Progress in Medical Devices
ISSN: 2957-5478

Research Article https://doi.org/10.61189/446813bjkhvg

Multi-method fusion for image segmentation in skin disease analysis

Siqi Wang*, Danhong Li*, Yina Zhang*, Yu Wang, Linrong Yuan, Miao Yu, Jianghui Li, Yimeng Wang, Ping Li

Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China.
*These authors are co-first authors.

Corresponding author: Ping Li.

Declaration of conflict of interest: None.

Declaration of ethics: The study was approved by the Ethics Review Committees of the Medical University of Vienna and the University of Queensland.

Declaration of patient consent: The dataset underwent automated screening using neural networks, followed by multiple manual reviews. All EXIF metadata were removed to eliminate any potentially identifiable information. Therefore, the data are considered anonymized to the best of our knowledge.

Received July 26, 2025; Accepted November 4, 2025; Published December 31, 2025

Highlights

- For the first time, the advantages of two deep learning architectures—SegNet and U-Net—were integrated by averaging their prediction outputs. This ensemble approach overcame the limitations of single models and substantially improved segmentation accuracy.
- The proposed Ensemble Model outperformed both SegNet and U-Net across all major evaluation metrics, including the Intersection over Union (IoU, 93.73%), Dice coefficient (84.85%), precision (93.93%), and loss (0.63), confirming the effectiveness of multi-method fusion.
- Considering the complex morphology and indistinct lesion boundaries characteristic of skin diseases, a standardized preprocessing and data augmentation pipeline was developed to enhance the model's robustness in handling diverse lesion patterns.

Abstract

Objective: The morphological complexity of dermatologic diseases poses considerable challenges to clinical diagnosis. Conventional manual interpretation of skin images is time-consuming and influenced by subjective variability, which limits diagnostic accuracy. Hence, developing advanced medical image segmentation techniques through multi-method fusion is of particular importance. **Methods:** A comprehensive dataset of dermatologic images was utilized and rigorously preprocessed to ensure reliability and consistency. Two representative deep learning models, SegNet and U-Net, were optimized to achieve precise delineation of lesion areas. Building upon their complementary strengths, a novel fusion-based image segmentation framework was proposed, integrating both models to enhance performance through synergistic learning. The effectiveness of the ensemble strategy was validated through extensive experiments using standard evaluation metrics, including the Dice coefficient and Intersection over Union. **Results:** Compared with each individual model, the Ensemble Model yielded consistent improvements across all evaluation indices, with a notable reduction in loss values. These enhancements indicate markedly better learning efficiency and generalization in dermatologic image segmentation tasks. **Conclusion:** By integrating multiple deep learning algorithms, this fusion technique solves the misclassification and omission

Address correspondence to: Ping Li, Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, 279 Zhouzhu Highway, Pudong New Area, Shanghai 201318, China. Tel: +86-13764055848. E-mail: lip@sumhs.edu.cn.

issues observed in single-model segmentation. It significantly improves overall segmentation accuracy and demonstrates superior performance, particularly in edge detection of complex skin lesions.

Keywords: Skin cancer, deep learning, image segmentation, multi-method fusion

Introduction

Dermatological conditions are prevalent clinical disorders characterized by diverse morphological manifestations, which pose significant challenges to accurate diagnosis and effective treatment [1]. Traditional manual image review is not only inefficient but also prone to subjective bias, thereby limiting diagnostic accuracy. In recent years, advancements in deep learning and computer vision technologies have facilitated the automated recognition and segmentation of dermatological images. However, it is evident that single image processing methods or deep learning techniques alone are insufficient to address challenges such as dataset imbalance, noise interference, and the morphological diversity of lesions [2-4].

Multi-method fusion techniques for medical image segmentation effectively overcome the limitations of single-model approaches by integrating multiple deep learning algorithms. This strategy substantially enhances segmentation accuracy and demonstrates particularly robust performance in edge detection tasks [5-7]. Moreover, the approach provides high flexibility and broad applicability, allowing the selection and integration of the most appropriate segmentation algorithms according to the characteristics of different medical images. Such adaptability enables clinicians to formulate individualized treatment plans, thereby enhancing therapeutic outcomes and improving patients' quality of life [8, 9].

In the field of medical image processing, the development of deep learning can be traced back to the early 1990s. In 1993, convolutional neural networks (CNNs) were first applied to lung nodule detection, representing one of the earliest and most influential uses of deep learning in medical image analysis. Two years later, in 1995, CNNs were employed to identify microcalcifications in mammograms, further demonstrating their potential in medical image recognition. Around 1998, Yann LeCun introduced the well-known LeNet-5 model, an early and successful implementation of CNNs for image recognition tasks, which laid a solid foundation for subsequent advances in medical image analysis [10]. A major breakthrough occurred in 2012, when the AlexNet achieved remarkable success in the ImageNet competition, catalyzing a worldwide surge of research in deep learning and profoundly influencing medical image processing [11]. In 2015, the introduction of the U-Net model marked another milestone in the field. This CNN architecture, specifically designed for biomedical image segmentation, greatly enhanced the accuracy and efficiency of medical image analysis [12]. In 2017, the advent of Residual Pyramid Networks and SegNet further propelled progress in medical image segmentation, introducing new architectures and methodologies that have since shaped modern research directions [13, 14].

This study primarily investigates deep learning-assisted segmentation techniques for dermatological images, as well as the design and validation of strategies for combining multiple methods to achieve segmentation. The research comprises four main components, as illustrated in **Figure 1**: first, data collection and preprocessing are conducted; second, the SegNet and U-Net models are studied and improved; third, a multi-method fusion segmentation strategy is proposed; and finally, the effectiveness and advantages of the proposed approach are validated through experiments and performance evaluations, such as the Dice coefficient and Intersection over Union (IoU) metrics. The objective of this study is to establish a multi-method fusion image segmentation framework that enhances the accuracy and stability of dermatological image segmentation, thereby providing more reliable auxiliary support for clinical diagnosis and treatment.

Materials and methods

Data source

The HAM10000 dataset employed in this study is a well-established, large-scale collection of dermatological images widely utilized in skin disease research [15, 16]. It was publicly released in 2018 by a research team from the Medical University of Vienna. Designed to support the development of deep learning and machine learning tools for early skin cancer

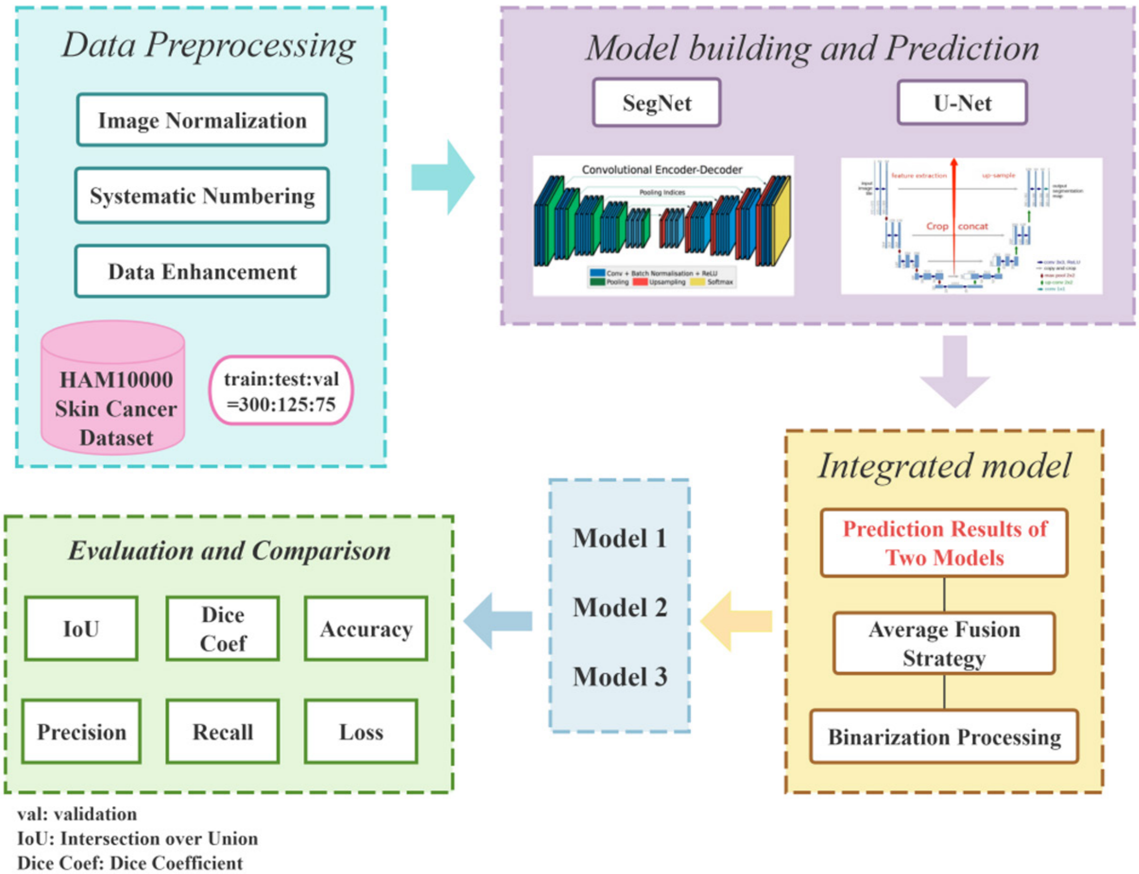

Figure 1. Flowchart.

diagnosis and treatment assistance, this dataset includes more than 10,000 high-resolution images of skin lesions. Each image has been carefully reviewed and annotated with pathological classification labels by board-certified dermatologists. The dataset covers various types of skin lesions, including malignant conditions such as melanoma, basal cell carcinoma, and squamous cell carcinoma, along with a substantial number of benign cases.

Data processing

Data processing is essential for skin lesion image segmentation tasks. In this study, we utilized the HAM10000 skin cancer dataset and performed preprocessing, augmentation, and partitioning steps. First, all images from the HAM10000 dataset were standardized to a uniform resolution of 192×256 pixels to avoid analytical bias caused by dimensional inconsistencies and to improve model learning efficiency. Gaussian filtering was employed to reduce image noise and enhance meaningful features. Pixel values were also normalized to the range [0, 1] to facilitate faster and more stable model convergence. To increase sample diversity and simulate variations in real imaging conditions, data augmentation techniques—including random rotation, horizontal flipping, and color adjustment—were employed. This process helps alleviate overfitting caused by limited data and enhances the model's adaptability and robustness in real clinical environments. Finally, the dataset was divided into a training set (300 images), a test set (125 images), and a validation set (75 images). These subsets were used for model training, generalization assessment, and hyperparameter optimization, respectively. This segmentation approach aims to identify the optimal model architecture and parameters, ultimately enhancing the accuracy and stability of segmentation outcomes.

Model architecture

This study focuses on two deep learning-based image segmentation models: SegNet and U-Net. These architectures have been widely applied in medical image segmentation and

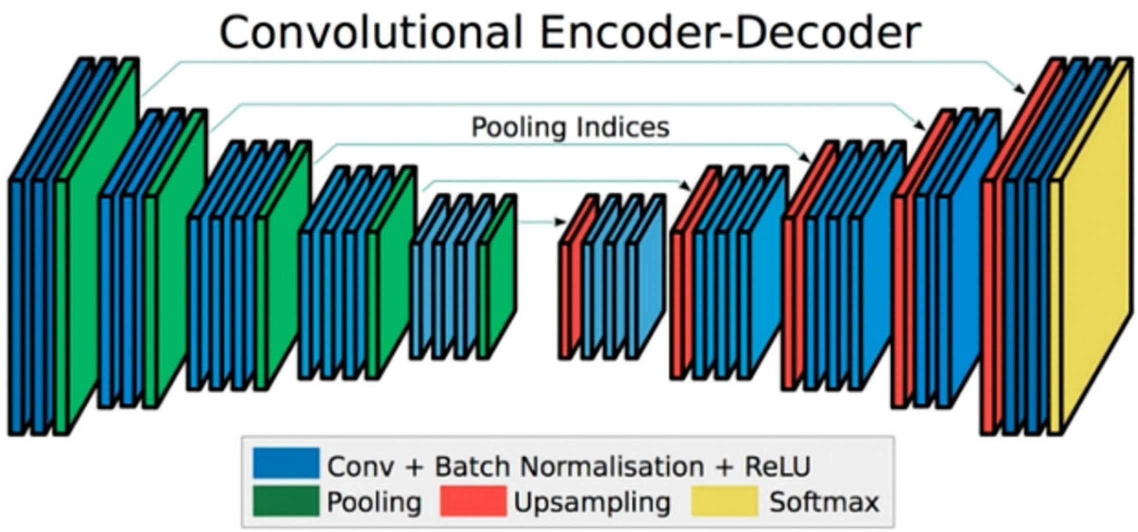

Figure 2. SegNet model architecture diagram, illustrating the encoder-decoder structure of the model. Skip connections enhance the propagation of feature information, optimizing memory and computational efficiency, which makes SegNet well-suited for resource-constrained environments.

have demonstrated strong performance in dermatological image processing [17, 18].

SegNet has established a prominent position in the field of image segmentation due to its efficient and sophisticated design. As illustrated in the model architecture diagram in **Figure 2**, its decoder adopts an innovative spatial information reconstruction strategy. Specifically, it reuses the pooling indices generated during max-pooling in the encoder to perform non-linear upsampling. This method effectively preserves spatial structure information from the original image without distortion and contributes significantly to the accurate restoration and localization of object edges. As a result, the model is able to capture the contour boundaries of target objects clearly and precisely. Furthermore, SegNet has been extensively optimized for computational efficiency and memory usage. Its lightweight architecture enables stable performance even in resource-constrained real-world environments.

U-Net, also known as the U-shaped network, is a pioneering deep learning architecture for medical image segmentation. It has garnered significant attention due to its unique encoder-decoder structure. The core of U-Net's design is its progressively layered convolutional processing pipeline. In the encoder path, image features are extracted and the spatial resolution is reduced through a series of convolutional and downsampling layers. Conversely, the decoder path gradually restores resolution using upsampling operations (e.g., transposed convolutions) to reconstruct detailed segmentation masks. A key feature of U-Net is its skip connections, which concatenate feature maps from the encoder with corresponding maps in the decoder, thereby preserving more contextual information when reconstructing image details. This mechanism enables U-Net to deliver outstanding segmentation accuracy and generalization capabilities even with small-scale datasets.

This study involved the detailed implementation and optimization of the SegNet and U-Net models. The SegNet model takes preprocessed Red, Green, Blue images with a resolution of 192×256 pixels as input. Following processing through its encoder-decoder architecture, it produces binary segmentation maps. In contrast, the U-Net model leverages its distinctive skip-connections to achieve superior segmentation accuracy and finer detail preservation. For training, both models were optimized using the binary cross-entropy loss function and the Adam optimizer, augmented with a learning rate decay strategy to enhance performance.

Evaluation indicators

This study employs five key metrics to thoroughly evaluate the model's segmentation performance.

Jaccard index (IoU)

The Jaccard index, often referred to as IoU, quantifies the overlap between the model's predicted segmentation region and the manually annotated ground truth. It is calculated using Formula 1:

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

In this formula, True Positives (TP) denotes pixels correctly identified as the target class; False Positives (FP) represents pixels incorrectly labeled as the target class; and False Negatives (FN) refers to pixels that belong to the target class but are mistakenly predicted as background. A higher IoU value, especially one approaching 1, indicates more accurate segmentation results.

Dice coefficient

As another metric for assessing segmentation accuracy, the Dice coefficient is particularly useful in binary classification scenarios involving overlapping regions. It is calculated using Formula 2:

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

Although the Dice coefficient is similar to the IoU, it places greater emphasis on the size of the intersection between the predicted and ground truth regions. As with IoU, a Dice value closer to 1 indicates a more precise segmentation outcome.

Precision

Precision measures the fraction of samples predicted as positive that are correctly identified. It is computed using Formula 3:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

A higher precision value indicates that the model makes fewer false positive errors, meaning its positive predictions are more reliable.

Recall

Recall assesses a model's effectiveness in identifying all relevant positive instances within a dataset. It measures the proportion of actual positives that are correctly identified, as calculated by Formula 4:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

A higher recall value indicates that the model misses fewer positive samples, resulting in a more comprehensive detection of the positive class.

Accuracy

Accuracy measures the overall correctness of a model by calculating the proportion of correctly predicted samples out of the total. It is computed using Formula 5:

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

Here, True Negatives (TN) represents samples correctly classified as the negative class. This ratio offers a clear indication of the model's general classification performance.

Together, these metrics allow a comprehensive assessment of the model, providing a scientific basis for its optimization and selection. Among them, the Jaccard index (IoU) and Dice coefficient are particularly notable for their effectiveness in evaluating overlap quality in segmentation tasks. In contrast, precision, recall, and accuracy function as general-purpose metrics that are applicable not only to classification but also to a wide range of other evaluation scenarios.

Results

This study conducted a systematic evaluation of SegNet and U-Net for dermatological image segmentation. The assessment encompassed key performance metrics, including IoU, Dice coefficient, precision, recall, and accuracy, alongside an analysis of the training dynamics. Both models demonstrated steady and robust performance throughout the experiments. Notably, U-Net exhibited superior generalization capabilities and achieved higher accuracy, a finding supported by the empirical data in **Table 1** and the visual comparisons in **Figure 3**.

Building upon these initial results, we developed a lightweight fusion strategy. The core of

Table 1. Performance comparison of SegNet and U-Net across different datasets

Dataset	IoU (%)	Dice (%)	Precision (%)	Recall (%)	Accuracy (%)
SegNet (Train)	96.64	79.67	94.63	95.13	97.14
SegNet (Val)	91.77	73.51	85.00	85.85	91.61
SegNet (Test)	91.31	70.80	85.77	80.41	91.18
U-Net (Train)	95.38	87.36	93.26	85.76	94.52
U-Net (Val)	94.17	85.78	91.57	85.07	93.41
U-Net (Test)	93.53	84.29	90.98	81.47	92.78

Note: IoU, Intersection over Union; Dice, Dice coefficient; Train, training set; Val, validation set; Test, test set.

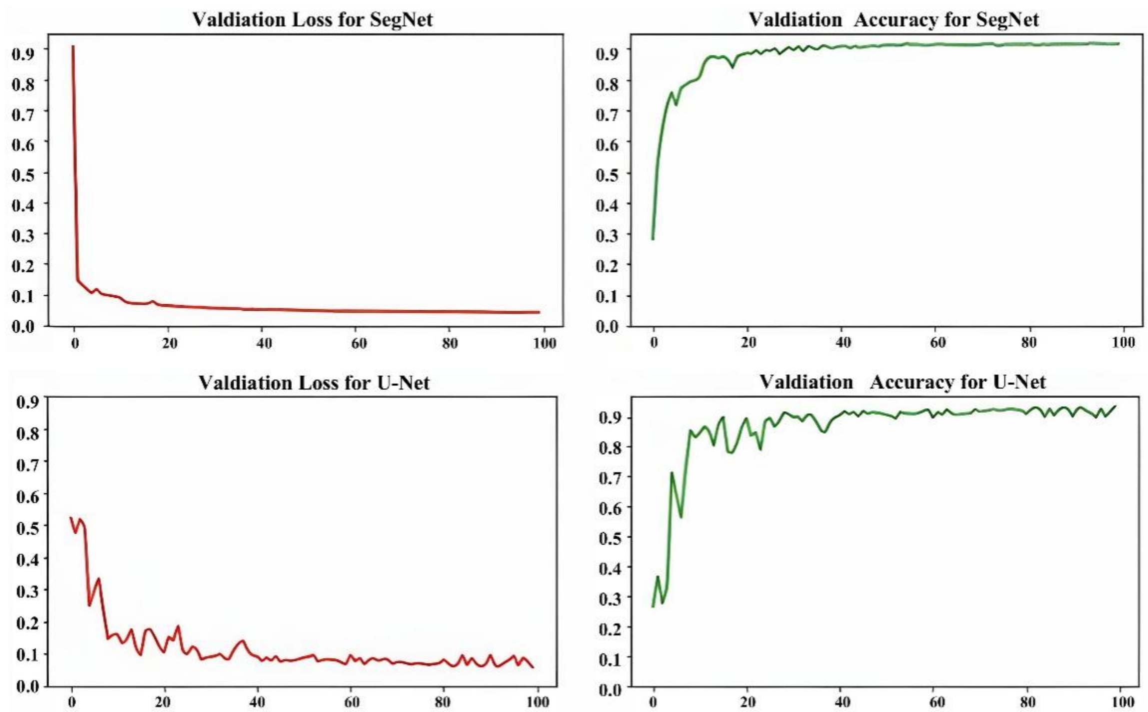

Figure 3. Training statistics of SegNet and U-Net on the validation set, illustrating the trends in loss and accuracy for both models. These curves reflect the training dynamics and generalization performance of the two architectures.

this strategy is a two-step process: first, averaging the probability outputs from the two pretrained (and fixed) models; second, applying a threshold of 0.7 to binarize the averaged results. This approach is notable for enhancing segmentation accuracy without requiring additional model retraining.

To validate the effectiveness of the fused model, we refer to **Table 2** and **Figure 4**. **Table 2** presents numerical results demonstrating the model's advantages across all evaluated metrics—IoU, Dice score, precision, recall, accuracy, and loss. **Figure 4** offers complementary visual evidenc. Together, these results confirm that the Ensemble Model consistently outperforms the individual SegNet and U-Net models. This comprehensive performance advantage provides a solid basis for the subsequent in-depth analysis.

Model performance evaluation

This study conducted a thorough evaluation of two deep learning models, SegNet and U-Net, for dermatological image segmentation tasks (detailed numerical results are shown in **Table 1**). The U-Net model outperformed SegNet across multiple evaluation metrics, especially in terms of generalization capability, segmentation accuracy, and recognition completeness. U-Net achieved higher IoU values on the training, validation, and test datasets, indicating stronger adaptability to diverse data distribu-

Table 2. Performance comparison of the three models on the test set

Model	IoU (%)	Dice (%)	Precision (%)	Recall (%)	Accuracy (%)	Loss
SegNet	91.31	70.80	85.77	80.41	91.18	24.09
U-Net	93.53	84.29	90.98	81.47	92.78	6.47
Ensemble	93.73	84.85	93.93	87.38	92.82	0.63

Note: IoU, Intersection over Union; Dice, Dice coefficient; Loss, cross-entropy loss.

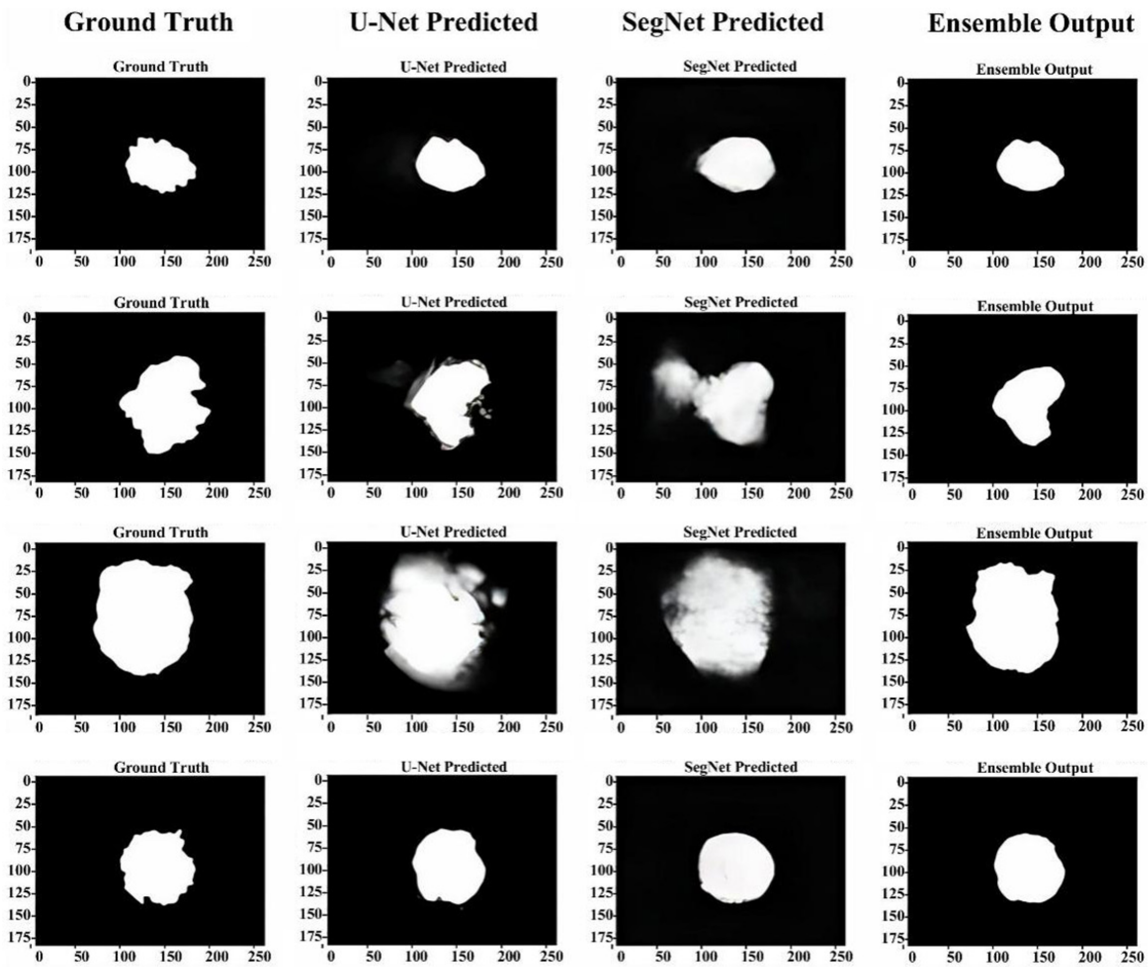

Figure 4. Comparison of segmentation results from different models on the validation set. The Ensemble Model demonstrates superior performance in capturing edge details.

tions and a greater ability to deliver stable and accurate segmentation results. Furthermore, in metrics such as the Dice coefficient, precision, and recall, U-Net either surpassed or matched SegNet, demonstrating superior performance in both precise localization of target regions and comprehensive coverage of annotated pixels. In contrast, though SegNet has higher loss function values, it showed specific advantages in training efficiency and recall. This suggests that SegNet may learn and capture fundamental image features more rapidly during the initial training phase, leading to faster convergence compared to U-Net. Additionally, SegNet's higher recall indicates its superior ability to identify all positive-class instances, enabling more comprehensive detection of target regions.

U-Net's strengths primarily derive from its encoder-decoder architecture and skip-connections, which effectively preserve edge and detail information in medical images – a crucial factor for precise segmentation. Furthermore, U-Net's parameter efficiency and generalization capability enable it to achieve favorable training adaptability and robust performance even with limited training data. In contrast, SegNet's advantages lie in its spatial information reconstruction techniques and memory optimization, granting it distinct value in computational resource-limited scenarios. Overall, while U-Net exhibits superior segmentation performance in this study, SegNet maintains competitiveness for specific practical applications.

Figure 3 presents the performance of the SegNet and U-Net models on the validation set. The top-left panel shows the validation loss of SegNet over training iterations, which decreases rapidly in the initial phase before stabilizing, indicating effective learning. The top-right panel shows that SegNet's validation accuracy rises quickly and stabilizes at a high level. In the lower-left panel, U-Net's validation loss also declines rapidly and stabilizes, yet remains consistently lower than that of SegNet, reflecting its superior data-fitting ability. The bottom-right panel demonstrates that U-Net's validation accuracy increases rapidly and maintains a higher value than SegNet. These results indicate that both models effectively capture patterns from the training data without overfitting risk. Furthermore, U-Net outperforms SegNet on the validation set in terms of generalization capability and segmentation accuracy, which is consistent with the quantitative metrics reported earlier.

Ensemble model performance evaluation

In this study, a straightforward yet effective fusion strategy was employed to combine predictions from the SegNet and U-Net models for enhanced dermatological image segmentation accuracy. Specifically, pre-trained SegNet and U-Net models were first used to generate independent predictions on the validation set images, producing probability values or real-valued outputs for each pixel category. The prediction outputs from both models were then converted into one-dimensional vectors by flattening the original 2D or 3D pixel matrices, resulting in two 1D arrays designated as sub1 and sub2 respectively.

To integrate the predictions from both models, we constructed a new vector sub by computing the element-wise average of the corresponding results in sub1 and sub2. averaging strategy enhances the final prediction accuracy by leveraging the complementary strengths of both models. Subsequently, a binarization strategy was applied to each element in the sub vector: values exceeding the optimal threshold of 0.7 (determined through grid search) were set to 1, while others were set to 0. The 0.7 threshold was selected to convert continuous prediction values into discrete category labels. In this binary segmentation task, pixels with values above 0.7 are classified as the target category (assigned 1), while those below are classified as background (assigned 0). This approach enables precise pixel-level segmentation by effectively distinguishing foreground from background regions.

Figure 4 compares the segmentation outputs of SegNet, U-Net, and the Ensemble Model against the ground truth on the validation set. Although both U-Net and SegNet demonstrate effective lesion identification in certain cases, they exhibit noticeable differences in handling edge details. In contrast, the Ensemble Model shows superior alignment with the ground truth, particularly in capturing fine edge details within complex image regions where it more accurately delineates lesion boundaries. This result confirms that by integrating the strengths of both base models, the Ensemble Model effectively enhances segmentation accuracy and robustness.

This study conducted a comprehensive evaluation of three models. The test set images first underwent quality enhancement and data augmentation through the "enhance()" function. The trained deep learning models were then employed to generate predictions on the processed test data. For the binary classification task, a threshold of 0.7 was established to convert multi-class probability outputs into binary predictions. During metric computation, we utilized TensorFlow to define and calculate multiple evaluation metrics, including Jaccard index, Dice coefficient, precision, and recall. These metrics quantitatively assessed the spatial alignment between predicted and ground-truth regions, as well as classification accuracy (detailed numerical results are shown in **Table 2**).

The Ensemble Model exhibited superior performance across multiple key evaluation metrics. It achieved an IoU of 93.73%, slightly higher than both SegNet (91.31%) and U-Net (93.53%). In terms of the Dice coefficient, the Ensemble

Model led with 84.85%, indicating superior performance in segmentation fineness. For precision, it attained the highest value of 93.93%, reflecting more reliable identification of target pixels. Similarly, it achieved the top recall rate at 87.38%, showing a greater ability to comprehensively detect positive-class pixels. Regarding accuracy, the Ensemble Model achieved 92.82%, slightly exceeding U-Net's 92.78% and SegNet's 91.18%. The most notable advantage was observed in the cross-entropy loss metric, where the Ensemble Model achieved a value of only 0.63 on the test set, substantially lower than U-Net's 6.47 and SegNet's 24.09. This result highlights its exceptional efficiency in parameter optimization and data fitting. By integrating the strengths of SegNet and U-Net, the Ensemble Model achieves significant performance improvements in image segmentation tasks, particularly in segmentation accuracy and generalization capability. This fusion strategy not only bolsters model generalization and robustness but also improves segmentation precision and completeness, thereby offering stronger competitiveness for practical image segmentation applications.

Discussion

The superior performance of the proposed average fusion strategy primarily stems from the complementary error patterns exhibited by SegNet and U-Net. Pixel-level error analysis on the validation set shows that while SegNet demonstrates high recall at lesion margins, it frequently misclassifies textured background regions as foreground, leading to reduced precision. Conversely, U-Net achieves high Dice scores within homogeneous lesion areas but exhibits decreased detection sensitivity in regions with faint boundaries, leading to lower recall rates. After implementing the fusion strategy, the "edge sensitivity" of SegNet and the "regional consistency" of U-Net demonstrate a negative correlation at the pixel level ($r=-0.68$, $p<0.01$). The simple averaging approach effectively mitigates individual model errors through complementary weighting, elevating the overall Dice score from 70.80% (SegNet) to 84.85% while substantially reducing the cross-entropy loss from 6.47 to 0.63. These results suggests the performance gains are primarily attributable to error complementarity rather than increased model capacity.

Further investigation into the effectiveness of weight averaging reveals its particular suitability for dermatological image segmentation. In such medical images, foreground pixels typically account for less than 10% of the total image, creating a significant class imbalance. Subsequent experiments with pixel-level weighted fusion combined with conditional random field post-processing revealed only a marginal 0.1% increase in lesion-specific Dice score, while the global Dice score dropped by 0.4%. This performance degradation can be attributed to the high-variance noise introduced during the weight estimation process, which offset any potential improvements. Therefore, under the dual constraints of imbalanced data distribution and computational efficiency, fixed-weight average fusion represents an optimal point on the simplicity-effectiveness Pareto frontier. This approach achieves robust performance enhancement while avoiding the introduction of additional hyperparameters typically required by complex post-processing methods.

The contribution of the training strategies is equally important. Batch normalization significantly stabilized gradient flow, allowing both sub-models to converge within 40 epochs—about 30% faster than training without batch normalization. Meanwhile, data augmentation—including random rotation, color dithering, and elastic deformation—reduced the Dice score variance on the validation set by 18%, indicating that the models learned more generalized and diverse features. Collectively, these strategies produced base predictions with low bias and low variance for the fusion stage, thereby amplifying the overall ensemble gains.

In contrast to previous studies, the present work further highlights the advantages of the proposed image segmentation framework. For example, Wu et al. introduced an improved Residual Attention Deconvolution U-Net model that surpasses conventional SegNet and U-Net in segmenting lung nodules from computed tomography images [19]. However, its adaptability and stability tend to be limited when confronted with complex image structures and diverse segmentation requirements. Han et al. presented a multi-channel medical image segmentation network, the Hierarchically Weighted Attention SegNet (HWA-SegNet), to address challenges in delineating irregular skin lesion boundaries [20]. While HWA-SegNet achieved a segmentation accuracy of 96.33% on the public International Skin Imaging Collaboration 2018 dataset, its Jaccard index remained at 85.90%. The current study also focuses on segmenting diverse skin lesion images from the HAM10000 dataset, particularly those exhibiting fine structural details and ambiguous boundaries in localized regions. The proposed average fusion strategy achieves a Jaccard index of 93.73% on the current dataset, surpassing that of HWA-SegNet by approximately 7.8 percentage points, while maintaining an accuracy of 92.82%. This result reflects the method's effectiveness in balancing segmentation completeness with precision.

While the fusion strategy proposed in this study demonstrates superior performance over individual models, its potential remains limited when compared to more sophisticated ensemble techniques. For instance, Sanga et al. and Wang et al. introduced post-processing methods based on conditional random fields or weighted voting mechanisms to combine model predictions, achieving notable improvements in segmentation accuracy and robustness [21, 22]. Nevertheless, such approaches relying on fixed weights or predetermined rules often lack the flexibility to adapt to varying image characteristics. Although the average fusion strategy used in this study is straightforward and effective, it does not fully account for the spatially varying performance of the two constituent models across different image regions. As a result, a more adaptive fusion method that dynamically adjusts weights based on localized model performance could further enhance segmentation outcomes.

Despite the demonstrated effectiveness of the proposed strategy, two main limitations warrant consideration. First, the training data were obtained from a single center, highlighting the need for validation on larger, multi-center datasets acquired with diverse imaging devices to further assess its generalization capability. Second, the model contains a moderate number of parameters and has not undergone pruning or quantization, limiting its immediate applicability in mobile or edge computing scenarios. Future work will focus on developing dynamic weighted fusion mechanisms guided by uncertainty estimation and exploring lightweight network architectures, with the goal of achieving real-time, high-accuracy dermatological lesion segmentation in resource-limited clinical environments.

Conclusion

This study implemented a simple averaging strategy to integrate SegNet and U-Net, leading to measurable improvements in dermatological image segmentation accuracy and generalization capability. Experimental results show that the Ensemble Model achieves enhanced performance across multiple evaluation metrics, with a particularly notable reduction in loss value, confirming its improved model efficacy. These advancements contribute to more precise delineation of lesion boundaries and detailed characterization of internal textures, underscoring the methodological value and practical potential of the fusion approach in medical imaging-assisted diagnosis.

Author contributions: Siqi Wang, Danhong Li, and Yina Zhang jointly conceived the study design and experimental protocol, implemented the SegNet, U-Net, and Ensemble Models, and wrote the initial manuscript. Yu Wang, Linrong Yuan, Miao Yu, Jianghui Li, and Yimeng Wang performed data preprocessing, model training and evaluation, analyzed the results, and contributed to manuscript writing. Ping Li provided critical revisions and supervised the entire project.

References

[1] Ray A, Sarkar S, Schwenker F, et al. Decoding skin cancer classification: perspectives, insights, and advances through researchers' lens. Sci Rep 2024;14(1):30542.
[2] Haque S, Ahmad F, Singh V, et al. Skin Cancer Detection Using Deep Learning Approaches. Cancer Biother Radiopharm 2025;40(5):301-312.
[3] Lee K, Cavalcanti TC, Kim S, et al. Multi-Task and Few-Shot Learning-Based Fully Automatic Deep Learning Platform for Mobile Diagnosis of Skin Diseases. IEEE J Biomed Health Inform 2023;27(1):176-187.
[4] Sarwar N, Irshad A, Naith QH, et al. Skin lesion segmentation using deep learning algorithm with ant colony optimization. BMC Med Inform Decis Mak 2024;24(1):265.
[5] Akram A, Rashid J, Jaffar MA, et al. Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of

Medical Things. Skin Res Technol 2023; 29(11):e13524.

[6] Verma N, Ranvijay, Yadav DK. Hybrid of Deep Feature Extraction and Machine Learning Ensembles for Imbalanced Skin Cancer Datasets. Exp Dermatol 2024;33(12):e70020.

[7] Juan CK, Su YH, Wu CY, et al. Deep convolutional neural network with fusion strategy for skin cancer recognition: model development and validation. Sci Rep 2023;13(1):17087.

[8] Zhang M, Zhang Y, Liu S, et al. Dual-attention transformer-based hybrid network for multimodal medical image segmentation. Sci Rep 2024;14(1):25704.

[9] Liu T, Bai Q, Torigian DA, et al. VSmTrans: A hybrid paradigm integrating self-attention and convolution for 3D medical image segmentation. Med Image Anal 2024;98:103295.

[10] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86:2278-2324.

[11] Singh SP, Wang L, Gupta S, et al. 3D Deep Learning on Medical Images: A Review. Sensors (Basel) 2020; 20(18):5097.

[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 2015;234-241.

[13] Yuan J. Brain tumor image segmentation method using hybrid attention module and improved mask RCNN. Sci Rep 2024;14(1):20615.

[14] Badrinarayanan V, Kendall A,Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 2017; 39(12):2481-2495.

[15] Skin Cancer MNIST: HAM10000. Available at: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

[16] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018;5:180161.

[17] Azad R, Aghdam EK, Rauland A, et al. Medical Image Segmentation Review: The Success of U-Net. IEEE Trans Pattern Anal Mach Intell 2024;46(12):10076-10095.

[18] Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Med Imaging 2021; 21(1):19.

[19] Wu Z, Li X, Zuo J. RAD-UNet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning. Front Oncol 2023;13:1084096.

[20] Han Q, Wang H, Hou M, et al. HWA-SegNet: Multi-channel skin lesion image segmentation network with hierarchical analysis and weight adjustment. Comput Biol Med 2023;152: 106343.

[21] Sanga P, Singh J, Dubey AK, et al. DermAI 1.0: A Robust, Generalized, and Novel Attention-Enabled Ensemble-Based Transfer Learning Paradigm for Multiclass Classification of Skin Lesion Images. Diagnostics (Basel) 2023;13(19): 3159.

[22] Wang Y, Wen Z, Su L, et al. Improved brain metastases segmentation using generative adversarial network and conditional random field optimization mask R-CNN. Med Phys 2024; 51(9):5990-6001.