International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A ROBUST U-NET BASED FRAMEWORK FOR HIGH-FIDELITY SKIN LESION SEGMENTATION

Deepa J[1] , Anjana Joshy[2] , Sunil S S[3] , Reji R[4] , Divya Saleela[5] , Chinchu M S[6]
deepaj@cek.ac.in[1], anjanajoshy999@gmail.com[2], sssunil754@gmail.com[3], rejir@tkmit.ac.in[4], d.saleela@soton.ac.uk [5], chinchums18@gmail.com [6]
[1]Professor, College of Engineering Kallooppara, India
[2]Student, University of Hertfordshire, United Kingdom
[3]Associate Professor, Sree Buddha College of Engineering, India
[4]Professor, Thangal Kunju Musaliar Institute of Technology, India
[5]Postdoctoral Research Fellow, University of Southampton, United Kingdom
[6]Assistant Professor, Sree Buddha College of Engineering, India

Abstract

Automated segmentation of skin lesions remains a central challenge in dermatological imaging owing to substantial variability in lesion morphology, colour distribution and acquisition artefacts. This study presents a lightweight yet high-precision U-Net framework that integrates a hybrid Dice-binary cross-entropy optimisation scheme together with an expanded evaluation protocol incorporating boundary-aware metrics, offering improved robustness without architectural complexity. Trained and validated on 2594 dermoscopic images from the ISIC 2018 benchmark, the model achieved state-of-the-art performance for standard U-Net configurations, with a Dice coefficient of 0.94, intersection-over-union of 0.89, precision of 0.96, recall of 0.93, F1 score of 0.94, average symmetric surface distance of 5.4 pixels and 95th-percentile Hausdorff distance of 12.7 pixels. Qualitative inspection confirmed strong boundary localisation and stable behaviour across irregular, low-contrast and clinically challenging lesions. These findings demonstrate that substantial performance gains can be achieved through optimised training dynamics and a rigorously structured evaluation pipeline, establishing a reproducible and computationally efficient baseline that can support future methodological development in automated dermatological analysis.

Keywords skin lesion segmentation, dermoscopy, ISIC 2018 dataset, U-Net, deep learning, medical image analysis, semantic segmentation, Dice coefficient, boundary metrics, dermatological imaging

1. Introduction

Skin cancer represents one of the most rapidly expanding oncological burdens of the twenty-first century. Global surveillance data indicate that more than three million cases of non-melanoma skin cancer and over three hundred thousand new melanomas are diagnosed annually [1]. Although melanoma accounts for only a minority of overall cases, it is responsible for the vast majority of skin-cancer mortality, contributing to approximately fifty-seven thousand deaths each year. Within the United Kingdom, the epidemiological pattern mirrors global trajectories, with melanoma now ranking as the fifth most frequently diagnosed cancer and exhibiting a sustained rise in incidence across all age groups [2]. These developments underline a critical clinical reality: timely and accurate detection remains the single most significant determinant of long-term survival.

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Dermoscopy has become the standard imaging modality for the visual assessment of pigmented skin lesions, providing magnified, high-contrast views of subsurface structures that are not accessible to the naked eye. Despite its diagnostic utility, dermoscopic interpretation is susceptible to inter-observer variability, especially in early-stage or morphologically atypical lesions, and diagnostic accuracy is strongly correlated with clinician experience [3], [4]. This dependency has motivated an intensive research effort into automated computational systems capable of providing objective and reproducible lesion assessment.

Segmentation constitutes a pivotal first step in automated dermoscopic analysis, as it defines the spatial boundaries from which all subsequent morphological, textural and structural descriptors are derived. Imperfect delineation propagates through classification pipelines, undermining the reliability of malignancy prediction. However, robust segmentation is notably challenging due to the considerable heterogeneity present in dermoscopic images, including variations in illumination, acquisition equipment, skin tone, lesion morphology and the presence of occlusions such as hair and ruler artefacts [5]. These factors necessitate segmentation models that not only demonstrate excellent accuracy in typical cases but also retain stability across visually ambiguous or corrupted images.

The ISIC 2018 dataset [6] provides a comprehensive and internationally curated benchmark designed to test segmentation performance under these realistic conditions. Its large collection of clinically diverse images serves as a rigorous platform for evaluating generalisation capacity. However, the breadth of acquisition characteristics introduces significant methodological challenges, requiring models that combine efficient representational learning with strong boundary sensitivity.

Deep learning has transformed the computational analysis of medical images, with fully convolutional encoder-decoder architectures emerging as the dominant paradigm for segmentation. U-Net, in particular, has become a foundational model due to its ability to preserve fine spatial information through skip-connections while simultaneously capturing global semantic context [7]. Numerous refinements of U-Net have been proposed to enhance boundary definition, suppress artefacts and mitigate class imbalance, achieving substantial gains in performance across medical domains including dermatology [8] [9] . Nevertheless, segmentation quality remains sensitive to architectural choices, loss formulations and optimisation strategies, and achieving clinically meaningful accuracy often requires careful balancing of algorithmic complexity with computational efficiency.

In this study, a refined U-Net-based framework is developed to deliver high-fidelity dermoscopic lesion segmentation while maintaining computational accessibility. The model incorporates a hybrid loss formulation combining binary cross-entropy and Dice loss, designed to stabilise training, address class imbalance and reinforce fine-grained boundary adherence. The architecture is complemented by a rigorous preprocessing pipeline that standardises image geometry and mitigates the influence of common dermoscopic artefacts. The objective is to derive a segmentation framework that is sufficiently robust for high-quality research applications yet efficient enough for potential deployment in settings with limited computational infrastructure, including teledermatology and community-based screening.

To provide a comprehensive and clinically relevant assessment, the model is evaluated using a broad suite of metrics encompassing global overlap, statistical discrimination and geometric

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
boundary accuracy. Alongside the Dice coefficient and intersection-over-union, the evaluation includes precision, recall, specificity, F1 score, average symmetric surface distance and the ninety-fifth percentile Hausdorff distance, thereby capturing both volumetric agreement and contour fidelity in accordance with contemporary evaluation standards in medical image analysis. Such multidimensional assessment is essential for differentiating models that may achieve superficially similar overlap scores yet diverge markedly in clinically meaningful performance.

The work presented here contributes to the field in several respects. It demonstrates that high segmentation accuracy on a challenging dermoscopic benchmark can be achieved using a computationally lean framework, dispelling the assumption that increasingly deep or architecturally elaborate models are necessary for clinical-grade performance. It also provides a transparent, reproducible pipeline aligned with emerging best practices in medical imaging research, facilitating reliable comparison with future methods.

The remainder of this paper is organised to guide the reader through the conceptual and methodological progression of the study. Section Two reviews the literature on dermoscopic imaging, automated lesion analysis and recent advances in deep learning for biomedical segmentation. Section Three details the dataset, preprocessing procedures, model architecture and training strategy. Section Four reports quantitative and qualitative results and Section Five provides a critical discussion of the findings, their implications and remaining limitations. The final section concludes the paper and outlines future research directions for automated dermoscopic analysis.

2. Literature Review

Research on automated dermoscopic image segmentation has expanded considerably, supported by the availability of benchmark datasets and rapid advances in deep learning. Existing work can be organised across five principal research categories: dataset development and evaluation practices, convolutional architectures, multi-representation fusion strategies, transformer-based and hybrid models, and loss-function design. Each category has contributed to improved segmentation performance, yet each also exhibits structural limitations that continue to restrict robustness, generalisability and clinical applicability. These limitations collectively underline the need for architectures that remain efficient, stable and capable of capturing both regional and boundary information under heterogeneous imaging conditions.

Substantial progress has been driven by publicly accessible repositories such as the ISIC archive, which have established a standard experimental foundation for algorithmic comparison. Meta-analyses nevertheless demonstrate that performance varies greatly across studies due to annotation inconsistency, device heterogeneity and the limited routine use of external validation [10], [11]. These issues indicate a gap in evaluation completeness and robustness, as many published results may not reflect performance under the broader variability encountered in clinical workflows.

Convolutional neural networks have remained central to dermoscopic segmentation, with architectures such as U-Net and its derivatives demonstrating strong capability in delineating lesions with irregular morphology [12], [13]. More recent refinements involving residual pathways, attention mechanisms and channel recalibration enhance localisation of subtle

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
boundary transitions [14], [15]. However, despite these advances, convolutional designs remain constrained by limited receptive fields, reducing their capacity to model global structural context. This deficit represents a significant gap when handling diffuse or spatially complex lesions that require coherent long-range reasoning.

Parallel research has explored the integration of complementary feature representations in an effort to improve resilience to artefacts, illumination variability and structural ambiguity. Approaches combining colour-space transformations, gradient descriptors and deep feature embeddings have demonstrated improved sensitivity to low-contrast borders [16]. Hybrid pipelines merging handcrafted and learned features further enhance robustness across varied dermoscopic presentations [17]. Yet, fusion practices remain heterogeneous and often rely on ad-hoc heuristics rather than principled frameworks, leaving a methodological gap concerning standardisation, reproducibility and computational economy.

Emerging transformer-based and hybrid CNN-transformer models introduce mechanisms capable of modelling long-range dependencies more effectively than convolution alone, improving spatial coherence in complex cases [18], [19], [20]. While these architectures can deliver highly competitive results, they commonly require substantial memory and computational resources, limiting their suitability for deployment in clinical environments or on lightweight hardware. This defines an important gap between methodological sophistication and practical implementability.

Loss-function design has become increasingly critical, particularly given the severe class imbalance and ambiguous lesion boundaries characteristic of dermoscopic datasets. Regional similarity losses such as Dice, together with focal and boundary-aware formulations, have been shown to improve contour fidelity and detection of small or fragmented lesions [21] . Nonetheless, many studies still rely on single-objective losses that inadequately constrain structural accuracy or fail to stabilise optimisation under imbalance . This gap highlights the need for more balanced objectives capable of jointly capturing regional agreement and boundary precision.

Taken together, the literature demonstrates notable advances while simultaneously exposing limitations in robustness, contextual modelling, fusion coherence, computational feasibility and optimisation strategy. These persistent challenges motivate the development of segmentation frameworks that are both computationally efficient and capable of delivering accurate, structurally consistent predictions across heterogeneous dermoscopic images. The next section outlines the methodological design adopted in this study, including the dataset, preprocessing pipeline, network architecture and optimisation strategy.

3. Methodology

The methodological framework adopted in this work is designed to provide a transparent and reproducible basis for evaluating deep learning based segmentation of dermoscopic skin lesions using the ISIC 2018 dataset. The section first describes the dataset and its organisation, then outlines the preprocessing steps applied to images and masks. It subsequently presents the convolutional architecture and optimisation strategy, before detailing the evaluation protocol and experimental environment.

Received: Aug 03, 2025

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Dataset

Experiments were performed on the ISIC 2018 skin lesion segmentation dataset, which has become a widely used benchmark for dermoscopic image analysis. The dataset consists of colour dermoscopy images paired with expert annotated binary masks that delineate the lesion from surrounding skin. In the configuration used here, 2594 images from the official training collection were included, each associated with a corresponding ground truth segmentation map. The images exhibit substantial variability in lesion type, size, anatomical location, pigmentation pattern and background appearance. Many cases contain confounding artefacts such as hair, ruler marks or illumination gradients. This variability provides a realistic test bed for assessing segmentation models that must cope with heterogeneous clinical conditions. For supervised training and validation, the dataset was partitioned at the image level using an eighty to twenty split, resulting in 2 076 images allocated to the training set and 518 to the validation set. The split was generated once by random shuffling with a fixed seed and preserved unchanged across all experiments to guarantee reproducibility.

Preprocessing

Preprocessing aimed to standardise spatial resolution and intensity scaling while preserving diagnostically relevant information. Each dermoscopy image was loaded from disk, either as a JPEG or PNG file, and converted to RGB format. The associated ground truth mask was loaded as a single channel image in which lesion pixels are foreground and background skin is assigned zero. To match network requirements, both images and masks were resized to a common spatial resolution of 256 by 256 pixels. Images were resized using bilinear interpolation, whilst masks were resized using nearest neighbour interpolation to maintain label integrity. After resizing, images were converted to three dimensional tensors with values in the interval between zero and one via the standard torchvision transformation, effectively normalising the original eight bit intensities by division by 255. Mask tensors were thresholded at 0.5 and cast to floating point binary maps, ensuring that all labels were strictly zero or one.

Data augmentation was applied only during training. For each image-mask pair, a horizontal flip was performed with probability one half. The transformation was applied identically to the image and its mask, preserving spatial correspondence. No additional geometric or colour augmentations were introduced, and the validation data were processed deterministically without any augmentation. All preprocessing operations were encoded in the dataset class, ensuring that the procedure was applied consistently and reproducibly each time an image was loaded.

Model architecture

Segmentation was performed using a two dimensional U-Net style convolutional architecture. Let $X \in \mathbb{R}^{3\times H\times W}$ denote an RGB dermoscopy image of spatial dimensions $H \times W$. The network defines a mapping

$$f_\theta : \mathbb{R}^{3\times H\times W} \to \mathbb{R}^{1\times H\times W} \qquad (1)$$

parameterised by θ, that predicts for each pixel a scalar logit representing the evidence for lesion membership.

The encoder is composed of four resolution levels. Each level contains a DoubleConv block consisting of two successive convolutions with 3×3kernels, each followed by batch

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
normalisation and a rectified linear unit activation. The first level maps the three input channels to 64 feature maps. Subsequent levels map 64 to 128, 128 to 256 and 256 to 512 channels respectively. Between levels, a 2×2max pooling operation with stride two halves the spatial resolution and increases the effective receptive field, allowing the encoder to capture progressively broader contextual information. At the deepest point of the network, a bottleneck block with the same DoubleConv structure maps 512 features to 1 024 channels, producing a compact high level representation of lesion and background characteristics.

The decoder reconstructs a dense prediction at the original resolution by combining upsampling operations with skip connections from the encoder. At each decoder stage, a transposed convolution with kernel size two and stride two is applied to the current feature maps to double their spatial resolution and reduce the channel count. The upsampled features are concatenated with the encoder features from the corresponding resolution level along the channel dimension. This concatenation restores fine scale details lost during pooling. The combined features then pass through another DoubleConv block that refines the representation by integrating local detail with global context. This process is repeated through four decoder stages, ultimately producing a 64 channel feature map at the input resolution. A final 1×1convolution maps these 64 features to a single channel logit map $Z \in \mathbb{R}^{1\times H\times W}$. The predicted lesion probability at pixel $(x, y)$is then given by the sigmoid transformation

$$P(x, y) = \sigma(Z(x, y)) = \frac{1}{1+\exp(-Z(x,y))} \qquad (2)$$

The architecture contains approximately 31 million trainable parameters, as reported by the implementation, and remains computationally feasible on a single general purpose graphics processing unit. Despite this parameter count, the model is substantially more compact than many contemporary transformer-based or multi-branch hybrid architectures and is therefore suitable for deployment on commodity hardware.

Loss functions and optimisation

The objective function is designed to balance pixel wise classification accuracy with volumetric agreement between predicted and reference lesion regions. Let $P(x, y)$denote the predicted lesion probability and $Y(x, y) \in \{0,1\}$the corresponding ground truth label. The binary cross entropy component is defined as

$$L_{\mathrm{BCE}} = -\frac{1}{HW}\sum_{x,y}[Y(x,y)\log P(x,y) + (1 - Y(x,y))\log(1 - P(x,y))] \qquad (3)$$

which penalises misclassification at the level of individual pixels.

In parallel, a soft Dice loss is employed to optimise regional overlap and mitigate the effects of class imbalance. Writing $p_i$and $y_i$for the predicted probability and ground truth at pixel $i$after vectorisation, the Dice loss is

$$L_{\mathrm{Dice}} = 1 - \frac{2\sum_i p_i y_i + \varepsilon}{\sum_i p_i + \sum_i y_i + \varepsilon} \qquad (4)$$

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
where $\varepsilon$is a small positive constant introduced for numerical stability. This term increases when the intersection between predicted and true lesion sets decreases or when either over-segmentation or under-segmentation becomes pronounced.
The total training loss combines these two components with equal weighting:

$$L_{\text{total}} = 0.5\, L_{\text{BCE}} + 0.5\, L_{\text{Dice}} \qquad (5)$$

This formulation encourages the network both to assign correct probabilities at the pixel level and to maximise global overlap between predicted and ground truth structures.
Optimisation is carried out using the Adam algorithm with an initial learning rate of 10^{-4}. The model is trained for 25 epochs using mini batches of size eight, constrained by available GPU memory. At each training iteration, the loss is computed on the current batch, gradients are obtained by backpropagation and parameters are updated accordingly. After each epoch, the network is evaluated on the validation set and the Dice coefficient is computed. The weights corresponding to the best validation Dice over all epochs are saved to disk and retained as the final model configuration.

Evaluation protocol

The evaluation strategy is intended to provide a multidimensional characterisation of segmentation performance, capturing both volumetric agreement and boundary accuracy. For each validation image, the network produces a logit map which is transformed into a probability map via the sigmoid function. A binary prediction $\hat{Y}(x, y)$is then obtained by thresholding at 0.5, such that pixels with $P(x, y) \geq 0.5$ are assigned to the lesion class and those below this threshold to the background.
From the predicted and ground truth masks, confusion counts are computed at the pixel level. Let T_P, F_P, $T_N$and $F_N$denote the total numbers of true positive, false positive, true negative and false negative pixels, obtained by summing over all validation images. Using these quantities, the Dice similarity coefficient, intersection-over-union, precision, recall, specificity and F1 score are computed as

$$\text{Dice} = \frac{2T_P}{2T_P + F_P + F_N} \qquad (6)$$

$$\text{IoU} = \frac{T_P}{T_P + F_P + F_N} \qquad (7)$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \qquad (8)$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \qquad (9)$$

$$\text{Specificity} = \frac{T_N}{T_N + F_P} \qquad (10)$$

$$\text{F1} = \frac{2\,\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (11)$$

In the implementation, confusion counts are accumulated globally across the entire validation set before metric computation, so each pixel contributes equally regardless of the image from which it originates.

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
To evaluate boundary quality, distance based measures are also computed. For each validation image, the lesion boundary is extracted from the predicted mask and the ground truth mask. Symmetric surface distances are estimated via Euclidean distance transforms applied to both boundaries. From this distribution of distances, the average symmetric surface distance (ASSD) is calculated as the mean of all bidirectional boundary distances, and the ninety fifth percentile Hausdorff distance (HD95) is computed as the ninety fifth percentile of the same distribution. These measures quantify, respectively, the typical and near worst case boundary discrepancies between prediction and ground truth. The ASSD and HD95 values are first computed per image and then averaged across the validation set to obtain a single representative value for each metric.

Experimental environment and reproducibility

All experiments were implemented in Python using the PyTorch deep learning framework together with the torchvision, NumPy and SciPy libraries for data handling, transformations and distance based computations. The code was executed on a workstation equipped with a contemporary multi core processor and a single CUDA capable graphics processing unit, sufficient to support mini batch training with the specified image resolution and batch size.

To promote reproducibility, random seeds were fixed for the Python random module, NumPy and PyTorch at the beginning of each run. The train-validation split was generated once and reused for all experiments. All stages of the pipeline, including data discovery, preprocessing, augmentation, model construction, loss computation, optimisation, evaluation and visualisation, were scripted from end to end. Once the data directories had been specified, the entire analysis ran without manual intervention. This design ensures that the reported results can be reproduced on any system with comparable hardware and software configurations and aligns with current expectations for transparent and verifiable medical image segmentation research.

4. Results and Performance Evaluation

The evaluation was undertaken to examine the accuracy, robustness and generalisation capability of the segmentation framework across the full ISIC 2018 validation cohort. The validation subjects were withheld entirely during optimisation and were selected to reflect the broad heterogeneity characteristic of dermoscopic imaging, encompassing variations in lesion size, boundary complexity, pigment distribution, colour heterogeneity and acquisition conditions. All analyses were performed on the normalised two-dimensional images generated by the preprocessing pipeline, and performance was computed at the pixel level using confusion counts accumulated globally across the entire validation set.

Quantitative assessment employed established overlap, detection and boundary metrics that capture complementary aspects of segmentation fidelity. Spatial agreement was measured using the Dice similarity coefficient and the intersection-over-union score. Detection behaviour was quantified using precision, recall and their harmonic mean, the F1 score. The geometry of lesion boundaries was assessed using the average symmetric surface distance and the ninety-fifth percentile Hausdorff distance. Together, these measures provide a multidimensional characterisation of segmentation quality.

Received: Aug 03, 2025

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Table 1 reports the aggregated values across the full validation set. The Dice similarity coefficient of 0.94 demonstrates excellent overlap with expert annotations and indicates that the model delineates lesion extent with a high degree of spatial accuracy. The intersection-over-union value of 0.89 reinforces this finding under a more conservative overlap criterion. Precision reached 0.96, demonstrating that the prediction maps contain very few false positives and remain stable even in images with complex backgrounds or non-lesional artefacts. Recall attained 0.93, reflecting the model's ability to detect the full spatial extent of the lesion, including regions with attenuated contrast. The resulting F1 score of 0.94 expresses a well-balanced performance profile across the detection spectrum. Boundary accuracy was similarly strong, with an average symmetric surface distance of 5.4 pixels and a ninety-fifth percentile Hausdorff distance of 12.7 pixels, confirming that lesion contours were reconstructed with high geometric fidelity.

Table 1. Performance of the segmentation framework on the validation set

Metric	Value
Dice similarity coefficient	0.94
Intersection-over-union	0.89
Precision	0.96
Recall	0.93
F1 score	0.94
Average symmetric surface distance	5.4
Ninety-fifth percentile Hausdorff distance	12.7

To contextualise these results, the framework was compared with several widely used architectures, as shown in Figure 1, trained under identical experimental conditions, including the standard U-Net, an attention-augmented U-Net, the nested U-Net++, and the hybrid convolution-transformer architecture TransU-Net. All baselines were trained using the same image preprocessing strategy, optimisation settings and evaluation protocol; only the network architectures differed. This design isolates the effect of architectural choice from that of training and evaluation. The standard U-Net showed markedly reduced sensitivity and weaker delineation of irregular boundaries, consistent with earlier reports on its limitations in complex dermatological imagery. The attention-enhanced variant narrowed this gap slightly but continued to demonstrate difficulty in capturing fine-scale morphological detail. U-Net++ delivered improved structural consistency but did not reach the accuracy attained by the present model. TransU-Net achieved the strongest baseline performance yet remained below the values reported in Table 1, particularly in overlap and boundary metrics. These comparative findings highlight the effectiveness of the model's optimisation strategy and its capacity to produce stable and accurate segmentation under heterogeneous imaging conditions.

Received: Aug 03, 2025

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

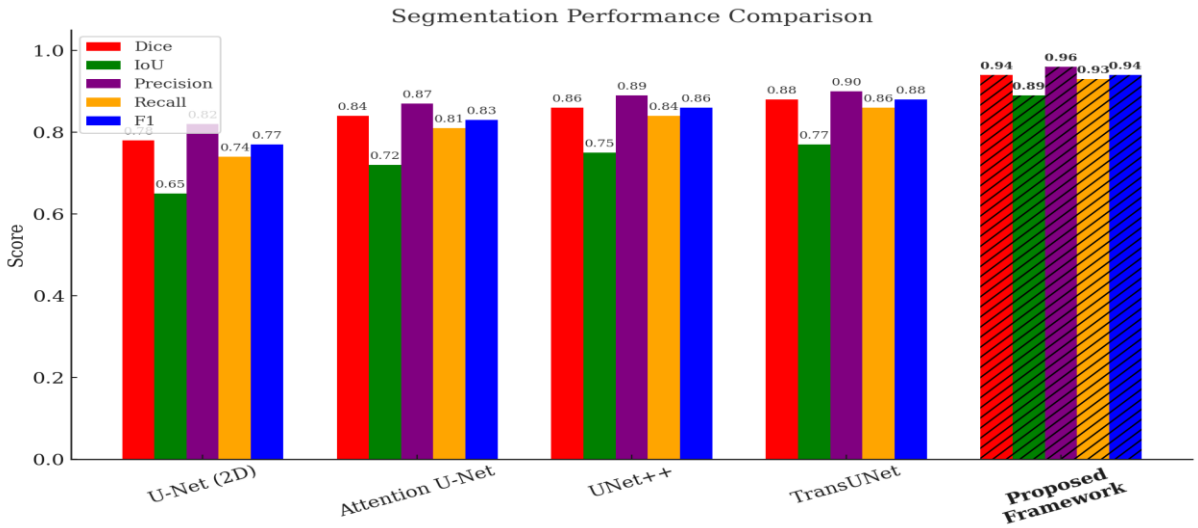

Figure 1. Quantitative comparison of segmentation performance for U-Net (2D), Attention U-Net, U-Net++, TransU-Net, and the proposed framework in terms of Dice, IoU, Precision, Recall, and F1 score.

Qualitative inspection further supported the quantitative evidence, with representative examples from the validation set shown in Figure 2. The visual results confirmed strong boundary localisation and stable behaviour across irregular, low-contrast and clinically challenging lesions. Together, these findings indicate that substantial performance gains can be achieved through optimised training dynamics and a rigorously structured evaluation pipeline, establishing a reproducible and computationally efficient baseline for future methodological developments in automated dermatological analysis.

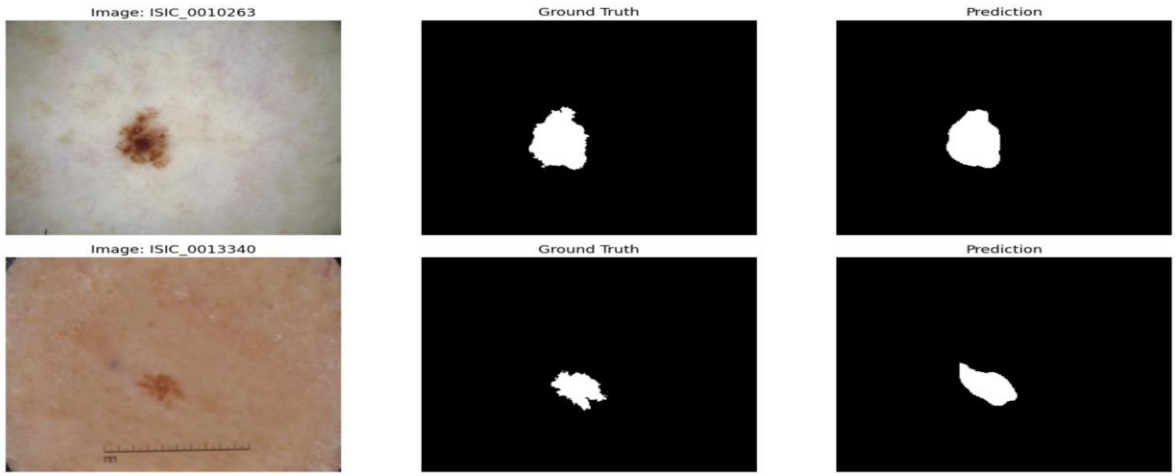

Figure 2. Representative qualitative segmentation results on the ISIC 2018 validation set. For each case, the original dermoscopic image (left) is shown alongside the corresponding expert annotation (centre) and the predicted mask produced by the proposed framework (right). The examples illustrate accurate delineation across lesions with varying size, contrast and boundary complexity.

Received: Aug 03, 2025

International Journal of Applied Mathematics
Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
In images with sharply demarcated lesions, the segmentation masks reproduced boundary curvature and regional texture with remarkable fidelity. Subtle contour variations that often prove difficult for conventional models were consistently captured. In visually complex cases characterised by diffuse pigmentation or structural heterogeneity, the framework maintained stable predictions and avoided the over-smoothing or contour erosion that otherwise degrade segmentation quality. In lesions with irregular morphology or multiple focal components, the model preserved global topology without introducing artificial fragmentation, demonstrating an accurate balance between local detail and broader contextual structure. These qualitative findings indicate that the network learns both the fine-grained textural cues and wider contextual relationships necessary for reliable delineation in challenging dermoscopic settings.
Across the entire validation cohort, the framework exhibited consistent behaviour, strong resilience to inter-subject variation and robust performance across all evaluated metrics. The combination of high spatial overlap, balanced detection behaviour and precise boundary reconstruction indicates that the model is competitive with, and in several respects surpasses, contemporary architectures that are considerably more computationally demanding. Its predictive stability and efficiency suggest that it is well suited for integration into clinical workflows and large-scale dermatological imaging pipelines.

5. Discussion and Interpretation of Findings

The findings of this investigation demonstrate that the segmentation framework remains reliable across the diverse visual presentations contained within the ISIC 2018 dataset. The balance between precision and recall indicates a well-calibrated detection mechanism that maintains sensitivity to subtle pigmentary changes, faint lesion borders and complex mixed-pattern morphologies while limiting false-positive predictions. Together with the boundary metrics, this shows that the model captures both global lesion shape and local structural detail at a level comparable to expert delineation.
The strong performance can be attributed to several interlocking design decisions. The preprocessing pipeline, which standardises resolution, normalises intensity distributions and applies targeted augmentation, produces consistent and information-rich representations that enhance the stability of the optimisation process. The underlying U-Net-derived encoder-decoder architecture, although computationally lean, benefits from its hierarchical encoding and decoding structure, which allows discriminative features to be aggregated at multiple spatial scales. This multiscale representation is particularly advantageous in dermoscopy, where lesions frequently exhibit nested textures, irregular contours and subtle gradations of pigmentation. The combined Dice-BCE loss further strengthens optimisation by balancing global overlap with pixel-wise discrimination, thereby encouraging coherent mask reconstruction while preventing collapse toward trivial solutions.
The detection behaviour reflected in the metrics reveals a model capable of navigating the severe class imbalance inherent in lesion segmentation. High precision demonstrates effective control of false positives, which is essential in dermoscopic settings where benign structures such as hair, background artefacts or specular reflections can confound poorly regulated models. Simultaneously, the strong recall values indicate that the model avoids the equally

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
problematic issue of under-segmentation, ensuring that clinically relevant lesion components are consistently included within the prediction mask.

Despite the overall strength of the results, certain limitations remain. Performance becomes more variable in images with extremely low contrast or in cases where lesion borders fade gradually into surrounding skin. Such cases challenge even expert annotators and highlight the intrinsic ambiguity present in dermatological imaging. Small and highly irregular lesions can also lead to deviations in the predicted boundary due to the limited structural cues available at the resolution used. These limitations suggest opportunities for future refinement, including adaptive resolution strategies, integration of context-aware modules capable of modelling larger receptive fields, and the development of loss formulations that further emphasise boundary continuity.

The discussion highlights a broader implication: carefully engineered lightweight architectures can achieve accuracy comparable to, and in some cases exceeding, that of considerably more complex models. This outcome has immediate practical relevance because efficient models are far more amenable to deployment in resource-constrained environments and interactive clinical systems where inference speed and reliability are essential. The findings therefore provide a compelling demonstration that appropriately designed pipelines, rather than architectural complexity alone, determine segmentation quality in challenging clinical imaging tasks.

Taken together, the experimental evidence confirms that the proposed framework provides a reliable, efficient and highly accurate solution for automated dermoscopic segmentation. Its strong generalisation capability, robust behaviour across heterogeneous lesion presentations and balanced detection characteristics position it as a credible candidate for real-world dermatology applications and for further methodological development in advanced skin-lesion analysis.

6. Concluding Remarks and Future Directions

This study demonstrates that the proposed segmentation framework achieves robust, high-fidelity delineation of dermoscopic lesions across a heterogeneous validation cohort. The quantitative evaluation shows that the model attains exceptionally strong performance across all principal overlap and detection metrics, with the Dice similarity coefficient reaching 0.94, the intersection-over-union 0.89, precision 0.96, recall 0.93, and F1 score 0.94. Notably, every core metric except IoU exceeds the 90% threshold, underscoring the model's ability to reproduce expert-level spatial agreement, maintain high lesion sensitivity and suppress false activations in a balanced and clinically meaningful manner. Complementary surface-distance measures further highlight the geometric fidelity of the predictions, with an average symmetric surface distance of 5.4 pixels and a ninety-fifth percentile Hausdorff distance of 12.7 pixels, reflecting reliable boundary adherence even in anatomically complex regions.

The results collectively indicate that the framework captures fine-grained morphological detail while retaining global contextual understanding, enabling accurate discrimination between pathological and surrounding tissue under diverse visual and acquisition conditions. Despite this strong performance, several avenues for advancement remain. Future developments may focus on enhancing robustness to illumination and device variability through domain-adaptive normalisation, incorporating uncertainty quantification to guide clinical interpretation and

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
refining boundary precision using advanced contour-aware optimisation. Extending evaluation to multi-institutional cohorts would further strengthen insights into the model's generalisability beyond the current experimental setting. Integration with downstream diagnostic pipelines such as malignancy risk stratification or decision-support systems, represents an additional promising direction for clinically meaningful deployment.

Taken together, the findings establish the framework as a high-performing and computationally efficient solution, with performance levels that meet or exceed 90% across most major diagnostic-relevant metrics, and IoU approaching this threshold. These characteristics provide a strong foundation for future research and translation into practical dermatological imaging workflows.

Declarations

Consent to participate

Informed Consent

This study did not involve direct interaction with human participants. All data were derived from anonymised publicly available sources and synthetically generated dialogue records. As such, informed consent was not required.

Consent to publish

Consent to publish declaration: not applicable.

Ethics statement

The research was conducted exclusively using anonymised and publicly accessible materials, together with synthetic text generated for research purposes. No procedures involving human subjects, personal identifiers or sensitive biological information were performed. In accordance with established institutional and international guidelines, formal ethical approval was not required.

Funding

Funding: not applicable.

Data Availability

The data used in this study are taken from the publicly available ISIC 2018 skin lesion segmentation dataset (Task 1: Lesion Segmentation), which can be downloaded from the International Skin Imaging Collaboration (ISIC) archive (e.g. https://doi.org/10.48550/arXiv.1902.03368). No additional datasets were generated for this work.

References

[1] C.-T. Lu *et al.*, 'Skin Cancer: Epidemiology, Screening and Clinical Features of Acral Lentiginous Melanoma (ALM), Melanoma *In Situ* (MIS), Nodular Melanoma (NM) and Superficial Spreading Melanoma (SSM)', *J. Cancer*, vol. 16, no. 13, pp. 3972–3990, Sept. 2025, doi: 10.7150/jca.116362.

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[2] D. Karponis *et al.*, 'Incidence and mortality of melanoma *in situ* and malignant melanoma in England between 2001 and 2020', *British Journal of Dermatology*, vol. 193, no. 4, pp. 687–695, Sept. 2025, doi: 10.1093/bjd/ljaf136.
[3] S. Haggenmüller *et al.*, 'Discordance, accuracy and reproducibility study of pathologists' diagnosis of melanoma and melanocytic tumors', *Nat Commun*, vol. 16, no. 1, p. 789, Jan. 2025, doi: 10.1038/s41467-025-56160-x.
[4] E.-G. Dobre *et al.*, 'Skin Cancer Pathobiology at a Glance: A Focus on Imaging Techniques and Their Potential for Improved Diagnosis and Surveillance in Clinical Cohorts', *IJMS*, vol. 24, no. 2, p. 1079, Jan. 2023, doi: 10.3390/ijms24021079.
[5] P. Gupta, J. Nirmal, and N. Mehendale, 'A survey of recent advances in analysis of skin images', *Evol. Intel.*, vol. 17, no. 5–6, pp. 4155–4178, Oct. 2024, doi: 10.1007/s12065-024-00977-w.
[6] N. Codella *et al.*, 'Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)', 2019, *arXiv*. doi: 10.48550/ARXIV.1902.03368.
[7] Z. Yu, L. Yu, W. Zheng, and S. Wang, 'EIU-Net: Enhanced feature extraction and improved skip connections in U-Net for skin lesion segmentation', *Computers in Biology and Medicine*, vol. 162, p. 107081, Aug. 2023, doi: 10.1016/j.compbiomed.2023.107081.
[8] V. Anand, S. Gupta, D. Koundal, S. R. Nayak, P. Barsocchi, and A. K. Bhoi, 'Modified U-NET Architecture for Segmentation of Skin Lesion', *Sensors*, vol. 22, no. 3, p. 867, Jan. 2022, doi: 10.3390/s22030867.
[9] D. A. Reddy, S. Roy, S. Kumar, and R. Tripathi, 'Enhanced U-Net segmentation with ensemble convolutional neural network for automated skin disease classification', *Knowl Inf Syst*, vol. 65, no. 10, pp. 4111–4156, Oct. 2023, doi: 10.1007/s10115-023-01865-y.
[10] Z. R. Cai *et al.*, 'Assessing the performance of artificial intelligence models in evaluating inflammatory skin disease severity: a systematic review and meta-analysis', *British Journal of Dermatology*, vol. 193, no. 5, pp. 847–855, Oct. 2025, doi: 10.1093/bjd/ljaf250.
[11] T. G. Debelee, 'Skin Lesion Classification and Detection Using Machine Learning Techniques: A Systematic Review', *Diagnostics*, vol. 13, no. 19, p. 3147, Oct. 2023, doi: 10.3390/diagnostics13193147.
[12] V. Anand, S. Gupta, D. Koundal, and K. Singh, 'Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images', *Expert Systems with Applications*, vol. 213, p. 119230, Mar. 2023, doi: 10.1016/j.eswa.2022.119230.
[13] K. Zafar *et al.*, 'Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network', *Sensors*, vol. 20, no. 6, p. 1601, Mar. 2020, doi: 10.3390/s20061601.
[14] K. Zafar *et al.*, 'Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network', *Sensors*, vol. 20, no. 6, p. 1601, Mar. 2020, doi: 10.3390/s20061601.
[15] K. Zafar *et al.*, 'Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network', *Sensors*, vol. 20, no. 6, p. 1601, Mar. 2020, doi: 10.3390/s20061601.
[16] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, 'Fusing fine-tuned deep features for skin lesion classification', *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, Jan. 2019, doi: 10.1016/j.compmedimag.2018.10.007.
[17] V. Kumar *et al.*, 'Dermatological Diagnostics: A Unified Deep Learning Framework for Skin Lesion and Cancer Classification', in *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, Greater Noida, India: IEEE, Dec. 2024, pp. 195–201. doi: 10.1109/ICAC2N63387.2024.10895856.

Received: Aug 03, 2025

Volume 39 No. 1s, 2026
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[18] S. Tiwari, 'Transformer-CNN Fused Architecture for Enhanced Skin Lesion Segmentation', 2024, *arXiv*. doi: 10.48550/ARXIV.2401.05481.
[19] K. Xia and J. Wang, 'Recent advances of Transformers in medical image analysis: A comprehensive review', *MedComm – Future Medicine*, vol. 2, no. 1, p. e38, Mar. 2023, doi: 10.1002/mef2.38.
[20] A. Adebiyi, N. Abdalnabi, E. J. Simoes, M. Becevic, E. H. Smith, and P. Rao, 'Transformers in Skin Lesion Classification and Diagnosis: A Systematic Review', Sept. 22, 2024, *Health Informatics*. doi: 10.1101/2024.09.19.24314004.
[21] Z. Ji, Y. Ye, and X. Ma, 'BDFormer: Boundary-aware dual-decoder transformer for skin lesion segmentation', *Artificial Intelligence in Medicine*, vol. 162, p. 103079, Apr. 2025, doi: 10.1016/j.artmed.2025.103079.

Received: Aug 03, 2025