Machine Learning-Driven Support Algorithm for Skin Ulcers Preliminary Diagnosis: A Lightweight Approach for Digital Images Semantic Segmentation and Color-Based Classification

Debora Beneduce, Guido Pagana, Fabrizio Bertone, Giuseppe Caragnano
Fondazione LINKS – Leading Innovation and Knowledge for Society
Torino, Italy
e-mails: {debora.beneduce, guido.pagana, fabrizio.bertone, giuseppe.caragnano} @linksfoundation.com

***Abstract*—This paper presents an automated pipeline for the detection, segmentation, and severity classification of cutaneous ulcers, addressing the clinical need for objective and remote wound monitoring. Despite increasing interest, real-time and interpretable Machine Learning tools in this domain remain scarce. We propose a hybrid solution combining classical image processing and Machine Learning techniques. Exploiting the guaranteed Convolutional Neural Network performance in binary segmentation tasks, a modified U-Net architecture, trained on grayscale digital images enhanced via Contrast Limited Adaptive Histogram Equalization, achieved high segmentation performance with an Intersection over Union of 0.82, Precision of 0.93, Recall of 0.89, and Dice coefficient of 0.88, using fewer than 2 million parameters. For severity classification, superpixel-wise brightness histograms were used to extract six discriminative features. A logistic regression model trained on these features reached a classification accuracy of 94%, effectively distinguishing between ulcer classes despite intra-class variability. The system offers robust performance with fast inference of 100 milliseconds per image and skin phototype-independence.**

Keywords-machine learning; convolutional neural network; teledermatology; skin ulcers monitoring.

I. INTRODUCTION

In recent years, the increasing demand for accessible and remote healthcare services has accelerated the development of telemedicine solutions. In particular, dermatology stands out as a field where early intervention can drastically reduce long-term complications, especially in the management of chronic wounds and skin ulcers [1]. Artificial Intelligence (AI) has demonstrated remarkable progress in dermatology, particularly in the automated detection and classification of pigmented lesions and melanoma, supported by large-scale studies and Deep Learning (DL) advancements [2][3]. However, despite the clinical relevance and growing incidence of chronic wounds, the application of AI to ulcer assessment remains comparatively underexplored, with relatively few high-quality studies and limited clinical integration [4][5][6]. This imbalance highlights the need for further research on AI-driven systems tailored specifically to the complex and heterogeneous nature of cutaneous ulcers. In addition, among different kinds of skin lesions, chronic skin ulcers represent a significant clinical concern due to their prolonged healing time and resistance to standard therapeutic interventions. These lesions are defined by their failure to progress through the normal stages of wound healing, often persisting for weeks or even months. Their development is frequently associated with underlying conditions, such as immobility, diabetes, and chronic venous insufficiency, making them prevalent in elderly and at-risk populations. Clinically, chronic ulcers are commonly categorized into three primary types: pressure ulcers, diabetic foot ulcers, and venous leg ulcers [7]. Pressure ulcers–also referred to as decubitus ulcers or bedsores–are caused by sustained mechanical pressure, typically over bony prominences, which results in localized ischemia and subsequent tissue necrosis. These lesions are especially common among bedridden or immobilized individuals and are a major source of morbidity and healthcare costs [8]. Diabetic foot ulcers, on the other hand, arise due to the interplay of peripheral neuropathy, ischemia, and repeated trauma in patients with diabetes mellitus. This kind of lesion represents one of the most severe complications of diabetes and is the leading cause of non-traumatic lower-limb amputations worldwide [9]. Venous leg ulcers are primarily the result of chronic venous insufficiency, where long-term increases in venous pressure cause fluid leakage, inflammation, and eventual skin breakdown. Venous leg ulcers are the most frequently occurring type of leg ulcer and are notorious for their tendency to recur and resist conventional treatment [10].

Despite the widespread occurrence and impact of these conditions, their clinical management remains highly reliant on subjective and manual evaluations. In most cases, wound assessments are carried out through visual inspection during in-person consultations, using basic tools, such as rulers, tracing paper, or planimetry software to measure wound size. Additionally, tissue characteristics, such as color, presence of exudate, or odor, are described qualitatively, introducing high inter-observer variability and limiting the precision required for effective longitudinal monitoring [11].

In light of these challenges, there is a growing need for objective, reproducible, and accessible tools that can assist healthcare providers in the accurate evaluation and follow-up of chronic wounds. This demand is further amplified by the global shift toward telemedicine and decentralized healthcare delivery. Patients with mobility limitations or those residing in remote areas could benefit greatly from systems that enable remote wound documentation and asynchronous specialist evaluation [12]. In response to these needs, this paper presents the development of a Machine Learning (ML)-based algorithm for the real-time elaboration of digital images to detect, classify and assess the severity of cutaneous ulcers. By leveraging AI techniques—specifically Convolutional Neural Networks (CNNs)—the system performs automate key tasks,

 ISBN: ISBNFILL

such as wound detection, semantic segmentation and severity estimation based on dominant color classification.

The current study is taking part of the SALUTEDERM research project, which is dedicated to explore telemedicine techniques and solutions to enhance skin lesion healing and skin care (see Acknowledgment section for project details).

II. Related work

In the clinical field of dermatology, the diagnostic process for cutaneous lesions entails a multi-step procedure involving lesion detection, morphological assessment, and subsequent classification and staging according to established severity criteria. Visual inspection and manual palpation by clinicians remain the gold standard for ulcer classification. In some cases, tools are also used to assess lesion depth. Accurate staging is essential to determine effective treatment and reduce healing time. However, misclassifications are not uncommon, due to factors, such as skin tone variation, patient age, and overall health status [13][14]. Despite standardized assessment protocols, wound classification by visual inspection and manual palpation exhibits substantial inter-rater variability, with reported agreement coefficients ranging from poor to moderate. This variability arises from differences in clinician experience, subjective interpretation of tissue characteristics, and patient-related factors, which together challenge the reliability and reproducibility of manual ulcer staging [15]. In the last few years, to support clinical assessments, the use of digital image analysis has emerged as a promising approach to improve the evaluation of chronic wounds.

Delegating the diagnostic responsibility from human experts to automated systems involves a sequential pipeline of three fundamental and cascading tasks:

1) *Segmentation task* – automated identification and delineation of the precise boundaries of the Region of Interest (ROI) that is the ulcerated region within an image.
2) *Classification task* – division of the ROI into different classes basing on the lesion severity.
3) *Severity assessment task* – evaluation of the main characteristics extracted from the ROI.

Among the parameters useful to describe and evaluate the staging and the damage progress, to consider the kind of the involved cutaneous tissues is mandatory. In fact, the European Pressure Ulcer Advisory Panel (*EPUAP*) classified pressure ulcers into four main stages as:

- *Grade I*: intact skin with non-blanchable redness, which may also present with warmth, hardness, or pain.
- *Grade II*: partial thickness skin loss involving the epidermis and/or dermis, appearing as a shallow ulcer or blister.
- *Grade III*: full thickness skin loss extending into subcutaneous tissue often showing slough presence (yellow), but not exposing bone, tendon, or muscle.
- *Grade IV*: full thickness skin loss with extensive destruction, tissue necrosis, or damage to muscle, bone, or supporting structures [16].

Automated image analysis methods—based on bag-of-features representations and ML classifiers, such as Support Vector Machines and K-Nearest Neighbors—have been widely adopted for segmentation and classification tasks. Despite promising accuracy levels, these models often struggle when dealing with complex wound structures involving mixed tissue types, increasing the likelihood of classification errors [17][18]. Early efforts in this domain focused on basic computer vision techniques, such as color-based segmentation and edge detection, has been applied to 2D photographic data. These methods enabled semi-automated estimation of wound dimensions but were limited in their ability to capture complex tissue characteristics or to generalize across different wound types and imaging conditions [19]. The advent of ML and DL has markedly advanced the field, particularly with the introduction of CNNs for medical image analysis. CNN-based models have demonstrated strong performance in tasks, such as wound segmentation, classification of tissue types (e.g., granulation, slough, necrosis), and even prediction of healing trajectories based on sequential imaging data [20][21]. These models offer significant advantages over traditional techniques by learning hierarchical features and complex patterns directly from raw image inputs, thereby reducing the need for manual feature engineering [22].

Among the studies most closely related to the topic is the work by Zahia et al. [23] who developed a CNN-based method to classify tissue types (granulation, necrosis, and slough) using 20 high-resolution images, which were cropped into 380,000 smaller RGB patches. These were manually segmented and preprocessed with masking, grayscale conversion, Otsu thresholding, and reflection correction. Their CNN had 3 convolutional layers with increasing feature maps and used ReLU activations. They reported high sensitivity and precision for granulation and necrosis tissues (greater that 80%) but lower for slough (less than 60%). García-Zapirain et al. [24] employed a two-stage approach using a 3D CNN (DeepMedic) to extract the region of interest and segment tissues from a dataset combining original and Medtec images. Pre-processing included Gaussian smoothing and HSI color-space transformation to handle lighting variation. The first network had dual pathways for ROI detection, and the second network used four input modalities, including a prior visual appearance model built using color probability and Euclidean distances. The system achieved strong performance with Dice Similarity Coefficient (DSC) and Area Under the Curve (AUC) values around 95%. Aldughayfiq et al. [22] leveraged YOLOv5 for real-time detection and classification of pressure ulcers by grade, demonstrating a precision of 78.1%, while Pereira et al. [25] highlighted the importance of perceptually uniform color spaces (CIELAB and CIELUV) for tissue discrimination, achieving 73.8% accuracy and an AUC of 0.82. More recently, Liu et al.[4] integrated deep learning (Inception-ResNet-v2) with a clinical questionnaire in a smartphone-based diagnostic tool for pressure ulcer assessment, achieving over 90% accuracy across both cellulitis detection and necrotic tissue grading. Recent contributions have further emphasized clinical applicability, real-time performance, and quantitative wound measurement. Ramachandram et al. [26] proposed a fully automated pipeline for wound and tissue segmentation on mobile devices, employing two CNNs to segment the

ulcer and classify tissue types, such as epithelial, granulation, slough, and eschar. Their models achieved an Intersection over Union (IoU) of 0.8644 for wounds and 0.7192 for tissue classification, while also quantifying the substantial inter- and intra-rater variability among clinicians. Liu et al. [27] combined U-Net and Mask R-CNN with LiDAR-based area measurement for pressure injuries, reporting a Dice coefficient of 0.8448 on external validation and a mean relative area error of 26.2% compared to manual measurements, highlighting both the potential and the current limitations of quantitative wound assessment in clinical settings. Carvalho et al. [28] explored CNN and Transformer-based architectures, including DeepLabV3+, SegFormer, and MedSAM, for segmentation and real-world wound measurement. Using reference markers to scale images, their pipeline achieved Dice scores above 92% and area estimation errors as low as 5.36% on private datasets, although performance decreased when the entire pipeline was applied under diverse imaging conditions.

Despite these advances, several challenges persist. Learning-based methods require large volumes of high-quality annotated medical images, which remain scarce due to high annotation costs, limited patient data, and ethical constraints [29]. Unlike imaging modalities such as brain, retinal, or chest CT, dermatological conditions have traditionally been assessed via direct visual inspection, complicating the creation of large-scale datasets. Additional challenges include differentiating among tissue types, dealing with ill-defined lesion boundaries, and ensuring robustness to variations in lighting, skin tone, and image quality [30][31]. Data augmentation strategies have been proposed to mitigate small dataset limitations [22][32]. Collectively, these studies indicate that automated segmentation, tissue classification, and measurement are promising, but fully integrated systems capable of robust performance across heterogeneous conditions, severity assessment, and telemedicine deployment remain an open research need.

III. METHODS

Building on the presented premises, the present study aims to develop an algorithm that first achieves high-quality semantic segmentation of skin ulcers, and subsequently enables color-based lesion severity evaluation, with limited computational demands. Indeed, achieving strong segmentation accuracy is crucial, but it must be balanced with low computational cost–implying a model with fewer learnable parameters and lightweight architecture–to ensure real-world efficiency [33]. For this reason, small network sizes and high inference speeds guaranteed by CNNs architectures have been preferred over superior segmentation accuracy performed by more complex methods, like transformer-based ones [34].

The proposed algorithm is structured as a sequential pipeline, in which each task takes as input the output of the previous one. Specifically, the process begins with a pre-processing stage to standardize and enhance the input data, followed by lesion segmentation with associated post-processing to refine the obtained masks, and finally lesion classification based on the segmented regions. This design ensures modularity, efficiency, and a clear propagation of information across tasks.

The entire system has been programmed in Python language and built, trained, validated and tested on two public databases from the 2021 MICCAI Foot Ulcers Segmentation Challenge, the first composed by over 1200 de-identified diabetic foot ulcers images and their respective labels [35], and the second created in collaboration with the AZH Wound & Vascular Center composed by 1109 cropped patches of foot ulcers [36].

A. Segmentation task

The first step of the pipeline is the ROI detection that consists in the recognition of the ulcer area and its borders highlighting.

Pre-processing: To perform the task, the dataset variability reduction is recommended as a starting pre-processing strategy. For this reason the images from the dataset need to be normalized both in terms of size and in terms of pixels values to make the training consistent and stable. Hence, a division by 255 – the maximum pixel value– is applied to each of the three color channels of the RGB color space, namely Red, Green and Blue, so that the operational range per pixel is now $[0, 1]$; then each image undergoes a `resize` operation to the standard 512×512 size. The size choice is a trade-off between the need to preserve as much information as possible and a reduced computational weight. It has also been demonstrated that the exclusion of the brightness information, which carries a big amount of variability and translates into the use of single-channel grayscale (GS) images instead of three-channels RGB ones, can further improve segmentation performance in dermatological studies by making the algorithm less sensitive to illumination artifacts [37]. For this reason, a comparison between ulcers detection on RGB and on GS images, both original and contrast-enhanced, will be evaluated.

Training: Among the most efficient and cost-effective neural network architecture employed for binary segmentation, U-Net demonstrates relevant potential. We implemented a deep U-Net variant tailored for high-resolution biomedical image segmentation tasks. The network adopts a symmetric encoder–decoder architecture with four levels of downsampling and upsampling, and includes skip connections to preserve spatial context and fine-grained features. The encoder consists of four convolutional blocks, each composed of two convolutional layers (kernel size 3×3, ReLU activation, He-normal initialization, and same padding), followed by 2×2 max pooling for downsampling. The number of filters doubles with each level, starting from 16 up to 128. A bottleneck layer with two convolutional layers and 256 filters processes the compressed representation. The decoder mirrors the encoder structure. Each upsampling step (via 32×2 upsampling) is followed by a concatenation with the corresponding encoder feature map (skip connection), and two convolutional layers that progressively reduce the number of filters back to 16. The final layer is a 1×1 convolution with sigmoid activation, producing a single-channel probability map for binary segmentation. This architecture balances depth and computational efficiency, maintaining the U-Net's ability to integrate multi-scale contextual information while enabling the extraction of

deeper features via the added encoder stage. The modified U-Net architecture proposed by the authors is depicted in Figure 1 and more details, as the number of layers, are reported for clarity in Table I.

TABLE I. ARCHITECTURE OF OUR U-NET-LIKE MODEL.

Layer(s) block	Feature Maps	Kernel Size
Input	1	-
Encoder 1	16 → 16	3×3, 3×3
Pooling 1	16	2×2 (max pooling)
Encoder 2	32 → 32	3×3, 3×3
Pooling 2	32	2×2 (max pooling)
Encoder 3	64 → 64	3×3, 3×3
Pooling 3	64	2×2 (max pooling)
Encoder 4	128 → 128	3×3, 3×3
Pooling 4	128	2×2 (max pooling)
Bottleneck	256 → 256	3×3, 3×3
Decoder 1	256+128 → 128 → 128	up 2×2, 3×3, 3×3
Decoder 2	128+64 → 64 → 64	up 2×2, 3×3, 3×3
Decoder 3	64+32 → 32 → 32	up 2×2, 3×3, 3×3
Decoder 4	32+16 → 16 → 16	up 2×2, 3×3, 3×3
Output	1	1×1 (sigmoid)

Training parameters, as the number of training epochs, batch size and learning rate, and data augmentation techniques have been tuned after several trainings. The training has been run across 50 epochs with $batch\_size = 2$ and $learning\_rate = 1e^{-4}$. The Adam optimizer has been set to take advantage of:

- Automatically regulated learning rate useful to manage ulcers borders that can produce different gradients in respect to other areas.
- Faster convergence in presence of a U-Net architecture.
- Low sensitivity to unbalanced classes (e.g. 90% background, 10% object).

Data augmentation techniques included clockwise/counterclockwise rotations, width/height shifts, zoom and horizontal/vertical flips and the fill mode was set as `'nearest'`. Due to this method, the volume of data has increased fourfold. Finally, the 80% of the whole dataset was split using a 90/10 ratio between training and validation sets and both RGB and grayscale enhanced images have been used for comparison in different trainings.

Unlike many previous studies, where ulcer analysis is performed in two sequential steps—first identifying a Region Of Interest (ROI) and then segmenting the wound within that ROI—we directly trained the U-Net to segment the ulcer from the entire image. This decision was motivated by both methodological and practical considerations. First, ROI detection introduces an additional preprocessing stage that may propagate errors and increase variability across images. Second, U-Net architectures have proven effective at simultaneously learning global contextual cues and local boundary information, allowing reliable segmentation even without prior cropping [38]. Finally, an end-to-end segmentation pipeline reduces complexity and enhances reproducibility, making the method easier to deploy in real-world clinical workflows.

Post-processing: The predicted binary masks generated by the algorithm on the test set (remaining 20% of the dataset) have been observed and analyzed. The biggest issue came from the background noise represented by disturbing elements present into the image and characterized by a range of colors similar to the one of the skin ulcers. Also, very small areas of healthy skin are sometimes wrongly identified as ROIs. After a `resize` operation at the original size of each input image, these considerations led to two main post-processing procedures:

1) Removal of very small object–For each mask, each segmented object with $area < area_{max}/6$, where $area_{max}$ is the area (expressed in pixels) of the biggest recognized object, is ignored.
2) Background removal through skin segmentation–In addition to different appearances of ulcers of different etiology and class, healthcare professionals usually take pictures in diverse ways, in heterogeneous light conditions and position. These factors further increase the variability of the dataset, which is already limited in size given the complexity of the problem. To limit confounding elements and reduce the amount of data to be analyzed, a strategy to isolate the affected limb by removing the background has been implemented. First, images are converted from *RGB* color-space to YC_rC_b one, useful to take advantage of the separation between the luminance information represented by the *Y* channel and the chrominance contribution expressed by the C_i channel, where subscript *i* can stand for red (C_r) or blue (C_b). The luminance represents the brightness level of the image, whereas the blue and red chrominances carry the color information by indicating the shift of blue and red channels from the luminance value [39]. As visible in Figure 2, the C_r channel shows the cutaneous area highlighted in respect to other objects into the image offering the strategy for skin segmentation.

B. Classification task

Among the various features used for skin chronic wound classification, color information plays a prominent role due to its partial correlation with the depth and extent of tissue damage [40]. Indeed, first stage lesions, partial tissue loss lesions, covered by slough lesions and necrotic lesions are characterized by colors ranging from light red to vivid red to yellow-tinged to brown/black. As the available dataset is made of diabetic foot ulcers, Stage I lesions are not represented and therefore are not recognized during the segmentation task. Since the classification module operates exclusively on segmented ulcers, the absence of Stage I lesions in segmentation directly implies that their classification is not included in the scope of the algorithm.

To enable chromatic differentiation of lesions, brightness histograms of each R, G and B channel intensities were computed and analyzed. The histograms show the possible intensity values, $[0, 255]$ along the x-axis, and the frequency of pixels exhibiting each intensity value along the y-axis. Figure 3 displays an example of the different distributions of colors intensities among different stages of ulcers.

To reduce the huge amount of data deriving from the

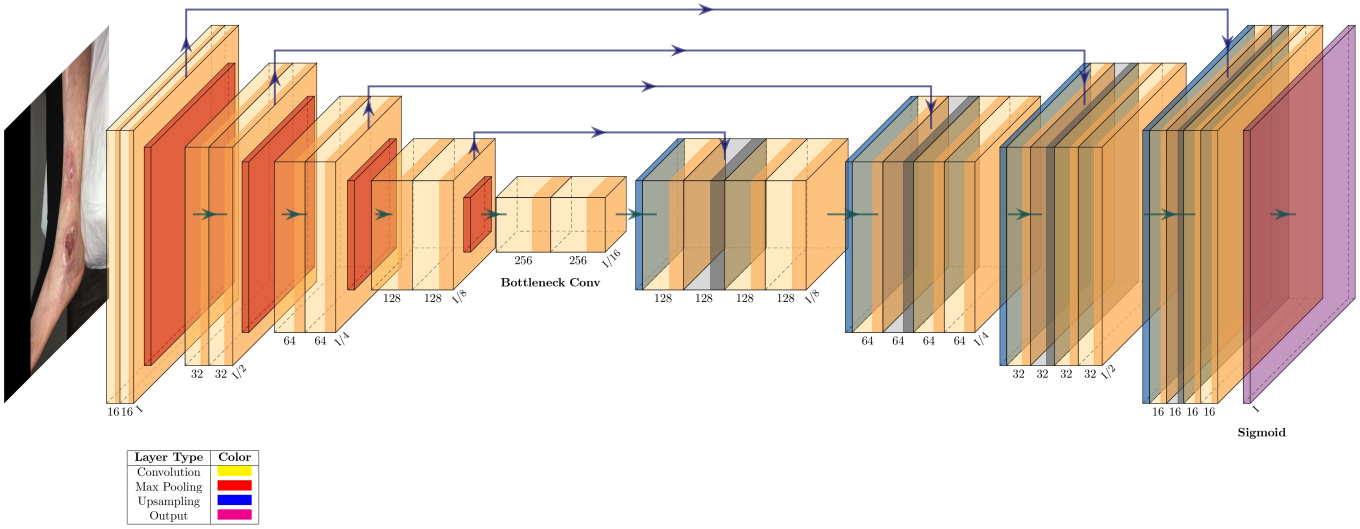


Figure 1. Modified U-Net architecture proposed by the authors.

Figure 2. RGB image [17], YC_rC_b image, Y channel, C_r channel, and C_b channel (from left to right).

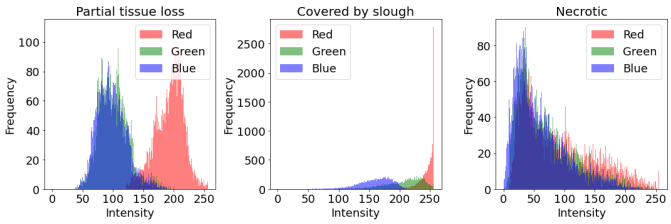


Figure 3. Partial tissue loss ulcer, necrotic ulcer and covered by slough ulcer and respective brightness R, G and B histograms (from left to right).

processing of each single pixel, the related introduced variability and because regions of pixels are more informative than single pixels, the Simple Linear Iterative Clustering (SLIC) algorithm–based on *Superpixel* technique–is performed. This method segments an image into a chosen number of regions–named superpixels– clustering pixels that appear similar according to some perceptual features relying on measures based not only on color similarity but also on the shape of the regions delimiting areas significant changes in intensity [41].

As the dataset does not provide any class or tissue information, the involvement of a medical expert for accurate labeling was crucial. Consequently, the entire database was carefully reviewed by the clinician, who selected 216 lesions deemed representative of all severity classes, while deliberately avoiding borderline cases that could introduce ambiguity or bias due to subjective interpretation. Each selected lesion was then manually classified under the expert's supervision, ensuring that the labels reflected both clinical relevance and consistency. From each of these lesions, six parameters has been extracted and, in addiction to the ground truth labels, fed into a multiple logistic regressor for the class prediction by splitting the set of labeled ulcers into 80% training set and 20% test set.

In the following section, the results achieved through the implemented workflow are presented.

IV. Results Analysis

The results are organized according to the pipeline introduced in Section II and Section III.

A. Segmentation task results

The described modified U-Net model has been trained across 50 epochs. A notably small $batch\_size = 2$ yielded more favorable training dynamics compared to larger batch sizes. The increased stochasticity in the gradient estimates, induced by the smaller batch size, likely acted as an implicit regularization mechanism, contributing to improved generalization and reduced overfitting also lowering GPU memory usage. The most performing trainings derived from GS images training, in particular on locally-contrast modified GS images via CLAHE method. Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image processing technique that enhances the contrast of images by applying histogram equalization locally, in small regions, rather than globally. CLAHE also incorporates a contrast-limiting step to prevent over-amplification of noise in homogeneous regions. It does this by clipping the histogram at a predefined threshold ($clip\_limit$ parameter) before redistributing the clipped pixels evenly. After several trials, the optimal value has been found to be $clip\_limit = 0.8$. An example of comparison among *RGB*, *GS* and *GS*-contrast enhanced outcomes is reported in Figure 4.

Despite the discrete performance achieved, the segmentation of confounding elements remained a challenge. For this reason, post-processing is compulsory both in terms of small object removal and in color-thresholding to discard out-of-skin segmented element. Since in some cases the lesion is extremely small, while in others similarly sized objects

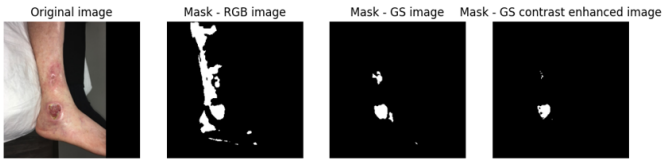


Figure 4. Example of comparison on a test image [35] between binary segmentation masks predicted by training on RGB, GS and GS-contrast enhanced images.

represent confounding elements, it is not feasible to define a global size threshold that is consistently valid across all images. To overcome this problem, it has been observed that the biggest recognized object is always represented by the real lesion to be segmented; in respect to this object, other selected areas, smaller than $\frac{1}{6}$ of the biggest identified area on the same image, are for sure segmentation errors. In case only one object is found, then it is considered as an ulcer unless it is discovered not to belong to the skin area detected through the second post-processing step. As illustrated in Section III, the YC_rC_b color-space can be helpful for skin segmentation. Indeed, by executing a global thresholding in the range of $[0.55, 0.70]$ for C_r channel, it is possible to easily remove the background (Figure 5). Each detected object outside the segmented skin is then ignored.

(a) (b)

Figure 5. Skin segmentation: (a) original image [35], (b) result after global thresholding on channel C_r.

At the end, the final best training provided good segmentation performance represented by $IoU = 0.82, Precision = 0.93, Recall = 0.89$ and $Dice\ Coefficient = 0.88$. Being in a biomedical field, the Precision metric gains more importance in respect to the Recall one since to reduce the amount of false negative pixels (in this case meaning non-detected ulcers) is considered more relevant than to have less false positive pixels (meaning healthy skin recognized as damaged). Thus, the elevated precision score is of particular relevance in study scenarios where accurate discrimination is as inherently difficult as relevant. Finally, the mean inference time per image is 100 ms.

It is important to emphasize that our modified U-Net achieves competitive results while maintaining a notably low number of trainable parameters — fewer than 2 million. Table II presents *IoU* scores reported by some of the most frequently cited studies within the same research domain, employing comparable ML-based segmentation methodologies, alongside their respective parameter counts. To ensure a fair and meaningful evaluation, we compared our proposed model against several widely used and high-performing U-Net variants in biomedical image segmentation, namely the standard U-Net, ResU-Net, Attention U-Net, and FU-SegNet. These architectures represent the most commonly adopted baselines in the field and cover different directions of improvement to the original U-Net design, such as residual connections, attention mechanisms, and enhanced decoder structures. This choice allows us to directly assess the specific contribution of our modifications within the same architectural family.

Finally, three examples are displayed in Figure 6 to show the segmentation outcomes. The first image also assesses the phototype-independence of the model, which should not be taken for granted. Although the dataset does not provide explicit information on skin phototype, a visual inspection of the database revealed that only about 5% of the images correspond to darker skin tones. Nevertheless, the first image on the left demonstrates that, despite this underrepresentation, darker phototypes do not compromise the robustness of the algorithm.

TABLE II. COMPARISON WITH STATE-OF-THE-ART ULCER SEGMENTATION RESEARCHES.

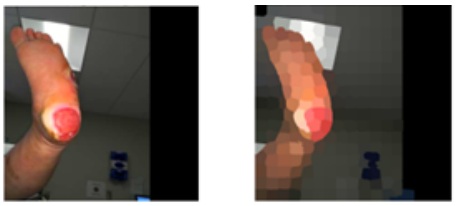
Model	IoU	Param.	Reference
Standard U-Net	0.68	7.8M	[38]
ResU-Net	0.72	8.9M	[42]
Attention U-Net	0.75	8.9M	[43]
FUSegNet	0.77	12.7M	[32]
Authors' Modified U-Net	0.82	1.9M	–

Figure 6. Examples of segmentation outcomes on test images [35].

B. Classification task results

The brightness histograms approach allows for the extraction of color distribution patterns within each lesion, facilitating the identification of relevant visual features, such as variations in redness, yellowness, or darkness associated with different tissue types or stages of wound healing. The use of this method would be computationally and time-consuming if applied pixel by pixel. For this reason, the SLIC algorithm paired with histograms analysis is the key of our solution to the classification problem. Considering *N* as the number of pixels in the input image, *K* as the desired number of superpixels, the approximate size of a superpixel will be N/K and for roughly equally sized superpixels there would be a center at every grid interval $S = \sqrt{\frac{N}{K}}$, resulting into a superpixel spatial extent about S^2. The input parameter for the SLIC algorithm are then the input image, the parameter *K* and the variable **m** that is a measure of superpixel compactness. For the current study, $K = 290$ and $m = 20$ have been demonstrated to guarantee the best tradeoff between minimizing color-variability and computational costs and avoiding the loss of original color

and shape information (Figure 7).



(a) (b)

Figure 7. SLIC technique: (a) original image [35], (b) SLIC processed image with $K = 290$ and $m = 20$.

Each superpixel has then been enumerated so that the interested areas could be manually selected by typing their respective number. RGB histograms were computed for each extracted superpixel, and six parameters–identified as the most discriminative–were evaluated. These consist of the median values of the three color channels and the intensity differences among them. The ground-truth labels and the respective six variables for each lesion have been exploited to train a multiple logistic regressor. The extracted variables for the training set are reported in Figure 8 to have an overview of the distributions. The total number of samples is higher than 216 because some lesion are mapped by more than a single superpixel. In addition, some samples from healthy skin areas has been analyzed and inserted into the figure to demonstrate, once more, how the distinction between injured and non-injured skin is not trivial. Indeed, the extracted parameters from healthy skin result into overlapping to all of the three ulcers classes. Moreover, the limited representation of the necrotic lesion class can be observed. This problem is counterbalanced by the evident color separation from the other classes.

Despite the quite considerable variability of the features given by different reasons (e.g. image acquisition device and the huge variety of ulcers) the extracted features demonstrated to be sufficiently discriminatory. Figure 9 present the data as mean ± Standard Error of the Mean (SEM) in order to provide a concise summary of class-level differences. While the SEM does not reflect the within-class variability, it effectively represents the precision of the estimated class means, allowing for clearer visualization of systematic differences between classes. This choice is further supported by the model's high classification accuracy of 94%, suggesting that, despite the high variability, the method proves to be both effective and reliable. The notable inter-class discrimination ability is, furthermore, expressed by the confusion matrix in Figure 10. The vertical axis represent the ground truth class, whereas the horizontal axis represents the class predicted by our model.

Classes are assessed according to the *EPUAP* definition and, as already discussed, class I is not considered as it is not represented by the dataset. The classification error reaches only 6% in the higher severity classes. Expanding the dataset with a larger number of labeled images would enable a more robust evaluation, potentially confirming that the majority of misclassifications are conservative–i.e., the predicted severity tends to be higher than the actual one–an outcome that is generally preferable in medical contexts to avoid underestimation of critical conditions.

With regard to the propagation of segmentation errors to the sequential classification task, although the segmentation model does not achieve 100% accuracy in terms of IoU, this does not cause a meaningful drop of performance. Segmentation inaccuracies mainly occur at the lesion borders, whereas in clinical practice the most severely affected tissue is often located in the central region of the ulcer — a region that tends to heal more slowly, and which becomes the critical determinant of severity. Consequently, since that central area is generally well captured by the model, the predicted class remains accurate and robust despite minor segmentation errors.

V. Conclusion and Future work

This work presented the design and evaluation of an end-to-end system for the automated analysis of cutaneous ulcers, addressing two critical tasks in the wound care pipeline: segmentation and severity classification. The proposed solution leverages a combination of classical image enhancement techniques, DL architectures, and lightweight ML classifiers to support clinicians in the timely assessment and monitoring of chronic wounds.

For the segmentation task, a modified version of the U-Net architecture was employed. The model, trained on grayscale images enhanced via the CLAHE algorithm, demonstrated superior performance compared to RGB-based approaches. CLAHE proved particularly beneficial in enhancing local contrast while avoiding noise over-amplification, allowing for more reliable lesion boundary detection. The model achieved a mean Intersection over Union (IoU) of 0.82, Precision of 0.93, Recall of 0.89, and a Dice coefficient of 0.88, which compare favorably to existing state-of-the-art solutions while maintaining a significantly lower number of parameters (< 2M). This makes the proposed model suitable for real-time deployment, especially in resource-constrained clinical or mobile environments. Despite the strong performance, certain challenges were observed in the segmentation of confounding elements, such as artifacts or visually similar skin regions. These were effectively addressed through a post-processing pipeline, which included object size filtering and color-based thresholding in the YC_rC_b color space. In particular, lesions were robustly distinguished from non-skin elements by analyzing the C_r channel, enabling the removal of out-of-context segmented areas.

For the classification task, the model relied on the extraction of color-based features from wound superpixels obtained using the SLIC algorithm. This method significantly reduced the computational burden compared to pixel-level analysis, while preserving the spatial and chromatic properties of the lesions. The most informative features–channel medians and inter-channel differences–were used to train a multiple logistic regression classifier, which achieved an overall classification accuracy of 94%. Notably, the model performed well despite significant intra-class variability due to differences in ulcer morphology, acquisition conditions, and lighting. The use of mean ± SEM to present class-level feature distributions

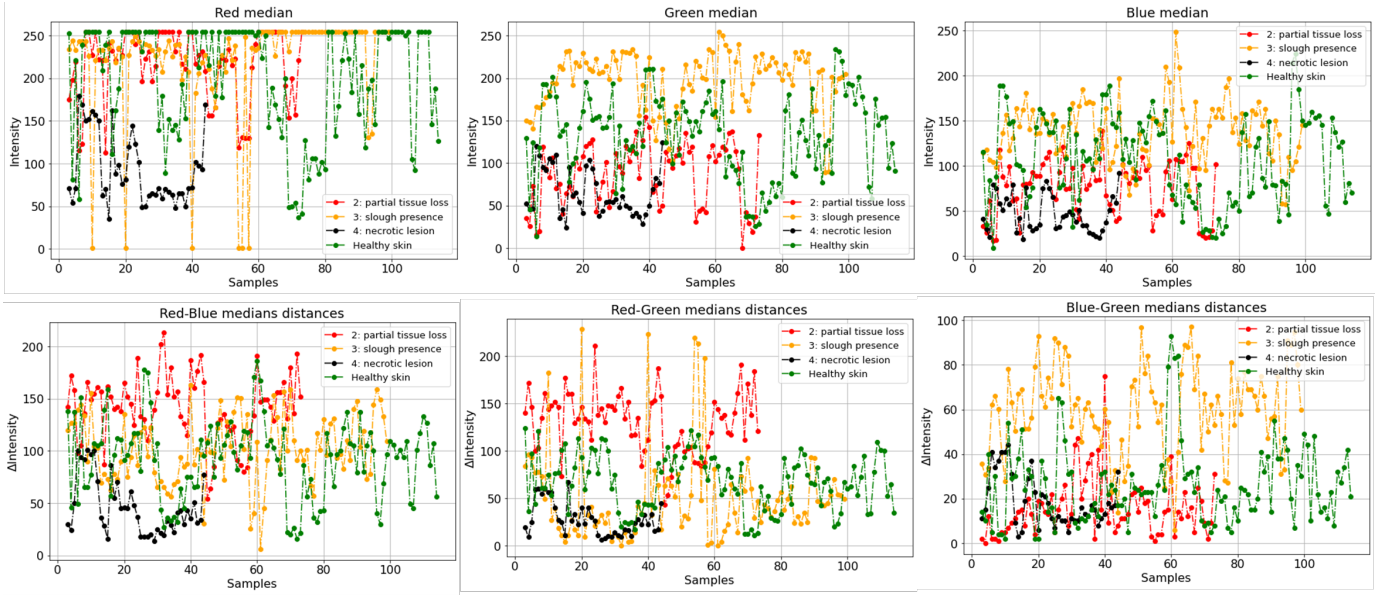


Figure 8. Extracted features for each class (training set): Red, Green and Blue medians (first raw, from left to right); Red-Blue, Red-Green and Blue-Green medians distances (second raw, from left to right).

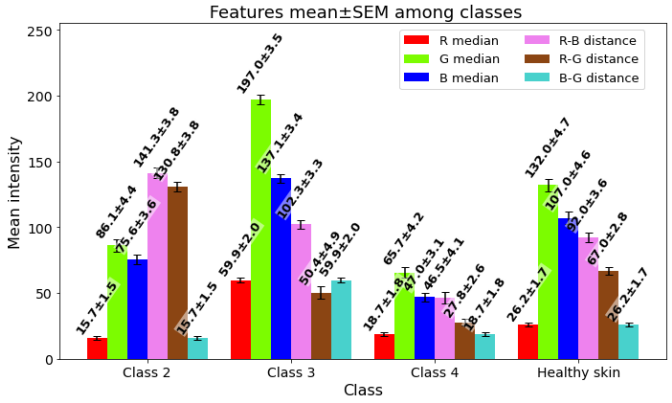


Figure 9. Mean of the extracted features for classification task.

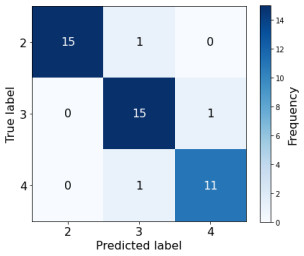


Figure 10. Confusion matrix for the classification task.

proved to be an effective strategy for visually conveying discriminative trends, even in the presence of overlapping data. Importantly, the classification error was predominantly observed in the higher severity classes, where conservative misclassifications are preferable in clinical settings, as they reduce the risk of underestimating potentially critical conditions. Moreover, the model demonstrated promising phototype-independence, which is a crucial factor for broad applicability in diverse patient populations.

Building upon the encouraging results obtained in both segmentation and classification tasks, future developments will focus on three main directions. First, authors plan to significantly expand and diversify the dataset, with particular emphasis on including underrepresented ulcer classes (such as EPUAP Stage I and various necrotic subtypes) and a broader range of patient skin phototypes. This will improve the generalizability of the model and ensure its reliability in real-world, heterogeneous clinical settings. Second, efforts will be directed toward extracting clinically interpretable outcome measures from the segmentation and classification results. These outcomes will be used to derive a quantitative severity score for each lesion, analogous to established metrics in the literature—such as the Photographic Wound Assessment Tool (PWAT)—but designed to be computationally lighter and fully automatable, thus more suited for integration into digital health systems. Third, all modules of the proposed pipeline will be integrated into a dedicated telemedicine device, currently under development by the authors. This device will be capable of acquiring standardized digital images of the wound, will incorporate the proposed algorithms to provide preliminary, automated assessments of wound presence and will evaluate the temporal evolution of the severity score.

The system has been trained, validated, and tested on anonymized images from a public dataset that already complies with General Data Protection Regulation (GDPR) requirements. Regarding its future clinical use, procedures for approval by the local ethics committee have already been initiated, and a dedicated protocol for data encryption and anonymization will be developed to ensure secure storage in the databases that will be progressively built.

The system is designed to function as a medical decision-support tool, especially in settings where specialist access is limited, such as home care or remote rural areas. This integrated approach–combining robust AI algorithms, clinically

meaningful outcomes, and practical hardware implementation–lays the foundation for a comprehensive and scalable solution in the emerging field of AI-assisted wound telemonitoring.

ACKNOWLEDGMENT

This research has been funded under the "Aggregazioni R&S – Salute" call issued by Assessorato Sviluppo Economico, Formazione e Lavoro, Trasporti e Mobilità Sostenibile, Ricerca, Innovazione e Trasferimento Tecnologico of the Aosta Valley Region, with the project SALUTEDERM (CUP B49J23008170009).

REFERENCES

[1] E. L. Eber, E. Arzberger, C. Michor, R. HofmannWellenhof, and W. Salmhofer, *Mobile Teledermatologie in der Behandlung chronischer Ulzera*, German, 2019. DOI: 10.1007/S00105-019-4397-5.

[2] N. Melarkode, K. Srinivasan, S. M. Qaisar, and P. Plawiak, "AI-Powered Diagnosis of Skin Cancer: A Contemporary Review, Open Challenges and Future Research Directions", *Cancers*, vol. 15, no. 4, p. 1183, Feb. 2023, Epub 2023 Feb 13. DOI: 10.3390/cancers15041183.

[3] C. Lei et al., "Convolutional Neural Network Models for Visual Classification of Pressure Ulcer Stages Cross-Sectional Study", *JMIR Medical Informatics*, vol. 13, no. 1, Mar. 2025. DOI: 10.2196/62774.

[4] T. J. Liu et al., "A pressure ulcers assessment system for diagnosis and decision making using convolutional neural networks", *Journal of the Formosan Medical Association*, vol. 121, no. 11, pp. 2227–2236, Nov. 2022, Epub 2022 May 4, ISSN: 0929-6646. DOI: 10.1016/j.jfma.2022.04.010.

[5] P.-H. Huang et al., "Development of a deep learning–based tool to assist wound classification", *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 76, no. 6, pp. 1462–1470, Jun. 2023, Epub 2023 Feb 10. DOI: 10.1016/j.bjps.2023.01.015.

[6] R. Niri et al., "Wound Segmentation with UNet Using a Dual Attention Mechanism and Transfer Learning", *Journal of Imaging Informatics in Medicine*, Jan. 2025, Published 23 Jan 2025. DOI: 10.1007/s10278-025-01386-w.

[7] R. G. Frykberg and J. Banks, "Challenges in the Treatment of Chronic Wounds", *Advances in Wound Care*, vol. 4, no. 9, pp. 560–582, Sep. 2015. DOI: 10.1089/wound.2015.0635.

[8] L. E. Edsberg et al., "Revised National Pressure Ulcer Advisory Panel Pressure Injury Staging System", *Journal of Wound, Ostomy and Continence Nursing*, vol. 43, pp. 585–597, 6 Nov. 2016, ISSN: 10715754. DOI: 10.1097/WON.0000000000000281.

[9] D. G. Armstrong, A. J. Boulton, and S. A. Bus, "Diabetic Foot Ulcers and Their Recurrence", *New England Journal of Medicine*, vol. 376, pp. 2367–2375, 24 Jun. 2017, ISSN: 0028-4793. DOI: 10.1056/nejmra1615439.

[10] T. F. O'Donnell et al., "Management of venous leg ulcers: Clinical practice guidelines of the Society for Vascular Surgery® and the American Venous Forum", *Journal of Vascular Surgery*, vol. 60, 3S–59S, 2 2014, ISSN: 10976809. DOI: 10.1016/j.jvs.2014.04.049.

[11] K. Ousey, B. Gilchrist, and H. Jaimes, "Understanding clinical practice challenges: a survey performed with wound care clinicians to explore wound assessment frameworks", *Wounds International*, vol. 9, no. 4, pp. 58–62, 2018, ©Wounds International 2018, [retrieved: August, 2025].

[12] C. M. Contreras et al., "Telemedicine: Patient-Provider Clinical Engagement During the COVID-19 Pandemic and Beyond", *Journal of Gastrointestinal Surgery*, vol. 24, no. 8, pp. 1692–1697, 2020. DOI: 10.1007/s11605-020-04623-5.

[13] M. C. Araujo, A. R. Silva, and F. A. Pereira, "A Systematic Review on Wound Classification Challenges in Clinical Practice", *Journal of Wound Care*, vol. 30, no. 3, pp. 132–140, 2021.

[14] M. Cabrera, E. Gómez, and A. Torres, "Pressure Injury Image Analysis with Machine Learning Techniques: A Systematic Review", *Journal of Biomedical Informatics*, vol. 107, p. 103 432, 2020.

[15] R. G. Sibbald, D. Krasner, and J. Lutz, "Challenges in wound assessment: Interrater variability and implications for clinical care", *Advances in Skin & Wound Care*, vol. 32, no. 8, pp. 388–395, 2019, [retrieved: July, 2025]. DOI: 10.1097/01.ASW.0000565797.24804.52.

[16] National Pressure Ulcer Advisory Panel and European Pressure Ulcer Advisory Panel, "Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline", International Guideline Development Group, Cambridge Media, Osborne Park, Western Australia, Tech. Rep., 2019, Also developed in collaboration with Pan Pacific Pressure Injury Alliance; third edition released 15 Nov 2019.

[17] L. Wang, X. Zhou, and H. Zhang, "Deep LearningBased Classification of Pressure Injuries Using Multimodal Imaging", *Theranostics*, vol. 15, no. 7, pp. 1662–1675, 2023.

[18] Y. Jiang, P. Liu, and J. Wang, "Automated Wound Classification Using Machine Learning Algorithms", *Computers in Biology and Medicine*, vol. 127, p. 104 057, 2020.

[19] L. B. Jørgensen, J. A. Sørensen, G. B. Jemec, and K. B. Yderstræde, "Methods to assess area and volume of wounds – a systematic review", *International Wound Journal*, vol. 13, pp. 540–553, 4 Aug. 2016, ISSN: 1742481X. DOI: 10.1111/iwj.12472.

[20] D. M. Anisuzzaman et al., "Image Based Artificial Intelligence in Wound Assessment: A Systematic Review", *Adv Wound Care (New Rochelle)*, vol. 11, pp. 687–709, 12 Dec. 2022, Epub 2021 Dec 20; published Sep 21 2021. DOI: 10.1089/wound.2021.0091.

[21] F. Veredas, H. Mesa, and L. Morente, "Binary tissue classification on wound images with neural networks and bayesian classifiers", *IEEE Transactions on Medical Imaging*, vol. 29, pp. 410–427, 2 Feb. 2010, ISSN: 02780062. DOI: 10.1109/TMI.2009.2033595.

[22] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "YOLO-Based Deep Learning Model for Pressure Ulcer Detection and Classification", *Healthcare*, vol. 11, no. 9, p. 1222, 2023, Received: 12 Mar 2023; Revised: 15 Apr 2023; Accepted: 22 Apr 2023; Published: 25 Apr 2023. DOI: 10.3390/healthcare11091222.

[23] S. Zahia, D. Sierra-Sosa, B. Garcia-Zapirain, and A. Elmaghraby, "Tissue classification and segmentation of pressure injuries using convolutional neural networks", *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 51–58, Jun. 2018, Epub 2018 Mar 3, ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2018.02.018.

[24] B. García-Zapirain, M. Elmogy, A. El-Baz, and A. S. Elmaghraby, "Classification of pressure ulcer tissues with 3D convolutional neural network", *Medical & Biological Engineering & Computing*, vol. 56, no. 12, pp. 2245–2258, Dec. 2018, Epub 2018 Jun 15, ISSN: 0140-0118. DOI: 10.1007/s11517-018-1835-y.

[25] S. M. Pereira, M. A. Frade, R. M. Rangayyan, and P. M. A. Marques, "Classification of color images of dermatological ulcers", *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 136–142, Jan. 2013, Epub 2012 Nov 15, ISSN: 2168-2194. DOI: 10.1109/TITB.2012.2227493.

[26] D. Ramachandram et al., "Fully Automated Wound Tissue Segmentation Using Deep Learning on Mobile Devices", *JMIR mHealth and uHealth*, vol. 10, e36977, 2022. DOI: 10.2196/36977.

[27] T. J. Liu et al., "Automatic Segmentation and Measurement of Pressure Injuries Using Deep Learning Models and a LiDAR Camera", *Scientific Reports*, vol. 13, p. 680, 2023. DOI: 10.1038/s41598-022-26812-9.

[28] D. Carvalho et al., "Enhancing Chronic Wound Assessment through Agreement Analysis and Deep Learning Models", *Scientific Reports*, vol. 15, p. 12 345, 2025. DOI: 10.1038/s41598-025-06703-5.

[29] N. Tajbakhsh et al., "Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation", *Medical Image Analysis*, vol. 63, p. 101 693, 2020. DOI: 10.1016/j.media.2020.101693.

[30] M. A. Kabir, N. Roy, M. E. Hossain, J. Featherston, and S. Ahmed, *Deep Learning for Wound Tissue Segmentation: A Comprehensive Evaluation using A Novel Dataset*, [retrieved: July, 2025], 2025. arXiv: 2502.10652 [eess.IV].

[31] H. Liu et al., "Current status, challenges, and prospects of artificial intelligence applications in wound repair theranostics", *Theranostics*, vol. 15, pp. 1662–1688, 2025, [retrieved: June, 2025]. DOI: 10.7150/thno.105109.

[32] Y. Patel et al., "Integrated image and location analysis for wound classification: a deep learning approach", *Scientific Reports*, vol. 14, 1 2024, ISSN: 20452322. DOI: 10.1038/s41598-024-56626-w.

[33] C. BroniBediako, J. Xia, and N. Yokoya, "RealTime Semantic Segmentation: A Brief Survey & Comparative Study in Remote Sensing", *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–33, 2023. DOI: 10.1109/MGRS.2023.3321258.

[34] J. Cheng, H. Li, D. Li, S. Hua, and V. S. Sheng, "A Survey on Image Semantic Segmentation Using Deep Learning Techniques", *Computers, Materials & Continua*, vol. 74, no. 1, pp. 1941–1957, 2023. DOI: 10.32604/cmc.2023.032757.

[35] C. Wang et al., *FUSeg: The Foot Ulcer Segmentation Challenge Dataset*, GitHub repository, [retrieved: May, 2025], 2021.

[36] D. M. Anisuzzaman, Y. Patel, J. Niezgoda, S. Gopalakrishnan, and Z. Yu, *AZH Wound Care Center Dataset (patches)*, GitHub repository, Preprocessed wound image patches from clinical data collected at the AZH Wound and Vascular Center, Milwaukee, WI, [retrieved: May, 2025], 2020.

[37] Y. N. Hwang, M. J. Seo, and S. M. Kim, "A Segmentation of Melanocytic Skin Lesions in Dermoscopic and Standard Images Using a Hybrid TwoStage Approach", *BioMed Research International*, vol. 2021, p. 5 562 801, 2021. DOI: 10.1155/2021/5562801.

[38] O. Ronneberger, P. Fischer, and T. Brox, "UNet: Convolutional networks for biomedical image segmentation", in *Medical Image Computing and ComputerAssisted Intervention – MICCAI 2015*, Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[39] International Telecommunication Union, *Recommendation ITUR BT.601-7: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*, https://www.itu.int/rec/R-REC-BT.601, Accessed: 2025-07-01, 2011.

[40] S. Li, A. H. Mohamedi, J. Senkowsky, A. Nair, and L. Tang, "Imaging in Chronic Wound Diagnostics", *Advances in Wound Care*, vol. 9, no. 5, pp. 245–263, 2020. DOI: 10.1089/wound.2019.0967.

[41] M. S. Nixon and A. S. Aguado, "8 - Region-based analysis", in *Feature Extraction and Image Processing for Computer Vision (Fourth Edition)*, M. S. Nixon and A. S. Aguado, Eds., Fourth Edition, Academic Press, 2020, pp. 399–432, ISBN: 978-0-12-814976-8. DOI: https://doi.org/10.1016/B978-0-12-814976-8.00008-7.

[42] M. Z. B. Jahangir, S. Akter, M. A. A. Nasim, K. D. Gupta, and R. George, *Deep learning for automated wound classification and segmentation*, Preprint, 2024.

[43] M. Alabdulhafith et al., "Automated wound care by employing a reliable UNet architecture combined with ResNet feature encoders for monitoring chronic wounds", *Frontiers in Medicine*, vol. 11, p. 1 310 137, 2024.