Instituto Superior
de Engenharia
Politécnico de Coimbra

DEPARTMENT OF SYSTEMS AND COMPUTER ENGINEERING

Segmentation of Wounds and Pressure Ulcers

Project Report to fulfill the Master's degree in Informatics Engineering

Specialization in Intelligent Data Analysis

Author

Diogo Cid Lino de Carvalho

Supervisor

Mateus Daniel Almeida Mendes

Carlos Manuel Jorge da Silva Pereira

Segmentation of Wounds and Pressure Ulcers

Coimbra, december 2024

INSTITUTO POLITÉCNICO
DE COIMBRA

INSTITUTO SUPERIOR
DE ENGENHARIA
DE COIMBRA

Resumo

Os profissionais de saúde enfrentam desafios ao avaliar e segmentar feridas em imagens médicas, uma tarefa essencial para controlar a evolução da cicatrização e planear tratamentos adequados. O presente trabalho propõe a aplicação de modelos de deep learning para a segmentação de feridas, explorando duas arquiteturas principais: YOLOv9 e Medical Transformer (MedT). O objetivo é identificar a melhor abordagem para automatizar a segmentação de diferentes tipos de feridas, reduzindo a subjetividade e o tempo associados aos métodos tradicionais. Para posteriormente obter as dimensões das feridas.

Ao realizar este estudo, foi preparado um conjunto de dados que combina imagens recolhidas de fontes online e de uma parceria com uma instituição de saúde. Essas imagens foram cuidadosamente anotadas para treinar os modelos propostos. A YOLOv9 destacou-se pela eficiência na identificação de marcas de calibração, enquanto o MedT demonstrou maior consistência na segmentação de feridas, aproveitando as capacidades de atenção global dos transformers.

Os resultados obtidos mostram que, apesar dos desafios, como falsos positivos e limitações do conjunto de dados, os modelos apresentam potencial significativo para melhorar a avaliação clínica de feridas. Este projeto representa um avanço na aplicação de inteligência artificial na área da saúde, contribuindo para o desenvolvimento de sistemas automáticos de segmentação que podem servir como ferramentas de apoio à decisão para profissionais de saúde.

Palavras-chaves: feridas, segmentação, *deep learning*, rede neuronais

Abstract

Healthcare professionals face challenges when assessing and segmenting wounds in medical images, an essential task for monitoring the progress of healing and planning appropriate treatments. This paper proposes the application of deep learning models for wound segmentation, exploring two main architectures: YOLOv9 and Medical Transformer (MedT). The aim is to identify the best approach for automating the segmentation of different types of wounds, reducing the subjectivity and time associated with traditional methods. In order to subsequently obtain the dimensions of the wounds.

To carry out this study, a dataset was prepared combining images collected from online sources and from a partnership with a healthcare institution. These images were carefully annotated to train the proposed models. YOLOv9 stood out for its efficiency in identifying calibration marks, while MedT showed greater consistency in segmenting wounds, taking advantage of the transformers' global attention capabilities.

The results obtained show that, despite challenges such as false positives and limitations of the data set, the models have significant potential to improve the clinical assessment of wounds. This project represents a breakthrough in the application of artificial intelligence in healthcare, contributing to the development of automated segmentation systems that can serve as decision support tools for healthcare professionals.

Keywords: wounds, segmentation, deep learning, neural networks

Epigraph

Success is getting what you want. Happiness is wanting what you get.
Dale Carnegie

Acknowledgments

It is with enormous gratitude that I thank the supervisors of this project, especially Professor Mateus, for all his dedication, patience and wisdom. His vision and continuous support were essential in transforming challenges into learning and achieving the objectives of this work.

To my family, my greatest foundation, I express my deepest gratitude for their unconditional love, understanding and support at every moment of this journey. This work is, above all, dedicated to you, with all my affection and gratitude.

My sincere thanks go to my friends, for their encouragement and constant support, and to MedicineOne, for sharing the problem and the challenge, which enriched and shaped this project with a more practical and relevant purpose.

I would also like to thank Ana Pais, who is part of the research center at Unidade Local de Saúde do Baixo Mondego, and the Board of Directors of ULSBM, EPE, for their support. To Dr. Ana Sofia Bento, Coordinator of the health unit at USF São Julião da Figueira, and a special thanks to Nurse Cardoso, for his generosity in sharing his time and knowledge, which were of enormous value to this work.

To all of you, my sincerest and deepest thanks.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AHRF	Associated Hierarchical Random Field
ATM	Attention-to-Mask
AZH	Advancing the Zenith of Healthcare
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Networks
COCO	Common Objects in Context
DICOM	Digital Imaging and Communications in Medicine
DFUC	Diabetic Foot Ulcer Challenge
DS	Detect-and-Segment
DSC	Dice Similarity Coefficient
FNN	Forward Neural Network
FP	False Positives
FN	False Negatives
FUSeg	Foot Ulcer Segmentation Challenge
GELAN	Generalized Efficient Layer Aggregation Network
GIMP	GNU Image Manipulation Program
IoU	Intersection over Union
LR	Learning Rates
LoGo	Local-Global
MCC	Matthews Correlation Coefficient
MedT	Medical Transformer
MICCAI	Medical Image Computing and Computer-Assisted Interventions
MIoU	Mean Intersection over Union
MS COCO	Microsoft Common Objects in Context
PGI	Programmable Gradient Information
PUSH	Pressure Ulcer Scale for Healing
ReLU	Rectified Linear Unit
RGB-D	Red, Green, Blue, and Depth
TN	True Negatives
TP	True Positives
TTA	Test-Time Augmentation
ULSBM	Unidade Local de Saúde do Baixo Mondego
USF	Unidade de Saúde Familiar
ViT	Vision Transformers
YOLO	You Only Look Once

1 Introduction

This chapter provides an overview of the significance of wound segmentation in healthcare, emphasizing the challenges involved and the motivation for using automated methods. The following sections introduce the problem, discuss the importance of accurate wound assessment, and present the objectives and goals of this research project.

1.1 Importance of Wound Segmentation

The segmentation of wounds and pressure ulcers is an important task in the field of medicine and healthcare, as acute wounds are one of the main causes of reduced quality of life for patients. If wounds are left untreated, they can cause serious complications that can lead to amputations or even death [1]. Regular examinations aim to minimize the damage and are essential to prevent deterioration. When healthcare professionals assess wounds, it is essential to accurately identify and isolate the wound area on a medical image and be able to separate unhealthy tissue from healthy tissue. This is important for monitoring the healing progress, determining the extent of the wound and planning appropriate treatments.
However, manual wound segmentation is a time-consuming and error-prone process. Nowadays, wound assessment is typically done visually, using a ruler in which a calculation is made by measuring the height times the width and the result is noted down [2]. Automating this process is highly desirable, as it can save time and reduce the possibility of human error. This is where image wound segmentation comes in.
Some of the challenges in segmenting wounds are represented by their appearance, since they can vary in terms of size, shape, location on the body, age, skin tone and healing time. The quality of the images, taking into account the lighting, the quality of the camera and the environment in which these images are captured, are also factors that make it difficult to accurately segment wounds.
The need to perform this action in real time with reliable results, especially in an emergency situation and in sterile environment adds another layer of challenge, as the segmentation algorithms need to be fast and efficient.

1.2 Project objectives

This thesis aims to improve the management of wounds in healthcare. The main objective of this work is to identify the optimal deep learning architecture for a system

to segment different types of wounds using manually altered images. The study first focuses on data collection, alteration, followed by a suggestion and implementation of wound segmentation models in which the wound area can be concluded using the best model. The main aim of this system will be to provide medical staff with a valuable tool to make quick and informed decisions about the size of wounds. By automating wound segmentation, the aim is to reduce the subjectivity, time and inter-observer variability associated with traditional assessment methods. The aim of this thesis is therefore to initiate an automated segmentation system for wounds with the potential to be widely used in clinical practice and serve as a valuable tool to support medical staff in the diagnosis and treatment of patients with wounds.

1.3 Document structure and organization

This document consists of seven chapters, each structured to provide a comprehensive understanding of the research problem and the proposed solutions. Chapter 1 serves as an introduction to the problem, as well as an overview of the research objectives. It highlights the importance of wound segmentation and presents the motivation behind the study. Chapter 2 explores the theoretical foundations essential to understanding the problem at hand, including the definition and main characteristics of wounds, as well as the Pressure Ulcer Scale for Healing (PUSH). In addition, the literature review methodology is presented and the main studies that support the current understanding of wound segmentation are highlighted.

Chapter 3 provides an overview of the commercial solutions currently available for wound measurement and management, presenting a comparative analysis of the main solutions on the market. Chapter 4 outlines the methodologies used to select and evaluate the proposed models. This chapter includes a description of the selected architectures, YOLO and MedT and discusses the evaluation metrics used to gauge their performance. In addition, the tools and technologies used in the development of this work are presented, including software frameworks, development environments and data annotation tools.

Chapter 5 explains the data collection and preparation process, as well as the structure of the final dataset. This chapter discusses the data collection methodology, the datasets initially researched and the data augmentation techniques used. A detailed overview of the dataset used for train, validation and testing is also provided. Chapter 6 presents the results of the experiments conducted with the selected models, including an analysis of the performance of the YOLO and MedT models. A comparative study of the models is also presented, highlighting their complexities and performance differences by comparing the results with the state of the art research.

Chapter 7 focuses on Wound Area Calculation, presenting the methodologies for pro-

cessing segmentation masks and calculating wound areas.

Finally, the Conclusion, with Chapter 8 closing the work, also outlines the limitations of the current study and suggests directions for future research.

2 Theoretical Foundations

In this chapter, we analyze the fundamental concepts needed to understand the context of this work, as well as the relevant articles that deal with wound segmentation. Firstly, the definition of a wound and its characteristics, followed by an analysis of the main techniques and methodologies applied in recent years in this field.

2.1 Definition of wound

S. Ather and K. G. Harding [3], show that a wound, by definition, represents an interruption in the integrity of the skin or other body tissues, resulting from a traumatic injury or a surgical incision. It is characterized by a breakdown in the protective function of the skin, involving a loss of continuity of the epithelium, and can affect the underlying connective tissues, such as muscles, bones and nerves. Wounds can be caused by a variety of causes, including physical trauma, cuts, exposure to chemicals, extreme temperature variations, friction, shear force, constant pressure or as a result of medical conditions such as ulcers or carcinomas.

As we can see in [4] one specific type of wound that is commonly observed is a pressure ulcer, like Figure 2.1. This condition occurs when the skin and deeper tissues under the skin are damaged due to the continuous application of pressure to a specific area for a prolonged period. This results in decreased blood flow to the affected area. Pressure ulcers are often associated with patients who are hospitalized for long periods of time and are unable to move, which increases the risk of developing this type of injury.

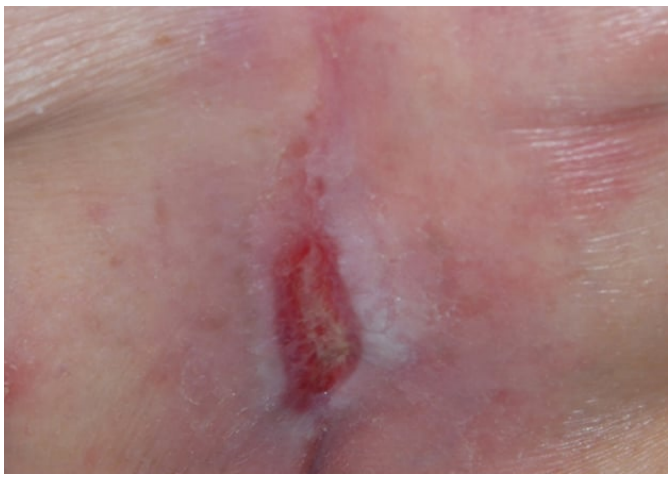

Figure 2.1: Example of pressure ulcers.[1]

.

2.1.1 Important characteristics of wounds

According to the Practical Guide [5], in order to manage a wound or pressure ulcer effectively, it is essential to consider three main parameters:

- **Wound/Pressure ulcer area**: The extent of the affected area is a crucial indicator of treatment progress. A smaller area suggests closer proximity to complete wound healing.
- **Amount of Exudate**: This fluid plays a vital role in healing, providing necessary nutrients and maintaining an ideal moisture balance, which facilitates the healing process. However, too much exudate can delay healing or even enlarge the wound area.
- **Wound/Pressure Ulcer Tissue Type**: The nature of the wound tissue is an important factor. Common categories include:
  1. Healed
  2. Epithelialized
  3. Granulated
  4. Devitalized
  5. Necrotic

  Each of these types reflects a different stage in the healing process and requires specific treatment approaches.

[1]Source: *https://www.mayoclinic.org/diseases-conditions/bed-sores/symptoms-causes/syc-20355893*, accessed on 2024-03-17

2.1.2 PUSH – Pressure Ulcer Scale for Healing

According to Sergio Almeida *et al.* in [6] the PUSH Scale (Pressure Ulcer Scale for Healing) is an objective method for assessing progress in the treatment of pressure ulcers (bedsores). This scale is used by healthcare professionals to monitor how the ulcer is evolving over time. The assessment is based on three main criteria:

- **Wound Area**: This part of the scale measures the size of the wound. The score ranges from 0 to 10, depending on the size of the affected area. For example, a wound that does not exist (0 cm^2) receives 0 points, and a wound larger than 24 cm^2 receives 10 points. Intermediate sizes have corresponding scores between 1 and 9, as follows:
  - 0 cm^2 - 0 points;
  - $<$ 0.3 cm^2 - 1 point;
  - 0.3 to 0.6 cm^2 - 2 points;
  - 0.7 to 1.0 cm^2 - 3 points;
  - 1.1 to 2.0 cm^2 - 4 points;
  - 2.1 to 3.0 cm^2 - 5 points;
  - 3.1 to 4.0 cm^2 - 6 points;
  - 4.1 to 8.0 cm^2 - 7 points;
  - 8.1 to 12.0 cm^2 - 8 points;
  - 12.1 to 24.0 cm^2 - 9 points;
  - $>$24.0 cm^2 - 10 points;
- **Amount of Exudate**: Exudate is the fluid that can accumulate in the wound. The score for this category ranges from 0 to 3. A wound with no exudate receives 0 points, while one with a large amount of exudate receives 3 points. Intermediate scores are given for small and moderate amounts of exudate, according to the following distribution:
  - Absent - 0 points;
  - Small - 1 point;
  - Moderate - 2 points;
  - Large - 3 points;
- **Tissue Type**: This criterion assesses the condition of the wound tissue. The score ranges from 0 to 4, based on the type of tissue present. For example, a wound with scar tissue (fully healed) receives 0 points, while one with necrotic (dead) tissue receives 4 points. Other types of tissue, such as epithelialization and granulation,

receive intermediate scores:

- Scarred - 0 points;
- Epithelialization - 1 point;
- Granulation - 2 points;
- Devitalized - 3 points;
- Necrotic - 4 points;

The total value on the PUSH Scale is calculated by adding up the scores in these three categories. It can range from 0 to 17, where 0 indicates a fully healed wound and 17 indicates a wound in severe condition. Regularly monitoring the PUSH score of a pressure ulcer helps healthcare professionals assess whether the wound condition is improving or worsening over time.

2.2 Methodology for literature review

Report sources

Table 2.1 shows the sources used to collect the state of the art, as well as the number of results and the number of articles selected.

Tabela 2.1: Date of search, website of reports, keywords used, number of results filtered and reports selected.

Date	Website	Keywords	Results	Chosen
2023-09-28	Google Scholar	wounds, segmentation neural, network	12000	3
2023-09-28	DBLP computer science bibliography	wound, segmentation	22	0
2023-10-14	Google Scholar	wounds, segmentation, deep learning	7430	11
2023-10-14	ScienceDirect	wounds, segmentation, deep learning	992	3
2023-10-14	Scopus	wounds, segmentation	0	0
2024-01-08	Google Scholar	transformers, segmentation, wounds	6770	8

Search strategy

The search strategy includes terms such as "segmentation", "deep learning", "wounds" and "pressure ulcers". In order to narrow down the results, the search filters were used, as well as the aforementioned criteria.

Inclusion and Exclusion Criteria

- Studies published between 2020 and 2023 will be included.
- Revised articles with more than 10 citations will be considered.
- Studies that do not directly address the relationship between segmentation and wounds, studies with small samples and studies not available in English will be excluded.

Review of Related Works

2.3 Review of Related Works

This section provides an overview of significant works related to wound segmentation using deep learning techniques.

2.3.1 Detect-and-segmentation: A deep learning approach to automate wound image

Scebba [7] developed a new deep learning approach to automate the segmentation of wound images. Called Detect-and-Segment (DS), this approach consists of three main steps:

- A first deep learning model is used to detect the location of the wound in the image.
- A second model is used to isolate the wound by centering it in the center of the image.
- Finally, a third model is used to segment the wound and identify all the pixels that belong to the wound.

The authors trained the DS approach models on six independent datasets with images of diabetic foot wounds. The dataset includes images with different types of wounds, complex backgrounds and variable lighting.
The DS approach was able to segment the wounds with high accuracy, even in images with complex backgrounds and different types of wounds. The Matthews Correlation Coefficient (MCC), a performance measure that takes into account both accuracy and sensitivity, was used. The results showed that the DS approach was able to improve the MCC from 0.17 to 0.85 on a test data set.
The article also states that the DS approach can be used to train wound segmentation models with up to 90% less training data, without affecting segmentation performance. In conclusion, the DS approach is considered a promising approach for automating wound image segmentation. The automation of wound segmentation can help clinicians to assess the condition of wounds faster, more accurately, and to make better treatment decisions.

2.3.2 Fully Automatic Wound Segmentation with Deep Convolutional Neural Networks

Chuanbo Wang *et al.* [8] use a convolutional neural network (CNN) to learn wound features and segment wounds from images. In this approach, a large dataset of wounds images is built with segmentation annotations made by wounds experts. This dataset consists of 1109 images of foot ulcers from 889 patients serving the CNN to be trained. The dataset includes images with different types of wounds, complex backgrounds and different lighting conditions.

In order to unify the size of the images in the dataset, the wound was first located by placing bounding boxes around it using a YOLOv3 model which serves as an object locator and was used for image-labelling. As pre-processing a series of techniques

were applied, cropping, zero-padding and data augmentation, in order to increase the training set, the result of these techniques, was a dataset of 5000 images.

The CNN structure proposed for wound segmentation is based on the MobileNetV2 architecture, which is a convolutional neural network that is widely used for computer vision tasks, lightweight, efficient and suitable for mobile applications. For the authors, the choice of this network can benefit professionals, doctors and patients by allowing instant segmentation of the wound and measurement of the wound area immediately after the photo is taken using mobile devices such as smartphones and tablets.

Precision, Recall and Dice were used as evaluation metrics, and other experiments were carried out on other models for comparison purposes. The models used were VGG16, SegNet, U-net and Mask-RCNN.

In the comparison made, the method proved to be effective in the field of image segmentation, always obtaining the best Dice score.

2.3.3 Automatic Foot Ulcer Segmentation Using an Ensemble of Convolutional Neural Networks

Amirreza Mahbod *et al.* [9] propose a method based on two convolutional neural networks, LinkNet and U-Net are to segment foot ulcers in medical images.
Some of the main steps this method is composed of:

- Data pre-processing: Medical images are pre-processed to remove noise and improve image quality.
- Segmentation: The pre-processed images are segmented using a set of two convolutional neural networks.
- Post-processing: The segmentation results are post-processed to remove false positives and improve segmentation accuracy.

The set of CNN used consists of the LinkNet architecture network and the UNet architecture network. According to the authors, LinkNet networks are lightweight and efficient, and UNet networks are accurate and robust.
Three public datasets were used, some of which are made up of more than one type of wound, from which foot ulcer image data was selected. Augmentation techniques were applied.
The Dice score was used as the main evaluation metric and the results showed that the method used is capable of segmenting ulcers with high accuracy, with a score of 92.07% in the foot ulcer segmentation challenge.
The authors conclude that the proposed method is a promising approach for the automatic segmentation of foot ulcers in medical images, having been ranked first in the

recent MICCAI 2021 FUSeg challenge.

2.3.4 Fully Automated Wound Tissue Segmentation Using Deep Learning on Mobile Devices: Cohort Study

Ramachandram *et al.* [2] present an approach for the automatic segmentation of wound tissues using deep learning on mobile devices.
The authors have developed two methods for segmenting wound tissues:
AutoTrace: This is a deep convolutional encoder-decoder model with attention mechanisms.
AutoTissue: This is a CNN based on the EfficientNetB0 architecture, with an encoder-decoder focused on segmenting wounds rather than skin.
The authors used a dataset of 465,187 images of wounds to train the CNN.
This dataset is one of the most complete and has been labelled by healthcare professionals. The CNN was able to learn a model of the wound tissues with high precision, with a hit rate of 0.8644.
The article also states that their proposal identifies four classes of wound tissue: granulation tissue, healing tissue, necrotic tissue and infection.
Accurate identification of wound tissue can be used to improve wound care, making it more effective and efficient.
Automatic segmentation of wound tissues can be a valuable tool for doctors when treating wounds. Accurate identification of wound tissue can be used to improve the effectiveness and efficiency of treatment, which can lead to better outcomes for patients.
For example, identifying necrotic tissue can help doctors remove dead tissue, which can speed up the healing process. Identifying infection can help doctors prescribe the appropriate treatment for the infection. Identifying granulation tissue and scar tissue can help doctors develop an effective treatment plan.
Automatic segmentation of wound tissue is still under development, but has the potential to be a transformative tool for wound care.

2.3.5 An Automatic Wound Detection System Empowered by Deep Learning

Muhammad Adnan *et al.* [10] explore deep learning to automatically detect and classify wounds. The article proposes a technique that uses the YOLO v3 model to detect, locate and classify wounds into four main categories: suture wound, cutting wound, open wound and normal skin. Experimental results show that the proposed technique

is more efficient and robust, with an accuracy of 99%.

The article addresses the importance of early wound detection for the healing process. A wound detection method based on the YOLO v3 model is proposed. It uses a dataset of 400 images which, after a data augmentation process, results in 1550 images. Experimental results show that the proposed method is efficient and robust, achieving an accuracy of 99%. The article shows a different approach to those used in the state of the art, but it is very contributory to this report.

2.3.6 Wound Detection by Simple Feedforward Neural Network

Marijanović *et al.* [11] present a solution based on simple feedforward neural network, but efficient considering it is not a CNN, designed for wound segmentation purposes. The main aspects of this study are as follow.
The feedforward neural network developed for this research was used for wound segmentation, using data from the MICCAI 2021 Foot Ulcer Segmentation Challenge (FUSeg).
The algorithm is one component of a larger system designed to analyze chronic wounds. This system incorporates a robotic arm, an RGB-D camera and a 3D scanner. The main neural network consisting of five fully connected layers is demonstrated. The first four layers use Rectified Linear Unit (ReLU) activation functions and the final layer uses a sigmoid activation function. Other networks with more layers were tested, containing seven and nine fully connected layers.
This choice of design explains the simplicity and effectiveness of the network in performing the segmentation task.
A significant contribution of this paper is the introduction of a method using a sliding window procedure that combines with the simple feedforward neural network classifier.
The metrics obtained in this experimental study were: recall value of 0.77, precision of 0.72 and i, F1 score of 0.74 for the combined models.

2.3.7 Semantic Segmentation of Smartphone Wound Images: Comparative Analysis of AHRF and CNN-Based Approaches

Ameya Wagh *et al.* [12] addresses the problem of analyzing wound images, which is known to be a relevant topic in medicine and in monitoring the evolution of wounds. The central objective of the article is to compare the performance of two different approaches to segment wounds in images: the Associated Hierarchical Random Field (AHRF) technique and approaches based on CNN, including FCN, U-Net and DeepLabV3.

To solve this problem, the authors carried out a systematic and comprehensive analysis, applying the aforementioned methodologies to wound images, making use of three different datasets, including data from the University of Massachusetts or from a public data collection on the internet. This study used pre-processing techniques such as data augmentation and also a technique to improve image contrast called Contrast-Limited Adaptive Histogram Equalization (CLAHE). These models were compared based on metrics such as segmentation accuracy (measured by dice score), inference time, amount of training data required and performance on different wound sizes and tissue types.
Focusing on the results of the CNN approaches, these networks show good results and were able to generalize well in images of wounds with different types of skin tissue and wounds, even with the presence of clutter in the background of the image. There is a fairly exhaustive study, but the DeepLabV3 network, in general, was the one that obtained the best results.

2.3.8 A Survey of Wound Image Analysis Using Deep Learning: Classification, Detection, and Segmentation

Ruyi Zhang *et al.* in the study [13] present a comprehensive analysis of the use of deep learning in wound image analysis. This article analyzes a publicly available dataset and explores different network models applied to wound classification, detection and segmentation tasks. The study highlights the challenges in wound image analysis and discusses future prospects in the field, addressing the importance of reliable data and effective pre-processing methods. In addition, significant contributions to the diagnosis and treatment of wounds are presented, demonstrating a major impact of the potential use of deep learning in health and medicine. As an example, in the case of wound classification, we can find studies of burns that identify an injury by dividing a body into 4 parts, and also by giving a level to the burn.

For wound segmentation, there are studies, as already mentioned, to estimate the total surface area of the burn.

For diabetic foot ulcers, we found studies that already compare transformers models with versions of Yolo that are superior to those found so far. For wound segmentation, there are studies, as already mentioned, to estimate the total surface area of the burn. The article mentions the best accuracy for burns, diabetic foot ulcers, pressure ulcers and other types of wounds. For these types of wounds, the segmentation is divided into two regions. In the case of burns, the region and depth are segmented to determine the different layers of skin affected. For diabetic foot ulcers, pressure ulcers and other types of wounds, a segmentation is made at the level of region and depth and skin. The best results were achieved by the DeeplabV3 network, with a dataset of 2893 images,

achieving an accuracy of 99.25.

The article also mentions that the network most used in the research studies and for segmentation is U-Net.

2.3.9 Deep learning approach based on superpixel segmentation assisted labeling for automatic pressure ulcer diagnosis

In [14] the authors present a detailed study on the use of deep learning for the automatic diagnosis of pressure ulcers. Labeling-assisted superpixel segmentation is used to classify ulcer tissues and segment wounds and re-epithelialization.

To put it in context, superpixel segmentation is a technique that groups neighboring pixels with similar characteristics into "superpixels" which simplifies the image and reduces computational complexity. On the other hand, re-epithelialization is a medical term that refers to the healing process of a wound, where epithelial cells grow over the wound to form a new layer of skin, which is crucial for assessing the progress of wound healing.

Five deep learning models were used (U-Net, DeeplabV3, PsPNet, FPN and Mask R-CNN) with the ResNet-101 encoder, trained on two datasets containing 2836 and 2893 images, respectively. The DeeplabV3 model showed the best performance in the detection and segmentation tasks, with a precision of 0.9915, recall of 0.9915 and accuracy of 0.9957. This study makes a significant contribution to the diagnosis and treatment of pressure ulcers, offering an automatic tool that can detect signs of healing, monitor the progress of healing, estimate the size of the wound and suggest the need to remove devitalized tissue.

2.3.10 SWS-NET: An Image Segmentation Framework For Chronic Wounds Based On Self-Supervised Learning

Jiyun Li *et al.* in [15] presents an innovative approach to image segmentation of chronic wounds. Using a model based on self-supervised learning, the study addresses the challenge of segmenting chronic wounds with limited data. The model, called SWS-NET, takes advantage of large sets of unlabeled data in the pre-training phase to learn useful representations, which are later adjusted with a smaller amount of labeled data. It involves using a set of unlabeled data so that the model learns to structure and understand the data on its own in order to understand the characteristics of the data, without the direct intervention of manual labeling.

A model can be trained to predict one part of an image based on another part. By hiding a section of the image and trying to predict that hidden section, the model learns to understand visual features without the need for labeling.

The experimental results show significant improvements in precision, recall and MIoU (Modified Intersection over Union) after self-supervised learning, demonstrating the effectiveness of the method compared to traditional supervised approaches. This study makes a significant contribution to the field of medical image segmentation, especially for chronic wounds, where labeled data is scarce.

2.3.11 Effectiveness of Semi-Supervised Active Learning in Automated Wound Image Segmentation

As noted in [16] a study on the use of semi-supervised active learning to segment wound images using CNN and presents a reliable decision support system for clinical use in wound assessment.

Semi-supervised active learning is a method that combines elements of supervised and unsupervised learning.

In order to be clearer in explaining this method, a model is initially trained with a small set of labeled (supervised) data. Subsequently, the model is used to predict labels on unlabeled (unsupervised) data. The model requests labels for the most informative or uncertain cases among the unlabeled ones, and these new labeled examples are added to the training set. This cycle continues, allowing the model to learn more efficiently and accurately with less manually labeled data.

The data used in the study was obtained from 474 patient records over two years, resulting in a total of 1564 wound images, which was called the "Deepskin" dataset. The images were captured using a smartphone digital camera, under uncontrolled lighting conditions and with various backgrounds and exposures, for clinical use. The patient population selected was heterogeneous, which means that the dataset includes samples of ulcers at different stages of healing and in different anatomical positions.

In the study, a variety of CNN architectures were explored, starting with lighter variants of U-Net and ending with more complex architectures such as PSPNet. The final choice was for a model similar to U-Net, using an EfficientNet-b3 model for the encoder, with corresponding adaptations for the decoder component.

The model achieved a Dice coefficient (DSC) of over 0.95 after less than 50 training epochs, demonstrating high efficiency and robustness.

2.3.12 DETR and YOLOv5: Exploring Performance and Self-Training for Diabetic Foot Ulcer Detection

R. Brüngel and C. Friedrich [17] investigated the effectiveness of the DETR and YOLOv5 object detection models in identifying diabetic foot ulcers.

The data used in the study came from the 2020 Diabetic Foot Ulcer Challenge (DFUC)

dataset. Initially, the dataset was cleaned to remove duplicate images, resulting in 1961 images with 2453 annotations. The DETR and YOLOv5 models were chosen in their largest available base versions and trained with pre-trained COCO weights. The standard data augmentation techniques incorporated in both models were used unchanged.

Data augmentation techniques were applied to both models, for DETR there is image scaling and random cuts during training. In the case of YOLOv5, there are changes in tonality, saturation and value, as well as rotation, translation, image scaling and other techniques.

Initial training was carried out with the pre-trained COCO weights, followed by a self-training stage, which involved extending the training data sets with predictions from the base models, creating pseudo-labels. This process aimed to improve the generalization of the models. The models were evaluated at different confidence levels, and YOLOv5 was also tested with test boosting (TTA).

The results showed that both DETR and YOLOv5 performed similarly in detecting diabetic foot ulcers. DETR achieved an F1-Score of 0.7355 and YOLOv5 an F1-Score of 0.7302, which increased to 0.7351 with the use of TTA. Self-training significantly improved the results, especially at lower confidence levels. YOLOv5 with TTA achieved the highest F1-Score of 0.7474 after self-training.

2.3.13 SegViT: Semantic Segmentation with Plain Vision Transformers

Bowen Zhang *et al.* in [18], unlike the articles presented so far, does not focus on wound segmentation, but rather on the development of an innovative method for semantic segmentation using Vision Transformers (ViT) without hierarchy, or without a neural network architecture such as CNNs. The proposed method, called SegViT, introduces the Attention-to-Mask (ATM) module to generate segmentation masks from attention maps, improving the efficiency of the process.

In addition, a "Shrunk" structure was developed that uses query-based samples for both down-sampling (QD) and up-sampling (QU). Essentially, it allows the size of the data processed by the Transformers to be manipulated, reducing the computational cost without significantly sacrificing the model's performance experiments have shown that SegViT outperformed similar methods on the ADE20K, COCO-Stuff-10K and PASCAL-Context datasets, reaching new levels of performance with lower computational cost.

2.3.14 Summary of the study of art

Table 2.2 shows the results of articles analyzed in the state of the art. sources used to collect the state of the art, as well as the number of results and the number of articles selected.

Tabela 2.2: Columns in Table 2.2 represent the reference, objective or problem that it aims to solves, dataset used, the best architecture, what kind of metrics used in segmentation and the best results achieved

Ref.	**Objective / Problem**	**Dataset**	**Best Architecture**	**Metrics**	**Best Results**
[7]	Wound image segmentation	SW-DFU/SW-SSD/ Medtec/SIH/FUSC	U-Net/DeepLab	MCC/IoU	0.85/0.75
[8]	Wound Segmentation	Own Dataset / Medetec	Mask-RCNN/ MobileNetV2 +CCL	Precision / Recall-Dice	98.40/94.27-94.05
[9]	Foot ulcer segmentation	Medetec/ Chronic Wound /FU-Seg	LinkNet-EffB1	Dice	92.09
[2]	Wound Tissue segmentation and classification	Own dataset	Own model "AutoTrace"and "Auto tissue"	MIoU	0.8644
[10]	Detection and classification	Medetec/ Handsurgery	YOLOv3	Precision / Recall / Dice	99%
[11]	Detection and segmentation	MICCAI 2021	Own model FNN	Precision / Recall / Dice	0.72 / 0.77 / 0.74
[12]	Segmentation	Massachusetts Medical Center	FCN / U-net / DeepLab V3	Dice	0.784 / 0.859 / 0.876
[13]	Classification, detection and segmentation	2893 images	DeepLab V3	Accuracy / IoU / Precision	99.25 / 98.87 / 98.88
[14]	Detection and segmentation	2863 images	DeepLab V3	Accuracy / Recall / Precision	0.99155 / 0.9915 / 0.9957
[15]	segmentation	Kaggle of chronic wound	Based on U-Net	Precision / Recall / IoU	0.941 / 0.776 / 0.844
[16]	Segmentation	Deepskin 1564 images	Own U-Net	Dice	0.96
[17]	Detection	DFUC	DETR / YOLOv5	Dice	0.7384 / 0.7474
[18]	Segmentation	ADE20K	SegVit	MIoU	0.50

These studies reflect the growing impact and potential of artificial intelligence and deep learning in medicine, especially in the area of wound analysis and treatment, offering faster, more accurate and efficient methods for diagnosing and monitoring the evolution of wounds.

The "Detect-and-segmentation"(DS) approach to automating the segmentation of wound images, developed by Scebba *et al.* [7], represents a significant advance in the field of medical image analysis. Using a set of three deep learning models, the DS method not only detects and centers the wound in the image, but also segments it accurately, achieving a Matthews Correlation Coefficient (MCC) of 0.85, demonstrating effectiveness even in images with complex backgrounds and diverse wound types.

Chuanbo Wang *et al.* [8] employ a CNN based on the MobileNetV2 architecture for ulcer segmentation, with promising results, especially in mobile applications, thanks to the lightness and efficiency of the chosen network.

Amirreza Mahbod *et al.* [9] propose a method based on two convolutional neural networks, LinkNet and U-Net, to segment foot ulcers, achieving 92.07% accuracy in the MICCAI 2021 FUSeg foot ulcer segmentation challenge.

Other studies include the development of automatic methods for segmenting wound tissue on mobile devices and the use of deep learning for automatic wound detection and classification. These advances demonstrate the growing usefulness of machine learning and image processing in the medical field, offering valuable tools for diagnosing and treating wounds.

A notable study [16] in the area of active semi-supervised learning for wound image segmentation is the work using CNNs, showing a reliable decision support system for clinical use in wound assessment.

The study "DETR and YOLOv5: Exploring Performance and Self-Training for Diabetic Foot Ulcer Detection" in [17] compares the DETR and YOLOv5 object detection models, showing that both are effective for detecting diabetic foot ulcers, each with its advantages in different usage scenarios. In this way, we get into the use of more recent technologies, the transformers, which is why the SegVit article [18] was incorporated into the state of the art, since we couldn't find any aimed at segmenting wounds.

The article "A Survey of Wound Image Analysis Using Deep Learning: Classification, Detection, and Segmentation" in [13] is a great contribution to those researching wound segmentation. This article, comparing it with the study carried out for the state of the art of this work and looking at table 2.2, we can find many points in common. We can see that both the models based on U-Net and Deeplab achieve a very high level of satisfaction, although it should be noted that the effectiveness of these models is intrinsically linked to the volume of datasets used for their training.

As a result of the latest survey of the state of the art, with the aim of finding articles

on transformers used in segmentation tasks, including a specific focus on wound segmentation, we can see from the previous analysis that comparative studies of transformer models are still limited. Transformers originally developed for natural language processing have been adapted to deal with the complexities of medical image segmentation[19], demonstrating positive performance when compared to traditional CNN, effectively capturing long-range dependencies in the data. This is particularly evident in the segmentation of 3D medical images and foot ulcers [20], where the integration of transformer-based models such as UNETR and Medical Transformer has shown promising results. These models take advantage of the self-attention mechanism to process images in a way that improves feature extraction, leading to more accurate segmentation results.

In addition, the articles highlight the versatility of transformers models in tackling various segmentation tasks, including semantic segmentation, instance segmentation and the specific challenge of wound localization and segmentation. The development of mobile applications [21] for wound localization using transform-based approaches highlights the practical applications and potential for real-world impact, especially in healthcare contexts. An emphasis on automatic wound segmentation using deep learning methods further illustrates the shift towards more sophisticated, AI-driven approaches to medical diagnosis and treatment planning.

When comparing pure transformer-based architectures with CNN-only architectures, we observed significant differences in terms of processing capacity and efficiency [22] [23]. Pure transformer architectures, like UNETR in [20] are excellent at capturing long-range dependencies and complex relationships in data, which is particularly useful for medical images that require detailed analysis. On the other hand, architectures that combine CNNs and transformers how is the case of TransUnet take advantage of CNNs for local feature extraction and transformers for global contextual analysis, offering a balance between local and global insight. CNNs, being more traditional, are more effective at identifying local patterns and textures, but can struggle with contextualization with long-range dependencies due to the design of the convolution.

3 Commercial solutions

After presenting the state of the art regarding the study of wound segmentation using deep learning algorithms, a survey was carried out of existing solutions already on the market. After some research, a set of applications that aim to solve the same problem were found.

3.1 Swift Skin and Wound

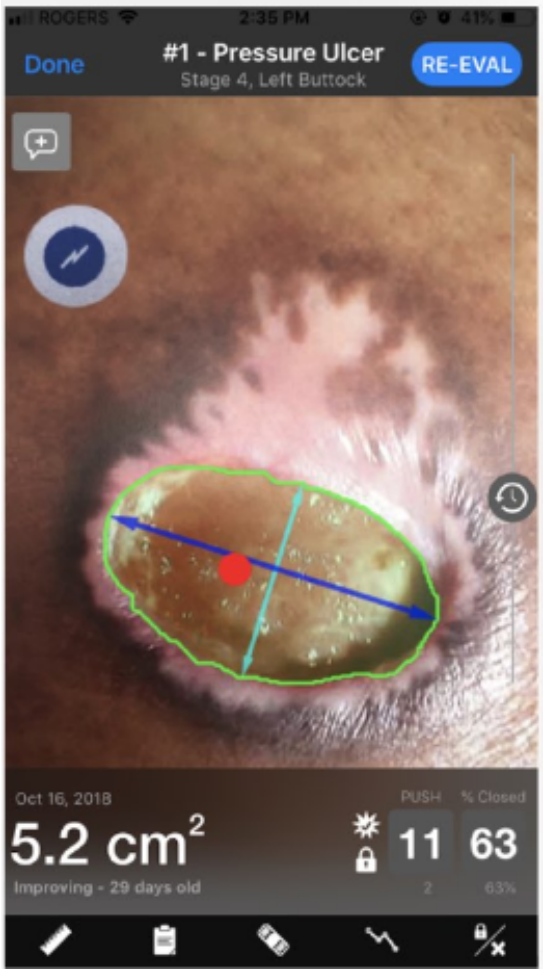

Figure 3.1: Swift Skin and Wound.[1]

Figure 3.1 represents the "Swift Skin and Wound"application that was developed by Swift Medical, a company based in Toronto, Canada. Swift Medical specializes in de-

[1]Source: *https://www.woundlocal.com/technology/swift*, accessed on 2024-03-28

veloping advanced digital technologies for wound care.

3.2 Imito Measure

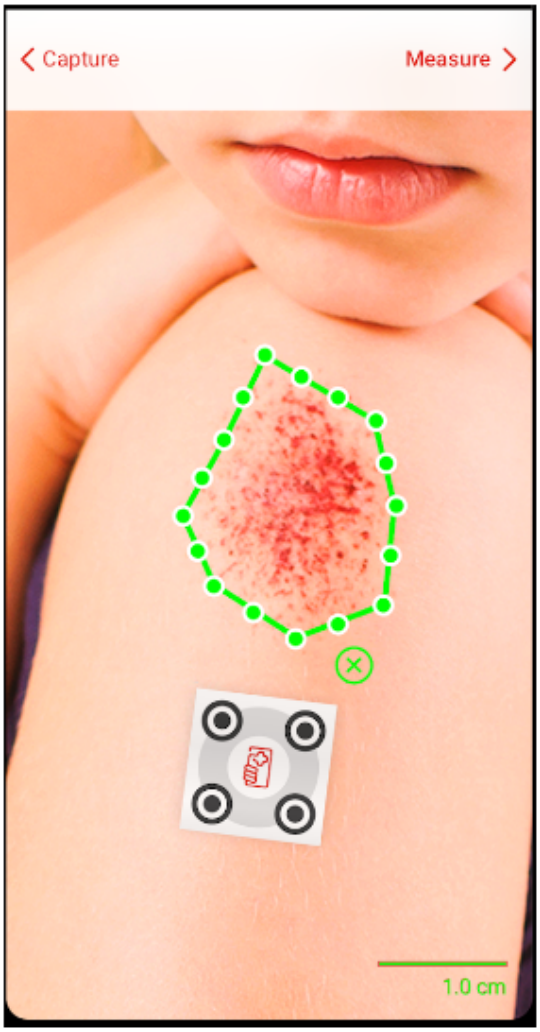

Figure 3.2: Imito Measure.[2]

The "Imito Measure" application shown in Figure 3.2 is another innovative healthcare application developed by Imito, a company based in Switzerland. This application is designed to assist healthcare professionals in the process of documenting and measuring wounds.

[2]Source: *https://imito.io/en/imitomeasure*, accessed on 2024-03-28

3.3 eKare inSight Healthcare

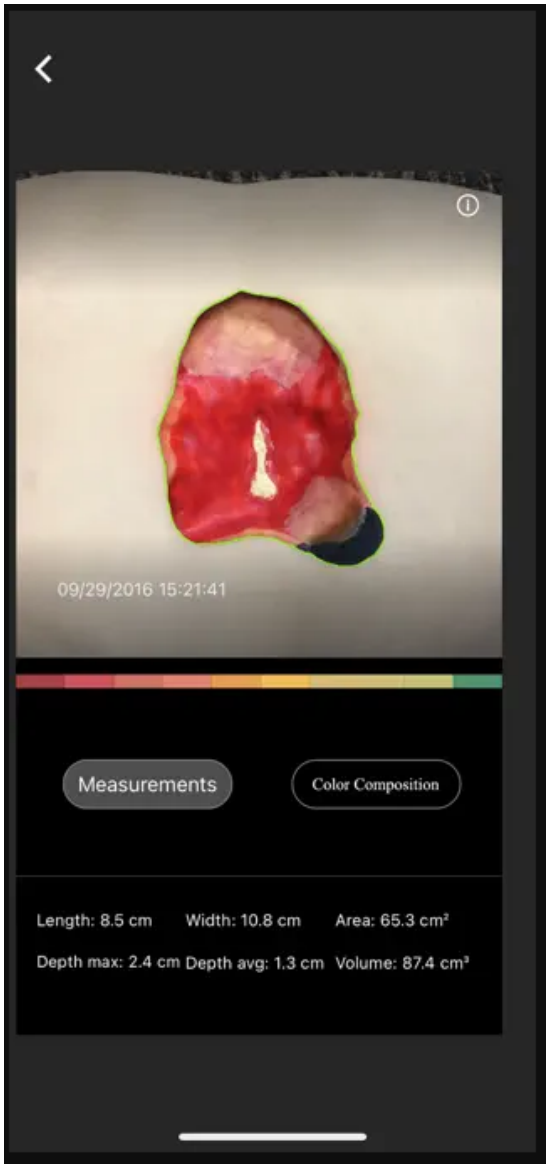

Figure 3.3: eKare inSight Healthcare.[3]

The application in 3.3 is an innovative solution in the field of wound care technology, developed by US-based eKare Inc., a company that focuses on digital technologies for healthcare. This application stands out for its use of advanced technology for the assessment and control of wounds.

[3]Source: *https://www.woundcare-today.com/journals/issue/wound-care-today/article/ekare-insight*, accessed on 2024-04-02

3.4 Wound Matrix

Figure 3.4: Wound Matrix.[4]

The WoundMatrix application shown in Figure 3.4 was developed by the company WoundMatrix, Inc., which specializes in developing technological solutions for wound care and management. The company is based in the United States and is known for being a pioneer in the creation of mobile telehealth and wound analysis technologies.

3.5 Wound Zoom

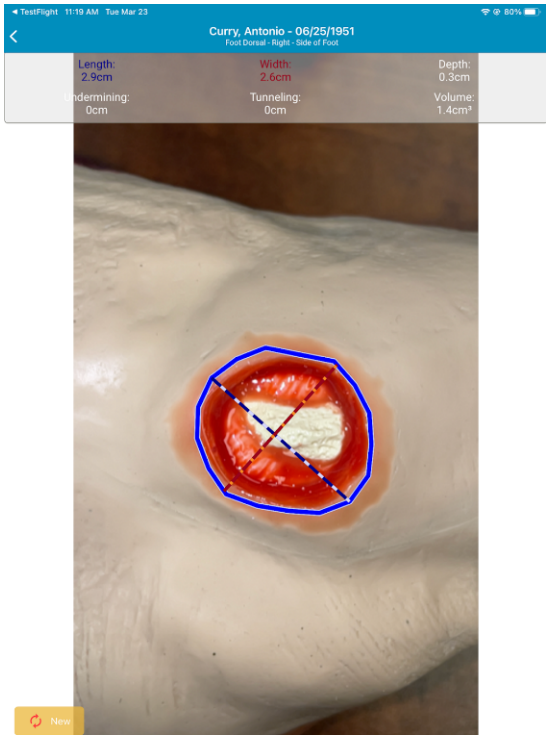

Figure 3.5: Wound Zoom.[5]

[4]Source: *https://www.woundmatrix.com/*, accessed on 2024-04-02

[5]Source: *https://play.google.com/store/apps/details?=com.perceptivesol.woundzoom.prohl=enpli=1*, accessed on 2024-04-14

The WoundZoom application in 3.5 is a technological solution for wound care and management. Developed by WoundZoom Inc., this tool offers advanced features to assist healthcare professionals in wound care. WoundZoom Inc. is based in the United States and is known for improving healthcare quality and operational efficiency.

3.6 Summary of comparative analysis of commercial applications

Of all the applications surveyed, the majority resort more specifically to the use of calibrated marks to calculate the area of the wound/pressure ulcer, as is the case with the applications Imito Measure, Wound Matrix and Wound Zoom. The eKare inSight Healthcare application makes use of extra hardware to obtain information from the images collected. On the other hand, the Swift Skin and Wound application uses neural networks to perform segmentation automatically as well as to obtain data on the dimensions of wounds/pressure ulcers, as can be seen in article [2].
In addition to the mobile solutions mentioned above, other desktop solutions were found, but they are rather rudimentary, the processing of the area calculation is done manually and, moreover, they are far from the intended focus, as they do not use deep learning technologies.
As for the "Swift Skin and Wound" app, the way it works is different, it makes use of some kind of mark for comparison, they use the placement of a mark next to the wound/pressure ulcer in order to be able to automatically calculate its dimensions. Once the mark has been placed, the outline of the wound is delineated and after this step the dimensional data about the wound is obtained.
All the information related to this section was only collected from the websites of each of the solutions found and other similar ones, given the scarcity of related information, as well as given the specificity of the applications found, since it is a niche. It was also not possible to try them out, as they are not available for free and most of them are not available on Marketplaces.

4 Methodology and Evaluation Framework

In this chapter, we discuss the selection of the most suitable model architectures for wound segmentation, as well as the metrics used to evaluate their performance. The chapter also presents the tools and technologies used throughout the study. The aim is to provide a comprehensive overview of the models, highlight the main features and present the main innovations in their architectures, along with the metrics and evaluation tools that supported the entire methodological process.

4.1 Model Selection

It is important to discuss the choice of model architectures. In this study, we considered YOLO v9 [24] and MedT [19] models, each of which represents a particular facet of the neural network spectrum. This choice is not arbitrary, but reflects a consideration of the specific needs of medical image segmentation and the potential of each architecture to meet those needs.

The YOLO v9 model, a continuation of the evolution of the YOLO architecture [25], is known for its speed and accuracy in real-time detection tasks. MedT is specialized in processing medical images as seen in the state of the art, but still without reference in the wound representation. This model uses attention mechanisms to focus on critical areas of the image and increase segmentation accuracy.

The choice of these models reflects the commitment to innovation and the exploration of technological frontiers in the field of computer vision applied to health. Each model brings unique and promising characteristics to the task of segmenting wound images, enhancing the search for effective and efficient solutions. The model was implemented using the TensorFlow framework with the Keras library, a high-level interface for building and training deep learning models [26].

The following is a detailed exploration of each architecture, based on their presentation articles.

4.1.1 YOLO

The central aim of the YOLOv9 [24], is to improve object detection performance by mitigating the challenges posed by information bottlenecks and the limitations of existing architectures in retaining critical data throughout the learning process.

YOLOv9's architecture is based on two fundamental innovations: Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN). PGI is designed to manage the changes required by deep networks to achieve various objectives effectively, while ensuring the integrity of the input data is preserved for reliable calculation of gradient information. GELAN, a new lightweight network architecture, is developed based on gradient path planning, proving to be efficient in models of different dimensions. These innovations aim to provide complete input information for the target task, increasing the model's ability to update the network weights accurately and maintain the features essential for executing the target task without additional inference costs.

The model's effectiveness was validated on the MS COCO dataset, emphasizing YOLOv9's ability to achieve superior object detection performance. The choice of dataset was guided by its relevance, annotation quality and suitability for the project, highlighting the model's robustness in handling diverse and complex object detection scenarios.

YOLOv9 demonstrated performance improvements in several metrics on the MS COCO dataset compared to its predecessors and contemporary models. It outperformed existing models in terms of parameter utilization and computational efficiency, while achieving higher accuracy rates.

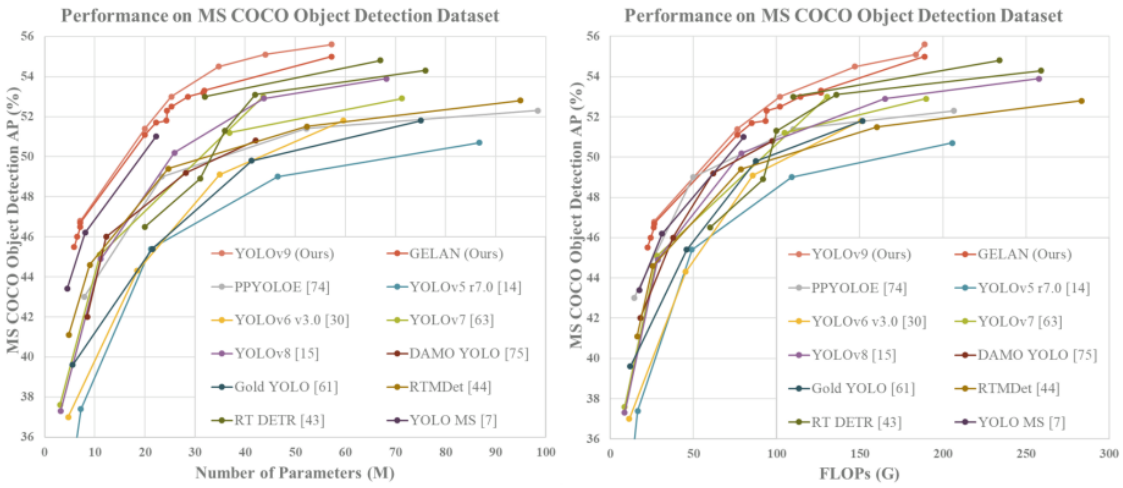

Figure 4.1: Performance of the different YOLO models and other architectures on the MS COCO dataset.[1]

Figure 4.1 compares different neural network architectures based on their performance on the MS COCO dataset for the object detection task.

[1]Source: *https://arxiv.org/pdf/2402.13616*, accessed on 2024-06-07

4.1.2 Medical Transformer

Jeya Maria Jose Valanarasu *et al.* in [19] propose a new approach to medical image segmentation using a Transformer-based architecture, called MedT. The main objective is to address the limitations of CNN in modeling long-range dependencies present in images, a critical aspect for accurate segmentation in medical applications.

MedT has innovated by introducing a gated axial-attention mechanism, which allows for more effective control of the influence of positional encodings during dependency modeling. In addition, the Local-Global (LoGo) training strategy is proposed to further improve performance, operating simultaneously on full images and patches, to learn global and local features respectively.

For the evaluation, MedT was trained and tested on three different medical image segmentation datasets: ultrasound images of the brain, microscopic images of glands and the MoNuSeg dataset for nuclei segmentation. MedT demonstrated superior performance, outperforming both CNN and other Transformer-based architectures in terms of F1 and IoU (Intersection over Union) metrics in medical segmentation tasks.

The study shows that the Medical Transformer, with its innovations of gated axial attention and the LoGo training strategy, presents a significant advance for medical image segmentation, offering a powerful tool for computer-aided diagnosis and image-guided surgery systems.

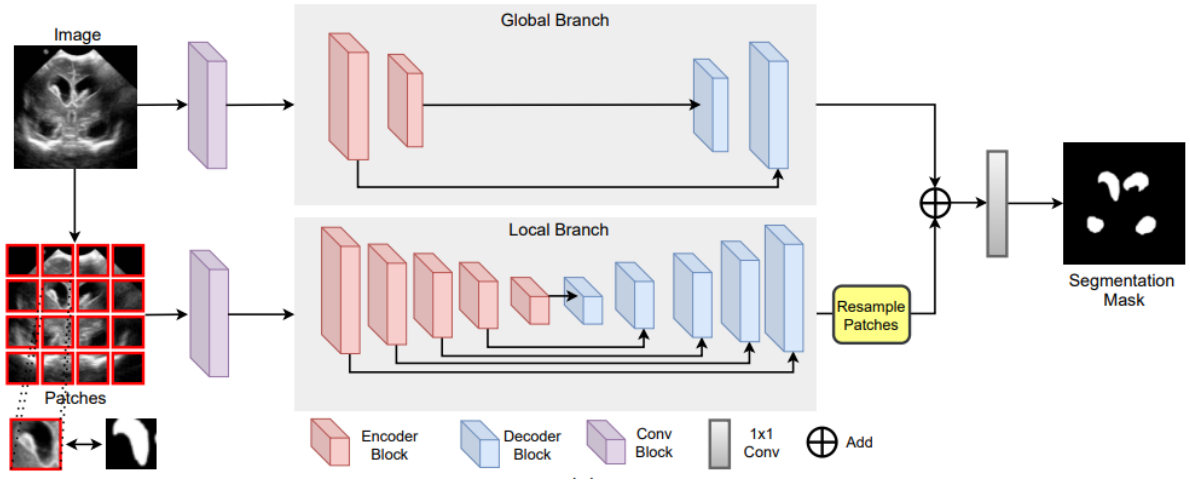

Figure 4.2: MedT architecture.[2]

In figure 4.2 we can see the diagram of MedT's main architecture.

[2]Source: *https://arxiv.org/pdf/2102.10662*, accessed on 2024-06-25

4.2 Evaluation Metrics

In multi-class classification and image segmentation problems, different metrics are used to evaluate the model's performance. In multi-class classification, each sample is assigned to a class based on the highest probability among all those chosen, and the metrics are applied individually to each class or averaged over all classes to obtain an overall score.

To understand the model's performance in detail, the following metrics are used, providing detailed insights into the performance of the model:

- **TP** (**True Positives**): Number of samples correctly diagnosed as being a specific lesion.
- **FP** (**False Positives**): Number of samples incorrectly diagnosed as being a specific lesion, when in fact they are not.
- **TN** (**True Negatives**): Number of samples correctly diagnosed as not being a specific lesion.
- **FN** (**False Negatives**): Number of samples incorrectly diagnosed as not being a specific lesion, when in fact they are.

To effectively organize and visualize these metrics, they can be presented in a confusion matrix, such as the Figure 4.3. This matrix serves as a basis for deriving important metrics and conclusions.

The confusion matrix is a table that shows:

- Rows, the actual categories of the samples.
- Columns, the categories predicted by the model.

The row-column intersection indicates the number of predictions for each combination of actual and predicted category.

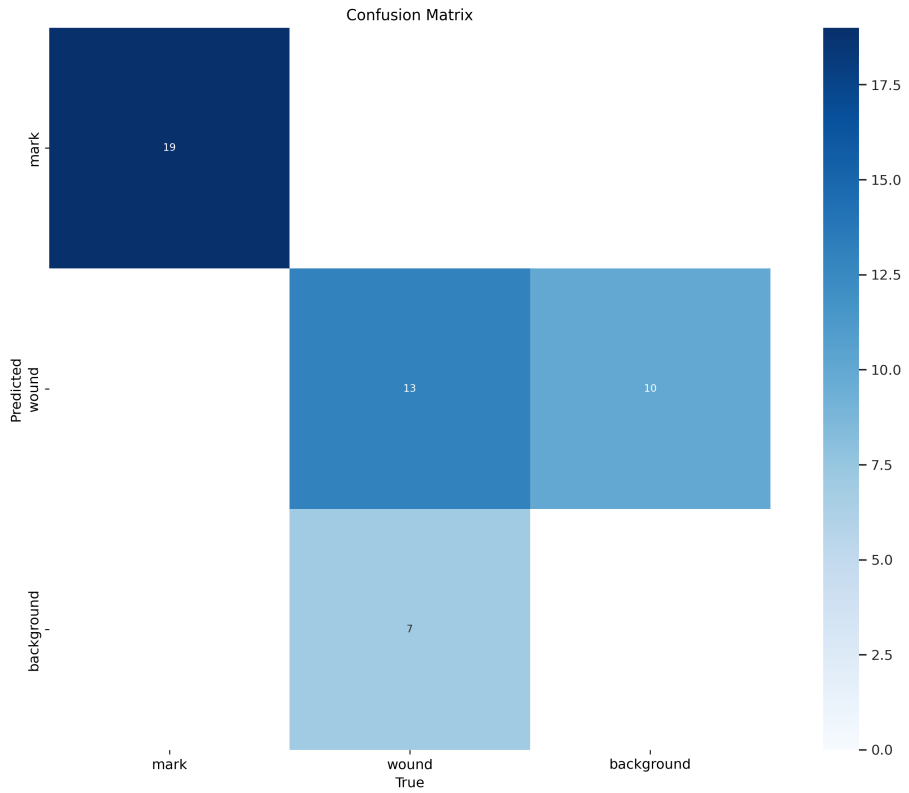

Figure 4.3: Example of the confusion matrix, the results shown are from the last training session carried out by the YOLO network in this work.

On the other hand, in image segmentation, metrics such as Precision, Recall and MIoU (*mean Intersection over Union*) are used.

- *Precision*: This metric evaluates the proportion of correct predictions among all the positive predictions made by the model. In other words, it is a measure of how accurate the positive results. Its calculation formula is shown in equation 4.1.

$$Precision = \frac{TP}{TP + FP} \tag{4.1}$$

- *Recall*: Also known as sensitivity, *Recall* measures the proportion of real positives that were correctly identified by the model. This indicates how effective in capturing relevant cases. We can see its calculation formula in equation 4.2.

$$Recall = \frac{TP}{TP + FN} \tag{4.2}$$

- MIoU (*Mean Intersection over Union*): This is a common metric in image segmentation tasks. *MIoU* calculates the average intersection over the union of the model predictions and the true labels. The higher the *MIoU*, the better the model's performance in accurately overlay of the area of interest on the images. The *MIoU* is

calculated using the formula shown in equation 4.3.

$$MIoU = \frac{TP}{TP + FP + FN} \tag{4.3}$$

In segmentation, the concepts of TP, FP, TN and FN are applied at pixel level: TP are pixels correctly identified as part of the class of interest, FP are pixels incorrectly marked as part of the class, TN are pixels correctly identified as not belonging to the class, and FN are pixels of the class of interest that the model has not detected. These are fundamental for calculating metrics such as Precision and Recall at pixel level, as well as for calculating the MIoU.

Therefore, while multi-class classification metrics provide a comprehensive view of the model's performance in correctly categorizing samples, segmentation metrics offer a detailed assessment of the accuracy and perfection of segmentation at the pixel level, ensuring that the model is accurate and reliable in its predictions.

4.3 Tools and Technologies

This chapter describes the tools and technologies used throughout this work to develop and implement the models, manage the data and prepare the wound segmentation experiments.

4.4 Machine Learning Frameworks and Libraries

4.4.1 PyTorch

PyTorch, an open source machine learning library developed by Facebook, was the main framework used to implement and train the models presented in the "Model Selection" chapter. This tool is widely used in computer vision and natural language processing applications and stands out for its ease of use compared to other libraries such as TensorFlow [26].

4.4.2 Support Libraries

Some additional libraries were used to support data processing and analysis:

- **Pandas**: For manipulating and analyzing tabular data.
- **NumPy**: For high-performance mathematical operations.
- **Scikit-Learn**: Used for data pre-processing and analysis.

4.5 Development Environment

4.5.1 Visual Studio Code (VS Code)

Code development was carried out in *Visual Studio Code* (VS Code), a powerful tool for editing and managing code, together with execution environments in the cloud.

4.5.2 Google Colaboratory

Google Collaboratory [27] (*Google Colab*) was used to train the YOLOv9 and MedT models. This is a Jupyter notebook environment that allows code to be executed in the cloud without the need for configuration, and also offers free access to GPUs, which makes the training process much faster than if it were done on personal computers. Figure 4.4 shows an example of this development environment.

Figure 4.4: Example of a Google Colab development environment.

4.5.3 GitHub

GitHub [28] was used as a code repository, especially important for the development of MedT, which required several adaptations to work on a multi-class problem. In this context, GitHub was essential for version management during development and debugging.

4.6 Tools for Data Annotation and Preparation

4.6.1 GIMP

GIMP [29], a free software program for creating and editing images, was used to put the calibration mark on the images in the training and validation datasets. This tool

was particularly useful in the data preparation process, as detailed in the chapter on the dataset.

4.6.2 Roboflow

Roboflow [30] is a platform for hosting and managing computer vision datasets. In this project, Roboflow played an essential role at various stages of the data pipeline:

- Data hosting: We used Roboflow to store the dataset created, consisting of two sets already on the platform and additional images sent by the health center.
- Image marking: Platform tools were used to draw the boundaries of the calibration marks on the training and validation images, as well as to segment the wounds in the test set. Figure 4.5 shows an example of using Roboflow for this task.
- Data augmentation: Roboflow was also used to apply data augmentation techniques to the dataset images in order to improve the robustness of the model. export: When it was necessary to export versions of the dataset to the Google Collaboratory project, we used Roboflow's export functionality, selecting the YOLOv8 PyTorch format.

Figure 4.6 shows an overview of the dataset divided into training images, validation and test images.

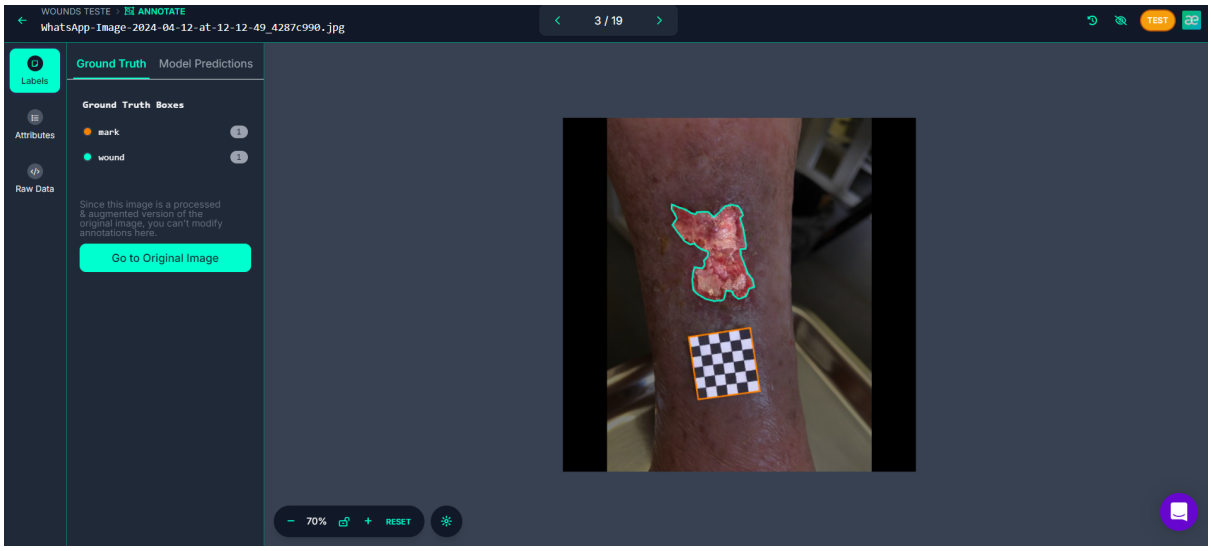

Figure 4.5: Using Roboflow to mark wounds sent by the health center.

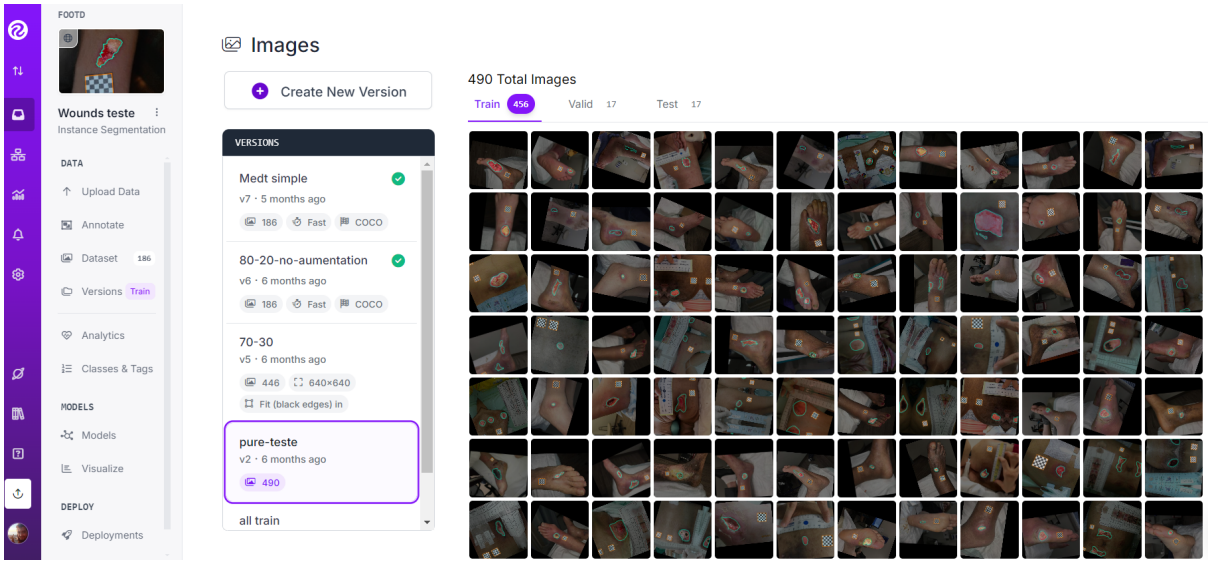

Figure 4.6: Overview of the *dataset* divided into training, validation and testing.

5 Dataset

This chapter describes the process of collecting and preparing the datasets used in this work, highlighting the collaboration with medical institutions and associations. The need for accurate and varied data for the development and testing of wound segmentation algorithms is fundamental, given the diversity of wounds that affect patients.

5.1 Methodology of Data Collection

This section describes the approach adopted for data collection in this study, highlighting the importance of ethics and transparency throughout the process.

Partnerships Established: In this study, strategic partnerships were established with reference entities in the health area and specialized professionals to ensure the relevance, rigor and applicability of the results. These contacts were essential to the success of the work, guaranteeing access to information, data and specialized technical support.

The Portuguese Wound Treatment Association (APTFeridas) was the first attempt at a partnership. Recognized as the main national reference association in the area of wound prevention and treatment, APTFeridas is widely known for its initiatives, such as workshops, promoting courses, organizing congresses and other training events. Further information about the association can be found on its official website: `www.aptferidas.com`.

The second entity contacted was the Unidade Local de Saúde do Baixo Mondego (ULSBM), an institution that includes the Figueira da Foz District Hospital, E. P. E., and the Health Centers. E., and the Figueira da Foz, Soure and Montemor-o-Velho Health Centers. The ULSBM is responsible for coordinating essential health services in the region, joining forces between different care units. More information about the ULSBM can be found on the official website: `www.hdfigueira.min-saude.pt`.

Through the ULSBM, we were able to establish a collaboration with the Figueira da Foz Health Center, which became a key partner in providing relevant clinical data and information for the study.

In addition to these institutional contacts, other health professionals were involved, such as doctors, nurses and health technicians

The process of formalizing these partnerships was conducted meticulously, ensuring

ethical and efficient cooperation between all parties involved. These collaborations represented an important step towards aligning the project with the practical needs of the health sector and ensuring the positive impact of its results.

Ethical approval process: To guarantee the ethical integrity of this project, it was necessary to submit the study to the rigorous analysis of the Ethics Committee of the ULSBM. The process began on 17-01-2024, with the first contacts made through emails and phone calls to establish communication with those responsible for the committee.

During this period, the documents required by the Ethics Committee were prepared and submitted. These included a detailed description of the study's scientific rationale, its objectives, the proposed methodology, the type of study carried out, the sites and institutions involved, as well as the participants and data collection instruments, where applicable. In addition, a statement of informed consent was developed for the users and statements of authorization for the use of data relevant to the study.

The process was conducted meticulously, ensuring compliance with applicable ethical and regulatory standards. After successive revisions and adjustments to the documents, final approval was granted on 13-05-2024.

Number of Participants and Inclusion/Exclusion Criteria: We specify the minimum number of participants to be included in the study, with strict inclusion/exclusion criteria. We have thoroughly assessed the risks and benefits for participants, ensuring ethical and responsible consideration.

Data Access and Informed Consent: We detail who will have access to the data collected and highlight whether or not informed consent must be obtained from all participants, respecting ethical and legal principles.

Ethical Approval and Personal Data Processing: In the ethical approval process, we carefully fill in the details relating to personal data processing operations, outlining responsibilities, data flow and associated risks. We identify the measures adopted to address possible risks, such as illegitimate access, unwanted modification and disappearance of data, always considering the potential impacts on data subjects.

This careful process of formalization and ethical approval ensures the scientific robustness and moral integrity of our research, contributing to the responsible and ethical conduct of the study.

5.2 Datasets Searched

In this section, we will explore the datasets considered during the research for this work on wound segmentation using deep learning techniques. The careful choice of dataset is fundamental to the effective training of computer vision models. We will discuss the main datasets evaluated, highlighting their characteristics and justifications for the

final selection.

AZH Dataset: The dataset used in this work is from the AZH dataset, which was compiled during a two-year clinical period at the AZH Wound and Vascular Center in Wisconsin. This dataset includes 730 images of wounds in .jpg format. The images vary in size, ranging from 320 to 700 pixels wide and 240 to 525 pixels high. These images represent at least four different types of wounds: pressure ulcers, diabetic wounds, venous ulcers and surgical ulcers. The images were taken with an iPad Pro (software version 13.4.1) and a Canon SX 620 HS digital camera. The labeling was carried out by wound specialists at the AZH wound and vascular center. In most cases, each image in the dataset is associated with a specific patient. However, there are cases where several photos of the same patient were taken on different parts of the body or at different stages of healing. In the latter case, the shapes of the wounds are different, so they are classified as separate images. This dataset is available on Github [28] through the following link https://github.com/uwm-bigdata/wound-segmentation

Kaggle Dataset [31]: This dataset includes 200 images of wounds in .jpg format, with dimensions of 512 by 512 pixels. The images specifically depict diabetic wounds and were selected due to the availability of labels and their visual quality. Although there is no detailed information on the devices used in the collection, the images show significant diversity in terms of background scenery, wound types and skin tones. The data can be accessed via the following link:
https://www.kaggle.com/datasets/mohamadtaher/wound-data

Roboflow Datasets [32][33]: The datasets made available comprise two distinct collections, one containing 817 images and the other 445 images, both in .jpg format and 640x640 pixel resolution. These collections are particularly relevant to our research, as they focus on diabetic wounds and pressure ulcers, areas of great interest in the medical field. A notable advantage of these datasets is the inclusion of polygonal annotations (labels) that allow for precise segmentation, as well as the high visual quality of the images. Although details of the capture devices are not available, the diversity in the images is evident in relation to the background contexts, the types of wounds represented and the variations in skin tone. It is important to note that, specifically for pressure ulcers, most of the images already include a scale factor, facilitating dimensional analysis. The datasets are accessible at the following addresses:
https://universe.roboflow.com/ta-bluys/segmentasiluka
https://universe.roboflow.com/hyunnununanna/augmented-dtpi-lastest

For a comprehensive evaluation, we will present a comparison Table 5.1 considering various aspects.

This comparative analysis aims to provide a clear view of the distinctive characteristics of each dataset, enriching the understanding of the final choice and highlighting the advantages of the dataset selected for the scope of this study. Detailed inspection of

Tabela 5.1: The columns represent each dataset searched, the number of images, the type of wounds covered, the source of the image collection, the image sizes, and the presence or absence of labels.

Ref.	**Nº Images**	**Type Wounds**	**Collection Source**	**Image Sizes**	**Label**
AZH	730	4	iPad Pro, Canon	Several	Yes
Kaggle	1010	1	N.A.	512x512	Yes
HealTech	1001	5	N.A.	331x331	No
Roboflow	817	1	N.A.	640x640	Yes
Roboflow	445	1	N.A.	640x640	Yes
Medetech	160	1	N.A.	224x224	Yes

the table initially led to a preference for Roboflow datasets.

5.3 Data Collection and Preparation Process

In the context of this study, the data collection and preparation stage plays a crucial role in obtaining reliable and representative information. Based on specific criteria above, the selection process aims to guarantee the diversity and quality of the images used. This section presents the criteria chosen for data selection.

Selection criteria: The data collection and preparation process is based on specific criteria to guarantee the quality and diversity of the images. Selections are based on images that have a height or width of more than 500 pixels, are accompanied by a polygon-shaped wound delimiting mask (label), where its position is at a perpendicular angle to the wound and show variety in terms of wound cases, skin tones, backdrops used and lesion sizes. This rigorous set of criteria ensures that the data set is representative and comprehensive, enabling a comprehensive and accurate analysis.

Data Preparation and Annotation: In addition to the criteria already mentioned for image selection, it is crucial to include the area not affected by the wound. This will allow the integration of a calibration mark, which is essential for accuracy in calculating the wound area and, consequently, for categorization using the PUSH scale. We have implemented a specific method involving the inclusion of this calibration mark, as detailed below.

This implementation consists of placing a physical mark adjacent to the wound or pressure ulcer, called a calibration mark. This mark is provided to the partner and in the case of the public dataset it was placed next to the wound by editing the image. By knowing the physical dimensions of the mark, we are able to accurately calculate the dimensions of the wound by correlating the actual dimensions of the calibration mark with the dimensions (in pixels) present in the captured image.

The calibration mark adopted is a checkerboard, as illustrated below (Figure 5.1):

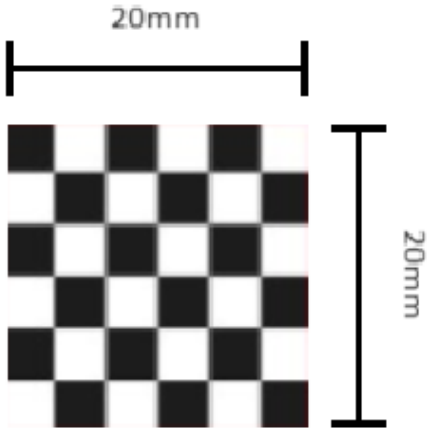

Figure 5.1: Calibration mark.

It has four rectangles positioned in the corners, twenty millimeters apart (end to end). The algorithm will calculate the dimensions of the wound using these rectangles, determining the number of pixels from one end of the squares to the other in two of the corners and then applying the proportion to the outlined wound. As for positioning during image capture, the calibration mark can be in any position within the photograph's field of view, ideally as perpendicular as possible to the angle of the photo. The algorithm must calculate the number of pixels between the rectangles mentioned, corresponding to the twenty millimeters of the physical mark.

In the partnerships, we provided a set of these targets printed on paper. The ideal dimensions of the chessboard were assessed to ensure its visibility in the photographs. These targets were printed on 250 gram sheets, resulting in 75 calibration marks per A4 page.

This approach offers a high degree of precision in obtaining measurements, although it requires placing the calibration mark close to the wound as a disadvantage, implying an additional (albeit minimal) effort compared to other solutions mentioned in the introduction.

Visual Presentation: Real Calibration vs. Artificial Calibration

To illustrate the process of implementing the calibration mark in the datasets, we visually present three images with a real calibration mark and three images in which the mark has been added artificially.

5.4 Final Dataset

In this section, we describe the final dataset that was developed from the research in partnership with health institutions. This dataset was created and prepared to guaran-

Figure 5.2: Real photos compared to manually altered photos. The first row of 3 photos are the real photos, while the bottom row are manually altered.

tee the quality of the results of the models used in wound segmentation. In addition, we discuss the data augmentation techniques applied to increase the variability of the data, promoting a better ability to generalize the models.

5.4.1 Dataset Structure and Composition

The final dataset was created from the images provided by the Unidade Local de Saúde do Baixo Mondego and the selection of Roboflow datasets. The images sent by the health center were collected as patients showed up at the institution, over a period of eight months, between January and August. This gradual collection ensured a diversity of cases and conditions, allowing the dataset to represent different types of wounds and pressure ulcers, which is essential for a model intended to be used in patient follow-up. The process of creating the dataset involved using the Roboflow tool to accurately segment the "Wound" and "Mark" regions, resulting in a rich and detailed dataset for the segmentation task. It was necessary to use Python code to transform the coordinates of the segmentations into images, as the models do not interpret the labels in the same way.

As a result, it was possible to create a dataset of 186 images in total, which were divided into training, validation and test sets. This division of the data is essential to guarantee a robust assessment of the models' performance, enabling independent validation and testing, in order to understand whether the model has sufficient generalization capacity to respond to the real problem of segmenting wounds in healthcare institutions. Correct wound segmentation is essential for monitoring and controlling the progress of each patient throughout treatment.

5.4.2 Data augmentation techniques

In order to enrich the training set and improve the robustness of the models, data augmentation techniques were applied which generated a final set of 490 images. These techniques included:

- **Horizontal Flip**: Horizontal mirroring of the images.
- **90° Rotation**: Clockwise and counter-clockwise rotation.
- **Random Rotation**: Between -26° and +26°.

The use of these techniques made it possible to increase the variability of the training images, ensuring that the model was exposed to different wound variations. This approach is crucial for increasing model results in order to generalize well to new examples, especially in real contexts such as health centers, where image capture conditions vary significantly.

5.4.3 Final Dataset Tables

The table below summarizes the number of images in the original dataset and after the data augmentation process.

Tabela 5.2: Composition of the Final Dataset - Original and with Data Augmentation

Dataset	**Original Dataset**	**Dataset with Data Augmentation**
Train	152 images	456 images
Validation	17 images	17 images
Test	17 images	17 images
Total	**186 images**	**490 images**

The creation of this dataset was essential to ensure that the model was developed with a solid base of real and varied data. The use of Roboflow to segment the wounds and calibration marks ensured consistent, high-quality annotation, which was essential for training the models.

With a total of 456 images being trained on the data augmented set, the aim is to assess the models' ability to generalize to new data coming from the partnership, replicating

real situations found in a healthcare center environment. These images cover a variety of wound types and stages of healing, which is essential for developing a model that can be used to help control patient treatment.

6 Results, Analysis and Discussion

This chapter presents the results obtained by the models applied to wound segmentation. Firstly, the individual results of each approach are presented, highlighting the performance of the YOLO and MedT architectures. This structure aims to highlight the adaptability of each model and its effectiveness in dealing with different scenarios.

The main focus is on identifying the most suitable sets of hyperparameters, taking into account the existing constraints and the impacts of the choices made during the adjustment process. The experiments for the image-based approaches, as well as the methodological detail of the entire process, are presented in the subsequent sections.

Finally, the chapter concludes with a comparative analysis of the two models, discussing their respective strategies and highlighting the main differences and similarities in terms of performance.

6.1 Image Segmentation Model Performance

6.1.1 Yolo

This section presents the results of the hyperparameter optimization for the YOLO model, version 9 c, considering images with a resolution of 640x640 pixels. In order to identify the best hyperparameter, the results obtained were analyzed for combinations of varying batch sizes (4, 8 and 16), and different learning rates (0.01, 0.001 and 0.0001).

The analysis is based on the average training and validation values of the main performance metrics, namely precision, recall and MIoU, in wound segmentation tasks.

These metrics were calculated using a dedicated function, which processes the predictions and the actual labels, computing separate metrics for each class in the model. The function implements the calculation of MIoU, precision and recall per class, using intersection and union operations between the predictions and labels to calculate IoU, and evaluating true positives, false positives and false negatives for the precision and recall metrics. This method guarantees a detailed and standardized evaluation of the model's performance.

After 10 epochs, the hyperparameter configurations with the greatest impact were analyzed. For each class in the model, the three best-performing configurations were selected for a more detailed analysis. This approach allows for a more targeted focus,

considering the factors that directly influence the performance of the segmentation tasks. By detailing the performance metrics of each class, we gain valuable insights into the efficiency of the model in different categories.

The results obtained for the configuration presented are described in Table 6.1, which shows the average training and test values, as well as the performance metrics for Precision, Recall and MIoU in the “Mark” and “Wound” classes.

The configuration with 10 epochs and a batch size of 16 performed the best for the "Mark" class in the test dataset, achieving a Precision of 0.928, Recall of 0.737, and MIoU of 0.775. These values demonstrate a strong capability of the model to detect and segment the "Mark" class accurately with this setup. However, for the "Wound" class, the performance was significantly lower, with a Precision of 0.113, Recall of 0.700, and MIoU of only 0.138, highlighting that this class remains challenging to segment effectively in this configuration.

For the batch size of 8, the results indicate a balanced but moderate performance across both classes. The "Mark" class achieved a Precision of 0.692, Recall of 0.895, and MIoU of 0.663 in the test dataset. The "Wound" class showed marginal improvement, with a Precision of 0.269 and MIoU of 0.074, though its Recall of 0.200 suggests that this configuration still struggles with consistent segmentation of this class.

The configuration with batch size 4 had mixed results. The "Mark" class achieved a Precision of 0.852, Recall of 0.305, and MIoU of 0.308 on the test set. However, for the "Wound"class, performance was notably poor, with a Precision of 1.000 but a Recall and MIoU of 0.000 and 0.009, respectively, indicating severe overfitting or class imbalance issues in this setup.

Based on these results, we set the batch size to 16. As no different results were detected when changing the learning rate, we continued with the next training sessions at a learning rate of 000.1, which is the default in the documentation.

Table 6.1: Analysis of the image-based approach using YOLO: results of the performance metrics for the training and test set, in the different classes with the batch value increased.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Test	Train	Test	Train	Test
epochs = 10 ; batch = 4	Mark	1.000	0.852	0.985	0.305	0.843	0.308
lr = 0.0001	Wound	1.000	1.000	0.000	0.000	0.195	0.009
epochs = 10 ; batch = 8	Mark	0.957	0.692	1.000	0.895	0.876	0.663
lr = 0.0001	Wound	0.802	0.269	0.571	0.200	0.340	0.074
epochs = 10 ; batch = 16	Mark	0.997	0.928	1.000	0.737	0.883	0.775
lr = 0.0001	Wound	0.330	0.113	0.810	0.700	0.391	0.138

After 30 and 100 training epochs, the hyperparameter configurations with the greatest impact were analyzed in detail. For each task in the model, the best performing configurations were selected for a more detailed analysis. This approach aims to explore the impact of a greater number of epochs on the performance of the "Mark" and "Wound" classes, taking into account metrics such as Precision, Recall and MIoU, with the results shown in Table 6.2. These results provide important insights into the efficiency of the model in the different classes and show how increasing the number of epochs influenced overall performance.

The results presented show that, after increasing the number of epochs to 30, the configuration with batch size = 16 and learning rate = 0.0001 obtained outstanding results in the "Mark" class, achieving a Precision of 0.998 in the test set, Recall of 0.947 and MIoU of 0.931. However, for the "Wound" class, despite the relative improvement compared to the previous analysis (table with 10 epochs), performance remained sub-optimal, with a Precision of 0.916 and MIoU of 0.382 in the test set. These values indicate that increasing the number of epochs has provided clear benefits for the "Mark" class, but the "Wound" class still represents a challenge, with a Recall of only 0.600 in the test.

By increasing the number of epochs to 100, the configuration with batch size = 16 and learning rate = 0.0001 continued to show significant results in the "Mark" class. Precision reached 0.989 and Recall was 1.000, with an MIoU of 0.958 on the test set, representing a clear improvement on training with 30 epochs. For the "Wound" class, the increase in epochs resulted in some improvements in Recall and MIoU, but still to a limited extent. While the Recall of the "Wound" class remained at 0.600 in the test, the MIoU showed a slight increase to 0.383. The Precision of the "Wound" class in the test set, however, fell to 0.782, suggesting that a greater number of epochs may be leading the model to overfit the "Wound" class, affecting its generalization capacity.

Therefore, it can be concluded that increasing the number of epochs brought significant benefits mainly for the "Mark" class, while the "Wound" class still shows difficulties when it comes to segmentation. Although the metrics showed improvements with 30 and 100 epochs, the "Wound" class did not obtain as marked an improvement as the "Mark" class. Thus, additional strategies, such as increasing training data or using regularization techniques, may be necessary to ensure consistent performance in both classes.

Table 6.2: Analysis of the image-based approach using YOLO: results of the performance metrics for the training and test set, in the different classes with an increase in the number of epochs.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Test	Train	Test	Train	Test
epochs = 30 ; batch = 16	Mark	0.937	0.998	1.000	0.947	0.922	0.931
lr = 0.0001	Wound	0.868	0.916	0.940	0.600	0.661	0.382
epochs = 100 ; batch = 16	Mark	0.936	0.989	1.000	1.000	0.913	0.958
lr = 0.0001	Wound	1.000	0.782	0.894	0.600	0.736	0.383

6.1.2 MedT

This section presents the results of hyperparameter optimization for the MedT network architecture. For this model, the batch size was set at 4 due to GPU memory limitations, and the images were resized to 256x256 pixels to ensure that training could be carried out efficiently within the hardware limitations. Initially, different learning rates were considered (0.01, 0.001 and 0.0001), and the main performance metrics - precision, recall and MIoU - were evaluated on the training and validation sets, applied to the wound segmentation task.

The results presented below include the values obtained for the training and validation sets during the first stages of training, with a view to optimizing the hyperparameters. In a final experiment, more training epochs were carried out by finetuning the model, and the performance values on the test set are also discussed, allowing a complete analysis of the model's generalization capacity.

The metrics were calculated using the CrossEntropy loss function, ensuring a consistent assessment of the model's performance. This approach made it possible to assess how the hyperparameters and training settings impacted the segmentation of the different classes, especially in terms of precision, recall and MIoU, providing a comprehensive view of the model's behaviour throughout the optimization and finetuning process.

The best hyperparameter tuning results for this segmentation network are shown in Table 6.3.

It can be seen that the configuration with learning rate = 0.001, batch size = 4 and epochs = 30 stands out as the best among the configurations tested, with a precision of 0.941 and a recall of 0.873 for the "Mark" class in the validation set, and an MIoU of 0.821. For the "Wound" class, the precision was 0.599 and the recall 0.545, with an MIoU of 0.388. These values indicate a consistent performance in segmenting the "Mark" class, showing that the model was able to generalize well to the validation set, while the "Wound" class shows slightly lower results, suggesting greater difficulty in segmenting this class.

However, with learning rate = 0.01, there was a drop in performance, especially for the "Wound" class. The precision was 0.387 in the validation set and the recall was 0.525, while the MIoU only reached 0.280. For the "Mark" class, precision was 0.914 and recall 0.738, with an MIoU of 0.685. These results suggest that a higher learning rate was not ideal, possibly leading the model to unstable convergence.

On the other hand, when the learning rate was set to 0.0001, the precision in the validation set was 0.834 for the "Mark" class and 0.556 for the "Wound" class. Recall reached 0.857 for "Mark" and 0.521 for "Wound". The MIoU was 0.790 for the "Mark" class and 0.293 for the "Wound" class. Compared to the 0.001 learning rate, the values are lower, showing that finetuning the learning rate to 0.0001 did not bring substantial improvements, especially for the "Wound" class, which maintained weaker performance.

It is worth noting that, in this case, there was no MIoU data available until the tenth epoch, so the number of epochs started at 30 for all configurations. This choice aimed to ensure a more consistent assessment of the model's performance after an adequate training period.

We can conclude that the best combination of hyperparameters for segmentation was learning rate = 0.001 and batch size = 4, which showed the best balance between accuracy, recall and MIoU, especially for the "Mark" class.

Table 6.3: Analysis of the image-based approach using MedT: results of the performance metrics for the training and validation set, in the different classes with an increased LR value.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Valid	Train	Valid	Train	Valid
epochs = 30 ; batch = 4	Mark	0.921	0.914	0.851	0.738	0.797	0.685
lr = 0.01	Wound	0.623	0.387	0.506	0.525	0.377	0.280
epochs = 30 ; batch = 4	Mark	0.914	0.941	0.909	0.873	0.845	0.821
lr = 0.001	Wound	0.758	0.599	0.654	0.545	0.525	0.388
epochs = 30 ; batch = 4	Mark	0.899	0.834	0.805	0.857	0.811	0.790
lr = 0.0001	Wound	0.681	0.556	0.584	0.521	0.498	0.293

Based on the results obtained previously, it was found that the "Wound" class had significantly lower performance, especially in the accuracy and MIoU metrics, suggesting that the model had difficulties in correctly identifying and segmenting this class. To mitigate this problem, a new series of experiments was carried out with adjustments to the weight balancing of the classes, with the aim of improving the performance of the "Wound" class. Adjusting the balance of the weights implies giving greater importance to the error made by the model when segmenting the "Wound" class. This means that when calculating the loss during training, errors related to the "Wound" class have a greater impact. Using the best hyperparameters found previously (learning rate = 0.001, batch size = 4, and epochs = 30), different weights were tested for the "Wound", "Mark" and "Background" classes, as shown in Table 6.4.

Initially, the "Wound" class was given a high weight in relation to the other classes (ranging from 5 to 20), while the weights of the "Mark" and "Background" classes were kept constant. The results show that although there was a clear attempt to favor the "Wound" class with higher weights, the impact was limited. In the best case scenario, the "Wound" class achieved an accuracy of 0.504 in train and 0.242 in validation with a weight of Wound = 5, but as the weight increased, the accuracy decreased, reaching only 0.354 in validation for a weight of Wound = 20. Even when increasing the weights of the "Wound" class, the MIoU remained low, with the best value achieved being 0.387 in training and 0.226 in validation for the Wound = 15 configuration, indicating that the change in weights did not provide substantial improvements in performance.

For the "Mark" class, performance remained relatively stable throughout the different weight settings, with minimal variations in accuracy, recall and MIoU values. Notably, the best precision and MIoU results for the "Mark" class were observed when the weight of the "Wound" class was 11, obtaining a precision of 0.886 and a MIoU of 0.830

in the validation set. This suggests that although the balancing attempts slightly favored the "Mark" class, the impact on the "Wound" class was not significant enough to justify the change in weights.

Based on these results, it can be concluded that adjusting the class weights did not have the desired effect on improving the segmentation of the "Wound" class. As such, it was decided to maintain the previous results, as the adjustments to the class balance did not bring significant gains in terms of precision, recall or MIoU for the most problematic class.

Table 6.4: Analysis of the image-based approach using MedT: results of the performance metrics for the training and validation set, trying to find the ideal weight for the "Wound" class.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Valid	Train	Valid	Train	Valid
Background = 1 Wound = 5 Mark = 1	Mark	0.879	0.873	0.880	0.892	0.782	0.797
	Wound	0.504	0.242	0.866	0.781	0.441	0.225
Background = 1 Wound = 7 Mark = 1	Mark	0.417	0.891	0.816	0.666	0.745	0.634
	Wound	0.887	0.172	0.821	0.766	0.364	0.166
Background = 1 Wound = 11 Mark = 4	Mark	0.818	0.886	0.947	0.935	0.789	0.830
	Wound	0.409	0.224	0.935	0.826	0.387	0.209
Background = 1 Wound = 15 Mark = 4	Mark	0.789	0.827	0.946	0.864	0.759	0.778
	Wound	0.382	0.243	0.944	0.905	0.369	0.226
Background = 1 Wound = 20 Mark = 4	Mark	0.770	0.903	0.945	0.845	0.737	0.758
	Wound	0.354	0.175	0.955	0.948	0.342	0.171

The results presented in Table 6.5 show the evolution of the MedT model's performance in segmentation tasks, by increasing the number of epochs from 30 to 100, keeping the hyperparameters constant, including batch size = 4 and learning rate = 0.001. This experiment, carried out without data augmentation and after finetuning, aimed to assess whether a greater number of epochs could improve performance in the "Mark" and "Wound" classes.

With 30 training epochs, the configuration resulted in moderate performance for the "Mark" class, achieving a precision of 0.914, Recall of 0.738 and MIoU of 0.685 in the validation set. For the "Wound" class, performance was more limited, with a precision of 0.387, Recall of 0.525 and MIoU of only 0.280 in the validation set. These results indicate that although the model managed to segment the "Mark" class reasonably well, there was significant difficulty in accurately identifying the "Wound" class.

By increasing the number of epochs to 100, the model showed a notable improvement, especially for the "Mark" class. Precision rose to 0.950, Recall was 0.961 and MIoU reached 0.914 in the validation set. These values show that a longer training time allowed the model to better capture the characteristics of the "Mark" class, resulting in a more accurate and consistent segmentation. For the "Wound" class, there were also important improvements, with precision increasing to 0.637, Recall to 0.632, and MIoU to 0.465 in the validation set. Although the improvement in the "Wound" class was not as significant as in the "Mark" class, it still represents progress compared to training with 30 epochs.

It can therefore be concluded that increasing the number of epochs had a positive impact on the performance of both classes, although progress in the "Wound" class is still limited. The results suggest that in order to improve the segmentation of the "Wound" class, it may be necessary to incorporate other strategies in addition to increasing the number of epochs, such as data augmentation techniques or the application of more robust regularization methods. In any case, the finetuning combined with the increase in epochs provided a considerable improvement, especially for the "Mark" class, showing the importance of longer training in complex segmentation tasks.

Table 6.5: Analysis of the image-based approach using MedT: results of the performance metrics for the training and validation set, separated by class and increasing the number of epochs.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Valid	Train	Valid	Train	Valid
epochs = 30 ; batch = 16	Mark	0.921	0.914	0.851	0.738	0.797	0.685
lr = 0.001	Wound	0.653	0.387	0.506	0.525	0.377	0.280
epochs = 100 ; batch = 16	Mark	0.962	0.950	0.958	0.961	0.923	0.914
lr = 0.001	Wound	0.877	0.637	0.846	0.632	0.751	0.465

The results presented in Table 6.6 show the performance of the MedT model on the test set after training for 100 epochs, using a batch size = 4 and learning rate = 0.001. These results reveal insufficient performance, highlighting the model's inability to generalize adequately to the test data. In particular, the precision, Recall and MIoU values show a clear discrepancy with the validation results, highlighting a lack of robustness in the model.

For the "Mark" class, the precision was 0.652, the Recall was 0.586, and the MIoU was only 0.483. These values indicate that the model had difficulty correctly identifying this class in the test set, resulting in a performance that is far below what was expected. For the "Wound" class, precision was extremely low (0.082), while recall was 0.670, suggesting that the model managed to identify many examples of this class, but with

an excess of false positives, reflected by the low precision and MIoU of only 0.078. This result is indicative of inconsistent behaviour and a lack of effective generalization, especially for the "Wound" class.

The absence of data augmentation may have contributed to a model that was unable to cope well with the variations present in the test set, resulting in poor performance. The next step will involve applying data augmentation techniques to enrich the training data and improve the model's ability to generalize, in the hope of obtaining more robust results for both sets of data.

Table 6.6: Analysis of the image-based approach using MedT: results of the performance metrics for the test set, separated by class in the largest number of epochs trained.

Hyperparameters	Class	Precision	Recall	MIoU
		Test	Test	Test
epochs = 100 ; batch = 16	Mark	0.652	0.586	0.483
lr = 0.001	Wound	0.082	0.670	0.078

6.1.3 Enhancing Segmentation Performance

This section presents the results of the most recent experiments conducted to improve the performance of the YOLO and MedT architectures in wound segmentation. The results were obtained by applying data augmentation techniques to the training data, including horizontal mirroring, 90° rotation (clockwise and counterclockwise), as well as random rotations between -26° and +26°.

YOLO

Before presenting the results of the last training session, we show the graph of segmentation loss during training (train/seg_loss) on the left-hand side (a) of the Figure 6.1.

Before presenting the results of the last training session, the Figure 6.1 shows the graph of loss per epoch during training and validation, representing "Train Loss" and "Validation Loss" respectively.

We can see a clear downward trend in the training loss (blue line), which is a desirable characteristic. Initially, the loss remains close to 1.0 and begins to decrease gradually towards the end of the 100 epochs. This behavior indicates that the model is adjusting well to the training data, progressively learning the patterns needed for accurate segmentation.

The validation loss (orange line) shows a high peak in the very first epoch, reaching a value close to 6.0, which indicates that the model initially had great difficulty in generalizing to the validation data. This behavior is common when the model is still "learning" the fundamental characteristics of the data and possibly finding variations in the examples that were not observed during training. However, after the first few epochs, the validation loss drops rapidly and, from the tenth epoch onwards, stabilizes at a level close to the training loss, suggesting that the model is beginning to adapt consistently to the validation data.

From the stability observed in both losses over the last 70-80 epochs, we can conclude that the model has reached a state of convergence, where the additional adjustment provided by training no longer generates substantial reductions in loss. The close behaviour between training and validation loss indicates that the model is not overfitting, suggesting a good ability to generalize to unseen data.

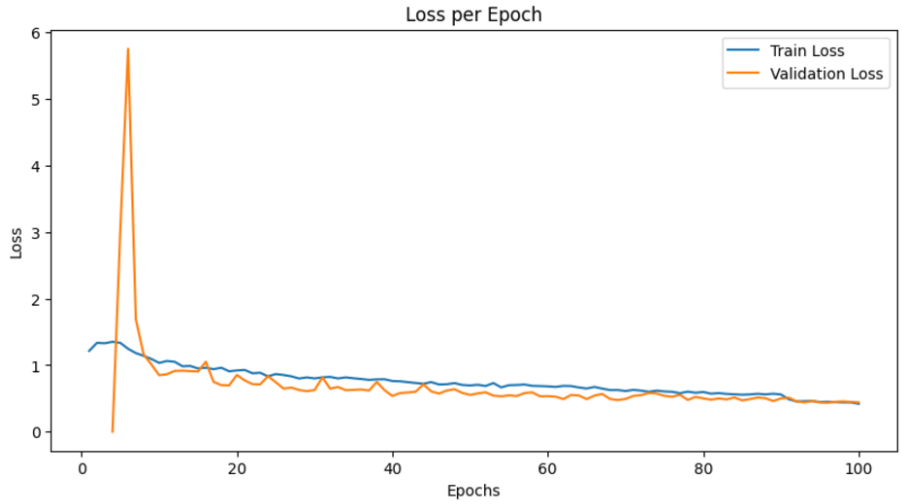

Figure 6.1: Evolution of Segmentation Loss in Training and Validation over 100 epochs.

In Table 6.7 we show the results obtained with the YOLO architecture, considering the different data sets (training and test) and the accuracy, recall and MIoU metrics for the "Mark" and "Wound" classes. Initially, with 10 epochs, batch size = 16, and learning rate = 0.0001, the results for both classes were extremely low, showing that the model was underfit, with a precision of only 0.113 for the "Mark" class and 0.008 for the "Wound" class in the test set. The MIoU was also very low, reaching only 0.030 and 0.007 respectively.

By increasing the number of epochs to 30, there was a significant improvement in the model's performance, especially for the "Mark" class, which obtained an accuracy of 0.979 and MIoU of 0.958 in the test set, with a recall of 1.000, indicating that the model was able to accurately and consistently identify this class. For the "Wound" class, the precision reached 0.760 and the MIoU was 0.400, suggesting an important evolution, although the performance of this class is still less than ideal.

With 100 epochs, the model continued to show positive results for the "Mark" class, with an accuracy of 0.932 and MIoU of 0.950 in the test, consolidating the stability of the results. For the "Wound" class, the accuracy remained around 0.759 and the MIoU was 0.356, demonstrating that although the data augmentation techniques and the greater number of epochs have brought improvements, generalization for the "Wound" class still presents challenges.

Table 6.7: Analysis of the image-based approach using Yolo: results of the performance metrics for the training and test set, separated by class, using data-augmentation and increasing the number of epochs.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Test	Train	Test	Train	Test
epochs = 10 ; batch = 16	Mark	0.162	0.113	0.631	0.157	0.111	0.030
lr = 0.0001	Wound	0.027	0.008	0.285	0.105	0.008	0.007
epochs = 30 ; batch = 16	Mark	0.979	0.979	1.000	1.000	0.922	0.958
lr = 0.0001	Wound	0.926	0.760	0.842	0.635	0.716	0.400
epochs = 100 ; batch = 16	Mark	1.000	0.932	1.000	1.000	0.923	0.950
lr = 0.0001	Wound	0.943	0.759	0.947	0.650	0.774	0.356

Figure 6.2 shows examples of prediction, with different results for the two images, as described below:

On the left-hand side (a), we observe a wound located in an area of skin with various signs and variations in texture, which could pose a challenge for detection. Despite this visual complexity, the YOLO model was able to correctly identify both the wound and the calibration mark. The wound was detected with a reliability of 0.8, while the mark was detected with a reliability of 0.9, suggesting that even in situations with many additional elements (such as signs), the model was able to distinguish the area of interest well.

On the right (b), we have a slightly more challenging situation. The wound is on one leg and there is a complex background with additional elements. In this case, the YOLO model identified the wound with a reliability of 0.9 in correctly segmenting the area. The calibration mark was detected with 0.9, suggesting that it was not affected by the complexity of the background.

An important point to note is that the model misidentified part of the background as a wound, with a high degree of reliability (0.6). This false positive shows that the model was fooled by the visual context, considering a part of the surrounding environment to be a lesion, which is a clear indication that backgrounds with many variations and elements can confuse the model.

This analysis suggests the need to guide healthcare professionals to reduce the complexity of the background during image capture. More uniform and neutral backgrounds, or even the absence of a background other than skin, help to improve the accuracy of the model, minimizing the occurrence of false positives and increasing confidence in the detection of wounds.

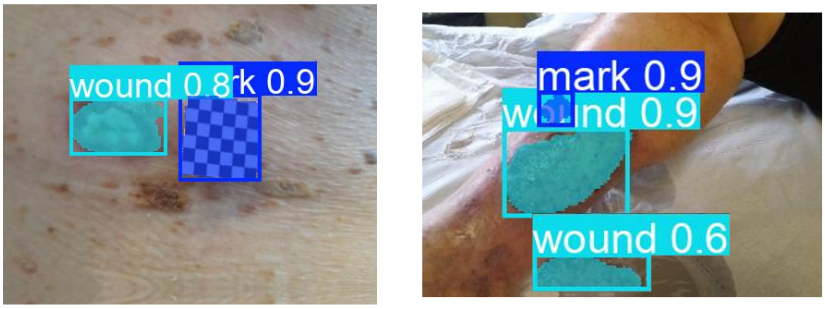

Figure 6.2: Prediction examples of the YOLO model applied to the task of segmenting wounds and calibration marks.

MedT

Before analyzing the results, we look at the evolution of training. Figure 6.3 shows the graph of Train Loss and Validation Loss over 100 epochs.

In the first 20 to 30 epochs, there is a significant drop in both training and validation loss. This is the expected behaviour in successful training, where the model begins to learn and adjust its weights to minimize errors.

After this initial adjustment phase, the losses stabilize. The validation loss begins to fluctuate slightly, but remains close to the training loss. It's a good sign that there isn't a big divergence between the training and validation errors, which suggests that the model isn't overfitting.

Fluctuations in the validation loss over time, especially after 40 epochs, can be normal in complex models. These oscillations can occur due to the natural variability of the validation data, which is more difficult to predict than the training data. However, the oscillations remaining close and not increasing significantly indicates that the model is managing to maintain its ability to generalize.

The results of the MedT network, shown in Table 6.8, show the positive impact of increasing the number of epochs and applying data augmentation techniques to the training and validation sets. With 30 epochs, batch size = 4 and learning rate = 0.001, the

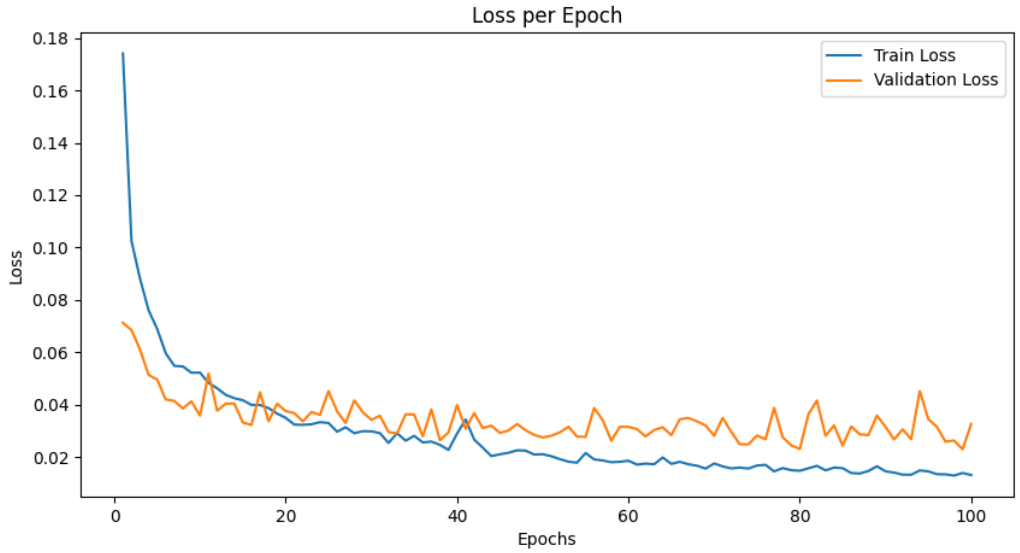

Figure 6.3: Graph of the evolution of the Train Loss and Validation Loss lines over 100 epochs.

"Mark" class achieved an accuracy of 0.896 and Recall of 0.951 in the validation set, while the "Wound" class showed an accuracy of 0.594 and a MIoU of 0.506. These values show an improvement compared to previous experiments, especially in the "Wound" class, although there are still difficulties in accurately segmenting this class.

At 100 epochs, the "Mark" class continued to show high results, with a precision of 0.977, Recall of 0.958 and MIoU of 0.937 in the validation set. The "Wound" class also showed improvements, with a precision of 0.699 and MIoU of 0.584, suggesting that the increase in the number of epochs, combined with data augmentation techniques, contributed to better segmentation of this class.

Table 6.8: Analysis of the image-based approach using MedT: results of the performance metrics for the training and validation set, separated by class, using data-augmentation and increasing the number of epochs.

Hyperparameters	Class	Precision		Recall		MIoU	
		Train	Valid	Train	Valid	Train	Valid
epochs = 30 ; batch = 4	Mark	0.945	0.896	0.933	0.951	0.884	0.865
lr = 0.001	Wound	0.750	0.594	0.699	0.812	0.550	0.506
epochs = 100 ; batch = 4	Mark	0.969	0.977	0.966	0.958	0.937	0.937
lr = 0.001	Wound	0.866	0.699	0.846	0.772	0.745	0.584

The results on the test set, presented in Table 6.9, show that for the "Mark" class, the model achieved a precision of 0.946, Recall of 0.936 and MIoU of 0.909, suggesting a good generalization capacity. However, for the "Wound" class, performance was

lower, with a precision of 0.672 and an MIoU of 0.482, highlighting the persistence of challenges in segmenting this class, even after using data augmentation techniques.

Table 6.9: Analysis of the image-based approach using MedT: results of the performance metrics for the test set, separated by class, using data-augmentation for the largest number of epochs trained.

Hyperparameters	Class	Precision	Recall	MIoU
		Test	Test	Test
epochs = 100 ; batch = 4	Mark	0.946	0.936	0.909
lr = 0.001	Wound	0.672	0.598	0.482

6.2 Comparative Study of Image-Based Models

This section presents the results of the comparison between the two architectures used in the wound segmentation task: YOLO and MedT. The aim is to provide a comprehensive overview of the efficiency and generalization capacity of each model, highlighting its strengths and limitations in segmenting the "Mark" and "Wound" classes.

6.2.1 Model complexity

Before analyzing the segmentation results, it is important to understand the complexity of the models in terms of the number of parameters each architecture has. The YOLO model has 27,897,120 parameters, while the MedT architecture has a total of 1,564,246 parameters. This discrepancy in terms of complexity indicates that the YOLO model is significantly heavier and requires more computing resources for training and inference, while MedT, being a lighter model, tends to be more efficient in terms of memory usage and processing speed.

The difference in the number of parameters also reflects the design philosophies of these architectures. YOLO was designed to be a powerful and robust model, capable of handling large volumes of data and complex object detection and segmentation tasks. MedT [19], on the other hand, is a transformer model aimed specifically at medical segmentation. It uses attention mechanisms that allow it to identify global and local patterns in medical images, which is essential for capturing complex details and subtle characteristics present in this type of data. These properties make MedT a particularly effective choice for segmentation tasks in the medical domain, where accuracy and sensitivity to detail are crucial.

6.2.2 Performance comparison

The results obtained for each model are shown in Tables 8.7, 8.8 and 8.9, and here we discuss the differences observed in the performance of YOLO and MedT, considering the training, validation and test sets.

General Model Performance

The results for the YOLO and MedT architectures show notable differences in terms of precision, recall and MIoU between the classes. For YOLO, the increase in epochs initially brought improvements, but there was a slight drop in the MIoU of the "Mark" class when going from 30 to 100 epochs, possibly due to overfitting. Precision and recall remained at 1,000 during both training sessions, but segmentation consistency decreased.

For the "Wound" class, increasing the epochs in the YOLO model had no significant impact, and the MIoU dropped from 0.400 (with 30 epochs) to 0.356 (with 100 epochs). In contrast, the MedT model showed better performance and consistency. With 100 epochs, MedT achieved MIoU of 0.937 for "Mark" and 0.584 for "Wound", improving both precision and recall in both cases. These results indicate that MedT was the most effective model, benefiting most from the increase in epochs, especially in the segmentation of the "Wound" class.

Influence of the Number of Epochs and Data Augmentation

The impact of increasing the number of epochs and applying data augmentation techniques was significant for both models. For YOLO, the use of data augmentation helped improve the robustness of the model, especially for the "Mark" class, where the model achieved a perfect recall of 1,000 and a high MIoU of 0.958 with 30 epochs. However, increasing the number of epochs to 100 brought no additional benefits and even resulted in a slight drop in MIoU, suggesting possible overfitting over time.

For MedT, the application of data augmentation techniques, combined with the increase in epochs, resulted in notable improvements for both the "Mark" and "Wound" classes. MedT was able to benefit more from a greater number of epochs, by avoiding overfitting and improving the performance of both classes, although the "Wound" class still proved to be a challenge, even after 100 training epochs.

6.2.3 Final considerations

The results of the comparative study suggest that the choice of the ideal model depends on the application context. YOLO is a good option when detailed segmentation and high accuracy are required, especially for well-defined classes such as "Mark", as long

as sufficient computing resources are available. However, its tendency to overfitting, especially with a high number of epochs, must be carefully monitored.

MedT is an interesting choice when resource efficiency is a priority, and shows good segmentation consistency, especially in complex classes such as "Wound". The improvements obtained by increasing the number of epochs and applying data augmentation show that MedT can adapt well to scenarios where the model needs to deal with variations in the input data.

For the "Wound" class, both architectures presented difficulties, suggesting that additional approaches, such as increasing the data set, applying regularization techniques, or even combining the strengths of YOLO and MedT, could help overcome the limitations observed and achieve more accurate and consistent segmentation.

6.3 Discussions

This work explored the use of artificial intelligence-based segmentation models for wound analysis, specifically comparing the YOLO and MedT architectures. The experiments were carried out on a dataset collected through a partnership with the ULSBM, containing images of wounds with calibration marks, with and without data augmentation techniques. The results showed significant progress and also highlighted challenges that still need to be addressed.

In the state-of-the-art review, we highlighted that no work was found that employed transformer architectures specifically for wound segmentation, which makes the application of the MedT architecture more important. The performance of MedT, although it showed a relatively low MIoU, should be viewed positively, especially considering the complexity of the problem and the size of the dataset. Transformer architectures, such as UNETR [34] and Medical Transformer [19], have already proven to be effective for segmentation in other areas of medicine, including 3D segmentation and DICOM images where the majority of the images are two-color. The results obtained in this study suggest that there is significant potential for transformers in wound segmentation tasks, but that there is still room for improvement, especially in terms of precision

Additionally, in the case of the YOLO model, we found that it has only been mentioned in the state of the art for detection tasks, not for wound segmentation. This motivated us to put it to the test in the context of segmentation, using the latest available version of the architecture. This choice not only broadens YOLO's scope of application, but also adds value to the work presented by exploring its performance in a scenario for which it was not originally designed.

Compared to the state of the art, CNN-based models, such as U-Net and Deeplab, have proven to be particularly efficient, achieving superior performance metrics, with Dice

values higher than 0.90 in several studies, such as Mahbod et al. [9] and [16]. In comparison, the MedT model used here did not achieve such high values for the wound class, partly due to false positives that occurred in scenarios with complex backgrounds or the existence of too much area as background. These results suggest that there are clear benefits to using hybrid methods that combine CNNs and transformers, such as those presented in some state-of-the-art studies, such as TransUNet, which takes advantage of both local feature extraction (CNN) and global contextual analysis (transformers).

In the study by Scebba et al. [7], the Detect-and-Segment (DS) approach stood out for its ability to segment wounds in images with complex backgrounds, achieving an MCC of 0.85, which indicates robust segmentation even in challenging situations. Compared to the results of this study, it was possible to observe that the YOLO and MedT models had difficulties with more complex backgrounds, resulting in false positives where parts of the background were incorrectly identified as wounds. This suggests that integrating the DS method to first segment the region of interest and then apply a more refined segmentation could significantly improve the accuracy of the results.

The results obtained with the YOLO model showed promising performance, especially in terms of precision and recall for the "Mark" class, reaching values of 1.0 in some experiments. This suggests that YOLO has a good segmentation capacity, especially for objects with clear and consistent visual characteristics, such as the calibration mark. However, when increasing the number of epochs from 30 to 100, there was a slight drop in MIoU, which may indicate overfitting in some cases.

The "Wound" class showed greater segmentation difficulties in both models. MedT showed a gradual evolution in performance, benefiting from the increase in the number of epochs, but the MIoU values for validation were still lower than those observed in state-of-the-art CNN models, such as U-Net [16]. MedT's performance, even with positive results, indicates that the application of regularization and data augmentation techniques is essential to improve generalization capacity, especially considering the complexity of wounds, which vary in color, texture and shape.

7 Wound Area Calculation

After arriving at the model with the best result, we move on to calculating the wound area. This is a fundamental step in clinical analysis, allowing a quantitative assessment of the size of the wound and its evolution over time. This chapter presents the methods used to calculate the segmented area of wounds in square centimeters (cm^2), from the images processed by the MedT model. The process involves analyzing the segmented masks, identifying the regions corresponding to the wound and the calibration mark, which is used as a reference and is present in all the images.

The use of deep learning in wound segmentation, combined with automatic area calculation, can offer a more efficient and accurate alternative to traditional methods, such as manual measurement with a ruler. This chapter details how the calculations were carried out and presents the results obtained.

7.1 Methodology

7.1.1 Segmentation Masks

The images of the wounds were processed by the MedT model, resulting in segmented masks. The masks contain different grayscale colors and each one represents a specific class:

- **Color 0**: Background.
- **Color 85**: Wound.
- **Color 170**: Calibration mark ($2\ cm \times 2\ cm$).

Scale calculation per pixel

With the pixel-area ratio determined, it was possible to calculate the wound area using the pixel count of the corresponding class (color 85). Since the calibration mark (color 170), which occupies an area of 4 cm² (2 cm × 2 cm), we calculated the conversion scale from pixels to square centimeters, its formula is:

$$\text{Area per pixel}(\text{cm}^2) = \frac{\text{Área per mark } (\text{cm}^2)}{\text{Number of pixels in the mark}}. \tag{7.1}$$

$$\text{Area per pixel} = \frac{4}{543} \approx 0.0073\,\text{cm}^2 \text{ per pixel.}$$

7.1.2 Mask Processing

Mask Reading and Pixel Counting

As mentioned, the masks were processed to count the number of pixels belonging to each color, corresponding to the identified classes.
The example below shows a list of pixel numbers separated by class. This example is the one shown in Figure 7.1.

- **Color 0** (background): 63,824 pixels.
- **Color 170** (Wound): 543 pixels.
- **Color 85** (mark): 1,169 pixels.

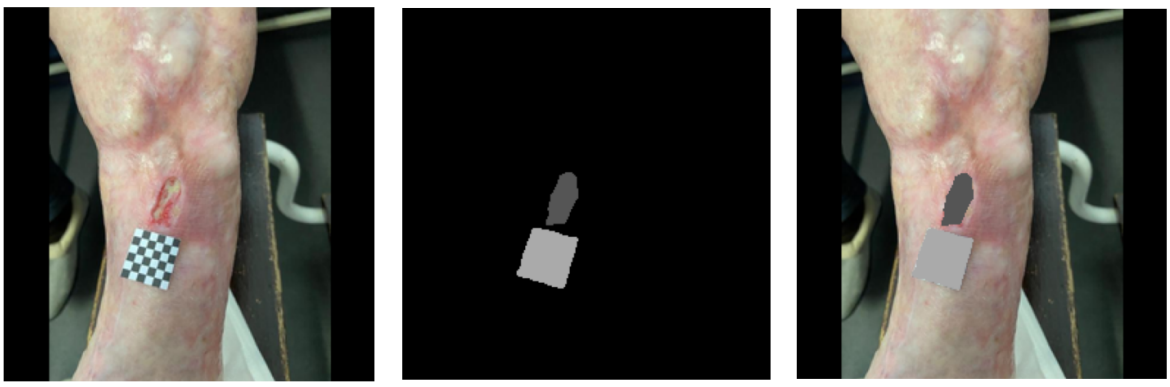

Figure 7.1: Analysis of the image segmentation belonging to the test dataset. Represented with (a) is the original image, image (b) is the segmentation predicted by the trained model. In image (c) is the predicted mask under the original image.

Calculating the Wound Area

We multiply the number of pixels belonging to the wound by the area of each pixel:

- **Area represented in Figure 7.1**:

$$\text{Wound area}(\text{cm}^2) = \text{Number of wound pixels} \times \text{Area per pixel}. \tag{7.2}$$

$$\text{Wound area} = 1,169 \times 0.0073 \approx 1.8580\,\text{cm}^2.$$

7.2 Results

Before concluding the analysis, we obtained Figure 7.2, which does not exist in any dataset, and put our model to the test.

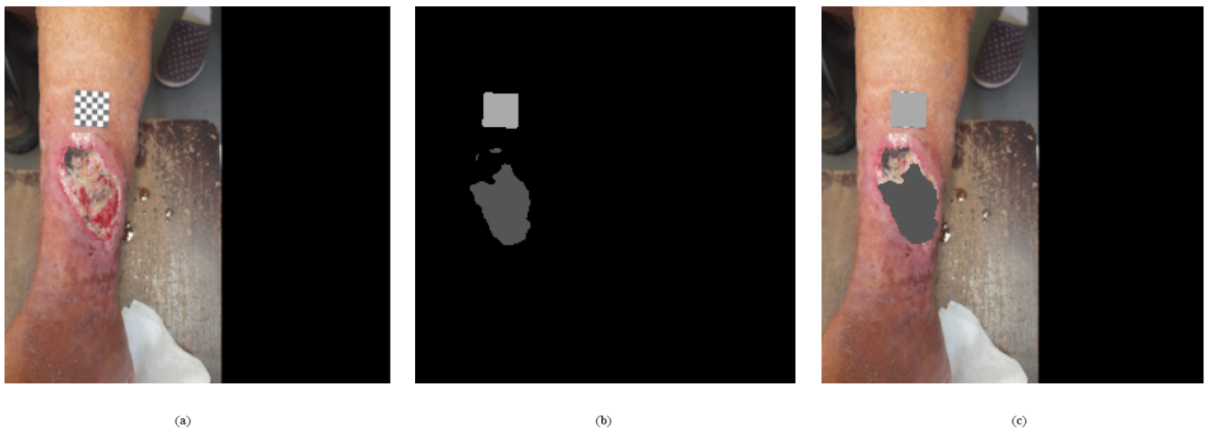

Figure 7.2: Analysis of the segmentation of the last image sent by the partnership, which has not yet been integrated into the test dataset. Represented with (a) is the original image, image (b) is the segmentation predicted by the trained model. In image (c) is the predicted mask under the original image.

The results of the wound area calculation are summarized below:

- **Area represented in Figure 7.1**:
  - Pixels of the wound (color 85): 543.
  - Wound area: **1.8580 cm²**.
- **Area represented in Figure 7.2**:
  - Pixels of the wound (color 85): 1,343.
  - Wound area: **10.2129 cm²**.

The results obtained show that the approach adopted is functional and accurate for calculating the area of wounds based on the masks generated by the MedT model. The use of the 2 cm $\times$ 2 cm calibration mark enables the conversion scale to be calculated accurately, allowing the model to generalize well even in scenarios where the images have different resolutions. However, we would point out that variations in segmentation or inaccuracies in identifying the reference mark can introduce errors into the area calculation. In Figure 7.2 we can see that the calculation of the wound was not as

accurate as in the first figure. We deduce that the lack of efficiency is due to the low number of wounds with necrotic skin in the training dataset and the poor quality of the image. This automated process has the potential to be integrated into clinical support systems, helping healthcare professionals to manage and treat wounds more efficiently and objectively.

8 Conclusion

In conclusion, this work has shown potential in the application of transformer-based models in wound segmentation, although there are still important challenges to overcome, such as the occurrence of false positives and the balance between overfitting and underfitting. The creation of a realistic dataset, including images collected through partnership and online images, was a significant contribution that allowed us to validate the models' ability to generalize to real cases in clinical contexts. This approach highlighted both the potential and limitations of the models when dealing with the diversity of cases found in health centers.

An important observation concerns the composition of the initial data set, which did not incorporate all the images made available by the health center. With the knowledge gained throughout this study, it became clear that the inclusion of more images representative of specific patient cases could have improved the performance of the models, making the dataset more diverse and suited to the practical challenges faced in clinical environments.

8.1 Limitations and Future Directions

Future research should explore:

The integration of hybrid approaches combining CNNs and transformers, taking advantage of both architectures. The application of the Detect-and-Segment method to improve accuracy, especially in cases of complex backgrounds, which currently generate false positives. Incorporating regularization techniques to improve the model's generalization capacity and reduce the observed overfitting effects. Expanding the dataset to include more cases provided by health centers, improving the representativeness of the examples and allowing the models to be trained more specifically for the types of wounds found in these institutions. To further assess whether the choice of dataset was the most appropriate. With a greater number of images collected through the partnership, the dataset can be expanded, enabling a detailed analysis of which specific characteristics of the data directly impact the performance of the models. Completing the PUSH scale using deep learning tools to automate the assessment of wound progression, providing more robust and accurate support to healthcare professionals. The validation of the models in real scenarios, integrating them into mobile applications, as suggested by the studies by Chuanbo Wang et al. [8] and others, to better understand

the limitations and potential in clinical environments.

References

[1] M. K. Dhar, T. Zhang, Y. Patel, and Z. Yu, "Fusegnet: A deep convolutional neural network for foot ulcer segmentation," *arXiv preprint arXiv:2305.02961*, 2023.

[2] D. Ramachandram, J. L. Ramirez-GarciaLuna, R. D. Fraser, M. A. Martínez-Jiménez, J. E. Arriaga-Caballero, J. Allport *et al.*, "Fully automated wound tissue segmentation using deep learning on mobile devices: cohort study," *JMIR mHealth and uHealth*, vol. 10, no. 4, p. e36977, 2022.

[3] S. Ather and K. G. Harding, "Wound management and dressings," *Advanced Textiles for Wound Care: A Volume in Woodhead Publishing Series in Textiles*, pp. 3–19, 1 2009.

[4] "Safe skin, safe patients: The value of patient hygiene - clinical consensus statement," *American Journal of Nursing*, vol. 123, pp. S1–S16, 6 2023. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/bed-sores/symptoms-causes/syc-20355893

[5] D. J. Leaper, G. Schultz, K. Carville, J. Fletcher, T. Swanson, and R. Drake, "Extending the time concept: What have we learned in the past 10 years?" *International Wound Journal*, vol. 9, pp. 1–19, 12 2012. [Online]. Available: https://www.ausmed.com.au/learn/guides/wound-care

[6] S. A. D. ALMEIDA, C. N. D. O. MOREIRA, and G. M. SALOME, "Pressure ulcer scale for healing in monitoring of wound healing in elderly patients with leg ulcer," *Revista Brasileira de Cirurgia Plástica*, vol. 29, pp. 120–127, 1. [Online]. Available: http://www.rbcp.org.br/details/1500/pt-BR

[7] G. Scebba, J. Zhang, S. Catanzaro, C. Mihai, O. Distler, M. Berli, and W. Karlen, "Detect-and-segment: A deep learning approach to automate wound image segmentation," *Informatics in Medicine Unlocked*, vol. 29, p. 100884, 2022.

[8] C. Wang, D. Anisuzzaman, V. Williamson, M. K. Dhar, B. Rostami, J. Niezgoda, S. Gopalakrishnan, and Z. Yu, "Fully automatic wound segmentation with deep convolutional neural networks," *Scientific reports*, vol. 10, no. 1, p. 21897, 2020.

[9] A. Mahbod, G. Schaefer, R. Ecker, and I. Ellinger, "Automatic foot ulcer segmentation using an ensemble of convolutional neural networks," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 4358–4364.

[10] M. Adnan, M. Asif, M. B. Ahmad, T. Mahmood, K. Masood, R. Ashraf, and C. N. Faisal, "An automatic wound detection system empowered by deep learning," in

Journal of Physics: Conference Series, vol. 2547, no. 1. IOP Publishing, 2023, p. 012005.

[11] D. Marijanović, E. K. Nyarko, and D. Filko, "Wound detection by simple feed-forward neural network," *Electronics*, vol. 11, no. 3, p. 329, 2022.

[12] A. Wagh, S. Jain, A. Mukherjee, E. Agu, P. C. Pedersen, D. Strong, B. Tulu, C. Lindsay, and Z. Liu, "Semantic segmentation of smartphone wound images: Comparative analysis of ahrf and cnn-based approaches," *IEEE Access*, vol. 8, pp. 181 590–181 604, 2020.

[13] R. Zhang, D. Tian, D. Xu, W. Qian, and Y. Yao, "A survey of wound image analysis using deep learning: classification, detection, and segmentation," *IEEE Access*, vol. 10, pp. 79 502–79 515, 2022.

[14] C. W. Chang, M. Christian, D. H. Chang, F. Lai, T. J. Liu, Y. S. Chen, and W. J. Chen, "Deep learning approach based on superpixel segmentation assisted labeling for automatic pressure ulcer diagnosis," *PLoS One*, vol. 17, no. 2, p. e0264139, 2022.

[15] J. Li, S. Yang, and Q. Chen, "Sws-net: An image segmentation framework for chronic wounds based on self-supervised learning," in *The 35th International Conference on Software Engineering and Knowledge Engineering, SEKE 2023, KSIR Virtual Conference Center, USA, July 1-10, 2023*, S.-K. Chang, Ed. KSI Research Inc., 2023, pp. 691–696. [Online]. Available: https://doi.org/10.18293/SEKE2023-134

[16] N. Curti, Y. Merli, C. Zengarini, E. Giampieri, A. Merlotti, D. Dall'Olio, E. Marcelli, T. Bianchi, and G. Castellani, "Effectiveness of semi-supervised active learning in automated wound image segmentation," *International Journal of Molecular Sciences*, vol. 24, no. 1, 2023. [Online]. Available: https://www.mdpi.com/1422-0067/24/1/706

[17] R. Brüngel and C. Friedrich, "Detr and yolov5: Exploring performance and self-training for diabetic foot ulcer detection," 06 2021, pp. 148–153.

[18] B. Zhang, Z. Tian, Q. Tang, X. Chu, X. Wei, C. Shen, and Y. Liu, "Segvit: Semantic segmentation with plain vision transformers," 2022.

[19] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, 2021, pp. 36–46.

[20] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 574–584.

[21] D. Anisuzzaman, Y. Patel, J. A. Niezgoda, S. Gopalakrishnan, and Z. Yu, "A mobile app for wound localization using deep learning," *IEEE Access*, vol. 10, pp. 61 398–61 409, 2022.

[22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[24] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024.

[25] ——, "Yolov9: Learning what you want to learn using programmable gradient information," 2 2024. [Online]. Available: https://docs.ultralytics.com/pt/models/yolov9

[26] "F. chollet, deep learning with python. simon and schuster, 2021."

[27] "Colab is a hosted jupyter notebook service that requires no setup to use and provides free access to computing resources, including gpus and tpus. colab is especially well suited to machine learning, data science, and education." [Online]. Available: https://colab.google/

[28] "This project aims at wound area segmentation from natural images in clinical settings." 2021.

[29] "The free open source image editor." [Online]. Available: https://www.gimp.org/

[30] "This project aims at wound area segmentation mohamad taher," 2023. [Online]. Available: https://universe.roboflow.com/lucas-245yq/wound-segmentation-ywxdj

[31] "This project aims at wound area segmentation mohamad taher," 2021.

[32] TA, "segmentasiluka dataset," https://universe.roboflow.com/ta-bluys/segmentasiluka, sep 2023, visited on 2024-03-10. [Online]. Available: https://universe.roboflow.com/ta-bluys/segmentasiluka

[33] hyunnununanna, "augmented-dtpi(lastest) dataset," https://universe.roboflow.com/hyunnununanna/augmented-dtpi-lastest, aug 2023, visited on 2024-04-01. [Online]. Available: https://universe.roboflow.com/hyunnununanna/augmented-dtpi-lastest

[34] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," 2021. [Online]. Available: https://arxiv.org/abs/2103.10504

Instituto Superior de Engenharia
Politécnico de Coimbra