Automatic Skin Lesion Segmentation Using an Enhanced U-Net Framework

Azhen Omar Jabbar and Asim Majeed Murshid
Department of Computer Science, College of Computer Science and Information Technology, University of Kirkuk, 36001 Kirkuk, Iraq
stcm23008@uokirkuk.edu.iq,2: dr.asim.majeed@uokirkuk.edu.iq

Keywords: U-Net, Medical Image Segmentation, Skin Lesion Segmentation, ISIC Dataset, Deep Learning.

Abstract: The given project offers an enhanced version of a U-Net-based framework for segmenting skin lesions with high precision because it is the best measure in the identification of melanoma at an early stage. ISIC2016, ISIC2017, and ISIC2018 datasets were used, including thousands of dermoscopic images with respective binary masks. The workflow ensemble with complex preprocessing (image resize, normalization, contrast enhancement, and severe data augmentation) was carried out to minimize the risk of the model being sensitive to changes in the appearance of the lesions. The training of the U-Net model involved a hybrid of Dice loss and binary cross-entropy loss, where the training was initiated to get a well-balanced performance between boundary and pixel-level accuracy. Morphological post-processing was used to clean up the small artifacts, fill the gaps, and smooth the mask borders to make the initial predictions more sophisticated. Specific measures of medical images, e.g., the Dice coefficient, Intersection over Union (IoU), precision, recall, and specificity, were employed to assess model performance. The data indicated very high segmentation accuracy in each of the datasets, which illustrated the robustness of the method in difficult cases that included low-contrast and irregularly shaped lesions. The approach is, thus, suitable for computer-aided diagnosis systems, which leads to more available and efficient skin cancer detection.

1 INTRODUCTION

The problem of image segmentation is a subject of computer vision. It is characterized by partitioning images or video shots into individual segments or objects and is used as a major ingredient in several systems, such as medical image analysis, self-driving cars, video surveillance, and augmented reality [1].

The segmentation of medical images allows one to outline organs, tissues, and abnormalities clearly so that they can be correctly diagnosed and treated. Deep neural networks have been developed to enhance efficiency and accuracy in computer-aided diagnosis, which has been the case as prevailing developments have allowed them to assist doctors and clinicians in detecting and analyzing important areas in medical images [2].

Several researchers are using deep learning models, machine learning algorithms [3], and medical image analysis techniques, which are often applied to raw data step-by-step [4]. Commonly used techniques in traditional medical image segmentation, such as edge detection and template matching methodologies, experienced major limitations as they were considerably susceptible to noise, poor contrast, and image distortions; hence, the lack of proper segmentation, especially in multifaceted medical images [5].

2 RESEARCH PROBLEM

The majority of the available medical image segmentation models grapple with the issue of processing complex and high-resolution dermoscopic images, and they usually tend to work best only when fed with a particular dataset. This explains why further improvement in feature extraction and the combination of both deep learning type and post-processing techniques are necessary to attain greater accuracy and generalization.

3 WORK CONTRIBUTIONS

Concepts in this paper propose an enhanced architecture approach to dermoscopic skin lesion automatic segmentation. The method will combine dropout regularization, combination loss (Dice + Binary Cross-Entropy), elaborate data augmentation, and morphological post-processing.

The enhancements solve some important difficulties facing medical image segmentation, including low contrast, image noise, and irregularities of lesion boundaries, resulting in better accuracy, robustness, and generalization.

4 RELATED WORK

The success of U-Net in medical image segmentation is attributable in part to its encoder-decoder structure and powerful local and global feature capture capability. Nonetheless, the standard U-Net continues to experience issues with insufficient contrast noisy background, and visibility of the lesion boundaries.

Scholars have come up with various enhanced versions to deal with these problems. nnU-Net, a self-adapting model, was systematically configured and automatically tuned by Isensee et al. [6] in an attempt to adapt preprocessing, architecture, and training measures to distinct medical image segmentation tasks. Such an automation enhances generalization and best performance in numerous public datasets.

Walsh et al. [7]used a standard U-Net to segment brain tumors in MRI scans, showing it can still deliver accurate and robust results even without major design changes. Çetiner et al.[8]proposed DenseUNet+ is a multi-modality imaging combined with dense connections within a U-Net architecture, with higher accuracy and boundary precision of brain tumor segmentation than that of the normal U-Net. Kashyap et al. [9] suggested a better U-Net connected with DenseNet-201 to find the glaucoma and optic cup in jelly images on retinas, with about 0.9882 and 0.9690 training and testing accuracy, respectively, surpassing the ordinary CNN architecture without manual tuning. Zhao et al. [10] introduced VM-Unet, which combines Vision Mamba with a state-of-the-art U-Net to perform medical image segmentation, with superior accuracy and extraction of global contextual information compared to traditional U-Net architectures. Qbal et al. [11] developed TESL-Net, a Transformer-based augmented CNN in skin lesion segmentation, demonstrating higher accuracies using the ISIC2016, ISIC2017, and ISIC2018 datasets than the traditional CNN-based solutions.

5 DATASET BRIEF DESCRIPTION

The data is composed of sweaty dermoscopic photos with skillfully judged masks. It is representative of numerous types of skin lesions, realistic, and intrinsically associated with the real world. Data is collected in ISIC datasets and is consistently standardized, providing consistent data used to train, validate, and test. The datasets related to this study were publicly and anonymously available at the time of the study: ISIC2016, ISIC2017, and ISIC2018. As such, no ethical approval was necessary, nor was patient consent to be obtained.

6 METHODOLOGY

The suggested methodology will start with the preparation and division of the dataset. Data preparation to analyze and model includes data ready [12]. The use of 3 well-known ISIC datasets (2016, 2017, and 2018) was realized, rated in a random procedure to a 70 percent training, 15 percent validation, and 15 percent test. This division guarantees impartial assessment, as well as allows the model to generalize to previously unseen pictures. The overall process of the proposed segmentation framework is illustrated in Figure 1.

6.1 Preprocessing

To increase the model's resilience and applicability, the images and their masks have undergone a comprehensive preprocessing chain. All the data were first resized to have a common resolution of 256×256 pixels to minimize the computation load as well as so that all the data were uniform in resolution. To minimize image noise and reject minor artifacts, a Gaussian blur filter was subsequently applied, so the model can concentrate on the main structural patterns. Then, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used for the enhancement of dermoscopic images with poor lighting or low contrast to create better visibility and delineation of the lesions. To further avoid overfitting and enhance data variation during training, data augmentation was applied in the form of horizontal and vertical flipping, random rotations, and brightness/contrast changes. This enabled the model to use more visual patterns in

learning and be more general in new instances. Lastly, pixel values were normalized to a fixed range of [0,1] so that image intensity values are consistent throughout all images and accelerate convergence with training (Fig. 2).

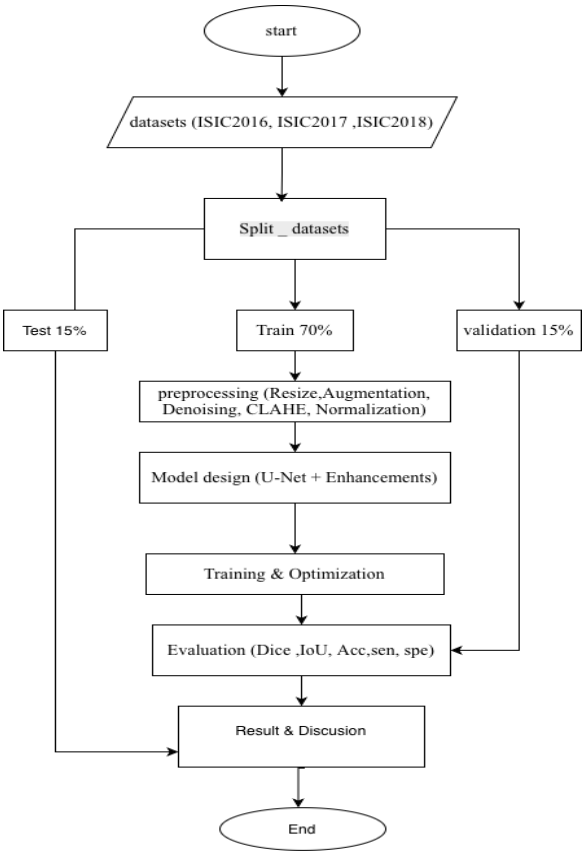

Figure 1: The General flowchart of the Proposed Segmentation Frame.

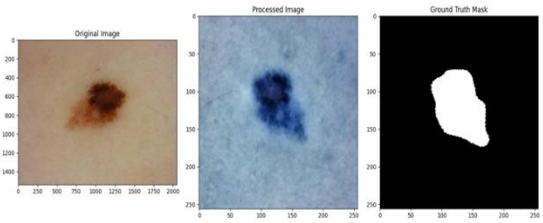

Figure 2: Preprocessing (original image, processed image, ground truth).

6.2 Model Design (U-Net Model)

The segmentation model implemented is the standard U-Net, consisting of three principal components: encoder, bottleneck, and decoder. This architecture can extract hierarchical features and leave fine spatial details that are required in segmenting the lesions accurately.

Input: The model accepts dermoscopic images in 256*256*3 dimensions, i.e., there are 256*256 dimensions and three channels, which are R, G, and B. Encoder: There are four blocks in the encoder. Both blocks consist of two Conv2D layers with ReLU activation and use the same padding, and use MaxPooling2D on the outputs to downsample in space. Each successive level will have twice as many filters as the previous one, with 64 through 512 filters.

A dropout with a rate of 0.5 is performed after the fourth block to limit overfitting.

Bottleneck The bottleneck is at the innermost part of the network and consists of two convolutional layers of 1024 filters, which serve as a compact representation of the original input image containing rich semantics. Decoder: The upsampling done by the decoder utilizes Conv2DTranspose. The upsampled features of the encoder process at each level are concatenated to the encoder layers through skip connections and allow more minute-level details to be retained. Each concatenated result is filtered through two Conv2D layers.

Output A 1x1 convolution and a sigmoid activation are used in the final layer, and the output surface is the binary identity of the area, 252x252x1, with a value indicating the likelihood of being a lesion.

6.3 Training Setup

All the experiments were handled by a MacBook Pro with an Apple M4 chip, 16 GB of unified memory, and GPU acceleration through the Apple Metal API, and were run using TensorFlow 2.16.2. This whole pipeline was written in Python 3.10, and random seeds were fixed so that they remained reproducible. Its training was carried out at a batch size of 4 and 256 x 256 x 3 input images. Adam optimizer was used with an initial learning rate of 5e-5. To balance the accuracy and region-level overlap, a composite loss function, consisting of Binary Cross-Entropy (BCE) and Dice loss, was used so that the algorithm converged better when working on imbalanced lesion data.

To avoid memorization bias, data were reshuffled at the beginning of each epoch. Random data augmentation (e.g., rotations, flips, and brightness/contrast) was performed on-the-fly to enhance generalization and mitigate overfitting. Two major Keras callbacks were incorporated into the training pipeline.

- The EarlyStopping: used to watch the performance loss on the validation set and halt

the training when no gains were noticed to keep training efficient.
- ModelCheckpoint: saved the best-performing model according to validation Dice coefficient.

Several libraries were used in the project: TensorFlow/Keras, which were used to implement and train models; OpenCV and Albumentations to preprocess and augment images; NumPy and Pandas to handle data; Matplotlib and Seaborn to visualize data; and Scikit-learn to evaluate metrics.

6.4 Validation

After every epoch, the validation was done with 15 percent of the dataset put aside before training. This step guaranteed that the performance of the model was observed on unknown data through training. Metrics such as validation loss and validation Dice score could be monitored, based on which one could understand whether there was an adequate generalization of the model or if it was overfitting. Gradients were never calculated on validation data, but instead, it was used as a cross-validation measure to tune hyperparameters and as a cross-validation measure to decide when to stop training. The superior models were selected based on their validation performance only, without any unfair and biased pre-selection of the models before testing.

6.5 Evaluation Metrics

To evaluate the performance of the proposed enhanced U-Net model, standard segmentation metrics were employed, including Accuracy, Dice Coefficient, Intersection over Union (IoU), Precision, Recall (Sensitivity), and Specificity. These metrics provide complementary insights into the model's ability to correctly identify lesion and non-lesion regions.

Accuracy measures the proportion of correctly classified pixels. However, it may be biased in imbalanced datasets where background pixels dominate.

Dice Coefficient (F1 Score) evaluates the overlap between predicted and ground-truth masks and is particularly suitable for medical image segmentation tasks involving irregular lesion boundaries.

Intersection over Union (IoU) provides a stricter overlap measure, penalizing both over-segmentation and under-segmentation.

Precision quantifies the proportion of correctly predicted lesion pixels among all predicted lesion pixels, reflecting the model's ability to reduce false positives.

Recall (Sensitivity) measures the proportion of actual lesion pixels correctly detected, which is critical in medical diagnosis to avoid missing lesions.

Specificity evaluates the model's ability to correctly identify non-lesion pixels, reducing false alarms in healthy regions.

All metrics were computed on the ISIC2016, ISIC2017, and ISIC2018 datasets to assess the robustness and generalization capability of the proposed model.

7 RESULTS AND DISCUSSION

In this section, the results of the proposed U-Net-based model in ISIC2016, ISIC2017, and ISIC2018 are displayed. This is accompanied by both quantitative and qualitative analyses to demonstrate the level of performance by the model.

7.1 Evaluation Metrics and Comparative Benchmarking

To holistically compare the segmentation accuracy of the proposed model with others, a total of eight quantitative measures were used. These are accuracy, precision, F1 score, Intersection over Union (IoU), Dice coefficient, mean IoU (mIoU), sensitivity (recall), and specificity. In the comparative analysis with reference to the TESL-Net model, however, only five of these key metrics, which include Accuracy, IoU, Dice Coefficient, sensitivity, and Specificity could be used to benchmark the analysis. The proposed model outperformed TESL-Net on all these typical parameters, as seen in Table 1, proving its superiority in lesion boundary detection, noise resistance, and generalization on the ISIC datasets.

7.2 Performance of Training and Validation

Loss and segmentation were monitored during the epochs to determine the training stability. The validation results after each epoch ensured that the model did not overfit and that generalization was maintained well.

Figures 3, 4, and 5: Training and validation curves of the suggested U-Net model. The left plot shows that the training and validation loss gradually decreased, implying that there is good convergence. The fine plot displays accuracy curves, where it is possible to see that there is a steady improvement, with no significant indicators of overfitting.

Table 1: Comparative performance between the proposed model and TESL-Net across ISIC datasets.

Datases	Model	Acc	Precise	F1 S	IOU	Dic	mIoU	Se	Sp
ISIC2016	TES_Net	96.40	-	-	89.51	93.43	-	94.55	97.02
	proposed	98.3	95.5	97.2	94.64	97.2	94.6	99.0	97.9
ISIC2017	TES_Net	95.80	-	-	86.91	90.09	-	91.10	97.29
	proposed	96.4	96.1	93.9	88.5	93.9	88.5	91.8	98.4
ISIC2018	TES_Net	96.23	-	-	90.56	94.22	-	95.02	97.21
	proposed	97.7	93.7	96.3	92.9	96.3	92.9	99.0	97.1

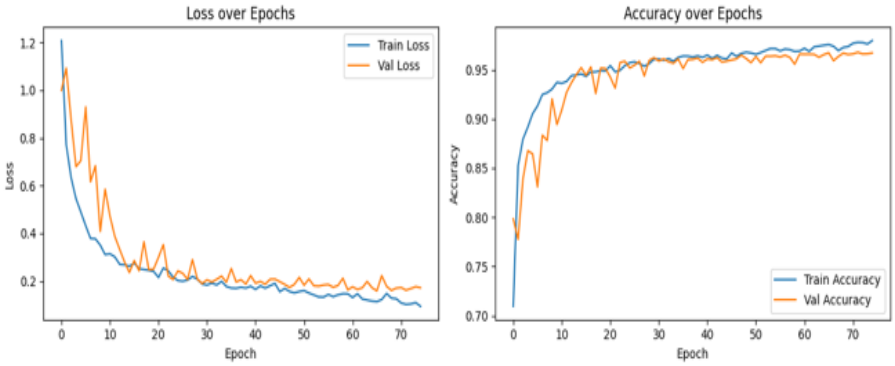

Figure 3: Training and validation accuracy curves over 75 epochs for ISIC2016.

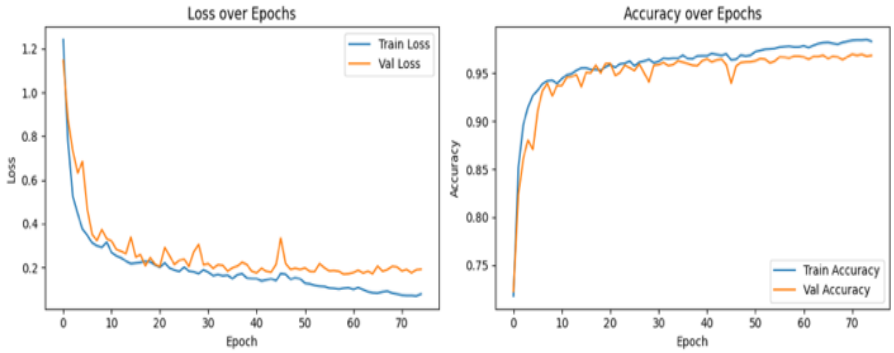

Figure 4: Training and validation accuracy curves over 75 epochs for IIC2017.

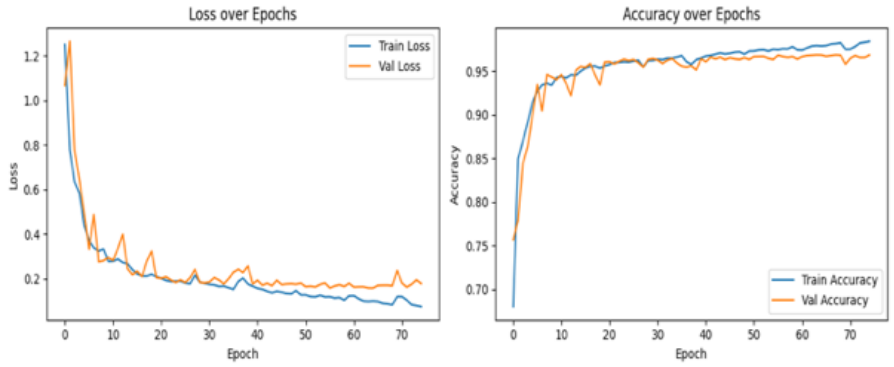

Figure 5: Training and validation accuracy curves over 75 epochs for ISIC2018.

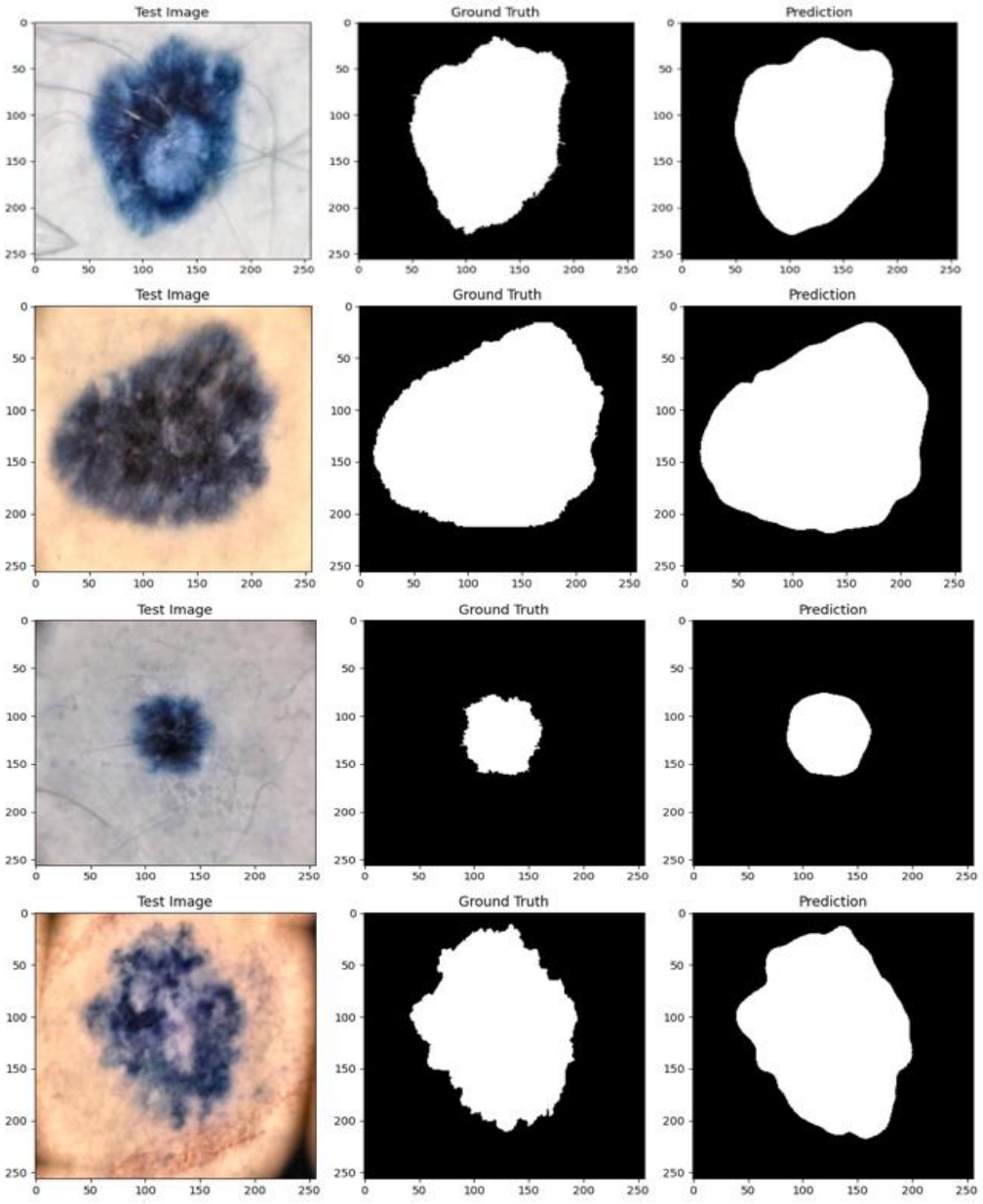

Figure 6: Qualitative segmentation results on the ISIC dataset.

7.3 Qualitative Results

A qualitative assessment of the performance about segmentation was also done, to complement the quantitative analysis of the same. Examples can be seen in Figure 3, where each triple of images has: (i) the original dermoscopic test image, (ii) the ground truth annotation that experts provided, and (iii) the predicted mask that can be generated by the proposed U-Net model. These outcomes prove that the model can successfully identify the boundaries of the lesion even in difficult circumstances, including the irregular shapes, color intensity, and blurred edges. The target masks agree well with the ground truth annotations, and this implies the effectiveness and accuracy of the model.

The other qualitative analysis was done by superimposing the predicted lesion contours with the ground truth boundaries, as shown in Figure 6. In both the samples, the contours of the expert-annotated

masks are provided in green (or blue in other instances), whereas the predicted contours provided by the proposed U-Net model are represented as red. The findings show that the expected lesion edges are close to the assigned lesion by the experts in various instances, comprising lesions with irregular shapes, smoothened borders, and different texture patterns (Fig. 7). This visual correspondence establishes the fact that this model is capable of not only an accurate prediction at pixel levels but also minuscule structural details, which further substantiates its prospective trustworthiness as a clinical context to interpret.

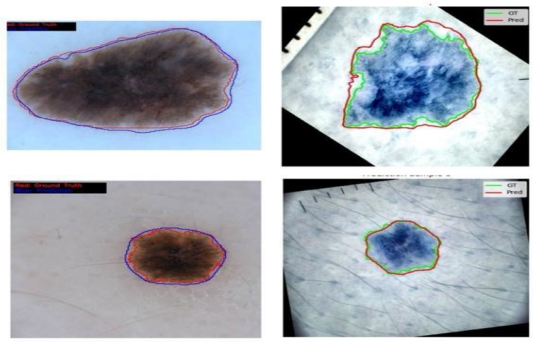

Figure 7: Contour-based qualitative segmentation results on the ISIC dataset.

8 CONCLUSIONS

In the current work, a modified U-Net architecture was proposed to achieve high segmentation accuracy while remaining computationally efficient. To address common limitations of U-Net, such as excessive parameter count and generalization issues, small convolutional filters were integrated into the bottleneck to reduce memory footprint, and dropout regularization was employed to prevent overfitting. Additionally, strong preprocessing (CLAHE, denoising) and on-the-fly augmentation (rotations, flips, brightness/contrast changes) were used to improve model robustness, even under limited or low-quality data conditions. The semantic gap between encoder and decoder was minimized through skip connections and hybrid loss optimization (Dice + Reverse Cross-Entropy), allowing better boundary precision and handling of small or unclear structures.

The model's efficiency was rigorously evaluated using a comprehensive set of pixel-based and structural metrics), including accuracy, precision, recall, specificity, F1-score, Dice, and IoU. These metrics were selected to reflect both regional overlap and boundary alignment. The proposed model achieved strong generalization across datasets:

- ISIC2016: Accuracy 0.983, Dice 0.972, IoU 0.946;
- ISIC2017: Accuracy 0.964, Dice 0.939, IoU 0.885;
- ISIC2018: Accuracy 0.977, Dice 0.963, IoU 0.929.

These consistent results confirm the model's effectiveness in overcoming traditional U-Net limitations and adapting to different lesion characteristics.

9 FUTURE WORK

Future improvement could include the incorporation of an attention mechanism to pick up on contextual information in dermoscopic images. Further exploration of transformer-based hybrid architectures or integration of the proposed U-Net with generative adversarial networks (GANs) is also likely to advance segmentation of low-contrast or overlapping lesions. It is also advisable that further experiments should be done using newer ISIC datasets (e.g. ISIC2019 and ISIC2020) and on real-world clinical data to assess model generalization. Further, shortening inference time to enable mobile or embedded use could be a future direction of real-time diagnosis applications.

REFERENCES

[1] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 7, pp. 3523–3542, Jul. 2022, doi: 10.1109/TPAMI.2021.3059968.

[2] F. Li et al., "PS5-Net: A medical image segmentation network with multiscale resolution," J. Med. Imaging, vol. 11, no. 1, p. 014008, Feb. 2024, doi: 10.1117/1.JMI.11.1.014008.

[3] H. W. Ahmed and A. M. Murshid, "Facial expression recognition using deep convolutional neural network," J. Comput. Anal. Appl., vol. 33, no. 8, 2024.

[4] I. Oksuz et al., "Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation," IEEE Trans. Med. Imaging, vol. 39, no. 12, pp. 4001–4010, Dec. 2020, doi: 10.1109/TMI.2020.3008930.

[5] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," arXiv:1705.03820, May 2017. [Online]. Available: http://arxiv.org/abs/1705.03820

[6] F. Isensee et al., “nnU-Net: Self-adapting framework for U-Net-based medical image segmentation,” arXiv:1809.10486, Sep. 2018. [Online]. Available: http://arxiv.org/abs/1809.10486

[7] J. Walsh, A. Othmani, M. Jain, and S. Dev, “Using U-Net network for efficient brain tumor segmentation in MRI images,” Healthcare Analytics, vol. 2, p. 100098, Nov. 2022, doi: 10.1016/j.health.2022.100098.

[8] H. Çetiner and S. Metlek, “DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation,” J. King Saud Univ. Comput. Inf. Sci., vol. 35, no. 8, Sep. 2023, doi: 10.1016/j.jksuci.2023.101663.

[9] R. Kashyap, R. Nair, S. M. P. Gangadharan, M. Botto-Tobar, S. Farooq, and A. Rizwan, “Glaucoma detection and classification using improved U-Net deep learning model,” Healthcare, vol. 10, no. 12, Dec. 2022, doi: 10.3390/healthcare10122497.

[10] J. Ruan, J. Li, and S. Xiang, “VM-UNet: Vision Mamba UNet for medical image segmentation,” [Online]. Available: https://github.com/JCruan519/VM-UNet

[11] S. Iqbal, M. Zeeshan, M. Mehmood, T. Khan, and I. Razzak, “TESL-Net: A transformer-enhanced CNN for accurate skin lesion segmentation,” in Proc. 25th Int. Conf. Digit. Image Comput.: Tech. Appl. (DICTA), 2024, pp. 313–320, doi: 10.1109/DICTA63115.2024.00054.

[12] H. A. Mohammad, “Hybrid deep learning techniques for improved anomaly detection in IoT environments,” Wasit J. Comput. Math. Sci., vol. 3, no. 4, pp. 62–77, 2024.