Automated 3D Wound Segmentation Using UV-based Feature Extraction and Deep Learning

Jeffrey Jenkins
Dept of EE & Computer Science
Catholic University of America
Washington, DC
54jenkins@cua.edu

Jonathan Nguyen
CHABSS
California State University
San Marcos, California
jonatngu@icloud.com

Lin-Ching Chang
Dept of EE & Computer Science
Catholic University of America
Washington, DC
changl@cua.edu

Salam Daher
Department of Informatics
New Jersey Institute of Tech
Newark, New Jersey
daher@njit.edu

Abstract_— Effective wound assessment requires accurate 3D wound segmentation, but annotated datasets are scarce due to privacy and the labor-intensive nature of manual labeling. This study presents a 3D wound segmentation model trained using synthetic wounds generated using a 3D scanner. Using advanced generative techniques, we created a diverse synthetic dataset for deep learning training, enabling precise 3D wound segmentation using the 2D UV-mapped texture from the 3D wound surface. Our approach yields positive results and is a novel alternative to traditional 2D image and 3D volume segmentation methods. This work paves the way for using synthetic data and multi-dimensional analysis to advance medical imaging workflows.

Keywords—Deep Learning, U-Net, Segmentation, Wound Imaging, Wound Image Processing.

I. Introduction

The healthcare industry faces a lack of objective data, which is critical for clinical decision making at both the data collection and interpretation stages so that it can adequately translate and improve patient outcomes. Current wound measuring methods lack accuracy and consistency. For example, wounds are often measured using rulers and Q-tips for length, width, and depth. Wound surface areas are estimated by Length x Width, which is suitable for measuring rectangles, but wounds are usually irregular; this approach is inaccurate and accounts for a significant overestimation of about 41% [1]. Wound measurement informs clinicians on course prediction, interventions, and treatment. Accuracy and consistency in measuring wounds are critical to facilitate proper assessments and treatments. Despite measurements being critical in assessing the patients' wound progression, existing methods are subjective, with high variability between providers and a lack of precision.

Using an off-the-shelf scanner (e.g., a smartphone), we can generate a 3D model of wounds, enabling healthcare providers to measure and track the wound dimensions consistently in 3D. Through interviewing medical practitioners, we have learned that this approach has substantially improved the accuracy and reliability of the wound measures [2,3] and reduces the time taken for wound assessment for healthcare providers by 50%. Healthcare providers also found the software to be user-friendly, trustworthy, and acceptable.

Currently, our system, known as ViMeT (Visualize, Measure and Track), requires manual tracing of the wound contour (perimeter) in 3D as the first step from which all other measures are automated. Based on the feedback from various stakeholders, we realized that integrating automated wound edge detection will significantly minimize the time required for the wound measurement process. To achieve this goal, this paper outlines an approach to establish an Artificial Intelligence (AI) data processing and training pipeline which will create models trained using real patients' wounds. However, as a first step, we have collected data from simulated wounds to train the model to detect the contour of simulated wounds. These simulations will help estimate the volume and types of real-world wound images needed, paving the way for developing and deploying a computationally and contextually scalable architecture to enable fully automated wound detection using real patient data, seamlessly integrated into our ViMeT system.

II. Background

Burns and wounds provide excellent examples of injuries that have tangible measures: Burns are estimated using what is called "The Rule of Nines" [4], which estimates the patient's body surface using the palm of the hand (one hand palm = 1%). The percentage of body burns determines treatment and next steps. Similarly, the depth of wounds is still measured using Q-tips [3], and the surface area is estimated by Length x Width, which works great for measuring rectangles, but wounds are usually irregular; that formula is inaccurate and accounts for a significant overestimation of about 41% [5,6]. These subjective methods are currently used in healthcare for measurement estimates; measuring burns and wounds occur on complex surfaces, which are hard for healthcare providers to measure accurately and consistently.

Wound measurement informs clinicians on course prediction, interventions, and treatment. A retrospective analysis of Medicare beneficiaries in 2018 revealed that approximately 8 million Americans suffer from wounds of different etiologies [7]. Regardless of the wound condition or where the care was provided, the healthcare cost associated with these incidents' ranges from $28.1 billion to $96.8 billion [7]. In addition, wounds, particularly pressure injuries, contribute to extended hospital stays and hospitalization costs—hospital-acquired pressure injuries associated costs exceed $26.8 billion yearly [8] these costs are not reimbursable since they are considered preventable injuries by the Center of Medicaid and Medicare Services. Consistency in measuring

wounds is critical to facilitate proper assessments and treatments.

Current wound measuring techniques include (1) Ruler and Q-Tips technique, (2) 2D image techniques, (3) 3D software technique. Wounds are frequently documented on 2D images taken digitally with smartphones or Polaroid cameras. These images are printed later and archived in patients' charts (e.g., binders) instead of being stored electronically or transferred to an Electronic Medical Record (EMR) as part of the patient's permanent record. When there is the capability to obtain images that interface directly with the EMR, the captured and uploaded images are in 2D. Unfortunately, there is no standard for picture taking (e.g., angle, lighting, resolution), resulting in significant variability in documenting the evidence of wounds pictorially. In addition, the depth and contour of the wounds are difficult to document pictorially. Moreover, the evaluator's documentation of wound measurements and appraisal is highly subjective and provider-driven (i.e., their description and measurement, the quality of the photos, and the points used for measurement). This subjectivity influences and affects decisions on treatment and interventions. Also, sometimes wound records are not readily available to other healthcare team members, further complicating the participation of healthcare providers in the assessment and documentation of the wound's course.

Without better tools to objectively evaluate and document wounds, patient outcomes are directly impacted by possibly affecting delays in appraising the course of wounds and treatment. For instance, different healthcare providers can use multiple modalities for measuring (i.e., which axis is chosen for length and width, especially when wounds have irregular shapes). Also, the depth of the wound is particularly challenging to measure or estimate, mainly since a cotton-tip applicator is commonly used as a measuring tool and no standard measurement instrument is currently used. Another challenge is access to the wound, mainly when wounds occur on non-flat surfaces or difficult areas to access. In other instances, the patient is not mobile enough to provide a clear view of the wound, making it challenging to assess wounds for one individual alone and compromising the assessment's acuity.

We developed a prototype that can objectively capture, visualize, measure and track any skin abnormalities including wounds and pressure injuries. We scan the patient's skin abnormalities using an off-the-shelf scanner (i.e., IReal2E, mobile phone), then we consistently measure and track the measurements. The scanning and measuring of 3D wounds are a new technology applied to a new situation intended to be used by healthcare providers and by healthcare educators. This technology applies to any type of wounds or lesions that appear on the skin. Overlapping photos of wounds are taken from multiple angles, and a 3D model (mesh geometry, texture) is generated from these photos using photogrammetry. The 3D model is then cleaned, and transformed (scale, position, rotation) and displayed in the software where healthcare providers can view the wound in 3D from multiple angles, then draw a contour around the boundaries of the wound. The length, width, depth, and area are generated automatically. The wounds are visualized and measured in 3D accurately representing shapes and colors. The current system is semi-automated and can keep track of the measures taken by patients and by providers over time. The ViMeT system provides measures along the surface which provides a sensitive indicator of wound progression that is not available in any other existing system. Our goal is to develop technology to fully automate the process.

III. Related Work

In the early stages of our work with ViMeT, we conducted a study at NJIT to evaluate the system that we developed [2]. In the IRB approved study, we asked 30 participants (NJIT students) to measure physically simulated wounds of pressure ulcers (loaner from Avkin [9]) vs. 2D images of the wounds vs. 3D scans of the wounds in our software. At this stage, we only had unconstrained measures. The intraclass correlation coefficient (ICC) was used to evaluate measurement reliability. These results showed that measuring wounds in software (ICC: 0.984) was more reliable than 2D images (ICC: 0.937), and also more reliable than physical measures (ICC: 0.934).

We conducted a study with 23 professional healthcare providers (registered nurses, nurse practitioners, physician assistants, and graduate nursing students), where they measured simulated wounds both physically (using the 3D prints) and virtually using our software. Results showed that our guided measurement method was significantly more accurate (physical accuracy: 93.29%, software accuracy: 97.68%), reliable (physical ICC: 0.833, software ICC: 0.957), and time efficient (49.26% time savings). The System had excellent usability (System Usability Scale score > 85), and the measurements were accepted by healthcare providers (91.54% acceptance rate). In addition, the length, width, and area measurements along the surface provided a sensitive indicator of wound changes ($p < 0.01$). One glaring issue remained, which was the fact that participants still had to manually draw the perimeter separating wound vs. non-wound first, before measuring or generating the rest of the measures (i.e. length, width, depth, surface area). Therefore, we have turned our focus to utilizing wound segmentation to fully automate the contour generation to improve the temporal efficiency and reliability of the measurements. Fig. 1 shows our whole development process.

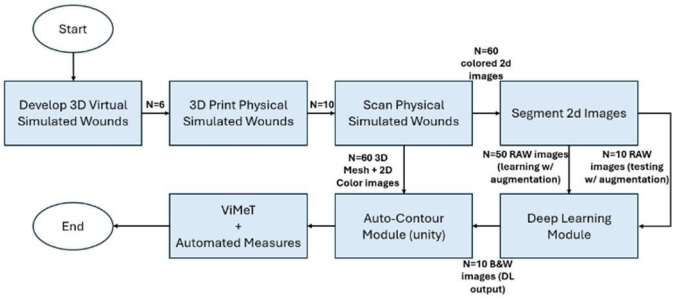

Fig. 1. - Block diagram for the development process.

Wound segmentation is not a new problem – it has been attempted with varying experimental setup. Recently, Wang et. al proposed a method of training a 2D segmentation model using a deep learning architecture based on RGB images of diverse

wounds [10]. This approach yielded formidable results, with a dice coefficient of 90.47%. However, this technique was focused on entire wound 2D images, where illumination, skin tone, and various other factors were reported to affect the accuracy and robustness of the model. Another deep learning method for 3D volumetric segmentation was published by Mahmud et. al, which made use of the Computed Tomography (CT) sandstone detection dataset, yielding an impressive 96.78% accuracy and 91.12% Intersection Over Union (IOU) [11]. This approach would extend nicely to additional monochromatic volume scans like MRI and perhaps ultrasound but would suffer with image features existing in the RGB domain. Later in our discussion we will indicate that wound segmentation should not be attempted with grayscale features, as it leads to a large drop in accuracy for all evaluated methods.

Our approach combines elements from these two related works mentioned above, along with a novel approach to utilize the UV-mapped RGB texture mapping onto a 2D image by the 3D scanner to influence the segmentation of the 3D object. In the next section, we will describe our data pipeline and basics of extracting the necessary data from the 3D object, then how we used this collection of information to create the segmentation training dataset and U-Net model.

IV. Materials and Methods

To build our dataset, we scanned 6 synthetic wound variations (W0-W5) using the PolyCam scanner [12] 10 times each. These wound variations were designed to simulate different stages of the healing process, with each variant containing realistic surface features as well as realistic geometric distortions around the site of the wound.

A. Capturing the 3D Synthetic wound sample

Fig. 2 is a representative example of how synthetic wounds appear, and rendered after scanning as a textured mesh.

Fig 2. - Virtual and 3D printed simulated wound. The left shows a top view of the mesh wireframe and its colored texture mapped. The center image shows different views of the virtual simulated wounds, and the left image shows the 3D printed wound.

Although this is a fabricated wound, Fig. 2 demonstrates the realism of our underlying data. Once a wound is scanned and exported, it produces a few files: a 3D object file containing the mesh's vertices, and a UV-mapped texture file where pixels map to a specific face on the 3D mesh.

B. Exporting UV-mapped texture and mask creation

To mitigate concerns about labeling bias, two independent annotators performed manual mask creation of the same 3D scans using different labeling tools; one labeler used Maya and the other used Blender. These annotations were saved as separate datasets for comparative evaluation. In each binary mask, the white pixels represent wound regions, and the black pixels indicate non-wound areas.

Fig. 3 illustrates the process of exporting a UV-mapped texture and how we generate the corresponding masks, which will be used later as training label for our deep learning model.

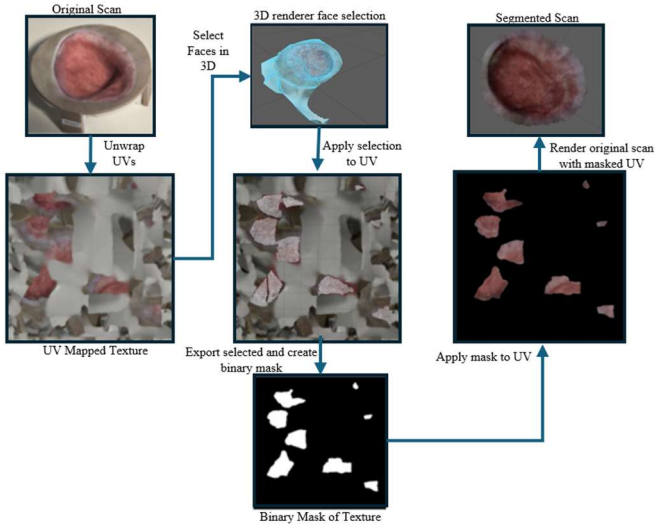

Fig. 3 - UV segmentation process. Our training data labels are created when exporting selected mesh into a binary mask.

C. Create U-Net architecture to support RGB images

We created a 3D U-Net architecture to extract features for training, using the UV-mapped texture as input and the corresponding binary mask as labels. Our development builds upon the widely used 2D U-Net architecture [13,14], recognized for its effectiveness in medical image segmentation due to its ability to capture fine-grained details while preserving spatial accuracy. Our architecture is shown in Fig. 4.

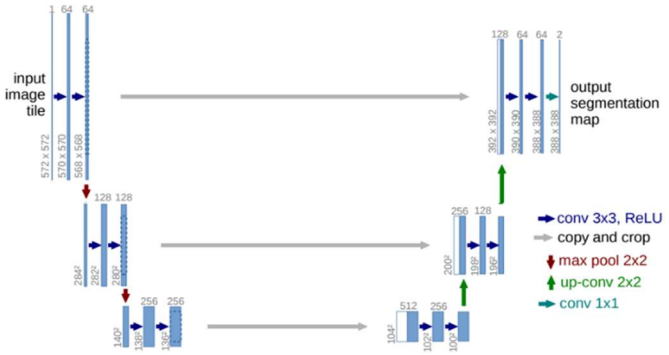

Fig. 4 - U-Net archicture: the min kernel size is 64, max is 512. He normalization used as the kernel initialize with relu activation, same padding, and a dropout value of .3.

We began by implementing the baseline U-Net with contracting and expansive paths to learn spatial and contextual features. This model was then tailored for wound segmentation by adjusting the number of layers and filters to match the complexity and resolution of our wound data. Due to the size of our input images (256×256×3), we modified the architecture by reducing the maximum number of filters in the deepest layer from 1024 to 512. This adjustment helps mitigate overparameterization, lowers computational demands, and

reduces overfitting risk, especially when working with limited datasets, while still preserving the model's capacity for effective feature extraction and segmentation. All original 8192x8192x3 RGB input images and their corresponding masks were resized and normalized to a resolution of 256x256x3, before being fed into the U-Net architecture.

D. Create Training data from wound scans

Prior to the training process, due to the small sample size of our data, we performed data augmentation, using only the translation, horizontal and vertical flip transformations, as shearing, rotation and scaling would most certainly create pixel deformation that leads to incorrect re-mapping post segmentation. This allowed us to generate additional synthetic data to be used for our model. Our run parameters were 10 epochs, 50 training steps and 10 validation steps.

V. Experimental Setup

To start our experiments, we organized all the scans into folders that correspond to the wound ID (W0-W5), within each folder was an image folder containing the 10 UV-mapping texture images from repeated trials on the same wound ID, and a label folder containing binary masks of the corresponding numbered UV-mapped texture image. This gave us a total of 60 ground truth images with which to perform our evaluation. Our evaluation contained several permutations of initial conditions and partitioning methods to test the viability of 3D segmentation by utilizing a model trained on 2D UV-mappings. We have made all the code supporting these experiments available on GitHub, the specific python implementation is not shown in this paper [15].

A. RGB vs. Grayscale

We attempted to build wound specific RGB models in grayscale to test the robustness of the wound features absent of color information. This experiment consistently yielded results where the RGB model outperformed grayscale models with the exact same training samples (and seed value) was greater than 15% accurate in all wounds.

B. No data augmentation vs. Data Augmentation

Then, we built wound specific RGB models without the use of data augmentation to test the robustness of the wound features with a limited dataset. This experiment consistently yielded results where the RGB models using data augmentation outperformed all the un-augmented models with the exact same training samples (and seed value) by greater than 36% accuracy in all wounds.

C. Wound specific models: Leave one out (LOO)

Next, we attempted to build wound specific RGB models by performing Leave One Out cross validation. The approach here was to withhold an entire folder, for instance W0 (10 labels, 10 RGB UV-mappings), for validation, while training with all the other wounds, W1-W5 (50 labels, 50 RGB UV-mappings).

D. Random sampling: Stratified K-Fold (SKF)

Finally, we ensure each fold has a balanced mix of subjects with different numbers of wound images. In addition, we balanced the splits based on a specific number of images per subject (subjects with 10 images vs. subjects with different counts). Then we create 5 folds where each fold contains a proportional representation of both groups, which prevents training bias and ensuring robust model validation across all subject types.

VI. Results and Discussion

In this section, we will perform data analysis to focus on evaluating the performance of the ML models. We calculate standard metrics such as Dice Similarity Coefficient (DSC), Jaccard, and IoU, to assess the segmentation accuracy of the ML-generated 2D contours against the ground truth [16]. Note that when a graph says 'Label 1' it is referring to training data created by labeler 1, and 'Label 2' refers to training data created by labeler 2, which are two distinct individuals. Each model experiment took roughly 1 hour to run on a 32GB CPU PC.

The results are as follows – Fig. 5 shows the accuracy of each experiment configuration for both Label 1 and Label 2 training sets. The accuracy gradually approaches maxima around epoch 40, and the most accurate model is the Leave One Out + stratified, boasting ~95% accuracy.

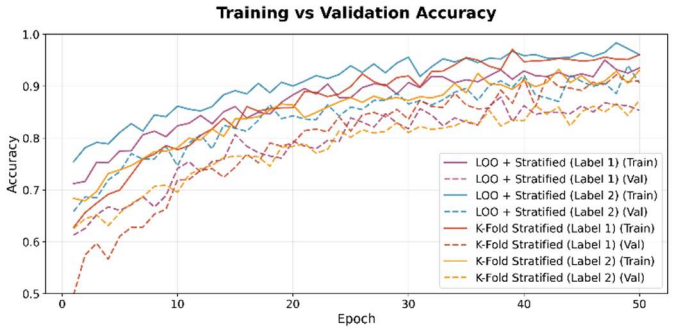

Fig. 5 - Training vs Validation accuracy for different model strategies.

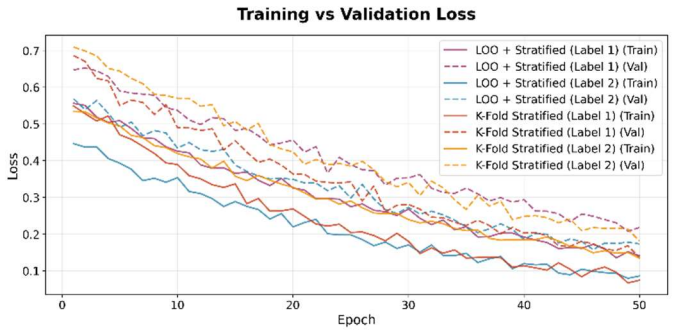

Fig. 6 - Training vs Validation loss for different model strategies

Above in Fig 6., the training and validation data behaves as expected, with the loss decreases over each epoch, eventually stabalizing at around 50 epochs. Future work will explore more epochs to find a minimum loss and maximal accuracy for this wound type.

Furthermore, overfitting analysis shows that the least amount of overfitting actually came from samples created by 'Label 1' which indicates that the labeler may matter. In addition, training

convergence shows that the speed of convergence does not depend on the model or labeler, since one of each labeler and one of each training strategy obtained consistent convergence.

Next, we explore the Wound segmentation performance using the IoU/Jaccard Index and Dice coefficient. It is evident below in both Fig 7. and Fig. 8 that the overall winner was Leave One Out + stratified with the labels provided by Labeler 2, but the results were fairly close across the board.

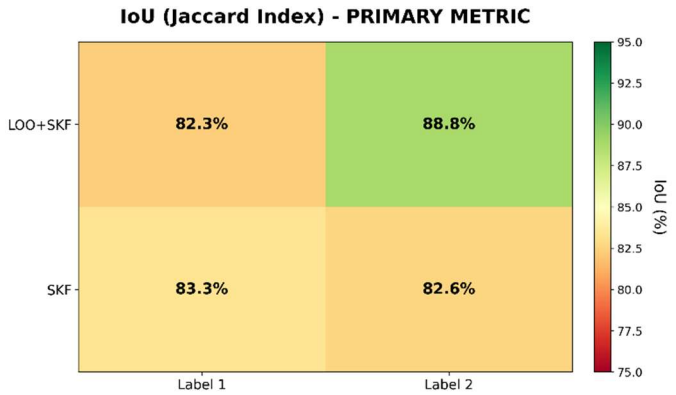

Fig. 7 - Performance Heatmap - IoU/Jaccard performance for different labelers and different model strategies.

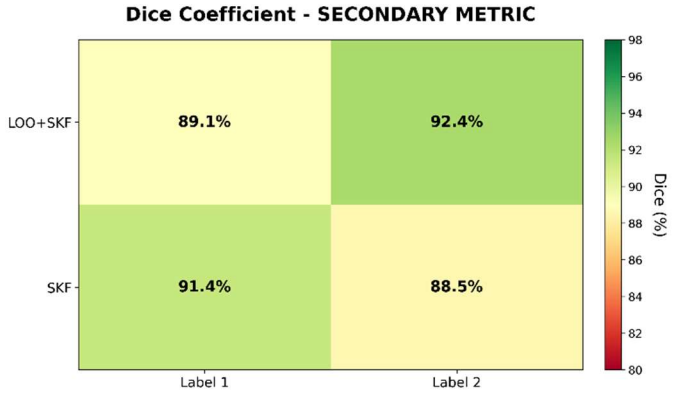

Fig. 8 - Performance Heatmap - Dice coefficient performance for different labelers and different model strategies.

While both SKF and LOO+SKF have are similar model behavior, the LOO+SKF offers more focused testing of features. It is clear from our work that utilizing UV-mapped textures is indeed a sustainable approach to performing 3D segmentation. These results give us confidence in moving forward with the next phases of our research and introducing our Deep Learning outcomes into the ViMeT pipeline to produce efficiencies for healthcare professionals.

VII. Future Work and Conclusion

In the future, we will be able to make use of the insight that specifically trained models gave us – which is that using stage specific models performs better than a model constructed from random wound samples. Additionally, varying conditions of the scan, such as the background, the skin tone, illumination, etc. will give us another dimension for analysis and we must explore what effect, if any, background features have on properly recognizing wound features.

The cell-phone scanning capability of PolyCam is also going to be an analysis vector in the next stage. It is our belief that the models built from the 3D scanning capability PolyCam offers will translate directly to the output of the cell-phone output. If this is the case, deploying a lightweight (but equally accurate) version of the model via mobile app could greatly speed up processing at the edge, directly on the cell-phone.

We can make use of similarity measures between an unknown 3D point cloud and co-registered point clouds of known wound types can be the first stage for recognizing which wound we may be looking at [17]. With this information, we could perform diffeomorphic registration to understand similarity to a particular wound healing stage and select that model as our segmentation model [18]. From there, establishing a pipeline to harmonize the processing pipeline for each type of wound we scan, as well as transformations we need to apply to it can be managed with workflow tools such as Apache Airflow, and monitored via a web UI [19].

In this paper, we have successfully demonstrated a novel approach to perform 3D segmentation of a 3D wound volume using a 2D UV-mapping texture. This work paves the way for contextually scaling the pipeline with automated segmentation of additional color based medical image volumes which rely on handheld scanners, such as many other wound types, or even dental scans. The application space of this technique can be computationally scaled as a cloud service for non-PolyCam devices – our model could be hosted and accessible via Application Programming Interface (API) or remote invocation throughout the world, from hospitals to the battlefield, where in both situations automatically understanding the priority of wound triage is timely and can save lives.

Acknowledgment

The authors would like to thank NJIT research assistant Nahreg Rastguelenian for providing segmentation masks, Dr. Dahlia Musa for building the surrounding pipeline for providing input, and handling our results, Christopher Frenchi for helpful discussion and our healthcare experts Prof. Frank Guido-Sanz and Prof. Mindi Anderson.

References

1. Rogers, Lee C., et al. "Digital planimetry results in more accurate wound measurements: a comparison to standard ruler measurements." (2010): 799-802.
2. D. Musa, F. Guido-Sanz, M. Anderson and S. Daher, "Reliability of Wound Measurement Methods," in IEEE Open Journal of Instrumentation and Measurement, vol. 1, pp. 1-9, 2022, Art no. 9700109.
3. D. Musa, F. Guido-Sanz, M. Anderson, D. A. Díaz and S. Daher, "Impact of Digital Guidance on Accuracy, Reliability, and Time Efficiency of Wound Measurements," in IEEE Open Journal of Instrumentation and Measurement, vol. 4, pp. 1-11, 2025
4. https://www.omnicalculator.com/health/parkland-formula
5. L. C. Rogers, N. J. Bevilacqua, D. G. Armstrong, and G. Andros, "Digital planimetry results in more accurate wound measurements: A comparison to standard ruler measurements," J. Diabetes Sci. Technol., vol. 4, no. 4, pp. 799–802, 2010.

6. E. Nichols, “Wound assessment part 1: How to measure a wound,” Wound Essentials, vol. 10, no. 2, pp. 51–55, 2015.
7. Sen, C. K. (2019). Human wounds and its burden: An updated compendium of estimates. Advances in Wound Care, 8(2), 39-48. https//: doi.org/10.1089/wound.2019.09463
8. Padula WV, Delarmente BA. The national cost of hospital-acquired pressure injuries in the United States. Int Wound J. 2019 Jun;16(3):634-640. doi: 10.1111/iwj.13071. Epub 2019 Jan 28. PMID: 30693644; PMCID: PMC7948545.
9. https://avkin.com/avwound-wound-simulator/
10. Wang, C., Anisuzzaman, D.M., Williamson, V. *et al.* Fully automatic wound segmentation with deep convolutional neural networks. *Sci Rep* **10**, 21897 (2020).
11. Mahmud BU, Hong GY, Mamun AA, Ping EP, Wu Q. Deep Learning-Based Segmentation of 3D Volumetric Image and Microstructural Analysis. Sensors (Basel). 2023 Feb 27;23(5):2640.
12. https://learn.poly.cam/hc/en-us
13. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI).
14. Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., & Merhof, D. (2024). Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
15. https://github.com/jjenki11/wound-segmentation.
16. Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1), 210.
17. Hutchinson EB, Schwerin SC, Radomski KL, Sadeghi N, Jenkins J, Komlosh ME, Irfanoglu MO, Juliano SL, Pierpaoli C. Population based MRI and DTI templates of the adult ferret brain and tools for voxelwise analysis. Neuroimage. 2017 May 15;152:575-589.
18. Irfanoglu MO, Nayak A, Jenkins J, Hutchinson EB, Sadeghi N, Thomas CP, Pierpaoli C. DR-TAMAS: Diffeomorphic Registration for Tensor Accurate Alignment of Anatomical Structures. Neuroimage
19. J. Jenkins, L. -C. Chang, E. Hutchinson, M. O. Irfanoglu and C. Pierpaoli, "Harmonization of methods to facilitate reproducibility in medical data processing: Applications to diffusion tensor magnetic resonance imaging," *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2016, pp. 3992-3994.