ISSN 2442-6571

Enhanced U-Net architecture with CNN backbone for accurate segmentation of skin lesions in dermoscopic images

Aqil Aqthobirrobbany [a,1], Resha Dwika Hefni Al-Fahsi [a,2], Indah Soesanti [a,3], Hanung Adi Nugroho [a,4,*]

[a] Department of Electrical and Information Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia
[1] aqilaqthobirrobbany1998@mail.ugm.ac.id; [2] resha.dwika.hefni.alfahsi@mail.ugm.ac.id; [3] indahsoesanti@ugm.ac.id; [4] adinugroho@ugm.ac.id
* corresponding author

ARTICLE INFO

Article history
Received October 16, 2023
Revised January 01, 2024
Accepted March 11, 2024
Available online August 31, 2024

Selected paper from The 2023 8th International Conference On Electrical, Electronics and Information Engineering (ICEEIE'23), Malang - Indonesia, September 28-29, 2023, https://iceeie.um.ac.id/. Peer-reviewed by ICEEIE'23 Scientific Committee and Editorial Team of IJAIN journal.

Keywords
Convolution neural network
Enhanced U-Net
Ensemble segmentation methods
skin Skin lesion segmentation

ABSTRACT

Addressing the critical public health challenge of skin cancer, particularly melanoma and non-melanoma, this study focuses on enhancing early diagnosis through improved automatic segmentation of skin lesions in dermoscopic images. The researchers propose an optimized U-Net architecture that integrates advanced convolutional neural networks (CNNs) with backbone models such as ResNet50, VGG16, and MobileNetV2, specifically designed to handle the inherent variability and artifacts in dermoscopic imagery. The method's effectiveness was validated using the ISIC-2018 dataset, and our U-Net model incorporating the VGG16 backbone achieved notable improvements in segmentation accuracy, demonstrating an accuracy rate of 0.93. These results signify significant enhancements over existing methods, emphasizing the potential of the proposed approach in aiding precise skin cancer diagnosis and detection. This study makes a valuable contribution to dermatological imaging by presenting an advanced method that substantially boosts the accuracy of skin lesion segmentation, addressing a crucial need in public health.

1. Introduction

Melanoma and non-melanoma represent two distinct forms of skin cancers. Statistical observations often indicate a lower incidence of non-melanoma compared to melanoma in many nations [1]. However, data from the GLOBOCAN database managed by the International Agency for Research on Cancer (IARC) revealed in 2018 that there were about 287,700 cases of melanoma, contrasting with a notably higher count of 1,042,100 cases of non-melanoma skin cancers [2]. Notably, the annual increase in deaths related to melanoma skin cancer (MSC) has been significant, at approximately 50% over the past decade [3], underscoring the growing public health concern posed by skin cancer.

Early detection of melanoma skin cancer (MSC) is paramount for achieving higher success rates in treatment and recovery [3]. However, autonomously segmenting skin abnormalities in dermoscopic photographs is a complex task, fraught with challenges due to the heterogeneity in skin characteristics, including lesion locations, morphologies, textures, sizes, and colors. Additionally, various extraneous factors, such as body hair and inconsistent lighting, further compound the intricacies of the segmentation process [4]. These patient-specific factors pose significant challenges to the accurate diagnosis and segmentation of skin lesions, necessitating the utilization of advanced and robust techniques to address the wide array of variations encountered in clinical settings.

In the initial stages of automated skin lesion segmentation, the accurate identification of skin lesion borders is crucial for effectively distinguishing lesions from the surrounding skin background. Traditionally, combinations of thresholding techniques, as described in previous studies [5], [6], have been employed to delineate these boundaries. Additionally, recent research [7] has utilized color space evaluation and clustering-driven histogram thresholding strategies to pinpoint the most suitable color channel for skin lesion segmentation. However, these approaches often required the extraction of pre-defined image features. For instance, Jafari [8] elucidated 10 distinct color characteristics from skin lesion images captured with standard cameras. The methodology utilized the fuzzy C-means approach to derive five colors based on chromatic variation and spatial arrangement, and another five based on color intensity. This extraction method significantly improved segmentation efficiency. In a similar vein, Pereira [9] introduced supplementary features, particularly borderlines, obtained through segmentation masks created by combining gradients and Local Binary Patterns (LBP). The integration of this feature was observed to enhance segmentation precision.

The research mentioned above focuses on feature extraction primarily relying on hand-crafted techniques to derive necessary attributes. This phase is crucial as it ensures a discerning representation of skin lesions for subsequent analyses [10]. However, these hand-crafted methodologies require a profound understanding similar to that of dermatologists and demand considerable time and varied methods to capture all essential features. To overcome the limitations inherent in these hand-crafted approaches, feature extraction through Deep Learning has been introduced, specifically employing Convolutional Neural Networks (CNNs). Inspired by the intricacies of the human visual cortex [11], CNNs are designed to leverage latent information within image datasets. Interestingly, they autonomously discern features and patterns, some of which might elude human detection [12]. It is noteworthy that over the past half-decade, CNN models have surged in prominence, accounting for 71% of machine learning applications in the analysis of pigmented skin lesions.

The utilization of deep learning in the analysis of pigmented skin lesions was initially introduced by Codella [13] in 2015. Since then, the prevalence of Convolutional Neural Networks (CNNs) has grown significantly, as demonstrated by the majority of participants in the 2017 ISIC challenge [14]. The strength of deep learning with CNNs lies in its end-to-end methodology, seamlessly integrating feature extraction and classification into a unified model architecture. The current trajectory of research is shifting towards multi-class segmentation, recognizing the nuanced clinical complexities involved in assessing pigmented skin lesions [15]. In recent years, numerous deep learning architectures have emerged to enhance segmentation efficacy. Notable among these are the Fully Connected Network (FCN) [16] and the U-Net [17]. After exploring various deep learning-based approaches for their effectiveness and accuracy, S. Manivannan [18] introduces a U-Net architecture augmented with a ResNet101 backbone, tailored for precise skin lesion segmentation. This model benefits significantly from advanced data augmentation techniques and leverages the efficient feature extraction capabilities of the ResNet architecture, achieving impressive Jaccard scores of 96.28% in training and 84.20% in testing, surpassing previous benchmarks in this domain. Similarly, V. Anand [19] presents an enhanced U-Net architecture designed specifically for the accurate segmentation of dermoscopic images. This enhancement involves modifications to the feature map's dimensions and the addition of more kernels to improve nodule extraction. The model's effectiveness was evaluated across various hyperparameters, including epochs, batch size, and optimizers. Testing was conducted on the augmented PH2 dataset to increase the available image pool, and the most effective performance was achieved using an Adam optimizer with a batch size of 8 for 75 epochs.

In the field of medical image segmentation, particularly in skin lesion analysis, the U-Net architecture has emerged as a standout performer. Its encoder-decoder design effectively extracts contextual features, leading to significant advancements in performance. This approach is increasingly acknowledged as the pinnacle in pigmented skin lesion segmentation, surpassing traditional techniques with superior performance metrics. Despite these strides, deep learning methodologies, including U-Net, encounter challenges such as reduced sensitivity and a notable lack of interpretability in results. These issues underscore the necessity for continuous refinement in deep learning approaches. Striking a balance between high efficiency and improved sensitivity, along with clearer interpretability, is crucial for ensuring their broader applicability and trustworthiness in medical diagnostics.

The primary advancements of the study encompass the following key elements:

- The study elevates the U-Net architecture by seamlessly integrating Convolutional Neural Networks (CNN) to achieve a more precise diagnosis and segmentation of skin lesions.
- Foundational models, including ResNet50, VGG16, and MobileNetV2, are strategically incorporated into the encoder segment of the U-Net architecture. This integration aims to augment the overall performance of the model.
- The enhanced U-Net model's performance is compared to that of the standard U-Net model, which lacks these sophisticated backbones.

The remainder of the paper is organized as follows: Section II provides an in-depth exploration of the methods and materials utilized in this study. It encompasses the preprocessing steps essential for skin tumor segmentation, elucidates the U-Net segmentation technique, and introduces the U-Net with backbone. In Section III, the researchers delve into the experiments, present the results, and engage in corresponding discussions. Finally, Section IV encapsulates the concluding remarks of this research

2. Method

A comparative methodology has been developed to segment skin lesions by contrasting the outcomes of using the U-Net method alone and the U-Net method with a fully automated backbone. Furthermore, an image resizing preprocessing strategy was employed to optimize the architectural performance. The research is organized into three phases: preprocessing, segmentation, and the presentation and comparison of results, as illustrated in the flowchart depicted in Fig. 1.

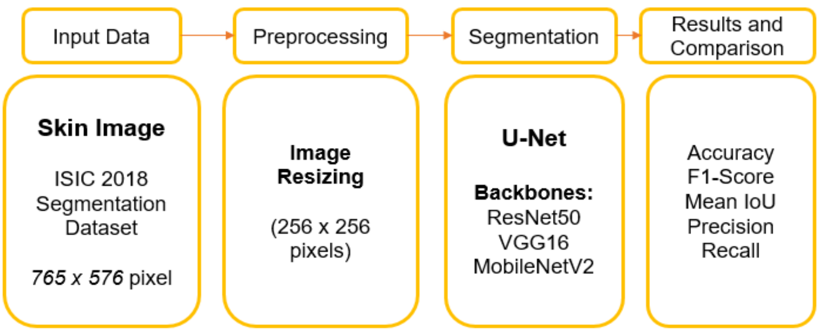


Fig. 1. Flow Diagram of the Study

2.1. Dataset

The ISIC2018 [14] dataset functions as a publicly accessible resource for both classification and segmentation tasks related to skin lesions. It holds a prominent position as a widely utilized dataset for training deep learning frameworks. Comprising a total of 2594 datasets and an equal number of ground truths, the dataset is divided into 1558 for training, 518 for validation, and 518 for testing, spanning seven distinct grades. In Fig. 2, exemplary images are presented alongside their corresponding ground truth segmentation images.

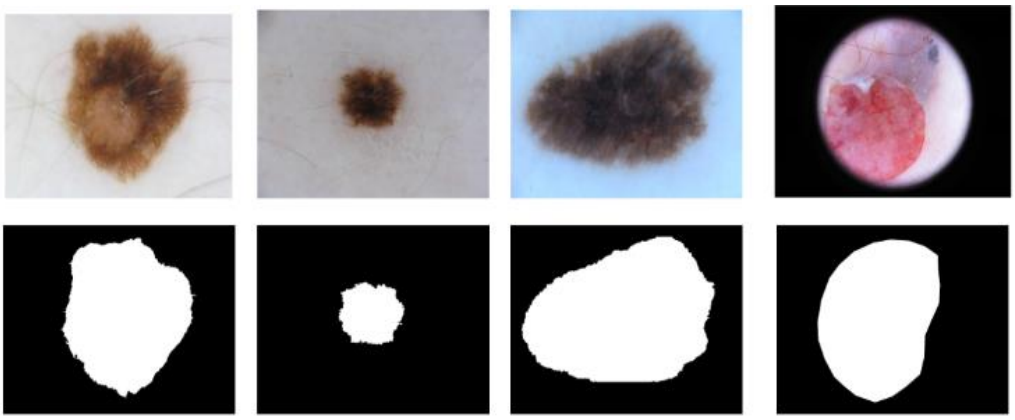


Fig. 2. Sample Images of ISIC 2018 Dataset [10]

2.2. Preprocessing

The preprocessing procedure plays a crucial role in enhancing image fidelity and improving segmentation efficacy in dermoscopic image analysis, as supported by prior research [20], [21]. In this study, the researchers implemented a resizing-type preprocessing technique to evaluate its impact on the architectural performance of the U-Net framework, particularly concerning the ISIC 2018 dataset of dermoscopic images. Resizing, a fundamental image processing technique, is utilized to adjust the spatial parameters of the image. This step involves transforming the images from their original dimensions of 765x576 pixels to a standardized size of 256x256 pixels. This decision was made to strike a balance between maintaining image detail and ensuring computational efficiency. The resizing process preserves the aspect ratio of the images, which is crucial for maintaining the integrity of dermoscopic features [22].

For resizing, the utilization of bicubic interpolation is essential [22], [23]. This technique entails the evaluation of multiple adjacent pixels surrounding each pixel in the original image, enabling the accurate determination of values in the resized image. Bicubic interpolation is selected for its capacity to generate smoother and visually coherent outputs in comparison to simpler interpolation methods. This attribute is particularly crucial for preserving the nuanced details in dermoscopic images.

As illustrated in Fig. 3, the resizing process is conducted by iterating through each pixel in the new 256x256 image, calculating its value based on the corresponding pixels in the original 765x576 image using bicubic interpolation. This meticulous process is applied across all color channels of the image to preserve critical color information essential for accurate dermoscopic analysis. Following this resizing procedure, the images are suitably prepared for introduction into the U-Net framework for segmentation tasks. The resized images, while smaller in dimensions, retain the essential features of the original images, thereby ensuring they are representative for analysis and computationally efficient for processing in the U-Net model.

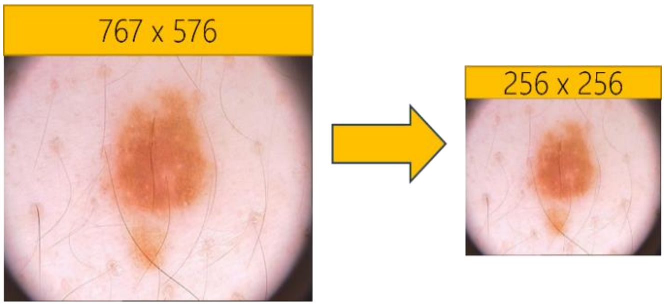


Fig. 3. Resizing Image [23]

2.3. Segmentation Methods

In this segment, the researchers present a succinct overview of the fundamental Convolutional Neural Network (CNN) and its various iterations frequently used in the analysis of medical images. CNN models commonly tackle three primary tasks in medical imaging, which can be broadly categorized as classification, detection, and segmentation. Notable architectures like GoogleNet, VGG, and ResNet predominantly emphasize classification by incorporating fully-connected layers alongside distinct classifiers. Models such as R-CNN, Faster R-CNN, and YOLO are typically employed for object detection, whereas FCNN and U-Net find extensive use in segmentation tasks.

A Convolutional Neural Network (CNN) consists of three essential layers: the input, hidden, and output layers. This structure draws parallels to conventional neural networks, such as Artificial Neural Networks (ANN). The key distinction between CNN and ANN is particularly evident in the hidden layer. In the architecture of a CNN, the hidden layer traditionally consists of three fundamental segments: the convolutional layer, the pooling (often referred to as subsampling) layer, and the fully connected layer [24]. Image segmentation involves the division of digital images into smaller groups known as segments based on pixel characteristics [17]. This study focuses specifically on skin cancer as our area of interest. To extract the region of interest corresponding to skin cancer, the researchers employ various methods, including the utilization of U-Net alone and U-Net with a backbone segmentation process. A concise description of this methodology is provided below.

2.3.1. U-Net architecture

Introduced in 2015, the U-Net model [17] has showcased its proficiency in precisely segmenting small targets and its adaptable network structure. With the rising demand for accurate segmentation in medical imaging, U-Net has garnered over 2500 academic citations, underscoring its growing recognition [25]. This network comprises two main components: the encoding and decoding sections. Given the limited size of the dataset in this study, the U-Net architecture was deliberately chosen.

To enhance the model's performance, modifications were made to the architecture, including the integration of pre-trained backbone models such as VGG16, ResNet50, and MobileNetV2 on the encoder side. This strategic adjustment has led to significant improvements in model performance. For a visual representation of the U-Net architecture, refer to Fig. 4.

In the U-Net architecture, the left convolutional section functions as the contracting path, while the right or deconvolutional section serves as the expansive path for achieving semantic segmentation. The convolutional network architecture consists of a repetitive pattern involving two 3×3 convolution operations followed by a Rectified Linear Unit (ReLU) activation, culminating in a downsampling process facilitated by a 2×2 max-pooling operation with a stride of 2. Conversely, the deconvolutional framework encompasses an upsampling procedure applied to the feature map derived from the contraction phase. The antecedent feature map produced during the contracting phase requires a subsequent 3×3 convolution followed by a ReLU layer. This entire process involves 23 profound convolutional layers, with the terminal layer designated to characterize each component feature vector relevant to its class. This architecture is designed to accept RGB input along with the corresponding binary mask. During training with weights, the architecture takes an RGB input image and its corresponding binary mask. The training process involves adapting weights through the utilization of stochastic gradient descent (SGD) [26].

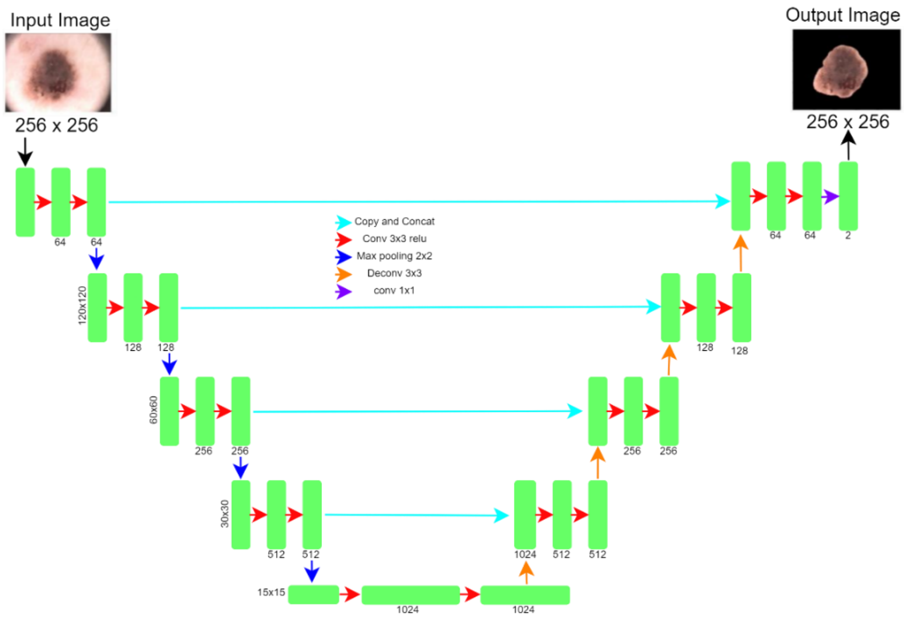


Fig. 4. U-Net architecture [17]

2.3.2. CNN backbone (VGG16, ResNet50 and MobileNetV2)

The VGG16 backbone consists of 16 convolutional layers, renowned for their stacked 3x3 kernels, which effectively extract highly localized features [27]. The VGG16 sequence concludes with a dimensionality reduction layer, succeeded by three fully connected layers, enhancing its robust feature extraction capabilities.

MobileNetV2 is specifically designed for mobile applications, with a focus on enhancing computational speed and efficiency without compromising accuracy. It incorporates advanced techniques such as separable convolution and a linear bottleneck, effectively reducing the model's parameter count while accelerating its operations [28].

The ResNet50 architecture is specifically designed to mitigate feature degradation in deep neural networks. Comprising 50 layers, it incorporates residual blocks to maintain initial information throughout the network's depth, thereby augmenting the learning of complex features [29].

Incorporating these backbones into the U-Net's encoder enhances the extraction of intricate features while maintaining high image resolution. The U-Net decoder effectively leverages the features from these backbones, resulting in superior segmentation outcomes. The adoption of this integrated approach in skin lesion segmentation facilitates the capture of diverse and complex image features, offering a more comprehensive representation of skin lesions. This methodology aims to refine the model's proficiency in distinguishing between healthy skin and lesions—an essential aspect for accurate diagnosis and early detection of skin cancers.

2.4. Pseudocode Training Data

The dataset employed in this study is ISIC-2018, comprising 2594 JPEG skin images accompanied by corresponding skin segmentation achieved through image masks. The original images have a resolution of 765 x 576 pixels. However, for further processing, the dimensions of the images are standardized to 240 x 240 pixels. The dataset is partitioned with 60% of the data allocated for training, 20% for validation, and the remaining 20% for testing purposes.

In the model processing, various hyperparameters are employed, including 100 epochs, a batch size set to 4, and a fixed learning rate of 1e-4. The chosen optimizer is Adam, with epsilon set to 1e-7, beta_1 set to 0.9, and beta_2 set to 0.999. To regulate the model training process, two essential callbacks are utilized: 'Reduce Learning Rate on Plateau' and 'EarlyStopping.' The 'Reduce Learning Rate on Plateau' strategy is implemented to adjust the learning rate downward if the model's performance shows stagnation over successive epochs. In this context, the learning rate decay is set at 0.1, and the patience level is set to 5 epochs. Meanwhile, 'EarlyStopping' is employed to halt the model training if there is no improvement in performance over several consecutive epochs. In this scenario, the patience level for 'EarlyStopping' is set at 10 epochs.

- ❑ Begin
- ❑ Load Dataset:
  - ✓ ISIC-2018:
    - ▪ 2594 JPEG Skin Image + Segmentation skin image mask
    - ▪ Resolution: 765 × 576 → Resized to 240 × 240
    - ▪ 60% training, 20% validation and 20% testing
- ❑ Hyperparameters
  - ✓ Epoch → 100
  - ✓ Batch size → 4
  - ✓ Learning rate → 1e–4
- ❑ Optimizer:
  - ✓ Adam
- ❑ Callback
  - ✓ Reduce LRO and Plateau
    - ▪ learning rate decay → 0.1
    - ▪ patience → 5 epochs
  - ✓ EarlyStopping
    - ▪ Patience → 10 epochs
- ❑ Train model using CNN architecture with training data
- ❑ Validate model using validation data and callbacks
- ❑ Test model using testing data
- ❑ Evaluate model performance using appropriate metrics

End

2.5. Performance Matrix

The segmentation algorithms evaluated in this study underwent an assessment of their performance, utilizing a meticulously chosen set of evaluation metrics. These metrics encompass the Jaccard Index (IoU), F1-Score, Sensitivity, Precision, and Accuracy [30], [31]. The selection of these metrics was guided by their demonstrated relevance and effectiveness in the field of medical image analysis, specifically in the context of skin lesion segmentation.

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$F1 - Score = \frac{2\, x\, precision\, x\, recall}{precision+recall} \tag{3}$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

$$Jeccard\ index(A,B) = \frac{|y_a \cap y_b|}{|y_a|+|y_b|-|y_a \cap y_b|} \tag{5}$$

Sensitivity (Eq. 1) is a crucial metric that gauges the algorithm's effectiveness in correctly identifying actual positives (true positives, TP). This measure is essential for evaluating the algorithm's capability to detect lesions accurately, ensuring that no lesions are overlooked, which is particularly vital in medical diagnostics. Precision (Eq. 2) assesses the accuracy of lesion detection by determining how many of the positively identified cases are truly positive. This metric aids in evaluating the specificity of the algorithm's results. The F1-Score (Eq. 3) is a comprehensive metric that combines Precision and Recall (Sensitivity), providing a balanced assessment of both aspects. It is particularly useful in situations where an equal emphasis on both Precision and Recall is desired, offering a holistic view of the algorithm's performance. Accuracy (Eq. 4) serves as an overarching effectiveness measure for the model. It indicates the proportion of correctly identified pixels in the image, offering insights into the overall performance of the algorithm in lesion detection.

The Jaccard Index, also known as the Intersection over Union (IoU) (Eq. 5), precisely measures the intersection between the predicted segmentation and the ground truth. This metric offers a direct evaluation of spatial accuracy in lesion delineation, providing valuable insights into the algorithm's effectiveness in precisely capturing the boundaries of lesions.

The training for each network was conducted on a GPU apparatus with the following specifications: (1) Hardware: The machine featured an Intel i5-5200 CPU running at 2.20GHz, an NVIDIA GeForce 930M GPU, and 12GB of DDR3 RAM. (2) Software: The training was carried out using TensorFlow.

3. Results and Discussion

The image segmentation method underwent testing using the ISIC-2018 image dataset [14], which comprises segmented images of skin lesions associated with seven disorders. Each image includes a manually delineated cancer area, denoted as the ground truth image in this study. Our experimental dataset encompassed 2594 skin cancer images, distributed for training (60% or 1558 images), validation (20% or 518 images), and testing (20% or 518 images) purposes. Each sample image has a resolution of 765 by 576 pixels and is formatted in JPEG.

3.1. Subjective Performance Evaluation

The Intersection over Union (IoU) criterion, commonly known as the Jaccard Index, serves as a quantitative measure to evaluate the agreement between a predicted image and the reference ground truth image. It quantifies the percentage of overlap by assessing the area of intersection—indicating where both images overlap—and the area of union, representing the combined coverage of both images. The IoU provides a relevant metric for assessing the congruence between predicted and reference ground truth images, particularly in terms of their overlapping regions.

For a thorough evaluation of the model's predictive accuracy, it is essential to compare the forecasted area with the reference ground truth area. This comparative analysis enables a comprehensive understanding of the model's proficiency in identifying the lesion zone within the image. Assessing the congruence between the forecasted and ground truth regions serves as a reliable indicator of the model's efficacy in precisely pinpointing the lesion domain.

Fig. 5 presents a sample case where the ground truth image has an area of 15,989.5 cm². The predicted area is compared using four different methods, resulting in prediction areas of 0, 16,123.5, 15,254, and 15,569.5, respectively. Additionally, the perimeter values of the object in the ground truth image and the prediction images are calculated as 521.55 and 665.55, 506.86, 502.62, and 543.83, respectively, for each method. By evaluating the IoU values, it is observed that the U-Net + ResNet50 and U-Net + VGG16 methods yield the highest IoU values, namely 0.99 and 0.97, respectively, indicating better performance in this particular sample.

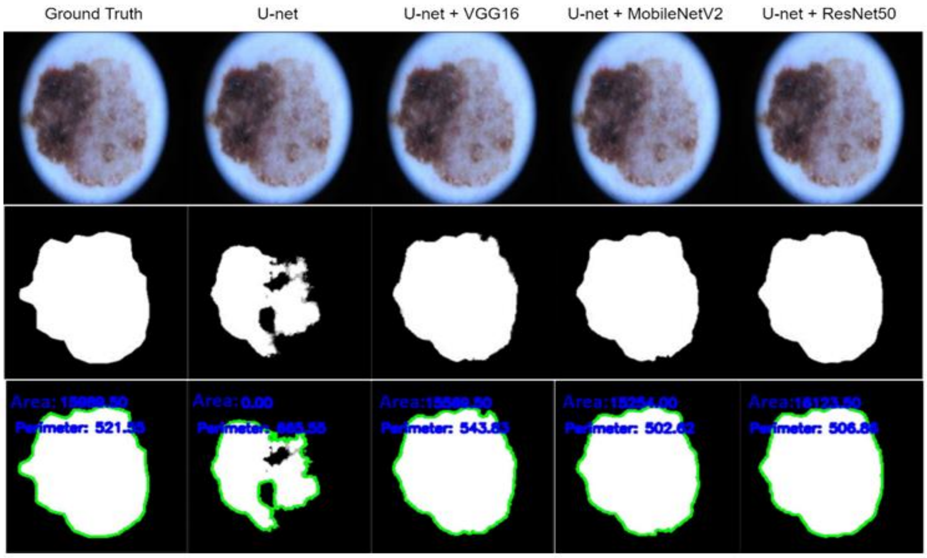


Fig. 5. Outcomes pertaining to the area and perimeter measurements for each respective method are presented

3.2. Quantitative Performance Evaluation

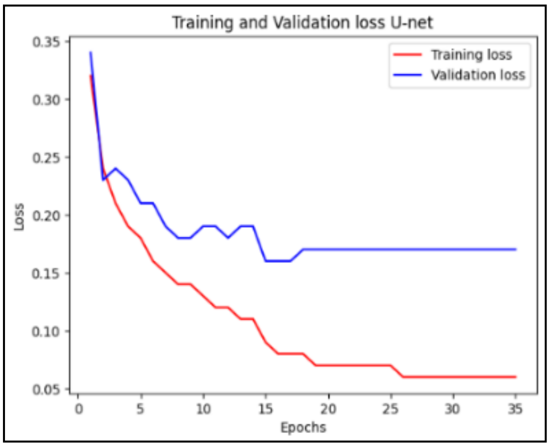
To mitigate potential biases in subjective evaluations, it is crucial to augment them with quantitative assessments employing diverse quality metrics. These metrics encompass Accuracy, F1-Score, IoU (Intersection over Union), Recall, and Precision. The incorporation of such quantitative metrics guarantees a more objective evaluation of segmentation outcomes, providing a comprehensive understanding of the effectiveness of the segmentation algorithms.

Based on the assessment results presented in Table 1, it is apparent that integrating the U-Net architecture with the VGG16 backbone enhances efficacy compared to the other three examined methodologies. Nevertheless, the performance differences among the models are relatively small. These findings align with the results of the loss functions obtained during the training process.

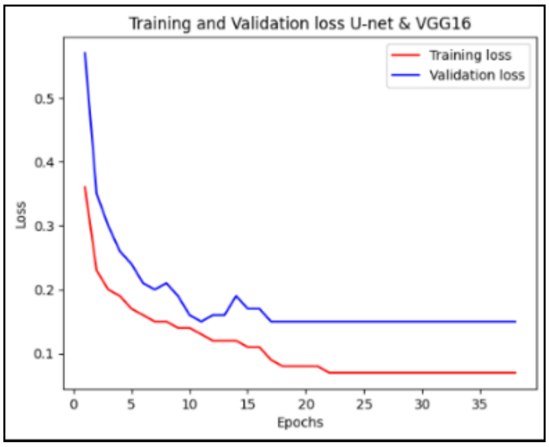
Table 1. Quantitative Performance Evaluation of Image Segmentation Method

Model	Acc.	F1-Score	IoU	Recall	Precision
U-Net [17]	0.9208	0.7275	0.6313	0.7321	0.8016
U-Net + VGG16 [32]	**0.9319**	**0.7802**	**0.6838**	**0.8161**	**0.8124**
U-Net + ResNet50 [33]	0.9278	0.7334	0.6401	0.7482	0.7894
U-Net + MobileNetV2 [34]	0.9183	0.7301	0.6236	0.7916	0.7639

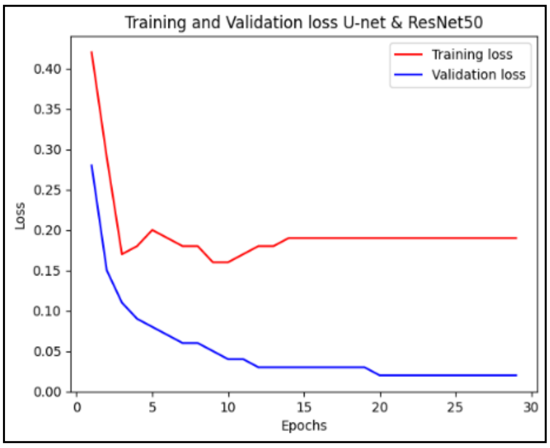
In Fig. 6, the training and validation losses associated with the U-Net and VGG16 architectures are illustrated, demonstrating that both models effectively minimize loss during training. It is important to note that a high validation loss may suggest overfitting, wherein the model struggles to generalize well to unseen data despite performing well on the training data.



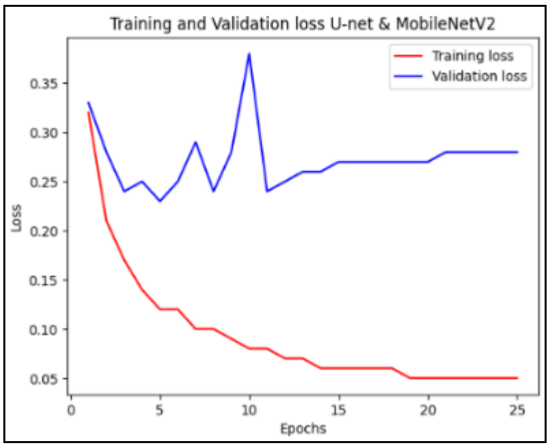
(1)



(2)



(3)



(4)

Fig. 6. Loss Function Results (1) U-net (2) U-net + VGG16 (3) U-net + ResNet50 (4) U-net + MobileNetV2

Fig. 7 presents the accuracy distribution across the four examined models. Notably, the U-Net+VGG16 configuration outperforms its counterparts in terms of accuracy. Conversely, the U-Net+MobileNetV2 model exhibits a narrower range of accuracy distribution, which tends to be lower compared to the other models. After comparing the loss function and validation loss, as well as examining the accuracy distribution of the four tested models, it is evident that the U-Net+VGG16 model outperforms the other three models. This conclusion is based on the higher accuracy values and the consistent range of accuracy values observed in the U-Net+VGG16 model.

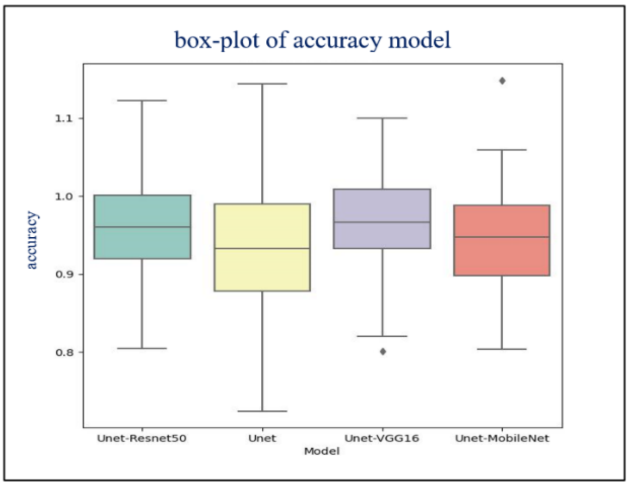


Fig. 7. Accuracy model box-plot results

4. Conclusion

In this study, the researchers present an augmented U-Net framework specifically designed for precise skin lesion segmentation in dermoscopic imagery. Our modification to the conventional U-Net involves the integration of VGG16, ResNet50, and MobileNetV2 CNN architectures into its encoder component. Our empirical findings highlight the superiority of the U-Net augmented with the VGG16 architecture over alternative methodologies. This enhanced model adeptly identifies contextual features, resulting in reliable segmentation outcomes. Quantitative assessment metrics, including Accuracy, F1-Score, IoU, Recall, and Precision, consistently support the claim that the U-Net-VGG16 architecture outperforms other examined models. The combination of U-Net and VGG16 architecture shows significant potential for automating skin lesion segmentation, a critical step in early detection and diagnosis of skin cancers. The high accuracy achieved by our proposed method suggests its potential to assist medical professionals in conducting precise diagnoses. However, it is essential to note that further research and validation on larger datasets are necessary to fully assess the generalization and robustness of our approach. The incorporation of advanced deep learning techniques, such as U-Net with a VGG16 backbone, has the potential to revolutionize skin cancer diagnosis and management, contributing to improved healthcare services and enhanced quality of life for patients. As part of the future work, the researchers intend to focus on incorporating explainable AI (XAI) into our deep learning model. Explainable AI aims to create more transparent algorithms, enabling users, especially healthcare professionals in our context, to understand and trust the decisions made by AI systems. This transparency is crucial in the medical field, as it facilitates a better understanding of the reasoning behind AI-driven diagnostic suggestions.

Acknowledgment

This research was financially supported by the Directorate of Research at Universitas Gadjah Mada, under the auspices of the Final Project Recognition Grant, Universitas Gadjah Mada, bearing the reference number 5075/UN1.P.II/Dit-Lit/PT.01.01/2023. Additionally, the Department of Electrical and Information Engineering at UGM also provided research grant assistance.

Declarations

Author contribution. Each author made an equivalent contribution to the principal content of this manuscript. All contributors have reviewed and endorsed the final version of the paper.

Funding statement. This research endeavor received financial backing from both the Directorate of Research at Universitas Gadjah Mada and the Department of Electrical and Information Engineering at UGM.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] "Skin cancer statistics," *World Cancer Research Fund International*, 2022. [Online]. Available at: https://www.wcrf.org/cancer-trends/skin-cancer-statistics/.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[3] Euromelanoma and Global Coalition for Melanoma Patient Advocacy, "2O2O Melanoma Skin Cancer Report," 2020. [Online]. Available at: https://melanomapatients.org.au/wp-content/uploads/2021/09/2020-campaign-report-GC-version-MPA_1.pdf.

[4] N. K. Mishra and M. E. Celebi, "An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning," *arXiv Comput. Vis. Pattern Recognit.*, p. 15, Jan. 2016. [Online]. Available at: https://arxiv.org/abs/1601.07843v1.

[5] L. Zhang, G. Yang, and X. Ye, "Automatic skin lesion segmentation by coupling deep fully convolutional networks and shallow network with textons," *J. Med. Imaging*, vol. 6, no. 02, p. 1, Apr. 2019, doi: 10.1117/1.JMI.6.2.024001.

[6] P. M. M. Pereira *et al.*, "Dermoscopic skin lesion image segmentation based on Local Binary Pattern Clustering: Comparative study," *Biomed. Signal Process. Control*, vol. 59, p. 101924, May 2020, doi: 10.1016/j.bspc.2020.101924.

[7] R. Garnavi, M. Aldeen, M. E. Celebi, G. Varigos, and S. Finch, "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Comput. Med. Imaging Graph.*, vol. 35, no. 2, pp. 105–115, Mar. 2011, doi: 10.1016/j.compmedimag.2010.08.001.

[8] M. H. Jafari, S. Samavi, S. M. R. Soroushmehr, H. Mohaghegh, N. Karimi, and K. Najarian, "Set of descriptors for skin cancer diagnosis using non-dermoscopic color images," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, vol. 2016-Augus, pp. 2638–2642, doi: 10.1109/ICIP.2016.7532837.

[9] P. M. M. Pereira *et al.*, "Skin lesion classification enhancement using border-line features – The melanoma vs nevus problem," *Biomed. Signal Process. Control*, vol. 57, p. 101765, Mar. 2020, doi: 10.1016/j.bspc.2019.101765.

[10] C. Barata, M. E. Celebi, and J. S. Marques, "A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 3, pp. 1096–1109, May 2019, doi: 10.1109/JBHI.2018.2845939.

[11] Q. Sang, Z. Shu, L. Liu, C. Hu, and Q. Wu, "Image quality assessment based on self-supervised learning and knowledge distillation," *J. Vis. Commun. Image Represent.*, vol. 90, p. 103708, Feb. 2023, doi: 10.1016/j.jvcir.2022.103708.

[12] M. Cullell-Dalmau, M. Otero-Viñas, and C. Manzo, "Research Techniques Made Simple: Deep Learning for the Classification of Dermatological Images," *J. Invest. Dermatol.*, vol. 140, no. 3, pp. 507-514.e1, Mar. 2020, doi: 10.1016/j.jid.2019.12.029.

[13] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images," in *Lecture Notes in Computer Science*

(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9352, Springer, Cham, 2015, pp. 118–126, doi: 10.1007/978-3-319-24888-2_15.

[14] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr. 2018, vol. 2018-April, pp. 168–172, doi: 10.1109/ISBI.2018.8363547.

[15] R. C. Maron *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *Eur. J. Cancer*, vol. 119, pp. 57–65, Sep. 2019, doi: 10.1016/j.ejca.2019.06.013.

[16] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, p. 114417, May 2021, doi: 10.1016/j.eswa.2020.114417.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[18] S. Manivannan and N. Venkateswaran, "Skin Lesion Segmentation using Residual U-NET," in *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, Mar. 2023, pp. 405–409, doi: 10.1109/SPIN57001.2023.10116287.

[19] 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.

[20] E. Uysal and G. E. Güraksin, "Computer-aided retinal vessel segmentation in retinal images: convolutional neural networks," *Multimed. Tools Appl.*, vol. 80, no. 3, pp. 3505–3528, Jan. 2021, doi: 10.1007/s11042-020-09372-w.

[21] M. Tamoor, A. Naseer, A. Khan, and K. Zafar, "Skin Lesion Segmentation Using an Ensemble of Different Image Processing Methods," *Diagnostics*, vol. 13, no. 16, p. 2684, Aug. 2023, doi: 10.3390/diagnostics13162684.

[22] B. K. Triwijoyo and A. Adil, "Analysis of Medical Image Resizing Using Bicubic Interpolation Algorithm," *J. Ilmu Komput.*, vol. 14, no. 1, p. 20, Apr. 2021, doi: 10.24843/JIK.2021.v14.i01.p03.

[23] D. Khaledyan, A. Amirany, K. Jafari, M. H. Moaiyeri, A. Z. Khuzani, and N. Mashhadi, "Low-Cost Implementation of Bilinear and Bicubic Image Interpolation for Real-Time Image Super-Resolution," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, Oct. 2020, pp. 1–5, doi: 10.1109/GHTC46280.2020.9342625.

[24] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019, doi: 10.1016/j.ymssp.2018.05.050.

[25] X.-X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "U-Net-Based Medical Image Segmentation," *J. Healthc. Eng.*, vol. 2022, no. 1, pp. 1–16, Apr. 2022, doi: 10.1155/2022/4189781.

[26] H. Akan and M. Z. Yıldız, "Development of new descriptor for melanoma detection on dermoscopic images," *Med. Biol. Eng. Comput.*, vol. 58, no. 11, pp. 2711–2723, Nov. 2020, doi: 10.1007/s11517-020-02248-z.

[27] S. Benyahia, B. Meftah, and O. Lézoray, "Multi-features extraction based on deep learning for skin lesion classification," *Tissue Cell*, vol. 74, p. 101701, Feb. 2022, doi: 10.1016/j.tice.2021.101701.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[29] S. Ali, M. Hassan, J. Y. Kim, M. I. Farid, M. Sanaullah, and H. Mufti, "FF-PCA-LDA: Intelligent Feature Fusion Based PCA-LDA Classification System for Plant Leaf Diseases," *Appl. Sci.*, vol. 12, no. 7, p. 3514, Mar. 2022, doi: 10.3390/app12073514.

[30] S. Mongkhonthanaphon and Y. Limpiyakorn, “Classification of titanium microstructure with fully convolutional neural networks,” *J. Phys. Conf. Ser.*, vol. 1195, no. 1, p. 012022, Apr. 2019, doi: 10.1088/1742-6596/1195/1/012022.

[31] D. Müller, I. Soto-Rey, and F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation,” *BMC Res. Notes*, vol. 15, no. 1, p. 210, Dec. 2022, doi: 10.1186/s13104-022-06096-y.

[32] A. A. Pravitasari, N. Iriawan, U. S. Nuraini, and D. A. Rasyid, “On comparing optimizer of UNet-VGG16 architecture for brain tumor image segmentation,” in *Brain Tumor MRI Image Segmentation Using Deep Learning Techniques*, Elsevier, 2022, pp. 197–215, doi: 10.1016/B978-0-323-91171-9.00004-1.

[33] S. Lefkovits and L. Lefkovits, “U-Net architecture variants for brain tumor segmentation of histogram corrected images,” *Acta Univ. Sapientiae, Inform.*, vol. 14, no. 1, pp. 49–74, Aug. 2022, doi: 10.2478/ausi-2022-0004.

[34] B. C. Bag, “Unet Mobilenetv2: Medical Image Segmentation Using Deep Neural Network (Dnn),” *J. Mech. Contin. Math. Sci.*, vol. 18, no. 1, pp. 21–29, Jan. 2023, doi: 10.26782/jmcms.2023.01.00002.