Contents

List of Figures

List of Tables

List of Abbreviations

2D 2-Dimensional

3D 3-Dimensional

AI Artificial Intelligence

ANN Artificial Neural Networks

CNN Convolutional Neural Network

CW Chronic Wounds

DFU Diabetic Foot Ulcers

DL Deep Learning

DM Diabetes Mellitus

FC Fully Connected

FCN Fully Convolutional Network

FCNs Fully Convolutional Networks

FN False Negative

FP False Positive

IoU Intersection over Union

ML Machine Learning

NN Neural Network

ROI Region of Interests

SVM Support Vector Machines

TP True Positive

1 Introduction

1.1 Medical Context

.....

1.2 Motivation and Problem Statement

n recent years, the use of smartphones and imaging technology in daily clinical practice, especially towards wound and DFU assessment has increased considerably. Clinicians can obtain additional information about the wound characteristics from digital image processing to improve diagnostic accuracy.

1.3 Aim and Objectives

.....

1.4 Thesis Contributions

.....

1.5 Thesis Organisation

.....

2 Background of Diabetic Foot Ulcers

2.1 Definition and Complications of Diabetes

Diabetes mellitus (DM), known as diabetes, is a serious and long-term health issue characterized by a disorder of metabolism resulting in high levels of glucose in the blood called hyperglycemia. In healthy individuals, blood glucose level should be in the range of 70 and 120 milligrams per deciliter. Hyperglycemia take place when the glucose value is greater than 140 milligrams per deciliter []. That occurs when the pancreas cannot produce any or enough insulin or when the body cannot effectively use the insulin it produces.

Insulin is a hormone produced by a large gland behind the stomach called pancreas and more precisely by the beta cells. This hormone is essential for the metabolism to transport glucose from the bloodstream into the body's cells where the glucose is converted to be used for energy. During digestion, all carbohydrates are broken down into glucose in the blood. When this happens, the amount of glucose in the blood rises. Since our body is designed to keep the level of glucose in your blood constant, beta cells trigger the pancreas to release more insulin into the bloodstream. Insulin must be present with the right amount to help glucose get into the cells. Once in the cells, glucose can be converted immediately into energy or stored to be used later. As glucose moves from the bloodstream into the cells, blood sugar levels start to drop. In people with diabetes, without enough insulin, glucose can't move from the bloodstream into the cells. Consequently, blood glucose level stays high and leads to hyperglycemia.

Diabetes is one of the biggest global health emergencies of the twenty-first century that has reached alarming levels. Nowadays, diabetes is affecting nearly half a billion people worldwide. According to the most recent 9th edition of the International Diabetes Federation (IDF) Diabetes Atlas report of 2019, more than 463 million people are actually living with DM, compared to 151 million in 2000 [1]. These large numbers impose a heavy economic burden on public health that drains global healthcare budgets due to high medical costs, productivity loss and premature mortality. Moreover, the increase in the economic burden of this epidemic is expected to continue to grow. The World Health

Organization (WHO) is estimating that the global number of diabetes prevalence will rise to reach 700 million by 2045. Despite the economic burden, diabetes also imposes a significant social impact on individuals and their families due to its chronic nature. Although diabetes has no cure, maintaining an active and healthy lifestyle can reduce the impact of diabetes. In addition, doctors can prescribe if needed some medications in form of drugs or injections that replace the hormone in the body to help lower blood sugar.

There are three main categories of diabetes which include type 1, type 2, and gestational. The first category occurs most frequently in children and adolescents due to an autoimmune condition where the immune system mistakenly attacks the beta cells in the pancreas [2]. In the second category, type 2 diabetes occurs when the pancreas does not produce enough insulin or when cells respond poorly to the produced insulin. It is commonly seen in adults but it is increasingly seen in adolescents with unhealthy lifestyle and physical inactivity. The final category is gestational diabetes that occurs only during pregnancy and it is caused by insulin resistance due to hormones produced by the placenta [3]. In this work, we will focus only on type 2 diabetes, which is by far the vast majority accounting for around 90 to 95% of all diabetes cases worldwide.

Over the long-term, diabetes in all forms cause serious health problems. Untreated high blood sugar affects large and small arterial blood vessels which can damage several body's organs especially heart, eyes, kidneys, foot, etc. Leading to major life-threatening complications such as cardiovascular diseases, kidney failure, blindness and foot problems.

2.2 Diabetic Foot Ulcers

Diabetic foot ulcers are the most serious complication of uncontrolled diabetes [4]. When left untreated, can lead to subsequent infections and may end in lower limb amputation [5]. The lifetime incidence of diabetic people that will develop foot ulcers at some point in their lives is around 19% and 34% [6] resulting in frequent hospitalizations. These ulcers are responsible for more admissions than any other diabetes's complications. DFUs management is often challenging because high glucose levels slow down the healing process which results in prolonged stays in hospital. Consequently, a full recovery of these lesions requires several weeks, months or even years in serious cases. Moreover, 60% of patients will develop a recurrent ulcer within three years of ulcer healing diabetic [7]. Furthermore, foot ulcers increase medical care costs and have a high burden of healthcare professionals and facilities. In addition, it is a source of major suffering for

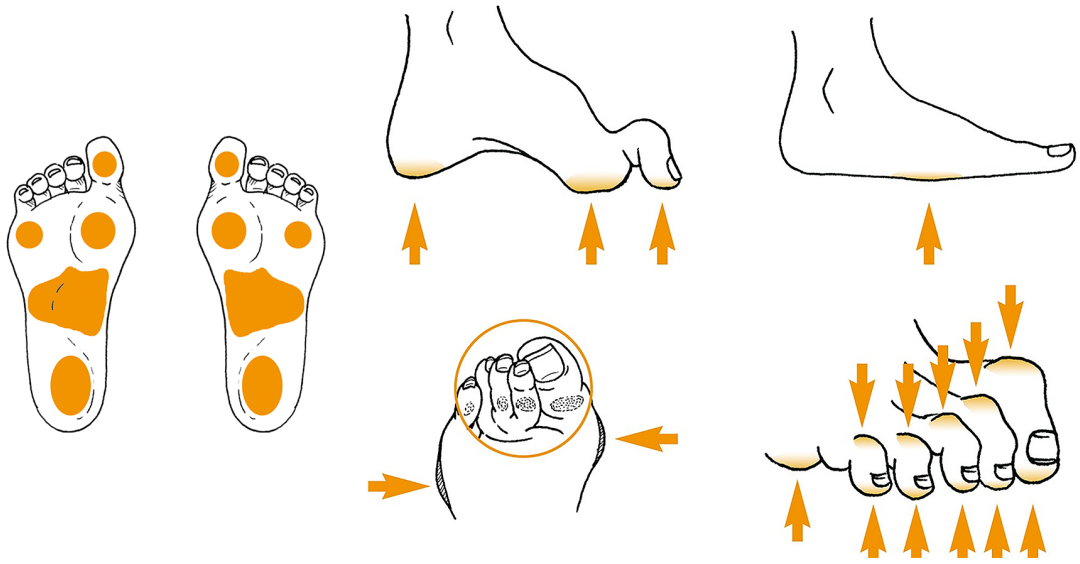

Figure 2.1: Usual locations of ulcers in the diabetic foot

patients, affecting their mobility and lowering quality of life.

Mostly located on areas at risk of the foot such as the bottom of the foot or along the top and bottom of toes (see Figure 2.1), these ulcers does not appear overnight. DFUs consist of lesions in the deep tissues that occur in gradual way and have a tendency to worsen over time if not well managed. They start small sores which gradually cause skin breakdown and precipitate ulceration. The risk of foot ulceration increases with age, poor glycemic control, duration of diabetes and smoking.

By far, the two primary underlying risk factors of foot ulcers are diabetic peripheral neuropathy (DPN) and peripheral artery disease (PAD) (see Figure 2.2). Peripheral diabetic neuropathy is the most common complication of elevated blood glucose levels over time. The global prevalence of DPN is estimated between 16% and 87% among adults with diabetes citeidf2019. It's characterized by a reduced or complete loss of sensation to pain in the lower extremities due to nerve damage. The nerve damage leads to an insensitive foot and allows minor injuries to go unnoticed which facilitate the development of chronic ulcers if undetected at the time. Poor blood circulation due to arterial diseases can also make it more difficult for ulcers to heal. Insufficient blood flow increases the risk of infection and necrosis. The result is a purely painful and ischemic foot [8]. However, most foot ulcers are of mixed etiology neuropathy and ischemia resulting in neuroischemic ulcer. Neuropathic ulcers heal withing a period of 20 weeks while neuroischemic ones take longer and end more often with an amputation [9].

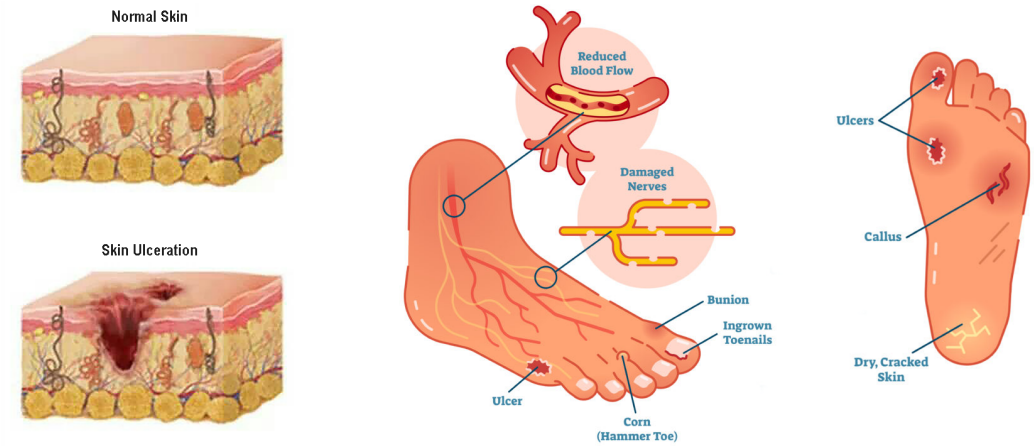

Figure 2.2: Risk factors of diabetic foot ulcer

However, once the diabetic foot ulcer is suspected, a fast treatment and an effective management are vital to avoid serious complications. The faster the healing is, the less chance for an amputation. It is expected that 85% of amputations can be avoided when an effective healing plan is adopted [10]. A delay in diagnosis results in higher morbidity and mortality due to increased amputation rate. The key element for a successful diabetic foot management is periodic visits to healthcare centers for a regular examination by experienced health professionals.

2.3 DFU in clinical practice

Diabetic foot ulcers require a systematic approach to provide an optimal intervention. Therefore, clinical assessment of ulcers is a complex process that helps to perform a successful diagnosis and to determine the progression towards healing. Monitoring the process of healing over time is primordial to develop a proper management plan and to select an efficient treatment, thereby assuring quality wound care.

Knowledge and fast decision making are crucial all along wound healing process. This process requires experienced clinicians that demonstrate high-level clinical decision-making skills. The standard approach of DFU assessment is typically based on visual examination. In current clinical practice, healthcare professionals often rely on visual observation and manual measurements to judge the healing progress of the wound. For an appropriate diagnosis, is necessary to quantify wound area variation and tissue composition as well [11].

2.3.1 Wound Measurement

Wound measurement is therefore an integral part in the assessment process. It provides a baseline from which clinicians can determine the progression of the ulcer according to its physical dimension. The measurements of the wound mainly include length, width, surface area, depth and volume. Many traditional methods exist of varying degrees of sophistication that can be divided in three main categories:

- One-dimensional based method: It is a quick and easy way to quantify wound perimeter that can be performed using a simple ruler or caliper. The most straightforward method consists of measuring the greatest perpendicular length and width. Then, the clinician multiplies these two linear dimensions to obtain the surface area [12]. However, this consider that the wound has a rectangular or squared shape and does not take into account wound irregularities [13]. Thus, the surface is usually overestimated by 44% compared the true area [14]. Another drawback is that each clinician may have his own intuitive way of choosing the exact points to measure even for the same wound. In addition, wounds with different shapes can still have the same dimensions, thus yielding the same estimated areas.

- Two-dimensional based method: It refers to measuring the surface area by tracing the ulcer boundaries onto a gridded transparent sheet placed over the wound bed with a permanent pen [15]. Then clinicians estimate the area by counting the number of squares within the wound edge based on the metric grid. This planimetric method is slightly more accurate and reliable than the ruler. Furthermore, the tracing process involves direct contact with the wound that can carry a high risk of infection and contamination and can also be time consuming.

- Three-dimensional based method: The three-dimensional measurement of the ulcer aims to evaluate the healing status from a volume perspective by quantifying the volume and depth. This technique is much more complicated than estimating the surface or perimeter. Traditionally, the volume is measured using a moldable material or saline gel injection inside the wound cavity [16] [15]. The required volume of the used material to fill the wound can theoretically give a good estimate of its real volume [17]. Regarding depth, it can be estimated by inserting a sterile cotton tipped swab in the deepest part of the wound. However, the volume measurement techniques are often complex for practitioners to handle and can be painful for the patient.

The major disadvantage of these traditional methods is that the measurements are subjective and imprecise. In addition, they are invasive and require direct contact with the wound which may lead to severe infection. Furthermore, there is a need for a non-

contact and objective technique for an accurate DFU measurement.

2.3.2 Tissue Analysis

Analyzing tissue composition is necessary thorough DFU assessment. This method is widely used in clinical settings to evaluate the condition of the wound and the healing progress. In daily practice, clinicians quantify the proportion of each tissue within the ulcer site in terms of percentages. This is done based on visual inspection using the Wound Healing Continuum. It's a color-based continuum utilizing the Black-Yellow-Red scheme. Where each tissue corresponds to a healing stage and shows a specific color. Most researchers divide these tissues into three main types: black necrosis, yellow slough and red granulation.

Necrotic tissue refers to dead skin cells that appears in a dark color mainly black. Its texture can be dry and hard or soft and wet. Necrosis in diabetes subjects lead to severe infection which slows down the healing process. An extensive presence of infected necrotic tissue is the main cause of lower limb amputation [18]. In most cases, necrotic tissue should be removed to facilitate the healing of DFU. This process is called debridement in clinical settings [19].

Slough is a yellow-colored infective tissue inside the wound bed, mostly composed of fibrin. It generally has a moist and soft texture. The color can sometimes vary from yellow, tan, green due to infection and hydration. A DFU typically cannot heal effectively with the presence of sloughy cells. Therefore, a debridement is required to allow new tissue to form and cover the wound bed [20].

Granulation tissue refers to new tissue that is created when the wound is healing properly. Healthy granulation tissue will be light red, while unhealthy granulation will appear in dusky red which indicate the presence of wound infection [20]. The growth of granulation tissue indicates that the body is generating a protective layer of flesh which is rich in small blood capillaries reflecting the red color of the tissue. A large amount of granulation denotes that the wound is recovering successfully. The final phase of wound healing is the appearance of a pink epithelial tissue. This tissue slowly grows over the granulating tissue providing a protective layer over the wound surface. Once created, the function of healthy skin is progressively restored in time.

Understanding color and texture of tissues help to establish most appropriate wound management plan as they progress throughout the healing process. Unfortunately, clinicians can occasionally struggle to estimate the exact proportion of each tissue within wound area. As with all assessment methods performed based on the naked human

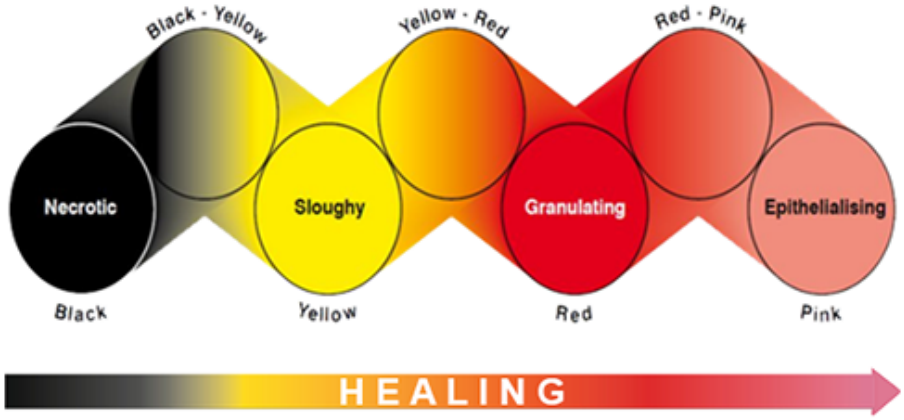

Figure 2.3: The wound healing continuum (WHC) [21]

eye, there is a high risk of subjectivity [22] since such evaluation depends on the experience and expertise of the healthcare professional. It is reported that the intra- and inter-expert variability of tissue identification is about 70% [23]. Moreover, DFUs heal gradually, so the detection of slight changes of the ulcer is often challenging even for experienced clinicians. Thus, tissue identification is quite challenging and difficult to perform in clinical settings [24]. Furthermore, this method is subjective, imprecise and expert-intensive. Hence, imaging techniques are essential to provide objective and precise tissue identification for a more accurate diabetic foot ulcer assessment.

2.4 Image-Based assessment systems

n recent years, the use of medical imaging technology for automatic chronic wounds assessment has increased considerably to become a common practice in clinical settings. It changed radically wound care practices. Imaging-based assessment does not require contact with the skin which reduce the risk of infection and contamination. Moreover, photography is particularly useful to ensure an objective and less time-consuming assessment for health professionals and patients. Following this trend, many commercial wound care devices are often used in medical environment, we stand for Aranz Silhouette [25], eKare Insight [26], MAVIS [27] and Wound Zoom by Stevens Point [28].

However, these medical devices are expensive and their use require a special training. Considering the emergence of low-cost smartphones, wound management by image analysis became an attractive option for clinicians. Modern smartphones have now more processing power and brilliant camera capabilities with high resolution. Due to advances

in miniaturization, they represent small computing devices especially with the help of advanced mobile internet and cloud computing.

In the past years, smartphones were mostly used for data recording to feed patient folder [29]. Nowadays, they are more often used as tracking devices to follow healing status of chronic wounds over time. Therefore, the efficiency of treatments and the speed of the healing process could be assessed and with the aid of clinicians more adapted therapy could be proposed. Thus, hospital stays will be reduced as well as healthcare expenses. Moreover, automatic assessment can help to enhance accuracy of wound assessment while reducing inter and intra-observer variability. This is particularly important when different healthcare practitioners take care sequentially of the same patient [30]. Clinicians can access to the wound healing status anywhere and anytime [31]. However, embedded image processing methods and computer vision techniques are required to perform automatic assessment using a smartphone.

Automated smartphone based wound care management has been of interest among the research community during last years. Many image-based mobile applications have been developed for automatic chronic wounds healing assessment including DFU [32]. Manuel Dujovny [33] designed a mobile application called "MOWA" which can extract ulcer size and wound edge dimensions. In a similar work, Hettiarachchi et al. [34] developed an android application for wound measurement and healing monitoring. These two methods require nurse intervention to draw manually the wound edge. In a more recent work, Yap et al. [35] developed an application called "FootSnap" to create a standard DFU database by capturing images of the sole of diabetic feet. Then, they extended their application to automatic detection and localization of the ulcers [36]. Nevertheless, these apps were developed specifically for wound measurement and detection. In another hand, Wang et al. [37] developed a semi-automatic user interface including wound measurements and tissue composition. Though the results showed high effectiveness and usability; their method is user-centered. The edge of the wound requires a freehand drawing of the initial contour by the nurse. Therefore, not a single system combines fully automatic wound segmentation and tissue analysis. Furthermore, the development of such a system would be rather complex with requirements to automatically detect the wound and extract the ROI from the background, then classify the tissues inside the wound bed, in the end produce wound measurements and the percentage of detected tissues.

3 Literature Review

3.1 Wound Segmentation

A critical first step in the development of intelligent systems for DFU management is wound detection and segmentation. Automatic detection of lesions in images is a key part of diagnosis that influences the outcome of the entire assessment process. Furthermore, it is often an important step in image-based DFU assessment pipeline. Typically, this task aims to delineate the regions of interest from the raw image so that the subsequent ulcer pixels are separated from the background and the other parts of the body which allows further quantitative analysis of clinical parameters such as wound physical measurements (perimeters, shape, area, ...), tissue analysis, 3D reconstruction, etc. Automatic wound detection and segmentation has been a hot topic in the field of medical imaging for many years. Several image processing studies have addressed wound segmentation using different approaches.

3.1.1 Traditional Image Processing Methods

At the very beginning, researchers performed wound segmentation using a broad bunch of traditional image processing algorithms. There exist two major segmentation approaches: edge based and region based. The main idea behind edge-based methods is extracting the wound boundary while region-based methods aim to detect common properties of the wound area that distinguish it from skin.

Mukherjee et al. [38] proposed a wound segmentation framework using fuzzy algorithm. The RGB images were transformed into HSI (hue, saturation, and intensity) color space and subsequently they applied the fuzzy divergence algorithm to the S channel of the images in the HIS color space for wound segmentation. Yadav et al. [39] segmented chronic wound areas by using K-means and Fuzzy C-means clustering algorithms. The experiments on a dataset of 77 images showed that Fuzzy C-means provided better results than K-means clustering. A different approach was proposed in [40], where the authors performed wound analysis using color segmentation by applying accelerated mean shift algorithm. In a similar approach, Wang et al. [41] proposed a smartphone

based wound analysis system for patients with diabetes. Ulcer segmentation were performed with the accelerated mean-shift algorithm to outline the wound based on skin color. However wound images were captured using an image capture box which was not practical in clinical settings.

The authors in [42] presented a methodology to segment wound area using color-based image analysis algorithms. They analyzed the relative color difference between the wound and the surrounding skin. Their methods consist on several steps. First, ROI is selected manually from the input images so that the wound occupies most of the cropped image, then a rectangle is selected from the center of the wound in the ROI that will be considered as Wound Reference (WR). After that, a polygon is drawn on the surrounding skin without including any part of the wound. All combinations of RGB color intensities from pixels outside this polygon are considered as Non-Wound Reference (NWR). Next, every pixel inside the polygon is compared with the (NWR) list using pixel color comparison algorithm (PCC) and only those pixels that have color combinations different from the list are classified as wound pixels. This method is not suitable for wounds that are small or nearly healed and it requires a lot of user interaction.

In 2017, Chang et al. [43] presented an interactive multi-modal system for real-time wound assessment including multiple imaging modalities: RGB, 3D depth, thermal, multi-spectral, and chemical sensing. The RGB images were used to segment wound pixels from the normal skin using GrabCut algorithm. First, a bounding box surrounding the wound is defined then two polylines are defined one for wound region and the second one for normal skin. The system then constructs an initial wound and background model using mean-shift. Finally, a graph partitioning algorithm calculates the wound segmentation mask. This procedure is repeated several times until a satisfactory segmentation is obtained. The results show promising results but at the expense of high user interaction. In a different approach based on contour detection[44], the authors designed a novel semi-automatic wound segmentation system. Seven segmentation algorithms were compared including edge-detection, region-growing, Livewire, active contours, and texture segmentation. A preprocessing step was required to reduce camera noise and normalize image intensity. The results showed that snake provide the higher accuracy but Livewire achieved the best performance. However, this method is highly interactive and its performance relies on the seed point provided by the user.

Segmentation accuracy of the above methods depend on many factors. Region-based methods depend on pixel color, texture and image intensity, while edge-based methods depend on the initial contour or the chosen seed point. Mainly, classical segmentation techniques are sensitive to noise and could not be applied to images with smooth tran-

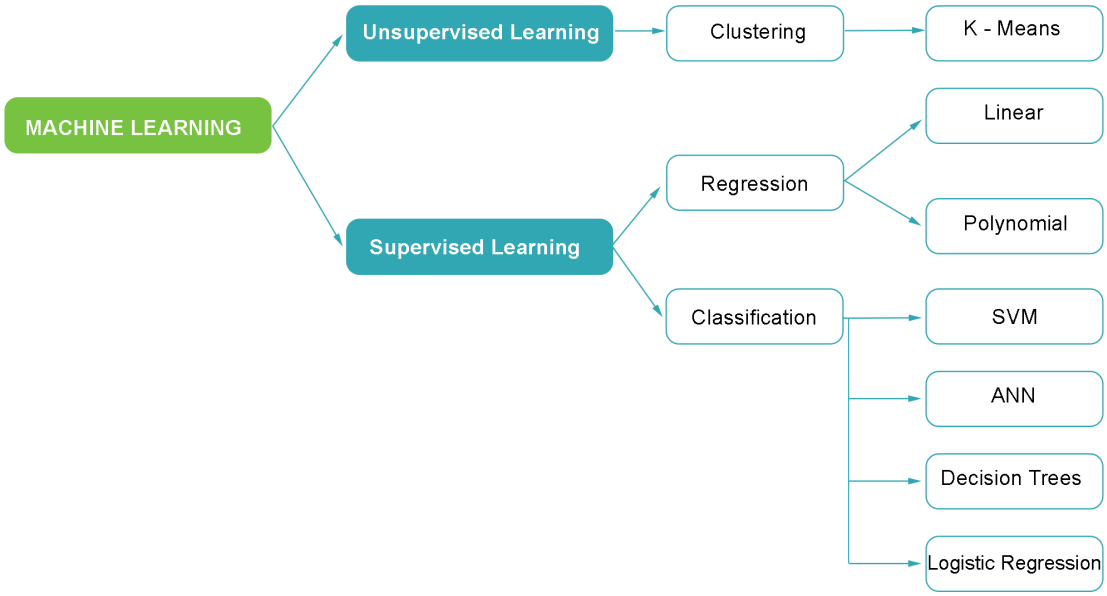

Figure 3.1: Types of Machine Learning algorithms

sitions or intensity variation, which may lead to over or under segmentation [45]. Thus, a preprocessing step including color calibration and noise removal are highly recommended. Moreover, it's difficult to distinguish between the wound and the background for low contrast images or between the shading of the real images. Hence, it is difficult to obtain accurate results for wound image segmentation using traditional image processing algorithms. In addition, these methods are time consuming, needs user interaction and did not provide an end-to-end processing.

3.1.2 ML Based Methods

As technology advanced, researchers moved from simple image processing methods to more advanced machine learning methods. ML is becoming an increasingly popular approach in medical image segmentation and have already shown promising accuracy and reliability.

Traditional ML algorithms are mainly divided into supervised and unsupervised learning (See Figure 3.1. DFU detection or wound detection in general, is usually performed by supervised ML. From a given training set, the algorithm extracts some interesting information from the input images corresponding to features. Most of feature extraction algorithms focus on detecting specific image properties such as texture, color, and shape [46]. Some popular feature descriptors include Local Binary Patterns (LBP) [47], His-

togram of Oriented Gradients (HOG) [48], Hough transform [49] and normalized RGB, L*u*v and HSV features using color histograms. These descriptors transform the input image into a set of features called feature vector which correspond to the higher-level representation of data in the given image. Then, a classifier is trained with these extracted feature vectors for classification of each input image into its corresponding label class. In the case of DFU detection, the output class labels will be ulcer and non-ulcer. According to Jiang et al. [50], the most popular supervised algorithm is support vector machine (SVM) [51]. Given a set of points, a linear SVM identify the optimal hyperplane leaving the largest separation margin possible between two classes by regrouping the points of the same class on the same side, while maximizing the distance of either class.

In [52] and [53], Kolesnik et al. employ an SVM classifier to differentiate between skin and wound tissue using combined color- and texture feature space with 9 color features and one textural feature. The classifier was trained on a very small dataset of six wound images using three different kernels. Radial kernel showed the best performance followed by the polynomial and linear one. The experimental results show that the combination of color and texture features reduces segmentation error as compared to the use of only color features. The performance of the proposed system is too sensitive to natural brightness variations and illumination conditions during acquisition of images. The obtained wound boundary is not precise and does not match with manual boundary of a clinician. Thus, even with both color and texture feature SVM does not guarantee a good boundary around wound. To compensate this drawback, the authors [54] tested the capability of deformable snake adjustment algorithm to refine SVM-generated wound boundary. SVM-generated contour is given as initial contour to perform snake adjustment. Snake refinement is only successful if the initial SVM contour lies in the vicinity of the true wound boundary.

In 2016, Wang et al. [55] proposed a cascaded two-stage classifier to determine the DFU boundary on images captured with a capture box with a pair of straight angled mirrors which reflect the image back to the camera, that help to provide controlled lighting, angle and range conditions. The two-stage classifier is trained in several steps. First, SLIC is used to divide training images into superpixels, then significant descriptors are extracted from these superpixels. Color, Local Dense SIFT (DSIFT) and Bag-of-Word (BoW) features were used as descriptors in the first stage, while color and wavelet based features were used in the second stage. The first stage aims to eliminate irrelevant regions, a set of binary SVM classifiers are trained and applied to the extracted superpixels to collect a set of incorrectly classified instances. Then the second stage is

used to distinguish wound region from healthy skin by training another binary classifier on the incorrectly classified set. Finally, wound contour was refined using Conditional Random Field (CRF). The proposed method is novel but is limited to DFUs that are present in the plantar foot surface and the use of the capture box is impractical in clinical settings.

In a more recent work [56], the authors performed wound segmentation using a superpixel based SVM. They first perform superpixel segmentation for each input image, 100 images were divided into superpixels using SLIC to obtain a total of 47,718 superpixel regions. Then, color and texture feature descriptors were generated form all generated superpixels using color histogram and LBP. The extracted features were used to perform SVM Classification in two stages for each superpixel; first one for background-skin classification and second one for wound-skin classification. Jaccard index was around 72% for a small test set of 10 images.

In general, the current studies have performed the segmentation task by extracting manually different features from wound images using texture and color descriptors followed by traditional ML algorithms such as SVM classifiers to categorize skin patches into ulcer and non-ulcer. Nevertheless, these methods have several drawbacks. Firstly, it is necessary to choose which features are important and best describe different classes in the given images. As the number of classes to classify increases, feature extraction becomes more and more unwieldy. Secondly, the handcrafted features can greatly be influenced by lighting conditions, illumination, image resolution and also skin shape and shades. In most cases, a color correction step is required to reduce color effects due to uncontrolled lighting conditions. Color calibration is often performed by inserting a color pattern in the field of view during acquisitions. Moreover, choosing an appropriate set of hyperparameters is crucial for machine learning models accuracy. This process involves a careful fine-tuning of the parameters, a domain expert has to step in and make the necessary adjustments which make them hard to implement in clinical settings. In practical terms, although their performance, basic ML methods are not robust enough due to their reliance on the handcrafted features with certain assumptions. Furthermore, an accurate segmentation cannot be achieved automatically using traditional machine learning approaches.

3.1.3 DL Based Methods

A more recent subset of ML known as Deep Learning (DL) has introduced the concept of end-to-end learning. DL is considered as the new backbone of artificial intelligence

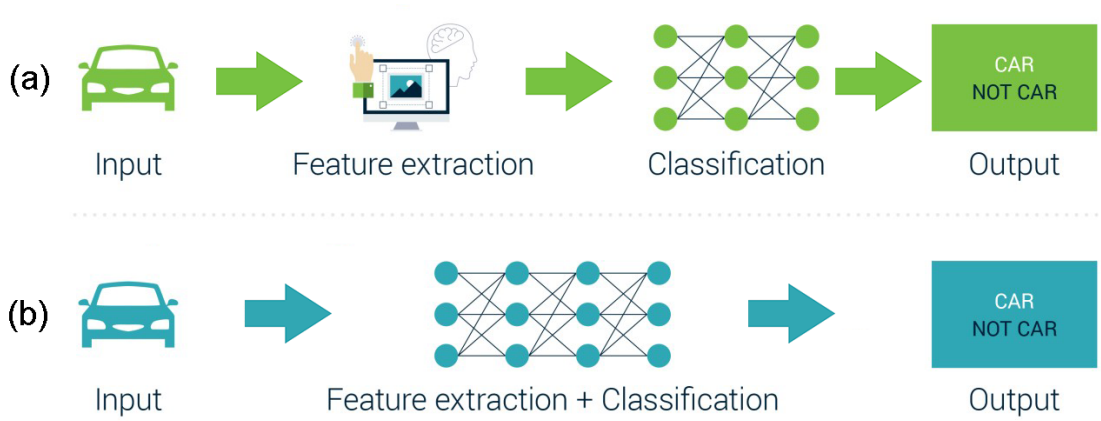

Figure 3.2: Comparison of ML workflow (a) against DL workflow (b)

that fuels all sorts of computer vision tasks that span across a wide range of fields.

DL models design is inspired by the functioning of the human brain providing more meaningful and powerful features. They use many artificial neurons, these neurons are connected with each other to form a layered structure called an "artificial neural network" that can learn and make accurate decisions on its own without any human intervention. In DL, feature engineering is done automatically and merged with the classification procedure. Thereby, DL changed dramatically the workflow of ML as shown in Figure 3.2. From a given dataset, DL model automatically discover the most salient and descriptive features of different classes in each input image throughout an iterative training process. Moreover, contrary to traditional ML algorithms which tend to be domain specific, DL has superior flexibility since CNN models can be re-trained using a custom dataset for any other task [57]. All these breakthroughs have led to higher performance compared to the traditional ML methods in various computer vision tasks such as image classification, semantic segmentation, object detection and Localization research fields. Specifically, with enhancements in device capabilities and the increase in computing capabilities using GPUs and cloud computing, as well as the availability of large public databases with a humongous amount of images for training neural networks like ImageNet [58]. More details about DL functioning as well as all different architectures are given in the next chapter.

The development of a such powerful algorithms had a tremendous influence in the medical field because of their accuracy and flexibility. After the success of deep learning in many real-world applications such as image classification, facial recognition, object detection and tracking, Natural language processing, etc. It is currently gaining a lot of

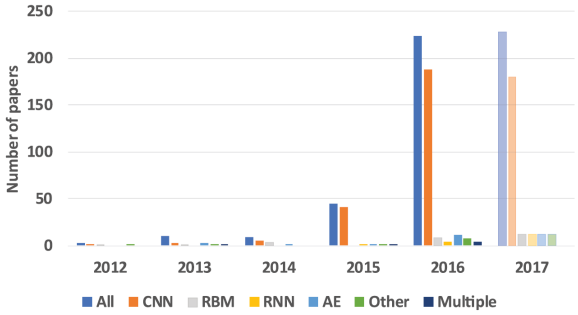

Figure 3.3: Breakdown of the papers included in this survey in year of publication

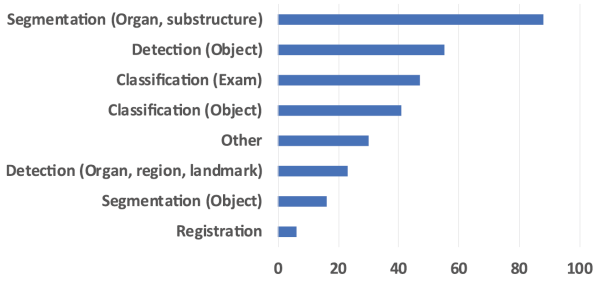

Figure 3.4: task addressed and application area

attention from medical imaging research community. A large number of recent review papers dedicated to deep learning application to medical image analysis were published [59], [60], [61], [62], [63], [64], [65]. Figure 3.3 depicts the rapid surge of interest in deep learning during recent years in terms of the number of papers published in the field of biomedical research that increased more than threefold in 2016 and 2017 compared to 2015. DL have yielded exciting solutions with outstanding performance for biomedical image analysis in various applications such as tumor segmentation, automatic lesion detection, classification of lung diseases, etc. DL models are ideally suited for medical image classification, segmentation, registration and object detection but the segmentation of organs and other substructures was one of the first areas in which deep learning made a major contribution to medical image analysis as shown in Figure 3.4.

Recently, DL has been widely implemented in the medical area to assist DFU and wound diagnosis. In literature, there are three common tasks for DFU image-based inspection using DL: Classification, detection and segmentation as illustrated in Figure 3.5. Ulcer segmentation remains challenging, it combines the challenges of object detection and segmentation.

3.1.3.1 DFU Classification

Several recent studies have been conducted to classify DFU images. Goyal et al [67] implemented a novel fast CNN architecture for DFU classification named as DFUNet. Two classes DFU and healthy skin have been assessed using a small training database of 397 images. Due to the limited size of their training set, the images were cropped to 256x256 patches, then they used data augmentation techniques including rotation, flipping, contrast enhancement and random scaling. The number of patches has increased by 15 times resulting in a total of 22605 patches. The proposed DFUNet is a 14 layers architecture

Figure 3.5: Examples of three common tasks for DFU image analysis using DL [66]

split into three main sections: the initialization layers inspired by GoogLeNet, parallel convolution layers for effective DFU discrimination, fully-connected layers and output classifier. DFUNet outperformed GoogleNet and AlexNet with a F-score of 93%. Some other studies have been carried out in literature to perform the classification task in chronic wound images. For instance, the following works were conducted on CNN-based classification into more than two Types. Shenoy et al. [68] proposed a CNNs method to classify wound images into nine different classes. The authors created a modified version of VGG16 network named WoundNet. The final network is the average of the outcomes from three individual WoundNets. This ensemble is called Deepwound. A dataset of 1335 wound images were pre-processed and augmented. Besides, a mobile application was designed to assist physicians and patients during wound healing evaluation. In a more recent work [69], the authors addressed the classification of wound images into multiple classes based on the type of the wound. A patch classifier with fine-tuned AlexNet architecture was designed to efficiently classify the wound patches.

Nevertheless, among the biggest limitations of DL based models for wound classification is that a large number of training samples is required for appropriate training to attain a good performance. A requirement that is costly and difficult to satisfy in the medical field. To address the issue of lack of training data for DFU classification, other studies adopted hybrid approaches. These approaches combine traditional ML and deep learning and take advantages of both methodologies. DL methods can extract relevant features effectively without any expert intervention. Meanwhile, ML methods can perform the classification task from the extracted features without large amount of data. The fusion of these two techniques offers a better performance on small datasets in comparison with DNNs.

Wang et al. [70] developed new ConvNet architecture to perform wound infection detection. The network was trained on 650 images from the New York University (NYU) database. Learned ConvNet features were used to fed an SVM classifier. Combined SVM classifier trained with CNN features achieved a total accuracy of 84.7%. Goyal et al. [71] used an ensemble CNN approach for combining the outputs of three well known deep networks (ResNet50, InceptionV3, and InceptionResNetV2) and fed it into an SVM for classification. 1459 DFU images were classified into infection or non-infection, and ischemia or non-ischemia classes. Traditional ML methods were compared to DNNs and ensemble CNN. Deep learning-based methods showed a better performance than the traditional classifiers. Furthermore, hybrid ensemble CNN reached the best performance with the accuracy values of 90% for ischemia and 73% for infection predictions. A recent study by Alzubaidi et al. [72] introduced a novel DNN called DFU-QUTNet for automatic classification of DFUs image patches into normal versus abnormal skin. The proposed network is designed based on the idea of increasing the width of the network while keeping the depth. To enhance the performance of their method, they cropped the ROI from 754 foot images to 224x224 patches to collect a total of 1609 patches including both classes. The number of training samples was multiplied by 13 using data augmentation transformations like flipping, rotating, and scaling. By combining SVM with the features extracted by DFU-QUTNet network they achieved a F1-score of 94.5% in classifying DFU images.

In a different approach, another study by the same authors [73] explored the benefit of Transfer Learning (TL) as an alternative solution to improve DNNs accuracy. They adopted a Dataset of 1200 DFU images that were cropped to 224x224 ROI to get a total number of 1477 patches including normal and abnormal skin. The main idea of transfer learning to use a model trained on a large dataset and transfer its knowledge which is basically the learning features to a smaller dataset from the same or a different domain (See Figure: 3.6). In this study, transfer learning utilized ResNet50 architecture in order to benefit from its object detection capabilities. Obtained results has shown that TL can help significantly improve the performance of DL when pretrained on a similar dataset from the same domain of the target dataset.

Seminal works on DFU classification has been primarily based on patch-wise image classification to benefit from the large number of extracted patches to boost the number of samples in datasets. However, the extracted patches do not provide enough insight about the segmentation task.

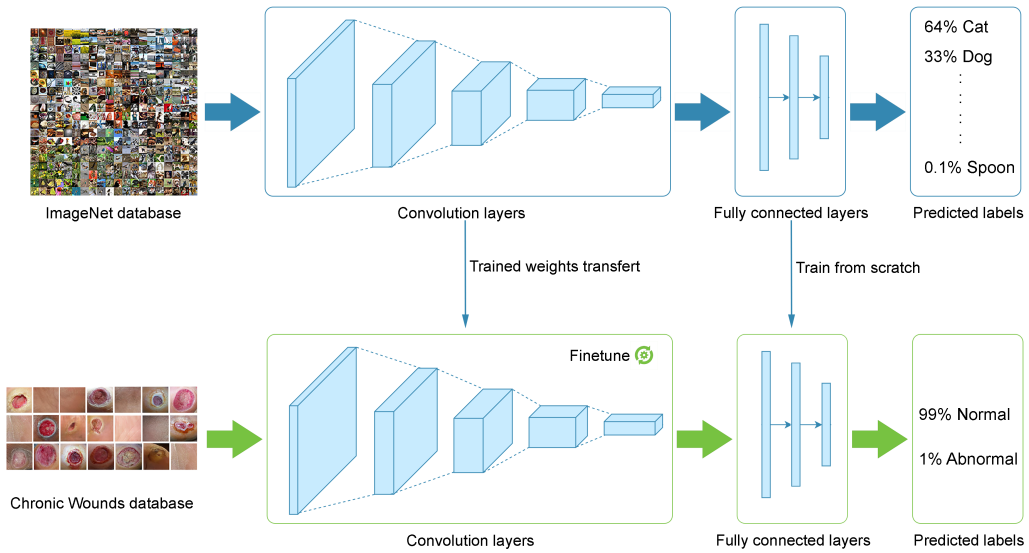

Figure 3.6: The concept of the transfer learning : a neural network is pre-trained on ImageNet dataset and subsequently trained on CWs images for abnormal skin classification.

3.1.3.2 DFU Localization

A series of recent studies has investigated wound and DFU localization using Deep Learning. Goyal et al. [36] performed real-time DFU localization using several deep learning networks for object detection: SSD with MobileNet backbone, SSD with InceptionV2, Faster R-CNN with InceptionV2 and R-FCN with Resnet 101. A large database of 1775 DFU images was considered in this study. Faster R-CNN with InceptionV2 backbone produced the most accurate results. No other evaluation metric was presented for ulcer detection on their dataset. Two studies looked at uses of YOLO for automatic detection and localization of diabetic wounds. Anisuzzaman et al. [74] presented a mobile wound detection method using YOLOv3. A lighter version of YOLOv3 named tiny-YOLOv3 has been used to turn the model into an iOS mobile application. They used AZH database with a total of 1,010 wound images including three types of ulcers: Diabetic foot ulcer, Pressure Ulcer and Venous Ulcer. Additionally, different data augmentation techniques have been considered to increase the number of images on the training set which results in a total of 4,050 image data. Prediction results on 52 test images show that YOLOv3 gives significantly better results compared to SSD model with a mAP value of 93.9%. Similarly, in [75] YOLOv2 with transfer learning was adopted to perform DFU detection. The model was designed using a shuffleNet as backbone of the YOLOv2. Experimental results proved that a combination of shuffleNet and YOLOv2

Figure 3.7: Samples of experts annotations from DFUC dataset

achieved better performance with a maximum localization scores of 0.94 and 0.95 instead of 0.75 and 0.86 for classical YOLOv2 model.

Recently, the emergence of the new DFUC database [76] have significantly promoted the proliferation of experimental practices in the field of automatic DFU localization. DFUC is a result of collaborative work between Manchester Metropolitan University and Lancashire Teaching Hospital and the Manchester University NHS Foundation Trust held in conjunction with the MICCAI 2020 conference. This challenge aims to improve the accuracy of DFU detection in real-world settings. This database consists of 4,000 images including DFU and other foot conditions. All images were annotated with the location of wounds using x_{min}, y_{min}, x_{max} and y_{max} coordinates as illustrated in Figure 3.7.

Several research papers were motivated by the Grand Challenge DFUC 2020. This review paper [77] presents a comprehensive comparison and evaluation of recent DL networks for object detection: Faster R-CNN, three variants of Faster R-CNN, YOLOv3, YOLOv5, EfficientDet and a new Cascade Attention Network on 2000 images from DFUC. All these methods required data augmentation to increase the number of training images. The best performance was obtained by the variant of Faster R-CNN using Deformable Convolution with an F1-Score of 74,34%. Cassidy et al. [76] compared three popular deep learning object detection networks: Faster R-CNN, YOLOv5 and EfficientDet. All networks performed the same but EfficientDet shows a better trade-off between recall and precision, which yields the best F1-Score. The authors in [78] proposed an adapted version of Faster R-CNN for DFU detection. The proposed network is a standard Faster R-CNN but with ResNet-50 as backbone and an IoU sample range of [0:0; 0:5) for negative ROIs. Additionally, 100 total ROIs were used for suggestion, 12 anchors including a scale of 64x64 and data augmentation techniques. Faster R-CNN for DFU detection improved metrics when compared to standard Faster R-CNN to achieve a mAP of 91.4% and F1-score of94.8%. Finally, in the most recent work of Goyal [79], the authors evaluated the performance of EfficientDet to detect DFU in the

DFUC2020 challenge dataset. They utilized the pre-trained weights of networks trained on the MS-COCO dataset, which consists of more than 80,000 images of 90 classes. They further refined the EfficientDet architectures with a score threshold of 0.5 and removed overlapping bounding boxes to minimize the number of false positive and false negative predictions. Color constancy algorithm was applied to cope with noise and lighting effects. Common types of data augmentation techniques were used to augment the DFUC database. However, none quantitative results are presented nor discussed in the paper.

Although results appear consistent in these prior works, the obtained metrics are not entirely satisfactory especially with the use a such big training set also considering that all methods applied data augmentation to increase even more the training set. The tested networks still struggle to differentiate wounds from other skin conditions.

3.1.3.3 DFU Segmentation

Most early studies as well as current work of DFU segmentation are based on a semantic segmentation approach using Fully Convolution Neural Networks (FCNs). Goyal [66] addressed automated segmentation of DFU and its surrounding skin by training the main architecture of FCNs (FCN-AlexNet, FCN32, FCN16 and FCN8) on a dataset of 705 images including 600 of DFU images. Experimental results revealed that FCN-16s performed the best compared to the other FCN architectures with a dice score of 79.4% for ulcer region and 85.1% for surrounding skin. A system based on MobileNet has been developed by li et al. in [80]. They used 950 wound images collected from the Internet and the hospital. Therefore, a preprocessing step was required to remove manually the background from the images using traditional methods based on watershed and dynamic threshold to improve the quality of the network. This step aims to extract the couple skin-wound from the background in order to simplify the segmentation task and to highlight the wound features. The obtained IoU index was of 85%. While another system by Wang et al. [81] performed fully automatic wound segmentation MobileNetsV2. Segmentation masks were filled to improve the true positive rate using connected component labelling (CCL). The network was trained on 5000 images with corresponding annotation masks after a data augmentation of 1109 DFU images. The proposed framework shows high efficiency and accuracy in wound image segmentation.

Another state-of-the-art semantic segmentation network is the fully convolutional network U-net. U-net is widely used in biomedical image segmentation and shows surprising results on small datasets. Several studies have investigated the robustness of U-net in

comparison to other FCNs. Cui et al. [82] found that a U-Net trained on 392 images can outperform a patch-based CNN trained on 4500 patches. However, the downside of this method is the requirement of a preprocessing step that includes artifacts removal and background removal performed by the GrabCut tool. In this study [83], they investigated ulcers segmentation through Mask-RCNN model which is an extension of Faster RCNN and compare it to U-net using a small dataset of 400 ulcer. An offline data augmentation step was considered to increase the size of the training dataset. Mask-RCNN with ResNet-101 as backbone the highest accuracy with an average precision of 50.84% which is considered as a very low performance.

In a study conducted by Zahia et al. [84], the authors presented a novel approach for automatic wound segmentation in pressure injuries images by combining DL based segmentation and 3D mesh of the wound. A Mask RCNN was trained on 210 2D wound images to get the external wound boundaries. Then, a segmentation of the wound in the 3D mesh was obtained by matching 2D and 3D views. Experimental results show that the proposed method achieved very close results while comparing physical wound measurements from the obtained segmentation masks to those retrieved by clinicians, but even so segmentation precision was solely 87%.

A comprehensive comparison among three DL-based models for semantic segmentation (FCN, U-Net and DeepLabV3) and Associative Hierarchical Random Fields (AHRF) for segmenting wound images [85]. Various metrics were considered in this review including DICE score, inference time, amount of training data etc. They found that for small datasets AHRF achieved good accuracy for small databases. In contrast, FCN, U-Net and DeepLabV3 are more accurate than AHRF for larger databases.

Most papers, perform a post-processing step after wound segmentation. This step aims to refine the segmentation result for more natural output. Some of the used post-processing methods are noise removal, edge softening and morphological operations like hole filling or area opening and closing by sequential dilation and erosion.

3.2 Tissue Classification

In addition to wound segmentation, tissue classification and characterization have been the subject of several studies on wound image analysis. This approach consists of classifying the distinct tissues withing the wound bed so that exact proportion of each tissue can be quantitatively estimated. Different authors have proposed various classification methods considering three main types of tissues (slough, granulation and necrosis) inspired by the red–yellow–black color evaluation model proposed by Arnqvist et al. [86].

Figure 3.8: Examples of tissue types in DFU images

Where, each color represents a different phase on the continuum of the wound healing process. Overall, granulation is red tissue that is healing properly, slough is yellow wet tissue that accumulates on the wound surface, and finally necrosis is black tissue that has died 3.8 [31].

3.2.1 Traditional Image Processing Methods

Based on several reviews of the current literature [87], [32] and [31], a few studies have addressed tissue classification using traditional image processing methods. Tissue identification is very complex even for the most experienced clinicians, estimating the exact percentage of each tissue can be very challenging with the presence of multiple types of tissues mixed together. Mostly, researchers focused on traditional image processing methods using color-based techniques with color histograms.

Choosing the most appropriate color space is primordial to achieve the best performance for color-based tissue analysis. RGB color space is known to be the most commonly used color representation system by researchers. Typically, images captured by a digital camera or smartphone are saved by default in RGB color space where R, G and B are considered as the primary colors of light. However, the value of each component strongly depends on the light intensity making this color space not perceptually uniform. For instance, two colors that appear visually close may be more separated in the RGB color space than two other colors that appear distant for the human eye [88]. Thus, RGB color representation is not the best for color analysis of wounds tissues.

An alternative is to convert the images from RGB model to a more suitable color space. Numerous considered studies used HSV (H: hue, S: saturation, V: value) or HSI (H: hue, S: saturation, I: intensity) color representation model that is much close to the way

humans perceive colors. In fact, "H" represents the pure color in the image, "S" is the degree of dilution of pure color by white light and obviously "I" is the intensity of light; which provides a most pronounced contrast discriminating between tissues. Calculation of probability map was used by Ahmad Fauzi et al. [89] for tissue separation based on distance in modified HSV space. In an early work, Perez et al. [90] performed tissue segmentation in leg ulcer images by analyzing the red, green, blue, saturation and intensity channels of the images. While, Delode et al. [91] performed tissue identification using a color conversion from RGB to HSV, which allowed them to detect black, red and yellow pixels corresponding to the different tissues inside withing the wound area. Some other color spaces such as CIE Lab, YDbDr and YCbCr have also been successfully used by researchers for color based wound analysis.

3.2.2 ML Based Methods

Machine learning methods have been widely applied by researchers to tissue segmentation problems in wound images. One of the first studies dealing with tissue classification using ML methods was carried out by Wannous et al. [92]. The authors employed a multi-view strategy based on an SVM to classify the tissues into granulation, slough and necrosis. Wound images were taken with a digital camera according to a specific protocol integrating several points of views for each single wound and a Macbeth color pattern was included for color correction. They selected 20 most relevant descriptors including color and texture that were used as input of the SVM classifier. The single view classification labels were mapped and merged into the multi-view 3D model [93]. The authors reported that the multiview strategy using the 3-D model help to enhance repeatability and robustness for tissue classification. In [94] and [23], the same author performed segmentation-driven classification approach using color and texture features and SVM classifier. Before feature extraction, three techniques (mean shift, JSEG and CSC) were compared to perform tissue segmentation of wound images. Experimental results showed that the segmentation-driven classification approach is more suitable than a pixel-based approach.

Mukherjee et al. [38] used a set of color and textural features to describe granulation, necrotic, and slough tissues. For every segmented wound region, 45 color channels were used for color feature extraction including 15 color spaces with considering the three-color components in each color space. Next, texture features were extracted from each of the 45 color channels using Shannon's entropy, local contrast features and local binary pattern. Finally, Bayesian classifier and SVM classifiers with various kernels were trained

for wound tissue classification. Their experiments on 74 wound images showed that SVM with 3rd order polynomial kernel provided the best results with an overall accuracy of 86.13%.

In [41] and [95], the authors applied a color segmentation algorithm by applying K-means clustering algorithm from CIE Lab color features after outlining the wound boundary with an accelerated mean-shift algorithm. Wang et al. reported that tissue classification is a hard task in real images due to the complexity of skin color and texture. For instance, many dark red granulation pixels were misclassified and considered as necrosis instead of granulation, while dark yellow slough pixels considered as granulation instead of slough tissue. In a similar approach by Patel et al. [96], texture-based features were extracted with local contrast features and local binary pattern. While color features were extracted with mean, skewness, standard deviation, kurtosis and variance. All these features were finally used for tissue classification purpose with K-means. In a more recent work [43], tissue classification is performed using color histogram features from RGB images. First, a mean shift segmentation is performed on the wound tissue pixels to obtain a set of fine-grained regions. Then, color histogram features are extracted from each region. A label with one of the predefined tissue types is set to each region using a random forest classifier. However, none quantitative results are presented nor discussed in these papers.

A hybrid system based on neural networks and Bayesian classifiers for automatic tissue identification in five different types namely: skin, healing, granulation, slough and necrosis has been proposed by Veredas et al. [97]. An adaptive mean shift procedure and a region-growing strategy were applied for ROI segmentation from wound images. Multi-layer neural network followed by a Bayesian classifier were performed to classify color and texture patterns extracted from the segmented area. The performance of proposed approach was about 78.7% of sensitivity and 91.5% of accuracy. In a similar research by the same authors [98], tissue categorization was carried out using three different ML methods namely: Neural Networks, SVM and Random Forest. The evaluation metrics on 113 photographs showed that the high overall accuracy rates of and 92.14% and 93.66% were obtained by SVM and Random Forest respectively. While Random Forest gives the highest metrics regarding granulation tissue, SVM gives the highest values when detecting necrotic, slough and healing tissues.

Recently, Chakraborty et al. [29] presented an automated tissue identification scheme after a fuzzy c-means clustering for wound segmentation from pressure and diabetic ulcer images considering Medetec database. Then, a set of color features were extracted from the segmented areas based on color intensity value using the Db channel in YDbDr

color space. In this paper, Linear Discriminant Analysis (LDA), Decision Tree (DT), Naïve Bayesian (NB) and Random Forest (RF) have been considered for comparison for classifying tissue patterns. The RF tissue classification approach provided the best tissue classification prediction on a total of 153 images with an overall accuracy of 93.75%. However, RF classifier performed poorly on necrosis tissue as compared to the four extant classifiers. In a previous work [99], the same author obtained an extremely efficient tissue classification using LDA based on three simple color features that reached 91.45% of overall accuracy regarding the three tissue types. Their method outperformed state-of-the-art methods (K-NN, fuzzy k-NN, K-means, SVM and Bayesian).

In a different approach [100], the authors proposed a tissue identification method based on color and texture analysis. Eleven features were considered to retrieve significant features including color spaces, local entropy, local range, local standard deviation, gray level co-occurrence matrix, SIFT and prior visual appearance. Next, non-negative matrix factorization (NMF) was applied in order to select the most significant features to reduce computation time. Tissue classification was done with the gradient boosted trees classifier (GBT) that was trained on 377 CWs images from Medetec and national pressure ulcer advisory panel website. The proposed method achieved an average accuracy of 96%.

Several ML wound tissue classification approaches have been proposed in the literature. The first step is wound segmentation using either automatic or semi-automatic methods. Then, an image pre-processing step is primordial for white balance estimation and color correction so tissue classification can be more effective. Finally, tissue classification step is then performed using ML classifiers (such as: SVM, RF, NB, K-means, etc.) by incorporating the extracted color and texture descriptors from prior segmented wound region. The most significant limitations of the ML based tissue classification studies can be summarized into two main issues. First, wound tissues are classified based on the selected features. Nevertheless, due to the complexity of skin color and texture of CW tissues, features extraction is a critical step that can be very challenging in ML based methods. Second, manual annotation of the ground truth tissues by clinical experts is very expensive in term of time and expertise and requires an exhaustive methodology to get significant and reliable annotations to compare the segmentation with. Therefore, as addressed by many similar studies, manual segmentation of tissues in wound images is prone to a high inter- and intra-observer variability. Thus, the classification performance evaluation is not trivial and can be highly inaccurate.

3.2.3 DL Based Methods

Recently, more sophisticated methods with deep learning have been the focus of the research in this area. New approaches involve convolutional neural networks for classification or fully convolutional neural networks for semantic segmentation. However, the lack of annotated wound tissue databases as mentioned before is still an ongoing problem. Therefore, most approaches are typically performed on a patch level and afterwards data augmentation techniques are applied to further increase the number of training samples.

A novel approach has been presented by Zahia et al. in [101] to handle small datasets in DL using a new CNN architecture that was trained and tested on a small dataset of 22 pressure wound images. The main idea is to divide the images into small 5x5 square patches to end up with 380,000 small images. These sub-images along with their ground truth labels were fed into the proposed CNN model for training the network to classify each input patch into one of the different tissue classes. The achieved performance was relatively high with an overall accuracy of 92.01%. Despite the high accuracy of their method, the system confuses some parts in the images that appears dark due to depth with necrotic tissue and slough tissue classification still challenging with the lowest precision rate of 77.90%. Rajathi et al. [102] presented a similar approach where 1250 varicose ulcer images were partitioned to 5x5 patches. This work consists of three main stages: data processing to remove flash light reflection, active contour for wounded area segmentation and a 4-layer CNN for tissue classification into four different types (granulation, slough, necrotic and epithelial). Collected patches and their corresponding ground truth are given as input to train the proposed CNN architecture. The proposed approach outperformed [38] and [93] with an overall accuracy of 99.55%. but the results were reported on a relatively small set of 5 images.

In this paper [103], the authors expanded the number of tissue types to seven instead of three classes, namely: Necrotic, Slough, Healthy Granulation, Infected, Unhealthy Granulation, Hyper Granulation and Epithelialization. Adding more tissue classes help to increase tissue identification accuracy and thus to obtain higher evaluation parameters. Nejati et al. proposed a hybrid deep learning approach considering a pre-trained AlexNet architecture for feature extraction and a linear SVM for patch-level classification. A dataset of 350 wound images was partitioned into 20x20 square patches that were resized to 227x227 to match the input size of AlexNet. The proposed method combining AlexNet and SVM classifier performed better than using SVM from classical feature descriptors such as RGB, HSV color histograms and LBP texture descriptor.

They achieved an acceptable overall accuracy of 86.40%.

Unlike existing approaches using square patches, Blanco et al. proposed a superpixel-driven method called QTDU [104]. A superpixel is an image patch that have more perceptual meaning than a regular square patch. In addition, pixels belonging to a given superpixel share similar tissue properties contrary to regular patches that can contain several tissue types. The proposed system includes three stages: ulcer segmentation, tissues labeling, and wounded area quantification. A database of 217 arterial and venous wound images were considered in this study. After superpixel segmentation, data augmentation was applied to each superpixel, which result in 44,893 superpixels labeled into four classes namely fibrin, granulation, necrosis, and not wound. For the training stage, the authors used the existing CNNs ResNet and InceptionV3 while adding six additional layers to the end of each network to deal with overfitting. Experiments showed that QTDU with ResNet generated better results than QTDU with InceptionV3. In addition, the proposed QTDU outperformed state-of-the-art ML approaches with an f1-score of 97.1%.

In a completely different approach [105], Maity et al. proposed a pixel-based wound tissue classification method, where a deep neural network classifier is trained based prior features extracted from different tissue regions. A total of 105 features were calculated for each pixel including color, texture, LBP, entropy, etc. Pixelbased feature extraction is implemented by running a mask window of 9×9 over each pixel of the tissue regions. Then, the neural network training was performed based on extracted features that achieved a very high accuracy rate of 99%. However, the authors did not give any details about the used neural network.

Another methodology to perform data augmentation has been investigated by Pholberdee et al. [106]. The authors proposed a color data augmentation method based on color mapping parameters from two camera pairs. This technique tripled the size of training samples, they moved from 180 images from Medetec database to over 400,430 samples for training set and 291,486 samples for validation set including granulation, slough, and necrosis tissue. The experimental results demonstrated that the proposed color data augmentation technique greatly improved tissue classification performance. However, the IoU index was merely 53% instead of 47%. As in Zahia et al., slough was the most challenging tissue to be identified.

On the other hand, Garcia-Zapirain et al. presented a complex classification framework with 3D convolution neural networks for pressure injuries images [107]. The proposed system is based on the fusion of the following feature modalities: original image in HIS color space, smoothed image with a pre-selected Gaussian kernel, prior appearance

image and current appearance image using linear combinations of discrete Gaussians (LCDG). These features are provided as inputs to the 3D DeepMedic CNN network. The proposed 3D CNN architecture consists eleven layers with four pathways one for each single modality. The network was trained on 193 images and reached a mean DICE coefficient of 92% for tissue classification. In similar research by Elmogy et al., five feature modalities were used with the 3D CNN. The authors added RGB image as an input alongside with prior model proposed in [108]. This architecture showed a similar DSC accuracy of 92% on the same dataset. In another work by the same authors [109], four modalities were considered; HIS image was replaced by the RGB image. The method achieved a DSC of 93% on 100 images.

4 Dataset and Techniques

4.1 Image acquisition

To explore the effectiveness of DFU segmentation and tissue classification using deep learning models, a wound database has been constituted over the previous three years in cooperation with two hospital centers, Hospital Nacional Dos de Mayo (Lima, Peru) and the diabetology service of CHRO Hospital (Orleans, France). The images are captured from different patients during several clinical visits with an approximative interval of two weeks between visits. For image acquisition, we used a smartphone camera of Xiaomi Note7 with a resolution of 1080 x 2340 and Samsung S10 with a resolution of 1440 x 3040. Additionally, an add-on temperature sensor FlirOne pro [110] is added to the acquisition system to capture instantly a pair of images providing both thermal and color modality. This camera can easily be plugged into a smartphone and offers high sensitivity that detects temperature differences down to 70 mK which can be very accurate to detect temperature changes due to inflammation or infection. A detailed overview of the used FliOne pro is presented in Table 4.1.

Figure 4.1: FlirOne Pro camera

Table 4.1: FlirOne pro overview

Resolution	1440 × 1080 pixels
Object Temperature Range	-20°C to +400°C
Thermal sensitivity	70 mK
Accuracy	3 °C or 5%
Weight	36.5 g
Dimensions (H x W x D)	68 × 34 × 14 mm

The acquired image pairs are taken using a multiview approach, and then used to generate a realistic 3D wound with measurement information including area, volume and temperature. An illustration of different acquisition modalities for one patient is

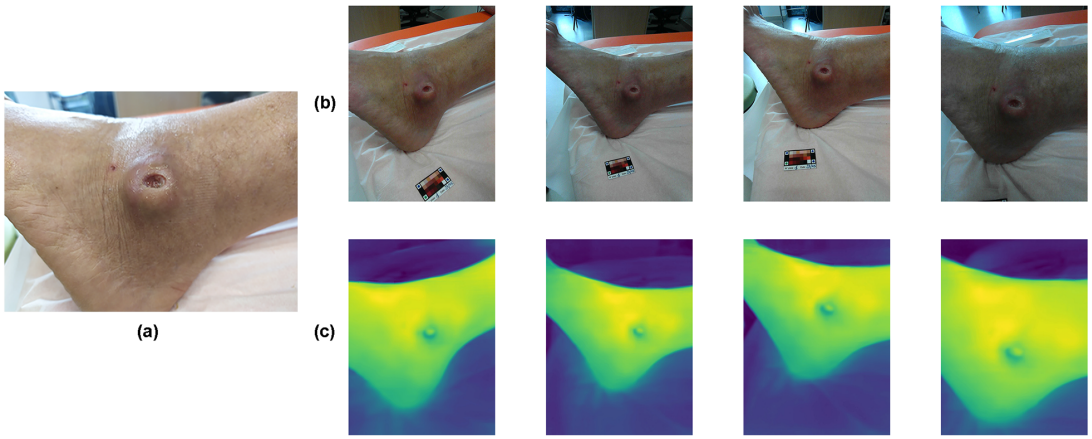

Figure 4.2: Acquisition modalities (a) Color image, (b) Multiview color modality and (c) Multiview IR modality

presented in (Figure. 4.2). Image acquisition is performed during clinical examination when the wound is totally cleansed to minimize the effects of infection, and so the tissues can be more visible. The acquisition protocol had to be fast and practical for clinical settings. Thus, we a proposed a very simple acquisition protocol according to each image modality:

- Color image modality: The acquisition protocol consists on capturing one image of the wound under uncontrolled lighting conditions while framing the ulcer area from a point of view as frontal as possible with a parallel orientation to the plane of the wound.

- Multiview color modality: The acquisition protocol consists on capturing a set of images that include several points of views by moving the camera around the wound; to the left, right, up and down.

- Multiview IR modality: Similarly, a set of images is captured from different perspectives. However, a preliminary requirement was to maintain a controlled environment where the temperature is set to room temperature.

In this dataset, we excluded blurry or dark images. Additionally, 92 CW images from ESCALE database [111] were also added to our database. Consequently, the collected database covers all CW pathologies including diabetic foot ulcers, burns, pressure injuries, etc with different shapes, sizes and healing stages (See Figure. 4.3). We collected some metadata regarding the patient's age, sex, identity, DFU measurements, site and the used dressing. However, this data was not considered in this work since our main objective was to perform accurate tissue and wound segmentation.

Figure 4.3: Sample images from our heterogeneous chronic wounds database

4.2 Expert Labelling

It is well known that accurate DL models with high prediction performance require massive amounts of labeled data with quality annotation. Furthermore, the quality of image annotation is an essential part of DL models training and success. Incorrect annotations if fed to the DL model can straightforwardly impact the network performance and can lead to inaccurate predictions. As an important component of DL models to be accurately trained for correct prediction, each image in the dataset must be accurately labeled. The process of manual labelling involves expert knowledge to obtain precise ground truth annotations. In our case, this process must be done by wound care specialists, which is time consuming and costly in term of expertise. Furthermore, manual tissue identification is a very difficult task for clinicians since tissue distribution withing the wound is usually irregular especially with the presence of a mixture of several tissue types inside the same wound. Some studies like [112] and [23] demonstrate that manual annotation of tissues in wound images by experts is prone to a high inter- and intra-observer variability which result in different ground truths for the same wound. Thus, some researchers developed their own annotation software. However, there is no public data annotation tool for wound tissue annotation and using a software like Photoshop is not practical for clinicians. So as solution, for an effective tissue annotation of the ground truth, we developed a GUI labelling interface to help clinicians in the labeling process with pre-segmenting the wound images using a superpixel algorithm. Then, clinicians can directly label the pre-segmented wound tissues. Consequently, this process will become easier, more accurate and less time consuming.

4.2.1 Superpixels Segmentation

Many image segmentation methods have been published in the literature during the last decades. Recently, the use of superpixel algorithms have become increasingly popular in many computer vision and Image processing applications for segmentation and semantic

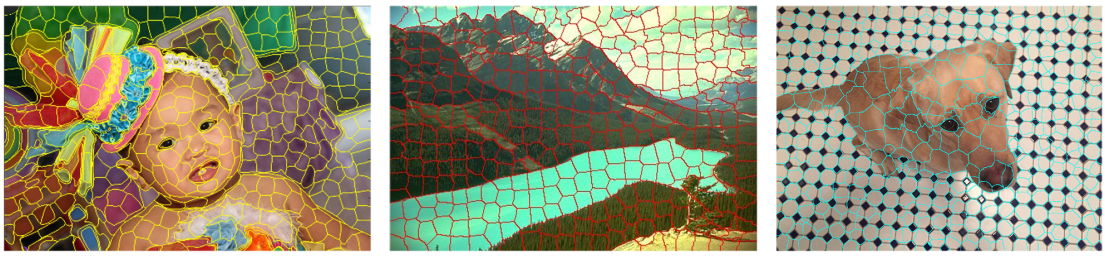

Figure 4.4: Samples of Superpixels Segmentation from literature

labelling especially in the biomedical imaging field. Recently, superpixel algorithms have become widely used in computer vision and Image processing application for segmentation and semantic labelling especially in the medical imaging field. First introduced by Ren and Malik in 2003 [113], these algorithms showed the ability to provide powerful segmentation results with high efficiency, especially in real-time vision systems.

Superpixels are the results of an image over-segmentation. The main idea behind these algorithms is to divide the image into a large number of segments perceptually similar, which are formed by considering similarity information according to many criteria such as color, texture, brightness, intensity, shape, etc. These segments carry more information about the structure of the scene than individual pixels. Additionally, they align better with the object's edges in the image than a regular square patches. A major advantage associated with these algorithms include high visually meaningful segments with lower computational complexity and less memory.

Superpixel algorithms can be broadly classified into graph-based and clustering-based approaches. Graph-based approaches treat the segmentation as a graph-partitioning problem where each pixel is considered as a node in the graph. While edge weights measure the similarity between adjacent pixels and help to relate the pairs of nodes. The most used algorithms in this category include Felzenszwalb [114] and normalized-cuts [115]. Whereas, clustering-based approaches are mainly based on traditional clustering techniques such as k-means for superpixel segmentation and are simpler to implement. There are various algorithms in this category such as SLIC [116], Watershed [117], Quick-Shift [118], etc.

4.2.2 Simple Linear Iterative Clustering (SLIC)

The Simple Linear Iterative Clustering (SLIC) is the state-of-the-art superpixel segmentation algorithm proposed in 2010 by Achanta et al. [116]. SLIC performs a local k-means clustering that generates clusters of pixels based on their proximity and color

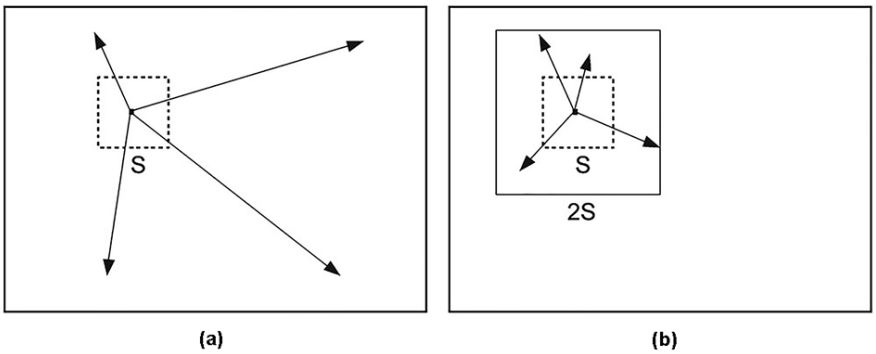

Figure 4.5: (a) Standard k-means searches, (b) SLIC searches within a limited region [119]

similarity in the combined five-dimensional space [l, a, b, x, y] considering the l, a, b values of the CIELAB color space and the spatial components x and y. Unlike K-Mean clustering method, similarity of any two pixels depends not only on color similarity but also on spatial distance. In addition, SLIC limited the search region to a neighborhood of 2S x 2S from each cluster center as can be seen in 4.5. Contrary to K-means where the distance was computed between each pixel and all cluster centers in the image. This reduced dramatically the number of distance calculation. Instead of using a simple Euclidean distance in the 5-D space, the authors proposed a weighted distance measure that combine two distances into a single measure. The proposed distance measure D_s is defined as:

$$d_{lab} = \sqrt{((l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2)} \tag{4.1}$$

$$d_{xy} = \sqrt{((x_k - x_i)^2 + (y_k - y_i)^2)} \tag{4.2}$$

$$D_s = d_{lab} + (\frac{m}{S}) * d_{xy} \tag{4.3}$$

where D_s is the sum of a distance in Lab color space d_{lab} and Euclidian distance d_{xy} normalized by the grid interval S. This combination results in more consistent superpixels of the with a better adherence to the edges in the image. The key parameters for SLIC are the number of superpixels to output K and compactness m.

First, a K number of superpixels is introduced. For an image of N pixels, the approximate size of each superpixel will be N/K. To produce equally sized superpixels, the K superpixels are sampled on a regular grid with an interval S = $\sqrt{(N/K)}$ between each

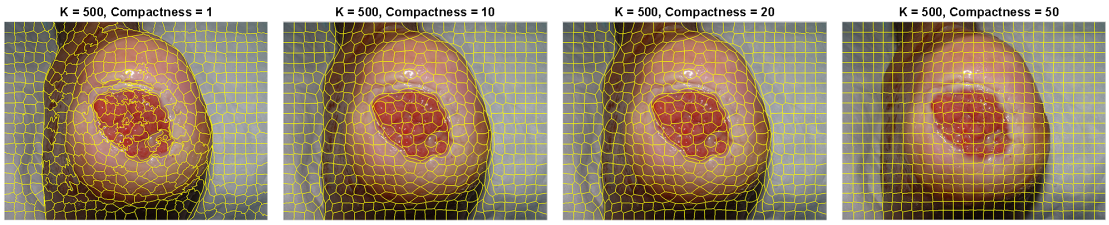

Figure 4.6: Samples of SLIC Segmentation with different compactness values

superpixel center. More details about superpixels repartition with SLIC algorithm can be found in [116]. Then, the compactness of the produced superpixels can be controlled by adjusting the parameter m in D_s. Compactness helps to balance color and space proximity during superpixel construction. Large values of m result in important space proximity, yielding to superpixels with more square shape. While, smaller values of m result in non-compact superpixels that adhere better to image boundaries but with more irregular shape. Figure 4.6 illustrates SLIC segmentation output when varying compactness value. This parameter can be greatly influenced by intensity and regularity of the boundaries in the image. In the same image, SLIC can produce smooth superpixels with regular size in the smooth areas and more irregular superpixels in the textured areas.

However, an improved version of SLIC has been proposed in [119]. So, instead of using the same compactness value for all superpixels in the image, this method adjusts adaptively the compactness for each superpixel independently. This generates regular shaped superpixels that adheres perfectly to the wound boundaries in both textured and non-textured regions. Thus, the only parameter left to manage is the number of superpixels K as can be seen in (Figure 4.7).

4.2.3 Annotation GUI

The proposed labelling GUI uses SLIC algorithm to automatically generate superpixel regions without any manual outlining. So, instead of setting a label for each pixel, the clinical expert simply has to assign a label to these pre-segmented regions to obtain at the end a pixel-wise annotation of the wound. An illustration of the developed wound tissue labelling software interface is presented in Figure. 4.8.

The interface is interactive and provide the user with various functionalities. First, the user can crop the image to select the ROI. This is particularly important for high resolution images to focus the segmentation on the wound tissues and thus reduce computational time. Clinicians have control over the number of desired segments to produce more accurate labeling. This value can be updated according to the size and the com-

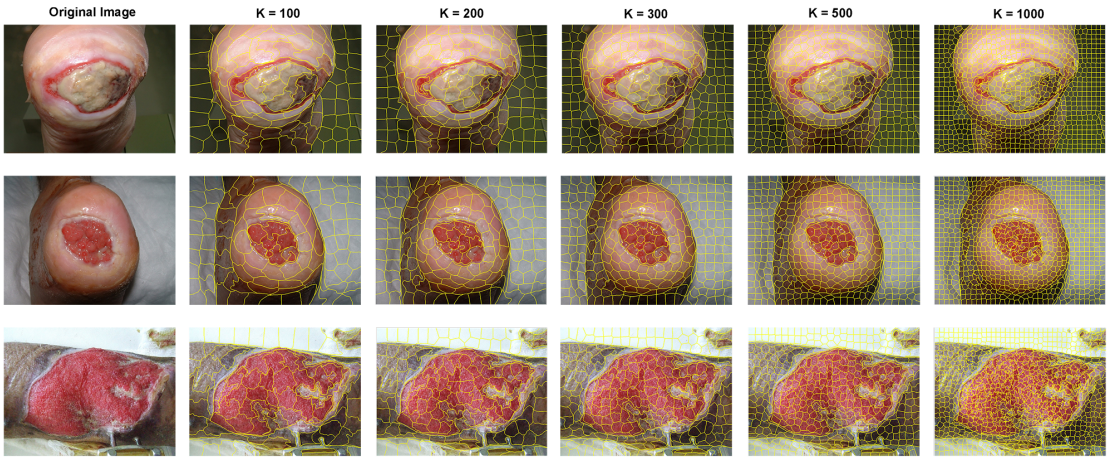

Figure 4.7: Samples of SLIC Segmentation with different numbers of superpixels

plexity of the wound. A high number of superpixels is required to segment accurately complex or big wounds while fewer superpixels are needed for smaller or less complex ones. The output of superpixels segmentation changes instantly in real time. Once the segmentation is done, the user has to assign a label color for each superpixel withing the wound. We have label buttons with a specific color for each tissue type; red for granulation, yellow for slough, black for necrosis and pink for healing tissue or epithelialization. An unknown label with white color is added to the list in case of the presence of non-tissue superpixels inside the wound bed like bones for example and can be used also when clinician is unable to detect precisely the tissue class. Finally, the annotation process is done by selection the corresponding label color for each selected superpixel. The color of the tissue changes immediately so clinicians can see the current set of labels in real time.

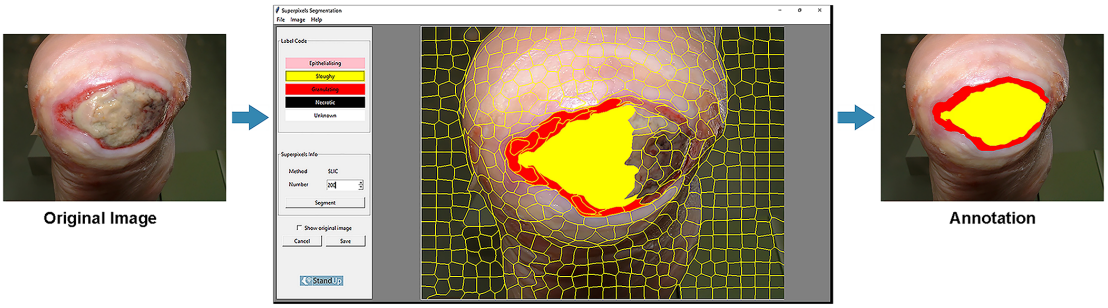

Figure 4.8: Graphical interface for the expert manual labeling

Clinicians reported that this GUI is very simple and easy to use. The annotation

process is much more accurate and less time consuming. This tool has been developed using Python3 [120] and relies solely on open-source libraries. Therefore, it can be adapted for any other dataset by changing only the labelling classes.

4.3 Deep Neural Networks

4.3.1 Artificial Neural networks

As we previously stated, Deep learning networks have surpassed Machine Learning approaches performance and have revolutionized several industries in the last decades. Today, DL has a widespread adoption in our everyday life for a variety of applications including forecast, natural language processing, fraud detection, speech recognition, computer vision and is showing great promise in the biomedical image analysis field where its precision is more and more approaching human level performance on various tasks.

However, DL is a complicated process based on artificial neural networks that has not appeared overnight, their evolution took decades. The history of DL dates back to more than 70 years ago when Warren McCulloch and Walter Pitts in 1943 modeled the first computational model for neural networks with electrical circuits inspired by the Biological Neuron of human brain [121]. Basically, a biological neuron receives a signal as input through the dendrites, processes it in the soma and then passes the output through the axon. In a more simplistic view, each single neuron in our brain receives signals and produce an instantaneous response. In a similar way, the proposed artificial neural is divided into two parts trying to mimic the working of biological neurons. The 'g' part similar to dendrite takes an input that can be excitatory or inhibitory then performs an aggregation. While the 'f' part takes a decision based on the aggregated value. The output is a binary value, '0' if the weighted sum exceeds threshold value 'b'

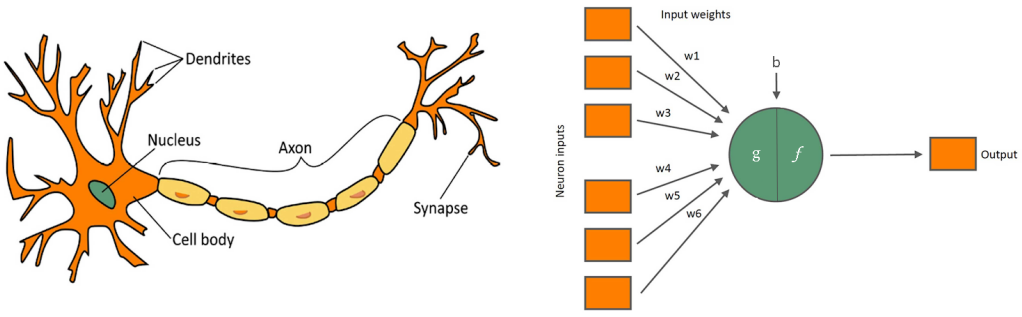

Figure 4.9: Biological neuron versus artificial neuron

and '1' otherwise. The structure of a biological neuron versus McCulloch-Pitts Neuron is shown in Figure 4.9.

The McCulloch-Pitts Neuron indeed suffers from some limitations and has been the subject of many critics. Yet it was the very first step towards the evolution of artificial neural network and deep learning. The major limitations are that the model lacked a mechanism for learning besides that weights and threshold were manually adjusted. Additionally, only binary inputs and outputs were allowed. These limitations however were reduced when the concept of perceptron was introduced by the psychologist Frank Rosenblatt in 1958 [122]. Mark I Perceptron was modeled on a McCulloch-Pitts neuron with a simple input output relationship but its originality lay in the ability to learn the correct weights automatically through a learning process that try to minimize the difference between desired and current output. The perceptron could also process non-boolean inputs. Therefore, it was the first trainable neural network with a single layer. Since that time, artificial neural networks were an active area of research and so deep learning was born.

Although Rosenblatt's perceptron seemed promising, it could only handle simple classification tasks for linearly separable classes. This simple single neuron model struggled to solve more difficult problems. Another big problem was that a single-layer perceptron cannot implement XOR logic function because the classes in XOR are not linearly separable. This is why the concept of hidden layers was introduced in the 1980s to overcome most issues of the Rosenblatt's Perceptron. Adding two or more layers to the perceptron allow the network to learn more complicated features and thus resolve more complex tasks. Thus, extra hidden layers of artificial neurons were added in between the input and the output nodes which transformed a single-layer perceptron into a Multi-Layer Perceptron (MLP). Another huge advance was the use of backpropagation to adjust the weights during the learning process [123]. Replacing a linear activation function with a nonlinear one such as sigmoid in an MLP could overcome the XOR problem case seen in the case of single-layer perceptron. Since then, the MLP has attracted the scientific community and numerous architectures of artificial neural networks continue to expand in order to solve more complex tasks related to real-world problems and so Deep Learning has evolved gradually as computers became more advanced.

In a simplistic view, an artificial neural network (ANN) consists of layers of interconnected nodes where each node is a perceptron that has an associated weight and threshold. The general structure of an ANN is shown in Figure 4.10. The input layer provides various forms of information about input data samples, while the output layer produces the final predicted result. In between, hidden layers represent the intermediary

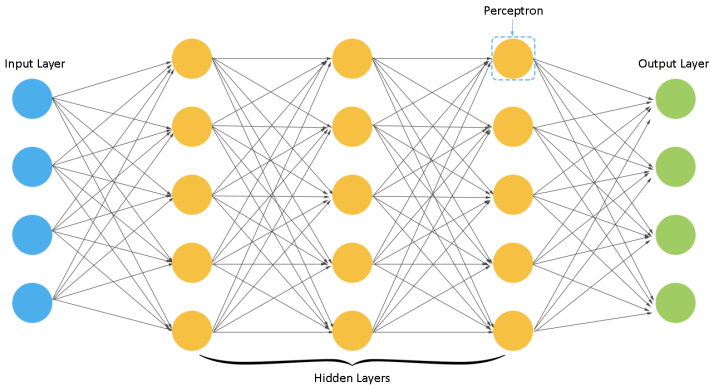

Figure 4.10: Multi-layer Perceptron

nodes where a single layer takes in a set of weighted input and produces output through. The outputs of each layer are used to feed the neurons of the next layer. Hidden layers identify salient features from the input data through a weight optimization process using an activation function. This process of optimizing weights is called training.

Once the ANN is structured, the next step is to train the model for a specific task. The algorithm starts with random initial weights. The training phase of the ANN consists on adjusting the weights for the output layer to provide low error on the training data and thus obtain accurate predictions on unseen data. Weights are updated after each iteration through the neuron in the network. This process is achieved by backpropagation learning algorithm which aims to minimize the squared error between the actual output values and the desired output values, in an iterative manner, so the network learns and improves. The complexity level of the ANN increases with every layer, the more hidden layers, the more the ability of the network to learn complex features. However, larger training datasets will be required to achieve a successful training and avoid overfitting. The network is classified as deep neural network when the model consists on a large number of hidden layers [124].

4.3.2 Convolutional neural networks

Convolutional neural networks (CNNs) are an extension of artificial neural networks presented above. Among different types of deep neural networks, convolutional neural networks are the most popular. CNNs methods have become popular in the recent years after Alexnet was introduced by Alex Krizhevsky in 2012 [125]. The research interest in CNNs increased ever since, more particularly for computer vision problems

including image classification, image segmentation, object detection. They have been also successfully applied to natural language processing, text analytics tasks and more.

CNNs are very similar to its predecessor artificial neural networks, both consist of a sequence of layers where every layer receives input from previous layer and provides output to the next one. However, the essential difference is that the CNN architecture utilizes the convolution operation. Rather than using fully connected hidden layers, CNNs consist of convolutional layers convolutional layers, pooling layers, fully connected layers, and normalization layers. In general, a CNN contains one or more convolutional layers, more convolutional layers mean a deeper network. In contrast to regular ANNs where each neuron is connected to every other neuron in the previous layer, in a CNN, only the last layers are fully connected. CNNs have the ability to maintain spatial information of the data with a few connections between the layers. This enables CNNs to effectively reduce the number of parameters in each layer.

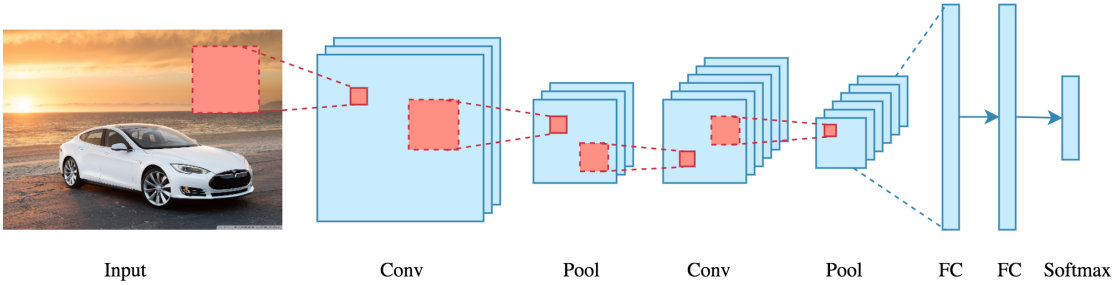

Figure 4.11: A graphic representation of the typical CNN architecture

A typical CNN architecture consists on a series of convolution and pooling operations occurring in an alternating way followed by one or more fully connected layers can be seen in Figure 4.11. The output is softmax in the case of a multiclass classification and sigmoid otherwise [124]. These layers can be thought as feature extractors, dimensionality reduction and classification layers, respectively. For example, given an input image and its label, the network applies the convolution layers to automatically extract meaningful features using filters. Extracted features become progressively more complex during the learning phase. In fact, the first convolution layer detects simple features such us edges and lines. While the following layers merge these features to learn more complex textures and patterns. Therefore, the result is a set of feature maps of various complexity. The next step is applying pooling layers to reduce computational complexity of the feature maps. Thus, reduce the number of parameters to learn and computational power requirements. Each sequence of convolution layer followed by pooling layer is a convolution. The parameters of these layers are learned over several iterations through a process

called training which aims to minimize the error between the output and ground truth label. Finally, for classification problems a couple of fully connected layers is required to connect the activations from the high-level features of the convolutions to the output layer. The number of neurons in the output layer is equal to the number of classes to be predicted.

- Convolutional layers: A convolutional layer is the main building block of CNNs that perform convolution operations to extract high-level features from the input data such as an image. Convolution make use of filters also known as kernels to slice across the image, each will produce a separate activation map. Intuitively, the result of the application of these filters is a map of activations also called feature map or convolved features that summarizes the present features throughout the input image. In fact, the major strength of CNNs is that they do not learn a single filter, they learn multiple features in parallel for a given input. Convolutional layers are not limited to the input layer but are also applied to the hidden layers for a wholesome understanding of the input images similar to how human would. Conventionally, the first convolutional layer allows the network to learn Low-Level features such as lines, edge, orientation and so on to move up to high-Level features like entire patterns as the depth of the network is increased.

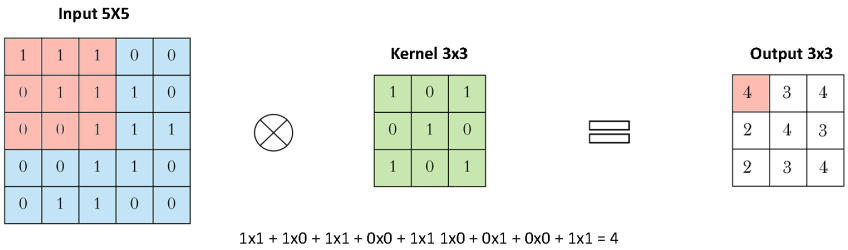

Figure 4.12: Convolution operation without padding

A convolution is a linear transformation that converts all the pixels in its receptive field into a single value. This operation involves a dot product between the convolution filter and the image pixels which is then summed to get one final output value. The two key hyperparameters of convolution operation include the filter size F and stride S. The filter size is usually smaller than the input which decrease the image size. The filter passes over all the pixels of the image to discover spatial

features anywhere in the image (See Figure 4.12). It moves from left to right with a given stride value across the complete width. This process is repeated from top to bottom with the same filter and stride value till it parses the entire image. The resulting output of the convolutional layer is a vector O called feature map.

However, since kernels are smaller than input, a few pixels from the border of the image might be lost when applying convolutional operation. The information on the boundaries is not preserved as well as the information in the center of the image. One straightforward solution to this problem is to use padding. The idea is to add extra pixels around to the edge of the input image. Consequently, the spatial output dimension will be equal to the input. Padding parameter is typically set to zero so the added pixels does not have any effect with the dot product when the filter is applied (See Figure 4.13).

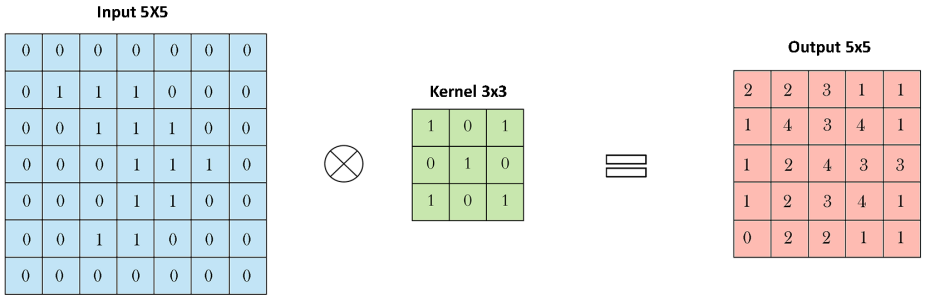

Figure 4.13: Convolution operation with padding

Besides regular convolution, there are two other types: dilated convolution and transposed convolution.

Dilated convolutions expand the receptive field of the network by inserting zero values into convolution kernels which provide a global view rather than finer details. Dilated convolutions are particularly used when object size is almost equal to the size of the image. This results in faster predictions and less computation cost. This type of convolution requires another parameter called the dilation rate which controls the up-sampling factor by defining a spacing between the values in a kernel. In practice, for a 3x3 kernel with a dilatation rate of 2 will result in 5x5 kernel while a dilatation rate of 3 will result in 7x7 kernel as shown in Figure 4.14. Whereas, for a dilation rate of 1, we get the same 3x3 kernel.

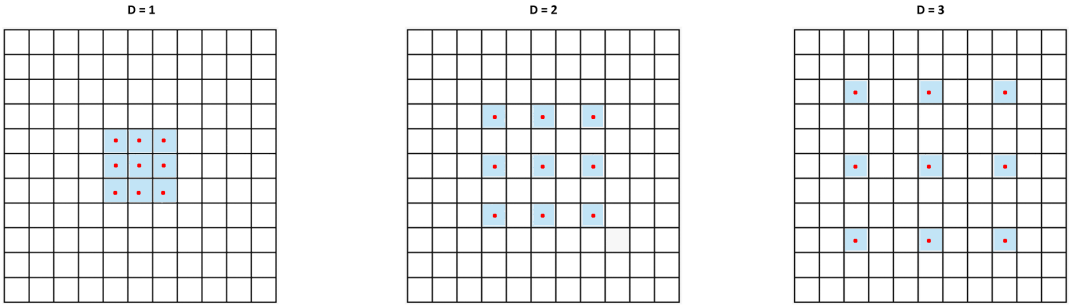

Figure 4.14: A 3x3 kernel with various dilatation rates: 1, 2 and 3 respectively

Transposed convolutions are used when output size is larger than input image. For instance, a 2x2 image and a 3x3 kernel while the output is of 4x4 (See figure 4.15). This type of convolution carries out a regular convolution but reverts its spatial transformation using dilation or a zero-padding strategy [126].

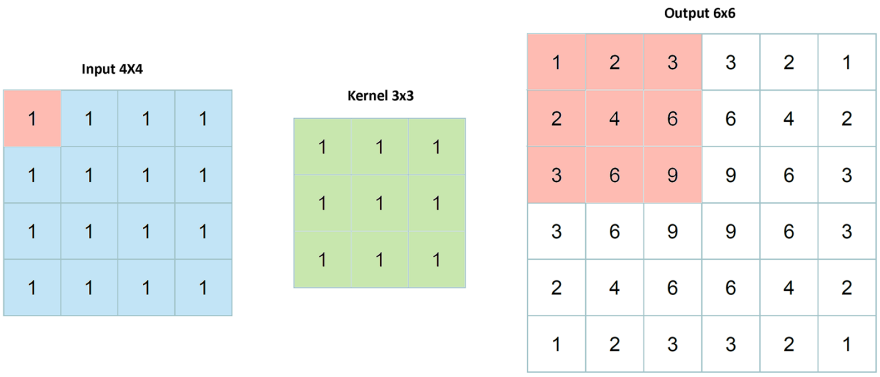

Figure 4.15: Example of transposed convolution

- Pooling layer: typically applied after a convolution layer, the pooling layer is a downsampling operation that reduce the spatial dimension of the convolved feature map while keeping the important information. The main objective is to discard redundant features and preserve relevant ones, and thus further reduce the number of parameters and speed up the computational power required to process the data. This downsampling reduces the height and width whilst keeping the depth intact. There are several types of pooling but max-pooling is the most popular one. Max

pooling slides a window over an input feature map resulting in a regular grid of squares and outputs the maximum value in each pooling window and the other values are discarded. Alike convolution, the window size and stride should be specified. The result of a max-pooling a filter size of 2 x 2 applied with a stride of 2 is illustrated in Figure 4.16. These parameters are commonly used in practice and reduce the dimension of feature maps by 75% compared to the original size [127].

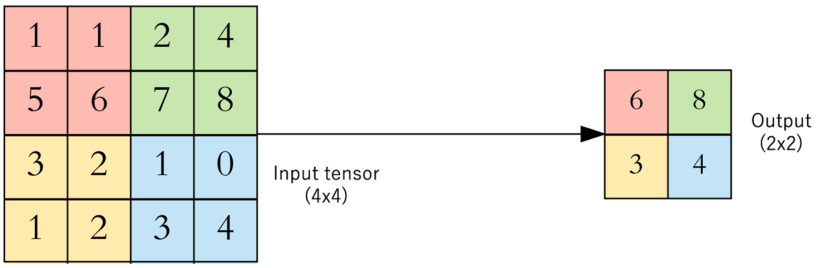

Figure 4.16: Max-pool with 2 x 2 window and stride 2

- Activation layer: is a non-linearity layer in a convolutional neural network that consists of an activation function that takes the feature map as input and creates the activation map as its output.

  The Formulas of the most commonly used activation functions are as follow :

  - Rectified Linear Activation (ReLU) :

$$\sigma(x) = max(0, x) \tag{4.4}$$

  - Logistic (Sigmoid) :

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4.5}$$

  - Hyperbolic Tangent (Tanh) :

$$\sigma(x) = \frac{e^{(x)} - e^{(-x)}}{e^{(x)} + e^{(-x)}} \tag{4.6}$$

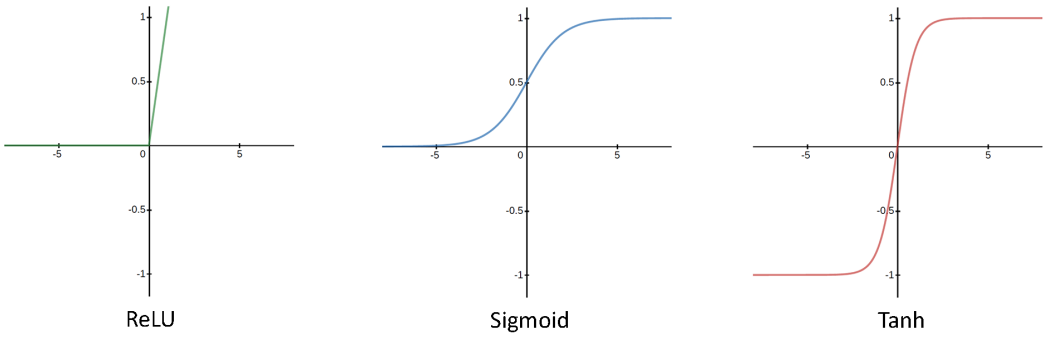

Figure 4.17: Activation Functions: ReLU, Sigmoid and Tanh

- Fully connected layers: fully connected layers are used as final layers of the network in classification problems. These layers use the high-level features produced by convolutions and generate class scores in order to correctly classify the objects in the images.

- Dropout : is by far the most used regularization technique for deep neural networks to prevent overfitting and improve generalization [128]. The idea is to temporally dropping out a number of neurons during training phase (See Figure 4.18). For example, if the dropout rate is around 0.5 this means that 50% of the neurons will be randomly ignored at each iteration. However, a dropped-out neuron at one iteration can be active at the next one. Dropout can be implemented on the input layer or any hidden layer in the network but not the output layer.

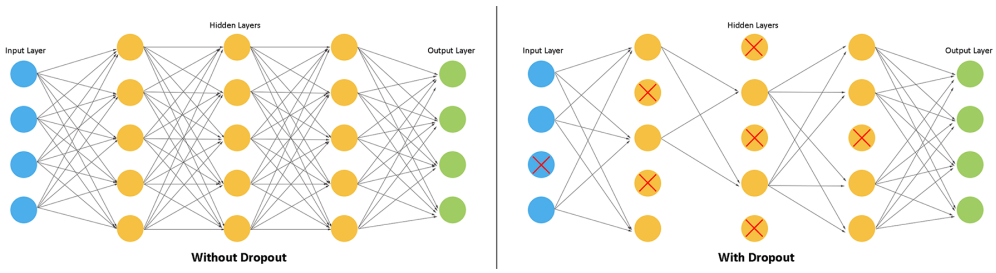

Figure 4.18: Dropout in Convolutional Neural Networks

Over the last 10 years, a various number of CNNs architectures made such significant contributions to the field and showed good results for challenging classification problems. Huge improvement in CNNs architecture have been achieved through architectural innovations as well as data availability and advances in the hardware technology. Some of

the most popular deep architectures include: LeNet, AlexNet, VGGNet, GoogLeNet and ResNet. These architectures have yielded the best results on the ImageNet classification challenge ILSVRC. ImageNet is a dataset of over 15 million labeled high-resolution images with around 22,000 categories (See Figure 4.19). ILSVRC image classification challenge uses a subset of ImageNet with 1000 images in each of the 1000 categories.

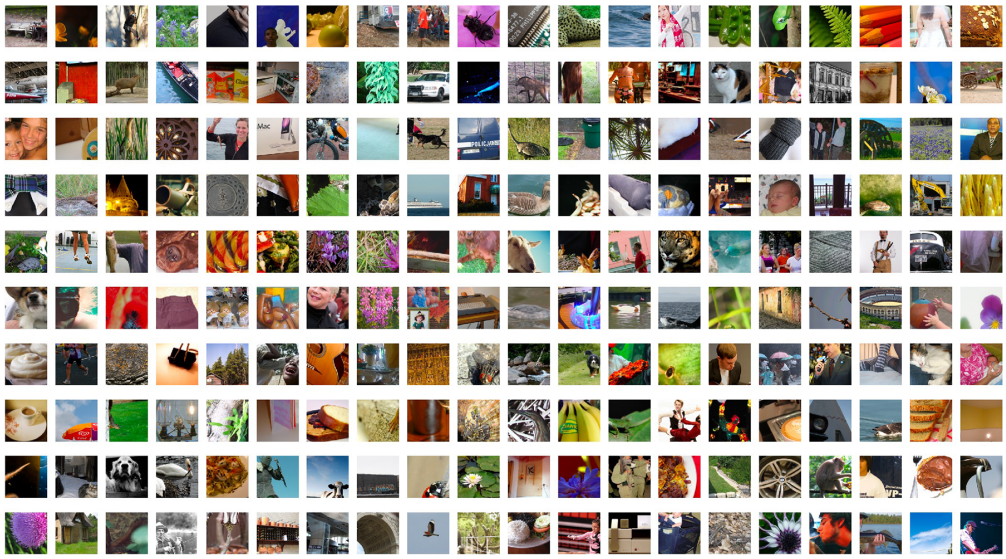

Figure 4.19: Samples from ImageNet Dataset

LeNet is one of the earliest image classification deep learning convolutional neural network proposed by Yann LeCun et al. in 1998 [129]. They model was first developed for handwritten digit recognition and has been then successfully applied to other image classification problems.

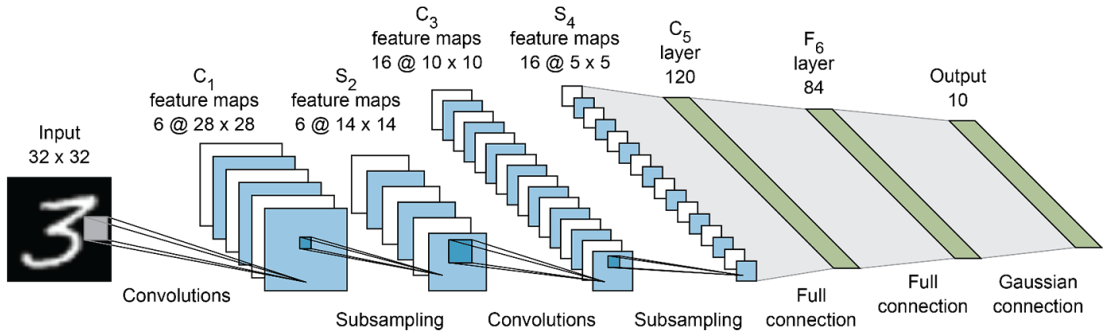

Figure 4.20: LeNet-5 Model Architecture

LeNet-5 is the most popular LeNet, its architecture is straightforward and easy to implement. The network consists of 5 layers with three sets of convolution layers with a combination of average pooling followed by two fully connected layers. At last, a Softmax classifier which classifies the images into corresponding class. The architecture of LeNet can be seen in Figure 4.20.

Proposed by Alex Krizhevsky et al., AlexNet [125] is a very popular convolutional neural network that won the ILSVRC challenge in 2012. The Alexnet architecture is much larger than the previous network LeNet. The model consists of eight layers: five convolutional layers followed by two fully-connected hidden layers and one fully-connected output layer with 1000 class labels (See Figure 4.21). AlexNet has the advantage of using Rectified Linear Units (ReLU) activation function instead of tanh after all the convolution and fully connected layers except last one. Additionally, AlexNet allows for multi-GPU training which reduced training time. Nevertheless, the network has 60 million parameters, thus, one major issue was the overfitting problem. Two methods were employed to reduce overfitting: data augmentation and dropout.

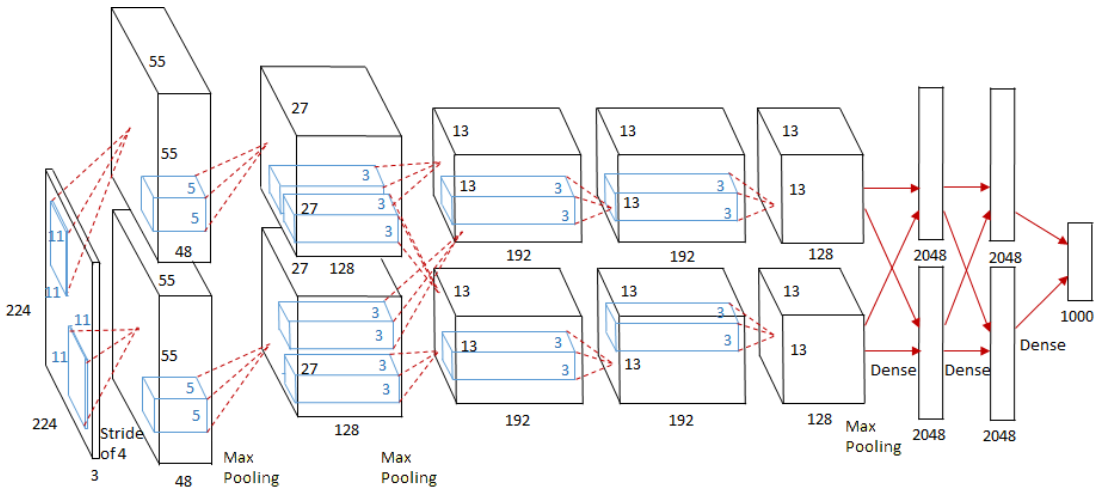

Figure 4.21: AlexNet Model Architecture

.

VGG is a successor of the AlexNet created by K. Simonyan et al. from the University of Oxford [130]. One major improvement of VGG, when compared to AlexNet is that large kernel-sized filters were replaced with multiple 3×3 kernel-sized filters in the first and second convolution. Two architectures were proposed VGG16 and VGG19. VGG16 model is composed of 16 layers comprising 5 blocks of convolution and max pooling

layers, followed by fully connected layers. Whereas VGG19 consists of 19 layers with extra convolution layers in the last three blocks (See Figure 4.22). VGG19 achieves slightly better performance but requests more memory. VGG-16 was the winner of the ILSVRC challenge in 2014.

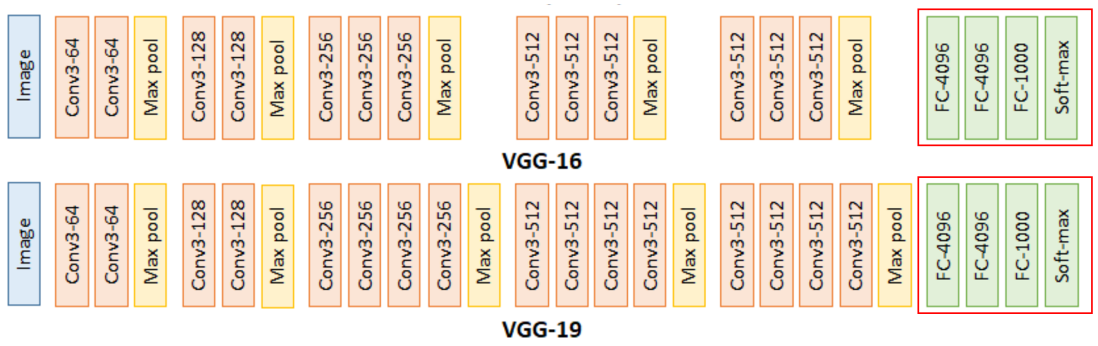

Figure 4.22: VGG-16 Architecture vs VGG-19 Architecture

.

GoogLeNet was proposed by researchers at Google was the winner at ILSRVRC in 2014 [131]. The GoogLeNet architecture is very different from previous state-of-the-art architectures. This model uses several small 1×1 convolutions in the middle of the architecture and global average pooling in order to create deeper architecture while drastically reduce the number of parameters. The deep architecture is composed by of 22 layers and 27 layers with pooling layers included (See Figure 4.23). However, the number of parameters was reduced from 60 million to 4 million in comparison with AlexNet.

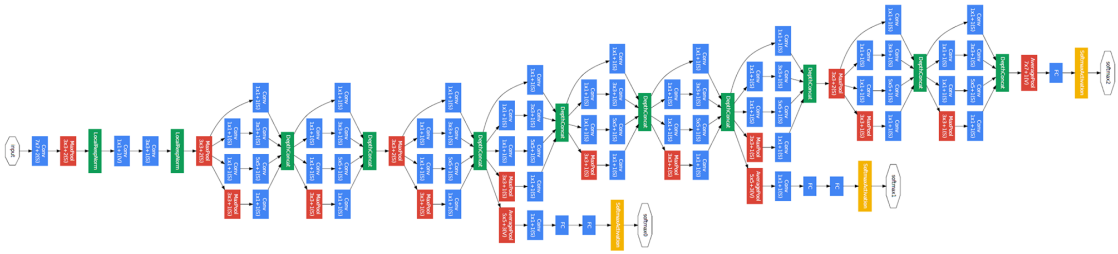

Figure 4.23: GoogLeNet Model Architecture

.

Another famous deep learning CNNs called Residual Neural Network (ResNet) [132] was proposed in 2015 by researchers at Microsoft Research. The network won the top position at the ILSVRC 2015 competition. The strength of the proposed approach is the use of skip connections. The skip connection skips some layers of the model and

connects directly to the output making it possible to train much deeper networks while minimizing the complexity. This network is inspired by VGG architecture in which the shortcut connection is added to convert the architecture into residual network. However, ResNet have fewer filters and lower complexity than a VGG network. ResNet has many variants based on the same concept but have different numbers of layers: ResNet34, ResNet50, ResNet101, ResNet152, etc. The architecture of ResNet50 can be seen in Figure 4.24.

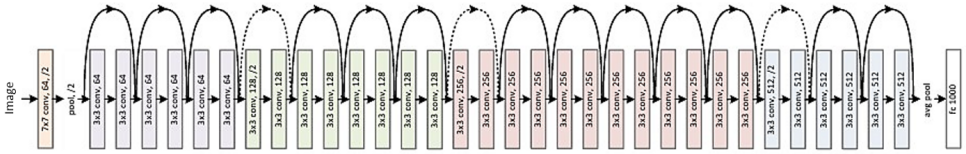

Figure 4.24: ResNet-50 Model Architecture

.

4.3.3 Fully convolutional networks

The relentless success of deep convolutional neural networks for image recognition, detection, and classification over the past few years, motivated computer vision and machine learning researchers to exploit their feature learning capabilities for semantic segmentation problems. Semantic pixel-wise segmentation is more and more being an active topic of research, fueled by the rise of publicly available challenging datasets [133], [134], [135], [136], [137], etc.

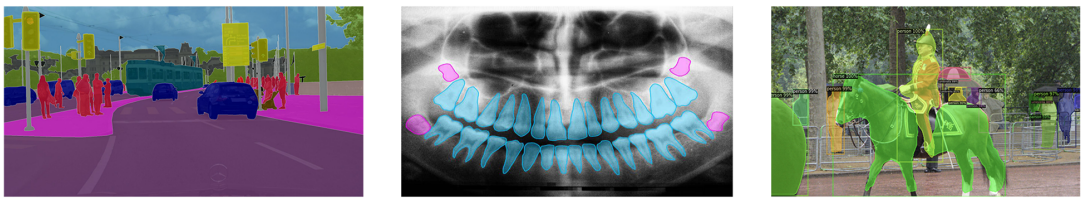

Figure 4.25: Examples of semantic segmentation applications

.

Contrary to image classification where we classify the entire image, in semantic segmentation the classification is performed at a pixel level. Semantic segmentation refers to the process of classifying each pixel belonging to a particular class label in an image

enabling a high-level understanding of the image. It assigns the same label color to all pixels in the image of a specific class. Semantic segmentation has been used successfully for various applications ranging from scene understanding and photo editing to augmented reality, autonomous driving and medical image processing, etc (See Figure 4.25).

In recent years, the great achievements of DL approaches for image classification were quickly transferred to the semantic segmentation task. These powerful deep CNN architectures were directly adapted to be suitable for pixel-wise labelling. Currently, Fully Convolutional neural networks (FCNs) have become the most successful state-of-the-art DL methods particularly designed for semantic segmentation. They only have convolutional and pooling layers which give them the ability to make pixel-wise predictions (See Figure 4.26).

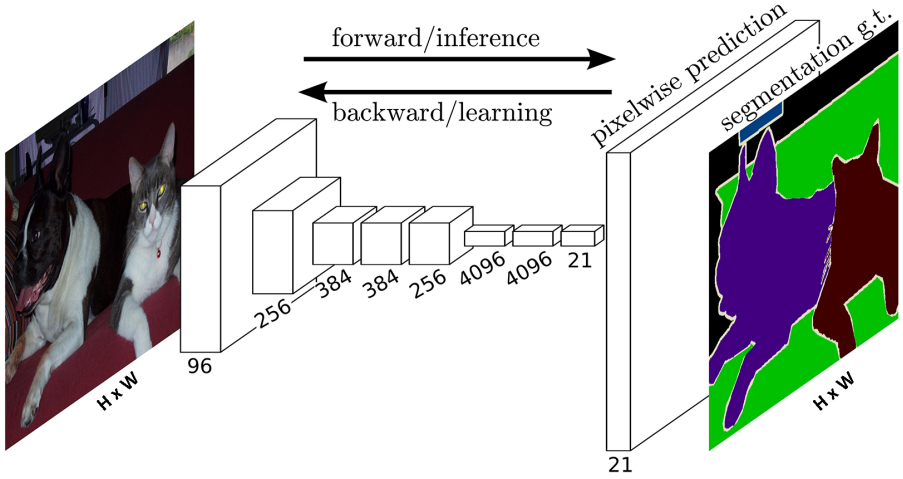

Figure 4.26: Fully Convolutional Network Architecture

.

FCNs are an extension of the classical CNNs introduced by Long et al. in 2015 [138]. The insight of the proposed approach was to take advantage of existing well-known CNN architectures seen in section above and used them as building blocks for these segmentation models, it is the case for: AlexNet, VGG16, GoogLeNet and ResNet. These powerful classification networks were transformed to fully convolutional models by replacing the fully connected layers with convolutional ones to produce dense pixel by pixel labeled outputs instead of classification scores (See Figure 4.27).

The general architecture of fully convolutional networks is based on an encoder-decoder structure where the down-sampling part of the network is called an encoder

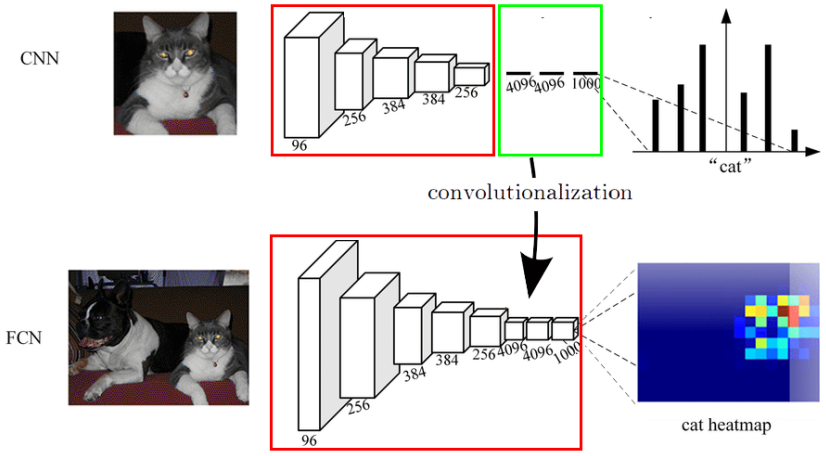

Figure 4.27: Comparison of a CNN for classification and a FCN which creates a heat map after convolutionalization

.

and the up-sampling part is called a decoder. The encoder is adopted for feature extraction while reducing the input spatial resolution which result in a lower resolution feature map. This part of the network typically employs a pre-trained CNN like VGG for example followed by the decoder part. The decoder upsamples low resolution feature representations extracted in encoding phase into a higher resolution to restore to the full-resolution segmentation map. The encoder uses convolution and pooling layers to reduce resolution of the feature map during down-sampling. Whereas up-sampling is mainly done through transpose convolutions to recover the original resolution of the input [139].

Different architectures of FCN have been developed over the last years. Long et al. [138] first proposed three variants of FCNs trained end-to-end for semantic segmentation namely FCN-32s, FCN-16s and FCN-8s. All these three FCN models use same encoder network based on the state-of-the-art VGG-16 that was pre-trained on the large ImageNet object classification dataset which is considered as the backbone architecture. However, the decoder network varies between these architectures, the up-sampling is done using transposed convolutions but with different pixel stride. Therefore, knowledge was directly transferred from VGG16 to perform semantic segmentation by replacing the fully connected layers of VGG-16 with fully convolutional layers using 1x1 convolution which produces a low-resolution heat map. Starting with the first architecture FCN-32, the up-sampling was performed at stride 32, predictions back to the same size of input image in a single step as can be seen in Figure 4.28. The obtained segmentation result

of FCN-32 was very rough and not smooth due to loss of location information during the encoder phase when the resolution of the input was reduced by a factor of 32. Thus, the decoder struggles to produce fine-grained segmentation.

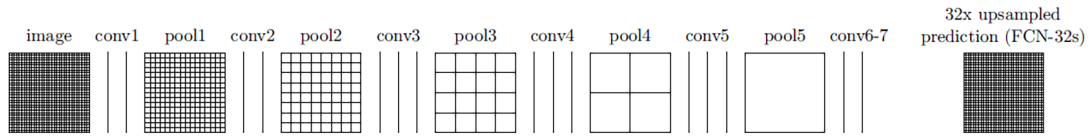

Figure 4.28: FCN-32 architecture with VGG16 backbone

.

To address this issue, the authors proposed adding skip connections in the up-sampling stage from previous layers and summing the two feature maps. This helps to recover more fine-grained details and indeed, generate accurate shapes for segmentation boundaries. Thus, two other architectures were proposed FCN-16 and FCN-8 illustrated in Figure 4.29. FCN-16 combines predictions from both the final layer and the pool4 layer with stride 16 resulting in finer details than FCN-32s (See Figure 4.30). FCN-8 performs even better with more precise boundaries by including predictions from one more previous pooling layer pool3 with stride 8.

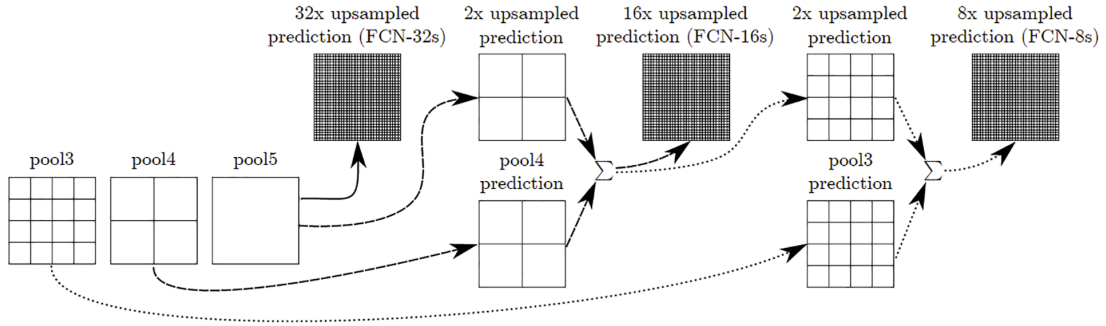

Figure 4.29: FCN-16 and FCN-8 variants architecture combining predictions from both the final layer and previous pooling layers

.

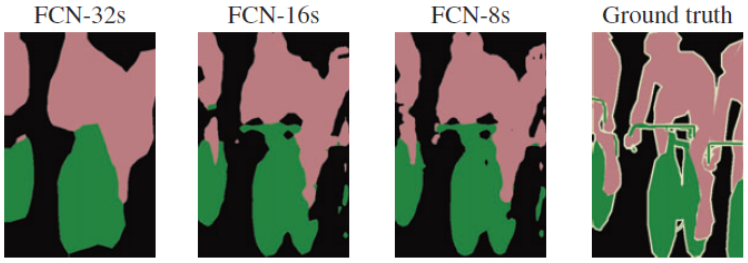

Figure 4.30: Segmentation outputs of FCN with different strides: 32, 16 and 8 [138]

.

Another popular end-to-end FCN architecture for semantic segmentation is U-net [140]. Specifically designed by Ronneberger et al. for biomedical image segmentation purpose to deal with small databases. The network drives its name from the symmetric U-shaped architecture. U-net is an encoder-decoder architecture which consists of a contracting path corresponding to the encoder that capture context followed by the expanding path that enables precise localization using transposed convolutions. The network architecture is illustrated in Figure 4.31.

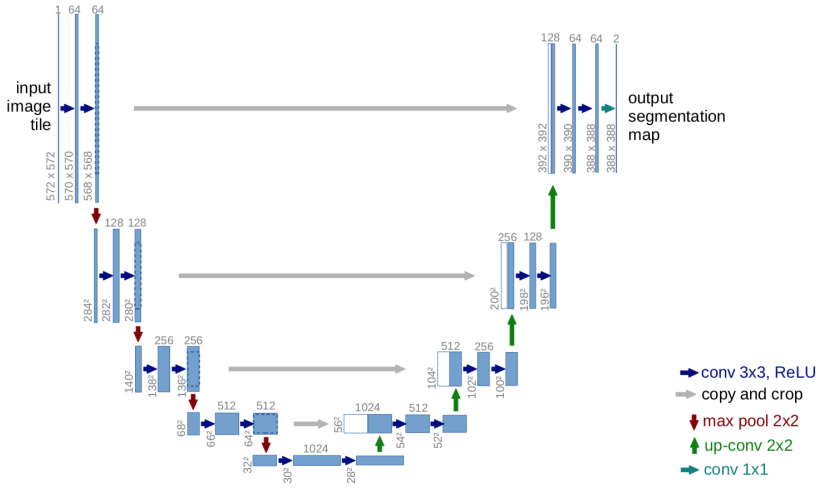

Figure 4.31: U-net Model architecture [140]

.

The contracting path is similar to a traditional convolutional neural network architecture with a successive layers of convolution operation followed by ReLU and then max pooling. The symmetric expanding path consists of deconvolution layers while pooling operators are replaced by up-sampling operators which increase the resolution of the

output. In order to localize, cropped feature map from the contracting path is combined with the up-sampled output. The final layer consists of a 1x1 convolution used to map each 64-component feature vector to the desired number of classes.

Segnet is another FCN architecture proposed in [141]. It is also an encoder-decoder architecture followed by a final pixel-wise classification layer. The encoder part is topologically identical to the a VGG-16 network. The key component of Segnet is the hierarchy of decoder part where each decoder corresponds to an appropriate encoder (See Figure 4.32). Decoders perform up-sampling of the input feature maps using max-pooling indices received from the corresponding encoder. This improved boundary delineation and reduced the number of parameters. The obtained feature maps are then convolved and fed to a softmax classifier to produce the final pixel-wise segmentation for an output with the same resolution as the input image.

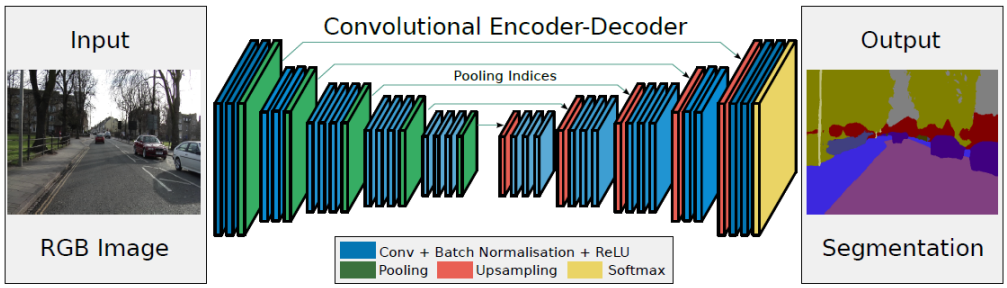

Figure 4.32: SegNet Model architecture [141]

.

In a different approach, DeepLab achieved state-of-the-art semantic segmentation. Four versions were proposed by Liang-Chieh Chen and the Google team. The two first versions of Deeplab, DeepLabv1 [142] and DeepLabv2 [143] use Atrous Convolution and Fully Connected Conditional Random Field (CRF). Atrous convolution also called dilated convolution provide larger output feature map without increasing the number of parameters, whereas CRF help to refine segmentation results. DeepLabv1 was based on the VGG-16 network while in DeepLabv2 VGG-16 was replaced with ResNet which is more complex and expressive. Additionally, DeepLabv2 implements an Atrous Spatial Pyramid Pooling structure (ASPP) to obtain multi-scale context information using parallel atrous convolution layers with different sampling rates. However, since both versions use CRF as a post-processing step, DeepLabv1 and DeepLabv2 are not considered as an end-to-end learning framework. Therefore, authors tried to rethink the

DeepLab architecture to come up with a more enhanced version called DeepLabv3 [144]. DeepLabv3 discards CRF post-processing and improves ASPP. It employs atrous convolution with up-sampled filters to deepen the network. DeepLabV3+ [145] utilizes the output of DeepLabV3 as the encoder output. In addition, DeepLabV3+ consists of a decoder module which improves the segmentation at object boundaries. The architecture of DeepLabV3+ is shown in Figure 4.33.

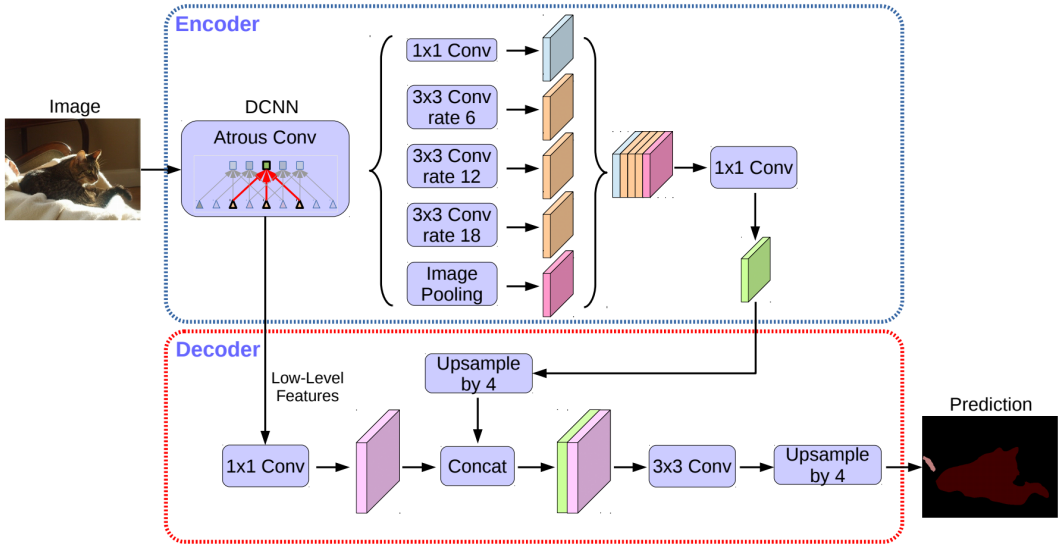

Figure 4.33: DeepLabv3+ Model architecture [145]

.

4.4 Evaluation Metrics

est

5 Wound Segmentation

5.1 Dataset

.....

5.2 Methodology

.....

5.2.1 Dealing with Small Dataset in DL: Data augmentation approach

. . . . . .

5.2.2 DFU segmentation

. . . . . .

5.2.3 Robust background removal with skin correction

. . . . . .

5.3 Results and Discussion

.....

5.4 Application

.....

5.4.1 Smartphone implementation (Min)

. . . . . .

5.4.2 3D chronic wounds segmentation

. . . . . .

5.4.3 Weakly Supervised Hamster's leishmaniasis Segmentation

. . . . . .

6 Tissue Classification

6.1 Dataset

.....

6.2 Methodology

.....

6.2.1 Pixel-wise tissue segmentation

. . . . . .

6.2.2 Superpixel-wise tissue classification

. . . . . .

6.3 Results and Discussion

.....

7 Conclusion and Future Works

Conclusion and Future Works.

7.1 Research Findings

.....

7.2 Future Works

.....

8 References

[1] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova, *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas," *Diabetes research and clinical practice*, vol. 157, p. 107 843, 2019.

[2] W.-P. You and M. Henneberg, "Type 1 diabetes prevalence increasing globally and regionally: The role of natural selection and life expectancy at birth," *BMJ open diabetes research and care*, vol. 4, no. 1, 2016.

[3] L. Guariguata, U. Linnenkamp, J. Beagley, D. Whiting, and N. Cho, "Global estimates of the prevalence of hyperglycaemia in pregnancy," *Diabetes research and clinical practice*, vol. 103, no. 2, pp. 176–185, 2014.

[4] N. C. Schaper, J. J. van Netten, J. Apelqvist, S. A. Bus, R. J. Hinchliffe, B. A. Lipsky, and I. E. Board, "Practical guidelines on the prevention and management of diabetic foot disease (iwgdf 2019 update)," *Diabetes/metabolism research and reviews*, vol. 36, e3266, 2020.

[5] J. Martın-Vaquero, A. Hernández Encinas, A. Queiruga-Dios, J. José Bullón, A. Martınez-Nova, J. Torreblanca González, and C. Bullón-Carbajo, "Review on wearables to monitor foot temperature in diabetic patients," *Sensors*, vol. 19, no. 4, p. 776, 2019.

[6] E. Everett and N. Mathioudakis, "Update on management of diabetic foot ulcers," *Annals of the New York Academy of Sciences*, vol. 1411, no. 1, p. 153, 2018.

[7] D. G. Armstrong, A. J. Boulton, and S. A. Bus, "Diabetic foot ulcers and their recurrence," *New England Journal of Medicine*, vol. 376, no. 24, pp. 2367–2375, 2017.

[8] L. A. V. Cajacuri, "Early diagnostic of diabetic foot using thermal images," PhD thesis, Université D'Orléans, 2013.

[9] K. Alexiadou and J. Doupis, "Management of diabetic foot ulcers," *Diabetes Therapy*, vol. 3, no. 1, p. 4, 2012.

[10] M. Edmonds, "Best practice guidelines: Wound management in diabetic foot ulcers," *Wounds International*, 2013.

[11] T. Wild, M. Prinz, N. Fortner, W. Krois, K. Sahora, S. Stremitzer, and T. Hoelzenbein, "Digital measurement and analysis of wounds based on colour segmentation," *European Surgery*, vol. 40, no. 1, pp. 5–10, 2008.

[12] C. Little, J. McDonald, M. Jenkins, and P. McCarron, "An overview of techniques used to measure wound area and volume," *Journal of wound care*, vol. 18, no. 6, pp. 250–253, 2009.

[13] U. Pavlovčič, J. Diaci, J. Možina, and M. Jezeršek, "Wound perimeter, area, and volume measurement based on laser 3d and color acquisition," *Biomedical engineering online*, vol. 14, no. 1, pp. 1–15, 2015.

[14] G. Schultz, D. Mozingo, M. Romanelli, and K. Claxton, "Wound healing and time; new concepts and scientific applications," *Wound repair and regeneration*, vol. 13, no. 4, S1–S11, 2005.

[15] L. B. Jørgensen, J. A. Sørensen, G. B. Jemec, and K. B. Yderstræde, "Methods to assess area and volume of wounds–a systematic review," *International wound journal*, vol. 13, no. 4, pp. 540–553, 2016.

[16] D. Jun, H. Choi, J. Kim, M. Lee, S. Kim, D. Jo, C. Kim, and D. Shin, "Efficacy of the mobile three-dimensional wound measurement system in pressure ulcer assessment," *Journal of Wound Management and Research*, vol. 15, no. 2, pp. 78–84, 2019.

[17] C. Ahn *et al.*, "Advances in wound photography and assessment methods," *Advances in skin & wound care*, vol. 21, no. 2, pp. 85–93, 2008.

[18] W. Marston, "Endpoint assessment for critical limb ischemia for cellular therapy," in *Regenerative Medicine for Peripheral Artery Disease*, Elsevier, 2016, pp. 137–156.

[19] A. Brown, "The role of debridement in the healing process.," *Nursing times*, vol. 109, no. 40, pp. 16–19, 2013.

[20] E. Nichols, "Describing a wound: From presentation to healing," *Wound Essentials*, vol. 10, no. 1, pp. 56–61, 2015.

[21] D. Gray, R. White, P. Cooper, and A. Kingsley, "Applied wound management and using the wound healing continuum in practice," *Wound Essentials*, vol. 5, pp. 131–139, 2010.

[22] J. Tulloch, R. Zamani, and M. Akrami, "Machine learning in the prevention, diagnosis and management of diabetic foot ulcers: A systematic review," *IEEE Access*, vol. 8, pp. 198 977–199 000, 2020.

[23] H. Wannous, S. Treuillet, and Y. Lucas, "Robust tissue classification for reproducible wound assessment in telemedicine environments," *Journal of Electronic Imaging*, vol. 19, no. 2, p. 023 002, 2010.

[24] A. F. M. Hani, L. Arshad, A. S. Malik, A. Jamil, and F. Y. B. Bin, "Detection and classification of granulation tissue in chronic ulcers," in *International Visual Informatics Conference*, Springer, 2011, pp. 139–150.

[25] *Aranzmedical silhouette*, `https://www.aranzmedical.com/`, 2018.

[26] *Woundworks insight*, `https://woundworks.com/`, 2018.

[27] P. Plassmann and T. Jones, "Mavis: A non-invasive instrument to measure area and volume of wounds," *Medical engineering & physics*, vol. 20, no. 5, pp. 332–338, 1998.

[28] *Wound zoom*, `https://perceptivesol.com/`.

[29] C. Chakraborty, "Computational approach for chronic wound tissue characterization," *Informatics in Medicine Unlocked*, vol. 17, p. 100 162, 2019.

[30] Y. Lucas, R. Niri, S. Treuillet, H. Douzi, and B. Castaneda, "Wound size imaging: Ready for smart assessment and monitoring," *Advances in wound care*, 2020.

[31] D. Marijanović and D. Filko, "A systematic overview of recent methods for non-contact chronic wound analysis," *Applied Sciences*, vol. 10, no. 21, p. 7613, 2020.

[32] D. Anisuzzaman, C. Wang, B. Rostami, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Image based artificial intelligence in wound assessment: A systematic review," *arXiv preprint arXiv:2009.07141*, 2020.

[33] M. Dujovny, "Pressure injury evolution: Mobile wound analyzer review," *J Biol Med Res*, vol. 2, no. 2, p. 12, 2018.

[34] N. Hettiarachchi, R. Mahindaratne, G. Mendis, H. Nanayakkara, and N. D. Nanayakkara, "Mobile based wound measurement," in *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, IEEE, 2013, pp. 298–301.

[35] M. H. Yap, K. E. Chatwin, C.-C. Ng, C. A. Abbott, F. L. Bowling, S. Rajbhandari, A. J. Boulton, and N. D. Reeves, "A new mobile application for standardizing diabetic foot images," *Journal of diabetes science and technology*, vol. 12, no. 1, pp. 169–173, 2018.

[36] M. Goyal, N. D. Reeves, S. Rajbhandari, and M. H. Yap, "Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1730–1741, 2018.

[37] S. Wang, Q. Zhang, W. Huang, H. Tian, J. Hu, Y. Cheng, and Y. Peng, "A new smart mobile system for chronic wound care management," *IEEE Access*, vol. 6, pp. 52 355–52 365, 2018.

[38] R. Mukherjee, D. D. Manohar, D. K. Das, A. Achar, A. Mitra, and C. Chakraborty, "Automated tissue classification framework for reproducible chronic wound assessment," *BioMed research international*, vol. 2014, 2014.

[39] M. K. Yadav, D. D. Manohar, G. Mukherjee, and C. Chakraborty, "Segmentation of chronic wound areas by clustering techniques using selected color space," *Journal of Medical Imaging and Health Informatics*, vol. 3, no. 1, pp. 22–29, 2013.

[40] M. Saratha and V. Mohana Priya, "Detection of diabetic wounds based on segmentation using accelerated mean shift algorithm," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 6, no. 2, pp. 201–206, 2016.

[41] L. Wang, P. C. Pedersen, D. M. Strong, B. Tulu, E. Agu, and R. Ignotz, "Smartphone-based wound assessment system for patients with diabetes," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 477–488, 2014.

[42] E. S. Papazoglou, L. Zubkov, X. Mao, M. Neidrauer, N. Rannou, and M. S. Weingarten, "Image analysis of chronic wounds for determining the surface area," *Wound repair and regeneration*, vol. 18, no. 4, pp. 349–358, 2010.

[43] M.-C. Chang, T. Yu, J. Luo, K. Duan, P. Tu, Y. Zhao, N. Nagraj, V. Rajiv, M. Priebe, E. A. Wood, *et al.*, "Multimodal sensor system for pressure ulcer wound assessment and care," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 1186–1196, 2017.

[44] P. Gholami, M. A. Ahmadi-Pajouh, N. Abolftahi, G. Hamarneh, and M. Kayvanrad, "Segmentation and measurement of chronic wounds for bioprinting," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1269–1277, 2017.

[45] M. Jena, S. P. Mishra, and D. Mishra, "A survey on applications of machine learning techniques for medical image segmentation," *Internationa Journal of Engineering & Technology*, vol. 10, no. x, pp. xxx–xxx, 2018.

[46] F.-A. Georgescu, D. Răducanu, and M. Datcu, "New mpeg-7 scalable color descriptor based on polar coordinates for multispectral earth observation image analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 987–991, 2017.

[47] M. Pietikäinen, "Local binary patterns," *Scholarpedia*, vol. 5, no. 3, p. 9775, 2010.

[48] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Ieee, vol. 1, 2005, pp. 886–893.

[49] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer vision, graphics, and image processing*, vol. 44, no. 1, pp. 87–116, 1988.

[50] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," *Stroke and vascular neurology*, vol. 2, no. 4, 2017.

[51] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.

[52] M. Kolesnik and A. Fexa, "Multi-dimensional color histograms for segmentation of wounds in images," in *International Conference Image Analysis and Recognition*, Springer, 2005, pp. 1014–1022.

[53] M. Kolesnik and A. Fexa, "Segmentation of wounds in the combined color-texture feature space," in *Medical imaging 2004: Image processing*, International Society for Optics and Photonics, vol. 5370, 2004, pp. 549–556.

[54] M. Kolesnik and A. Fexa, "How robust is the svm wound segmentation?" In *Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006*, IEEE, 2006, pp. 50–53.

[55] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu, "Area determination of diabetic foot ulcer images using a cascaded two-stage svm-based classification," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2098–2109, 2016.

[56] T. Biswas, M. F. A. Fauzi, F. S. Abas, and H. K. Nair, "Superpixel classification with color and texture features for automated wound area segmentation," in *2018 IEEE Student Conference on Research and Development (SCOReD)*, IEEE, 2018, pp. 1–6.

[57] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Science and Information Conference*, Springer, 2019, pp. 128–144.

[58] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[59] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[60] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology*, vol. 10, no. 3, pp. 257–273, 2017.

[61] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, "Deep learning in medical imaging: General overview," *Korean journal of radiology*, vol. 18, no. 4, p. 570, 2017.

[62] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.

[63] M. Biswas, V. Kuppili, L. Saba, D. R. Edla, H. S. Suri, E. Cuadrado-Godia, J. R. Laird, R. T. Marinhoe, J. M. Sanches, A. Nicolaides, *et al.*, "State-of-the-art review on deep learning in medical imaging.," *Frontiers in bioscience (Landmark edition)*, vol. 24, pp. 392–426, 2019.

[64] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, 2021.

[65] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimedia Tools and Applications*, pp. 1–34, 2021.

[66] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg, "Fully convolutional networks for diabetic foot ulcer segmentation," in *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, IEEE, 2017, pp. 618–623.

[67] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H. Yap, "Dfunet: Convolutional neural networks for diabetic foot ulcer classification," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 5, pp. 728–739, 2018.

[68] V. N. Shenoy, E. Foster, L. Aalami, B. Majeed, and O. Aalami, "Deepwound: Automated postoperative wound assessment and surgical site surveillance through convolutional neural networks," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2018, pp. 1017–1021.

[69] B. Rostami, D. Anisuzzaman, C. Wang, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Multiclass wound image classification using an ensemble deep cnn-based classifier," *Computers in Biology and Medicine*, p. 104 536, 2021.

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

[71] S. K. Das, P. Roy, and A. K. Mishra, "Recognition of ischaemia and infection in diabetic foot ulcer: A deep convolutional neural network based approach," *International Journal of Imaging Systems and Technology*, 2021.

[72] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, "Dfu_qutnet: Diabetic foot ulcer classification using novel deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15 655–15 677, 2020.

[73] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, J. Santamarıa, Y. Duan, and S. R Oleiwi, "Towards a better understanding of transfer learning for medical imaging: A case study," *Applied Sciences*, vol. 10, no. 13, p. 4523, 2020.

[74] D. Anisuzzaman, Y. Patel, J. Niezgoda, S. Gopalakrishnan, and Z. Yu, "A mobile app for wound localization using deep learning," *arXiv preprint arXiv:2009.07133*, 2020.

[75] J. Amin, M. Sharif, M. A. Anjum, H. U. Khan, M. S. A. Malik, and S. Kadry, "An integrated design for classification and localization of diabetic foot ulcer based on cnn and yolov2-dfu models," *IEEE Access*, vol. 8, pp. 228 586–228 597, 2020.

[76] B. Cassidy, N. D. Reeves, P. Joseph, D. Gillespie, C. O'Shea, S. Rajbhandari, A. G. Maiya, E. Frank, A. Boulton, D. Armstrong, *et al.*, "Dfuc2020: Analysis towards diabetic foot ulcer detection," *arXiv preprint arXiv:2004.11853*, 2020.

[77] M. H. Yap, R. Hachiuma, A. Alavi, R. Brüngel, B. Cassidy, M. Goyal, H. Zhu, J. Rückert, M. Olshansky, X. Huang, *et al.*, "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation," *Computers in Biology and Medicine*, p. 104 596, 2021.

[78] A. L. da Costa Oliveira, A. B. de Carvalho, and D. O. Dantas, "Faster r-cnn approach for diabetic foot ulcer detection.," in *VISIGRAPP (4: VISAPP)*, 2021, pp. 677–684.

[79] M. Goyal and S. Hassanpour, "A refined deep learning architecture for diabetic foot ulcers detection," *arXiv preprint arXiv:2007.07922*, 2020.

[80] F. Li, C. Wang, X. Liu, Y. Peng, and S. Jin, "A composite model of wound segmentation based on traditional methods and deep neural networks," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[81] C. Wang, D. Anisuzzaman, V. Williamson, M. K. Dhar, B. Rostami, J. Niezgoda, S. Gopalakrishnan, and Z. Yu, "Fully automatic wound segmentation with deep convolutional neural networks," *Scientific Reports*, vol. 10, no. 1, pp. 1–9, 2020.

[82] C. Cui, K. Thurnhofer-Hemsi, R. Soroushmehr, A. Mishra, J. Gryak, E. Domınguez, K. Najarian, and E. López-Rubio, "Diabetic wound segmentation using convolutional neural networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 1002–1005.

[83] H. Gamage, W. Wijesinghe, and I. Perera, "Instance-based segmentation for boundary detection of neuropathic ulcers through mask-rcnn," in *International Conference on Artificial Neural Networks*, Springer, 2019, pp. 511–522.

[84] S. Zahia, B. Garcia-Zapirain, and A. Elmaghraby, "Integrating 3d model representation for an accurate non-invasive assessment of pressure injuries with deep learning," *Sensors*, vol. 20, no. 10, p. 2933, 2020.

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

[86] J. Arnqvist, J. Hellgren, and J. Vincent, "Semiautomatic classification of secondary healing ulcers in multispectral images," in *9th International Conference on Pattern Recognition*, IEEE Computer Society, 1988, pp. 459–460.

[87] S. Zahia, M. B. G. Zapirain, X. Sevillano, A. González, P. J. Kim, and A. Elmaghraby, "Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods," *Artificial intelligence in medicine*, vol. 102, p. 101 742, 2020.

[88] V. Godeiro, J. S. Neto, B. Carvalho, B. Santana, J. Ferraz, and R. Gama, "Chronic wound tissue classification using convolutional networks and color space reduction," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, 2018, pp. 1–6.

[89] M. F. A. Fauzi, I. Khansa, K. Catignani, G. Gordillo, C. K. Sen, and M. N. Gurcan, "Segmentation and automated measurement of chronic wound images: Probability map approach," in *Medical Imaging 2014: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 9035, 2014, p. 903 507.

[90] A. A. Perez, A. Gonzaga, and J. M. Alves, "Segmentation and analysis of leg ulcers color images," in *Proceedings International Workshop on Medical Imaging and Augmented Reality*, IEEE, 2001, pp. 262–266.

[91] J. Delode, E. Rosow, C. Roth, J. Adams, and F. Langevin, "A wound-healing monitoring system," *ITBM-RBM*, vol. 22, no. 1, pp. 49–52, 2001.

[92] H. Wannous, S. Treuillet, and Y. Lucas, "Supervised tissue classification from color images for a complete wound assessment tool," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2007, pp. 6031–6034.

[93] H. Wannous, Y. Lucas, and S. Treuillet, "Enhanced assessment of the wound-healing process by accurate multiview tissue classification," *IEEE transactions on Medical Imaging*, vol. 30, no. 2, pp. 315–326, 2010.

[94] ——, "Efficient svm classifier based on color and texture region features for wound tissue images," in *Medical Imaging 2008: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 6915, 2008, 69152T.

[95] L. Wang, P. C. Pedersen, D. M. Strong, B. Tulu, E. Agu, R. Ignotz, and Q. He, "An automatic assessment system of diabetic foot ulcers based on wound area determination, color segmentation, and healing score evaluation," *Journal of diabetes science and technology*, vol. 10, no. 2, pp. 421–428, 2016.

[96] S. Patel, R. Patel, and D. Desai, "Diabetic foot ulcer wound tissue detection and classification," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, IEEE, 2017, pp. 1–5.

[97] F. Veredas, H. Mesa, and L. Morente, "Binary tissue classification on wound images with neural networks and bayesian classifiers," *IEEE transactions on medical imaging*, vol. 29, no. 2, pp. 410–427, 2009.

[98] F. J. Veredas, R. M. Luque-Baena, F. J. Martın-Santos, J. C. Morilla-Herrera, and L. Morente, "Wound image evaluation with machine learning," *Neurocomputing*, vol. 164, pp. 112–122, 2015.

[99] C. Chakraborty, B. Gupta, S. K. Ghosh, D. K. Das, and C. Chakraborty, "Telemedicine supported chronic wound tissue prediction using classification approaches," *Journal of medical systems*, vol. 40, no. 3, p. 68, 2016.

[100] A. Khalil, M. Elmogy, M. Ghazal, C. Burns, and A. El-Baz, "Chronic wound healing assessment system based on different features modalities and non-negative matrix factorization (nmf) feature reduction," *IEEE Access*, vol. 7, pp. 80 110–80 121, 2019.

[101] S. Zahia, D. Sierra-Sosa, B. Garcia-Zapirain, and A. Elmaghraby, "Tissue classification and segmentation of pressure injuries using convolutional neural networks," *Computer methods and programs in biomedicine*, vol. 159, pp. 51–58, 2018.

[102] V. Rajathi, R. Bhavani, and G. Wiselin Jiji, "Varicose ulcer (c6) wound image tissue classification using multidimensional convolutional neural networks," *The Imaging Science Journal*, vol. 67, no. 7, pp. 374–384, 2019.

[103] H. Nejati, H. A. Ghazijahani, M. Abdollahzadeh, T. Malekzadeh, N.-M. Cheung, K.-H. Lee, and L.-L. Low, "Fine-grained wound tissue analysis using deep neural network," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 1010–1014.

[104] G. Blanco, A. J. Traina, C. Traina Jr, P. M. Azevedo-Marques, A. E. Jorge, D. de Oliveira, and M. V. Bedo, "A superpixel-driven deep learning approach for the analysis of dermatological wounds," *Computer methods and programs in biomedicine*, vol. 183, p. 105 079, 2020.

[105] M. Maity, D. Dhane, C. Bar, C. Chakraborty, and J. Chatterjee, "Pixel-based supervised tissue classification of chronic wound images with deep autoencoder," in *Advanced Computational and Communication Paradigms*, Springer, 2018, pp. 727–735.

[106] N. Pholberdee, C. Pathompatai, and P. Taeprasartsit, "Study of chronic wound image segmentation: Impact of tissue type and color data augmentation," in *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, 2018, pp. 1–6.

[107] B. Garcıa-Zapirain, M. Elmogy, A. El-Baz, and A. S. Elmaghraby, "Classification of pressure ulcer tissues with 3d convolutional neural network," *Medical & biological engineering & computing*, vol. 56, no. 12, pp. 2245–2258, 2018.

[108] M. Elmogy, B. Garcıa-Zapirain, C. Burns, A. Elmaghraby, and A. Ei-Baz, "Tissues classification for pressure ulcer images based on 3d convolutional neural network," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 3139–3143.

[109] M. Elmogy, B. Garcıa-Zapirain, A. S. Elmaghraby, and A. El-Baz, "An automated classification framework for pressure ulcer tissues based on 3d convolutional neural network," in *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, 2018, pp. 2356–2361.

[110] `https://www.flir.com/products/flir-one-pro/?model=435-0007-02`.

[111] H. Wannous, Y. Lucas, S. Treuillet, and B. Albouy, "A complete 3d wound assessment tool for accurate tissue classification and measurement," in *2008 15th IEEE International Conference on Image Processing*, IEEE, 2008, pp. 2928–2931.

[112] M. F. A. Fauzi, I. Khansa, K. Catignani, G. Gordillo, C. K. Sen, and M. N. Gurcan, "Segmentation and management of chronic wound images: A computer-based approach," in *Chronic Wounds, Wound Dressings and Wound Healing*, Springer, 2018, pp. 115–134.

[113] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Computer Vision, IEEE International Conference on*, IEEE Computer Society, vol. 2, 2003, pp. 10–10.

[114] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.

[115] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[116] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," Tech. Rep., 2010.

[117] P. Neubert and P. Protzel, "Compact watershed and preemptive slic: On improving trade-offs of superpixel segmentation algorithms," in *2014 22nd international conference on pattern recognition*, IEEE, 2014, pp. 996–1001.

[118] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *European conference on computer vision*, Springer, 2008, pp. 705–718.

[119] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[120] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697.

[121] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[122] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[123] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[124] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[125] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[126] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.

[127] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[128] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[129] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[130] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

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

[132] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[133] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Letters*, vol. 30, no. 2, pp. 88–97, 2009.

[134] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European conference on computer vision*, Springer, 2012, pp. 746–760.

[135] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 2012, pp. 3354–3361.

[136] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, Springer, 2014, pp. 740–755.

[137] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.

[138] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[139] Z. Wojna, V. Ferrari, S. Guadarrama, N. Silberman, L.-C. Chen, A. Fathi, and J. Uijlings, "The devil is in the decoder: Classification, regression and gans," *International Journal of Computer Vision*, vol. 127, no. 11, pp. 1694–1706, 2019.

[140] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[141] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[142] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[143] ——, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[144] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

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

9 Discussed Papers

10 Acknowledgments

Write your acknowledgments.