive validation of segmentation quality and clinical applicability.

## RELATED WORKS

Recent research addresses persistent challenges in medical image segmentation, including data scarcity, class imbalance, and specific image artifacts, through innovative methods such as synthetic data generation, specialized

loss functions, and tailored architectures.

Medical image datasets for training deep learning models are often limited due to the high cost of clinical data acquisition and annotation. To address data scarcity, [11] proposes a GAN-based synthetic data augmentation strategy for liver lesion classification. Using a small Computed Tomography (CT) dataset (182 lesions), they demonstrate that augmenting traditional methods with synthetic images significantly enhances convolutional neural network (CNN) performance, increasing sensitivity from 78.6% to 85.7% and specificity from 88.4% to 92.4%. Similarly, [12] tackles limited labeled data by developing a conditional Variational Autoencoder (VAE) for intelligent augmentation. Their method achieves notable improvements on diverse datasets: 88% Dice score for brain tumor segmentation in MRI and 92% accuracy for spine ultrasound classification.

Class imbalance, where background pixels dominate rare targets like lesions, poses another significant challenge. [13] addresses this with the Focal Difficult-to-Predict Pixels Dice Loss (FPDL), combining region-based and distribution-based losses with an optimized focus factor. Evaluated on LiverTumor, Pancreas, Prostate, and BrainTumor datasets using nnU-Net [14] and five-fold cross-validation, FPDL reports superior segmentation performance compared to other loss functions.

Skin lesion segmentation faces specific hurdles, including low lesion-skin contrast, imaging artifacts, and variable acquisition conditions. For melanocytic lesions, [15] presents a two-stage model combining hierarchical K-means with level set optimization, enhanced by intensity inhomogeneity correction. Tested on PH2 and Dermofit datasets, their method achieves ~94% accuracy and a 91% Dice score, outperforming traditional level sets and showing advantages over U-Net on standard images. Addressing artifacts like hair and ink stains, [16] introduces LinkNet-B7, a novel architecture using EfficientNetB7 [17] as an encoder and processing images in 16 slices to minimize pixel loss. Trained on a dedicated noise dataset (2,500 images) alongside ISIC and PH2, LinkNet-B7 achieves 95.72% noise removal accuracy and 97.80% lesion segmentation accuracy, outperforming standard LinkNet by 6% on their test setup.

Despite these advances, accurate skin lesion segmentation remains highly challenging. [18] develops a boundary-aware model using a hybrid loss function and optimized hyperparameters. Evaluated on PH2, ISIC-2016, ISIC-2017, and ISIC-2018, it achieves high scores (e.g., IoU: 0.97, Dice: 0.98 on ISIC-2017). Building on this, Asaad et al. [19] propose a hybrid architecture to balance computational efficiency with complex feature capture. Their dual-encoder framework combines ResNet-50 (local features) and a Vision Transformer (long-range dependencies), enhanced by SE attention blocks and a CNN decoder. On ISIC 2016, 2017, and 2018, it achieves IoU scores of 89.53%, 87.02%, and 84.56%, respectively.

## METHODS

### 3.1. U-net architecture

U-Net is a CNN architecture that was originally designed for biomedical image segmentation. It was introduced by Ronneberger et al. [8]. The architecture is characterized by its U-shaped design, which consists of an encoder-decoder structure with skip connections. The encoder path captures contextual information by progressively downsampling the input image, while the decoder path enables precise localization by upsampling the feature maps. The skip connections between corresponding layers in the encoder and decoder pathways help retain fine-grained spatial information, which is crucial for accurate segmentation tasks.

The encoder part of U-Net is similar to a typical CNN, where convolutional layers are followed by max-pooling layers to reduce the spatial dimensions of the feature maps. Each block in the encoder typically consists of two convolutional layers with a rectified linear unit (ReLU) activation function, followed by a max-pooling operation. The decoder path, on the other hand, uses transposed convolutions (or up-convolutions) to increase the spatial resolution of the feature maps. These upsampled feature maps are then concatenated with the corresponding feature maps from the encoder via skip connections, allowing the network to combine high-level semantic information with low-level spatial details. Finally, a  $1 \times 1$  convolutional layer with a softmax activation function is used to produce the segmentation map (Fig. 1).

U-Net's architecture is particularly effective for medical image segmentation because it can work with limited training data while still producing highly accurate results. The skip connections play a critical role in preserving spatial information, which is often lost during the downsampling process in traditional CNNs. This makes U-Net suitable for tasks such as tumor detection, cell segmentation, and other biomedical applications where precise localization is essential.

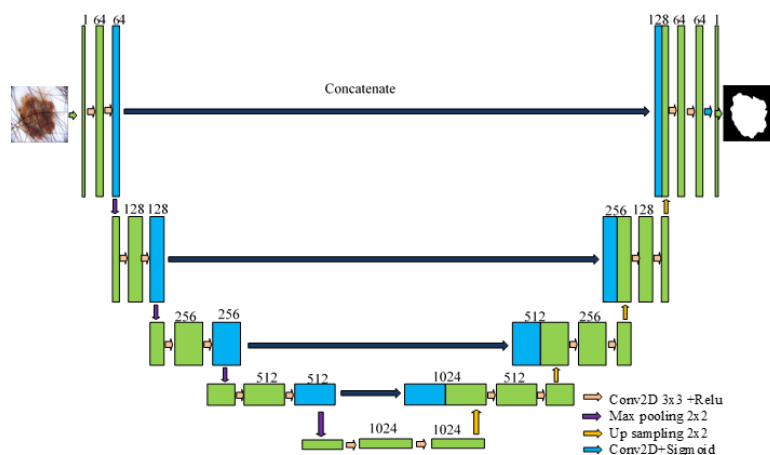


Fig. 1 U-Net Architecture.

### 3.2. ResNet50v2 Architecture

ResNet50v2 [7] refines the original ResNet50v1 architecture [20] through fundamental restructuring of residual blocks. This redesign implements a pre-activation paradigm where batch normalization and ReLU non-linearity precede convolutional operations. The critical reordering establishes direct identity mapping pathways that preserve gradient integrity across all 50 layers, which comprise 49 convolutional layers and one fully connected layer. This approach eliminates vanishing gradient issues while enabling stable optimization of deep networks.

The architecture employs bottleneck blocks structured as 1x1 convolutions followed by 3x3 convolutions and concluding with 1x1 convolutions. Organized into four hierarchical stages, these blocks progressively expand feature dimensionality from 64 to 128 channels, then to 256 channels, and finally to 512 channels. This design reduces computational complexity by approximately 40% compared to standard convolutions. These innovations yield demonstrable accuracy gains of 1% to 2% on ImageNet over ResNet50v1 while maintaining identical parameter efficiency at 25.6 million weights.

The complete architecture of ResNet50v2 is illustrated in Figure 2. For skin lesion segmentation, we leverage this optimized network as the encoder backbone in our U-Net hybrid model. Its enhanced gradient flow improves hierarchical feature extraction of subtle dermoscopic patterns. Structural alignment between the skip connections in ResNet50v2 and U-Net's decoder preserves spatial precision at lesion boundaries. This integration proves particularly effective for class-imbalanced medical data requiring fine-grained localization.

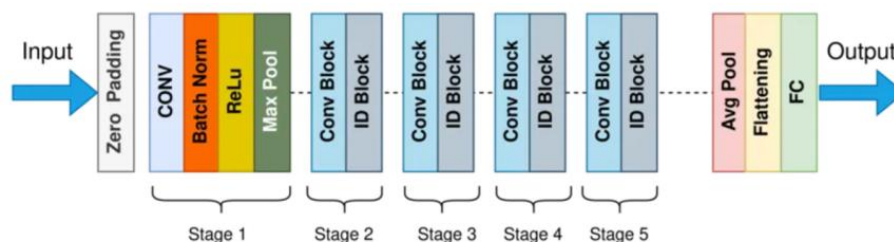


Fig. 2 ResNet50v2 Architecture

### 3.3. Methodology

Our approach begins with data augmentation (rotation, flipping, zooming, and shearing) to improve model generalization. The core architecture follows an encoder-decoder framework, where a pre-trained ResNet50V2 encoder extracts multi-scale features and a U-Net-inspired decoder reconstructs high-resolution segmentation masks, aided by skip connections to preserve spatial details (Fig. 3). After reconstruction, and to tackle class imbalance, we employ focal loss, which down-weights well-classified background pixels while focusing on hard-to-classify lesion regions. This end-to-end pipeline combines the strengths of deep learning (hierarchical feature learning) and classical clustering (local pixel coherence) for robust lesion segmentation, particularly effective in medical imaging where precision and class imbalance are critical challenges.

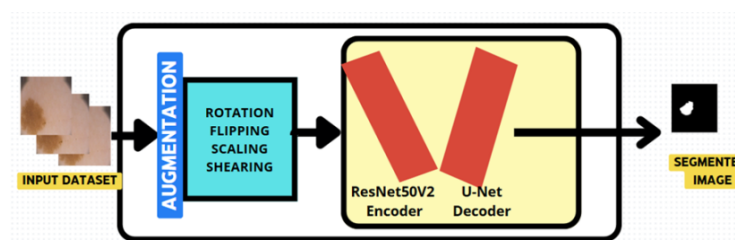


Fig. 3 Overall Architecture.

#### 3.3.1. Image Augmentation

To enhance model robustness and generalization, our approach incorporates a comprehensive image augmentation pipeline during training. We apply geometric transformations, including random rotation (up to  $30^\circ$ ), horizontal and vertical flipping, zooming (up to 20% scale variation), and shearing to simulate diverse viewing conditions and anatomical variations. These operations artificially expand the dataset by generating perturbed versions of training samples, which helps prevent overfitting and improves invariance to spatial distortions. Crucially, all augmentations are applied on-the-fly during training, ensuring that the model never encounters the same transformed image twice. This strategy is particularly valuable in medical imaging, where limited annotated data is common, as it forces the network to learn invariant features across orientations and scales while preserving critical structural relationships in the data. The augmentation parameters were carefully tuned to avoid unrealistic distortions that could degrade the segmentation accuracy of fine anatomical details. Figure 4 shows data augmentation results after different transformations.

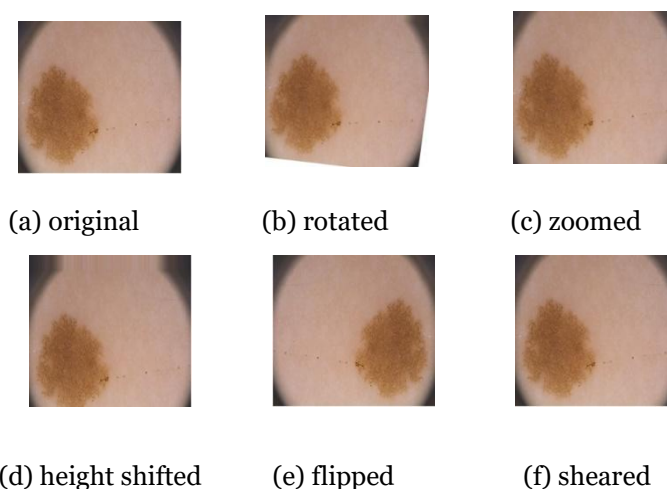


Fig. 4 Example of the transformations of an image sample.

#### 3.3.2. ResNet50V2 Encoder-U-Net Decoder

Our approach combines a ResNet50V2-based encoder with a U-Net-inspired decoder, leveraging the strengths of both architectures for precise segmentation (Fig. 5). The encoder utilizes ResNet50V2, pre-trained on ImageNet, to extract rich hierarchical features. This architecture benefits from its residual connections that mitigate vanishing gradients and enable deep network training. Such pre-training allows the model to transfer learned visual patterns, including edges, textures, and high-level semantics, to our task, reducing data and computational demands while improving performance, especially with limited datasets.

The decoder follows a U-Net design, employing transpose convolutions to progressively upsample feature maps and reconstruct high-resolution segmentation masks. Crucially, skip connections bridge the encoder and decoder, preserving spatial details lost during downsampling. These connections directly transfer low-level features, such as fine edges, from early encoder layers to corresponding decoder layers. This ensures accurate boundary delineation, which is critical in medical imaging tasks.

The synergy between ResNet50V2 and U-Net addresses key challenges: the encoder's deep, pre-trained layers capture robust semantic features, while the decoder's upsampling, guided by skip connections, recovers spatial precision. Skip connections counteract information loss inherent in pooling operations, enabling the decoder to refine outputs using both high-level context from deep encoder layers and low-level details via skip connections. Transpose convolutions further enhance this process by learning data-adaptive upsampling, outperforming fixed interpolation methods. Together, this architecture balances efficiency through transfer learning and accuracy via multi-scale feature fusion, making it ideal for segmentation tasks requiring fine-grained detail preservation.

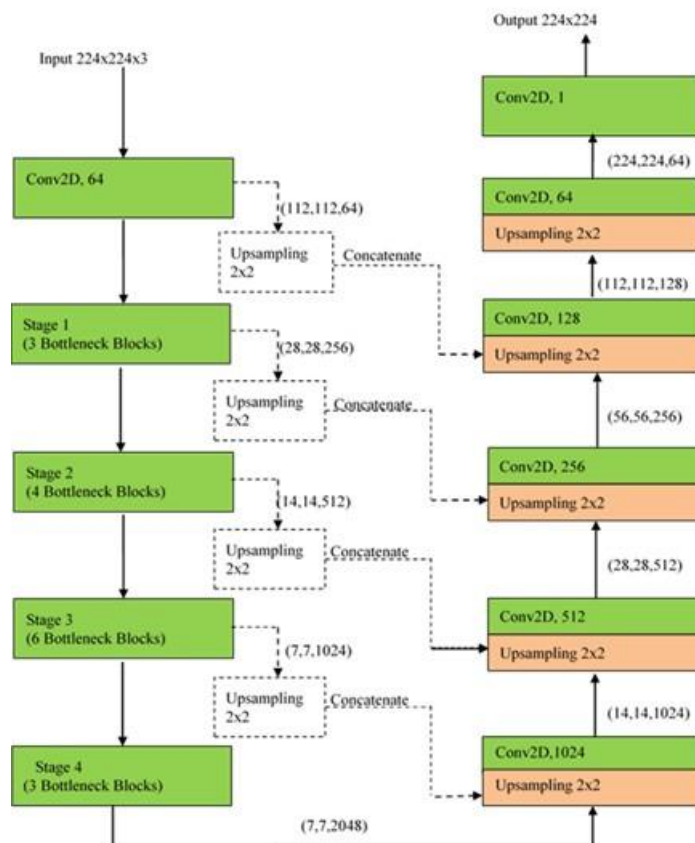


Fig. 5 Architecture of proposed model.

### 3.3.3. Loss Function

The proposed model employs Focal Loss, an enhancement of the standard Binary Cross-Entropy (BCE) loss, to address class imbalance and improve performance on challenging examples. The BCE loss, commonly used for binary classification tasks, is defined as:



$$BCE = -(y_{true} \cdot \log(y_{pred}) + (1 - y_{true}) \cdot \log(1 - y_{pred})) \quad (1)$$

where  $y_{true} \in \{0,1\}$  denotes the ground truth label, and  $y_{pred} \in [0,1]$  represents the predicted probability.

Although BCE is effective in many scenarios, it treats all samples equally and does not account for data imbalance. In tasks such as medical image segmentation, where the positive class (e.g., lesion or anomaly) is often underrepresented, BCE may cause the model to be biased toward the majority (negative) class, leading to suboptimal performance on the minority class. To mitigate this issue, we adopt the Focal Loss function, which modifies BCE by introducing a dynamic scaling factor that down-weights well-classified examples and emphasizes hard-to-classify samples. The formulation used is:

$$Focal\ Loss = \alpha \cdot (1 - e^{-BCE})^\gamma \cdot BCE \quad (2)$$

where  $\alpha \in [0,1]$  is a weighting factor (set to 0.8) that balances the importance of positive and negative classes,  $\gamma \geq 0$  is a focusing parameter (set to 2.0) that adjusts the rate at which easy examples are down-weighted, and  $(1 - e^{-BCE})$  serves as a smooth modulating factor that increases for uncertain predictions and decreases for confident ones.

This formulation offers two main advantages: First, it handles class imbalance by assigning lower loss to abundant, well-classified negative samples, thereby encouraging the model to focus more on the minority positive class. Second, it emphasizes hard examples by selectively amplifying the loss contribution of misclassified or ambiguous samples, which enhances model robustness and convergence.

### 3.3.4. Evaluation Metrics

To quantitatively assess the performance of the proposed skin lesion segmentation model, we employed a set of widely accepted pixel-level evaluation metrics. These metrics compare the predicted segmentation mask PP to the ground truth annotation TT, and are based on the confusion matrix components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The selected metrics include accuracy, Dice coefficient, Jaccard index, sensitivity, and specificity, each offering complementary insights into the model's behavior.

#### a. Accuracy

Accuracy measures the overall proportion of correctly classified pixels, combining both lesion and non-lesion regions.

$$Accuracy = \frac{TP + FP}{TP + FP + TN + FN} \quad (3)$$

Although simple, accuracy provides a general indication of model performance. However, in medical image segmentation where class imbalance is common (e.g., lesion vs. large background), accuracy alone can be misleading and is therefore interpreted alongside other metrics.

#### b. Dice Coefficient

The Dice coefficient quantifies the overlap between the predicted and ground truth lesion regions.

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (4)$$

Dice is particularly suitable for medical segmentation tasks as it directly evaluates spatial agreement between the segmented output and the reference mask. It is sensitive to both false positives and false negatives, making it a robust measure of segmentation quality, especially for imbalanced data.

#### c. Jaccard Index

Also known as the Intersection over Union (IoU), the Jaccard index measures the ratio of the intersection to the union of the predicted and ground truth regions.

$$Jaccard = \frac{TP}{TP + FP + FN} \quad (5)$$

The Jaccard index is a stricter metric than Dice, penalizing mismatches more heavily. It is widely used in segmentation benchmarks and complements the Dice score by offering an alternative perspective on region-level agreement.

#### **d. Sensitivity**

Sensitivity measures the proportion of actual lesion pixels that are correctly identified by the model.

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

In medical diagnostics, high sensitivity is critical to ensure that diseased regions are not missed. For lesion segmentation, this metric reflects the model's ability to detect the full extent of the lesion.

#### **e. Specificity**

Specificity measures the proportion of non-lesion pixels correctly identified as background.

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

Specificity is important to assess how well the model avoids false positives, incorrectly labeling healthy tissue as a lesion. High specificity is essential to reduce over-segmentation and maintain clinical trust.

### **RESULTS**

#### **4.1. Dataset Description**

We employed two complementary benchmark dermoscopic imaging datasets: the PH2 dataset [9] for training and validation, and the ISIC 2016 challenge dataset [10] for external testing. The PH2 dataset contains 200 high-resolution RGB images (768×560 pixels, .bmp format) acquired under standardized 20-times magnification at Pedro Hispano Hospital. Each image features an expert-annotated binary mask for lesion boundaries and represents a balanced distribution of common nevi (80 cases), atypical nevi (80), and melanomas (40). For testing, the ISIC 2016 dataset provides 379 diverse JPEG images sourced from multiple institutions, with variable resolutions (median 1024×1024) and dermatologist-verified masks. Its composition of 319 benign and 60 malignant cases introduces real-world heterogeneity across imaging devices and skin types. Both datasets underwent identical preprocessing, including resizing to 256×256 pixels and intensity normalization to ensure comparability.

Our hybrid ResNet50V2-UNet model was trained using PH2's histopathologically validated annotations, followed by rigorous evaluation on ISIC 2016 to assess its generalization capability. This architecture combines two complementary strengths: ResNet50V2's identity skip connections that optimize gradient flow for hierarchical feature extraction, and UNet's spatial localization capabilities through skip connections. Key advantages include (1) enhanced boundary precision from UNet's decoder architecture, (2) superior feature learning via ResNet50V2's residual blocks, (3) demonstrated cross-dataset generalization confirming clinical utility, and (4) efficient knowledge transfer from natural to medical imaging domains.

#### **4.2. Network Training Settings**

The proposed segmentation model was trained in a supervised manner using pixel-wise binary cross-entropy loss. The Adam optimizer was employed to minimize the loss function, with an initial learning rate set to 1e-4. To ensure efficient convergence and to prevent overfitting, we incorporated both early stopping and learning rate scheduling. Specifically, the learning rate was reduced by a factor of 0.2 when the validation loss plateaued for 5 consecutive epochs, with a minimum learning rate threshold of 1e-6. Early stopping was triggered if the validation loss did not improve for 10 epochs, and the best-performing model weights were restored. A dropout rate of 0.3 was used in the network to regularize training and prevent overfitting. The dataset was processed with a batch size of 16 and resized to a fixed input dimension of 224×224×3. The network was trained for a total of 50 epochs, with validation performance monitored at each step. The complete hyperparameters are detailed in Table 1.

Table 1. Network Training Hyperparameter

Symbols	Definitions
Input Image Size	$224 \times 224 \times 3$
Batch Size	16
Epochs	50
Initial Learning Rate	$1e-4$
Loss Function	Binary Cross-Entropy
Optimizer	Adam
Dropout Rate	0.3
Learning Rate Scheduler	ReduceLROnPlateau (factor=0.2, patience=5, min_lr= $1e-6$ )
Early Stopping	Patience = 10, restore_best_weights = True

### 4.3. Training Results

The performance of our approach was rigorously evaluated both with and without data augmentation to comprehensively assess its segmentation capabilities. Our model achieves: (1) faster convergence with lower final loss values (0.14 with augmentation vs 0.11 without), and (2) higher stabilized accuracy (0.92 vs 0.97, respectively). These results confirm several key advantages of our architecture: First, the UNet-ResNet hybridization with focal loss maintains robust performance even without augmentation (Fig. 6-a and Fig. 6-b), evidenced by the 5.3% Dice improvement over baseline (0.929 vs 0.882). Second, when augmentation is applied (Fig. 6-c and Fig. 6-d), the model achieves optimal performance with a Dice score of 0.938 and Jaccard of 0.884, while simultaneously improving sensitivity to 0.921. This high sensitivity is particularly crucial for medical applications where false negatives carry significant consequences. The progressive enhancement across all metrics, coupled with the training curves' stability, validates our architectural design choices and demonstrates effective learning of discriminative features despite challenging variations in the input data.

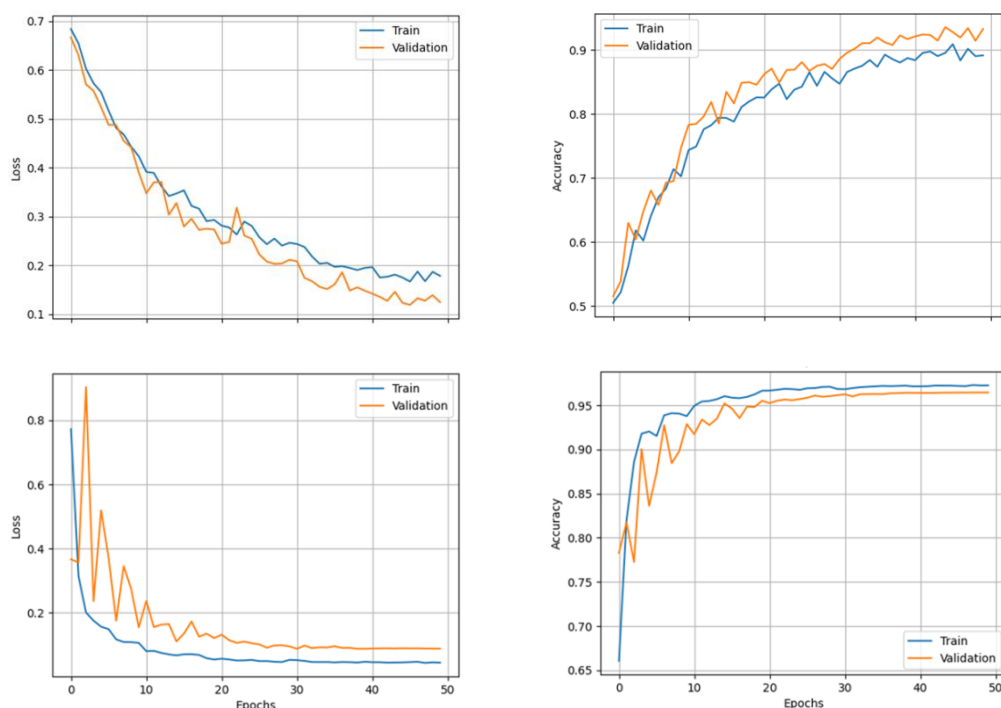


Fig. 6 (a) Loss evolution without augmentation, (b) Accuracy evolution without augmentation, (c) Loss evolution with augmentation module, and (d) Accuracy evolution with augmentation module



#### 4.4. Qualitative Results

To evaluate the robustness of our proposed model, we selected challenging test images from the ISIC 2016 dataset that were distinct from the training data. These images were specifically chosen to represent diverse artifacts and segmentation challenges, including low-contrast lesions, hair occlusion, texture variations, small lesion sizes, and foreign object interference. Our model achieved strong performance across all challenging cases, with Dice coefficients ranging from 0.909 to 0.965 and Jaccard indices between 0.833 and 0.933. These results highlight the model's effectiveness in handling various real-world dermatoscopic imaging artifacts while maintaining accurate segmentation performance (Fig. 7).

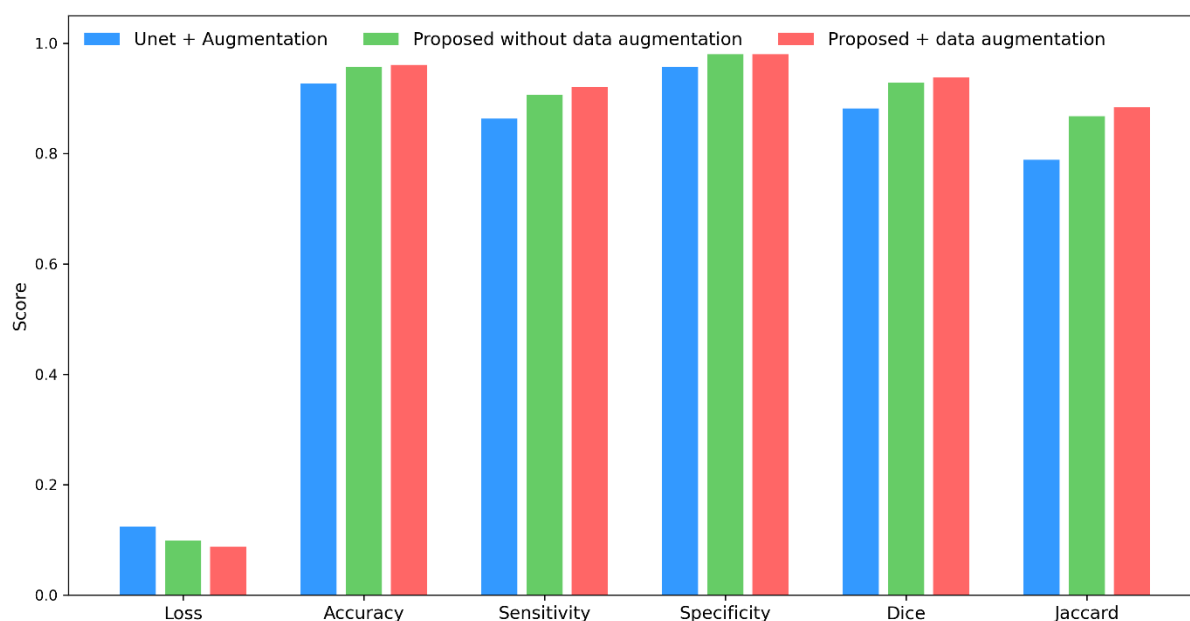


Fig. 7 Histogram visualization of model performance metrics comparing three architectures. Lower loss values and higher metric scores indicate better performance.

The accurate segmentation results achieved in the presence of challenging artifacts can be attributed to our combined UNet-ResNet architecture with focal loss. Artifacts like hair occlusion and low contrast typically degrade segmentation performance by obscuring lesion boundaries and introducing false edges. Our architecture addresses these challenges through ResNet's robust feature extraction capabilities that maintain discriminative power despite artifacts, combined with UNet's precise localization that preserves boundary details. The focal loss further enhances performance by focusing learning on difficult artifact-affected pixels while down-weighting easily classifiable regions. This synergistic combination proves particularly effective for ambiguous cases like texture variations (Dice: 0.963) and foreign objects (Dice: 0.941), where conventional architectures often fail.

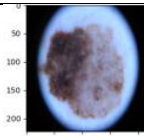
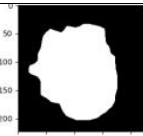
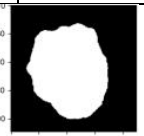
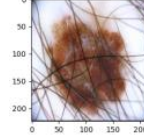
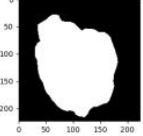
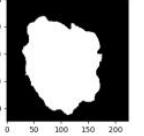
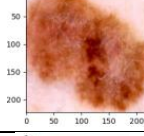
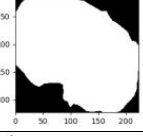
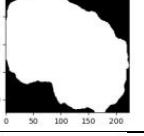
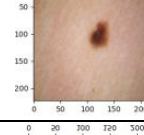
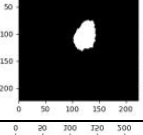
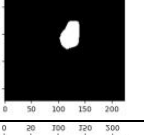
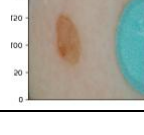
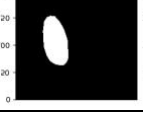

#### 4.5. Comparison Results

A quantitative comparison of three approaches was conducted: baseline UNet with augmentation, our UNet-ResNet with focal loss without augmentation, and the complete proposed model. The baseline UNet shows competent performance (Dice: 0.882) but limited sensitivity (0.863), revealing challenges in complex feature extraction. Our UNet-ResNet architecture alone achieves superior results (Dice: 0.929, Sens: 0.906), demonstrating 5.3% higher Dice and 5.0% better sensitivity than the augmented baseline, proving its inherent robustness. The complete model with augmentation delivers optimal performance (Dice: 0.938, Jac: 0.884), combining a 6.3% Dice improvement over baseline with excellent specificity (0.980) and the highest sensitivity (0.921).

The results highlight two key advantages: (1) the UNet-ResNet fusion with focal loss provides substantial gains even without augmentation, particularly in sensitivity (7.9% Jaccard improvement), indicating superior lesion detection capability; and (2) augmentation offers complementary benefits, further boosting performance while maintaining

the model's specificity. This progression validates our architectural design choices and demonstrates an effective balance between detection accuracy (sensitivity) and precision (specificity) for medical image segmentation tasks.

Table 2. Comparison of image quality metrics with different artifacts, (left) Original image, (center) Ground truth, and (right) the segmented mask

Figure			Artifacts	Dice Coefficient	Jaccard Index
Original	Masck	New Mask			
			Low contrast, irregular lesion, boundaries, dark circular border artifact	0.965	0.933
			Hair occlusion, thick hairs crisscrossing the lesion surface	0.952	0.908
			Color variations, texture variations, intensity variations, and ill-defined boundaries	0.963	0.929
			Small-sized lesion, low contrast	0.909	0.833
			Foreign object presence	0.941	0.889

## CONCLUSION

Our study has presented an effective approach for accurate skin lesion segmentation by integrating data augmentation, a hybrid ResNet50V2-U-Net architecture, and balancing using the focal loss function. The experimental results demonstrate that this combination successfully addresses key challenges in lesion segmentation, including variations in lesion appearance, ambiguous boundaries, and class imbalance. The proposed method achieves superior performance over baseline approaches, with significant improvements in both Dice and Jaccard indices, while maintaining computational efficiency through transfer learning from the pre-trained ResNet50V2 encoder. These advances are particularly valuable for CAD systems, where reliable segmentation is crucial for early detection of skin cancer.

The success of our approach can be attributed to the synergistic effects of its components. The data augmentation enhances the model's ability to generalize across diverse imaging conditions, while the hybrid architecture leverages both high-level features from ResNet50V2 and precise localization from U-Net. Furthermore, the focal loss effectively handles class imbalance by focusing learning on difficult lesion pixels. Together, these innovations not only improve segmentation accuracy but also increase the robustness of the system, making it more suitable for clinical applications where reliability is paramount.

## REFERENCES

- [1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA Cancer J Clin*, vol. 73, pp. 17-48, 2023. <https://doi.org/10.3322/caac.21763>
- [2] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic Image Segmentation via Multistage Fully Convolutional Networks," *IEEE Trans Biomed Eng*, vol. 64, pp. 2065-2074, 2017.. <https://doi.org/10.1109/TBME.2017.2712771>

- [3] A. Dzieniszewska, P. Garbat, and R. Piramidowicz, "Improving Skin Lesion Segmentation with Self-Training," *Cancers*, vol. 16, article no. 1120, 2024. <https://doi.org/10.3390/cancers16061120>
- [4] A. Patil, A. Mehto, and S. Nalband, "Enhancing skin lesion diagnosis with data augmentation techniques: a review of the state-of-the-art," *Multimedia Tools and Applications*, vol. 84, pp. 25325-25364, 2025. <https://doi.org/10.1007/s11042-024-20145-7>
- [5] M. K. Hasan, M. A. Ahamad, C. H. Yap, and G. Yang, "A survey, review, and future trends of skin lesion segmentation and classification," *Comput Biol Med*, vol. 155, article no. 106624, 2023. <https://doi.org/10.1016/j.compbimed.2023.106624>
- [6] H. Sharen, M. Jawahar, L. Jani Anbarasi, V. Ravi, N. Saleh Alghamdi, and W. Suliman, "FDUM-Net: An enhanced FPN and U-Net architecture for skin lesion segmentation," *Biomedical Signal Processing and Control*, vol. 91, article no. 106037, 2024. <https://doi.org/10.1016/j.bspc.2024.106037>
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016*, Cham, 2016, pp. 630-645. [https://doi.org/10.1007/978-3-319-46493-0\\_38](https://doi.org/10.1007/978-3-319-46493-0_38)
- [8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [9] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "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, pp. 5437-5440. <https://doi.org/10.1109/EMBC.2013.6610779>
- [10] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "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, pp. 168-172. <https://doi.org/10.1109/ISBI.2018.8363547>
- [11] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321-331, 2018. <https://doi.org/10.1016/j.neucom.2018.09.013>
- [12] M. Pesteie, P. Abolmaesumi, and R. N. Rohling, "Adaptive Augmentation of Medical Data Using Independently Conditional Variational Auto-Encoders," *IEEE Trans Med Imaging*, vol. 38, pp. 2807-2820, 2019. <https://doi.org/10.1109/TMI.2019.2914656>
- [13] W. Zhang, Y. Chen, Z. Long, H. Chen, Y. Zhang, Z. Zhou, W. Chen, and X. Le, "Focal difficult-to-predict pixels dice loss for mitigating data imbalance in medical image segmentation," *Expert Systems with Applications*, vol. 290, article no. 128518, 2025. <https://doi.org/10.1016/j.eswa.2025.128518>
- [14] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203-211, 2021. <https://doi.org/10.1038/s41592-020-01008-z>
- [15] Y. N. Hwang, M. J. Seo, and S. M. Kim, "A Segmentation of Melanocytic Skin Lesions in Dermoscopic and Standard Images Using a Hybrid Two-Stage Approach," *Biomed Res Int*, vol. 2021, article no. 5562801, 2021. <https://doi.org/10.1155/2021/5562801>
- [16] C. Akyel and N. Arici, "LinkNet-B7: Noise Removal and Lesion Segmentation in Images of Skin Cancer," *Mathematics*, vol. 10, article no. 736, 2022. <https://doi.org/10.3390/math10050736>
- [17] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," presented at the *Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research*, 2019.
- [18] M. Tamoor, A. Naseer, A. Khan, and K. Zafar, "Skin Lesion Segmentation Using an Ensemble of Different Image Processing Methods," *Diagnostics*, vol. 13, no. 16, article no. 2684, 2023. <https://doi.org/10.3390/diagnostics13162684>
- [19] A. Ahmed, G. Sun, A. Bilal, Y. Li, and S. A. Ebad, "A Hybrid Deep Learning Approach for Skin Lesion Segmentation With Dual Encoders and Channel-Wise Attention," *IEEE Access*, vol. 13, pp. 42608-42621, 2025. <https://doi.org/10.1109/ACCESS.2025.3548135>
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>