APPLICATION OF AI USING U-NET IN SKIN LESION SEGMENTATION

Serra Aksoy

Abstract:

Skin lesion segmentation is a critical task in dermatology, essential for the accurate diagnosis and treatment of various skin conditions, including skin cancer. The precise identification of lesion boundaries in medical images significantly helps in early detection and effective management of these conditions. In this study, a U-Net model was employed to perform segmentation of skin lesions, using its advanced encoder-decoder architecture and skip connections to capture fine details and spatial hierarchies within the images. The model yielded an overall accuracy of 95.47%, a precision of 97.21%, a recall of 84.04%, and an F1 score of 90.15%. These results affirm the U-Net model's proficiency in accurately segmenting skin lesions. The precise boundary delineation provided by the model can help healthcare professionals detect malignant lesions, which often exhibit irregular boundaries, thereby improving diagnostic accuracy. The integration of this model into clinical practice can enhance diagnostic accuracy and efficiency, reduce the workload on healthcare professionals, and improve patient outcomes. The promising performance of the U-Net model emphasizes its potential to revolutionize dermatological diagnostics and support healthcare professionals in delivering timely and precise patient care.

Keywords: *skin lesion, segmentation, U-Net, deep learning, dermatology, convolutional neural networks (CNNs)*

1. Introduction

Skin lesion segmentation is the creation of a detailed map of the skin's landscape to guide dermatologists in understanding and treating conditions like melanoma and basal cell carcinoma accurately. Just as a cartographer carefully marks boundaries and landmarks on a map, segmentation identifies and outlines lesions within skin images. This helps dermatologists gauge the seriousness of diseases, monitor how they change over time, and determine if treatments are working effectively. They use a mix of traditional methods like setting thresholds and finding edges, and cutting-edge techniques, such as deep learning with CNNs, U-Net, and Mask R-CNN, which act like advanced GPS systems for medical images. Challenges include the diverse appearances of lesions, noisy image data, and the need for algorithms that can handle large amounts of data efficiently.

Successfully segmenting lesions not only aids in making informed medical decisions but also drives forward research and improves remote medical consultations by providing clear and reliable information for analysis.

2. Related Works

The skin, our body's largest organ, plays a vital role as our main barrier against the outside world, covering about 16% of our total body weight. Because it's constantly exposed to the environment, the skin is susceptible to a range of issues, from common abnormalities like lesions to various appendages [1, 2]. Skin cancer, including types like squamous cell carcinoma, basal cell carcinoma, and malignant melanoma, poses a serious global health threat. Among these, malignant melanoma is particularly aggressive and is a leading cause of cancer-related deaths worldwide. In 2020 alone, estimates from the American Cancer Society indicated around 100,350 new cases of skin cancer and more than 6,500 deaths attributed to this disease. These numbers underscore the urgent need for effective prevention, early detection, and treatment strategies to combat this significant health issue [1, 3].

Skin lesion segmentation is crucial in medical diagnostics, playing a role in detecting and treating cancer early. However, accurately delineating skin lesions presents challenges due to their diverse appearances and sometimes ambiguous borders, especially in benign or nevus lesions. This complexity makes distinguishing them from healthy skin tissues challenging, underscoring the necessity for advanced segmentation techniques that can precisely identify and differentiate different types of skin lesions.

In recent years, deep learning-based algorithms for image segmentation have made significant advancements over traditional methods, aiming to enhance medical diagnostics by reducing subjective factors and improving efficiency in clinical settings [4–7]. Researchers have focused on developing efficient and stable segmentation algorithms to improve performance. These efforts are crucial for accurately identifying and distinguishing various medical conditions from imaging data [8–10].

The primary frameworks for medical image segmentation today include decoder-encoder structures like U-Net and DeepLab, which leverage dilated convolutions for handling different scales and improving object boundary accuracy [11, 12].

U-Net, known for its simplicity and utility, serves as a foundation in the field but has limitations in complexity and computational demands, particularly when integrating more complex structures like UNet++ and ResUNet++ [13–15]. Models such as DeepLabV3 and DeepLabV3++, based on dilated convolutions, aim to address larger-scale targets but often require substantial data and computational resources for effective training [16, 17]. Multiscale feature fusion is crucial for enhancing segmentation accuracy, blending local and global information through structures like FPN [18], MANet [19], LinkNet [20], and PSPNet [21], which integrate information across different levels to improve overall segmentation quality. Despite these advancements, challenges remain in optimizing feature extraction algorithms, balancing shallow high-resolution information loss, and addressing bottlenecks in encoder layers where vital contextual details are aggregated. This is critical for advancing segmentation performance in medical imaging.

Researchers are using traditional machine learning techniques to improve how medical images are segmented, focusing on making these methods more accurate. Jaisakthi et al. [22] have developed a method that combines Grab-cut and K-means clustering to outline skin lesions like melanoma. They've also added steps to enhance image quality, like smoothing out irregularities and standardizing lighting before analyzing the images. Meanwhile, Mohanad Aljanabi et al. [23] have come up with a technique that uses artificial bee colonies to find the best way to distinguish between healthy skin and lesions. This method uses fewer steps and can pinpoint lesions with a lot of precision, which makes it a powerful tool for diagnosing skin conditions. Both approaches show how researchers are finding new ways to use machine learning to help doctors get more accurate results when they're looking at medical images. Berseth et al. [24] created a special type of computer program called a U-Net to help doctors find skin problems in pictures of skin. They used a method called ten-fold cross validation to make sure their program was accurate. Meanwhile, Mishra et al. [25] came up with a different way to use computers to find skin problems in pictures of skin.

Properly diagnosing and screening for skin disorders relies heavily on accurately pinpointing the affected areas. Segmentation of skin lesions is tricky due to factors like their diverse shapes, their closeness to normal skin, and the presence of hair. To tackle these challenges effectively, we introduce a cutting-edge method that uses deep learning to automate the segmentation of skin lesions. This innovative approach aims to simplify and improve the accuracy of diagnosing skin conditions, ensuring better care and outcomes for patients dealing with dermatological issues.

Table 1. Train and Validation Data

	Train	**Validation**	**Total**
Images	8,016	1,999	10,015
Masks	8,016	1,999	10,015
Total	16,032	3,998	20,030

3. Material and Method

3.1. Data Preparation and Augmentation

The HAM10000 dataset, consisting of 10,015 dermoscopic images of skin lesions along with their corresponding segmentation masks, is used in this study.

This dataset includes a comprehensive collection of critical diagnostic categories of pigmented lesions, such as actinic keratoses and intraepithelial carcinoma (AKIEC), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC).

For effective model training and validation, the dataset is divided into two subsets: a training set and a validation set. The training set consists of 8,016 images and their respective masks, while the validation set comprises the remaining 1,999 images and masks. This split ensures a robust training of the model, giving the model the chance to see a lot of data and evaluation of the model's performance on unseen data (Table 1).

The images and their corresponding masks are loaded and preprocessed using a custom dataset class, 'SkinLesionDataset'. This class is designed to facilitate the retrieval of image-mask pairs from specified directories. Within the class, the '__init__' method initializes the dataset by specifying the directories for images and masks, along with any transformations to be applied. The '__len__' method returns the total number of images, and the '__getitem__' method retrieves the image and mask at the specified index. The masks are converted to binary format, where the pixel value of 255 is replaced with 1 for compatibility with the segmentation task.

To enhance the model, data augmentation techniques are employed on the training images. The transformations applied to the training set include resizing the images to 224x224, rotation with a probability of 1.0, horizontal flipping with a probability of 0.5, vertical flipping with a probability of 0.1, and normalization with mean and standard deviation set to [0.0, 0.0, 0.0] and [1.0, 1.0, 1.0] respectively, with the maximum pixel value set to 255.0.

For the validation set, a simpler augmentation pipeline is employed to maintain consistency during model evaluation. This pipeline involves resizing the images to the same dimensions as the training images and normalizing the pixel values using the same mean, standard deviation, and maximum pixel value settings as the training set. These preprocessing and augmentation steps are crucial for improving the model's generalization capability and performance in accurately segmenting skin lesions.
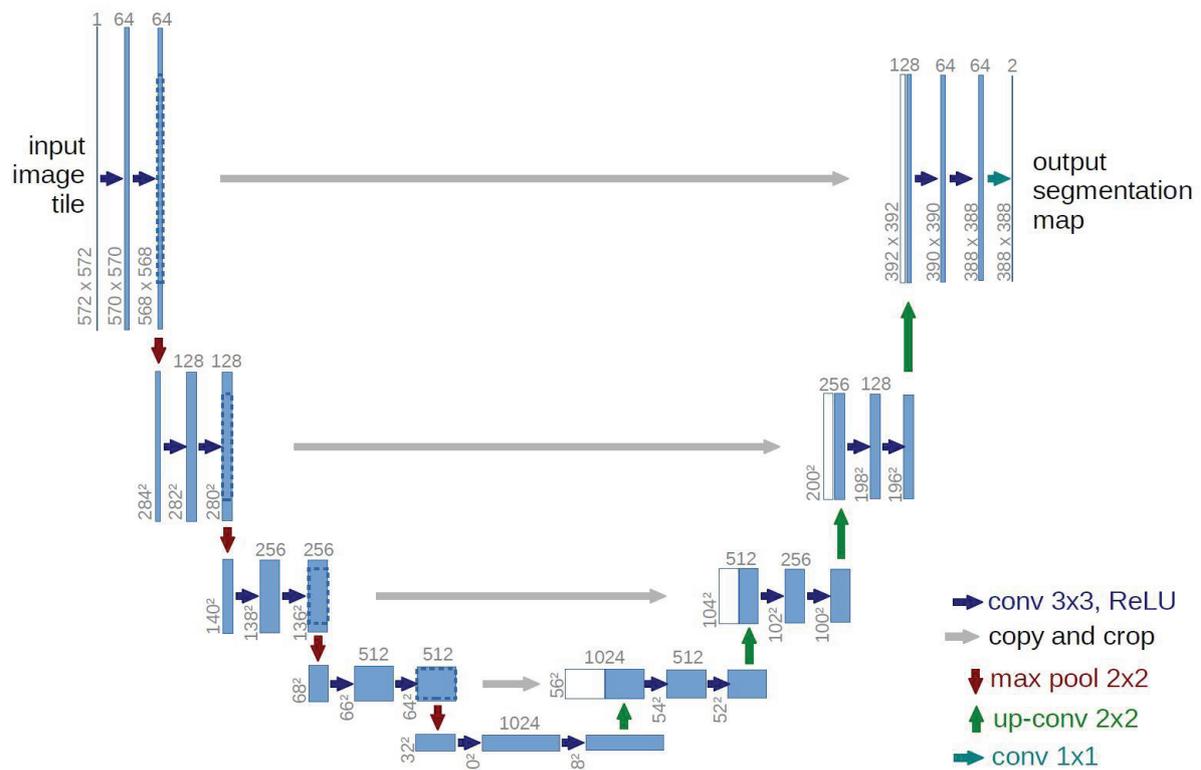
Figure 1. U-Net Architecture

3.2. Proposed Model

The U-Net architecture was chosen for its well-established effectiveness in biomedical image segmentation. The model features a symmetric encoder-decoder structure that excels in precise localization through its skip connections.

These connections enable the integration of detailed, high-resolution features from the encoder with upsampled features in the decoder, which is crucial for accurate skin lesion segmentation. While Fully Convolutional Networks (FCNs) offer end-to-end segmentation, they may lose fine-grained details because they lack skip connections. The DeepLab models use atrous convolutions and spatial pyramid pooling to capture multi-scale context but are computationally intensive and require large datasets. SegNet is efficient in memory usage through its use of max-pooling indices, though it might compromise on localization accuracy. Attention U-Net, an advanced version of U-Net, incorporates attention mechanisms to highlight relevant features but adds complexity. Despite these alternatives, U-Net remains the optimal choice for its proven performance and suitability in managing high-resolution medical images, ensuring robust and accurate segmentation crucial for early diagnosis and treatment of skin conditions.

The proposed model for segmenting skin lesions is the U-Net architecture, a highly effective convolutional neural network (CNN) designed specifically for biomedical image segmentation tasks. The U-Net architecture is characterized by its symmetric encoder-decoder structure, which facilitates precise localization required for accurate segmentation. The encoder, or contracting path, is responsible for capturing the context of the input image by progressively downsampling it through a series of layers.

Each layer comprises two 3x3 convolutional operations followed by batch normalization and ReLU activations and concludes with a 2x2 max-pooling operation that halves the spatial dimensions while doubling the number of feature channels. This downsampling process allows the network to capture increasingly abstract and high-level features of the image.

On the other hand, the decoder, also called the expansive path, strives to recreate the spatial information of the input image. This involves upsampling layers using transposed convolutional layers, followed by the concatenation of the features from the encoder network using skip connections. These skip connections are significant as they allow the model to retain the high-resolution information that is discarded during the downsampling process. After concatenation, each upsampling step includes two 3x3 convolutional operations with batch normalization and ReLU activations, similar to the encoder. This process reduces the number of feature channels while increasing the spatial dimensions, effectively recovering the original image size (Figure 1).

At the deepest part of the network, the bottleneck layer processes the smallest spatial representation with the highest number of feature channels, implemented using the same double convolution approach. The final layer of the U-Net architecture employs a 1x1 convolution to map the feature maps to a single output channel, which is essential for binary segmentation tasks. The output is then passed through a sigmoid

activation function, which squashes the values to the range [0, 1], producing a probability map that indicates the likelihood of each pixel belonging to the lesion.

For this segmentation task, the masks are converted to a binary format where pixel values of 255 are replaced with 1, ensuring compatibility with the binary classification required by the model. The sigmoid activation function's output is thresholded at 0.5 to produce a binary mask, where pixels with a probability greater than 0.5 are classified as part of the lesion (assigned a value of 1), and those with a probability of 0.5 or less are classified as background (assigned a value of 0).

By using the U-Net architecture and incorporating appropriate data augmentation techniques, the proposed model aims to achieve high accuracy and robustness in segmenting skin lesions. This robust segmentation capability is crucial for facilitating early diagnosis and treatment of various skin conditions, ultimately contributing to improved patient outcomes.

3.3. Experimental Setup

The experimental setup was conducted on a high-performance computing environment equipped with an Intel Core i9 processor and an NVIDIA GeForce RTX GPU. This setup ensured the efficient handling of the computational demands of deep learning tasks, particularly the extensive matrix operations required for training convolutional neural networks. The experiments were programmed using PyTorch, a widely used deep learning framework known for its dynamic computational graph and flexibility, which allowed the implementation and experimentation with the U-Net architecture.

Data preparation began with the HAM10000 dataset, which consists of 10,015 dermatoscopic images of various skin lesions and their corresponding segmentation masks. The dataset was divided into training and validation sets, with 8,016 images allocated for training and 1,999 images for validation.

A custom dataset class, 'SkinLesionDataset' retrieves the image-mask pairs, converting the masks to a binary format where pixel values of 255 are replaced with 1 to ensure compatibility with the segmentation task. The training set was augmented using rotation, vertical and horizontal flips, and transformed similarly to the validation set, while the validation set was only resized, converted to tensors, and normalized. The train and validation data loaders were created with a batch size of 32. The U-Net architecture was employed for the segmentation task. During model training, the binary cross-entropy loss function was used to optimize the model parameters, minimizing the difference between the predicted probability map and the ground truth mask. The optimization process used the Adam optimizer with a learning rate of 0.001. The training process involved iterating over batches of training data, where the images and masks were loaded, augmented, and passed through the model to generate predictions. The predictions were thresholded at 0.5 to produce

binary masks, which were then compared with the ground truth masks to compute evaluation metrics (Figure 2).

Validation of the results involved carefully classifying the model's predictions into categories such as True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). This was done by comparing each pixel in the predicted binary masks with the corresponding pixel in the ground truth masks. TP referred to pixels correctly identified as part of the lesion, FP to pixels mistakenly identified as part of the lesion, TN to correctly identified background pixels, and FN to pixels where the lesion was missed. The initial classification was handled automatically through Python scripts, which compared every pixel's prediction to its true label.

To ensure accuracy, this automated process was supplemented with manual checks where experts reviewed selected results to confirm the classifications and ensure they genuinely reflected the model's performance. This thorough approach ensured that metrics like accuracy and the Dice coefficient accurately represented the model's segmentation abilities and provided reliable insights into its performance.

Model performance was assessed using accuracy and the Dice coefficient. Accuracy measured the proportion of correctly predicted pixels (both lesion and background) out of the total number of pixels, providing a basic measure of overall correctness. The Dice coefficient, also known as the F1 score, was particularly useful for evaluating segmentation tasks, as it measured the overlap between the predicted mask and the ground truth mask. The Dice coefficient was calculated as twice the overlap between the predicted and ground truth masks, divided by the total number of pixels in both masks (Equation 1). Furthermore, the confusion matrix was visualized as a heatmap, with ground truth labels on the rows and predicted labels on the columns, which highlighted where the model succeeded, which means high TP and TN and where it struggles, which is high FP and FN in accurately segmenting objects. Unlike traditional classification tasks where predictions are made for entire instances, segmentation involves predicting labels for each pixel. The confusion matrix for segmentation, known as a pixel-wise confusion matrix, categorizes predictions into four groups: True Positives (correctly identified object pixels), False Positives (incorrectly identified as object pixels), True Negatives (correctly identified background pixels), and False Negatives (incorrectly identified as background pixels). In addition to the confusion matrix, plots depicting metrics like Dice score, training loss, accuracy, and validation loss and accuracy are essential. Together, these visualizations offer a comprehensive view of a segmentation model's progression, helping to refine and optimize it for more precise and dependable segmentations in applications.
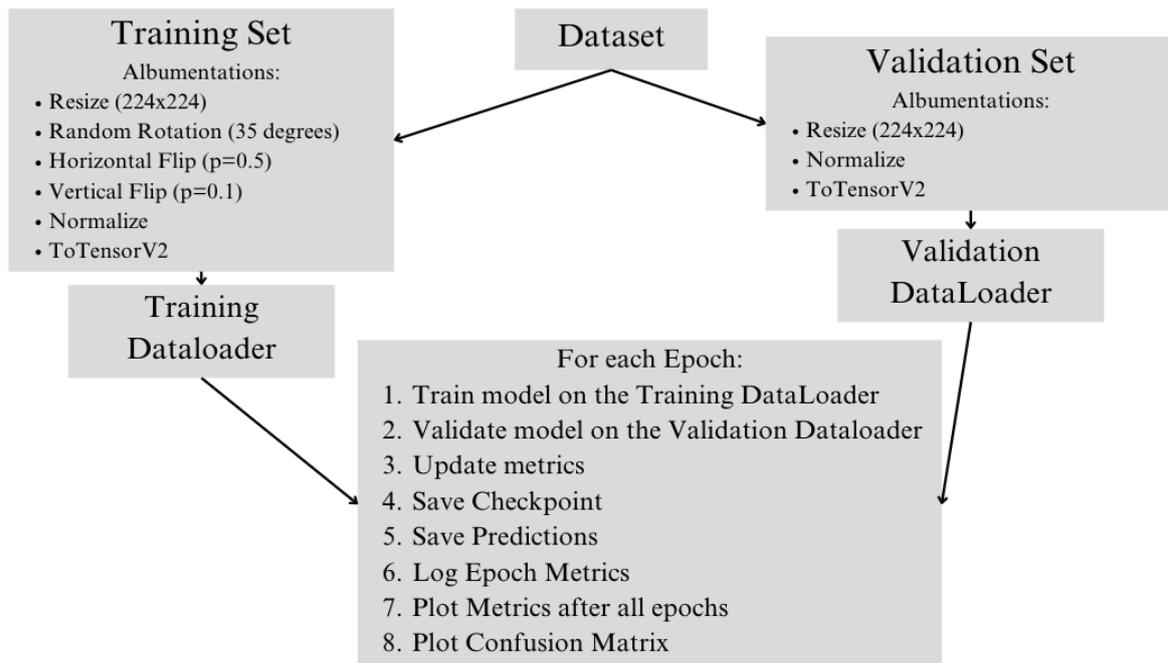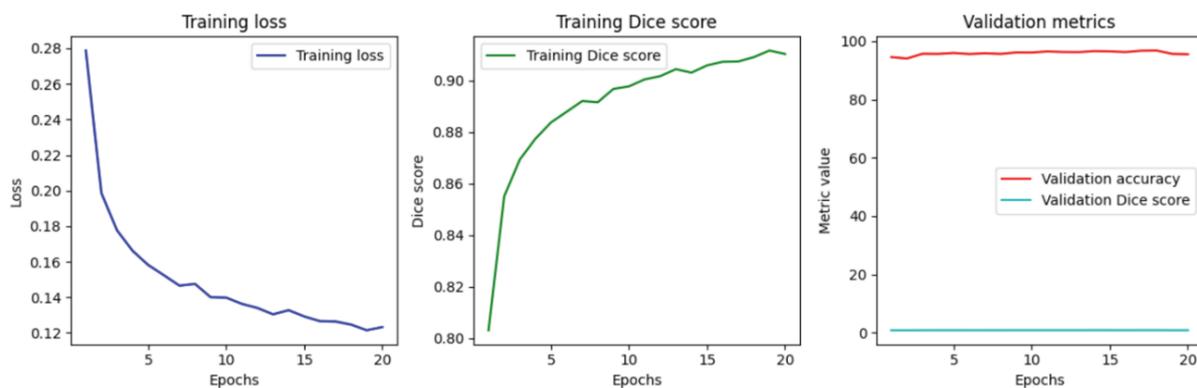
Figure 2. Experimental Setup

Figure 3. Training Loss, Dice Score, and Validation Metrics over 20 Epochs

4. Results and Discussion

4.1. Results

The performance of the U-Net model for skin lesion segmentation was evaluated over 20 epochs, employing a comprehensive set of metrics to monitor and analyze its effectiveness. Throughout the training process, a consistent and substantial decrease in training loss was observed, starting from an initial value of 0.2788 in the first epoch and progressively reducing to 0.1232 by the twentieth epoch. This steady decline indicates the model's increasing ability to minimize prediction errors effectively. Concurrently, the training Dice score exhibited continuous improvement, beginning at 0.8032 and rising to 0.9102 by the final epoch, emphasizing the model's increasing performance in accurately segmenting skin lesions. Validation metrics demonstrated remarkable stability, with validation accuracy remaining consistently high and validation Dice scores showing minimal fluctuation throughout the training epochs. This stability suggests that the model maintained robust generalization capabilities and effectively avoided overfitting (Figure 3).

The confusion matrix further illustrated the model's performance, showing a high number of true positives (20,781,788) and true negatives (74,978,631), along with relatively lower counts of false positives (595,742) and false negatives (3,945,663). From this confusion matrix, several key classification metrics were derived: an overall accuracy of 95.47%, a precision of 97.21%, a recall of 84.04%, and an F1 score of 90.15%. These metrics collectively indicate that the model achieved a high level of accuracy, demonstrating its reliability in correctly identifying both lesion and non-lesion pixels. The high precision value underscores the model's ability to minimize false positives, thereby ensuring that most of the identified lesions are indeed true lesions. The recall value, reflecting the model's effectiveness in identifying true positive cases, suggests that while some false negatives were
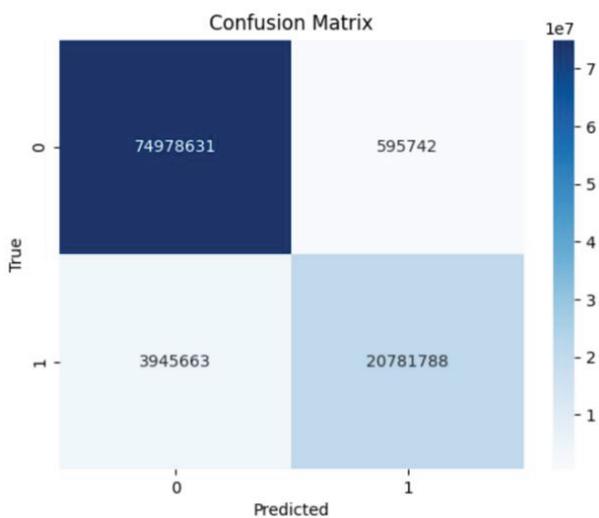
Figure 4. Confusion Matrix for Skin Lesion Segmentation

present, the majority of actual lesions were accurately detected. The F1 score, which balances precision and recall, further confirms the model's robust performance and its proficiency in handling the task of skin lesion segmentation (Figure 4).

Generally, the results affirm the U-Net model's high performance in medical image segmentation, demonstrating both accuracy and reliability in identifying skin lesions. The findings show the model's potential as a valuable tool in clinical settings, where accurate and efficient lesion segmentation is critical for diagnosis and treatment planning.

4.2. Discussion

To judge the model's accuracy, predictions were compared to ground truth masks pixel by pixel, using metrics like accuracy and the Dice coefficient. While these numbers provide a clear picture of how well the model performs in a controlled environment, applying them to real-world settings requires more consideration. Real-world effectiveness depends on how well the model adapts to the variety and complexity of skin conditions seen in everyday clinical practice. High accuracy and Dice scores in testing are promising, but it's essential to evaluate how the model performs with different patient populations and clinical scenarios.

There are some limitations to the proposed approach. The reliance on the HAM10000 dataset means the model might not cover the full range of skin conditions seen in real life, which could affect its performance with less common or atypical cases. While the U-Net architecture is effective for segmentation, it might miss subtle details or struggle with extreme variations in lesion appearance, potentially limiting its robustness outside of the dataset used.

Integrating AI into medical practice requires careful consideration of various factors. Error tolerance is crucial; false positives or negatives can impact patient care, so the AI should act as a supportive tool rather than a replacement for clinical judgment. It should

provide an additional layer of analysis, allowing doctors to make more informed decisions while maintaining compassionate and patient-centered care. The AI's role is to support and enhance medical practice, not to replace the human touch. Explainability is key for both patients and medical professionals. The U-Net model's structure, with its clear encoder-decoder framework, helps make its operations somewhat transparent. However, making the AI's decision-making process more understandable through visual aids like heatmaps or attention maps can help bridge the gap between complex technology and practical use. This transparency helps build trust and ensures that both patients and doctors can effectively use the model.

The benefits of this approach are substantial. By improving the accuracy and efficiency of skin lesion segmentation, the model can support earlier diagnosis and better treatment outcomes. It enables medical professionals to handle large amounts of data more effectively, reducing the risk of diagnostic errors and supporting consistent evaluations. The model's impact can be further enhanced through ongoing improvements, real-world testing, and collaboration with clinicians to ensure it meets practical needs and improves patient care.

5. Conclusion

In conclusion, the U-Net model demonstrated excellent performance in the task of skin lesion segmentation, as evidenced by the consistently decreasing training loss, increasing Dice scores, stable validation metrics, and high classification metrics derived from the confusion matrix. These results support the model's high performance in accurately segmenting medical images and identifying skin lesions.

Looking ahead, the potential applications of this model in healthcare are vast and promising. The deployment of such an advanced segmentation tool can significantly help healthcare professionals in several ways. Firstly, it can enhance the accuracy and efficiency of diagnosing skin conditions, enabling quicker and more precise identification of lesions. This can be particularly beneficial in early detection of skin cancers, where timely diagnosis is crucial for effective treatment and improved patient outcomes.

Additionally, the use of U-Net in automated diagnostic systems can help reduce the workload on dermatologists, allowing them to focus on more complex cases and patient care. By providing reliable segmentation results, the model can serve as a valuable second opinion, assisting clinicians in making informed decisions. This can be especially useful in remote or under-resourced areas where access to specialized medical expertise is limited.

Furthermore, integrating the U-Net model into telemedicine platforms can facilitate remote consultations and continuous monitoring of patients with chronic skin conditions. Patients can capture images of their skin lesions using mobile devices, and the

model can provide immediate analysis, aiding in ongoing disease management and follow-up.

Future research and development should focus on refining the model's performance and expanding its application to other types of medical imaging. Collaborative efforts between data scientists, clinicians, and healthcare professionals are essential to ensure the model's robustness, reliability, and integration into clinical workflows. Ethical considerations, such as patient privacy and data security, must also be addressed to build trust and acceptance among users. Overall, the implementation of advanced models like U-Net in clinical practice holds the potential to revolutionize dermatology and improve patient care through enhanced diagnostic accuracy, efficiency, and accessibility.

Author Contribution SA helped in conceptualization, software, validation, writing—original draft, resources and data curation. SA helped in methodology, writing—review & editing, supervision, revision, funding acquisition and formal analysis.

Data Availability All the datasets used in experimentation are available online on the internet.

Conflict of Interest Authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work submitted for publication.

AUTHOR
Serra Aksoy[*] – Institute of Computer Science, Ludwig Maximilian University of Munich (LMU), Oettingenstrasse 67, 80538 Munich, Germany, e-mail: serra.aksoy@campus.lmu.de, https://orcid.org/0009-0001-2391-9437.

[*]Corresponding author

References

[1] J. Li et al., "DSEUNet: A lightweight U-Net for dynamic space grouping enhancement for skin lesion segmentation," *Expert Systems with Applications*, vol. 255, Art. no. 124544, Dec. 2024; doi: 10.1016/j.eswa.2024.124544.

[2] R. J. Hay et al., "The global burden of skin disease in 2010: An analysis of the prevalence and impact of skin conditions," *Journal of Investigative Dermatology*, vol. 134, no. 6, pp. 1527–1534, Jun. 2014; doi: 10.1038/jid.2013.446.

[3] P. Mathur et al., "Cancer statistics, 2020: Report from national cancer registry programme, India," *JCO Global Oncology*, no. 6, pp. 1063–1075, Nov. 2020; doi: 10.1200/GO.20.00122.

[4] J. Wu, E.Z. Chen, R. Rong, X. Li, D. Xu, and H. Jiang, "Skin lesion segmentation with C-UNet," *Proc. 41st Annu. Int'l Conf. IEEE Eng. Med. Bio. Soc. (EMBC)*, 2019, pp. 2785–2788; doi: 10.1109/EMBC.2019.8857773.

[5] P. Tang et al., "Efficient skin lesion segmentation using separable U-Net with stochastic weight averaging," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 289–301, Sep. 2019; doi: 10.1016/j.cmpb.2019.07.005.

[6] R. Iranpoor, A.S. Mahboob, S. Shahbandegan, and N. Baniasadi, "Skin lesion segmentation using convolutional neural networks with improved U-Net architecture," *Proc. 6th Iranian Conf. Signal Processing and Intelligent Systems (ICSPIS)*, 2020, pp. 1–5; doi: 10.1109/ICSPIS51611.2020.9349577.

[7] L. Liu, L. Mou, X.X. Zhu, and M. Mandal, "Skin lesion segmentation based on improved U-Net," *Proc. IEEE Can. Conf. of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1–4; doi: 10.1109/CCECE.2019.8861848.

[8] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017; doi: 10.1016/j.media.2017.07.005.

[9] L. Qian, H. Huang, X. Xia, Y. Li, and X. Zhou, "Automatic segmentation method using FCN with multi-scale dilated convolution for medical ultrasound image," *Vis. Comput.*, vol. 39, no. 11, pp. 5953–5969, Nov. 2023; doi: 10.1007/s00371-022-02705-w.

[10] G. Zhan, W. Wang, H. Sun, Y. Hou, and L. Feng, "Auto-CSC: A transfer learning based automatic cell segmentation and count framework," *Cyborg Bionic Systems*, vol. 2022, 2022, Art. no. 9842349; doi: 10.34133/2022/9842349.

[11] X. Li, H. Huang, H. Zhao, Y. Wang, and M. Hu, "Learning a convolutional neural network for propagation-based stereo image segmentation," *Vis. Comput.*, vol. 36, no. 1, pp. 39–52, Jan. 2020; doi: 10.1007/s00371-018-1582-y.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation" *arXiv*, 2015; http://arxiv.org/abs/1505.04597

[13] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "U-Net++: A nested U-Net architecture for medical image segmentation," *in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov et al., Eds., *Lecture Notes in Computer Science*, vol. 11045, Springer, 2018, pp. 3–11; doi: 10.1007/978-3-030-00889-5_1.

[14] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018; doi: 10.1109/LGRS.2018.2802944.

[15] D. Jha et al., "ResUNet++: An advanced architecture for medical image segmentation." *arXiv*, 2019; http://arxiv.org/abs/1911.07067

[16] L.C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic Image segmentation." *arXiv*, 2017; http://arxiv.org/abs/1706.05587

[17] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation." *arXiv*, 2018; http://arxiv.org/abs/1802.02611

[18] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection." *arXiv*, 2017; http://arxiv.org/abs/1612.03144

[19] T. Fan, G. Wang, Y. Li, and H. Wang, "MA-Net: A multi-scale attention network for liver and tumor segmentation," *IEEE Access*, vol. 8, pp. 179656–179665, 2020; doi: 10.1109/ACCESS .2020.3025372.

[20] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," *Proc. IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4; doi: 10.1109/VCIP.2017.8305148.

[21] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network." *arXiv*, 2017; http://arxiv.org/abs/1612.01105

[22] S.M. Jaisakthi, P. Mirunalini, and C. Aravindan, "Automated skin lesion segmentation of dermo-scopic images using GrabCut and k-means algo-rithms," *IET Computer Vision*, vol. 12, no. 8, pp. 1088–1095, Dec. 2018; doi: 10.1049/iet‑cvi.20 18.5289.

[23] M. Aljanabi, Y.E. Özok, J. Rahebi, and A.S. Abdul-lah, "Skin lesion segmentation method for der-moscopy images using artificial bee colony algo-rithm," *Symmetry*, vol. 10, no. 8, p. 347, Aug. 2018; doi: 10.3390/sym10080347.

[24] M. Berseth, "ISIC 2017 - Skin lesion analysis towards melanoma detection." *arXiv*, 2017; doi: 10.48550/ARXIV.1703.00523.

[25] R. Mishra and O. Daescu, "Deep learning for skin lesion segmentation," *Proc. IEEE Int'l Conf. Bioin-formatics Biomedicine (BIBM)*, 2017, pp. 1189–1194; doi: 10.1109/BIBM.2017.8217826.