RESEARCH ARTICLE

Robust Skin Lesion Segmentation Using a Hybrid CNN Model and Compound Scaling

Maame G. Asante-Mensah[1], Charles Roland Haruna[1,*], Stephen Julius Atsu[1], Sandro Kwame Amofa[1] and Akua Gyanba Morgan Acquah[1]

[1] Department of Computer Science and Information Technology, University of Cape Coast, Cape Coast, Ghana

Abstract

Medical image processing by computers has become an area of growing importance, with researchers shifting their attention to various groundbreaking state-of-the-art models that aim to ease the pressure on doctors. Notwithstanding, image segmentation and feature extraction are critical yet challenging tasks in clinical diagnostics, hindered by issues such as noise and artifacts, limited annotated data, and the need for high computational efficiency. This research focuses on developing an EfficientNet-U-Net model, which aims to improve medical image segmentation for detecting skin lesions utilizing the International Skin Imaging Collaboration Challenge (ISIC) 2018 dataset. The model combines the EfficientNet encoder, optimized for proficient feature extraction via compound scaling, with a U-Net-style decoder that uses transposed convolutions and skip connections to preserve spatial information. The approach also utilizes a hybrid loss function that combines Dice and Binary Cross-Entropy, as well as an adaptive learning rate schedule. Experimental results indicate that the proposed model performs well for the fully automated identification and delineation of skin lesions with the potential for clinical use. The proposed model achieves performance comparable to that of existing skin lesion segmentation methods.

Keywords: medical imaging, EfficientNet-UNet, skip connection, skin lesion, image segmentation.

1 Introduction

Cutaneous malignant melanoma represents a major global health burden, with its incidence increasing in many populations. More than 324,000 new cases and over 57,000 deaths due to melanoma have been reported worldwide in 2020 alone, and the figures are expected to continue to increase until 2040 [1, 2]. In numerous developing countries, there is a lack of dermatologists, which aggravates the situation as patients are usually diagnosed late and the outcomes are poor [2]. Among other types of skin cancer, melanoma has the greatest risk of death due to its rapid spread when not detected at an early stage, and proper delineation of lesion boundaries is thus a priority in clinical care [1]. The standard imaging method to investigate the skin lesion is dermoscopy, which is a non-invasive method; however, the accuracy of interpretation differs significantly among clinicians, and this discrepancy has been an issue in dermatology

Submitted: 23 April 2026
Revised: 24 May 2026
Accepted: 26 August 2026
Published: 20 October 2026

Vol. 2, **No.** 4, 2026.
10.62762/TMI.2026.565544

***Corresponding author:**
✉ Charles Roland Haruna
charuna@ucc.edu.gh

Citation
Asante-Mensah, M. G., Haruna, C. R., Atsu, S. J., Amofa, S. K., & Acquah, A. G. M. (2026). Robust Skin Lesion Segmentation Using a Hybrid CNN Model and Compound Scaling. *ICCK Transactions on Machine Intelligence*, 2(4), 213–223.

© 2026 ICCK (Institute of Central Computation and Knowledge)

since the beginning of the century [3]. Deep learning-based methods have demonstrated excellent results in automating this analysis in recent years, and hybrid segmentation models, in particular, have results comparable to those of expert analysis on standard benchmark data [4, 5].

One of the biggest issues with medical imaging is that it lacks a reliable reference standard because of a large degree of inter- and intra-observer variability caused by the subjectivity of its interpretation [6]. Conventionally, image segmentation in hospitals has been relying on manual labelling performed by experts in the field of radiology and pathology. For example, studies developing automated segmentation systems for renal carcinoma have highlighted that manual delineation is time-consuming and subject to significant inter-observer variability among radiologists [7, 8], motivating the shift toward automated deep learning approaches. This creates challenges for humans interpreting images from MRI or CT scans, leading to misinterpretations and potentially adverse outcomes. Hence, it is imperative that image segmentation helps to characterise a patient's condition and abnormality represented on a medical image and acts as a base for many important medical applications including detecting tumors, segmenting organs, and diagnosing diseases. Furthermore, deep learning with convolutional neural networks (CNN) is able to classify skin lesions with comparable accuracy to that of board-certified dermatologists, which opens up the potential for fully automated diagnostic workflows, thereby increasing access to care beyond the physical clinic [9].

As mentioned previously, CNNs are part of the deep learning sub-field of computer science and their ability to accomplish tasks involving medical imaging, as well as many other fields, has exhibited remarkable capabilities. The use of deep learning has improved accuracy and efficiency of medical imaging tasks through a broad spectrum of algorithmic approaches—spanning convolutional networks, encoder-decoder architectures, and attention mechanisms—thereby advancing pattern recognition, segmentation, and clinical decision-making across diverse imaging modalities [10]. The existing methods have been proven to be insufficient for maintaining the imaging of critical applications such as medical images; deep learning provides the latest techniques for image segmentation which are based on previous techniques [11].

U-Net is the best way to segment medical images because it captures context in a contracting path while keeping spatial information in a symmetrical expanding state [12]. Recent lateral changes made to performance improvements on the modified U-Net model used new architectures, including (EfficientNet) U-Net encoder building and reduced computational costs. Previous studies have demonstrated that integrating EfficientNet as the encoder of U-Net can enhance feature extraction and segmentation performance while improving computational efficiency compared with the conventional U-Net architecture. CNN-based approaches have demonstrated substantial performance improvements in skin lesion classification, achieving diagnostic accuracy comparable to board-certified dermatologists, thereby illustrating the broad clinical impact of modern convolutional networks across both segmentation and classification tasks in dermatology [13]. The changes, such as CNN-based approaches and self-configuring frameworks, have been demonstrated to give optimal performance when added to medical imaging instances where they have very high performance in large international medical imaging benchmarks, such as the 2020 BraTS brain tumor segmentation challenge [14], and yield the best results in the field of both brain lesion segmentation and tumor segmentation.

In general, U-Net represents one of the most popular designs for performing image segmentation in healthcare due to its encoder-decoder style architecture, and more specifically, the symmetric expansion section which allows for accurately identifying spatial locations of pixels. Although U-Net has demonstrated state-of-the-art results across numerous biomedical tasks, traditional U-Net architectures utilize conventional convolutional layers, resulting in lower computational efficiency and performance when extracting more complicated features compared to modern architectural designs (e.g., EfficientNet) utilized by other convolutional neural networks. The addition of EfficientNet to U-Net implementation will introduce complexity to the model which in turn will enable the model to be more capable of extracting high-quality features with efficient computational processes, hence, be usable in real-world clinics [7]. The extension of the U-Net model through multi-scale feature extraction strategies—which capture fine-grained local detail alongside broader contextual information at multiple resolutions—represents one of the

effective approaches to image segmentation in the healthcare industry, providing strong support for the development of clinical decision support systems [15].

2 Related works

U-Net and EfficientNet are both excellent choices for image segmentation as these two different networks complement each other perfectly with regard to their individual strengths. Tan and Le [16] developed EfficientNet with the goal of scaling CNNs more sensibly, and the result is a network that gets more accurate without the usual cost of ballooning parameter counts. The U-Net [12] uses a U-shaped architecture and has utilized skip connections resulting in preserved detailed features during image segmentation. By using EfficientNet in combination with U-Net, the extraction capabilities of EfficientNet and the segmentation capabilities of U-Net combine to create a mechanism providing increased accuracy and efficiency. The DoubleU-Net structure has been continuously more effective as a result of outperformance when compared to traditional U-Net architecture (standard) along with several other baseline models used within the context of medical imaging segmentation tasks; thus, because of this consistent comparison, DoubleU-Net provides a reliable/generalizable baseline model for solving many complex biomedical imaging problems [17].

Only a limited number of studies have investigated architectures that combine EfficientNet encoders with U-Net-style decoders. Trinh et al. [18] developed an EfficientNet-B4-based U-Net incorporating a residual refinement module and a specialized loss function for biomedical image segmentation. UNet++ implemented a modified way of utilizing skip connections to exploit multi-scale features through a nested dense skip pathway [19, 20] in order to create a more complex model based on the original U-Net design. These studies demonstrate the potential of combining modern CNN encoders with U-Net-style decoders for medical image segmentation [7]. Other research works have demonstrated success in using hybrid architectures to reach their respective goals as evidenced by Abedalla et al. [21], who showed strong performance of multiple U-Net models with different backbones (i.e., one being EfficientNet-B4) in the SIIM-ACR Pneumothorax Segmentation Challenge; likewise, Vong and Dinh [22] proved that using an EfficientNet-B4 encoder with UNet++ architecture produced results of higher accuracy than standard approaches to pneumothorax detection in chest X-ray images.

Ruan et al. [23] developed EGE-UNet for skin lesion segmentation by integrating a lightweight attention module and a group aggregation bridge. This model emphasizes efficiency without sacrificing image segmentation quality because the model maintains a high performance on datasets like (International Skin Imaging Collaboration Challenge) ISIC 2017 and ISIC 2018 [24], while reducing computational complexity. This shows EfficientNet's strength in resource-limited settings, such as mobile health and outperforming traditional Transformer models with fewer parameters. Recognizing the significant computational complexity of channel attention models, Wang et al. [25] proposed a novel approach to channel attention, the Efficient Channel Attention (ECA) module, which uses a local cross-channel interaction method to reduce the computational burden in channel attention significantly without sacrificing performance. These light modules of attention are useful for computer vision, where efficient architectural design [23] is desirable.

To enhance the segmentation performance, the state-of-the-art U-Net++, designed in 2018 by Zhou et al. [19], is integrated with the EfficientNet backbones, as mentioned in [22]. The recent developments point to two aspects of image segmentation: improved efficiency and improved quality (segmentation) [26] from new architectures specifically designed for use in vision transformer architectures in resource-constrained edge and mobile devices. The proposed method leads to better segmentation of medical images for applications such as organ-specific structure identification and lesion detection in medical images using EfficientNet's efficiency and transformer-related techniques to optimise for edge devices. It is an approach that demonstrates either the evolution of EfficientNet-UNet or its applicability in the real world. Recently, CNNs and Transformers have been combined as a promising strategy to tackle the problem of local feature extraction and global context understanding, and widely used [27].

2.1 Strengths and Limitations of U-Net in Medical Image Segmentation

When the U-Net model was introduced to medical imaging research in 2015, it was considered one of the best deep learning models ever developed, thanks to its unique encoder-decoder structure with skip connections that help maintain spatial information in segmentation processes. Many studies have shown

that the U-Net model performs well in various applications (e.g., segmentation of brain tumors [28] or lung diseases [29]). Multiple investigations have confirmed that U-Net can perform tasks reliably and successfully because of the fact that it is able to give proper results even if used in standard form with no significant changes made related to complicated features in 3D medical imaging segmentation [30]. However, U-Net has not been without flaws, including the propensity for overfitting due to minimal training sample size, a frequent reality in the medical imaging community. Although such architectures as U-Net are effective, they use the simple convolution operations that are based on local neighborhoods. As Lee et al. [31] note, focusing on the local receptive areas only can result in an insufficient amount of comprehension regarding the entire situation, which contributes to handling the entire image in the medical field. Most applications of U-Net are not useful in reducing the large class imbalance, and therefore, result in the under-segmentation of smaller or fewer parts of anatomy. This has prompted a significant amount of work on how to improve the U-Net architecture in terms of attention mechanisms and multi-level context gating strategies that selectively attend to more meaningful features [32, 33], more complexity of the loss function, and hybrid models to improve accuracy and robustness.

3 Methodology

3.1 Proposed Model Architecture

The proposed architecture of the model is based on an EfficientNet [16] backbone with a U-Net [12] architecture (to improve the segmentation of medical images). U-Net is known to capture the spatial details from the input image via its encoder-decoder design. However, the standard U-Net architecture relies on simple convolutional encoders, which may be suboptimal for capturing complex feature hierarchies. In order to overcome the limitations of the encoder, we opted to replace it with another encoder using EfficientNet, which is a powerful convolutional neural network (CNN) created by Tan and Le [16] and used in many recent medical segmentation systems for its effective feature extraction capabilities [34].

According to Tan and Le [16], when it comes to superb performance, EfficientNet has demonstrated superiority in feature extraction by optimizing depth, width and resolution scaling through a compound coefficient strategy. EfficientNet achieves higher accurate performance on multiple vision tasks with fewer trainable parameters. Again, it captures high-level semantic features effectively, which helps improve segmentation accuracy in medical imaging tasks. Lastly, EfficientNet minimizes overfitting and improves model adaptability to different datasets.

3.2 Architectural Modifications for Enhanced Segmentation Performance

The proposed changes made to the model to improve the results of segmentation include the addition of transpose convolution and the use of EfficientNet as an encoder. Upsampling feature maps in the decoder section of the model is done through the use of a layer called transposed convolutions that are used to help generate a segmentation mask by means of reconstructing a segmentation mask from a deep feature representation. EfficientNet-B0 served as the encoder in the model. The encoder had already demonstrated effective features when performing the task due to the high degree of feature extraction accomplished by the encoder through its use of batch normalization, which stabilizes training by normalizing feature distributions, and ReLU activations, which introduce non-linearity and help mitigate vanishing gradient issues during optimization of this neural network. The model was architecturally influenced by the U-Net model with EfficientNet used to extract features and sequential transposed convolutions to upsample the feature maps used to create the final segmentation probability map. Figure 1 illustrates our proposed EfficientNet-UNet architecture.

3.3 Data Augmentation Techniques for Enhancing Model Training

To improve the generalizability of the model and to avoid overfitting, several data augmentation techniques were used:

- Geometric transformations (horizontal flipping 50%, vertical flipping 20% and rotating images between $\pm 30°$ with a 50% chance) can help improve the model's ability to generalise to previously developed unlabelled data; therefore, enabling it to produce more accurate results based on less annotated data compared to the original [35], because they replicate the natural variability found in lesion orientation and position typically observed in clinical practice during medical image segmentation and provide additional variability by simulating the natural variability typically seen in lesion orientation and position entered into models for lesion segmentation

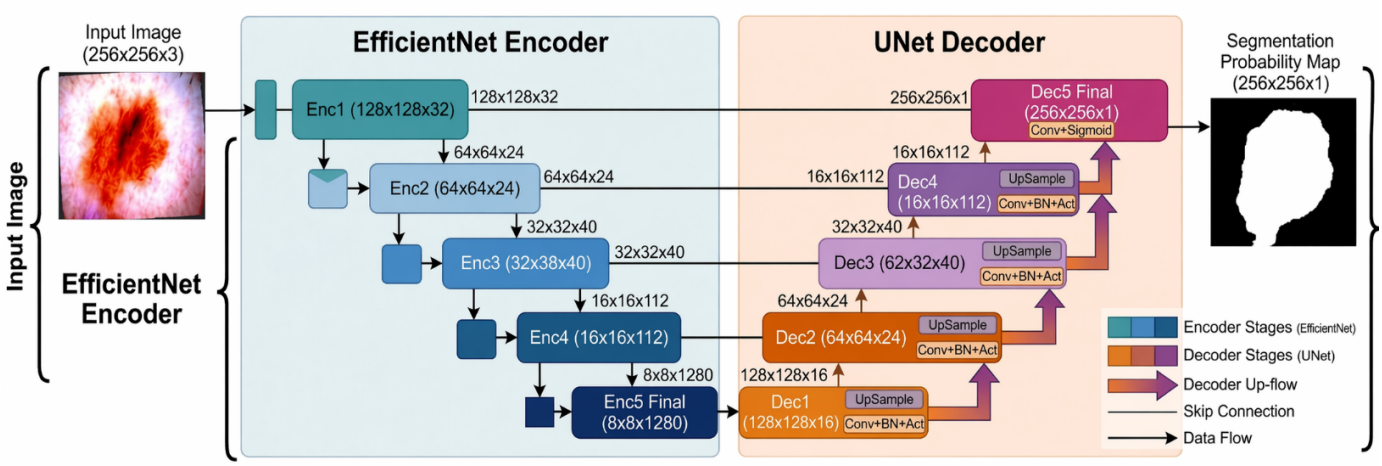


Figure 1. Proposed EfficientNet-UNet architecture, combining an EfficientNet-B0 encoder with a U-Net-style decoder linked by skip connections.

during training.

- Colour space adjustments: creation of random changes to the brightness and contrast (20%) at all instances gives a greater pool of training samples; colour space augmentations (brightness and contrast) create variability to match the wide variety of illumination conditions and factor in imaging device differences encountered across different imaging devices and environments; consequently, aiding the consistency of the model [5].

- Normalization during the training process makes the values of the pixels stable during training by normalizing the values according to the statistical values of ImageNet, for example the mean value (0.485, 0.456, 0.406) and the standard deviation value (0.229, 0.224, 0.225). This normalization is applied as a standard preprocessing step when using encoders pretrained on ImageNet, such as EfficientNet, in order to align the input distribution with the statistics seen during pretraining. The practical implications of such transfer learning choices for feature representation and fine-tuning dynamics in medical imaging—including the nuanced conditions under which ImageNet pretraining proves beneficial—have been examined empirically in prior work [36]. In this study, there are 1200 pairs of image-masks divided into 80% and 20% for training and validation respectively, providing the capability of fully testing the performance.

4 Experiments and results

4.1 Implementation Specifications

We leveraged Google Colab's environment where we used its GPU for training. During the training settings, the key hyperparameters used were a batch size of 4 samples per batch, with the learning rate initially established at 1×10^{-4}, with an adaptive decay approach applied, which reduces the learning rate by half after every 10 epochs. The Adam optimization algorithm was utilized for updating the model's parameters. The model underwent training for a total of 50 epochs.

4.2 Loss Function and Assessment Metrics

To tackle the issue of class imbalance typical in medical image segmentation, a hybrid loss function is utilized:

The Dice Loss evaluates how well the predicted segmentation overlaps with the actual segmentation, i.e., it measures the spatial agreement of the predicted to the actual segmentation.

$$\text{Dice Loss} = 1 - \frac{2\sum P_i \cdot G_i + \epsilon}{\sum P_i + \sum G_i + \epsilon} \tag{1}$$

where P_i is the predicted probability for pixel i (the output of the model after applying a sigmoid activation function); G_i is the ground truth binary mask for pixel i (1 for lesion, 0 for background); ϵ is a small constant (typically 1) added to the numerator and denominator of the equation to prevent division by zero.

The Binary Cross-Entropy (BCE) Loss increases the accuracy of classification on a pixel level by penalizing the classifier for making incorrect predictions based on the ground truth.

$$\text{BCE Loss} = -\frac{1}{N}\sum [G_i \cdot \log(P_i) + (1 - G_i) \cdot \log(1 - P_i)] \quad (2)$$

where N is the total number of pixels in the image, G_i is the ground truth label for pixel i (1 for lesion, 0 for background), and P_i is the predicted probability that pixel i is a member of the lesion class as produced by the sigmoid activation function applied to the output of the model. The log terms in the equation result in heavy penalties for making confident but inaccurate predictions, which makes BCE a useful loss function for learning the pixel boundaries on a case-by-case basis.

Model performance is evaluated using the following metrics:

Dice Similarity Coefficient (DSC): This metric measures the effectiveness of segmentation through the calculation of the ratio of two times the area of intersection over the sum of the areas of predicted and true areas, making it a good measure of overlap.

$$\text{DSC} = \frac{2\sum(P_i > \tau) \cdot G_i + \epsilon}{\sum(P_i > \tau) + \sum G_i + \epsilon} \quad (3)$$

P_i is the predicted probability of pixel i, and τ is the threshold value used for binarizing the prediction into a binary mask (in practice, $\tau \approx 0.5$). G_i is the binary mask of pixel i as ground truth (i.e., $G_i = 1$ for pixels that are lesions; $G_i = 0$ for pixels that are background) and ϵ is a smoothing value for avoiding division by zero. The DSC metric will yield a score of 1 if the predicted lesion exactly overlaps with the true lesion, whereas a score of 0 represents no overlap.

Intersection over Union (IoU):

$$\text{IoU} = \frac{\sum(P_i > \tau) \cdot G_i}{\sum(P_i > \tau) + \sum G_i - \sum(P_i > \tau) \cdot G_i + \epsilon} \quad (4)$$

where P_i, τ, G_i and ϵ are as defined above. The IoU metric looks at both areas and thus gives a more stringent measure than the DSC metric for comparing the predicted and true lesions.

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{\text{Total Pixels}} \quad (5)$$

where $TP = \sum(P_i > \tau) \cdot G_i$ is the count of correctly predicted lesion pixels (true positives) and $TN = \sum(1 - (P_i > \tau)) \cdot (1 - G_i)$ is the count of correctly predicted background pixels (true negatives). Total Pixels is the total number of pixels in the image. The numerator totals the accurately predicted lesion pixels (i.e., true positives) and accurately predicted background pixels (i.e., true negatives), so accuracy will be affected by imbalanced classes in datasets that include far more background than lesion pixel instances.

4.3 Evaluation of Convergence and Loss Patterns

During the entire training of the model there was a constant reduction of the loss function indicating that learning was being successfully achieved. The minimum of the Dice loss indicates that the predicted segmentation masks are more aligned to the true segmentation masks. Although the validation loss demonstrates a consistent trend, there have been fluctuations, which could mean that overfitting might be occurring; thus, dropout and weight decay are being used to mitigate those potential problems. This experiment tests the efficiency of using the EfficientNet-UNet model for medical image segmentation and applies this experiment specifically to the ISIC 2018 dataset [37], as it is a standard benchmark used to validate skin lesion segmentation. The ISIC 2018 dataset was selected because its pixel-level ground-truth masks (compared to the ISIC 2020 and ISIC 2024 datasets) make it the best recent image dataset that can be used to validate true skin lesion segmentation tasks.

The proposed EfficientNet-UNet model was trained and evaluated on a subset of 1200 image-mask pairs drawn from the ISIC 2018 Task 1 training set, selected to provide a computationally tractable development and verification environment while preserving class diversity across lesion types. Preprocessing and augmentation of data were performed in accordance with the information presented in Section 3.3, such as geometric transformations, colour space manipulations, and ImageNet-based normalisation, which enhance the resilience of the models and reduce the risk of overfitting to the data at hand. Transposed convolutions were used for upsampling in the decoder, while skip connections preserved spatial details across encoder and decoder stages. The implementation was carried out using PyTorch with GPU acceleration where available, a batch size of 4, a learning rate of 1×10^{-4}, and the Adam optimiser. The training

was performed for 50 epochs, and the Dice and Binary Cross-Entropy (BCE) loss function was used for optimisation, while performance was evaluated using the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and accuracy.

4.4 Results

The EfficientNet-UNet model showed outstanding results on the validation dataset from ISIC 2018. The results obtained mainly were: Dice Similarity Coefficient (DSC): 0.9234, Intersection over Union (IoU): 0.8577, and accuracy: 0.9651. These results show that the predicted segmentation masks have a substantial overlap with the actual masks, showing they are able to draw the lesion boundaries effectively, as indicated by the Dice score. The IoU, which is closely related to the Dice score (Dice = 2*IoU/(IoU+1)), reinforced the quality of segmentation, and the high accuracy indicated good classification of pixels, albeit it was less indicative of the accuracy of segmentation than the Dice score, because of the high background contribution in medical images.

Visual results, as shown in Figure 2, verified that the model could identify boundaries of the lesions in most of the examples, with minor inaccuracies at the edges or smaller lesions. The segmentation probability maps produced by the last transposed convolutional layer had smooth gradients and high confidence in the lesion regions, as shown in the output of the sigmoid activation function. The findings indicated that the architectural improvements, including the feature extraction of EfficientNet and the upsampling strategy of U-Net, are able to overcome the challenges of medical image segmentation.

The relationship among the computed metrics conforms to expected mathematical relationships. The Dice score (F1-score) of 0.9234 is in line with the calculated IoU through the recognized conversion. This consistency confirms correct metric implementation. This high accuracy rate of 96.51% is normal for such medical image segmentation problems due to the dominance of background pixels in the evaluation, although such a metric provides little insight into the quality of the segmentation result.

4.5 Comparative Performance Analysis Compared to Previously Published Results on the ISIC 2018 Dataset

As illustrated in Table 1, our EfficientNet-UNet demonstrates comparable performance (Dice: 0.9234) with significantly fewer parameters than the transformer-based DuAT (0.923) and without reliance on polar transformations, unlike Polar Res-U-Net++ (0.9253). This approach is feasible in situations where computational capabilities are constrained. However, the integration of lightweight attention strategies (such as DF-RAM in RMSM-UNet [38]) could potentially narrow the existing performance gap.

Table 1. Comparison of the proposed method with previously reported segmentation models on the ISIC 2018 dataset.

Model	Dice	IoU	Key Features	Authors
Polar Res-U-Net++	0.9253	–	Polar image transformations	Benčević et al. [39]
DuAT	0.9230	0.8670	Dual-aggregation transformer	Tang et al. [40]
RMSM-UNet + DF-RAM	0.9152	–	Dual-attention mechanism	Chowdary et al. [38]
BAT	0.9120	0.8430	Boundary-aware transformer	Wang et al. [41]
EfficientNet -UNet	**0.9234**	**0.8577**	**EfficientNet-B0 backbone**	**Proposed method**

As shown in Table 1, the proposed EfficientNet-UNet achieved a Dice score of 0.9234 and an IoU of 0.8577 on the validation set. The relationship between these two metrics is consistent with the theoretical conversion between Dice and IoU, where $\mathrm{IoU} = \mathrm{Dice}/(2 - \mathrm{Dice})$. The proposed model demonstrates segmentation performance comparable to previously reported methods on the ISIC 2018 dataset, including Polar Res-U-Net++, DuAT, RMSM-UNet with DF-RAM, and BAT. In particular, its Dice score is close to those reported for Polar Res-U-Net++ (0.9253) and DuAT (0.9230), indicating competitive segmentation performance. However, these comparisons should be interpreted with caution because the reported results were obtained under potentially different data partitions, preprocessing procedures, training strategies, and evaluation protocols. Therefore, the values presented in Table 1 are intended to provide contextual comparison rather than a strictly controlled head-to-head ranking.

However, several limitations and opportunities for further improvement were identified. Although the proposed model achieved a Dice score of 0.9234, further improvements in segmentation performance remain possible, particularly for challenging cases involving small lesions and low-contrast boundaries. Minor fluctuations observed in the validation loss may reflect variations in lesion appearance, size, and image quality within the ISIC 2018 dataset. Furthermore, pixel accuracy alone may not adequately reflect

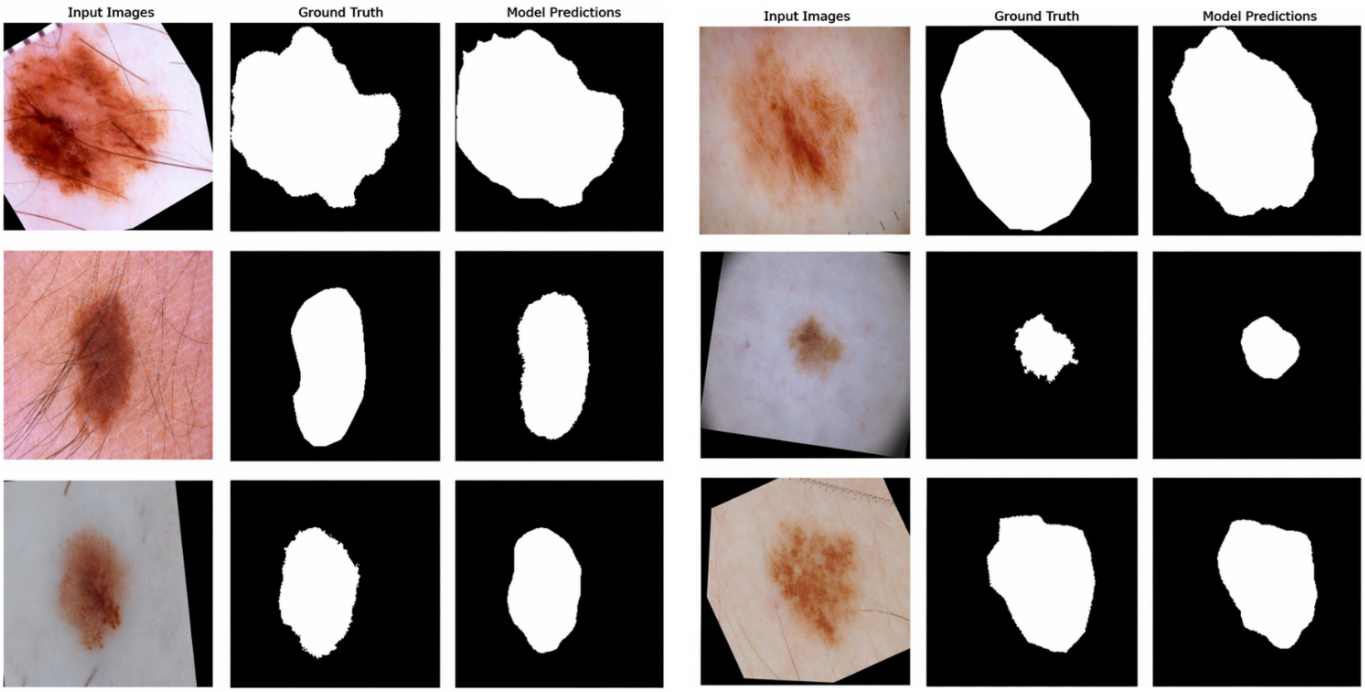


Figure 2. Sample segmentation output from the EfficientNet-UNet model on the ISIC 2018 dataset.

segmentation errors in lesion regions because of the class imbalance between lesion and background pixels. Qualitative inspection of the predicted masks showed that some small lesions and poorly defined boundaries were not segmented accurately. Future improvements could therefore investigate alternative loss functions, boundary-aware learning strategies, and appropriate post-processing techniques to improve segmentation of challenging lesion regions. Furthermore, the performance metrics reported in this study were obtained on the validation set; future work should evaluate the model on a fully held-out test partition to provide a more rigorous and unbiased assessment of generalization performance.

Future work should evaluate the proposed framework using larger and more diverse datasets and investigate deeper EfficientNet variants, such as EfficientNet-B1 and EfficientNet-B2, to determine whether additional model capacity can further improve segmentation performance. Incorporating attention mechanisms or boundary-aware learning strategies may also improve the delineation of small lesions and low-contrast boundaries. In addition, controlled ablation studies comparing EfficientNet-based encoders with conventional U-Net and ResNet-based encoders would help quantify the specific contribution of compound scaling and efficient convolutional operations to segmentation performance. Further evaluation on independent external datasets is also necessary to assess the generalizability of the proposed model. Finally, deployment-oriented experiments measuring model size, inference latency, memory consumption, and computational cost on resource-constrained or edge devices would provide a more rigorous assessment of its practical applicability.

5 Conclusion

In this work, we investigated the efficiency of our proposed EfficientNet (combined with U-Net's localizing ability) architecture for the task of skin lesion segmentation. By using features extracted by EfficientNet as input for localizing the lesion effectively, the model was able to identify and predict skin lesions using Dice, IoU, and pixel accuracy as proof of its capacity to effectively localize skin lesions. Adding transposed convolution and regularization provided consistent and generalizable results to the model while addressing issues of class imbalance and overfitting.

The compound scaling approach in the EfficientNet encoder helped in extracting meaningful features from input images, which prevented the model from becoming too computationally expensive. Though the model had fewer parameters, high accuracy was observed throughout training. As far as decoding is concerned, the use of transposed convolutions and skip connections helped restore spatial information in the input images, which allowed the model to determine boundary lesions with the accuracy required by the task. The generalization capabilities of the model also

improved with the addition of data augmentation and regularization techniques (e.g., dropout and batch normalization). The model achieved an overall Dice score of 0.9234 when applied to ISIC-2018 data; therefore, it appears to be a successful approach for segmenting skin lesions using machine learning techniques. The proposed model has the potential for significant utility in clinical practice.

For future studies, some areas for research will include expanding the ISIC-2018 dataset, evaluating at a more in-depth level the performance of the developed EfficientNet models, and testing the degree of improvement to classification accuracy by employing various forms of attention mechanism(s).

Data Availability Statement

The ISIC 2018 skin lesion segmentation dataset used in this study is publicly available through the International Skin Imaging Collaboration (ISIC) archive at https://challenge.isic-archive.com. The trained model weights and other data generated or analysed during this study are available from the corresponding author upon reasonable request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable. This study used only publicly available, de-identified dermoscopic image data (ISIC 2018) and did not involve direct patient recruitment, clinical intervention, or collection of new human subject data.

References

[1] Arnold, M., Singh, D., Laversanne, M., Vignat, J., Vaccarella, S., Meheus, F., ... & Bray, F. (2022). Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA dermatology, 158*(5), 495-503. [CrossRef]

[2] Liu, M., Wu, M., Liu, X., Zhou, J., Lan, Y., Zhang, H., ... & Li, J. (2023). Assessing the quality of care for skin malignant melanoma on a global, regional, and national scale: a systematic analysis of the global burden of disease study from 1990 to 2019. *Archives of Dermatological Research, 315*(10), 2893-2904. [CrossRef]

[3] Mirikharaji, Z., Abhishek, K., Bissoto, A., Barata, C., Avila, S., Valle, E., ... & Hamarneh, G. (2023). A survey on deep learning for skin lesion segmentation. *Medical Image Analysis, 88*, 102863. [CrossRef]

[4] Mustafa, S., Jaffar, A., Rashid, M., Akram, S., & Bhatti, S. M. (2025). Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet-Random Forest for enhanced segmentation and classification. *PloS one, 20*(1), e0315120. [CrossRef]

[5] Manzoor, K., Gilal, N. U., Agus, M., & Schneider, J. (2025). Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. *Digital health, 11*, 20552076251351858. [CrossRef]

[6] Schmidt, A., Morales-Alvarez, P., & Molina, R. (2023, October). Probabilistic modeling of inter-and intra-observer variability in medical image segmentation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 21040-21049). IEEE. [CrossRef]

[7] Abdelrahman, A., & Viriri, S. (2023). EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images. *Frontiers in Computer Science, 5*, 1235622. [CrossRef]

[8] Bachanek, S., Wuerzberg, P., Biggemann, L., Janssen, T. Y., Nietert, M., Lotz, J., ... & Uhlig, J. (2025). Renal tumor segmentation, visualization, and segmentation confidence using ensembles of neural networks in patients undergoing surgical resection. *European Radiology, 35*(4), 2147-2156. [CrossRef]

[9] Manole, I., & Tiplica, G. S. (2024). Integrating Artificial Intelligence in dermatology: progress, challenges and perspectives. *Romanian Medical Journal, 71*(2), 183. [CrossRef]

[10] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis, 42*, 60-88. [CrossRef]

[11] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence, 44*(7), 3523-3542. [CrossRef]

[12] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing. [CrossRef]

[13] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017).

Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. [CrossRef]

[14] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods, 18*(2), 203-211. [CrossRef]

[15] Su, R., Zhang, D., Liu, J., & Cheng, C. (2021). Msu-net: Multi-scale u-net for 2d medical image segmentation. *Frontiers in Genetics, 12*, 639930. [CrossRef]

[16] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, 6105-6114. https://proceedings.mlr.press/v97/tan19a.html?ref=ji

[17] Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., & Johansen, H. D. (2020, July). Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)* (pp. 558-564). IEEE. [CrossRef]

[18] Trinh, M. N., Nguyen, V. D., Pham, V. T., & Tran, T. T. (2023). An EffcientNet-encoder U-Net joint residual refinement module with Tversky–Kahneman Baroni–Urbani–Buser loss for biomedical image segmentation. *Biomedical Signal Processing and Control, 83*, 104631. [CrossRef]

[19] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging, 39*(6), 1856-1867. [CrossRef]

[20] Ummadi, V. (2022). U-Net and its variants for Medical Image Segmentation: A short review. *arXiv preprint arXiv:2204.08470*. [CrossRef]

[21] Abedalla, A., Abdullah, M., Al-Ayyoub, M., & Benkhelifa, E. (2021). Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures. *PeerJ Computer Science, 7*, e607. [CrossRef]

[22] Vong, K. M., & Dinh, T. B. (2021). Pneumothorax segmentation in chest X-rays using UNet++ and EfficientNet. *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 1-4. [CrossRef]

[23] Ruan, J., Xie, M., Gao, J., Liu, T., & Fu, Y. (2023, October). Ege-unet: an efficient group enhanced unet for skin lesion segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 481-490). Cham: Springer Nature Switzerland. [CrossRef]

[24] Lesion Segmentation on ISIC 2018 (Ranking from SOTA). (n.d.). Retrieved March 2025, from Papers with Code. https://paperswithcode.com/sota/lesion-segmentation-on-isic-2018

[25] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020, June). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (pp. 11531-11539). IEEE. [CrossRef]

[26] Li, X., Zhu, W., Dong, X., Dumitrascu, O. M., & Wang, Y. (2025, April). EViT-Unet: U-Net like efficient vision transformer for medical image segmentation on mobile and edge devices. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE. [CrossRef]

[27] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*. [CrossRef]

[28] Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentation based on U-net: A review. *Journal of Imaging Science & Technology, 64*(2), 1. [CrossRef]

[29] Siddique, N., Sidike, P., Elkin, C., & Devabhaktuni, V. (2020). U-Net and its variants for medical image segmentation: theory and applications. *arXiv preprint arXiv:2011.01118*. [CrossRef]

[30] Park, M., Oh, S., Park, J., Jeong, T., & Yu, S. (2025). ES-UNet: efficient 3D medical image segmentation with enhanced skip connections in 3D UNet. *BMC medical imaging, 25*(1), 327. [CrossRef]

[31] Lee, H., Park, J., & Hwang, J. Y. (2020). Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 67*(7), 1344-1353. [CrossRef]

[32] Zhang, C., Deng, X., & Ling, S. H. (2024). Next-gen medical imaging: U-Net evolution and the rise of transformers. *Sensors, 24*(14), 4668. [CrossRef]

[33] Asadi-Aghbolaghi, M., Azad, R., Fathy, M., & Escalera, S. (2020). Multi-level context gating of embedded collective knowledge for medical image segmentation. *arXiv preprint arXiv:2003.05056*. [CrossRef]

[34] Tiwary, P. K., Johri, P., Katiyar, A., & Chhipa, M. K. (2025). Deep learning-based MRI brain tumor segmentation with EfficientNet-enhanced UNet. *IEEE Access, 13*, 54920-54937. [CrossRef]

[35] Rais, K., Amroune, M., Haouam, M. Y., & Benmachiche, A. (2025). Enhancing medical image analysis through geometric and photometric transformations. *arXiv preprint arXiv:2501.13643*. [CrossRef]

[36] Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in neural information processing systems, 32*.

[37] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data, 5*, 180161. [CrossRef]

[38] Chowdary, G. J., Yathisha, G. V. S. N. D., Suganya, G., & Premalatha, M. (2021). Automated skin lesion segmentation using multi-scale feature extraction scheme and dual-attention mechanism. *arXiv preprint arXiv:2111.08708*. [CrossRef]

[39] Benčević, M., Galić, I., Habijan, M., & Babin, D. (2021). Training on polar image transformations improves biomedical image segmentation. *IEEE Access*, *9*, 133365–133375. [CrossRef]

[40] Tang, F., Xu, Z., Huang, Q., Wang, J., Hou, X., Su, J., & Liu, J. (2023, October). DuAT: Dual-aggregation transformer network for medical image segmentation. In *Chinese conference on pattern recognition and computer vision* (*PRCV*) (pp. 343-356). Singapore: Springer Nature Singapore. [CrossRef]

[41] Wang, J., Wei, L., Wang, L., Zhou, Q., Zhu, L., & Qin, J. (2021, September). Boundary-aware transformers for skin lesion segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 206-216). Cham: Springer International Publishing. [CrossRef]



Maame G. Asante-Mensah holds a Bachelor's degree in Information Technology from the University of Cape Coast, Ghana, and a Master of Engineering in Computer Science from the University of Electronic Science and Technology of China. She earned her PhD in Computational and Data Science and Engineering from the Skolkovo Institute of Science and Technology (Skoltech), Russia, where her research focused on deep learning and tensor completion for biomedical signal reconstruction. She is currently a lecturer in the Department of Computer Science and Information Technology at the University of Cape Coast. Her research interests include artificial intelligence, deep learning, computer vision, tensor completion, and randomized algorithms for multilinear algebra. (Email: gasante-mensah@ucc.edu.gh)



Charles Roland Haruna obtained his Doctor of Engineering in Computer Science and Technology from the University of Electronic Science and Technology of China in June 2020. He is a Senior Lecturer in the Department of Computer Science and Information Technology at the University of Cape Coast, Ghana, where he teaches both undergraduate and postgraduate courses. In 2022, he established the Big Data Research Lab, focusing on research in big data technological infrastructure, data repositories, and applications. He has published extensively in international journals and conference proceedings and serves as a peer reviewer for IEEE Access as well as a Program Committee member for Frontiers in Cyber Security conferences. His research interests are Artificial Intelligence for Big Data Management, DB4AI, Distributed Database Management Systems, Query Optimization, SQL similarity metrics. (Email: charuna@ucc.edu.gh)



Stephen Julius Atsu is an M.Phil. student in Computer Science at the University of Cape Coast (UCC), Ghana. He currently serves as an ICT instructor at St. Paul's Senior High School and Minor Seminary. His research interests include artificial intelligence, machine learning, computer vision, database systems, and cybersecurity, with a particular focus on performance optimization and intelligent systems. He has developed practical software solutions for educational institutions and is actively engaged in research on hybrid indexing strategies, computer vision and machine learning. (Email: steveatsu@gmail.com)



Sandro Kwame Amofa is affiliated with the Department of Computer Science and Information Technology at the University of Cape Coast, Cape Coast, Ghana. He received his training in computer science and has research experience spanning blockchain technology, secure health data sharing, and applied machine learning. His prior work includes contributions to blockchain-based architectures for personal health record management, decentralized identity systems for smart cities, and privacy-preserving data sharing frameworks, published in IEEE and Springer venues. His current research interests include deep learning, medical image analysis, and the application of intelligent computing systems in healthcare. (Email: sandro.amofa@ucc.edu.gh)



Akua Gyanba Morgan Acquah is a Research Associate and a Ph.D. student at the Department of Computer Science and Information Technology at the University of Cape Coast, Cape Coast, Ghana. She holds a bachelor's degree in Information Technology and a master's in Computer Science degree from the same university. (Email: abiney@ucc.edu.gh)