AUTOMATIC ASSESSMENT OF BURN INJURIES USING ARTIFICIAL INTELLIGENCE

by

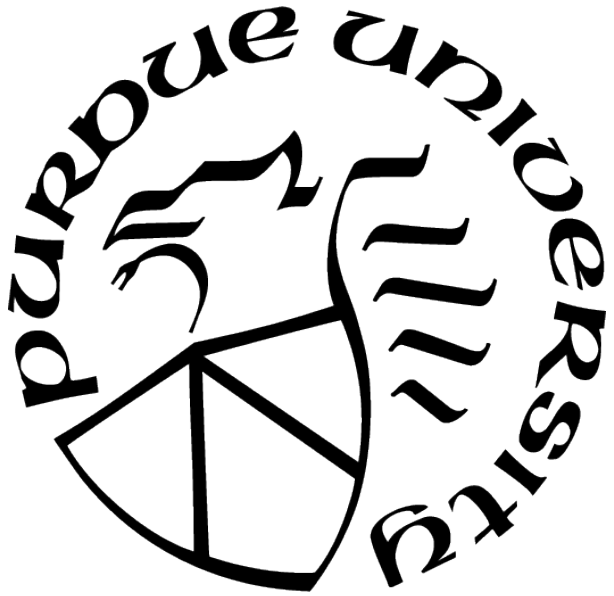
Daniela Chanci Arrubla

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Industrial Engineering



School of Industrial Engineering

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Juan P. Wachs, Chair

School of Industrial Engineering

Dr. Ramses Martinez

School of Industrial Engineering

Dr. Chandan K. Sen

Indiana University School of Medicine

Approved by:

Dr. Abhijit Deshmukh

This dissertation is dedicated to my family.

ACKNOWLEDGMENTS

First, I would like to express my gratitude to my advisor, Dr. Juan Wachs. He showed me how to properly conduct research, and provided his support and motivation during all my time at Purdue, especially during the writing of this thesis. I would also like to thank the rest of my committee for helping me in this process, Prof. Ramses Martinez, for his insightful comments and suggestions, and Dr. Chandan K. Sen, for providing valuable knowledge and materials used in this dissertation.

I would also like to express my gratitude to all the members of the ISAT lab. Special thanks to Juan Antonio Barragan for being my company and support during this journey, and to Naveen Madapana and Glebys Gonzalez for being my mentors since the beginning.

Last, my sincere thanks to my parents, my sister, my family, and my friends here at Purdue for their company and endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ABA	American Burn Association
AI	Artificial Intelligence
ASM	Angular Second Moment
CAD	Computer-Aided Detection or Diagnosis
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CT	Computer Tomography
DNN	Deep Neural Network
FCN	Fully Convolutional Network
FN	False Negatives
FP	False Positives
FPS	Frames Per Second
GA	Global Pixel Accuracy
GAN	Generative Adversarial Network
GLCM	Gray Level Co-ocurrence Matrix
HUSD	Harmonic B-mode Ultrasound
IDE	Integrated Development Environment
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IR	Infrarred Tomography
IUSM	Indiana University School of Medicine
JI	Jaccard Index
JPG	Joint Photographic Group
LDI	Laser Doppler Imaging
LSI	Laser Speckle Imaging
MRI	Magnetic Resonance Imaging
mIoU	Mean Intersection Over Union
OCT	Optical Coherence Tomography

OPS	Orthogonal Polarization Spectral
PET	Positron Emission Tomography
PNG	Portable Network Graphics
ReLU	Rectified Linear Unit
ResNet	Residual Network
RNN	Recurrent Neural Network
ROI	Region of Interest
STD	Standard Deviation
SVM	Support Vector Machine
TBSA	Total Body Surface Area
TN	True Negatives
TP	True Positives
VGG	Visual Geometry Group

ABSTRACT

Accurate assessment of burn injuries is critical for the correct management of such wounds. Depending on the total body surface area affected by the burn, and the severity of the injury, the optimal treatment and the surgical requirements are selected. However, such assessment is considered a clinical challenge. In this thesis, to address this challenge, an automatic framework to segment the burn using RGB images, and classify the injury based on the severity using ultrasound images is proposed and implemented. With the use this framework, the conventional assessment approach, which relies exclusively on a physical and visual examination of the injury performed by medical practitioners, could be complemented and supported, yielding accurate results. The ultrasound data enables the assessment of internal structures of the body, which can provide complementary and useful information. It is a noninvasive imaging modality that provides access to internal body structures that are not visible during the typical physical examination of the burn. The semantic segmentation module of the proposed approach was evaluated through one experiment. Similarly, the classification module was evaluated through two experiments. The second experiment assessed the effects of incorporating texture features as extra features for the classification task. Experimental results and evaluation metrics demonstrated the satisfactory results obtained with the proposed framework for the segmentation and classification problem. Therefore, this work acts as a first step towards the creation of a Computer-Aided Diagnosis and Detection system for burn injury assessment.

1. INTRODUCTION

Burn injuries are considered one of the most complex type of traumatic injuries, and are a principal cause of morbidity and mortality (Herndon et al., 1989; "World Health Organization", 2018). These injuries can be caused by heat, contact, friction, radiation, electrical or chemical sources. However, according to the American Burn Association (ABA), fire/flame and scalds correspond to the 43% and 34% of burn admissions to burn centers, respectively ("American Burn Association", 2015). It is estimated that 1.25 million people are treated each year for burns in The United States, and 40,000 are hospitalized for the treatment of these injuries resulting in high medical costs, approximately $7.9 billion per year (Browning and Cindass, 2019; Jeschke et al., 2012).

Wound care is one of the most crucial elements in the management of burn injuries. Prompt and accurate burn assessment is critical to decide the optimal management strategy, improve the patient's outcomes, and decrease the risk of infection and mortality. In addition, it enables the prioritization of patients that require a surgical procedure in order to heal, and need to be transferred to a specialized burn facility (Jeschke et al., 2020). In this regard, it is essential to determine the burned surface area and the burn depth. However, both tasks continue to constitute a major clinical challenge.

Burns can be classified into three categories: (1) Superficial burns, which only involve the epidermis. They do not blister and they are usually red, warm, dry, painful, and blanch when pressure is applied. They can heal without scarring within 5 to 10 days. (2) Partial-thickness burns, which affect the epidermis and the dermis. They are very painful, may be swollen, and appear red, moist and blistered. Typically, they heal spontaneously within 14 to 21 days with minimal to no scarring. (3) Full-thickness burns, which affect the epidermis, the dermis, and underlying subcutaneous tissue. Some may also involve deeper structures, such as muscle, tendon, and bone. There is usually minimal to no pain, and they look dry, leathery, white or brown/black. They heal with severe scarring and contractures, and the patients might experience itch, pain, loss of function, restricted movement, and dissatisfaction with their physical appearance. In order to prevent these outcomes, they require early excision and grafting (Finnerty et al., 2016; Kagan et al., 2013; Warby and Maani, 2019).

In the case of the burn size estimation, the manual used methods, such as the Wallace Rule of Nines and the Lund and Browder chart, tend to significantly overestimate the total body surface area (TBSA) burned, with a mean overestimation accuracy of 170% (Harish et al., 2015; Zuo et al., 2017). In the case of burn depth, it is determined by clinical assessment based on appearance, blanching to pressure, sensation to pin prick, and bleeding on needle prick. This visual and tactile inspection approach introduces inter-subject variability, especially when partial-thickness burns are involved (Zuo et al., 2017). Hence, for this task, experienced burn surgeons achieve an accuracy of 67-76%, value that decreases to 50% for inexperienced practitioners (Brown et al., 1998; Despo et al., 2017). This issue is exacerbated by the fact that burn depth underestimation or overestimation leads to poor clinical outcomes. The first results in delayed healing times and significant hypertrophic scarring, and the second results in unnecessary surgery (Cubison et al., 2006).

Different non-invasive imaging technologies have been implemented to improve the accuracy of the clinical evaluation for the burn assessment task by supporting the medical decision-making process, but none of them have been widely adopted by clinicians. Laser Doppler Imaging (LDI) is the most known technique. It consists of a noncontact scanning in which a laser light is pointed at erythrocytes in movement to obtain the blood perfusion. As a result, a color map of the perfusion, which corresponds to the burn depth, is generated based on the existing correlation between blood perfusion and burn depth (Monstrey et al., 2008). However, this maps can be difficult to interpret and since this technology uses light/optic principles, the accuracy can be greatly affected by the heterogeneity and curvature of the tissues, skin tone, topical dressings, blisters, and ambient light (Devgan et al., 2006). Another known technique is thermography, which uses the inverse correlation between burn wound temperature and depth, given that deeper wounds are colder due to less vascular perfusion. Nevertheless, evaporative loss of heat to the environment and the presence of granulation tissue have a negative impact in the accuracy of this technique (Devgan et al., 2006).

In this work, we explore the use of Artificial Intelligence (AI), especially deep learning, texture analysis and ultrasonography, aiming to overcome the mentioned drawbacks in the burn assessment process. Since the dramatic progress in AI to address complex decision-making problems, several computer-aided decision support systems have been developed in

the area of skin lesions and wounds, which includes burn injuries. Deep learning models have been implemented to obtain the burn depth from clinical images (Cirillo et al., 2019; Tran et al., 2016). However, similar to LDI, skin conditions and ambient light can introduce errors in the system. Thus, we leverage ultrasound imaging to obtain physical and hemodynamic characteristics of the skin and subcutaneous burned tissue that enable a reliable burn depth classification (Sen et al., 2016).

B-mode ultrasound is a rapid and noninvasive imaging technique widely used in the medical field. The basic operation principle involves the transmission of small pulses of ultrasound echo into the region of interest (ROI) of the body, some of them are then reflected back to the transducer as they reach body tissues of different acoustic impedances, to be processed and combined for the image generation (Narouze, 2018). To reduce the noise artifacts that are usually present due to reflected echoes of different frequencies, harmonic B-mode ultrasound is used (HUSD). In this modality, harmonic frequency echoes, more specifically, the second harmonic frequency is used to generate a better-quality image (Gnyawali et al., 2015a; Narouze, 2018; Ziegler and O'Brien, 2002).

We leverage HUSD, burn injury images, and deep learning to support the burn assessment process. In this thesis, we create a customized labelled dataset from online sources, and train a modified U-Net model from scratch to perform semantic segmentation of burn injury images. In addition, we train different supervised deep learning models using a transfer learning approach and HUSD data from porcine subjects, and compare their performance in the prediction of the burn depth. The proposed work benefits the burn assessment process by enhancing the accuracies for the burn size estimation and the burn depth classification through the use of a reliable and convenient imaging technique and AI capabilities, and eliminating the inter-subject variability for both tasks.

1.1 Problem Statement

Implement a framework to support burn injury assessment using deep learning and two imaging modalities. This includes: segmenting the burn area based on a pixel-wise classifica-

tion approach, predicting the burn depth with multiple deep learning models, and comparing the performance of such models.

1.1.1 Research Questions

RQ1: How can deep learning be used for burn area size estimation using images obtained in uncontrolled conditions?
This question addresses the implementation of a semantic segmentation model to classify the pixels into three different classes: burn injury, skin, and background.

RQ2: What is the best model for burn depth assessment that leads to a higher classification accuracy?
This question relates to how well widely used deep learning models perform the burn depth classification task for HUSD data.

RQ3: Does the integration of texture features lead to a better burn depth classification accuracy?
This question explores how the inclusion of texture features into the previous implemented models can impact their performance.

1.2 Thesis Structure

The rest of the thesis document is divided in the following sections. Chapter 2 presents a literature review that summarizes relevant topics for this work. Chapter 3 provides a description of the methodology used to address the research questions. Chapter 4 describes the experiments and the results obtained. Finally, chapter 5 summarizes the conclusions and future work.

1.3 Summary

In the first section of this chapter, we presented the motivation to explore the use of ultrasound imaging and deep learning for medical decision-making support in the field of

burn injury assessment. Then, the three research questions that describe what will be addressed in this thesis were introduced.

2. LITERATURE REVIEW

This chapter is an overview of the past and state-of-the-art research on the areas related to this thesis, namely automatic detection, classification and segmentation of burn injuries. Each section presents a brief background of the topic to be discussed, followed by relevant examples found in the literature. First, computer-aided detection or diagnosis in medical images is described. In the development of such systems, two main components must be considered: segmentation and classification. These constitute the second and the third section, respectively. In each, the corresponding problem, segmentation or classification, and the studied approaches are described, focusing on deep learning-based models. General examples in the medical domain are presented, followed by those that explore the application of these approaches to wound care and burn assessment in particular. The last three sections discuss three techniques used in this thesis: transfer learning, ultrasonography, and image texture analysis. Lastly, the corresponding examples are presented.

2.1 Computer-Aided Detection or Diagnosis in Medical Imaging

Computer-Aided Detection or Diagnosis (CAD) is a major area of research in the medical field. Even though computing involves any form of digital analysis, the efforts are typically targeted towards medical image analysis. These systems take an image as an input and generate an output that is used to support medical decision-making process, serving as a second opinion for the practitioner (Doi, 2007). In this regard, it has been demonstrated that they can improve diagnostic accuracy and decrease inter-subject variability while reducing the time taken in the process. Consequently, CAD systems have been adopted by medical institutions and have been incorporated into the clinical routine (Cheng et al., 2016; Shiraishi et al., 2011).

There is a significant number of deep learning architectures that have been proposed and implemented in the last years for medical imaging and diagnostics purposes, as well as a variety of imaging modalities used in medical imaging, such as Ultrasound, Computer Tomography (CT), Magnetic Resonance Imaging (MRI), and X-Rays. This has led to the development of several deep learning-based CAD systems for an ample range of applications

in this field (Mansour, 2018; Pisano, 2020). Such systems have outperformed traditional machine learning-based CAD systems and are the state-of-the-art in medical image analysis. Segmentation and classification are two fundamental components of any CAD system because they can provide accurate and relevant information about the evaluated lesion or disease (Jackowski and Goshtasby, 2005). Both of them are included in this thesis, and they are described throughout this section.

2.2 Semantic Image Segmentation

Semantic segmentation is a widely known computer vision technique in which each pixel of an image is classified into a class (Taghanaki et al., 2021). Therefore, it is possible to identify different objects present in the image based on the clusters formed by the pixels that belong to the same class, leading to a better understanding of the image (Guo et al., 2018). This technique is used in several critical applications, such as medical image analysis (Rezaei et al., 2017), augmented reality (Ko and Lee, 2020), human-computer interaction (Kalayeh et al., 2017), autonomous vehicles (Treml et al., 2016), scene understanding (Gupta et al., 2015), satellite imaging (Hordiiuk et al., 2019), and precision agriculture (Milioto et al., 2018). In the following subsections, deep learning-based methods and the essential role of semantic segmentation in medical images, especially in skin wounds, which includes burns, are discussed.

2.2.1 Deep Learning-based Semantic Segmentation

Before deep learning models became popular for image related problems, traditional image processing and computer vision techniques were used to tackle this task. They leveraged colors, textures, edges, and contours. Some of the well-known traditional methods that can be found in the literature include: thresholding (Kohler, 1981), Markov random fields (F. Yang and Jiang, 2003), and k- means clustering (Ng et al., 2006). However, after the significant improvement in the accuracy for computer vision tasks introduced by deep learning models in 2012, the focus of the semantic segmentation problem shifted, and the development of such models became the standard (Garcia-Garcia et al., 2017). In this regard,

several methods and architectures have been proposed in recent years. They use a variety of techniques and can be highly specialized for specific applications. Nevertheless, the backbone of these methods is usually based on one of the commonly known deep learning architectures, mainly Visual Geometry Group (VGG) (Simonyan and Zisserman, 2014) and Residual Network (ResNet) (He et al., 2016).

State-of-the-art deep learning segmentation models for semantic segmentation can be classified into multiple categories based on specific characteristics. In terms of the type of learning, they can be classified into three different categories: supervised learning, weakly-supervised learning, and semi-supervised learning models (Hao et al., 2020).

Supervised learning methods are the most commonly used in classification tasks, given that there is sufficient labeled data to train a computational model (Nasteski, 2017). For the semantic segmentation task, the dataset is composed of images and their corresponding masks. By mask, we refer to the ground truth, which is an annotated image in a pixel-wise manner according to the different objects that we are interested in identifying (Garcia-Garcia et al., 2017). In this regard, the model learns a function that is able to predict the mask based on an input image (Goodfellow et al., 2016). Long et al. (2015) proposed the first semantic segmentation method using Convolutional Neural Networks (CNN) without dense layers. It is known as Fully Convolutional Networks (FCN). When it was first introduced, there were three outstanding characteristics: (1) it can be trained in an end-to-end manner, (2) it is highly efficient, and (3), it is able to take variable-sized inputs and produce equivalent-sized outputs. Several modifications of FCN for semantic segmentation have been proposed to expand its capabilities and improve the resulting accuracy (Lin et al., 2017; W. Liu et al., 2015; Shuai et al., 2016; Yu and Koltun, 2015). Additionally, among other widely-used models, DeepLabv3 (Chen et al., 2017) and Mask-RCNN (He et al., 2017) are leading techniques in image segmentation.

In contrast, weakly-supervised and semi-supervised learning do not depend as much on pixel-wise annotations. In the former, the segmentation is performed using images with image-level labels corresponding to the objects present in the image (Ahn and Kwak, 2018). In this regard, the model predicts the label for each pixel of the image based on weak annotations that lack information corresponding to the location and shape of the objects

that will be identified. The use of class activation maps (Ahn and Kwak, 2018), seeded region growing (Z. Huang et al., 2018), expanded receptive fields (Y. Wei et al., 2018), and common object features (X. Wang et al., 2018) have been explored. In semi-supervised learning, only a tiny fraction of the data is annotated in a pixel-wise manner, and the remaining data is unlabeled. Typically, the proposed models leverage adversarial learning techniques (Hung et al., 2018; Souly et al., 2017). Both learning approaches, weakly supervised and semi-supervised, constitute a trending research line since they address one critical challenge in semantic segmentation: the availability of labeled datasets. This is challenging due to the cost and time required for the annotation process. However, such techniques have not been able to outperform supervised learning methods (Y. Wei et al., 2018).

Besides the classification based on the type of learning, Minaee et al. (2021) proposed a classification based on the AI technique used in the implementation of the models. Some of the techniques are: FCNs, graphical models, encoder-decoder blocks, feature pyramid networks, region based CNNs, dilated convolutions, recurrent neural networks (RNN), attention gates, generative adversarial networks, and active contour models. In this work, we use an encoder-decoder-based architecture which will be described in the subsection below. Additionally, a pixel-wise labeled dataset is used, i.e., the supervised learning approach described above is followed.

2.2.1.1 Encoder-Decoder-based Models for Semantic Segmentation

In general, an encoder-decoder architecture is constituted by two major blocks. The first one extracts the latent representation of an input, and from this, the second one generates the desired output. They are known as encoder and decoder, respectively (Cho et al., 2014). Several models have leveraged this architecture, introducing the modifications required for the specific task of semantic segmentation (Badrinarayanan et al., 2017). In this regard, the decoder recovers the spatial information lost in the encoder block due to the convolution and pooling operations present, and achieves a pixel-wise classification (Chen et al., 2018). This family of models is the state-of-the-art in Medical Image Segmentation (Tam, 2020). As expected, they have drowned the attention of researchers in this field, given the critical

nature of the segmentation task in the development of Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx) systems. Pertaining to this thesis, the most well-known architecture in this domain, U-Net, is described below.

U-Net

The U-Net architecture was proposed by Ronneberger et al. (2015) and was specifically targeted towards medical applications. Fundamentally, it tackles one of the most significant challenges AI faces in this domain: the limited accessibility to sufficient labeled data (Diaz et al., 2021). In other words, it is able to achieve satisfactory results using a small-sized dataset. In this regard, the architecture exploits data augmentation, upsampling operations, and skip connections. The last two enable the simultaneous use of spatial and semantic information leading to a high segmentation accuracy (Alom et al., 2018).

In this symmetrical architecture, the encoder and decoder blocks can be identified, and they are presented as the contracting and expansive paths, respectively. The basic block of the contracting path consists of two 3×3 convolutions, a rectified linear unit (ReLU) activation, and a 2×2 max pooling layer. As a result, the spatial resolution is downsampled to half its original size, and the number of feature channels is doubled in each step, obtaining the semantic information. In contrast, the basic block of the expansive path consists of a 2×2 up-convolution that reduces the number of feature channels by 50%, a concatenation with the cropped feature map from the same hierarchical level in the contracting path, two 3×3 convolutions, and a ReLU activation. The concatenation of the feature maps is achieved through skip connections and contributes to having a precise location. In addition, the last layer of the network is a 1×1 convolutional layer that produces the required number of output channels, which is equivalent to the number of classes in a determined problem (Ronneberger et al., 2015; Siddique et al., 2020). The described architecture is depicted in Figure 2.1.

U-net has had a dramatic impact on the state-of-the-art. Diverse medical and non-medical applications have been developed using this semantic segmentation model. For example, it has been used for the segmentation of vegetation using satellite images (Flood et

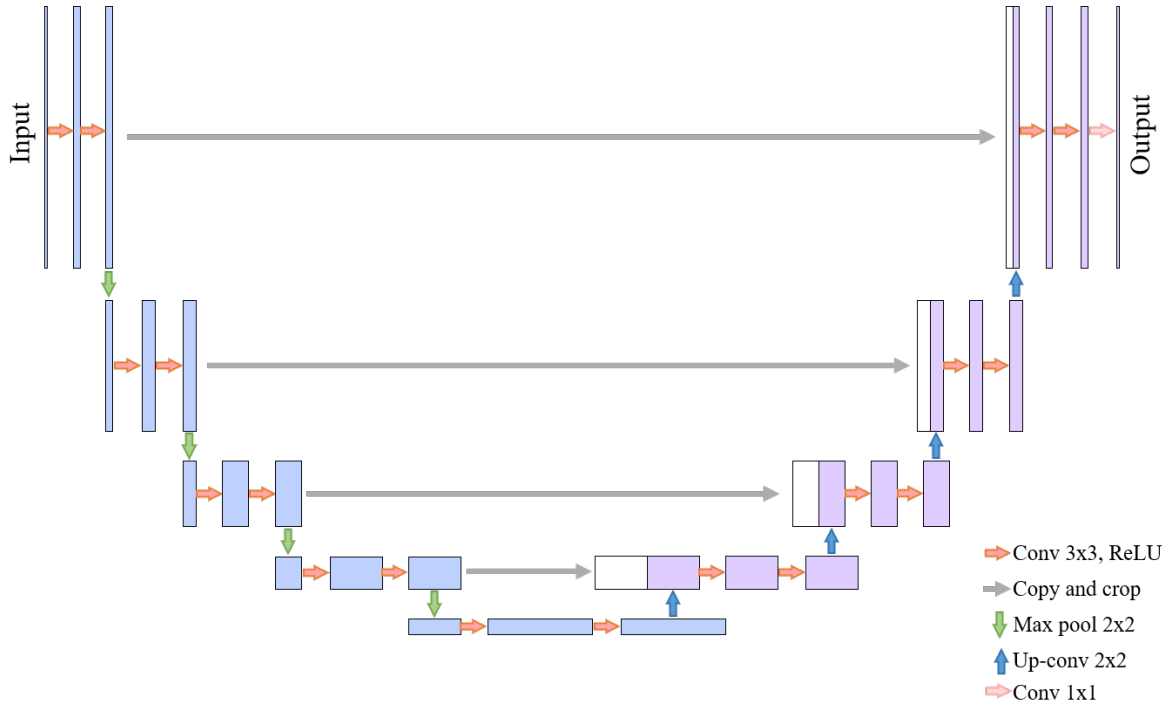


Figure 2.1. U-Net original architecture

al., 2019), for smart home developments (Stephan and Santra, 2019), and for road detection (X. Yang et al., 2019). In the medical domain, which was the initial motivation for the development of this architecture and is related to this work, it has been used with 2D and 3D data. Some applications include blood vessel segmentation (Xiancheng et al., 2018), cell detection and counting (Falk et al., 2019), and ulcers and tumors detection (Chae et al., 2021; Dong et al., 2017). Additionally, the impact of this architecture is also seen in the number of variants introduced in the last years. To name a few, the original U-Net has been combined with Recurrent Neural Networks (RNN) (Alom et al., 2018), nested and dense skip connections (Z. Zhou et al., 2018), and attention gates (Oktay et al., 2018).

2.2.2 Semantic Segmentation in Medical Images

Medical image segmentation is a fundamental task in the area of medical image analysis, and it is also one of the most challenging ones due to the inherent variability of the human body. Its main goal is to detect and characterize specific lesions, anatomical structures, or

ROI, either at the macroscopic level, such as organs and body parts, or at the microscopic level, like tissues and cells. As a result, it enables the obtention of critical information about the location, area, shape, among others (Hesamian et al., 2019). This is highly beneficial for early detection, diagnosis, monitoring, and treatment of different diseases or complex conditions that the body might present (T. Zhou et al., 2019).

In the literature, there is an abundance of deep learning-based models for segmentation of different imaging modalities, such as Ultrasound, CT, MRI, Positron Emission Tomography (PET), X-Ray, RGB-D, RGB, or even for hybrid modalities. Additionally, most efforts in this field are directed towards cancer screening, thoracic ROI, and brain disorders, using slices of MRI and CT. For example, Rezaei et al.(2017) proposed an end-to-end architecture based on Generative Adversarial Networks (GANs) for the segmentation of heterogeneous brain tumors using MRI images. Similarly, Li et al. (2018) implemented a variation of U-Net for volumetric data to segment liver tumors using CT scans, and later, Almotairi et al. (2020) proposed a modification of SegNet (Badrinarayanan et al., 2017) for the same problem. Lastly, Cai et al. (2020) recently proposed a model based on Mask R-CNN for detection and segmentation of lung nodules using the LUNA 16 dataset that contains annotated thoracic CT scans.

In addition to the examples mentioned above, semantic segmentation can be utilized for various lesions and conditions of the human body related to the skin. Creating a dataset for skin related segmentation problems can be accomplished by acquiring the pictures from a smartphone or a camera. In spite of the simplicity of the process, datasets with skin images are not abundant. Melanoma detection and segmentation is the only application that stands out in this field since the advent of deep learning. In the following subsection, we discuss the use of semantic segmentation for the particular case of wound care.

2.2.2.1 Semantic Segmentation for Wound Care

The skin is the largest organ in the body and it is responsible for the protection of internal organs and the regulation of the body temperature. It is constantly exposed to the outer world, making it susceptible to injuries and infections. Such injuries can sometimes

progress into serious wounds that can require specialized treatment (Sumithra et al., 2015). One essential piece of information in wound care is the area of the wound. This helps to determine the optimal treatment, in addition to being a valuable measurement for the monitoring and assessment of the wound healing process.

In this regard, some researchers have explored the use of semantic segmentation in wound care. This problem can be defined as a foreground-background segmentation in which the wound is the ROI, i.e., the foreground. For instance, in 2015, a CNN with an encoder-decoder architecture was proposed to segment different types of wounds. It could be trained in an end-to-end manner and used raw RGB images as inputs (C. Wang et al., 2015). Later, a Deep Neural Network (DNN) based on MobileNets (Howard et al., 2017) was implemented by Li et al. (2019). In addition to the RGB inputs, this model used location maps encoded by location kernels as an extra input to enhance the segmentation accuracy.

In this thesis, we focus exclusively on burn injuries, which can result in complex wounds. In this regard, obtaining an accurate burn size in terms of TBSA is crucial to determine the amount of fluid resuscitation required and select the optimal management plan (Jeschke et al., 2020). The use of deep learning-based semantic segmentation for the assessment of burn injuries is a new area of research and thereby the number of publications is currently scarce. For example, Jiao et al. (2019) implemented a model based on the Mask-RCNN architecture, using ResNet101 with atrous convolution as the backbone. For this purpose, they created their own annotated dataset composed of RGB pictures obtained with smartphones at the Wuhan Hospital; however, such dataset is not publicly available. Similarly, Liu et al. (2021) created a dataset from burn patients in a hospital in China using cameras and smartphones, and annotated the data with an image annotation tool known as LabelMe (Wada, 2016). This dataset was expanded by applying data augmentation operations, yet it is not publicly available either. For their model, they implemented an FCN with an encoder-decoder based architecture using different backbone networks. Lastly, Chauhan & Goyal (2020a) developed an architecture that leverages ResNet101 low-level features and atrous convolutions using a dataset created with online images. It was a binary problem in which the burn injury was one class, and the rest of the image was the other class (healthy tissue).

2.3 Medical Image Classification

Image classification is an essential task in the field of computer vision. It consists of assigning a label to an image based on the available categories for a specific problem. When applied to the medical domain, it is commonly known as medical image classification, a widely investigated area in deep learning (Lai and Deng, 2018). As in the segmentation problem, the inputs can be generated by different imaging modalities, including Ultrasound, CT, MRI, PET, and X-Rays; however, the main goal is to precisely determine whether the region of interest is affected by a particular disease or lesion. Consequently, it is one main component of any CAD system (Miranda et al., 2016).

It is possible to break down this problem into two main steps: (1) the extraction of meaningful features that describe the images, and (2) the classification of the medical images based on such features (Shen et al., 2017). Before the introduction of deep learning-based models, the traditional method for the step (1) – determining useful features - required a manual extraction and selection of the features by an expert. This approach had major drawbacks as it is extremely time-consuming, relies on prior knowledge, and introduces inter-subject variability. In a way, it is more an art than a science. Unsurprisingly, the results of the classification step following this method were not satisfactory, and thus, not suitable for the implementation of CAD systems (Yadav and Jadhav, 2019).

However, over the last decade, deep learning-based models enabled the implementation of end-to-end architectures that output the corresponding label using the raw pixels of the medical image exclusively. In this regard, the two main steps of the classification problem are combined into a single network. The convolutional layers of the model can automatically extract high-level features that are used as inputs of the fully connected layers to complete the classification task (W. Wang et al., 2020). As a result, high accuracies can be obtained which can even outperform human diagnosis in different medical applications (Razzak et al., 2018).

Numerous supervised and unsupervised deep learning-based models have been studied to diagnose a variety of diseases and lesions. Diabetic retinopathy detection, classification of breast tumors, lung cancer detection, and Alzheimer's early detection are among the most

common targets (Cai, Long, et al., 2020). For instance, Qummar et al. (2019) proposed a stacked CNN that combined five common pretrained CNN architectures to classify color fundus images, from a publicly available dataset, according to the different diabetic retinopathy stages to enhance the early detection of this disease. For breast cancer, Fujioka et al. (2019) evaluated the performance of GoogleNet Inception v2 (Szegedy et al., 2015) to classify breast masses into benign and malignant using ultrasound breast images and achieved high diagnostic accuracy. Additionally, Ding et al. (2019) leveraged fluorine 18 fluorodeoxyglucose PET scans of the brain and GoogleNet Inception v3 (Szegedy et al., 2016) pretrained on ImageNet (Deng et al., 2009) for Alzheimer's disease accurate prediction.

2.3.1 Deep Learning-based Burn Depth Classification

In the initial assessment and management of burn injuries, it is crucial to establish the burn depth accurately. It is a decisive factor for the selection of the wound care treatment, the healing process, and the burn patient's outcomes. However, this is a challenge even for experienced burn surgeons (Gomez and Cancio, 2007). The reason for it is that there is a high variability in the appearance of burn injuries, and in some cases, the damage caused by them is located in inner skin layers that are not visible from outside. In this regard, and considering the rapid advances in medical imaging technologies, different imaging modalities have been explored to obtain complementary information that can support the burn depth assessment and improve the accuracy of the task (Sen et al., 2016). These modalities include color photographs, ultrasound, LDI, infrared thermography (IR), laser speckle imaging (LSI), optical coherence tomography (OCT), and orthogonal polarization spectral (OPS) imaging (Thatcher et al., 2016).

In the medical literature, there are recent studies that combine deep learning-based models and machine learning models with a particular imaging modality, and evaluate the results obtained for the burn depth classification task. Cirillo et al (2019) created a labeled dataset of RGB burn images using a high-performance digital camera to reduce the effect of the variable lighting. They compared the performance of VGG16, GoogleNet, ResNet50, and ResNet101, pretrained on ImageNet, to classify the burn injury into the different categories.

A similar approach was followed by Chauhan & Goyal (2020b), but introducing an extra step to boost the accuracy. They implemented a body part classification model that allowed them to carry out the burn depth classification based on the specific characteristics of the body region in which the injury was located. However, the use of RGB images can negatively impact the accuracy due to factors like lighting conditions and skin color. Besides color images, additional imaging modalities have been studied using a traditional machine learning approach. In this regard, OCT (Singla et al., 2018), spatial frequency-domain imaging (Rowland et al., 2019), and ultrasound (Lee et al., 2020) have been utilized for feature extraction, combined with Support Vector Machine (SVM) models for the burn depth classification task.

2.4 Transfer Learning

Deep learning-based models are commonly known for being data-hungry, i.e., they require large datasets, in the scale of tens of thousands, to achieve satisfactory results. However, for some domains or specific applications, the creation of such datasets is very challenging due to the lack of expertise, time, and cost that this task implies (Weiss et al., 2016). This is the case with medical images, in which, typically, there are small datasets that are not sufficient to effectively train deep learning models for segmentation or classification problems (Shen et al., 2017).

To overcome this limitation, we leverage transfer learning. Transfer learning can be defined as a technique that applies the knowledge obtained from one domain with a great number of training samples to a target domain with limited training data (Tan et al., 2018). In this regard, the weights obtained after training a model with the original domain data can be reused following two different approaches. The first one is to use the pretrained model as a feature extractor. For this, all the layers of the model are frozen except for the last fully connected layer, which is replaced by a new fully connected layer specifically suited for the target domain. Then, the small dataset is used to train the added layer. The second one is known as fine-tuning. In this approach, the weights of the pretrained model are used as initialization, and the small dataset is used to retrain the whole model. However, it is also

possible to freeze the shallow layers. As in the first approach, the last fully connected layer is replaced with one suited for the target domain (F.-F. Li, 2021).

Typically, in computer vision, the dataset used to initially train the model is ImageNet, a massive dataset of natural images with 1000 categories (Deng et al., 2009). Even though ImageNet and medical image datasets are unrelated, it has been proven that CNNs that have been pretrained with this dataset are effective and lead to satisfactory results in medical applications. This applies to both transfer learning approaches, feature extraction and fine-tunning. In this regard, the use of ImageNet pretrained Inception-ResNet v2 architecture as a feature extractor was explored by Byra et al. (2018). They assessed the level of steatosis using CNN-based features obtained from B-mode ultrasound images of the liver. Similarly, da Nobrega et al. (2018) evaluated the performance of different CNN models pretrained on ImageNet as feature extractors. They obtained high accuracies for differentiating benign lung nodules from malignant lung nodules using lung CT images as inputs. Recently, Ahuja et al. (2021) leveraged transfer learning and data augmentation for COVID-19 detection in chest CT scans using the ResNet18 architecture.

2.4.1 Pretrained Deep CNN Architectures for Image Classification

Since the introduction of AlexNet (Krizhevsky et al., 2012) and its remarkable success in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), several deep CNN architectures have been proposed. Among the most used for image classification, VGG (Simonyan and Zisserman, 2014), Inception (Szegedy et al., 2015), ResNet (He et al., 2016), and DenseNet (G. Huang et al., 2017) are very common. They have been pretrained with ImageNet, and the weights are publicly available. In this thesis, different ResNet and VGG architectures are utilized.

VGG

Simple architecture that exploits the use of small-sized convolution filters (3×3) stacked together (Simonyan and Zisserman, 2014). Additionally, it has max-pooling layers and fully

connected layers. Besides obtaining satisfactory results in image classification tasks, it is also effective for semantic image segmentation tasks, being used as the backbone of several segmentation architectures. However, it can be computationally expensive due to the massive number of trainable parameters Khan et al., 2020.

ResNet

This architecture also exploits the use of small-sized convolution filters. It is constituted by several convolutional layers and one fully connected layer at the end. However, it introduced the use of residual connections, which enabled the creation of deeper CNNs without the vanishing gradients problem. These shortcuts connect the input of each basic block to its output (See Figure 2.2), facilitating the training of the model (Brodzicki et al., 2020).

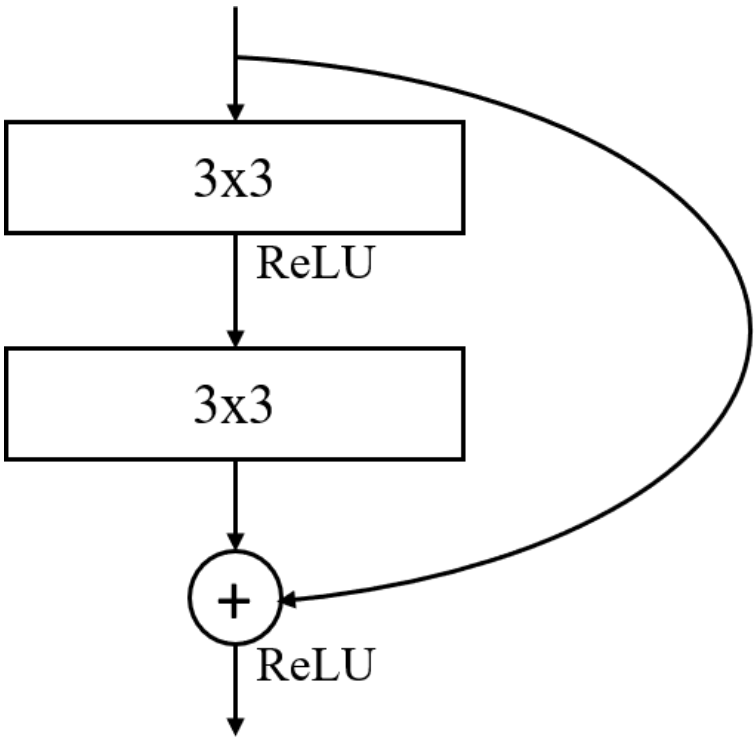


Figure 2.2. Residual block of ResNet34 architecture

2.5 Ultrasonography in Wound Care

Ultrasound imaging is a noninvasive, versatile, portable, low-risk, and cost-effective imaging modality widely used in the medical domain. It provides a real-time internal view of the body that is beneficial for diagnosing different diseases and conditions (Alfred et al., 2020). In medical ultrasound, the B-mode is the preferred modality. It leverages sound waves (1-20MHz) and the acoustic impedance of tissues to generate 2D grayscale images of the ROI (Chan and Perlas, 2011). In this regard, it can be used for several applications in different organs and tissues in the human body. For example, screening of aortic aneurysm in the abdomen (Sisó-Almirall et al., 2017), detection of foreign bodies in the patient's anatomy (Tantray et al., 2018), and diagnosis of ocular diseases (Gottlieb et al., 2019).

Because skin and underlying tissues can be characterized using B-mode ultrasound, such modality can be used to identify and quantify the changes caused by a wound. Kalus et al. (1979) explored the application of this imaging modality for burn depth assessment in two burn patients. Even though they used a modified conventional ultrasound machine, they concluded that it was an acceptable approach. Another early study was carried out by Bauer & Sauer (1989) using animal models. They concluded that the use of B-mode ultrasound images was only relevant and accurate during the first 6 hours after the burn injury occurred.

However, during the last decade, the rapid progress in technology has introduced significant improvements in medical ultrasound imaging. As a result, it is possible to obtain high-resolution images and more accurate diagnoses. Recently, Professor Chandan K. Sen and his associates at Indiana University School of Medicine (ISUM) explored the use of B-mode HUSD for assessing the severity of burn injuries and monitoring the healing process. They corroborated the results obtained with histological analysis, demonstrating that B-mode HUSD is an effective and advantageous imaging modality to assess burn depth (Gnyawali et al., 2015b; Sen et al., 2016). Lastly, Lee et al. (2020) took advantage of B-mode ultrasound combined with traditional machine learning techniques to classify different burn depths based on texture features, obtaining satisfactory results. However, they used ex vivo porcine tissue for their study, which significantly simplifies the problem.

2.6 Image Texture Analysis

Texture is one of the most important features in image processing and computer vision. It can be described by the spatial arrangement of intensity levels of the pixels in a specific region of the image (Bharati et al., 2004). In this regard, it can measure properties, such as smoothness, coarseness, frequency, uniformity, regularity, and density. Therefore, texture analysis can be used to segment different regions of an image and classify the obtained regions. To achieve this, extracting characteristic texture features using one of the following approaches: structural analysis, statistical analysis, model-based analysis, and transform-based analysis (Armi and Fekri-Ershad, 2019) is often required.

The structural approach assumes that the image is constituted by a combination of patterns and provides a high-level symbolic description, yet understandable, of the image. For this, the definition of primitives, which are repeated elements that constitute the image, and spatial arrangements of the selected primitives are required. However, given that this approach is more useful for simple images in which patterns can be easily extracted, it is not commonly used (Ţălu, 2012; Wechsler, 1980). In contrast, statistical methods are the most used in texture analysis. They leverage the spatial distribution of grey levels to extract texture features. Depending on the number of pixels used in the calculation of these features, the extraction methods can be classified into first, second, and higher order statistics (Materka, 2004). The first-order uses only one pixel, and the features are extracted from the histogram. Some examples are average, variance, skewness, median, and kurtosis (Galavis et al., 2010). The second-order uses the interaction between a pair of pixels at a determined distance and orientation from one to the other. The most important features are usually extracted from the Gray Level Co-occurrence Matrix (GLCM) that was proposed by Haralick et al. (1973). The higher-order uses more than two pixels to extract the features, but they are not widely used. Lastly, the model-based approaches exploit stochastic models, while the transform-based approaches leverage the Fourier, Gabor, and Wavelet transforms (Castellano et al., 2004).

In medical image analysis, statistical methods are widely used for detection, segmentation, and diagnosis tasks. They have been applied to different imaging modalities, and have

provided accurate results when combined with AI. In dermatology, they have been widely used to detect melanomas and psoriasis (Kolkur and Kalbande, 2016). Besides dermatology, they have been used to detect cancer or malignancies in histological samples extracted from different parts of the body like the brain, neck, and cervix (Fujima et al., 2019; L. Wei et al., 2017). Additionally, they have been adopted for the segmentation of brain tumors in MRI images (Usha and Perumal, 2019). Regarding CT scans, statistical methods have been used to detect lung and liver cancer and classify brain hemorrhages (Jony et al., 2019; Nawresh and Sasikala, 2020. Related to this thesis, second-order texture features have been used in B-mode ultrasound images. Some of the most common applications are in the detection and assessment of different types of cancer, including the most common ones, breast, prostate, and thyroid cancer (Faust et al., 2018). Such texture features have also been used to classify various liver diseases Xu et al., 2019. Finally, as mentioned in the previous section, Lee et al. (2020) used texture features with a traditional machine learning approach to classify burn injuries. Our work will explore the use of deep learning and texture analysis, by combining CNN-based features with texture features in order to determine the burn depth based on B-mode HUSD images.

2.7 Summary

This chapter presented state-of-the-art research in topics related to the development of CAD systems based on medical images. A review of AI approaches in semantic image segmentation and image classification was presented, with a particular focus on the medical domain and their applicability to the assessment of burn injuries. Additionally, transfer learning, ultrasonography, and image texture analysis techniques were discussed in the context of burn injury diagnosis and related clinical areas.

3. METHODOLOGY

In this section, the methods used for the automatic assessments of burn injuries are discussed. All of them belong to one of the two main problems tackled in this thesis. The first one is the semantic segmentation problem, which is illustrated in Figure 3.1. The second one is the burn depth classification problem, which is summarized in Figure 3.5. In each of these figures, four major components are identified. Three of them are primary components of any AI-related application: dataset organization, model training, and performance evaluation. Therefore, both problems share these three components. However, each of them is customized for the designated task. For example, the evaluation metrics used for the segmentation problem are applied in a pixel-wise manner. In contrast, the ones used for the classification problem are applied in an image-level manner. In the following subsections, a fully comprehensive description of every method and approach used in this work, including implementation details, is provided.

3.1 Automatic Burn Injury Segmentation

The measure of the body surface area affected by a burn wound is a critical element in the assessment and treatment of this complex injury. This body surface area is often overestimated by practitioners, and furthermore, its estimation is negatively affected by the presence of inter-subject variability. Automatic calculation of the burn area is an alternative approach that has the potential to overcome these limitations. In this thesis, as a first step towards this goal, the use of deep learning-based semantic segmentation models with RGB images for burn wound segmentation is explored.

In addition to the identification of the burn injury in the image, it is necessary to identify the pixels that belong to healthy skin, for accurate clinical diagnosis. More specifically, this process can be described as follows: let X be a burn injury image N pixels, $X = \{x_{\mathrm{i}}\}_{\mathrm{i}\epsilon N}$, where each pixel needs to be assigned one label from the following set, L=$\{$*Burn wound, Healthy Skin, Background*$\}$. The goal is to come up with the best predictor that given an arbitrary pixel can infer the correct label. To this end, in this section, the construction process of our dataset, which is suitable for this task, and the utilized data augmentation

strategy to assist in the process are described. Also, the implemented network architecture for the semantic segmentation task, with the corresponding training specifications, is presented. Lastly, the evaluation metrics for the assessment of the model performance are described. This is summarized in the pipeline illustrated in Figure 3.1

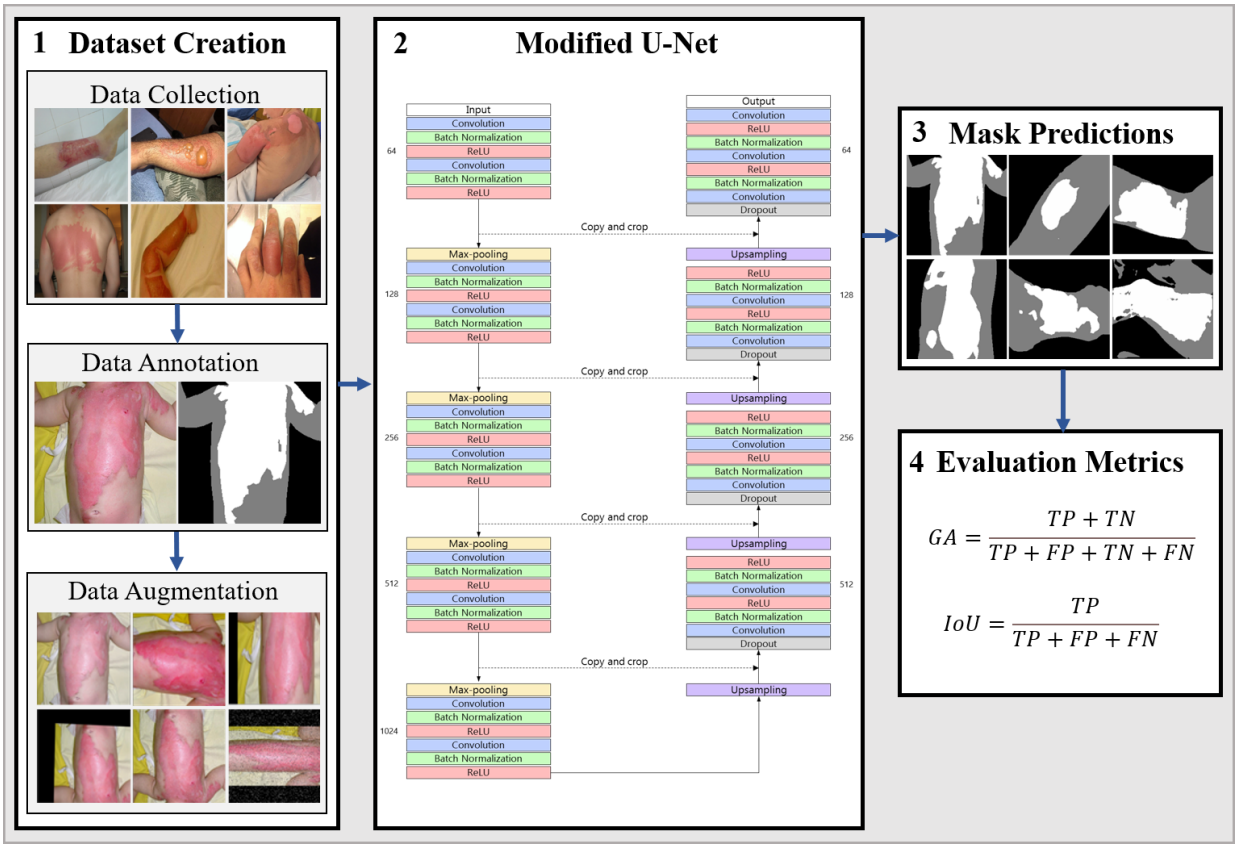


Figure 3.1. General pipeline of the semantic segmentation problem.

3.1.1 Dataset Creation

Currently, there is no publicly available dataset containing RGB burn images with their corresponding pixel-wise annotations for the semantic segmentation task. Therefore, in this work, the Google search engine was used to create our dataset by collecting RGB images available online. The keywords used were burn, burn injury, burn wound, scald, partial-thickness burn, full-thickness-burn, second degree burn, and third degree burn. Based on the clinical assessment guidelines for burn injuries, the only inclusion criteria was the depth

of the burn. In this regard, only images containing partial-thickness and full-thickness burn injuries were included since they are the ones considered for the TBSA estimation (Jeschke et al., 2020). As a result, a dataset of 390 images that were resized to 256×256 pixels each was obtained. This value was selected based on the resolution of the collected images and because it has shown satisfactory results for semantic segmentation problems. Additionally, the dataset was randomly divided into 270 images for training, 60 images for validation, and 60 images for testing.

As it is shown in Figure 3.2, the images collected were taken under very different conditions and formats. They were obtained using different devices, e.g., digital cameras and smartphones, in different lighting conditions, and from various distances and angles. This can be seen as an advantage, as the variability can help to achieve better generalizations, boost the robustness of the model, and help to make more accurate predictions under a variety of conditions.

Figure 3.2. Examples of images included in the training dataset.

3.1.1.1 Data Annotation

In order to follow a supervised learning approach, corresponding pixel-wise annotations for the collected data were required. For this, the user-friendly online platform, SuperAnnotate ("SuperAnnotate", 2020), which allows to assign the corresponding label to each pixel in the image, was utilized. After completing the annotation process, the generated color masks were exported and converted to greyscale masks in the Portable Network Graphics (PNG) format. The three semantic classes, burn wound, healthy skin, and background, are depicted in Figure 3.3. In the grayscale mask, they can be identified as the white, gray, and black regions, respectively.

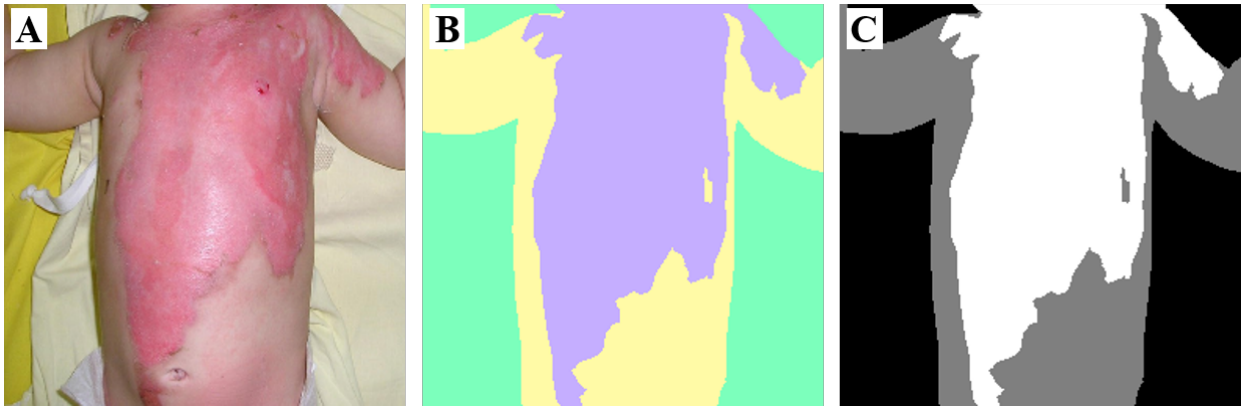


Figure 3.3. Sample of annotated training image. (a) RGB image, (b) Generated color mask, and (c) Grayscale mask.

3.1.1.2 Data Augmentation

Data augmentation is a technique widely used to increase the number of observations from small datasets acting as sources, such as the one previously described. It consists on the application of different transformations to the original images to create additional samples of the training dataset and avoid overfitting. We exploited this technique and created a data augmentation pipeline with several transformations using the python library imgaug (Jung et al., 2020). In this regard, from 3 to 5 of the selected transformations were randomly applied to some images, and the corresponding masks were modified accordingly. The trans-

formations are vertical flipping, horizontal flipping, multiplication of the pixels by a random value, addition of a random value, addition of Gaussian noise, addition of salt and pepper noise, application of Gaussian blur, average blurring, median blurring, multiplication of the saturation, gamma correction, logarithmic contrast adjustment, dropping pixels, affine transformation, perspective transformation, rotation, and cropping and padding. From the ones that apply a similar operation to the image, such as blurring, only one of the transformations could be selected for each image. After the data augmentation was completed, the size of the training dataset increased to 810 samples.

3.1.2 Network architecture

The semantic segmentation network is a slightly modified U-Net architecture (Ronneberger et al., 2015), which will be referred to as regularized U-Net. The original model was explained in detail in subsection 2.2.1.1, and it is illustrated in Figure 2.1. It is constituted by a contracting path and an expansive path, which play the roles of encoder and decoder, respectively. Each path has a basic block that contains convolutional layers and ReLU activation functions. However, there is a key difference between the contracting path's basic block and the expansive path's basic block. The first one contains max-pooling operations, while the second one contains upsampling operations. Additionally, the network exploits skip connections that allow the expansive path to use the feature maps generated in the corresponding block of the contracting path.

Above, an overview of the original U-Net architecture was presented. In this work, minor modifications were introduced to enhance the segmentation results. One modification was the addition of a batch normalization layer in each block of the contracting and expansive paths. This resulted in a more efficient training process and improved the generalization capabilities of the model. The second modification was introduced to avoid overfitting and further boost the performance of the model in our small dataset. In this regard, dropout layers were added to the basic blocks of the expansive path as a regularization strategy. The modified architecture is shown in Figure 3.4.

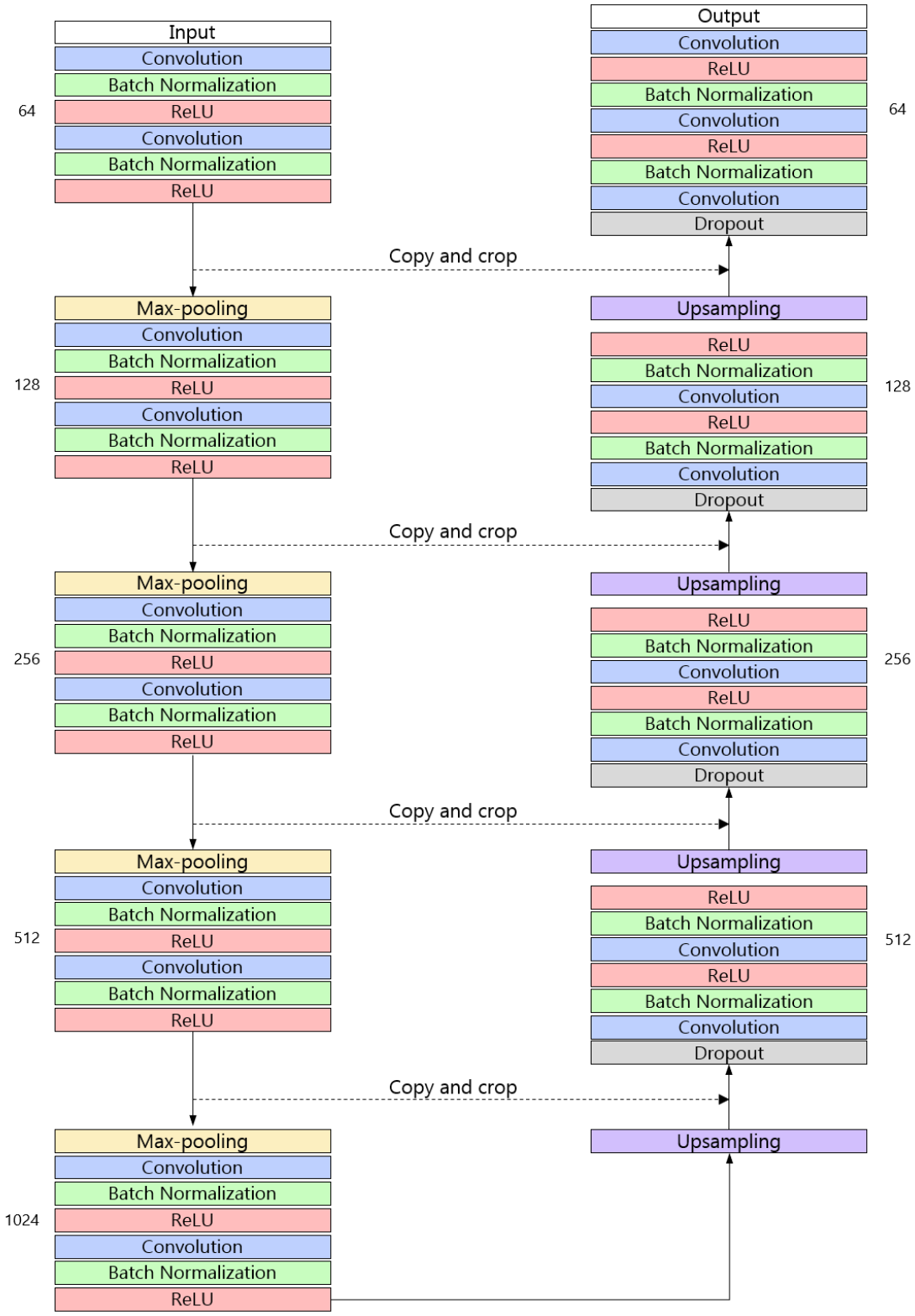


Figure 3.4. Modified U-Net architecture.

3.1.3 Training Specifications

Before starting the training process, images were pre-processed as follows: all the images used as inputs of the model were normalized in the interval $[0, 1]$. Then, the mean was subtracted from each pixel to center the data at a zero mean.

The model was implemented using the deep learning framework PyTorch, version 1.8.1 (Paszke et al., 2019). Additionally, we used Intel Core i7 7700HQ CPU with NVIDIA GeForce GTX 1050Ti GPU with 4GB to train the model in a Windows 10 operating system. All the weights were initialized randomly, i.e., the network was trained from scratch. The cross-entropy loss was selected as the loss function. It can be defined as follows

$$Loss = -\sum_{\mathrm{i}=1}^{N} y_{\mathrm{i}} \cdot log\hat{y}_{\mathrm{i}} \tag{3.1}$$

where N is the number of classes, in this case $N = 3$, y_{i} is the target, and $\hat{y}_{\mathrm{i}}$ is the output of the model. To minimize this loss function, the Adam optimizer, a widely known optimization algorithm, was used. Lastly, after experimenting with different values for the remaining hyperparameters, the ones used to train the model for 100 epochs are: $1e^{-}3$ initial learning rate, exponential learning rate decay with a 0.95 rate decay, dropout rate of 0.2, and batch size set to 4.

3.1.4 Evaluation Metrics

To quantitatively assess the segmentation performance of our model, two evaluation metrics were calculated for the samples of the testing dataset, which was specifically created for this problem. The selected standard metrics are global pixel accuracy and mean intersection over union. These are widely used to evaluate semantic segmentation models.

The global pixel accuracy (GA) metric can be defined as the percentage of pixels that were classified correctly. It is computed as the number of correct pixel predictions over the total number of pixels, as follows

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.2}$$

where TP makes reference to the true positives, TN to the true negatives, FP to the false positives, and FN to the false negatives.

The mean intersection over union (mIoU) measures the average of the intersection of the ground truth mask and the predicted mask over the union of both. Therefore, it is required to calculate first the IoU, also known as the Jaccard index (JI), of each semantic class and then find the average. The resulting value is contained in a $[0, 1]$ interval, and a higher value indicates a better model performance.

$$IoU = \frac{TP}{TP + FP + FN} \tag{3.3}$$

3.2 Burn Depth Classification

The initial care and selected treatment for burn patients rely heavily on the severity of the injury. At the same time, the success of the healing process and the achievement of satisfactory patient outcomes depend on the received treatment. In this regard, it is fundamental to have an accurate burn depth diagnosis; however, it remains to be a clinical challenge.

The burn depth classification task was formulated as a deep learning classification problem using medical images. Every image x_{i} is assigned a label y_{i}, which is selected based on the available categories. In this work, there are three categories: skin, partial-thickness injury, and full-thickness injury. These constitute the set L. In this section, the required steps to accomplish the proposed burn depth classification task are described. Similar to the

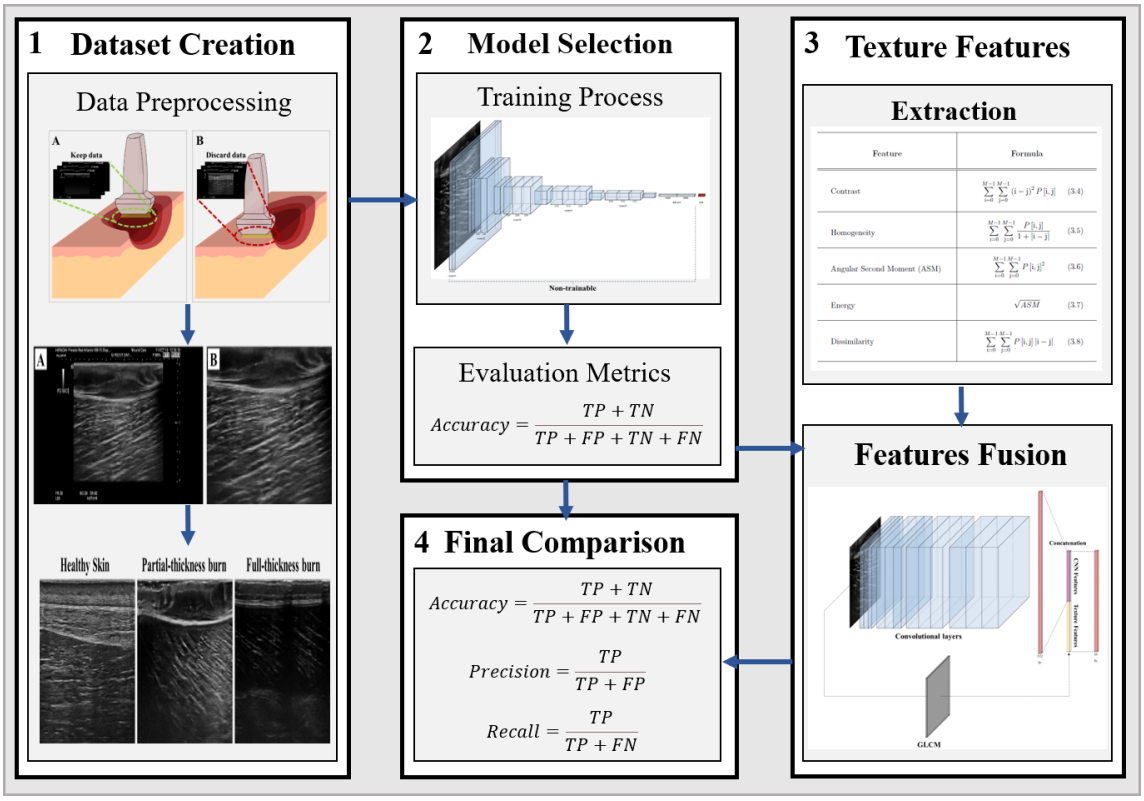


Figure 3.5. General pipeline of the burn depth classification problem.

previous section, we discuss the construction of the dataset and describe the model's training process. However, in this case, we compare the performance of different models, exploiting transfer learning, and we explore the use of hand-crafted features. The general pipeline of this classification task is depicted in Figure 3.5.

3.2.1 Dataset Creation

The dataset for the classification problem was created using B-mode HUSD images. This data was obtained from Dr. Chandan K. Sen and his associates at Indiana University School of Medicine. For this purpose, an electrically heated burning device was created to guarantee the repeatability and reproducibility of the generated injuries (Kim et al., 2016). Additionally, a porcine model was selected as the test subject, given its similarity to humans in terms of the skin wound healing process (Roy et al., 2014). More specifically, Yorkshire pigs, a breed of swine, were used for this study. The protocol developed and utilized by

Dr. Sen's group to generate the required burn injuries can be summarized in the following steps. First, the porcine subjects were sedated and anesthetized using a general inhalation anesthetic drug, and their dorsal region was shaved and prepared. Then, using the burning device mentioned above, eight burn injuries were created in a controlled and precise manner. The size of each wound was 2×2 inches, four of them were located at one side of the dorsal spine, and the other four were located at the other side. The created wounds were treated with Tegaderm (3M) held by V.A.C drape, and were then wrapped with Vetrap and Elastikon (3M). Lastly, B-mode HUSD data was collected with a linear transducer at 6.5MHz (Gnyawali et al., 2015b; Kim et al., 2016; Sen et al., 2016). The ultrasound device used in this process was the Hitachi Aloka Noblus device.

The collected data for the burn depth classification problem addressed in this thesis was provided according to the aforementioned step. This data is constituted by B-mode HUSD videos of healthy skin, partial-thickness injuries, and full-thickness injuries of 4 pig subjects. In addition, the burn depth was validated with a histological analysis of obtained biopsies by Dr. Chandan K. Sen and his team (Gnyawali et al., 2015b).

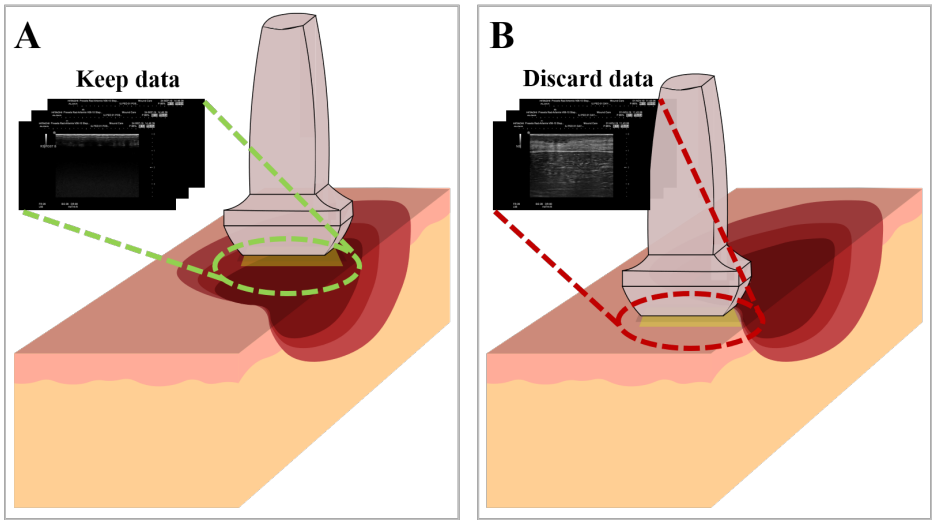


Figure 3.6. Video trim step. Video fragments corresponding to scenario (a) were used for the frame extraction process. Video fragments corresponding to scenario (b) were discarded.

3.2.1.1 Data Preprocessing

The dataset was created using frames extracted from the B-mode HUSD videos. For this, some preprocessing steps were required. The first step was exclusively applied to the videos corresponding to partial-thickness and full-thickness injuries. In this regard, before extracting the required frames, the videos were trimmed to remove the fragments that showed healthy skin or less severe burned tissue surrounding the wound (see Figure 3.6). This step guaranteed that the ground truth assigned to the frames that were utilized as model inputs was correct.

Then, frames were extracted from the trimmed videos, which had a frame rate of 20 frames per second (FPS). One problem is that consecutive frames contained the same information, i.e., there were minimal to no changes between them. For this reason, only two frames per second were extracted to minimize such redundancy. Last, the resulting frames were cropped to discard the black frame usually present in ultrasound data since it does not provide any valuable information for the classification task. This is depicted in Figure 3.7.

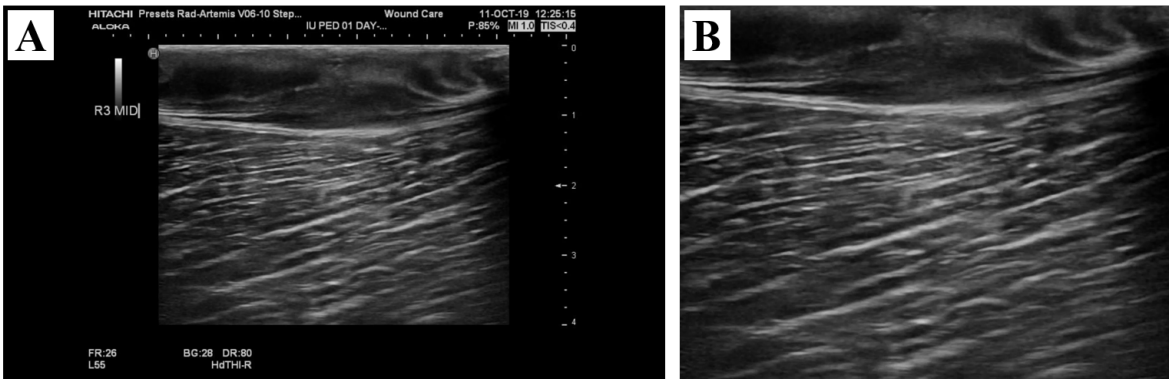


Figure 3.7. Black frame removal. (a) is the original extracted frame, and (b) is the cropped frame. The images in the dataset have the (b) format.

At the end of this process, a balanced dataset of 1,635 B-mode HUSD images was obtained. The three classes are healthy skin, partial-thickness injury, and full-thickness injury (Figure 3.8). In contrast to the data partition approach followed in the segmentation problem, a random partition could not be followed for the creation of the training, validation, and testing datasets for the burn depth classification problem. The reason is that multiple

samples belong to the same pig subject, and they can even belong to the same injury. In this regard, if a random partition was to be applied and one of the frames was assigned to the training dataset, and another frame extracted from the same injury was assigned to the testing dataset, an error would be introduced into the model. As a consequence, high accuracies would be obtained not reflecting the generalization capability of the algorithm. Therefore, the frames corresponding to the first three pigs were used for the training and validation datasets, while the frames corresponding to pig number four were used exclusively for the testing dataset.

Figure 3.8. Examples of each dataset class.

3.2.2 Training Deep Learning Models

Five different models were trained for the burn depth classification problem and their performance was compared in order to select the one that yields the best classification accuracy. In this regard, two widely used architectures were adopted: ResNet and VGG. From the first one, three models, ResNet18, ResNet50, and ResNet101, were selected. Similarly, from the second one, two models, VGG16 and VGG19, were selected. All of them have been previously trained on the ImageNet dataset, and the resulting weights are publicly available. In addition, PyTorch, the deep learning framework that was used for all the implementations, facilitates the use of these pretrained models directly.

Additionally, this thesis leverages transfer learning following a feature extraction approach. To this end, all the layers were frozen for the five models and the last fully connected

layer was replaced in all of them to fulfill the needs of this work. As a result, the number of output channels from this last layer changed from 1000 to 3. In this classification problem, there are three classes, while in the ImageNet dataset, there are 1000 classes. Using this approach, the selected models' performance as off-the-shelf feature extractors was compared.

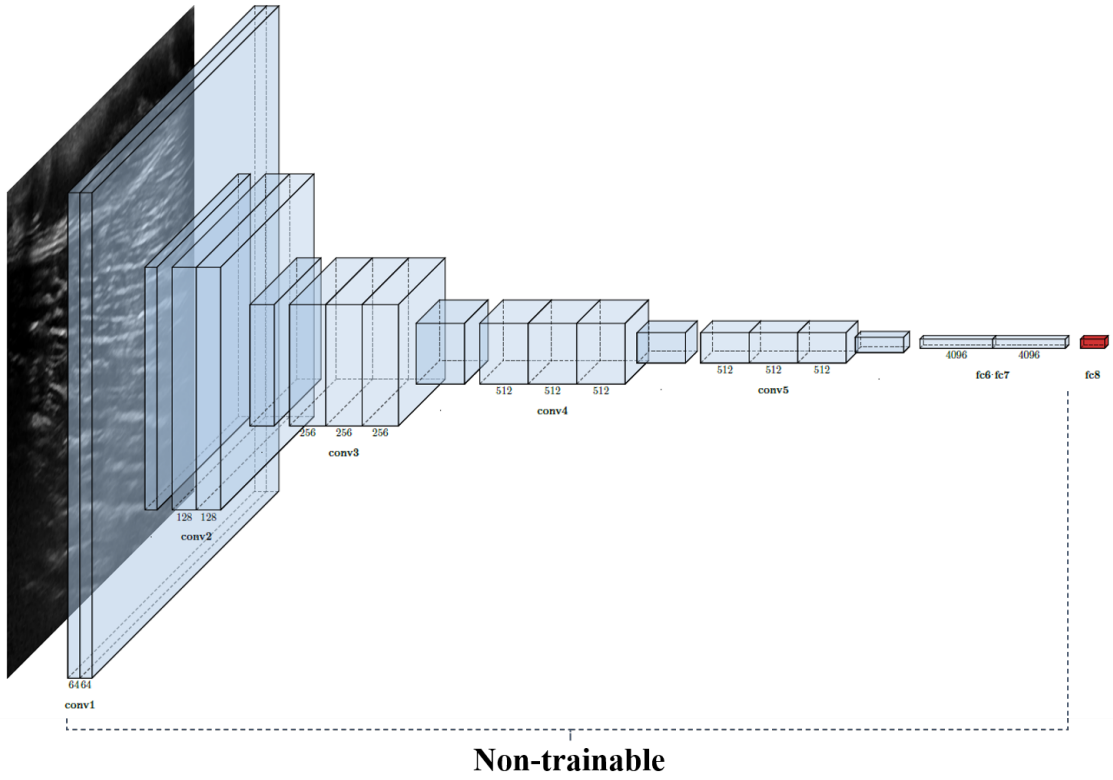


Figure 3.9. VGG16 architecture as feature extractor.

Figure 3.9 shows a VGG16 model used as an off-the-shelf feature extractor. The light blue layers contain the non-trainable weights and biases. Consequently, they were not updated during the training process. The only trainable weights were the ones that belong to the last fully connected layer of the model, depicted in red. Even though VGG16 is used as the example for this subsection, this technique was applied in the same way to the other CNN architectures, i.e., ResNet. As a result, it allowed to address the burn depth classification problem for the B-mode HUSD inputs based on the features extracted using the weights obtained with the ImageNet dataset.

For the training process, the same computer as the one described in subsection 3.1.3 was used. This is, a computer with a Windows 10 operating system, Intel Core i7 7700HQ CPU, and NVIDIA GeForce GTX 1050Ti GPU with 4GB. In addition, the cross-entropy loss was also selected as the loss function (Equation 3.1) with the Adam optimizer. Finally, the settings for the other hyperparameters were defined as follows: 150 epochs, $1e^{-}5$ learning rate with no decay, and batch size set to 8.

The evaluation metrics used to select the best model are described below. In this regard, the architecture with the best overall performance was ResNet. This refers to ResNet18, ResNet 50, and ResNet101.

3.2.3 Texture Features Extraction

The ultrasonography imaging modality, specifically the B-mode, generates grayscale images with different textures. The use of texture features has been studied to classify ultrasound images in a diagnostic context. In this thesis, the use of texture features combined with CNN-based features was explored for the burn depth classification problem. The hypothesis to be tested was that the inclusion of these hand-crafted features could improve the classification accuracy. To this end, the GLCM was computed, as well as the Haralick texture features that can be calculated from it.

Skimage, the image processing package for Python, was used. It enabled the obtention of the symmetric normalized GLCM. This is a squared matrix of dimensions $M \times M$, where M refers to the number of gray levels present in the image. Specifically for our case, $M = 256$. The matrix $P\,[\mathrm{i}, \mathrm{j}]$ is obtained based on the interaction of a pair of pixels with gray levels i and j, at a determined distance d and angle θ from each other. In our work, $d = \{1, 2\}$, $\theta = \{0, 45, 90\}$, and the Haralick features selected are defined in Table 3.1.

There were two values for d, three values for θ, and five Haralick texture features. Therefore, in total, 30 texture features were extracted and subsequently fused with the CNN-based features following the approach explained in the following subsection.

Table 3.1. Haralick texture features definition.

Feature	Formula
Contrast	$\sum_{\mathrm{i}=0}^{M-1}\sum_{\mathrm{j}=0}^{M-1}(\mathrm{i}-\mathrm{j})^2 P[\mathrm{i},\mathrm{j}]$ (3.4)
Homogeneity	$\sum_{\mathrm{i}=0}^{M-1}\sum_{\mathrm{j}=0}^{M-1}\frac{P[\mathrm{i},\mathrm{j}]}{1+\lvert\mathrm{i}-\mathrm{j}\rvert}$ (3.5)
Angular Second Moment (ASM)	$\sum_{\mathrm{i}=0}^{M-1}\sum_{\mathrm{j}=0}^{M-1}P[\mathrm{i},\mathrm{j}]^2$ (3.6)
Energy	$\sqrt{ASM}$ (3.7)
Dissimilarity	$\sum_{\mathrm{i}=0}^{M-1}\sum_{\mathrm{j}=0}^{M-1}P[\mathrm{i},\mathrm{j}]\lvert\mathrm{i}-\mathrm{j}\rvert$ (3.8)

3.2.4 CNN-based Features and Texture Features Fusion

The texture features were treated as additional features for the proposed hybrid model based on ResNet18 and the selected fusion approach. They were extracted from the GLCM matrix, and then they were concatenated with the CNN-based features extracted by the convolutional layers of the model. Typically, ResNet18 extracts 512 features that the fully connected layer uses to make the corresponding predictions. However, in the proposed model, this last fully connected layer was modified to take these 512 features and output a feature vector of dimension 30, $F_{CNN} = \{c_1, c_2, c_3, ..., c_{30}\}$. This vector was then concatenated with the 30 texture features extracted from the GLCM matrix for each B-mode HUSD image input, $F_T = \{t_1, t_2, t_3, ..., t_{30}\}$. This guaranteed an equivalent participation for each kind of

features, i.e., the CNN-based features and the texture features, in the final hybrid feature vector of dimension 60, $F_H = \{F_{CNN}, F_T\}$. The additional enhancements and the fusion approach are illustrated in Figure 3.10. Last, a fully connected layer was added at the end of the architecture. The newly created vector is the input of this fully connected layer, which serves as the classification layer with three output channels.

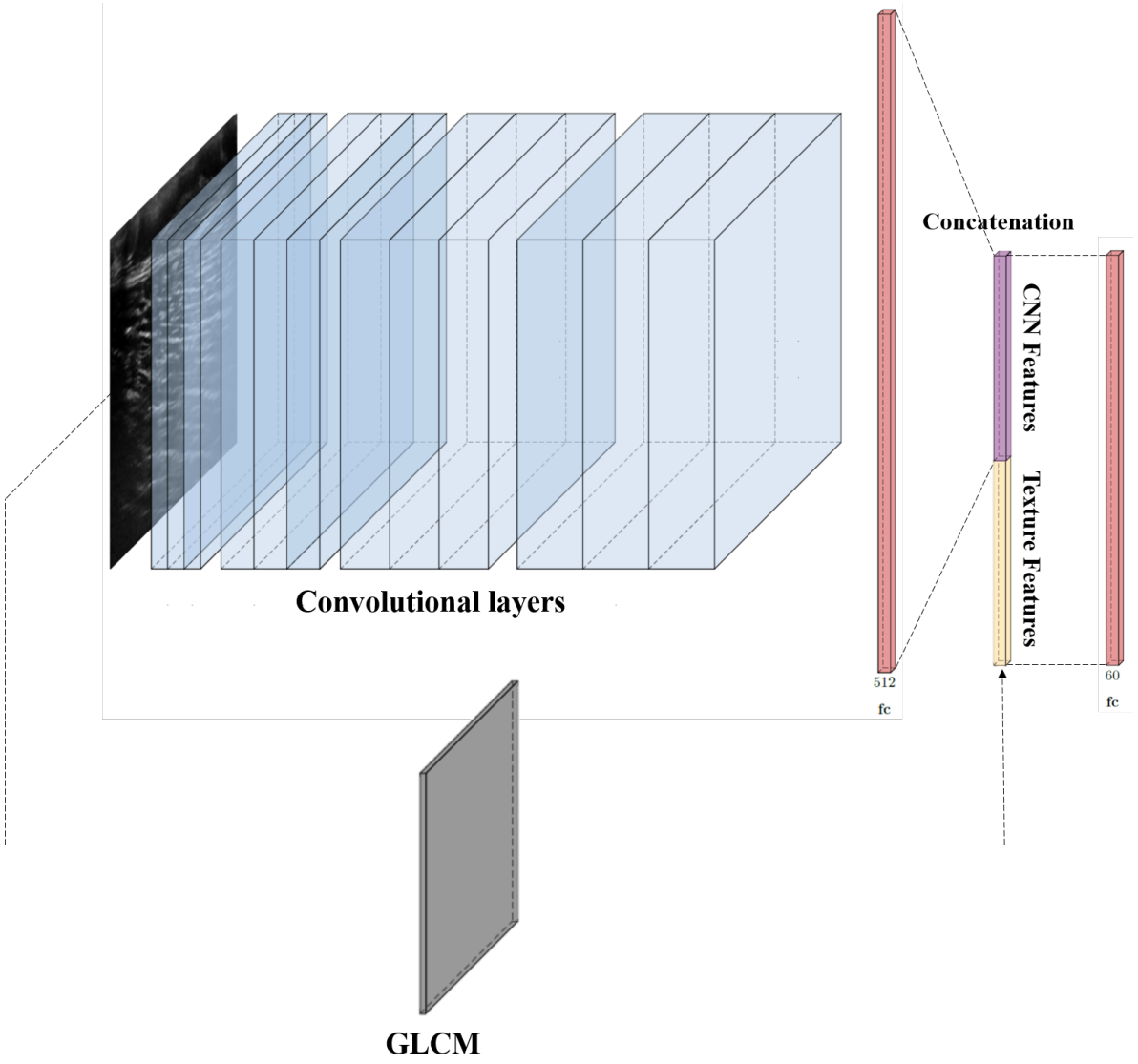


Figure 3.10. Features fusion strategy.

The training process is identical to the one described for the deep learning models, i.e., ResNet and VGG, before adding the texture features. This includes cross-entropy loss, Adam optimizer, 150 epochs, 1e^{-}5 learning rate with no decay, and batch size set to 8.

3.2.5 Evaluation Metrics

To evaluate and compare the performance of the classification models, four evaluation metrics were selected: classification accuracy, precision, recall, and F1-score. The first one was described above, and the formula is defined by Equation 3.2. The second one can be defined as the ability of the model to classify as positive, given that they are actually positive.

$$Precision = \frac{TP}{TP + FP} \tag{3.9}$$

The third one is the ability of the model to recognize the positives among all the possible positives. It is also known as sensitivity.

$$Recall = \frac{TP}{TP + FN} \tag{3.10}$$

And the last one is a combination between precision and recall.

$$F1 - Score = \frac{2 \times precision \times recall}{precision + recall} \tag{3.11}$$

3.3 Summary

This chapter discussed in detail the approach that was proposed and implemented for the assessment of burn injuries. For the two mentioned problems, the pipeline for the creation of each dataset to the implementation and training of the multiple models, to the selected evaluation metrics were presented. In this regard, for the semantic segmentation problem, a

labeled dataset composed of RGB images was created, and a regularized U-Net architecture was proposed. Similarly, for the burn depth classification problem, a dataset containing B-mode HUSD images was created and used to train multiple deep learning models. Also, an approach to incorporate texture features into the classification pipeline and evaluate their effect was proposed. In the next chapter, the experimental results obtained with the described methodology are discussed.

4. EXPERIMENTS AND RESULTS

This chapter presents the experiments conducted to evaluate the proposed approach for the burn injury assessment problem. The first experiment focused on the semantic segmentation challenge in this thesis. The objective of the experiment was to evaluate the performance of the regularized U-Net architecture based on the global accuracy and the mean IoU. The second and third experiments were focused on the burn depth classification challenge in this work. The objective of the second experiment was to compare the classification performance of different widely used deep learning models that leveraged transfer learning. Based on the results obtained, the best model was selected and used for the last experiment. In this last experiment, the effects of introducing hand-crafted features obtained from the B-mode ultrasound inputs for the classification task were explored. A detailed description of each experiment and the obtained results are provided in the following subsections.

4.1 Burn Injury Segmentation Task Evaluation

The main objective of this experiment was to evaluate the segmentation performance of the proposed model. For this, the training process of the model using the training and validation portions of the previously collected dataset was first assessed. They contain 270 and 60 images, respectively, partial-thickness and full-thickness burn injuries, before the data augmentation step. All the images in such dataset are RGB images of size 256×256, with their corresponding annotated masks. Both, the image and the mask, are PNG images. Additionally, data augmentation techniques were used to increase the number of samples in the training dataset. Then, the performance of the model was evaluated using the testing portion of the dataset, which remained unseen by the model until this last step. Similar to the validation dataset, it contains 60 RGB images of size $256 \times x256$, and their corresponding annotated masks.

To achieve a quantitative evaluation, the results obtained with our model and the ones obtained with the original U-Net architecture were compared. In this regard, this original architecture was implemented exactly as it is described by Ronneberger et al. (2015) in their paper *U-Net: Convolutional Networks for Biomedical Image Segmentation*. It was imple-

mented using PyTorch, version 1.8.1, and it was trained following the same hyperparameter values explained in section 3.1.3, which are summarized in Table 4.1. In addition, a Windows 10 operating system with Intel Core i7 7700HQ CPU with NVIDIA GeForce GTX 1050Ti GPU with 4GB was also utilized.

Table 4.1. Hyperparameter specifications for training the modified and the original U-Net architectures. The hyperparameters marked with an asterisk (*) apply exclusively for the modified U-Net.

Hyperparameter	**Value**
Optimizer	Adam
Epochs	100
Initial learning rate	0.001
Learning rate decay	Exponential
Decay rate	0.95
Decay Step	1
Batch size	4
Dropout rate*	0.2

The training process of deep learning models can display minor variations due to uncontrollable factors. These can have an effect on the evaluation metrics computed for the testing phase of the experiments. In this regard, to achieve a robust evaluation and extract significant inferences, each model of this experiment was trained for six times. This value was selected after running a power analysis using R-Studio to guarantee the obtention of a 90% power. R-Studio is an integrated development environment (IDE) for R, which is a programming language and software for statistical applications. Consequently, using a 0.9

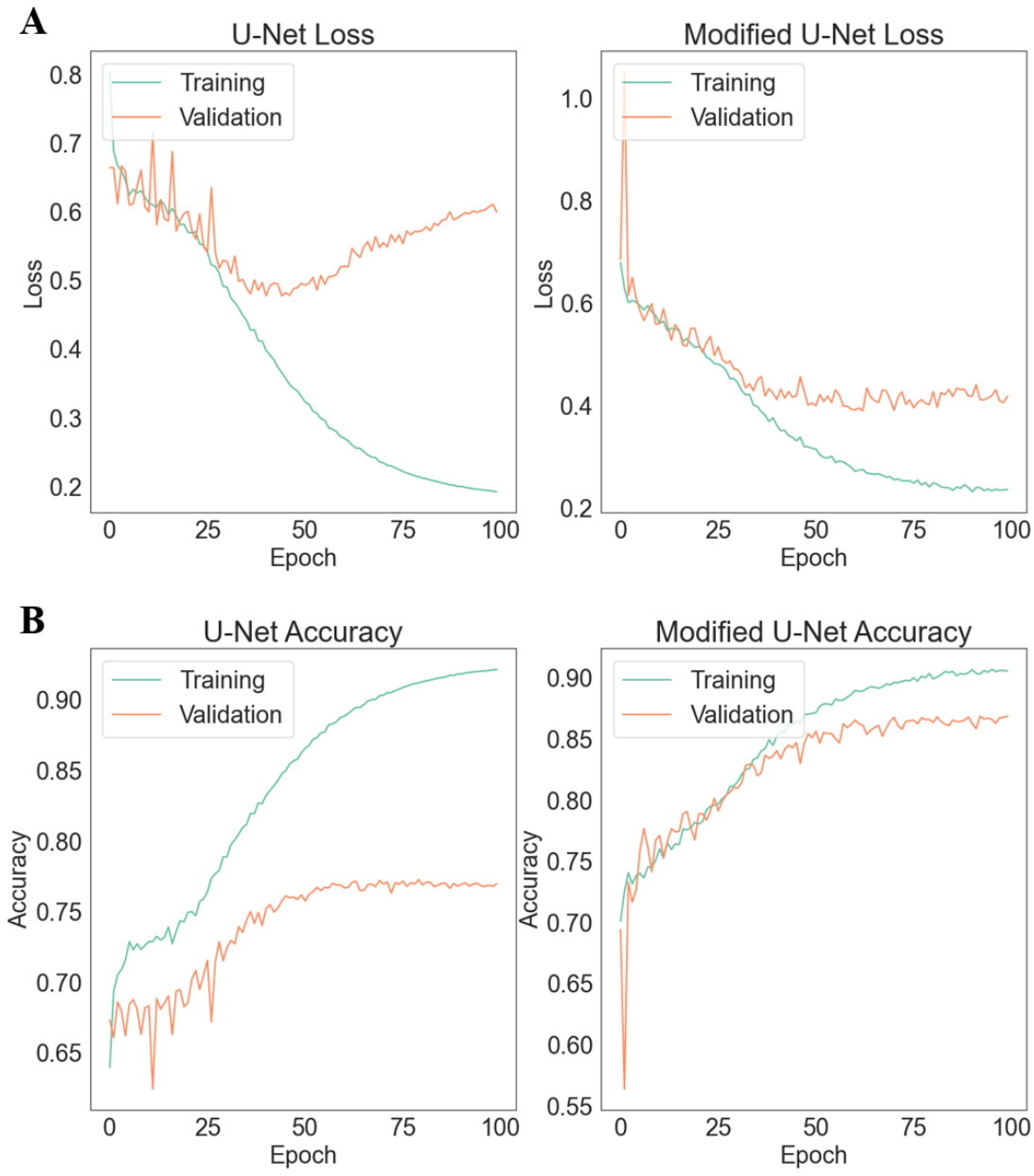


Figure 4.1. Training results of the original U-Net architecture and the modified U-Net architecture. (a) Loss for training and validation datasets. (b) Accuracy for training and validation datasets.

power, there is a 90% probability of detecting the differences between the original U-Net model and the regularized U-Net for the semantic segmentation task.

In the following subsections, the results obtained for both phases of this experiment, training and testing, including the selected evaluation metrics, and the corresponding statistical analysis are presented.

4.1.1 Models Training Process Assessment

In this subsection, the training process of the original U-Net architecture and the regularized U-Net architecture proposed for the semantic segmentation task are compared. Figure 4.1 shows the loss and the accuracy for the training and validation datasets for 100 epochs. This is taken from one of the six training iterations to provide a visualization of the process. As it can be observed, in the original U-Net loss graph, the model is suffering from overfitting. Approximately, after epoch 25, the training loss kept decreasing at the same rate as previous epochs. In contrast, the validation loss experienced a reduction in the decreasing rate. Then, right before epoch 50, it started to increase, demonstrating the poor generalization capabilities of the model for the created dataset. While data augmentation transformations were used to increase the size and diversity of this created dataset, such behavior was expected since it remained a very small-sized dataset. It is important to emphasize that these models were trained from scratch. After the data augmentation step there were a total of 810 samples. The original U-Net architecture was introduced to achieve satisfactory results in small datasets. However, it did not use any regularization strategy to enhance this feature.

In contrast, for the regularized U-Net, the reduction of the gap between the training loss and validation loss was notorious. This indicates that the introduction of the batch normalization layers and the dropout layers had a positive effect in the model's performance, addressing the overfitting concern. The validation loss reached a plateau after epoch 50, while keeping a small gap between the losses. In addition, in both accuracy graphs (see Figure 4.1 (b)), the difference in the results was evident. The training accuracy had almost an identical behavior, and reached a similar accuracy for both models, over 0.9. However, this is not the case for the validation accuracy. For the original U-Net model it reached a

plateau approximately at a 0.75 accuracy, 0.1 below the accuracy reached by the regularized U-Net. This suggests that the proposed architecture is a more robust model for the semantic segmentation task in unseen data, which is discussed in the following subsection.

4.1.2 Semantic Segmentation Experimental Results

To assess the performance of both models for the testing dataset, two evaluation metrics were computed: global accuracy, and mean IoU. These were calculated for each iteration of the model, i.e., there were six results for each metric. In Table 4.2 the average and the standard deviation (std) of the obtained values for the semantic segmentation task are presented.

Table 4.2. Semantic segmentation evaluation metrics.

	Original U-Net		**Modified U-Net**	
	Average	**Std**	**Average**	**Std**
Global Accuracy	0.783	0.033	**0.872**	0.008
Mean IoU	0.664	0.041	**0.784**	0.012

For the overall accuracy, there is an increase of 0.09 from the original U-Net to the regularized U-Net. Following this same trend, the mean IoU experimented an increment of 0.12. This indicates that the modified architecture has a better overall performance for this particular dataset. The obtained results were expected given the behavior of the models during the training process. In this regard, the validation dataset accuracy of the regularized U-Net reached a higher value than the original U-Net, and the loss decreased constantly until reaching the plateau, exhibiting a better learning behavior than the original U-Net. Lastly, some qualitative results obtained with the proposed model are displayed in Figure 4.2. These particular samples were selected to show the robustness of the model to variable burn depths, lighting conditions, body parts, distances, and angles. Consequently, the capabilities of the model to successfully segment the burn area and the skin from the background were demonstrated.

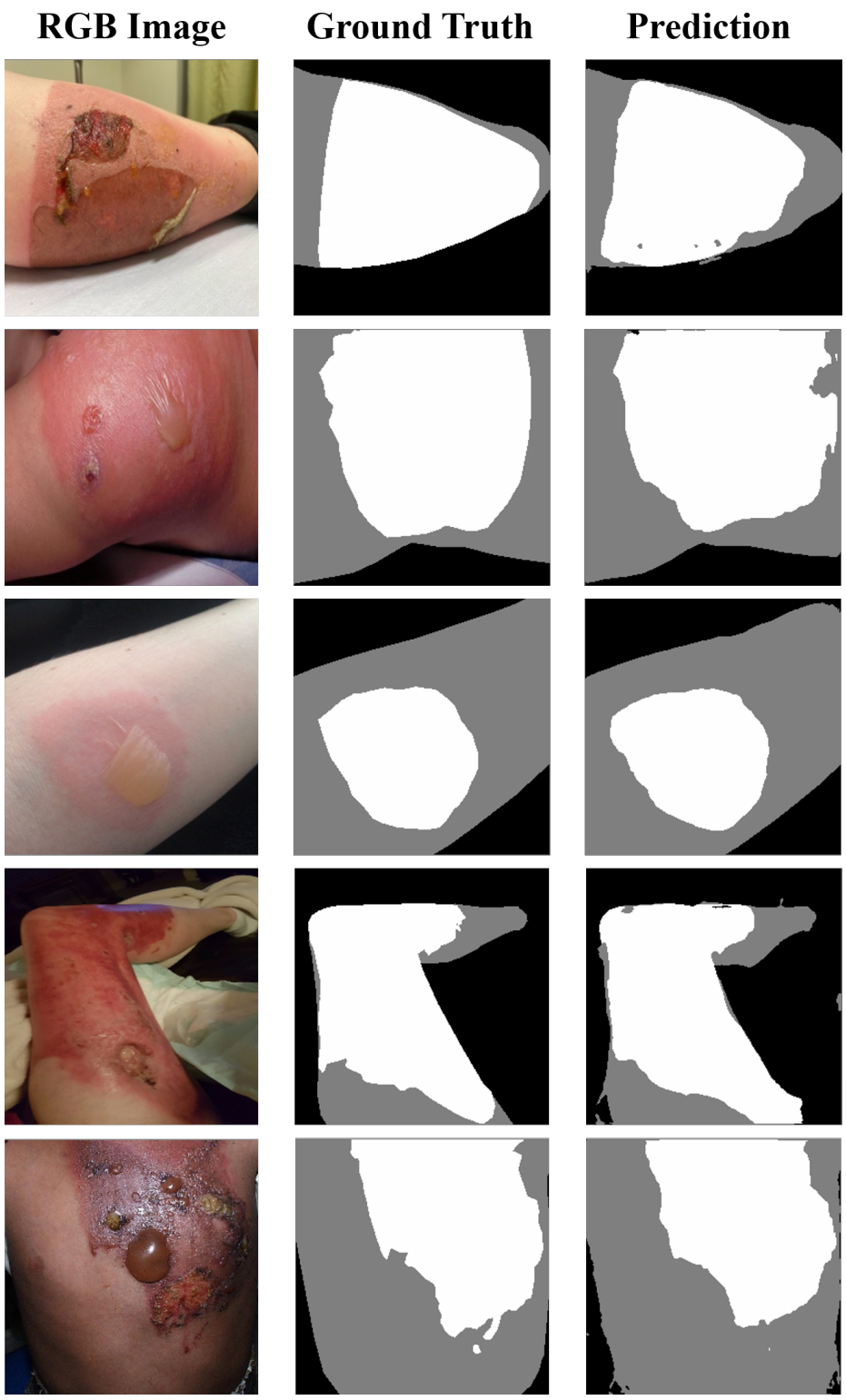


Figure 4.2. Examples of the semantic segmentation results with our proposed model.

4.1.3 Statistical Analysis

In this subsection, the differences obtained for both models are studied to determine if they are significant. Even though there is a considerable difference among them, a statistical analysis is required. For this purpose, a two samples t-test was conducted for each of the evaluation metrics, with a significance level of 0.05. The results obtained for the overall accuracy are summarized in Table 4.3. The DF refers to the degrees of freedom, number that is calculated by subtracting one from the number of observations per group, in this case, six. From the resulting t-statistic and p-value, it can be concluded that the difference between the mean of the global accuracies obtained for the original U-Net architecture and the regularized U-Net is indeed significant.

Table 4.3. Two-sample t-test for global accuracy.

T-Value	DF	p-value
-6.47	5	**0.001**

Similarly, the results obtained for the other evaluation metric, mean IoU, are summarized in Table 4.4. Unsurprisingly, the difference in the means for this metric is also significant given that the obtained p-value is less than the selected significance level.

Table 4.4. Two-sample t-test for mean IoU.

T-Value	DF	p-value
-6.88	5	**0.001**

It was statistically proven that the proposed model yields better results than the original U-Net architecture. Besides, as depicted in Figure 4.2, it was successful in the semantic segmentation task. This claim is based on the fact that the ground truth mask is highly similar to the the predicted mask, fact that is confirmed by the evaluation metrics results.

Therefore, the effectiveness of the proposed approach to extract the burn areas and the skin from RGB images was demonstrated using qualitative and quantitative methods.

4.2 Comparison of Deep Learning Models for Burn Depth Classification

The goal of this experiment was to select the deep learning model that yields the best burn depth classification results. To this end, the balanced dataset previously constructed with the B-mode ultrasound data obtained from pig subjects was adopted. In this regard, each of the three classes, skin, partial-thickness burn, and full-thickness burn, contained the same number of samples. All of them were saved in the Joint Photographic Group (JPG) format. Additionally, the training and the validation datasets were constructed with the data obtained from the first three pigs. On the other hand, the testing dataset was created using the data extracted from the fourth pig. This way, the introduction of errors in the classification models was prohibited. A detailed description of the dataset is shown in Table 4.5, summarizing the number of samples available in each category.

Table 4.5. Burn depth dataset description.

	Skin	**Partial-thickness**	**Full-thickness**	**Total**
Training	425	425	425	1275
Validation	55	55	55	165
Training	65	65	65	195

The models selected to conduct the comparison were ResNet18, ResNet50, ResNet101, VGG16, and VGG19. Their performance as off-the-shelf features extractors was compared. To this end, a transfer learning approach was implemented, and the last fully connected layer of each model was replaced with a new one specifically modified for this particular problem, i.e., a layer with three output channels. The training and validation datasets previously described were used for this step. Then, the testing dataset was used for the quantitative comparison of the performance of the models for the burn depth classification task based on the computed evaluation metrics. As it was stated on section 3.2.2, the utilized hardware

consists of a Windows 10 operating system with Intel Core i7 7700HQ CPU with NVIDIA GeForce GTX 1050Ti GPU with 4GB. The hyperparameters used are summarized in Table 4.6.

Table 4.6. Hyperparameter specifications for training deep learning models selected.

Hyperparameter	**Value**
Optimizer	Adam
Epochs	150
Learning rate	1e-5
Learning rate decay	None
Batch size	8

Similar to the semantic segmentation problem, to guarantee that the results obtained in this experiment were significant, a power analysis was conducted in RStudio. Based on the resulting sample size required to obtain a 90% power, and the capabilities of the available hardware in terms of memory, each model was trained eight times. The experimental results are presented in the subsections below.

4.2.1 Burn Depth Classification Experimental Results

In this subsection, the results obtained for the selected evaluation metrics are presented. The training process was repeated eight times for each of the models. Then, the result of each iteration and the testing dataset described above were used to conduct a quantitative evaluation of the models' performance. Based on that, the accuracy, precision, recall, and F1-Score were calculated. Therefore, given that there were 5 deep learning models, 8 iterations per model, and 4 evaluation metrics, at the end of this step, 160 results were obtained.

The selected evaluation metrics were computed based on the confusion matrix. In Figure 4.3, the confusion matrix for one of the iterations of the ResNet101 model is presented. In the case of a perfect prediction, all the values outside the principal diagonal of this matrix are zero. For the presented example, most of the testing dataset samples were classified correctly, except for some observations off the diagonal of the confusion matrix. Out of the 65 images in the skin category, only one of them was classified as a partial-thickness burn injury. For the images of partial-thickness burns, five of them were classified incorrectly. Three were classified as a full-thickness burn, and the remaining two were classified as healthy skin. Last, for the images of full-thickness burns, four of them were classified as partial-thickness burn injuries. Based on these results, the most challenging classification category for the model was the partial-thickness one. This is not surprising considering that this category is also the most challenging one for practitioners and burn surgeon experts during the visual and physical examination of burn injuries.

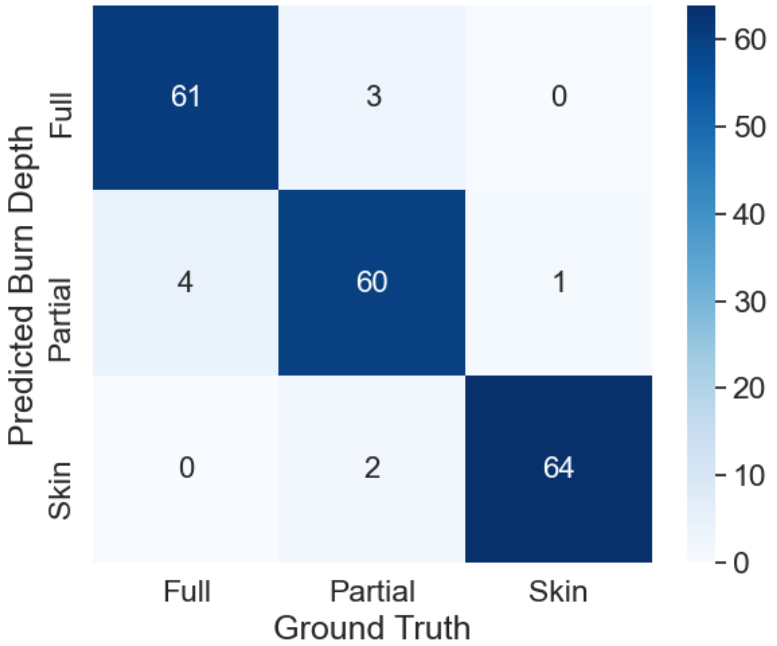


Figure 4.3. Unnormalized confusion matrix for ResNet101 iteration.

Table 4.7 presents the average of the evaluation metrics obtained for the burn depth classification task. ResNet101 achieved the best results for each of the computed metrics, which means that it has the best overall performance. Considering the fact that ResNet101 is the deepest architecture from the ones assessed, these results indicate that the burn depth classification improves with deeper learning models. Nevertheless, the size of our dataset limits the depth of the models that can be used for this particular problem due to the overfitting error.

Table 4.7. Burn depth classification evaluation metrics.

	ResNet18	**ResNet50**	**ResNet101**	**VGG16**	**VGG19**
Accuracy	0.920	0.926	**0.935**	0.913	0.913
Recall	0.899	0.899	**0.914**	0.829	0.879
Precision	0.918	0.926	**0.933**	0.912	0.911
F1-Score	0.905	0.908	**0.920**	0.898	0.896

4.2.2 Statistical Analysis

In this subsection, it was studied whether the differences in the means of the evaluation metrics are significant. The results obtained for the eight iterations of the assessed models were used to perform a one-way ANOVA test for each metric. The tests were formulated as follows: The five models used for the classification task were the levels of the test, and the evaluation metrics were the studied response in each of the conducted tests. In addition, the significance level was set to 0.05. All the tests shared the same hypothesis:

$$H_0 : \mu_{ResNet18} = \mu_{ResNet50} = \mu_{ResNet101} = \mu_{VGG16} = \mu_{VGG19}$$
$$H_a : At\ least\ two\ \mu_{\mathrm{i}}\ are\ different$$

The results of the five one-way ANOVA tests performed are summarized in Table 4.8.

All the resulting p-values were less than 0.05, which indicates that the differences among the models for each of the evaluation metrics are significant. Therefore, a Tukey's test was

Table 4.8. One-way ANOVA tests results.

	F-value	p-value
Accuracy	3.92	**0.010**
Recall	2.89	**0.036**
Precision	4.37	**0.006**
F1-Score	3.25	**0.023**

performed using the statistical software Minitab to determine which models have significantly different means. These results are summarized in Figure 4.4. with a graph for each evaluation metric. Each line represents an interval for a pairwise comparison. If this interval does not contain zero, i.e., if the horizontal line does not cross the dotted line, there is a significant difference between the two compared models. Therefore, it is evident that for all the metrics, except for recall, there is a significant difference between ResNet101 and both VGG models. For recall, this difference only occurs between ResNet101 and VGG19. These results imply that using any of the evaluated ResNet models leads to the same burn depth classification results. Therefore, for the last experiment, ResNet18 was selected to assess the effects introduced by the texture features.

4.3 Assessment of the Effects Texture Features in the Classification Model

This experiment explored the effects of introducing hand-crafted features, obtained from the B-mode HUSD images, as extra features for the burn depth classification problem. The ResNet18 architecture pretrained on ImageNet was selected as the base of the model implemented for this experiment given that no significant differences were found when compared to the ResNet101 model. There was one main reason for this selection. The modification and addition of the fully connected layers that enable the introduction of the texture features for the classification problem caused an increment in the number of trainable parameters of the model. For ResNet18, this increment was not as dramatic as it was for the ResNet101 model. Therefore, for the latter, the constructed dataset was not large enough for the train-

ing phase, and as a consequence, there was a severe overfitting that could not be addressed with common regularization strategies. This was alleviated by the use of ResNet18.

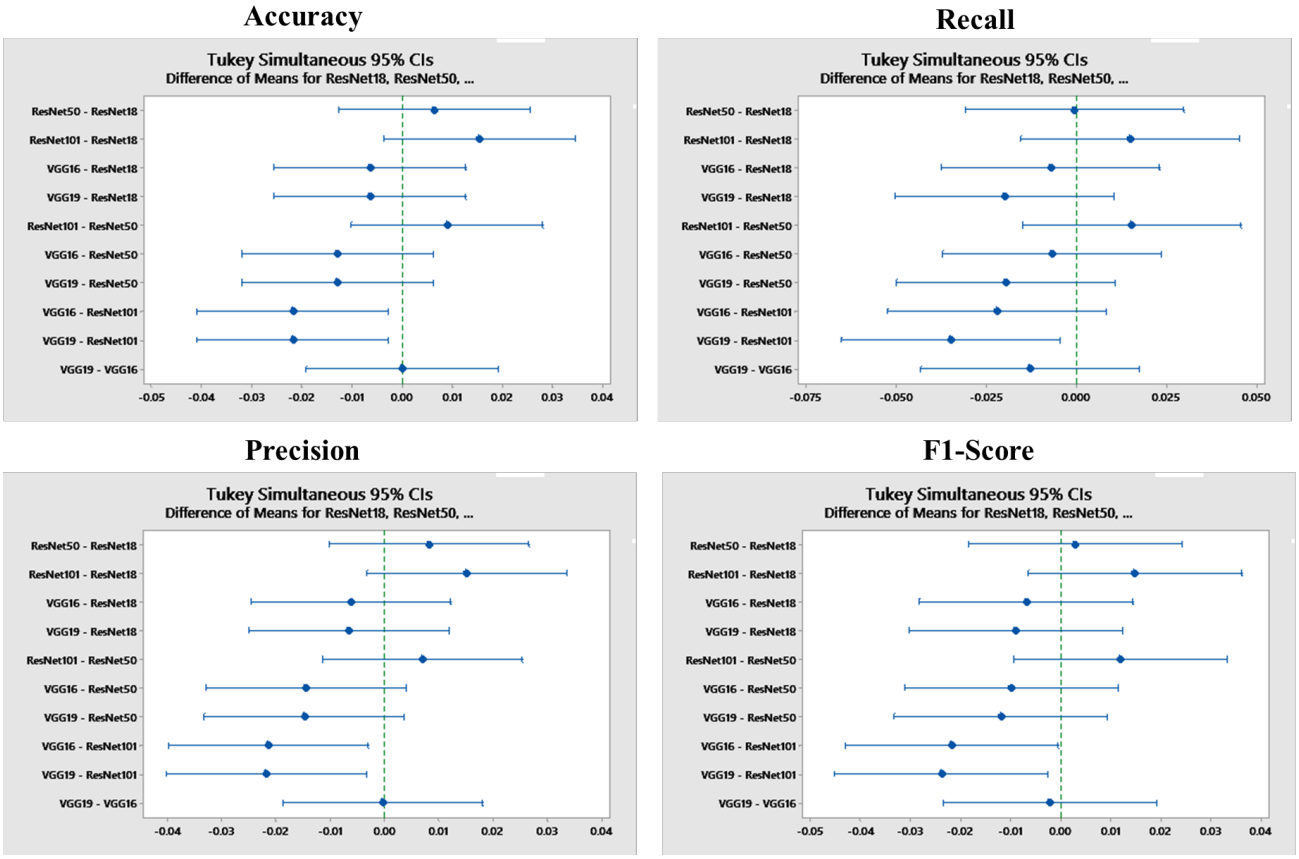


Figure 4.4. Tukey's test results for the evaluation metrics.

The dataset used for this experiment is the one created using the B-mode HUSD images, which was utilized in the previous experiment. This is, a balanced dataset containing three classes, skin, partial-thickness burn, and full-thickness burn. Additionally, the hyperparameters settings were also the same as the ones from the previous experiment. These are summarized in Table 4.6.

The CNN-based features were obtained using the convolutional layers of the ResNet18 model, and the texture features were extracted from the GLCM, using Python. Then, a hybrid feature vector was created by concatenating the CNN-based and the texture features. This created vector served as the input to the last fully connected layer of the implemented model, that in turn, served as the classifier. Therefore, the proposed model is an end-to-end

architecture in which the extraction and concatenation of the features, and the classification occurs internally.

To conduct a quantitative evaluation of the results obtained with the inclusion of the texture features, five training and testing iterations were required. As in the previous experiment, a power analysis was performed using RStudio to determine the number of iterations. Nevertheless, the computational requirements of this experiment were higher, thus, the minimum number of iterations needed to obtain significant results was selected. Additionally, the evaluation metrics described by Equations 3.2, 3.9, 3.10, and 3.11 were calculated. In the following subsections, the results obtained for the GLCM, the selected evaluation metrics, and the statistical analysis are presented.

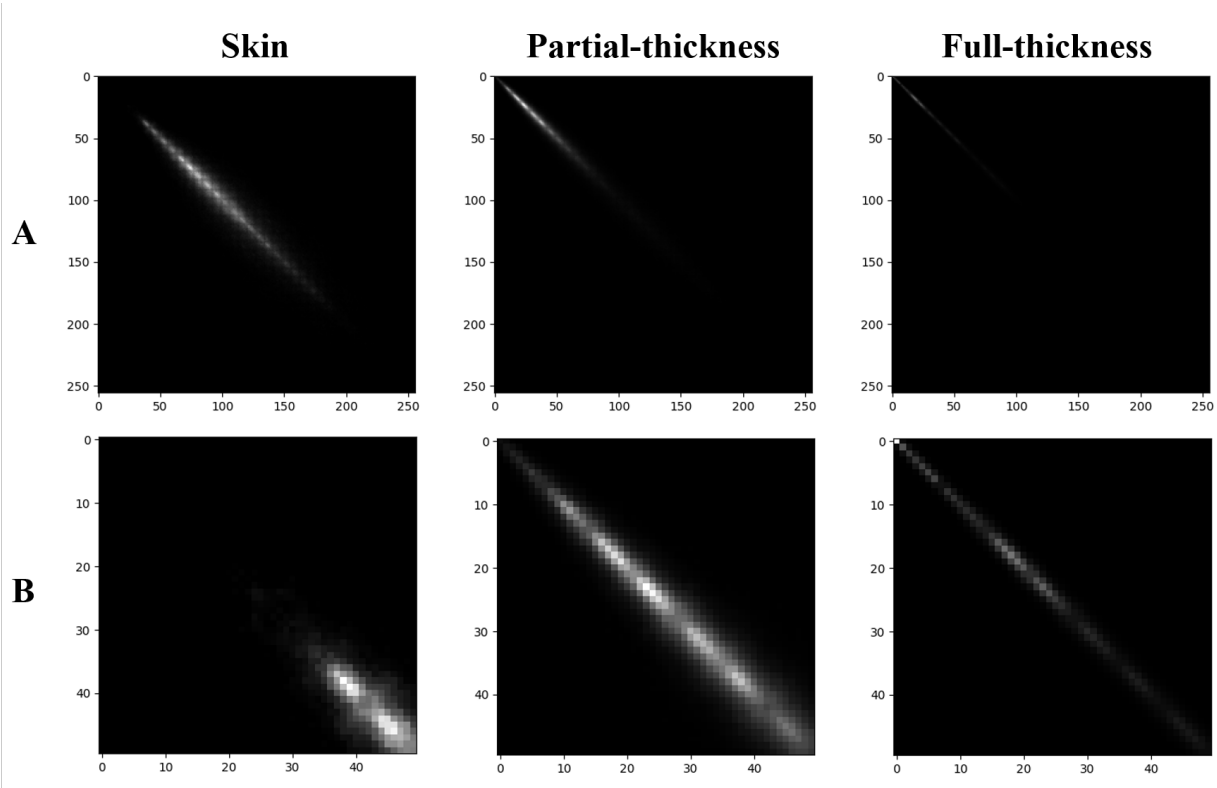


Figure 4.5. Example of GLCM for each class. (a) Complete GLCM with 256 gray levels. (b) Close up of the GLCM from level 0 to 50.

4.3.1 GLCM Calculation

The texture features used in this experiment were extracted from the GLCM. This matrix was computed in Python with the image processing package skimage. The distances and angles selected were $d = \{1, 2\}$ and $\theta = \{0, 45, 90\}$, respectively. In addition, the five texture features extracted were contrast, homogeneity, ASM, energy, and dissimilarity, which are defined in Table 3.1. Figure 4.5 shows the normalized symmetric GLCM for one sample of each class. These were calculated for $d = \{1\}$ and $\theta = \{90\}$.

There is a notorious difference in the GLCM obtained for each class. The size of the gray region in the GLCM matrix is inversely proportional to the severity of the burn. In this regard, the gray region corresponding to the skin is bigger than the one corresponding to the partial-thickness, which in turn is bigger than the one corresponding to the full-thickness burn. Given that this gray region depicts the interaction of two pixels with an associated gray level, it can be inferred that the contrast and the dissimilarity, which were selected as texture features, decrease with the severity of the burn. Therefore, they can effectively characterize the B-mode HUSD images for each class.

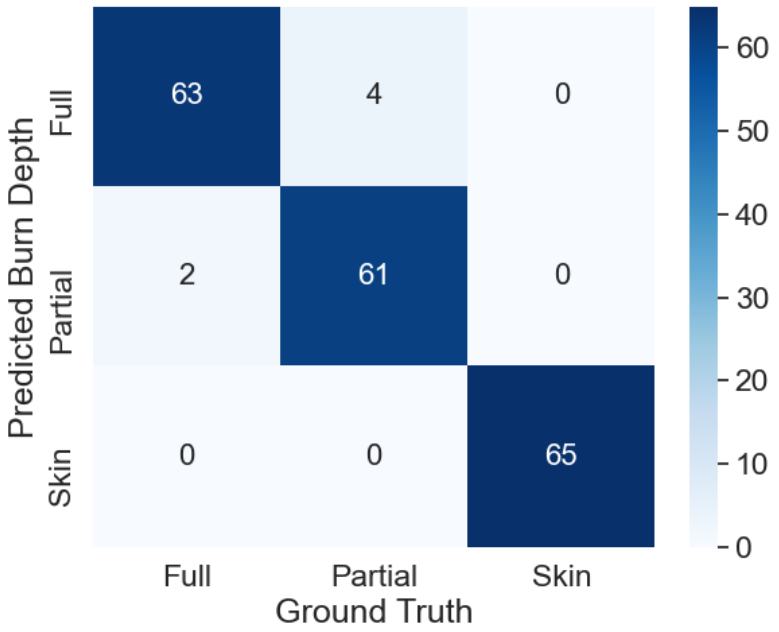


Figure 4.6. Unnormalized confusion matrix for hybrid model.

4.3.2 Burn Depth Classification Using Hybrid Features

After the training phase of the proposed hybrid model, the testing phase was carried out using the unseen testing dataset. The accuracy, recall, precision, and F1-score were obtained for each iteration. These metrics are calculated based on the confusion matrix. Figure 4.6 presents the confusion matrix obtained for one of the iterations of the hybrid model.

As occurred in the example presented in Figure 4.3 for the ResNet101 model, the most challenging category for the classification task was partial-thickness burn. Out of the 65 images in this category, four of them were classified as a full-thickness burn injury. Thus, in this example, the incorrect predictions of the model were always an overestimation of the real burn depth. For the images of the full-thickness class, two of them were classified as a partial-thickness injury. Last, a perfect classification was achieved for the healthy skin class. Based on the confusion matrix, the average of the evaluation metrics presented in Table 4.9 was obtained for the hybrid model.

Table 4.9. Burn depth classification evaluation metrics for hybrid model.

	Hybrid Model
Accuracy	0.957
Recall	0.911
Precision	0.931
F1-Score	0.918

These results were compared to the ones obtained with the ResNet models. The VGG models were not included in this experiment since it was previously demonstrated that both of them achieve a significant lower overall performance than the ResNet101 architecture. However, there was no significant difference found within the ResNet architectures included in this thesis. As mentioned above, the hybrid model, i.e., the model that incorporates the texture features, was only trained for five iterations in this experiment. Considering this, only the first five iterations from ResNet18, ResNet50, and ResNet101 were utilized to con-

duct the comparison. The average obtained with these five iterations for each evaluation metric is presented in Table 4.10. Based on these results, the best overall performance corresponds to the one achieved by the hybrid model, as previously hypothesized. Therefore, it was demonstrated that the inclusion of the texture features and the fusion strategy implemented benefited the burn depth classification problem. It is worth mentioning that due to hardware and dataset limitations, the hybrid model was developed having ResNet18 as a base. Consequently, the CNN-based features were obtained using the convolutional layers of ResNet18. Nevertheless, the introduction of the texture features resulted not only in a better overall performance than the ResNet18 architecture, but also than the ResNet50 and ResNet101 architectures.

Table 4.10. Burn depth classification evaluation metrics for ResNet architectures.

	ResNet18	**ResNet50**	**ResNet101**
Accuracy	0.916	0.927	0.930
Recall	0.895	0.901	0.908
Precision	0.914	0.927	0.928
F1-Score	0.901	0.909	0.914

4.3.3 Statistical Analysis

To determine whether the differences observed in the subsection above are significant, a statistical analysis was conducted. In this subsection, the statistical software Minitab was used to perform a one-way ANOVA test for each of the evaluation metrics computed. The tested levels are the hybrid model, ResNet18, ResNet50, and ResNet101. Additionally, the significance level was set to 0.05. The results obtained are summarized in Table 4.11.

There is only one p-value less than the significance level, 0.05, the one that corresponds to the accuracy. The differences found for the remaining three evaluation metrics are not significant. Therefore, they are not considered for further analysis. Following the significant

difference found for the accuracy, a Tukey's test was performed to establish which models lead to statistically different results.

Table 4.11. One-way ANOVA tests results.

	F-value	**p-value**
Accuracy	10.15	**0.001**
Recall	0.96	0.437
Precision	2.26	0.120
F1-Score	1.42	0.247

The results of the Tukey's test pairwise comparison are shown in Figure 4.7. It is observed that three intervals do not cross the zero line. These correspond to the pairs conformed by the hybrid model with each of the ResNet architectures, ResNet18, ResNet50, and ResNet101. This indicates that there is a significant difference among the hybrid model and the three ResNet architectures studied. These results demonstrated that the use of the model that was complemented with the texture features leads to the best results in terms of accuracy. The dataset used for this experiment and the previous one was balanced. This is worthwhile noting given that the results obtained for the accuracy could be misleading if the models were trained and tested with an unbalanced dataset. However, this is not the case in this experiment. Therefore, selecting the hybrid model as the best model for the burn depth classification problem based on the accuracies obtained was appropriate.

4.4 Summary

In this chapter, the results for the experiments conducted to evaluate the performance of the models proposed for the semantic segmentation problem and the burn depth classification problem were presented. In this regard, global pixel accuracy and mean IoU were calculated for the regularized U-Net and the original U-Net. For both metrics, the best results for the semantic segmentation problem were achieved with the regularized U-Net. Additionally, for the burn depth classification problem, three ResNet architectures and two VGG architectures

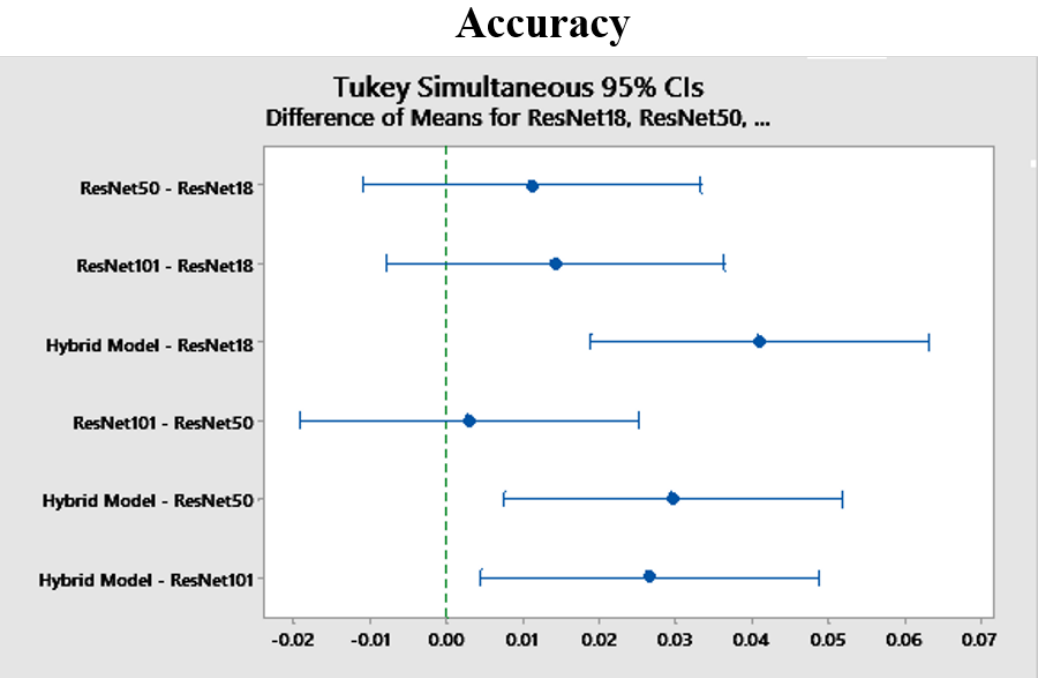


Figure 4.7. Tukey's test results for accuracy.

were compared based on the accuracy, recall, precision, and F1-score. The results indicated that the ResNet101 model is statistically better than VGG16, and VGG19. However, no significant differences were found among ResNet18, ResNet50, and ResNet101. Last, to assess the effects of the introduction of the texture features to the proposed model for the classification task, the performance of the three ResNet architectures was compared with the one corresponding to the hybrid model. The results showed a significant difference for the accuracy, indicating that the proposed hybrid model yields the best classification results. In general, the satisfactory results presented throughout this section demonstrated that the use of AI, in particular the framework proposed in this thesis, is suitable and beneficial for the assessment of burn injuries.

5. CONCLUSIONS AND FUTURE WORK

In this thesis, an AI-based framework for automatic assessment of burn injuries was developed to support the medical decision-making process. This approach was constituted by two main elements: a semantic segmentation module, and a burn depth classification module. For the first element, the burn injury area and surrounding healthy skin were extracted from RGB images acquired under a variety of conditions. A regularized U-Net architecture was implemented to assign the corresponding label to each pixel of the image based on the three categories available: burn, skin and background. For the second element, the severity of the injury was assessed and the corresponding burn depth was predicted. In this regard, multiple widely used deep learning models were implemented and compared. Their performance, more specifically, the accuracy, was later boosted by the introduction of texture features as extra features and the implementation of an effective feature fusion strategy. This AI-based approach automatically predicted pixel-level and image-level classifications that were critical in the burn assessment task. These predictions can be utilized to support the medical decision-making process that results in the selection of the optimal burn management strategy and surgical needs. Thus, the work presented constitutes a first step towards for the development of a Computer-Aided Diagnosis or Detection system for an accurate and prompt initial assessment of burn injuries. This development could decrease waiting times and avoid unnecessary delays in the burn evaluation process. Additionally, it could output satisfactory diagnostic results, generating an increase in the burn depth classification accuracy, and addressing the overestimation in the calculation of the burn superficial area. Last, this development could also decrease the inter-subject variability, the healing time, and the mortality and morbidity rates. This, in turn, can lead to the obtention of better outcomes for the wound care process, and satisfactory results for the burn patient.

In addition, the proposed framework is also a first step towards the reduction in the need of burn surgeon experts for burn assessment. A system, as presented in this thesis, would allow nurses, medics and generalists to assess burn injuries. Such a system could be deployed in a tablet or a smartphone, which have embedded cameras and are compatible with USB handheld ultrasound scanners. These are the only devices required for the implementation

of the proposed approach. The first enables the obtention of RGB images required for the segmentation step. The second one enables the obtention of B-mode ultrasound images required for the burn depth classification step. The utilization of such devices does not require profound expertise. Any medical practitioners, for example, would be capable of using a digital camera to take a picture following given instructions, and feed the diagnostic system presented. In addition, most of these practitioners are also trained to use ultrasound devices for the assessment and diagnostics of several pathologies. Thus, in the case of a disaster, in austere environments, or even in normal circumstances in which burn surgeons are not available, the practitioner would conduct the burn assessment process independently and use the output of the automatic system as a complement to a physical and visual examination that is performed anyway. The expert burn surgeons, non-expert practitioners, and burn patients could all benefit from the proposed approach.

The results obtained for the semantic segmentation task showed that the implemented regularized U-Net architecture is robust and effective for the segmentation of the burn, skin, and background. This is a particular challenging segmentation task given the variability in the appearance of this type of wounds, and the inherent variability of the human anatomy. The results obtained for the second experiment of the burn depth classification problem demonstrated that using an end-to-end architecture that leverages CNN-based features combined with hand-crafted features, i.e., texture features, constitutes an attractive alternative for this specific task. Complementary information can be acquired from the two types of features included in the model, yielding satisfactory results. The proposed model not only outperforms widely used deep learning architectures; it also outperforms medical practitioners in the burn depth classification task. Finally, the proposed approach showed that the use of ultrasonography impacts positively the assessment of burn injuries. Having access to information of internal tissues and organs that are not visible from the outside benefits the assessment process without incurring in invasive techniques, such as histological analysis of biopsies.

5.1 Limitations

In this thesis, the architecture for the semantic segmentation problem, as well as the one for the burn depth classification problem, were trained and tested using datasets created for these particular tasks. Both of them included a limited number of samples which can restrict the capabilities of the model. In this regard, the evaluation metrics computed to assess the performance of the proposed and implemented models during the testing phase, were obtained using exclusively this small dataset. In a real setting, a wider variety of burn injuries in terms of appearance and severity characteristics could be encountered. Even though there was a significant effort to include a wide variety of burn injuries with different characteristics to enhance the robustness of the model, the results obtained are only applicable to the dataset used. The dataset with RGB images for the segmentation task is available and can be found in the following GitHub repository https://github.com/dchancia/Burn_Assessment. To enhance the capabilities of the proposed approach, boost the overall performance of the model, and guarantee satisfactory results in real life settings, it is a requirement to increase the size of the dataset. This is a challenging task considering the barriers to overcome for accessing the data. First, it is not a widely studied topic in the field of medical image analysis, and as a result, there are no publicly available datasets, reason why one of the datasets created in this thesis was made available. Second, it is difficult to obtain the data directly from specialized burn facilities, emergency rooms, and hospitals given the sensitivity of the information, the patients' privacy concerns, and the emergent nature of burn injuries treatment which requires immediate action. Additionally, the labelling and annotation process is a time-consuming task that requires significant expertise in burn injuries. Even if a weakly-supervised approach or a semi-supervised approach is implemented in future attempts, instead of the supervised approach proposed in this thesis, manually annotated data and the required expertise to do it is still necessary. Considering the mentioned hurdles, it is evident that augmenting the size of both datasets is indeed a challenging task. However, without this effort, the creation of the intended Computer-Diagnosis or Detection system for burn injuries faces critical obstacles that can restrict the subsequent adoption of the system in clinical settings.

Currently, the two main elements of the proposed framework, i.e., the segmentation module and the classification module, are totally independent from each. This is an aspect in which the future attempts can focus. In this work, the RGB image corresponding to the ultrasound data is not available. However, even if the image was available, it is not suitable for the segmentation task. The reason is that the burn injuries used for the obtention of the ultrasound data were created with the same shape, a 2x2 inches square. The combination of both modules could be beneficial as features obtained using the RGB image, after the semantic segmentation is performed, can be also included in the classification framework.

5.2 Future Work

There are two main aspects to be considered for the future work. For the proposed framework, more specifically, for the semantic segmentation part, it is clear that the model can only segment the body regions that are shown in the RGB image. However, for the calculation of the TBSA, it is necessary to provide the model with a full visualization of the body. Thus, the semantic segmentation module of the proposed framework cannot be used for the calculation of the TBSA as is. It is necessary to implement an intermediary step in which it is possible to obtain a 2D visualization of the whole body to extract the burned area from the semantic segmentation outputs. To this end, the implementation of a function capable of mapping the human body from a 3D input, or multiple 2D inputs, to a single 2D output with high accuracy is required. One possible approach to achieve this is explained as follows. Obtain several views of the body from different angles, using a regular digital camera or a depth camera, and conduct an alignment and reconstruction of the 3D model of the human body. Subsequently, convert the reconstruction to a 2D image suitable for the already developed segmentation algorithm. As a result, it is possible to obtain the whole surface area of the body in an accurate manner. It is worth mentioning that the reconstruction of the human body could be achieved using a single frontal take obtained with a depth camera. However, this surface area must include both, the burn injuries for the model to extract the area of such burn injuries, and the area that belongs to the healthy

skin. Then, it can separate both of those regions of interest from the background to achieve the TBSA calculation.

The second aspect for future work is related to the fact that the classification module of the proposed framework uses B-mode HUSD data exclusively from porcine models. Even though there are relevant similarities between the healing process of a pig subject and a human subject, each has its own anatomical structures and body characteristics. The final aim of this project is the development of a reliable automatic system for the assessment of burn injuries that can be adopted in clinical settings, i.e., for the human body. Thus, the classification module needs to be carefully adapted for humans in the future.

5.3 Research Questions

The methodology proposed and the experimental results obtained throughout this thesis addressed the three research questions that were formulated at the beginning. A brief discussion for each of them is presented below.

RQ1: How can deep learning be used for burn area size estimation using images obtained in uncontrolled conditions?

To answer this question, a suitable dataset was created. More specifically, RGB images of diverse burn injuries available online were collected, preprocessed and annotated. Such annotation refers to a pixel-wise labelling of the image that enables the identification of the burn region, the healthy skin region, and the background. Additionally, a data augmentation strategy was proposed to increase the size of the mentioned dataset. This enabled the training of a deep learning based segmentation model following a supervised learning approach. The model proposed for this approach was a regularized version of the U-Net architecture that alleviated the overfitting concerns regarding the small-sized dataset. The resulting prediction masks enabled the partition of the image in three categories: burn, healthy skin, and background. The model was tested using an unseen testing dataset and satisfactory results were obtained. In this regard, after the segmentation task, the region corresponding

to the burn, and the one corresponding to the skin can be used to determine the burn size as a percentage value for the body surface depicted in the RGB image. Note that if a 2D visualization of the complete body surface is provided, the proposed framework could be used to directly obtain the TBSA.

RQ2: What is the best model for burn depth assessment that leads to a higher classification accuracy?

To answer this question, a different burn injury dataset was created using B-mode HUSD data obtained from porcine models. A group of widely used deep learning models were trained for the burn depth classification task using the created dataset and following a transfer learning approach. These models produced satisfactory results for this particular task (over 90% accuracies), that were then used to statistically compare the performance of such models based on four evaluation metrics. Significant differences were found for the ResNet101 architecture and both, VGG16 and VGG19. However, no differences were found for the three ResNet architectures studied; ResNet18, ResNet50, and ResNet101. Therefore, the three of them yield the higher classification accuracy.

RQ3: Does the integration of texture features lead to a better burn depth classification accuracy?

To answer this question, a deep learning model based on the architecture of ResNet18 was proposed and implemented. This model enabled the incorporation of texture features for the burn depth classification task. The input for the last fully connected layer of the model, i.e., the classifier, was a hybrid vector that contained 30 CNN-based features and 30 texture features, to guarantee an equivalent participation of both types of features in the classification task. The CNN-based features were extracted by the convolutional layers of the model, and the texture features were extracted from the computed GLCM. Even though there were no significant differences for the recall, precision, and F1-score, the proposed hybrid model achieved a classification accuracy mean of 95.7

REFERENCES

Ahn, J., & Kwak, S. (2018). Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4981–4990.

Ahuja, S., Panigrahi, B. K., Dey, N., Rajinikanth, V., & Gandhi, T. K. (2021). Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices [Publisher: Springer]. *Applied Intelligence*, *51*(1), 571–585.

Alfred, C., Demi, L., Muller, M., & Zhou, Q. (2020). Ultrasound Imaging: A Silent Hero in COVID-19 and Lung Diagnostics [Publisher: IEEE]. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *67*(11), 2194–2196.

Almotairi, S., Kareem, G., Aouf, M., Almutairi, B., & Salem, M. A.-M. (2020). Liver tumor segmentation in CT scans using modified segnet [Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *20*(5), 1516.

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*.

American Burn Association. (2015). https://ameriburn.org/who-we-are/media/burn-incidence-fact-sheet/

Armi, L., & Fekri-Ershad, S. (2019). Texture image analysis and texture classification methods-A review. *arXiv preprint arXiv:1904.06554*.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation [Publisher: IEEE]. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2481–2495.

Bauer, J., & Sauer, T. (1989). Cutaneous 10 MHz ultrasound B scan allows the quantitative assessment of burn depth [Publisher: Elsevier]. *Burns*, *15*(1), 49–51.

Bharati, M. H., Liu, J. J., & MacGregor, J. F. (2004). Image texture analysis: Methods and comparisons [Publisher: Elsevier]. *Chemometrics and intelligent laboratory systems*, *72*(1), 57–71.

Brodzicki, A., Jaworek-Korjakowska, J., Kleczek, P., Garland, M., & Bogyo, M. (2020). Pre-Trained Deep Convolutional Neural Network for Clostridioides Difficile Bacteria Cytotoxicity Classification Based on Fluorescence Images [Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *20*(23), 6713.

Brown, R., Rice, P., & Bennett, N. (1998). The use of laser Doppler imaging as an aid in clinical management decision making in the treatment of vesicant burns [Publisher: Elsevier]. *Burns*, *24*(8), 692–698.

Browning, J. A., & Cindass, R. (2019). Burn Debridement, Grafting, and Reconstruction. *StatPearls [Internet]*. StatPearls Publishing.

Byra, M., Styczynski, G., Szmigielski, C., Kalinowski, P., Michalowski, L., Paluszkiewicz, R., Ziarkiewicz-Wróblewska, B., Zieniewicz, K., Sobieraj, P., & Nowicki, A. (2018). Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images [Publisher: Springer]. *International journal of computer assisted radiology and surgery*, *13*(12), 1895–1903.

Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation [Publisher: AME Publications]. *Annals of translational medicine*, *8*(11).

Cai, L., Long, T., Dai, Y., & Huang, Y. (2020). Mask R-CNN-based detection and segmentation for pulmonary nodule 3D visualization diagnosis [Publisher: IEEE]. *IEEE Access*, *8*, 44400–44409.

Castellano, G., Bonilha, L., Li, L., & Cendes, F. (2004). Texture analysis of medical images [Publisher: Elsevier]. *Clinical radiology*, *59*(12), 1061–1069.

Chae, J., Hong, K. Y., & Kim, J. (2021). A Pressure Ulcer Care System For Remote Medical Assistance: Residual U-Net with an Attention Model Based for Wound Area Segmentation. *arXiv preprint arXiv:2101.09433*.

Chan, V., & Perlas, A. (2011). Basics of ultrasound imaging. *Atlas of ultrasound-guided procedures in interventional pain management* (pp. 13–19). Springer.

Chauhan, J., & Goyal, P. (2020a). Convolution neural network for effective burn region segmentation of color images [Publisher: Elsevier]. *Burns*.

Chauhan, J., & Goyal, P. (2020b). Deep Learning based fully automatic efficient Burn Severity Estimators for better Burn Diagnosis. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Cheng, J.-Z., Ni, D., Chou, Y.-H., Qin, J., Tiu, C.-M., Chang, Y.-C., Huang, C.-S., Shen, D., & Chen, C.-M. (2016). Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans [Publisher: Nature Publishing Group]. *Scientific reports*, *6*(1), 1–13.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Cirillo, M. D., Mirdell, R., Sjöberg, F., & Pham, T. D. (2019). Tensor decomposition for colour image segmentation of burn wounds [Publisher: Nature Publishing Group]. *Scientific reports*, *9*(1), 1–13.

Cubison, T. C., Pape, S. A., & Parkhouse, N. (2006). Evidence for the link between healing time and the development of hypertrophic scars (HTS) in paediatric burns due to scald injury [Publisher: Elsevier]. *Burns*, *32*(8), 992–999.

da Nóbrega, R. V. M., Peixoto, S. A., da Silva, S. P. P., & Rebouças Filho, P. P. (2018). Lung nodule classification via deep transfer learning in CT lung images. *2018 IEEE 31st international symposium on computer-based medical systems (CBMS)*, 244–249.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, 248–255.

Despo, O., Yeung, S., Jopling, J., Pridgen, B., Sheckter, C., Silberstein, S., Fei-Fei, L., & Milstein, A. (2017). BURNED: Towards efficient and accurate burn prognosis using deep learning.

Devgan, L., Bhat, S., Aylward, S., & Spence, R. J. (2006). Modalities for the assessment of burn wound depth [Publisher: Open Science Co.]. *Journal of burns and wounds*, *5*.

Diaz, O., Kushibar, K., Osuala, R., Linardos, A., Garrucho, L., Igual, L., Radeva, P., Prior, F., Gkontra, P., & Lekadir, K. (2021). Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools [Publisher: Elsevier]. *Physica Medica*, *83*, 25–37.

Ding, Y., Sohn, J. H., Kawczynski, M. G., Trivedi, H., Harnish, R., Jenkins, N. W., Lituiev, D., Copeland, T. P., Aboian, M. S., Mari Aparici, C., et al. (2019). A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain [Publisher: Radiological Society of North America]. *Radiology*, *290*(2), 456–464.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential [Publisher: Elsevier]. *Computerized medical imaging and graphics*, *31*(4-5), 198–211.

Dong, H., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. *annual conference on medical image understanding and analysis*, 506–517.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al. (2019). U-Net: Deep learning for cell counting, detection, and morphometry [Publisher: Nature Publishing Group]. *Nature methods*, *16*(1), 67–70.

Faust, O., Acharya, U. R., Meiburger, K. M., Molinari, F., Koh, J. E., Yeong, C. H., Kongmebhol, P., & Ng, K. H. (2018). Comparative assessment of texture features for the identification of cancer in ultrasound images: A review [Publisher: Elsevier]. *Biocybernetics and Biomedical Engineering*, *38*(2), 275–296.

Finnerty, C. C., Jeschke, M. G., Branski, L. K., Barret, J. P., Dziewulski, P., & Herndon, D. N. (2016). Hypertrophic scarring: The greatest unmet challenge after burn injury [Publisher: Elsevier]. *The Lancet*, *388*(10052), 1427–1436.

Flood, N., Watson, F., & Collett, L. (2019). Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia [Publisher: Elsevier]. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101897.

Fujima, N., Homma, A., Harada, T., Shimizu, Y., Tha, K. K., Kano, S., Mizumachi, T., Li, R., Kudo, K., & Shirato, H. (2019). The utility of MRI histogram and texture analysis for the prediction of histological diagnosis in head and neck malignancies [Publisher: BioMed Central]. *Cancer Imaging*, *19*(1), 1–10.

Fujioka, T., Kubota, K., Mori, M., Kikuchi, Y., Katsuta, L., Kasahara, M., Oda, G., Ishiba, T., Nakagawa, T., & Tateishi, U. (2019). Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network [Publisher: Springer]. *Japanese journal of radiology*, *37*(6), 466–472.

Galavis, P. E., Hollensen, C., Jallow, N., Paliwal, B., & Jeraj, R. (2010). Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters [Publisher: Taylor & Francis]. *Acta oncologica*, *49*(7), 1012–1016.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.

Gnyawali, S. C., Barki, K. G., Mathew-Steiner, S. S., Dixith, S., Vanzant, D., Kim, J., Dickerson, J. L., Datta, S., Powell, H., Roy, S., et al. (2015a). High-resolution harmonics ultrasound imaging for non-invasive characterization of wound healing in a pre-clinical swine model [Publisher: Public Library of Science]. *PLoS One*, *10*(3), e0122327.

Gnyawali, S. C., Barki, K. G., Mathew-Steiner, S. S., Dixith, S., Vanzant, D., Kim, J., Dickerson, J. L., Datta, S., Powell, H., Roy, S., et al. (2015b). High-resolution harmonics ultrasound imaging for non-invasive characterization of wound healing in a pre-clinical swine model [Publisher: Public Library of Science]. *PloS one*, *10*(3), e0122327.

Gomez, R., & Cancio, L. C. (2007). Management of burn wounds in the emergency department [Publisher: Elsevier]. *Emergency medicine clinics of North America*, *25*(1), 135–146.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.

Gottlieb, M., Holladay, D., & Peksa, G. D. (2019). Point-of-care ocular ultrasound for the diagnosis of retinal detachment: A systematic review and meta-analysis [Publisher: Wiley Online Library]. *Academic Emergency Medicine*, *26*(8), 931–939.

Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks [Publisher: Springer]. *International journal of multimedia information retrieval*, *7*(2), 87–93.

Gupta, S., Arbeláez, P., Girshick, R., & Malik, J. (2015). Indoor scene understanding with rgb-d images: Bottom-up segmentation, object detection and semantic segmentation [Publisher: Springer]. *International Journal of Computer Vision*, *112*(2), 133–149.

Hao, S., Zhou, Y., & Guo, Y. (2020). A brief survey on semantic segmentation with deep learning [Publisher: Elsevier]. *Neurocomputing*, *406*, 302–321.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification [Publisher: Ieee]. *IEEE Transactions on systems, man, and cybernetics*, (6), 610–621.

Harish, V., Raymond, A. P., Issler, A. C., Lajevardi, S. S., Chang, L.-Y., Maitz, P. K., & Kennedy, P. (2015). Accuracy of burn size estimation in patients transferred to adult Burn Units in Sydney, Australia: An audit of 698 patients [Publisher: Elsevier]. *Burns*, *41*(1), 91–99.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Herndon, D. N., Barrow, R. E., Rutan, R. L., Rutan, T. C., Desai, M. H., & Abston, S. (1989). A comparison of conservative versus early excision. Therapies in severely burned patients. [Publisher: Lippincott, Williams, and Wilkins]. *Annals of surgery*, *209*(5), 547.

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges [Publisher: Springer]. *Journal of digital imaging*, *32*(4), 582–596.

Hordiiuk, D., Oliinyk, I., Hnatushenko, V., & Maksymov, K. (2019). Semantic segmentation for ships detection from satellite imagery. *2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)*, 454–457.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Huang, Z., Wang, X., Wang, J., Liu, W., & Wang, J. (2018). Weakly-supervised semantic segmentation network with deep seeded region growing. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7014–7023.

Hung, W.-C., Tsai, Y.-H., Liou, Y.-T., Lin, Y.-Y., & Yang, M.-H. (2018). Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*.

Jackowski, M., & Goshtasby, A. (2005). A Computer-Aided Design System for Segmentation of Volumetric Images.

Jeschke, M. G., Kamolz, L.-P., Sjöberg, F., & Wolf, S. E. (2012). *Handbook of burns volume 1: Acute burn care* (Vol. 1). Springer.

Jeschke, M. G., van Baar, M. E., Choudhry, M. A., Chung, K. K., Gibran, N. S., & Logsetty, S. (2020). Burn injury [Publisher: Nature Publishing Group]. *Nature Reviews Disease Primers*, *6*(1), 1–25.

Jiao, C., Su, K., Xie, W., & Ye, Z. (2019). Burn image segmentation based on Mask Regions with Convolutional Neural Network deep learning framework: More accurate and more convenient [Publisher: Oxford University Press]. *Burns & trauma*, *7*(1), s41038–018.

Jony, M. H., Johora, F. T., Khatun, P., & Rana, H. K. (2019). Detection of Lung Cancer from CT Scan Images using GLCM and SVM. *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 1–6.

Jung, A. B., Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Rui, Z., Borovec, J., Vallentin, C., Zhydenko, S., Pfeiffer, K., Cook, B., Fernández, I., De Rainville, F.-M., Weng, C.-H., … Laporte, M., et al. (2020). *Imgaug*. https://github.com/aleju/imgaug
Online; accessed 01-Feb-2020

Kagan, R. J., Peck, M. D., Ahrenholz, D. H., Hickerson, W. L., Holmes IV, J., Korentager, R., Kraatz, J., Pollock, K., & Kotoski, G. (2013). Surgical management of the burn wound and use of skin substitutes: An expert panel white paper [Publisher: Oxford University Press]. *Journal of Burn Care & Research*, *34*(2), e60–e79.

Kalayeh, M. M., Gong, B., & Shah, M. (2017). Improving facial attribute prediction using semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6942–6950.

Kalus, A., Aindow, J., & Caulfield, M. (1979). Application of ultrasound in assessing burn depth [Publisher: Elsevier]. *The Lancet*, *313*(8109), 188–189.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks [Publisher: Springer]. *Artificial Intelligence Review*, *53*(8), 5455–5516.

Kim, J., Dunham, D., Supp, D., Sen, C., & Powell, H. (2016). Novel burn device for rapid, reproducible burn wound generation [Publisher: Elsevier]. *Burns*, *42*(2), 384–391.

Ko, T.-y., & Lee, S.-h. (2020). Novel Method of Semantic Segmentation Applicable to Augmented Reality [Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *20*(6), 1737.

Kohler, R. (1981). A segmentation system based on thresholding [Publisher: Elsevier]. *Computer Graphics and Image Processing*, *15*(4), 319–338.

Kolkur, S., & Kalbande, D. (2016). Survey of texture based feature extraction for skin disease detection. *2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, 1–6.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097–1105.

Lai, Z., & Deng, H. (2018). Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron [Publisher: Hindawi]. *Computational intelligence and neuroscience*, *2018*.

Lee, S., Ye, H., Chittajallu, D., Kruger, U., Boyko, T., Lukan, J. K., Enquobahrie, A., Norfleet, J., De, S., et al. (2020). Real-time burn classification using ultrasound imaging [Publisher: Nature Publishing Group]. *Scientific reports*, *10*(1), 1–13.

Li, F., Wang, C., Peng, Y., Yuan, Y., & Jin, S. (2019). Wound segmentation network based on location information enhancement [Publisher: IEEE]. *IEEE Access*, *7*, 87223–87232.

Li, F.-F. (2021). Convolutional Neural Networks for Visual Recognition. http://cs231n.stanford.edu/index.html

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., & Heng, P.-A. (2018). H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes [Publisher: IEEE]. *IEEE transactions on medical imaging*, *37*(12), 2663–2674.

Lin, G., Shen, C., Van Den Hengel, A., & Reid, I. (2017). Exploring context with deep structured models for semantic segmentation [Publisher: IEEE]. *IEEE transactions on pattern analysis and machine intelligence*, *40*(6), 1352–1366.

Liu, H., Yue, K., Cheng, S., Li, W., & Fu, Z. (2021). A Framework for Automatic Burn Image Segmentation and Burn Depth Diagnosis Using Deep Learning [Publisher: Hindawi]. *Computational and Mathematical Methods in Medicine*, *2021*.

Liu, W., Rabinovich, A., & Berg, A. C. (2015). Parsenet: Looking wider to see better. *arXiv preprint arXiv:1506.04579*.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Mansour, R. F. (2018). Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy [Publisher: Springer]. *Biomedical engineering letters*, *8*(1), 41–57.

Materka, A. (2004). Texture analysis methodologies for magnetic resonance imaging [Publisher: Les Laboratoires Servier]. *Dialogues in clinical neuroscience*, *6*(2), 243.

Milioto, A., Lottes, P., & Stachniss, C. (2018). Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. *2018 IEEE international conference on robotics and automation (ICRA)*, 2229–2235.

Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey [Publisher: IEEE]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Miranda, E., Aryuni, M., & Irwansyah, E. (2016). A survey of medical image classification techniques. *2016 International Conference on Information Management and Technology (ICIMTech)*, 56–61.

Monstrey, S., Hoeksema, H., Verbelen, J., Pirayesh, A., & Blondeel, P. (2008). Assessment of burn depth and burn wound healing potential [Publisher: Elsevier]. *burns*, *34*(6), 761–769.

Narouze, S. N. (2018). *Atlas of ultrasound-guided procedures in interventional pain management*. Springer.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, *4*, 51–62.

Nawresh, A. A., & Sasikala, S. (2020). An Approach for Efficient Classification of CT Scan Brain Haemorrhage Types Using GLCM Features with Multilayer Perceptron. *ICDSMLA 2019* (pp. 400–412). Springer.

Ng, H., Ong, S., Foong, K., Goh, P.-S., & Nowinski, W. (2006). Medical image segmentation using k-means clustering and improved watershed algorithm. *2006 IEEE southwest symposium on image analysis and interpretation*, 61–65.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pisano, E. D. (2020). *AI shows promise for breast cancer screening*. Nature Publishing Group.

Qummar, S., Khan, F. G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z. U., Khan, I. A., & Jadoon, W. (2019). A deep learning ensemble approach for diabetic retinopathy detection [Publisher: IEEE]. *IEEE Access*, *7*, 150530–150539.

Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future [Publisher: Springer]. *Classification in BioApps*, 323–350.

Rezaei, M., Harmuth, K., Gierke, W., Kellermeier, T., Fischer, M., Yang, H., & Meinel, C. (2017). A conditional adversarial network for semantic segmentation of brain tumor. *International MICCAI Brainlesion Workshop*, 241–252.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241.

Rowland, R. A., Ponticorvo, A., Baldado, M. L., Kennedy, G. T., Burmeister, D. M., Christy, R. J., Bernal, N. P., & Durkin, A. J. (2019). Burn wound classification model using spatial frequency-domain imaging and machine learning [Publisher: International Society for Optics and Photonics]. *Journal of biomedical optics*, *24*(5), 056007.

Roy, S., Elgharably, H., Sinha, M., Ganesh, K., Chaney, S., Mann, E., Miller, C., Khanna, S., Bergdall, V. K., Powell, H. M., et al. (2014). Mixed-species biofilm compromises wound healing by disrupting epidermal barrier function [Publisher: Wiley Online Library]. *The Journal of pathology*, *233*(4), 331–343.

Sen, C. K., Ghatak, S., Gnyawali, S. C., Roy, S., & Gordillo, G. M. (2016). Cutaneous imaging technologies in acute burn and chronic wound care [Publisher: NIH Public Access]. *Plastic and reconstructive surgery*, *138*(3 Suppl), 119S.

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis [Publisher: Annual Reviews]. *Annual review of biomedical engineering*, *19*, 221–248.

Shiraishi, J., Li, Q., Appelbaum, D., & Doi, K. (2011). Computer-aided diagnosis and artificial intelligence in clinical imaging [Issue: 6]. *Seminars in nuclear medicine*, *41*, 449–462.

Shuai, B., Liu, T., & Wang, G. (2016). Improving fully convolution network for semantic segmentation. *arXiv preprint arXiv:1611.08986*.

Siddique, N., Sidike, P., Elkin, C., & Devabhaktuni, V. (2020). U-Net and its variants for medical image segmentation: Theory and applications. *arXiv preprint arXiv:2011.01118*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singla, N., Srivastava, V., & Mehta, D. S. (2018). In vivo classification of human skin burns using machine learning and quantitative features captured by optical coherence tomography [Publisher: IOP Publishing]. *Laser Physics Letters*, *15*(2), 025601.

Sisó-Almirall, A., Kostov, B., Navarro González, M., Cararach Salami, D., Pérez Jiménez, A., Gilabert Solé, R., Bru Saumell, C., Donoso Bach, L., Villalta Martí, M., González-de Paz, L., et al. (2017). Abdominal aortic aneurysm screening program using hand-held ultrasound in primary healthcare [Publisher: Public Library of Science San Francisco, CA USA]. *PLoS One*, *12*(4), e0176877.

Souly, N., Spampinato, C., & Shah, M. (2017). Semi supervised semantic segmentation using generative adversarial network. *Proceedings of the IEEE international conference on computer vision*, 5688–5696.

Stephan, M., & Santra, A. (2019). Radar-based human target detection using deep residual u-net for smart home applications. *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 175–182.

Sumithra, R., Suhil, M., & Guru, D. (2015). Segmentation and classification of skin lesions for disease diagnosis [Publisher: Elsevier]. *Procedia Computer Science*, *45*, 76–85.

SuperAnnotate. (2020). https://superannotate.com/download?utm_term=annotate%20image&utm_campaign=Traffic-Download-Free+Annotation+Tool&utm_source=adwords&utm_medium=ppc&hsa_acc=7527629942&hsa_cam=11645645186&hsa_grp=118910718208&hsa_ad=480447910704&hsa_src=g&hsa_tgt=kwd-538080331152&hsa_kw=annotate%20image&hsa_mt=b&hsa_net=adwords&hsa_ver=3&gclid=Cj0KCQiAqdP9BRDVARIsAGSZ8AlIDYq
h1psC03klQFo5Q2wCYCaU7uGoaAva-EALw_wcB

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.

Taghanaki, S. A., Abhishek, K., Cohen, J. P., Cohen-Adad, J., & Hamarneh, G. (2021). Deep semantic segmentation of natural and medical images: A review [Publisher: Springer]. *Artificial Intelligence Review*, *54*(1), 137–178.

Ţălu, Ş. (2012). Texture analysis methods for the characterisation of biological and medical images [Publisher: Bioflux SRL]. *Extreme Life, Biospeology and Astrobiology*, *4*(1), 8–12.

Tam, C. (2020). Machine Learning towards General Medical Image Segmentation.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *International conference on artificial neural networks*, 270–279.

Tantray, M. D., Rather, A., Manaan, Q., Andleeb, I., Mohammad, M., & Gull, Y. (2018). Role of ultrasound in detection of radiolucent foreign bodies in extremities [Publisher: Springer]. *Strategies in Trauma and Limb Reconstruction*, *13*(2), 81–85.

Thatcher, J. E., Squiers, J. J., Kanick, S. C., King, D. R., Lu, Y., Wang, Y., Mohan, R., Sellke, E. W., & DiMaio, J. M. (2016). Imaging techniques for clinical burn assessment with a focus on multispectral imaging [Publisher: Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA]. *Advances in wound care*, *5*(8), 360–378.

Tran, H. S., Le, T. H., & Nguyen, T. T. (2016). The degree of skin burns images recognition using convolutional neural network. *Indian J. Sci. Technol*, *9*(45), 1–6.

Treml, M., Arjona-Medina, J., Unterthiner, T., Durgesh, R., Friedmann, F., Schuberth, P., Mayr, A., Heusel, M., Hofmarcher, M., Widrich, M., et al. (2016). Speeding up semantic segmentation for autonomous driving [Issue: 7]. *MLITS, NIPS Workshop*, *2*.

Usha, R., & Perumal, K. (2019). SVM classification of brain images from MRI scans using morphological transformation and GLCM texture features [Publisher: Inderscience Publishers (IEL)]. *International journal of computational systems engineering*, *5*(1), 18–23.

Wada, K. (2016). *Labelme: Image Polygonal Annotation with Python*. https://github.com/wkentaro/labelme

Wang, C., Yan, X., Smith, M., Kochhar, K., Rubin, M., Warren, S. M., Wrobel, J., & Lee, H. (2015). A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. *2015 37th annual international conference of the ieee engineering in medicine and biology society (EMBC)*, 2415–2418.

Wang, W., Liang, D., Chen, Q., Iwamoto, Y., Han, X.-H., Zhang, Q., Hu, H., Lin, L., & Chen, Y.-W. (2020). Medical image classification using deep learning. *Deep Learning in Healthcare* (pp. 33–51). Springer.

Wang, X., You, S., Li, X., & Ma, H. (2018). Weakly-supervised semantic segmentation by iteratively mining common object features. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1354–1362.

Warby, R., & Maani, C. V. (2019). Burns Classification.

Wechsler, H. (1980). Texture analysis—a survey [Publisher: Elsevier]. *Signal Processing*, *2*(3), 271–282.

Wei, L., Gan, Q., & Ji, T. (2017). Cervical cancer histology image identification method based on texture and lesion area features [Publisher: Taylor & Francis]. *Computer Assisted Surgery*, *22*(sup1), 186–199.

Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., & Huang, T. S. (2018). Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7268–7277.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning [Publisher: SpringerOpen]. *Journal of Big data*, *3*(1), 1–40.

World health organization. (2018). https://www.who.int/news-room/fact-sheets/detail/burns

Xiancheng, W., Wei, L., Bingyi, M., He, J., Jiang, Z., Xu, W., Ji, Z., Hong, G., & Zhaomeng, S. (2018). Retina blood vessel segmentation using a U-net based Convolutional neural network. *Procedia Comput Sci*, 1–11.

Xu, S. S.-D., Chang, C.-C., Su, C.-T., & Phu, P. Q. (2019). Classification of liver diseases based on ultrasound image texture features [Publisher: Multidisciplinary Digital Publishing Institute]. *Applied Sciences*, *9*(2), 342.

Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis [Publisher: Springer]. *Journal of Big Data*, *6*(1), 1–18.

Yang, F., & Jiang, T. (2003). Pixon-based image segmentation with Markov random fields [Publisher: IEEE]. *IEEE Transactions on Image Processing*, *12*(12), 1552–1559.

Yang, X., Li, X., Ye, Y., Lau, R. Y., Zhang, X., & Huang, X. (2019). Road detection and centerline extraction via deep recurrent convolutional neural network U-Net [Publisher: IEEE]. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(9), 7209–7220.

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion [Publisher: Elsevier]. *Array*, *3*, 100004.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer.

Ziegler, L., & O'Brien, R. T. (2002). Harmonic ultrasound: A review [Publisher: Wiley Online Library]. *Veterinary Radiology & Ultrasound*, *43*(6), 501–509.

Zuo, K. J., Medina, A., & Tredget, E. E. (2017). Important developments in burn care [Publisher: LWW]. *Plastic and reconstructive surgery*, *139*(1), 120e–138e.

VITA

Daniela Chanci Arrubla

School of Industrial Engineering, Purdue University

Education

B.S., Biomedical Engineering, 2018, Universidad EIA, Medellin, Colombia.

M.S., Industrial Engineering, 2020, Purdue University, West Lafayette, Indiana, United States.

Research Interests

Medical Image Analysis, Signal and Image Processing, Computer-Aided Diagnosis, Computer Vision, Machine Learning.

PUBLICATIONS

1. J. Barragan, **D. Chanci**, D. Yu, and J. Wachs, "SACHETS: Semiautonomous Cognitive Hybrid Emergency Teleoperated Suction.," presented at the *International Symposium on Robot and Human Interactive Communication.*

2. M. S. de Medeiros, D. Goswami, **D. Chanci**, C. Moreno, and R. V. Martinez, "Washable, breathable, and stretchable e-textiles wirelessly powered by omniphobic silk-based coils," *Nano Energy, p. 106155, 2021.*

3. **D. Chanci**, N. Madapana, G. Gonzalez, and J. Wachs, "Correlation Between Gestures' Qualitative Properties and Usa-bility metrics," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2020, vol. 64, no. 1, pp. 726–730.*

4. M. S. de Medeiros, **D. Chanci**, and R. V. Martinez, "Moisture-insensitive, self-powered paper-based flexible electronics," *Nano Energy, vol. 78, p. 105301, 2020.*

5. M. Sala de Medeiros, **D. Chanci**, C. Moreno, D. Goswami, and R. V. Martinez, "Waterproof, breathable, and antibacterial self-powered e-textiles based on omniphobic triboelectric nanogenerators," *Advanced Functional Materials, vol. 29, no. 42, p. 1904350, 2019.*