Towards efficient wound management: An automatic Deep Learning model for accurate wound image segmentation

Aleksandar Miladinović[1], Alessandro Biscontin[1], Andrea Bonini[2], Francesco Bassi[2], Simone Kresevic[2], Alessandra Raffini[2], Katerina Iscra[2], Agostino Accardo[2], Miloš Ajčević[2]

1 Institute for Maternal and Child Health IRCCS "Burlo Garofolo", Trieste, Italy

2 Department of Engineering and Architecture, University of Trieste, Trieste, Italy

aleksandar.miladinovic@burlo.trieste.it

Abstract. Untreated wounds may lead to serious complications, underscoring the need for consistent monitoring. In clinical practice, wound imaging offers a potential tool for objective assessment. However, obtaining quantitative analysis requires accurate semantic image segmentation, a task that is often time-consuming. Despite deep learning has shown effectiveness in image segmentation, several challenges remain, especially in occluded or low-light images. This study aims to provide a novel semantic segmentation approach for wound images, designed to overcome some of these limitations. A semantic segmentation framework that combines a Convolutional Neural Network-based encoder with a lightweight multilayer perceptron decoder was proposed. A dataset consisting of 451 images was employed for the training, validation, and testing of a semantic segmentation architecture. During the training phase, the model achieved a pixel accuracy of 99.02%, with 98.82% on the validation set and 98.27% on the test set. The corresponding F1 scores were 0.97 for both training and validation, and 0.95 for the test set. The deigned framework and the produced model provided a valuable solution for automatic wound assessment through precise image segmentation approach. The proposed method offers a remarkable contribution to telemedicine, providing a significant advancement for automatic wound management.

Keywords: Wound healing monitoring, Semantic segmentation, Deep Learning.

1 Introduction

Chronic wounds significantly diminish patients' quality of life [1]. If left untreated, they can result in severe complications, including limb amputation and even death [2]. Regular monitoring and appropriate treatment are crucial in preventing such deterioration. Besides the treatment, effective wound management involves continuous assessment,

which is fundamental to track healing progress and determine the best treatment approach. Typically, this assessment relies on visual inspection of wound characteristics, such as size, depth, and tissue growth. To enhance assessment and decision-making process, clinicians have integrated wound imaging using cameras and smartphones into wound evaluation protocol. These images are acquired during each examination, allowing professionals to track healing progress by comparing them with previous images. However, while wound imaging provides a more objective tool, it does not fully eliminate subjectivity in documentation. In fact, it may introduce additional variability due to the non-quantitative nature of visual comparisons and to external factors.
In recent years, the field of medical image analysis has experienced significant advancements, also in the field of wound assessment. Accurate measurement and segmentation of wound images play a pivotal role in diagnosing, treatments and monitoring healing processes. The complexities inherent in skin ulcer segmentation arise from the diverse characteristics of wounds, including variable shapes, sizes, and surrounding skin conditions [3]. Conventional techniques such as manual segmentation and simple thresholding often fail to meet the accuracy requirements for effective clinical applications, leading to a growing interest in automated, deep learning-based approaches [4]. Various methods have been employed for image segmentation in dermatological applications, predominantly grounded in traditional image processing techniques and machine learning paradigms [5, 6]. Methods such as color-based segmentation, contour detection, and superpixel algorithms have shown some efficacy in isolating wounds from background interference [7]. However, these methods often struggle with generalization across varied datasets and fail to perform well under adverse image conditions, such as non-homogeneous backgrounds or complex lighting [8, 9]. Additionally, many existing models focus more on retrieval and classification tasks rather than delivering the precision required for accurate segmentation [10]. The advent of deep learning has revolutionized the landscape of medical segmentation tasks; architectures like Fully Convolutional Networks (FCNs) and U-Net have gained prominence for their ability to capture intricate spatial relationships and learn semantic mappings between input images and their respective segmentation masks [11]. This has increased segmentation accuracy in challenging environments, establishing a benchmark for wound detection [4]. Despite these advancements, several challenges remain, including dependence on large, well-annotated datasets, susceptibility to misclassification in occluded or low-light conditions, and poor inter-observer agreement [4]. To address this, our approach involves refining a semantic segmentation architecture exploiting Contrast Limited Adaptive Histogram Equalization (CLAHE) as a preprocessing step. CLAHE is instrumental in enhancing image contrast, particularly in medical images, facilitating improved visibility of relevant structures in wound images that might otherwise be obscured [12]. This preprocessing technique enables the network to focus better on the wound areas, thereby improving segmentation outcomes [4]. The integration of CLAHE is expected to mitigate challenges related to illumination variations and enhance the overall robustness of the segmentation model.

The aim of this study is to develop an automated model for accurate wound segmentation, capable of effectively distinguishing wounds and linear scale markers from background elements. Through this work, we seek to provide a viable method that not only enhances accessibility and efficiency in wound assessment but also sets a foundation for future explorations of deep learning applications in clinical settings.

2 Materials and methods

2.1 Dataset definition

A dataset consisting of 451 images was employed for the training, validation, and testing of a semantic segmentation architecture. Additionally, an external test set comprising 98 images of the same category was utilized to further evaluate the model. The images were initially captured using smartphone cameras or similar devices and subsequently cropped to dimensions of 512x512 pixels. For each image, a corresponding ground truth mask was meticulously created. These generated masks highlighted two categories of regions of interest: "wound" and "scale marker." Furthermore, a third class, designated as "background," was established to include all other non-essential objects present within the image. The respective RGB triplets assigned to the wound, scale marker, and background regions were (0, 0, 0), (255, 255, 255), and (127, 127, 127) respectively.

2.2 Preprocessing

Prior to feeding the images into the network, a preprocessing stage was applied. First, noise reduction was performed applying a bilateral filter with a sigma spatial of 15, a sigma color of 15 and a filter size of 3px. To improve contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE) was employed. The images were initially transformed from the RGB color space to the LAB color vector. To accomplish this, the RGB vector (normalized in the range [0,1]) was first linearized by applying the inverse of the standard gamma correction:

$$R_{lin} = \begin{cases} \frac{R}{12.92}, & if\ R \leq 0.04045 \\ \left(\frac{R + 0.055}{1.055}\right)^{2.4}, & if\ R > 0.04045 \end{cases} \tag{1}$$

$$G_{lin} = \begin{cases} \frac{G}{12.92}, & if\ G \leq 0.04045 \\ \left(\frac{G + 0.055}{1.055}\right)^{2.4}, & if\ G > 0.04045 \end{cases} \tag{2}$$

$$B_{lin} = \begin{cases} \frac{B}{12.92}, & if\ B \leq 0.04045 \\ \left(\frac{B + 0.055}{1.055}\right)^{2.4}, & if\ B > 0.04045 \end{cases} \tag{3}$$

Where R, G and B are the value of the red, green and blue channels of the image. Then those vectors were multiplied by the transformation matrix:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R_{lin} \\ G_{lin} \\ B_{lin} \end{bmatrix} \quad (4)$$

Subsequently, to pass from XYZ space to the CIE LAB space (with the standard illuminant D65) the following transformations are applied:

$$L = 116\, f\left(\frac{Y}{Y_n}\right) - 16 \qquad (5)$$

$$a = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \qquad (6)$$

$$b = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \qquad (7)$$

Where Xn, Yn, Zn are the reference values for the D65 illuminant and

$$f(t) = \begin{cases} t^{1/3}, & if\ t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & otherwise \end{cases} \qquad (8)$$

Finally, the CLAHE algorithm was applied to the only perceptual lightness channel L of the images. The window size to implement the local adaptation was 32x32 px^2 and the maximum slope of the probability density function of the resulting image was clipped to 0.5. Results of the preprocessing stage are shown in Figure 1.

2.3 Semantic Segmentation Network

A semantic segmentation framework was explored for wound image analysis, integrating a Convolutional Neural Network (CNN)-based encoder with a lightweight multi-layer perceptron decoder. The study focused on optimizing performance by varying the CNN backbone, replacing the encoder with multiple ResNet variants pre-trained on the ImageNet dataset. Only the best-performing model on the internal test set was retained for further evaluation. The output layer of the network was designed to match the dimensions of the input tensor, and a softmax activation function was applied in a depth-wise manner to generate a probability distribution for each pixel within the mask. Subsequently, the argmax function was applied to determine the predicted class for each pixel, categorizing it as either "wound," "scale marker," or "background."

2.4 Training, validation and testing

Eighty images from the dataset were designated for internal testing, while the remaining three hundred and seventy-one images were allocated into training and validation sets, with 80% assigned to training and 20% to validation. The external test set was employed for additional metric evaluation. Data augmentation techniques were applied to the training set, allowing for transformations such as vertical flips, horizontal flips, and random rotations (specifically at angles of 90, 180, or 270 degrees). The batch size was configured to 8, with the maximum number of epochs set at 500. A weighted categorical focal loss was utilized as the loss function, with the gamma value set to 2.5. Weights for the loss were assigned as follows: 0.45 for the "wound" category, 0.45 for the "scale marker" category, and 0.1 for the "background" category. Adam optimizer with an initial learning rate of 5×10^{-4} and a weight decay of 1×10^{-6} was utilized. To enhance training efficiency, early stopping was implemented, which halted the training phase if the validation loss failed to improve for more than eight epochs. Additionally, a learning rate scheduler was incorporated, whereby the learning rate was reduced to 0.25 times its value if the validation loss did not show improvement over a span of five epochs. The model that exhibited the lowest validation loss during the training phase was saved for subsequent evaluation on both the internal and external test sets.

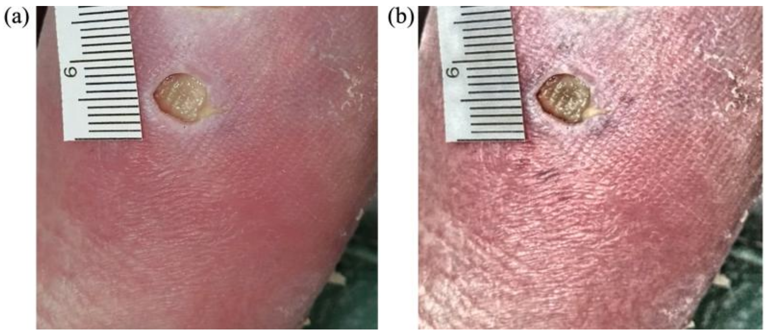


Fig. 1. Wound image before (a) and after (b) the preprocessing

2.5 Performance metrics

To evaluate the performance of the proposed model, accuracy and F1 score were computed at both the training and evaluation stages. While the accuracy of an image was assessed at the pixel-level, the F1 score was first calculated for each class and then the total score for the image (F1i) was obtained with the weighted average of the wound and the scale marker classes:

$$F1_i = \frac{3}{4} F1_w + \frac{1}{4} F1_s \quad (9)$$

Where $F1_w$ and $F1_s$ are the F1 scores for wound and scale marker classes, respectively. To assess the overall performance of the fine-tuned model, an average F1 score was also computed as the average of the F1 scores:

$$F1_i = \frac{1}{n}\sum_{i=1}^{n} F1_i \qquad (10)$$

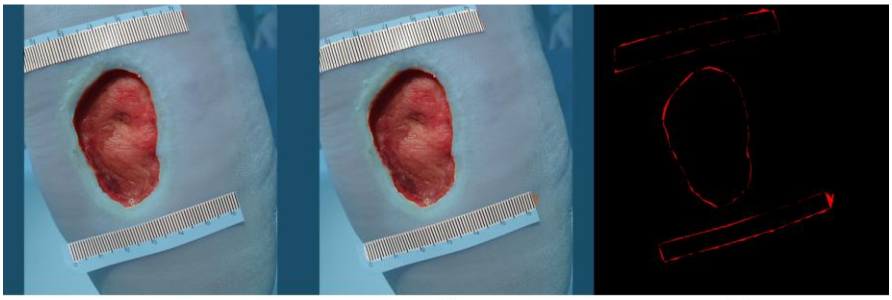
3 Results

The best-performing model employed ResNet152 as the CNN backbone in the encoder. Table 1 presents the performance metrics (including accuracy, F1 score, and loss) across the training, validation, internal test, and external test sets for the best finetuned model. The model reached a pixel accuracy of 99.02% during training, while achieving 98.82% on the validation set and 98.27% on the test set. The corresponding F1 scores were 0.97 for both the training and validation sets, and 0.95 for the test set. The loss values were 7×10^{-4} for training, 1.2×10^{-3} for validation, and 2.2×10^{-3} for the test set. Finally, for this multi-class image segmentation problem, the weighted F1 scores were 0.87 and 0.85 on the internal and external test sets, respectively.

Table 1. Accuracy, F1 Score, and Loss for Training, Validation, and Test Sets

Metric	Training	Validation	Internal test set	External test set
Accuracy (%)	99.02	98.82	98.27	-
F1 Score	0.97	0.96	0.96	-
F1 Score (Weighted)	-	-	0.87	0.88
Loss	0.7×10^{-3}	1.2×10^{-3}	2.2×10^{-3}	-

The result of segmentation using the proposed model, along with a comparison to the ground truth of two exemplificatory cases, is shown in Figure 2.



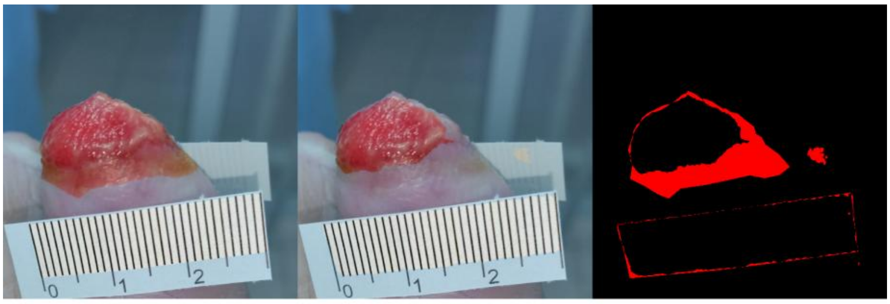
(a) (b) (c)

Fig. 2. Example of wound segmentation results produced by our model. (a) Segmentation output generated by our model, (b) ground truth annotation, and (c) visualization of differences between the model prediction and the ground truth, highlighting the detected discrepancies in red.

4 Discussion

Chronic wounds have a profound impact on patients' quality of life and, if not properly managed, can lead to severe complications such as limb amputation or even mortality. Effective wound management extends beyond treatment alone; it requires continuous and quantitative monitoring and assessment, to track healing progression and optimize therapeutic strategies. The necessity of automatic evaluation tools is highlighted by the lack of inter-observer agreement and precision [13].

The field of medical image segmentation has witnessed rapid advancements in recent years, primarily due to the evolution of deep learning techniques. In particular, techniques based on encoder-decoder architectures have become gold standards in medical imaging due to their architecture's ability to learn hierarchical features from complex datasets [14]. These techniques could be efficiently applied in the wound assessment field. Despite these advancements, several issues persist, such as dependence on large, annotated datasets and complexity of medical images themselves. Moreover, clinical applicability of segmentation algorithms depends on their robustness across diverse datasets. Variability in image acquisition modalities and patient demographics poses significant challenges, emphasizing the critical need for models to be both generalizable and interpretable [16].



(a) (b) (c)

Fig. 3. Example of a segmentation failure due to the specific characteristics of the wound. (a) ground truth annotation; (b) predicted segmentation mask; (c) visualization of differences between the model prediction and the ground truth.

To address these challenges and further improve segmentation performance in complex medical imaging scenarios, this work provides novel deep learning architectures along with a specifically designed pre-processing step for medical image segmentation. By doing so, it aims to enhance both the effectiveness and functionality of medical image segmentation, with a particular focus on anatomically complex and difficult-to-detect structures, such as chronic wounds. The methodology involves the usage of an encoder-decoder architecture tailored for images derived from diverse medical modalities, alongside with custom preprocessing pipeline. Our model proved capable of detecting wounds with notable precision in most cases, although minor issues persist when the differentiation between damaged and healthy tissue is not clearly identifiable from the initial image (Figure 3).

Our model showed remarkable performance with an accuracy of 0.98 and a F1-weighted of 0.87. Moreover, our results demonstrate strong generalization, as indicated by minimal performance degradation between the training and test sets.

The application of our segmentation technique could significantly support clinicians in assessing wound structures, providing a fast and reliable tool to make up for time-consuming processes and detection problems related to external factors. Moreover, considering that the healing of wounds is normally a long-lasting process, an evaluation method capable of providing accurate performances with smartphone-acquired camera is necessary to enable patients to stay at home and be frequently monitored at the same time [6]. This would lead to the applicability of our wound evaluation tool in telemedicine.

In conclusion, in this study by developing an automatic deep learning model for accurate wound image segmentation that demonstrates clinically relevant accuracy, contributes to the ongoing evolution of advanced wound assessment and management and lays the groundwork for future clinical applicability.

References

1. Lindholm C, Searle R (2016) Wound management for the 21st century: combining effectiveness and efficiency. Int Wound J 13:5–15. https://doi.org/10.1111/iwj.12623
2. Sen CK (2019) Human Wounds and Its Burden: An Updated Compendium of Estimates. Adv Wound Care (New Rochelle) 8:39–48. https://doi.org/10.1089/wound.2019.0946
3. Foltynski P, Ladyzynski P (2023) Internet service for wound area measurement using digital planimetry with adaptive calibration and image segmentation with deep convolutional neural networks. Biocybern Biomed Eng 43:17–29. https://doi.org/10.1016/j.bbe.2022.11.004
4. Alabdulhafith M, Ba Mahel AS, Samee NA, et al (2024) Automated wound care by employing a reliable U-Net architecture combined with ResNet feature encoders for monitoring chronic wounds. Front Med (Lausanne) 11:36977. https://doi.org/10.3389/fmed.2024.1310137
5. Chino DYT, Scabora LC, Cazzolato MT, et al (2020) Segmenting skin ulcers and measuring the wound area using deep convolutional networks. Comput Methods Programs Biomed 191:105376. https://doi.org/10.1016/j.cmpb.2020.105376
6. Foltynski P, Ciechanowska A, Ladyzynski P (2021) Wound surface area measurement methods. Biocybern Biomed Eng 41:1454–1465. https://doi.org/10.1016/j.bbe.2021.04.011
7. Blanco G, Traina AJM, Traina C, et al (2020) A superpixel-driven deep learning approach for the analysis of dermatological wounds. Comput Methods Programs Biomed 183:105079. https://doi.org/10.1016/j.cmpb.2019.105079
8. Hsu JT, Chen YW, Ho TW, et al (2019) Chronic wound assessment and infection detection method. BMC Med Inform Decis Mak 19:. https://doi.org/10.1186/s12911-019-0813-0

9. Wang J, Zhang H, Yi Z (2023) CCTrans: Improving Medical Image Segmentation with Contoured Convolutional Transformer Network. Mathematics 11:. https://doi.org/10.3390/math11092082
10. Zhang R, Tian D, Xu D, et al (2022) A Survey of Wound Image Analysis Using Deep Learning: Classification, Detection, and Segmentation. IEEE Access 10:79502–79515. https://doi.org/10.1109/ACCESS.2022.3194529
11. Ramachandram D, Ramirez-GarciaLuna JL, Fraser RDJ, et al (2022) Fully Automated Wound Tissue Segmentation Using Deep Learning on Mobile Devices: Cohort Study. JMIR Mhealth Uhealth 10:36977. https://doi.org/10.2196/36977
12. Marcato B, Ferraz C, Verri W, et al Automatic area estimation of mice wound images., 1-8. https://doi.org/10.5753/kdmile.2024.241973
13. Kekonen A, Bergelin M, Eriksson J-E, et al (2017) Bioimpedance measurement based evaluation of wound healing. Physiol Meas 38:1373–1383. https://doi.org/10.1088/1361-6579/aa63d6
14. Siddique N, Paheding S, Elkin CP, Devabhaktuni V (2021) U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. IEEE Access 9:82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020
15. Antonelli M, Reinke A, Bakas S, et al (2022) The Medical Segmentation Decathlon. Nat Commun 13:. https://doi.org/10.1038/s41467-022-30695-9
16. Tajbakhsh N, Jeyaseelan L, Li Q, et al (2020) Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med Image Anal 63:101693. https://doi.org/10.1016/j.media.2020.101693