UNIVERSIDADE D
COIMBRA

ANA CATARINA FONTES DE OLIVEIRA

Segmentation of lungs on CT : tools to aid Radiotherapy planning

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Pedro H. Abreu (Advisor)
Inês Domingues (Co-Advisor)

Coimbra, September 2019

This work was developed in collaboration with:

Centre for Informatics and Systems of the University of Coimbra

Instituto Português de Oncologia do Porto

Abstract

Radiotherapy planning plays a decisive role in cancer's treatment management. Many side effects come from the exposure of normal tissue to radiation during therapy, all the way from small acute side effects such as tiredness, to long term sequelae, like another type of cancer. To minimize this exposure, it is necessary to contour the organs at risk. However, this task is typically performed manually on a slice-by-slice basis, being consequently very time-consuming and susceptible to high intra and inter-subject variance and human errors.

In this way, this line of work aims to help the clinicians in this difficult and repetitive task by implementing algorithms that delineate non-pathological lungs. In this project, two lung segmentation algorithms are presented for Computed Tomography scans: the Iterative Region Growing algorithm and a U-Net Convolutional Neural Network model. One relies on image processing techniques, as intensity projection and region growing. This pipeline starts by isolating each lung. Then, three techniques for seed placement are explored. Lastly, an update on the traditional region growing algorithm is developed, allowing it to automatically discover the best threshold parameter value for each case. The other algorithm is a U-Net deep learning architecture model, that takes advantage of the distinctive ability of Convolutional Neural Networks to find hidden patterns present in the lungs without requiring feature extraction and selection.

The results obtained for the three different techniques for seed placement were, respectively, 74%, 74% and 92% of DICE with the Iterative Region Growing algorithm. The results for the U-Net model were 91% for the same metric.

Future work includes more tests on bigger and more diverse databases, analyzing the effect of morphology operations on the results and the effect of the hyperparameter optimization techniques on the network.

Key-words: Automatic Segmentation, Radiotherapy Planning, Organs at Risk, Lungs, 3D, Computerized Tomography.

Resumo

O planeamento da radioterapia desempenha um papel decisivo na gestão do tratamento do cancro. Múltiplos efeitos secundários advêm da exposição de tecido normal à radiação durante a terapia, desde cansaço a outros tipos de cancro. De modo a minimizar esta exposição, é necessário fazer o contorno dos órgãos em risco. No entanto, esta é uma tarefa tipicamente executada manualmente, fatia a fatia, e, portanto, demorada e suscetível a erros humanos e grandes variações intra e inter-individuais.

Deste modo, este estudo tem como objetivo ajudar os clínicos numa tarefa tão difícil e repetitiva como esta, através da implementação de algoritmos para a delimitação de pulmões não-patológicos. Neste projeto, dois algoritmos de segmentação para pulmões são apresentados para exames de Tomografia Computorizada: o algoritmo iterativo de *region growing* e uma Rede Neuronal Convolucional U-Net. Um deles é baseado em técnicas de processamento de imagem, como a projeção de intensidades e o *region growing*. Esta pipeline começa por isolar cada pulmão. Depois, três técnicas para colocação da semente são exploradas. Por último, uma alteração no algoritmo standard de *region growing* é criada, permitindo que este descubra automaticamente o valor do parâmetro de *threshold* mais adequado para cada caso. O outro algoritmo é um modelo de *deep learning* com uma arquitetura em U-Net, que explora as capacidades das Redes Neuronais Convolucionais para encontrar padrões escondidos presentes nos pulmões sem exigir a extração e seleção de *features*.

Os resultados obtidos para as três técnicas de posicionamento da semente foram, respetivamente, 74%, 74% e 92% de DICE para o algoritmo Iterativo de *Region Growing*. Os resultados para o modelo U-Net foram 91% para a mesma métrica.

No futuro, mais testes poderiam ser feitos em bases de dados maiores e mais diversas, estudando o efeito de operações morfológicas nos resultados e de técnicas de optimização de hiperparâmetros da rede neuronal.

Palavras-Chave: Segmentação Automática, Planeamento de Radioterapia, Órgãos em Risco, Pulmões, 3D, Tomografia Computorizada.

Acknowledgments

Gostaria de expressar o meu sincero agradecimento às seguintes pessoas:

Ao Professor Pedro Henriques Abreu, por me ter dado a oportunidade de trabalhar com ele, pela honestidade e confiança demonstrada e pela responsabilidade depositada. Acredito que todos os desafios colocados e perguntas levantadas foram imprescindíveis para este projeto e para o meu desenvolvimento academico e profissional. Foi um gosto conhecer as pessoas deste laboratório e de aprender com elas. E por isso estou extremamente grata.

À Inês Domingues, por todo o apoio, conhecimento, pela celebração das pequenas vitorias e incentivo quando as coisas não corriam tão bem. As suas questões, rigor, confiança, paciência e conhecimento deram-me sempre força para continuar e questionar.

À Joana, gostaria de agradecer o apoio, confiança e oportunidade. Os seus ensinamentos e paciência para me ajudar a começar foram imprescindiveis, assim como a simpatia. Ao Igor, por todo o empenho em me atualizar na cultura asiatica, pelo entusiasmo e paciência! Um obrigada aos dois por todo o companheirismo. Que tenham toda a sorte do mundo no futuro que se avizinha!

À minha família, por todo o apoio, por todo o empenho, incentivo e motivação. Por serem um exemplo e me darem sempre força e confiança para explorar. Não ha palavras para demonstrar todo o agradecimento.

Ao João Roque, por la estares, sempre. Por todos os conselhos, toda a motivaç ão, força e por me deixares sonhar. Estou extremamemente grata.

Aos meus amigos, por todos os momentos que me proporcionaram, por acreditarem em mim e me apoiarem, um profundo obrigada.

Contents

List of Tables xiii

List of Figures xv

List of Abbreviations xviii

1 Introduction 1
- 1.1 Contextualization . . . 1
- 1.2 Motivation . . . 2
- 1.3 Objectives . . . 3
- 1.4 Research Contributions . . . 3
- 1.5 Document Structure . . . 3

2 Background Knowledge 5
- 2.1 Medical Image modalities . . . 5
- 2.2 Preprocessing techniques . . . 5
  - 2.2.1 Contrast stretching or Normalization . . . 6
  - 2.2.2 Histogram equalization . . . 6
  - 2.2.3 Adaptive Histogram equalization . . . 6
  - 2.2.4 Resize . . . 7
  - 2.2.5 Resampling . . . 7
  - 2.2.6 Morphology operations . . . 8
    - 2.2.6.1 Dilation . . . 8
    - 2.2.6.2 Erosion . . . 8
    - 2.2.6.3 Opening & Closing . . . 8
- 2.3 Segmentation algorithms . . . 9
  - 2.3.1 Thresholding approaches . . . 9
  - 2.3.2 Deformable models . . . 9
  - 2.3.3 Atlas-guided approaches . . . 10
  - 2.3.4 Markov Random Field models . . . 11
  - 2.3.5 Clustering approaches . . . 11
  - 2.3.6 Region Growing . . . 11

2.3.7 Classifiers . . . 12
2.4 Data augmentation techniques . . . 17
2.5 Performance metrics . . . 18
2.6 Statistical tests . . . 20

3 State of the Art **21**
3.1 Other segmented areas . . . 28
3.2 Conclusions . . . 30

4 Material and Methods **33**
4.1 Dataset Description . . . 33
4.2 Region Growing . . . 34
4.2.1 Separation between right and left lungs . . . 35
4.2.2 Placement of the seed . . . 36
4.2.2.1 Method 1 . . . 36
4.2.2.2 Method 2 . . . 37
4.2.2.3 Method 3 . . . 37
4.2.3 Iterative Region Growing . . . 38
4.3 U-Net . . . 40

5 Results and Discussion **43**
5.1 Region Growing . . . 43
5.2 U-Net . . . 46
5.3 Comparison . . . 49
5.4 Overview . . . 51

6 Conclusions and Future work **55**

Appendices **59**
A Morphology operations . . . 61
B Equalization studies . . . 63
C Results of Comparison . . . 65
D Scientific Paper . . . 69

Bibliography **71**

List of Tables

3.1 Overview of the works presented. 32

4.1 Number of manually annotated regions of interest of the dataset. 33

5.1 Seed placement performance 43
5.2 Segmentation performance by Region Growing 44
5.3 Segmentation performance by Region Growing, considering valid seeds only 45
5.4 Parameters tested in different configurations. 46
5.5 U-Net Configurations 47
5.6 Results for each U-Net configuration 48
5.7 Results of the U-Net 48
5.8 Comparison of Iterative Region Growing and U-Net results 50
5.9 Comparison of the algorithms 53

List of Figures

2.1 Neural Network 13
2.2 Convolutional Neural Network 15
2.3 U Net architecture 17

4.1 Ground truth examples 34
4.2 Lung Segmentation pipeline - Region Growing technique. 34
4.3 Mask HU 35
4.4 Separation between right and left lungs 36
4.5 Resulting volumes after the separation 36
4.6 Seed location by Methods 1 and 2 37
4.7 Planes of the image 38
4.8 Seed location by Method 3 38
4.9 Lung Segmentation methodology - Region Growing technique. 40
4.10 U Net architecture 41

5.1 Example of the results of lung segmentation by Region Growing 45
5.2 Validation loss graph 49
5.3 Boxplot of U-Net CNN results 49
5.4 Boxplot comparing U-Net and Iterative Region Growing 50
5.5 Example of segmentation comparing U-Net and Iterative Region Growing . 51

A.1 Morphological operation example 61

B.1 Equalization studies for patient 13 63

C.1 Patient 1 results 65
C.2 Patient 2 results 65
C.3 Patient 3 results 66
C.4 Patient 4 results 66
C.5 Patient 5 results 66
C.6 Patient 6 results 66
C.7 Patient 7 results 67
C.8 Patient 8 results 67

C.9 Patient 9 results 67
C.10 Patient 10 results 67
C.11 Patient 11 results 68
C.12 Patient 12 results 68
C.13 Patient 13 results 68
C.14 Patient 14 results 68

List of Abbreviations

3D ACACM 3D Adaptive Crisp Active Contour Model.

ABD Absolute Border Distance.
ACM Active Contour Model.
ANN Artificial Neural Networks.
AVD Percentage Absolute Volumetric Difference.

BraTS Brain Tumour Segmentation.

CNN Convolutional Neural Network.
CT Computed Tomography.

DICE Dice coefficient.
DL Deep Learning.
DRLSE Distance Regularized Level Set Evolution.

EACM Edge-based Active Contour Model.

FCN Fully Convolutional Networks.

HD Hausdorff Distance.
HGG High-Grade Glioblastomas.
HU Hounsfield Units.

IOU Intersection Over Union.
IPO Institute of Oncology of Porto.
IRG Iterative Region Growing.
ITK Insight Toolkit.

JAC Jaccard index.

LBF Local Binary Fitting.
LCV Local Chan–Vese model.
LGDF Local Gaussian Distribution Fitting.
LGG Low-Grade Gliomas.
LIF Local Image Fitting.

LRBAC Localizing Region-Based Active Contour.

MAD Mean Absolute Distance.
MRF Markov Random Field.
MRI Magnetic Ressonance Imaging.

OAR Organs at Risk.
OR Oversegmentation Rate.

PRV Planning Organ at Risk Volume.
PVD Percentage of Volume Difference.
PVO Percentage of Volume Overlap.

RACM Region-Based Active Contour Model.
RG Region growing.
ROI Regions of Interest.
RP Radiotherapy Planning.

Sp Specificity.

TPr True Positive rate.

UR Undersegmentation Rate.

VS Volumetric Similarity.
VTK Visualization Toolkit.

1

Introduction

Medical imaging is the process of representing the human body, at a microscopic level (the cells and tissue) or macroscopic level (the organs and its disposition in the body), in an anatomical (body disposition) or functional (body metabolism) way. The images can be captioned by different means: by light (p.e.: endoscopy, OCT), sound (p.e.:ultrasound), magnetism (p.e.:MRI), radioactive pharmaceuticals (p.e.:PET, SPECT) or X-rays (p.e.:CT). With the goal of obtaining more accurate diagnostics or therapeutic assessments, image processing techniques can be applied to these representations, like segmentation [1].

Segmentation is the process of partitioning an image into its constituent regions or objects, or, the other way around, to group individual elements present in an image by similar properties, like texture or context. Generally speaking, segmentation can be a difficult, labour-intensive, error-prone and time-consuming procedure [1].

1.1 Contextualization

Radiation therapy has a dominant role in cancer treatment and has always been a major part of the effort to cure cancer patients. The main goal of radiotherapy is to deliver a prescribed dose to the target volume while sparing normal tissue [2]. Since radiotherapy is a personalized and localized treatment, not only the definition of the tumour and target volumes are vital to its successful execution, but also the definition of Organs at Risk (OAR), in order to minimize its exposure [3]. Contouring these regions is, however, a time-consuming part of radiotherapy treatment planning [4] since in current clinical practice, this important task is typically performed visually on a slice-by-slice basis with very limited support of automated segmentation tools, thus being prone to high intra and inter-subject variance and other human errors.

Diagnostic imaging is crucial in Radiotherapy Planning (RP) [5]. Computed Tomography (CT) is usually used as the basis for radiotherapy for two main reasons: (1) since it includes density information, it enables the improvement of dosimetry calculations accuracy, and (2) allows to locate the patient in relation to the treatment machine, as it is more reliable on illustrating the shape and position when compared with other image modali-

ties. Besides, as opposed to other imaging procedures, the patient can be scanned in the treatment position, which can be an advantage [5]. Therefore, this modality can play an important role in the early detection and treatment management of different types of cancer, such as lung cancer [5].

1.2 Motivation

In 2019, 1,762,450 new cancer cases and 606,880 cancer deaths are projected to occur in the United States [6]. Cancer is one of the deadliest diseases in Europe, the second-largest cause of death [6]. There have been significant advances in the treatment against this group of diseases over time [6, 7]. However, assuming the absence of competing causes of death, the cumulative risk of developing cancer before the age of 75 years is still 21.4% and, for dying, 17.7%, globally, thus representing such a concern in public health and an enormous burden in developed societies [7]. Thus, early diagnosis and treatment would help to minimize such a burden and concern.

At least half of the cancer patients require radiotherapy at some time in their condition and, according to the Cancer Services Collaborative, radiotherapy alone is responsible for 78% of non-surgical cancer cures [3]. When in treatment of cancer with radiotherapy, tumour and normal tissue are intricately linked, so when a radiation dose is applied to the tumour, some normal tissue will be inevitable irradiated [5]. If radiation in this tissue is lowered, whether in volume or dose, then a reduction in number and severity of complications will result from the therapy. Therefore, detail in RP is vital to the effectiveness and outcome of therapy. To decrease the level of radiation in normal tissue, in some circumstances a margin of uncertainty around an OAR is added, creating a Planning Organ at Risk Volume (PRV), such as the case of the spinal cord.

Although this project uses private data acquired from CT scans of patients with Hodgkin lymphoma, provided by the Institute of Oncology of Porto (IPO), our focus was on the segmentation of lungs tissue. Sensitivity to toxicities of the lungs and heart limits the radiation dose escalation in some tumours, like the oesophagal or gastroesophageal ones. Excessive radiation on lungs tissue enhances the risk of severe pneumonitis [8]. However, no guideline or atlas is available, and there is a big variation in electron density and the corresponding CT number. Lung segmentation can reveal itself as a difficult task: trachea, bronchus and small size vessels often appear on segmentation results and the respiratory motion affects the proper delineation. Delineating the normal lung used for dose computation was not yet standardized, according to [9], who proposed a guideline for OAR's contouring. The underestimation of this volume due to inappropriate segmentation may have two outcomes: the participation exclusion of a patient from a clinical trial when he(she) would otherwise be eligible to or overestimation of the lung toxicity and, as a result, the unnecessary limitation of the dose prescribed [9].

1.3 Objectives

The main goal of this thesis is to develop algorithms to help the specialists in the delimitation of the lungs for radiotherapy planning purposes, using information from CT scans. The identification of OAR is crucial for minimizing their exposure to radiation. To achieve this goal, there are 3 major objectives to be fulfilled:

- Verify the efficiency of traditional approaches on lung segmentation for Radiotherapy Planning
- Explore the use of a Deep Learning architecture in this segmentation task
- Compare the differences of the results from deep learning models over traditional models in this particular problem

1.4 Research Contributions

Part of the work of this project resulted in the publication of the paper:

Ana Catarina Oliveira, Inês Domingues, Hugo Duarte, João Santos, and Pedro H. Abreu. ”Going back to basics on volumetric segmentation of the lungs in CT: a fully image processing based technique”. IbPRIA 2019: 9th Iberian Conference on Pattern Recognition and Image Analysis (Submitted on April 30th, 2019; accepted on May 24th, 2019; presented on July 4th, 2019).

1.5 Document Structure

This document is organized as follows:

In Chapter 2 some useful knowledge that is going to be the basis for the understanding of this work is explained.

Chapter 3 presents a quick showdown of the state of the art research on lung segmentation in CT. It also refers to some machine learning models that are currently being used with great success in segmentation tasks in other types of medical images, and how they are being implemented.

Then, Chapter 4 describes the data used for the development of this work, as well as the adopted methods in the two segmentation approaches. More precisely, Section 4.2 describes the proposed Region Growing method in all its components: separation of right and left body volumes in Section 4.2.1; three methods for seed placement in Section 4.2.2;

and the new, iterative, region growing technique in Section 4.2.3. Successively, in Section 4.3 the U-Net pipeline is presented.

Chapter 5 discusses the most meaningful achieved results for each method and presents a comparison of both techniques.

The document finishes in Section 6 with some conclusions and directions for future work.

2

Background Knowledge

In this section some useful concepts and techniques are explained for the understanding of the state of art review (Chapter 3) and the methods developed (Chapter 4).

2.1 Medical Image modalities

There are several imaging modalities that try to represent the human body, and each one is defined taking into account the means by which it is captured: light, sound, magnetism, radioactive pharmaceuticals or X-rays. Medical modalities can be divided into two major groups: the anatomical or the functional modalities. The first ones are used to examine anatomical abnormalities, while the latter gives information about the cellular activity and molecular information on the biology of many pathologies [10]. In this project, CT scans will be analyzed, due to its anatomical representation and because it is one of the most used diagnostic imaging devices in radiotherapy planning.

In a CT procedure, an x-ray tube is rotated around the body to produce cross-sectional images, while on the opposite location of the tube a detector array collects the projected data. Then all the x-ray projections are integrated together, in conformity with the principle that the density of the tissue passed through by the x-ray beam can be quantified by the estimation of the attenuation coefficient. Therefore, the grey levels of a CT slice image match the x-ray attenuation, which reflects the proportion of x-rays scattered or absorbed as passing through each voxel. X-ray attenuation is correlated with the composition of the tissue being imaged and the x-ray energy [11].

2.2 Preprocessing techniques

Just like any other imaging process, CT scans contain noise and other image artefacts that can affect the quality of the segmentation and hamper the extraction of features. Therefore, segmentation pipelines can include preprocessing steps to remove unwanted

elements, for visual enhancement of certain features, isolation of individual elements or for joining disparate elements [1].

A good image should have pixels that are bright and dark in the same proportion. In other words, the intensity values of an image should have an uniform distribution, so that the image is not too bright or too dark [12]. Using an image like this, with low contrast, makes the separation of regions based on differences in intensity values more difficult [13]. Histogram equalization, and other methods such as contrast stretching or adaptive histogram stretching, among many others, can help to make these differences more significant, increasing the contrast.

2.2.1 Contrast stretching or Normalization

In this operation, quite similar to the histogram equalization, the image is rescaled to include all the intensities values that fall within the range specified by the user, using a linear transformation. Typically, there is very little weight at the ends of the whole value range of the input image, so the chosen range is usually between the 2nd and 98th percentiles, which will lead to an improvement in the global contrast but may saturate some pixels [14, 15].

2.2.2 Histogram equalization

Another option would be the histogram equalization. Considering a bright image, an histogram of the intensity values will show a distribution where most of the pixels are confined to high values. Histogram equalization will stretch the histogram to either ends, in order to uniform the distribution. This transformation function involves mapping the input pixels from the brighter region to output pixels in full region, spreading as evenly as possible in the new distribution and ensuring that all parts of the value range are equally represented. This operation can be an advantage in images whose histogram is confined to a particular region, improving the global contrast, but it fails in images with large intensity variations, having both bright and dark pixels, which may lead to an over-enhancement of noise or artefacts [12, 14].

2.2.3 Adaptive Histogram equalization

Alternatively, in cases where both bright and dark pixels are present, the adaptive histogram equalization may be used. In this transformation, the image is divided into small blocks (tiles) and subsequently each of these tiles is histogram equalized, improving the local contrast of each block. In a small area, the histogram distribution is confined to a

small region, unless there is noise. In this case, to avoid its amplification, some algorithms apply contrast limiting [12].

2.2.4 Resize

As the name suggests, resize is the process of converting an image from its original size to another size, without changing the number of pixels. This process can be helpful to uniform the size of every picture of a dataset or to zoom in a specific part of an image, for example. The operation is based on an interpolation method, which defines how the values in the final image are computed from pixels that do not correspond directly [16].

2.2.5 Resampling

An image is a sampling of a continuous signal. Resampling is the process of changing the original sampling using geometrical transformations and it can be applied with numerous goals [17, 18]:

- the adaptation of an image to a new set of coordinates (for example, when rotating an image);
- the application of texture to surfaces in computer-generated imagery without explicit modelling the texture;
- in medical and remotely sensed imagery, the registration of an image with a coordinate system such as another image, a map or any other reference;
- the reconstruction of a continuous image from a discrete one and a transformation matrix;
- the registration of images from different sources with one another, e.g. from different CT scans using different spacing;
- the registration of images acquired at different times, to analyze the temporal evolution, such as the evaluation of injuries healing process over weeks or months;
- the geographical mapping, e.g. changing from one projection to another;
- the photomosaicing: creating an image from many smaller images;
- the geometrical normalization for image analysis, e.g. normalization to an isotropic resolution.

This operation can be decomposed in three simpler sub-processes: reconstruction of a continuous intensity surface using a discrete image, transformation of the resultant surface, and sampling of that transformed surface to produce a new discrete image [17]. Simplifying, resampling is the process of estimating the most appropriate intensity value

for each pixel in the final image, based upon one or more pixels in an initial image and a transformation matrix [17]. Using the right interpolating function to perform this transformation between pixels will ultimately determine the image quality of the new image [18].

2.2.6 Morphology operations

In some cases, images may have small imperfections based on the shape, such as noise, artefacts, small holes. In other cases, there are small individual elements one would like to join, or the opposite, isolating. Morphology operations are intended to remove such imperfections by applying a structuring element (a kernel) to an input image.

2.2.6.1 Dilation

Dilation is a process of convolution of an image with a kernel, usually a small solid square. As the kernel is passed through the image, a pixel (i,j) (the centre of the kernel) is set to the maximum over all pixels in the neighbourhood. This leads to a brightening of the image, enlarging bright regions and shrinking dark ones. In general, dilation tends to smooth concavities [16].

2.2.6.2 Erosion

Erosion is the converse operation of dilation. The eroded image is the result of a convolution of a kernel through the image, but in this case, the pixel at (i,j) is set to the minimum over all pixels on the neighbourhood. This leads to a shrinking of bright regions and the enlarging of dark ones, having a tendency to smooth away protrusions [16].

2.2.6.3 Opening & Closing

The opening and closing operations are combinations of erosion and dilation operators. In the first, it erodes first and then dilates. As the name implies, opening operation leads to a bigger separation between objects, e.g. cells in a microscope image, allowing easier counting. Its more prominent effect is to attenuate isolated outliers that are higher than their neighbours. In closing operation, the dilation happens first and then the erosion. This is usually performed in connected-component algorithms aiming at reducing artefacts or noise. Generally in medical image algorithms is performed an erosion or closing operation to eliminate elements from noise and subsequently an opening operation to connect nearby large regions [16].

From all these preprocessing techniques, resampling was performed many times in the taken approaches. Although not implemented in both pipelines, initial studies were performed comparing histogram and adaptive histogram equalization techniques, as well as contrast stretching.

2.3 Segmentation algorithms

There exist many different ways of segmenting objects, each one encompassing a particular type of knowledge (statistics, anatomical, machine learning, etc.). Multiple reports [19, 20, 21] make use of various approaches in simultaneous, which can reveal an advantage in overcoming specific limitations.

Segmentation algorithms can be applied in different ways regarding dimensionality. Usually, 2D methods are applied in 2D images and 3D methods in 3D volumes, but in some cases, 2D methods are applied to slices of 3D volumes, sequentially [22]. This allows to reduce the computational complexity and memory requirements that 3D approaches demand, but can lead to inconsistent final results. There is also studies made on patches of 2D images, using them as 3D volumes.

According to Dzung Pham *et al.* [22], there are eight categories of segmentation algorithms: thresholding, region growing, classifiers, clustering, Markov random field models, neural networks, deformable models and atlas-guided approaches. Of all approaches, a Region Growing algorithm and an Artificial Neural Network were implemented and for that reason, these ones will be reviewed in more detail.

2.3.1 Thresholding approaches

Thresholding is the most basic technique for segmenting an image. It works by taking one or more values as a reference, creating a condition (membership function) and defining a binary image according to it.These values or thresholds can be defined using tools like the histogram of intensity values and cumulative distribution function, where, in some cases, it is possible to notice a division between classes. Therefore, thresholding does not take into account the spatial characteristics of an image, only its intensities. This method can also be used as an initial step in a sequence of image-processing operations.

2.3.2 Deformable models

Deformable models are a very used approach to delineate region boundaries. In a simple definition, deformable models start to expand/adjust an initial contour or model to the image under the influence of internal and external forces [22]. These methods combine

geometry, physics and approximation theory. The geometry is used to represent the initial object shape, the physics restrains the adjustment of the shape over space and time, and optimal approximation theory describes the underlying formal mechanisms on which the model is fitted to the measured data [23].

First, a closed curve or surface is placed near the desired boundary. Subsequently, the initial boundary suffers an iterative relaxation process, through internal forces that are computed from within the curve or surface and allow it to expand smoothly during deformation. These forces are defined according to the elasticity theory at the physical level [23] and comprise elastic energy and bending energy, responsible for shrinking and bending the contour, respectively [24].

As the model expands, the energy grows, and frequently include terms that constrain the smoothness or symmetry of the surface [23]. Using the classical optimal approximation theory, external potential energy functions are derived from the image to drive the model to fit the data, defining its contours along the boundaries [24] and generating external forces [2, 22, 23].

Active Contour Models (ACMs) are a type of deformable models [22, 24], whose primary goal is to define an initial contour around the boundaries of the object as a parametric curve, and distort it toward the desired boundaries by minimizing the functional energy [2]. ACMs can be distributed in two main types: Edge-based Active Contour Models (EACMs) and Region-Based Active Contour Models (RACMs). EACMs creates small variations in segmentation [24] by working with edge information, which is generated using gradient or edge extraction methods [25]. This approach can be, however, sensitive to noise, weak edges and intensity inhomogeneity [25]. On the other hand, RACMs uses statistical information for controlling the contour near the object boundary throughout the evolution. For this reason, RACMs are more robust to the disadvantages of EACMs [24, 25].

This approach can be advantageous due to their ability to create closed parametric curves or surfaces from images and their robustness to noise or spurious edges. However, they require manual interaction to define appropriate parameters and place the initial model [22].

2.3.3 Atlas-guided approaches

When a standard atlas or template of the object to segment is available, atlas-guided approaches can be a powerful tool. The reasoning behind it is based on the creation of a reference atlas, and successive segmentation of new ones using this reference-frame. The atlas can be created by manually segmenting an image or through the compilation of anatomic information from multiple previously segmented images. To segment new images, a standard atlas-guided approach treats segmentation as a registration process, an “atlas warping”. This means discovering a one-to-one transformation that describes the

mapping of the pre-segmented atlas to the new image requiring segmentation, maximizing the similarity. Due to anatomical variability, linear transformations may not be adequate, which leads to the use of sequential application of linear and nonlinear transformations to overcome this issue [22]. If more than one training image/map is used, for each voxel there will be a set of possible labels, which are combined by decision fusion in order to define the label.

This type of approach can face limitations when segmenting complex structures with variable size, shape and properties. Also, it requires expert knowledge to build the database [26]

2.3.4 Markov Random Field models

These models are not segmentation methods per se, but statistical models, being used within segmentation algorithms for that purpose [22]. Considering the set of all the possible labellings for each voxel and for each image, Markov Random Field (MRF) models take a probability measure on this set and select the most likely one. The goal is to maximize the probability of picking the most correct labelling. Images usually have homogeneous regions, that share similar properties (like intensity). These models are able to capture such constraints. By calculating and minimizing the energy of all potential configurations between pixels, it is able to separate regions in spatial clusters according to similar properties [27]. Therefore, structures that consist of a few voxels have a low probability of occurring and consequently be segmented under an MRF assumption [22].

2.3.5 Clustering approaches

Clustering models are machine learning models that are used to agglomerate data without labels based on their similarity, reducing variance. Since they do not require training data, the algorithm iteratively alternates between segmenting the image and adapting the properties of each class [22]. Clustering methods do not directly incorporate spatial modelling, hence being more sensitive to noise and intensity inhomogeneities. Regardless, the lack of spatial modelling can reveal itself more advantageous for fast computation. Although there is no need of labelled data, it requires an initial segmentation or initial parameters, such as the number of classes. Examples: k-means, the fuzzy c-means algorithm, expectation-maximization algorithm.

2.3.6 Region Growing

Region growing (RG) algorithms try to group pixels in larger regions based on connectivity criteria, like the intensity or texture present in the images. The standard procedure in RG algorithm (Algorithm 1) is to take a pixel as seed point and, following a criterion of

connectivity, e.g. the intensity, it cluster pixels by expanding the seed and the segmented region. In this way, different regions start to appear and to be identified as one [1]. The selection of the seed pixel can be performed in many ways, but usually, this is done by manual input. Next, the method adds a neighbour to the segmented region if the difference between the value of the neighbour pixel and the average value of the pixels already in the region does not exceed a threshold. This threshold needs to be manually set, case by case, as it often depends not only on the problem but also on the image in question [28].

Algorithm 1 Traditional Region Growing Algorithm for the segmentation of a 3D CT scan.
Inputs:
Seed vector, $s = [s_x,s_y,s_z]$
Volume to be segmented, CT
Tolerance threshold, Th
Output:
Volumetric mask, $Mask$, with the same size as CT

```
Initialise Mask as a volume with the same size as CT, filled with zeros
Initialise Checked as a volume with the same size as Mask, filled as false
Initialise NeedsCheck as an empty stack

Set Mask at s to one
Set Checked at s to true
Add neighbour coordinates of s to NeedsCheck.

while NeedsCheck is not empty do
    Pop a point p from NeedsCheck
    Set Checked at p to true

    Calculate m_s, the average of CT grey values in the points where Mask = 1
    Retrieve m_p, the grey value of CT in p

    if |m_s − m_p| < Th then
        Set Mask at p to one
        Add neighbour coordinates of p to NeedsCheck
    end if

end while
```

This method can be expensive in terms of computation, yet it is capable of taking advantage of several image properties such as spatial and intensity information directly from the image to detect its different boundaries [1].

2.3.7 Classifiers

Classifiers are pattern recognition algorithms that are used to predict the class of a given pixel (pixel-classification). As pattern recognition methods, they extract rules and pat-

terns from data with labels and seek the better partition of a feature space derived from the image. A feature space is the range of space where all features can be represented. In images, the most common feature space are the intensities themselves. By analyzing training data with known labels, the model adjusts weights, thresholds and other parameters in order to maximize performance [22, 16]. The Support Vector Machine and Artificial Neural Networkss (ANNs) are some of the most known classifiers.

Artificial Neural Networks

From the first steps, the computer was designed to be able to imitate the reasoning from human brain efficiently, performing all different types of functions and to retrieve conclusions, as Turing wrote in *Computing Machinery and Intelligence* in 1950. As the brain is composed of complex networks of linking neurons that communicate with each other through synapses, transferring information from a location to another, interconnecting it and assimilating new knowledge, so does an ANN. In biology, each neuron assigns a priority to some synapses, reinforcing the size and support to the stronger ones. In ANNs, this selection is performed through weights. Plus, a bias can be added, that will shift the activation function away from the origin to grant a better representation of the data. The similarities can be seen in Figure 2.1, that represents a neural network.

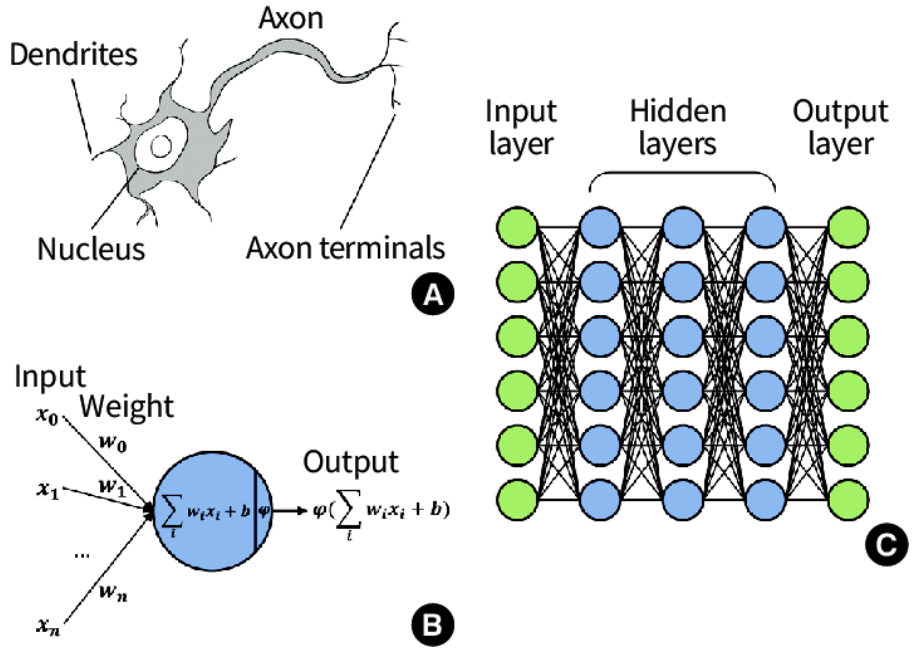


Figure 2.1: Biological Neuron (A), Artificial Neuron (B) and Neural Network (C). Figure from [29].

Each network can be made of one or more layers of interconnected neurons, who simulate biological learning by performing elementary computations when a signal is propagated. An ANN architecture is frequently constituted by an input layer, an output layer and one or more intermediate layers, also known as hidden layers. The first will be the layer that will receive the data, while the output one is going to return the result of the network. The

hidden layers are in charge of processing and adapting the input data through transformations that will lead to a particular outcome that can be interpreted by the output layer.

In traditional ANNs, feature vectors are first given to the input layer of the network. A weigh is assigned to each feature of the signal, and then the weighted sum of all the features is passed through an activation function (plus an optional bias). Then the output is fed to another neuron, if there is an additional layer, or can be the result of the network, if it is in the final layer. This activation function applies a transformation or a limit to the weighted sum, and according to the result, it defines the activation or not of that neuron.

Learning is achieved in each iteration through the adjustment of weights attributed to the connections and the bias that minimize the error between the obtained and the expected output of the training samples. This process is known as backpropagation and relies on the optimizer function, which is responsible for defining the type of update, and on the loss function, that measures the error between the predicted and the true values. These functions are selected based on the type of data and problem context. Later, the network is tested using new samples. Due to the high number of interconnections of ANNs, spatial information can be effortlessly integrated into its classification procedures [11, 22].

Deep Learning

A network can have many architectures that reflect the connections between the neurons in the multiple layers. models are ANNs with three or more layers, each with many neurons. Over the last decade, Deep Learning (DL) has been exponentially increasing popularity and being considered the future of image processing, from landmark detection and tissue segmentation to diagnosis and prognosis [30], and other data processing areas such as Natural Language Processing, Speech Recognition, Social Network filtering, Drug Discovery or Genomics [31].

With the increase of the computational power and available data, due to the increasing modernization and use of imaging devices, this type of models can be now used and achieve state of art performances, once impossible [30]. These algorithms are able of exploiting large amounts of data, identifying hidden and complex patterns and representing it into high-level features based on the low-level ones [30, 32]. In comparison with other methods, they have the ability to discover discriminant and important features without human intervention (feature engineering) and learn the optimal attributes from the available images [33]. Moreover, DL models are more robust to variations in position, rotation, scale, perspective, among others [30]. Such high-level features are, however, hard to explain. ANNs and DL, in particular, are known to be black-box models and their lack of interpretability is a critical concern in the area, with doctors discarding these models on their work for their lack of transparency, explaining the hard acceptance and adopting in medicine of DL methods [30]. They require large amounts of labelled data, increasing its performance as more data is fed to the network. However, when not available, there is the risk of overfitting and hard parametrization, which will lead to bad performances in

new non-trained data. Data augmentation techniques can be a possible solution when this is the case [30].

Convolutional Neural Networks

In image analysis, the most used DL method used is a Convolutional Neural Network (CNN) [30]. In fact, CNNs represent a breakthrough approach in the analysis of medical image, in areas such as organ/tumour detection and classification [34]. After being invented in 1980 [35], the architecture was brought back to life by Krizhevsky *et al.* [36], in 2012, with the AlexNet architecture, when the team won the ImageNet challenge. Their work used a network of 8 layers and millions of parameters, trained with 1 million training images.

CNNs are typically used for classification tasks, where the model determines to which single class the input image belongs. However, in certain tasks such as biomedical image processing, the desired output should include localization, which means that for each pixel should be assigned a classed label [37]. Despite that, thousands of labelled training images are often an unattainable prerequisite/requirement in biomedical tasks.

CNNs also have multiple architectures, such as AlexNet, LeNet, VGG19, GoogLeNet, ResNet, FCN [33]. The name comes from convolutional layers, which detect local features at different positions in the input image and apply a series of convolution filters or kernels. By analyzing small portions of the input image at a time, each of them is able to explore different patterns, encoding more abstract features gradually over layers, from low-level to high-level structures [11]. The result will be the generation of multiple feature maps, one for each layer. The kernel size reflects the size of the receptive field of the image that is being read and extends along the entire depth of the volume.

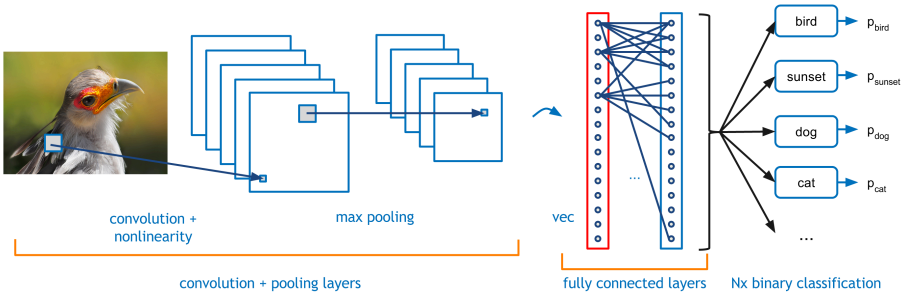


Figure 2.2: A typical Convolutional Neural Network. Figure from [38].

A typical CNN architecture is constituted of three types of layers, convolution, pooling and fully connected layers (Figure 2.2). At the end of each convolutional layer, an activation function, such as sigmoid, tanh or ReLu, is applied to the resulting feature map, introducing the non-linearity property to uncover nonlinear features. Pooling layers are responsible for non-linearly down-sampling the result of the previous layer. The most common function used is the max-pooling: it divides the input into multiple non-overlapping rect-

angles, and then outputs the maximum value of each sub-region. The key reason for this operation is the idea that the precise location of the feature is less important than the coarse location.Indeed, taking advantage of pooling layers will reduce the spatial size of the representation, parameters, the memory footprint and amount of computation required for implementing the network and will help manage overfitting. On the other hand, fully connected layers have neurons connected to all activations from the previous layer, which allows them to combine the resulting feature maps into a feature vector, creating global semantic information. Usually, a softmax function is frequently used in the output layer in classification problems to determine the final output. In general, a CNN is composed of series of convolution filters separated by pooling layers and completed by one or more fully connected layers [11].

The structure of CNNs allows them to learn effective hierarchical feature representations that are commonly present in image data, but they still generate coarse inference, low-resolution outputs and robustness issues [34]. For dealing with the coarse inference and low-resolution problem, Fully Convolutional Networkss (FCNs) arose. FCN are up-sampling networks, in which the fully connected layers are replaced by convolutions and deconvolutional layers are added to perform up-sampling. In other words, upconvolutional models are able of transforming an abstract low-resolution input into a high-resolution output [34]. This idea led Ronneberger *et al.* [37] to create the widely known U-Net architecture, further evolved to a 3D version, 3D U-Net [11]

U-Net Convolutional Network

The U-Net architecture (Figure 2.3) allows the network to intrinsically recombine different scales throughout the entire network.

In short, it combines a context aggregation pathway, responsible for encoding increasingly abstract representations of the input (high-resolution feature maps), with a symmetric localization pathway, obtained throughout multiple upsamplings of the low-resolution maps [37, 32]. Moreover, it recombines both representations by using multi-path connections, ending up with high dimensional features precisely localized on a feature map. The network can be divided into two parts:

- In the contracting path, a typical convolutional network architecture is applied to analyze the whole image. As mentioned earlier, this consists of multiple convolutional layers interspersed by max-pooling layers;
- The expansion part is responsible for providing high-resolution localized accuracy [34], allowing the network to make a pixel-wise classification. For each step in this path, seven operators are applied: 1) an upsampling operator is used instead of a pooling one, followed by 2) convolution (”up-convolution”), that decreases the number of feature channels to half, 3) a concatenation with the equivalent feature map from the contracting path and finally with 4) two convolution layers, each followed by a

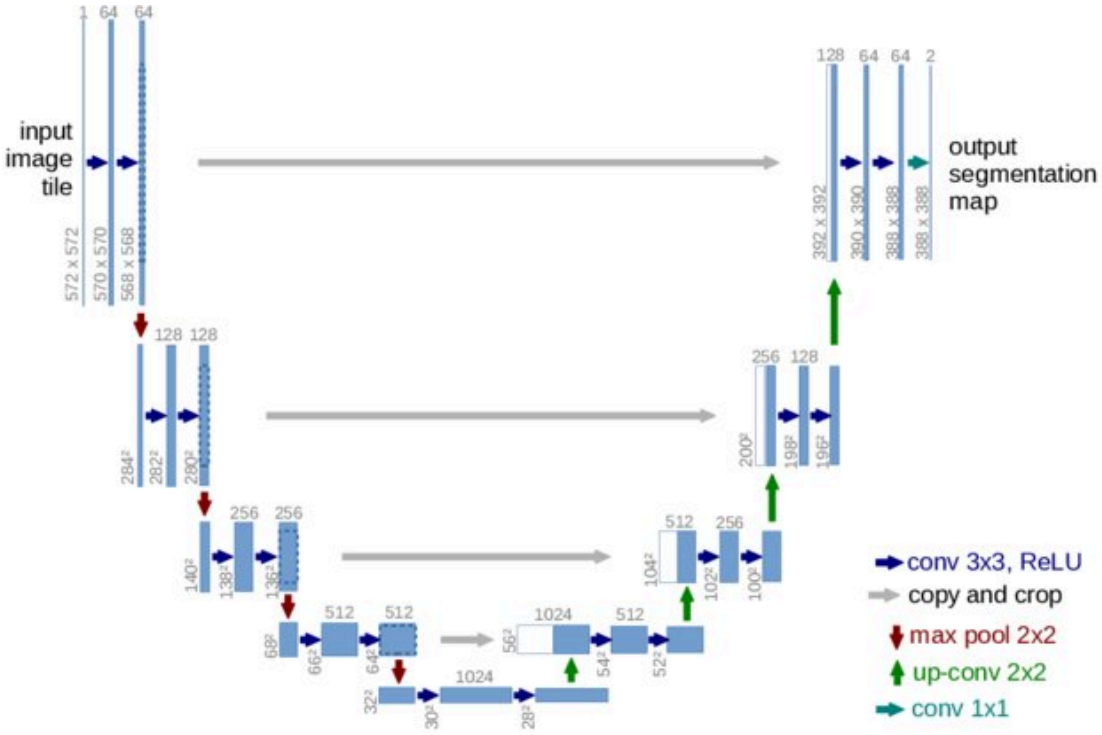


Figure 2.3: U Net architecture, aggregating high-level information from the context pathway (left) with the precise location from the localization pathway (right). Figure from [37]. "Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower-left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations."

ReLU.

It is important to notice the effect of the concatenation step: it allows the network to localize the most relevant features, by allowing the next convolution layer to learn to compile a more precise output based on the information from both the contracting and the expanding path (high-resolution features and upsampled output, respectively) [39]. Such connections are usually complex coarse-to-fine pathways, able to connect activations from earlier stages with later ones [40]. The resultant architecture reveals to be more or less symmetric, conceding a U-shape to the network. No fully connected layer is used in the original structure of Ronneberget *et al.*, but a 1x1 layer convolution layer is placed in the end, to map the final feature vector to the number of possible labels. Due to their ability to achieve good performances in pixel-classification and the efficient use of GPU memory, U-Nets are adequate for image segmentation tasks [41].

2.4 Data augmentation techniques

Due to the reduced number of training datasets in the medical field, and especially in medical imaging, it has become a necessity to create techniques that allowed the use of

the data available to a greater extent, particularly in DL models. To this end, many data transformations can be applied to provide more training examples [34]: scaling; shift; rotation; flipping; colour; cropping; random elastic deformations; drop-out layers [37, 42], amongst others.

As such operations allow the network to learn invariance to these deformations, the chosen methods should take into account the context and type of data being used, since the resultant images may not be biologically plausible. Moreover, deformation is the most common variation in real tissue, and for that reason is important to train the network to predict its effect [37].

2.5 Performance metrics

Many different metrics can be used to evaluate the result of segmentation tasks, that reflect different qualities of the segmented result.

Four metrics were chosen, in the present work, to evaluate the results: Dice coefficient, Jaccard index, True Positive rate, and Volumetric Similarity. While the first three are overlap based, the last one is volume based [43]. These metrics were chosen due to their complementarity. Dice coefficient and Jacquard index are suitable when in the presence of outliers; True Positive Rate for when recall is important; and Volumetric Similarity is appropriate both in scenarios with outliers and when the volume is important [43].

All of these metrics can be derived from the four basic cardinalities of the confusion matrix:

- TP: Voxels correctly considered to belong to the lung;
- FP: Voxels incorrectly considered to belong to the lung;
- TN: Voxels correctly considered not to belong to the lung;
- FN: Voxels incorrectly considered not to belong to the lung.

The Dice coefficient (DICE), also called the Overlap Index or F-score/F_1-measure, is the most frequently used metric. It can be defined as:

$$DICE = \frac{2TP}{2TP + FP + FN} \tag{2.1}$$

The Jaccard index (JAC) or Intersection Over Union (IOU) is defined as the intersection divided by the union:

$$JAC = \frac{TP}{TP + FP + FN} \tag{2.2}$$

True Positive rate (TPr), also called Sensitivity or Recall, measures the portion of positive voxels in the ground truth that are also identified as positive by the segmentation being

evaluated:

$$TPr = \frac{TP}{TP + FN} \tag{2.3}$$

This metric is sensible to segments size, and it penalises errors in small segments more than in large segments. Volumetric Similarity (VS) is a measure that considers the volumes of the segments to indicate similarity:

$$VS = 1 - \frac{|FN - FP|}{2TP + FP + FN} \tag{2.4}$$

Although the following metrics were not used in the present work, they are used to evaluate the results presented on the State of Art (Section 3), and for this reason, a brief description of each metric will be provided:

- Percentage of Volume Overlap (PVO), index showing segmentation accuracy, equivalent to DICE;
- Percentage Absolute Volumetric Difference (AVD) and Percentage of Volume Difference (PVD), metrics showing segmentation dissimilarity;
- Warping Error, boundary metric, designed to tolerate discrepancies over boundary location, penalize topological errors, and can also be applied directly as a cost function for learning boundary detection [44];
- Rand-error, a measure of disagreement between two clusters or segmentations, it evaluates whether the overall cluster of pixels into different segmentations is correct. It is robust to small variations in boundaries, but merging two objects or splitting of an object will increase significantly the error [44];
- Oversegmentation Rate (OR), as the name implies, represents the rate of voxels that are present as truth in the segmented result but not on the ground truth [19] ;
- Undersegmentation Rate (UR), indicates the relative amount of tissue that is considered as true in ground truth but is not present in the segmentation result [19];
- Absolute Border Distance (ABD), indicates the spatial similarity between segmentation boundaries;
- Specificity (Sp), also called True Negative Rate, indicates the portion of negative voxels that are correctly identified as such;
- Matthews Correlation Coefficient, balanced measure to analyze the correlation between two segmentations;
- Spearman Correlation Coefficient, a non-parametric metric to assess a possible linear association between two continuous variables;
- Accuracy, the proportion of voxels that are correctly identified among the total number of cases examined;

- Hausdorff Distance (HD), the largest minimum distance between two sets of points, that gives a better evaluation of the local behaviour [45, 2];
- Mean Absolute Distance (MAD), the average of the absolute distance between two contours, most suited for the evaluation of the global correspondence between two contours [45, 2];
- Convergence time, time required for a model to get closer to a certain metric value, over iterations.

2.6 Statistical tests

Statistics allows one to critically assess the quality of the results of a method, considering external factors to the model, such as different datasets or randomness. In this project, the following statistical tests were performed:

- Shapiro-Wilk test, evaluates the normality of the distribution;
- Friedman's ANOVA, non-parametric test to evaluate differences in multiple distributions with dependent samples;
- Wilcoxon Signed-Rank test, non-parametric test to evaluate if two distributions with dependent samples have statistical differences.

Since pair-wised tests were also performed to evaluate the performances in more than two models, the Bonferroni correction was applied to the Wilcoxon Signed-Rank test.

3

State of the Art

A search was performed for works from different approaches on lung segmentation in CT volumes published in the past 15 years that used approaches such as the ones described in the previous chapter. Additionally, some segmentation works on other areas of study were reviewed due to the relevance of their techniques and the results on the problem they tried to resolve.

In 2006, Sun *et al.* [21] developed a tool to reconstruct volume throughout CT slices, for segmentation and visualization of the lung volume. First, a preprocessing step was performed, in which CT slices were filtered by an anisotropic filter algorithm, which allowed to enhance the signal-to-noise ratio and eliminate noise. To allow the visualization 3D of the volume data thought entire CT slices, a interpolation method combined with volume rendering was used. After the preprocessing phase, a lung mask was generated with base on 3D RG algorithm, using three growing criteria, mainly based on region homogeneity and region aggregation by using either data values or gradient magnitudes. The threshold for the execution of the criteria was selected based on intensity and gradient magnitudes as well. To start the segmentation, they developed two algorithms for seed initialization, one manual and one automatic. For the manual seed, the user had two options: the interface allowed the user to extract any slice from the volume and select the seed interactively, or, alternatively, the user could draw a circle or rectangle to specify a region, inside of which all pixels would be marked as seeds. For the automatic seed method, a clustering algorithm was used: a fuzzy C-means, that divided the image into three labels: lung, background and other tissues. The seeds would be later validated regarding a Hounsfield Units (HU) interval, automatically obtained from valley points of the volumetric histogram. After the verification, the algorithms checked the 26 closest voxels (8 for the seed slice and 9 for the adjacent slices). If it was valid accordingly to the three criteria defined, the voxel was assigned to the segmentation result. Finally, a 3D morphological closing operation was performed on the lung mask, resulting in a complete cavity-free lung volume. To view the 3D segmented lung volume at any angle, a 3D visualization tool was also designed, with Visualization Toolkit (VTK). Using a dataset of 20 (single-detector) CT images, the results obtained were 88.5% of Volume overlap (PVO) and 11.5% of Volumetric Difference (PVD), when compared with the golden standard (ground truth).

Then in 2014, aiming to adjust a statistical shape model on the lung, especially on a lung with several pathologies, Birkbeck *et al.* [46] suggested the use of landmarks or contextual constraints from neighbouring anatomies, such as the curvature of nearby ribs or the distance to spine. The team proposed a learning-based algorithm using statistical classifiers. The algorithm detected adjacent organs and rib segments, using them as geometric constraints and appearance cues for the lung mesh segmentation. The pipeline had 5 steps: organ detection, rib points extraction, pathological lung surface refinement, identification of pathological lungs and final refinement. On the first step, the method detected surfaces and corresponded them with a template surface (a statistical shape model), previously trained with data. As explained before, some pathologies change dramatically the appearance of the lungs. To overcome the issue, the algorithm estimated the shape on a contrast adjusted input image. Later on, to extract the rib points, discriminative classifiers were applied, by creating a per-voxel rib probability map and obtaining the connected components of the thresholded mask. Next, an area filter was applied, and components out of the range [2000,9000] mm^3 were excluded. The segments that were not excluded would be like bent cylinders, but if they were not, a final test on the sizes of the segment along its principal directions was applied. A combination of image appearance cues with spatially anatomical constraints discovered from ribs and organs was used for the detection of pathological boundaries. The mesh initialized in the first step (and in correspondence with a known atlas) was refined by projecting it onto the shape subspace and then performing deformations on its vertexes. The proposed method used a discriminative model as a classifier so that the appearance term could be learned from training data. On the 4^{th} step, for the determination of pathological lungs, the algorithm checked if the anatomical limits between the lung/rib surface and lung/bottom organ surface satisfied the following conditions: for the lung/rib surface, the lungs were identified as pathological if the lung regions adjacent to the rib surface were too far from the ribs; in the case of lung/bottom organ surface constraint, if the surface was not close enough from the adjacent organ, then it would be considered pathological. Finally, on a 5^{th} step the mesh previously defined was refined so that it could be used to obtain voxel-level fine-scale detail. The pipeline was trained on a total of 381 lungs dataset (185 right and 196 left), the majority of them healthy (only 9 of them had between 25%-50% of unhealthy tissue). For testing, 50 lungs with pathologies were used. The results showed that the mean and median errors decreased after using the pathological pipeline, while the DICE increased. For the basic pipeline, the results showed an average DICE of 84.5%, while the lung pathological pipeline achieved a 90% and the full pipeline achieved an average of 92.5%.

Another well-known approach for lung segmentation is the thresholding technique. Due to the high percentage of air and low density, the lungs appear as dark regions in CT scans. However, in case of pathology, a part of the lung may be more dense, due to masses or nodules, and contrast with the lung background [11]. Based on the correlation

between radiographic density and the atomic number of a given material, it is possible to associate different ranges of radio densities for each tissue and pathology. For example, hyperinflated lungs ([−1000,−900] HU), normally aerated ([−900,−500] HU), poorly aerated ([−500,−100] HU) and nonaerated ([−100,50] HU) [47].

In light of this, Zhou *et al.* [48] proposed an automated method consisting of a combination of an enhanced grey-level thresholding algorithm with a refinement scheme, the latter based on an ACM. The system was trained on a dataset of 25 patients of the H. Lee Moffitt Cancer Center and Research Institute in Tampa, USA, with a total of 2342 slices. With the purpose of comparison, a traditional thresholding tool, Definiens (a commercial software), was also applied to the dataset, followed by morphology operations and contour analysis. The threshold interval used for the mask was [-1000,-400] HU. Instead of searching for the optimal threshold that would better fit the segmentation, the team applied a histogram-based image stretch processing, distributing uniformly the value intensities of the image slice.For this purpose, as prior-mentioned (Chapter 2), the range limits for the stretching were determined by the maximum and minimum intensity value of the bands visible on the histogram. The results showed enhanced images contrasts, leading to better results of the threshold method. In their approach, there was no need to initialize the threshold, as the Otsu algorithm automatically determined the best one, by separating the two classes visible on the histogram and minimizing its intra-class variation. In the end, although the threshold approach allowed the separation of lung area from non-lung areas, to obtain the lungs was necessary to apply a contour analysis. This analysis searched for the exterior contour, separating the body from the background, by looking for the region with maximal exterior contour. Afterwards, by checking the pixels inside this contour (the body), it searched for regions with intensities lower than the optimal threshold and with an external contour big enough (when compared, e.g. with the trachea size). At the end of this process, the lung region was segmented. Thresholding based segmentation still presents difficulties on boundaries and the areas with vessels, causing holes and inaccurate borders of segmentation. However, hilar vessels and other structures may be necessary for the detection of diseases. These areas are usually fine structures, that can be viewed as textures, which were detected by a texture-aware ACM, incorporating texture information, intensity properties and structural features of lungs. The textural descriptor was used to smooth the intensity variations in the texture region. Next, both the textures detected and the results from the segmentation were incorporated in a convex active contour. To assess the results and compare them with the ground truth, DICE was used to evaluate the overlapping results. The DICE of the 10 CT scans evaluated was 84.6% for the reference method and 90.59% for the proposed method.

Having in count the advantages and disadvantages of multiple simple techniques above described in Section 2, Shi *et al.* [19] proposed a pipeline that is the combination of simple (in part) effective approaches. The team proposed a three-phase approach, based on image processing (filtering and threshold approach), lung region initial segmentation (tho-

rax extraction with RG and random walk for lung extraction) and lung region refinement (rolling-ball algorithm for hole filling and lung contour correction with iterative weighted averaging and adaptive curvature threshold). The pipeline was tested on a dataset of 23 lung CT scans with 883 2D slices. The preprocessing phase aimed to reduce the noise, reduce (smoothing) the intra-region variance while preserving inter-region edges of the images. Using an improved anisotropic diffusion filtering method, such as a guided filter or bilateral filters, allowed to overcome the risk of blurring edges and losing details, that often occur in Gaussian filter, mean filter or median filter. The filtering process applied an edge detection step, to encourage intra-region smoothing while preserving inter-region edges. A comparison with other filtering methods such as the Gauss filter was made, achieving a peak signal-to-noise ratio of 63.1342, (contrasting with the 61.4568 of the Gauss filter). Next, a binarization of the image was made using an Otsu's adaptive threshold method. The segmentation phase started with the segmentation of the thorax versus background, that could include external artefacts to the patient bodies, such as the treatment couch. To overcome this issue, a RG algorithm was used, and the background image was subtracted from the binarized image. For lungs extraction and delineation, a random walk algorithm was used, due to the distribution of density of thorax tissues. At the end of the second phase, the extracted lungs still had a severe amount of holes and parts of regions were also excluded. A rolling-ball algorithm was used to fill the holes. Lastly, to redefine the contours of the segmented lungs, a curvature-based correction method was used on the preliminary contours. The results were evaluated according to the Oversegmentation Rate, the Undersegmentation Rate and the Absolute Border Distance. Comparing the results with RG and ACM alone, the proposed method achieved a OR of 1.87%, UR of 2.36% and 0.62% of ABD, contrasting with 2.1%, 2.71% and 0.72 % from the RG algorithm and 1.9%, 2.38% and 0.64% from the ACM. Evaluating the OR, results showed that the proposed method achieved 98.4%, whereas that obtained using the threshold-based method was 94.1%, the RG OR was 95.3%, the ACM was 94.4%, the ACM with a curvature-based correction method was 95.8%, and the random-walk-based method was only 93.8%.

Also using simple approaches, in 2017, Nobrega *et al.* [47] aimed to develop a 3D lung RG segmentation method based on several tools available with ITK and VTK modules, using a dataset composed of 30 full Chest CT exams. First, a Gaussian filter was used to attenuate noise and small discontinuities on the first slice of the CT scan. Then the authors assure that the seed used for the RG algorithm was inside/connected to the lungs by comparing with the HU values theoretically associated. For this purpose, the image was filtered with a Binary Threshold Image Filter, using a [-1000,-500] HU range. In this way, only the structures with these values would be considered. After this step, a 2D RG algorithm was applied, using the local minimums as seeds. In a fourth phase, an area filter was performed, discarding structures with less than 80 or more than 700 pixels. If at the end of this step there were no more structures left, the filtering restarted and the

interval range was adjusted. Finally, the only remaining structures could be or the lungs or the trachea. However, the latter and other respiratory airways (such as the bronchi) are usually analyzed separately from the lung parenchyma [46]. To exclude this organ, two RG algorithms were performed. Firstly, by finding the closest remaining structure to the centre of the image, the team was able to detect the trachea. Using the trachea centroid as a seed, a 3D RG algorithm segmentation was applied. The process stopped after not finding any more voxels inside the [-1000,-500] HU interval. After this step, a second 3D region algorithm was performed, but with an additional condition: when the RG rate exceeded twice of its mean value, the algorithm stopped the segmentation. This limit was empirically defined after realizing that the rate maintained constant in the trachea segmentation, but suddenly increased when it reached the lungs. Finally, the trachea was removed by subtracting the trachea volume (second segmentation result) from the lungs+trachea volume (first segmentation result). The proposed algorithm detected all the 30 trachea's successfully, and results were evaluated using the TPr, Sp, F-Score (F) and Matthews Correlation Coefficient (MCC). Compared with two OsiriX software algorithms, AOSP (Automatic Segmentation with Osirix Software plugin) and SOST (Semi-Automatic segmentation with Osirix toolbox), the proposed 3D RG method achieved average performance rates of TPr = 98.09%, Sp = 99.87%, F = 98.62% and MCC = 98.44%; while the AOSP method presented: TPr = 98.97%, Sp = 99.17%, F = 95.98% and MCC = 95.65% and the SOST method: TPr = 98.66%, Sp = 99.30%, F = 97.17% and MCC = 96.82%.

It is also possible to combine machine learning techniques with ACM to segment the lungs efficiently. Filho *et al.* [49] proposed a 3D Adaptive Crisp Active Contour Model (3D ACACM) for the segmentation of 3D CT lung volumes, using ITK and VTK functions. Unlike other parametric ACMs, the proposed method was able to segment complex objects in 3D, by moving the points of the model with base on the information from the voxels and model shape. By calculating the internal energy using 3D model information, like distances, they were able to prevent the points from moving uncoordinatedly in relation to the neighbouring voxels, which lead to an improvement in stability. For the external energy, the authors used the 3D Adaptive Crisp method, which detected the origin of the edges obtained with the 3D traditional external energy. When the volume did not increased after two successive iterations, the algorithm stopped and the model was considered stable. The dataset used was composed of 40 chest CT scans of healthy volunteers and patients with various pathologies. The efficiency of the model was evaluated by comparing the presented method with a 2D method previously proposed by their team [50], and was tested using F-score. The obtained results with the F-measure were 99.14% ±0.18 (with a processing time of 3.20 ±0.38 minutes), while the 2D model achieved 96.33% ±0.42 in 12.52 ±2.10 minutes. The algorithm presented an increase in the stability and an acceleration on the convergence of the 3D model. The authors also compared the model with three other methods: automatic 3D RG method (3D RG), semi-automatic segmentation using 3D OsiriX toolbox (EUOT) and Level-Set Algorithm based on the Coherent Propa-

gation Method, of MIA plugin for OsiriX (LSCPM). The developed algorithm obtained an f-measure of 99.22%, and the other methods achieved, respectively, 98.57%, 98.53% and 98.73%.

More recently, due to the success of machine learning algorithms, and particularly in deep learning-based approaches such as CNNs, an increasing number of works in this area have been proposed.

Using a U-Net architecture, Brahim Skourt *et al.* [51] proposed an end-to-end lung CT segmentation in pathological lungs. An effective lung segmentation method can be quite challenging, especially if the lung parenchyma presents an irregular structure or appearance. Is these cases, the nodules and blood vessels need to be segmented with the lung parenchyma, while the bronchus regions need to be separated from the lung tissue [51]. The dataset used consisted of diagnostic and lung cancer screening thoracic CT scans with marked-up annotated lesions as ground truth. As a preprocessing step, the images were cropped, removing any information that didn't belong to the area of study. Afterwards, the images were used to train a U-Net network. The group achieved results of 0.9502 of DICE index. The segmentations obtained do not contain parts of trachea or bronchus regions while maintaining the lesions and blood vessels, showing the efficiency of the method.

With the purpose of assessing the generalization of a segmentation method for different OAR, Astaraki *et al.* [2] exploited the reliability of segmenting OARs using three 2D local region-based semi-automatic segmentation algorithms (Localizing Region-Based Active Contour (LRBAC), Local Chan–Vese (LCV) and local Gaussian distribution fitting (Local Gaussian Distribution Fitting (LGDF))). Using a dataset of 20 CT scans of patients diagnosed with different types of cancer, including prostate, esophagus, lung and pancreatic cancer, in a total of 1340 slices, the team focused on the segmentation of eight OAR throughout the whole body (the bladder, rectum, kidney, clavicle, humeral head, femoral head, spinal cord, and lungs). Manual segmentation of the OAR volumes was also performed to be used as a reference, except for the lung delineation, for which the oncologists employed automatic segmentation tools. Firstly, the user selected three to four points around the desired organ in each slice. Later on, the initial contour was automatically interpolated and fitted between these points, reaching the boundary by minimizing the energy functional (LRBAC, LCV or LGDF). With the purpose of evaluating the applicability and accuracy of these functionals, all three models were tested against the eight OARs. An important feature of these models was their ability to segment cancer-related images with non-uniform intensity distributions, which usually restricts the algorithms usage on organs whose pixel intensities are distinguishable from adjacent organs. To evaluate the results, the DICE, the Hausdorff distance (HD), the mean absolute distance (MAD), the percentage absolute volumetric difference (AVD) and the PVD were chosen as metric measurements. Due to poor results, the femur and rectum were not considered further.

The results showed that the highest inter-observer DICE similarity criteria were obtained for the lung (0.94 ± 0.04) and bladder (0.93 ± 0.03), while the lowest agreements were observed for the spinal cord (0.81 ± 0.04) and clavicle (0.87 ± 0.02). For each OAR, the most accurate results were: the mean ± standard deviation of the DICE 0.9 ± 0.04 for the bladder (LGDF), 0.92 ± 0.12 for the humeral head (LRBAC), 0.81 ± 0.04 for the spinal cord (LCV), 0.84 ± 0.04 for the clavicle (LCV), 0.90 ± 0.04 for the kidney (LGDF), and 0.94 ± 0.02 for the lung (LGDF). According to the results, LGDF could achieve more precise segmentations, in OARs where high-intensity gradient and intense inhomogeneity were present. When comparing with other active contours techniques, particularly edge-based models, these three functionals tended to be less sensitive to initialization, being however sensitive to initial contours. They appeared to achieve good deformability and to have significant ability to delineate OARs when adequate intensity information exists.

It is common for a segmentation algorithm to oversegment when the boundaries of an organ are blurry. In 2018, Nithila *et al.* [24] developed a technique named Selective Binary and Gaussian filtering-new Signed Pressure Force (SBGF-new SPF) to overcome this issue in CT lung images. This technique was able to detect the external boundary of the lung and effectively stop contours, even at blurry boundaries, therefore preserving the boundary of the lung lobe. As explained in Chapter 2, there are two types of ACM: edge-based and region-based. The latter uses statistical information to control the contour inside and outside of the object boundary throughout the evolution, requiring two forces, the internal and external one. The authors of this study exploited the usage of different contouring models, such as Distance Regularized Level Set Evolution (DRLSE), Local Binary Fitting (LBF), Local Gaussian Distribution Fitting (LGDF), Local Image Fitting (LIF) and the proposed Selective Binary and Gaussian Filtering with new Signed Pressure Force (SBGF-new SPF). This method automatically placed the initial contour on the CT. Initially, it penalized level set function to be binary, and regularized it, making the evolutionary stable. However, it did not change the functionality of evolution, just reduced the number of iterations. The author compared the proposed model with four other ACM (DRLSE, LBF, LGDF and LIF). The results showed that the model converged at 150 iterations, taking 17 s, while DRLSE converged after 1018 iterations and 490s, the LBF at 100 iterations and 454s, LGDF at 1000 iterations and 811s, and finally LIF taked 150 iterations and 762s. The SBGF-new SPF method obtained an accuracy of 98.95% when compared with the ground truth.

The training phase of most of the supervised machine learning algorithms requires a large dataset of labelled/annotated data. This data is usually produced manually, and therefore demands a lot of time from the user to delineate it, and increases the variance and the risk of errors. This year, Mingjie Xu *et al.* [20] developed a workflow to reduce the workload of manually preparing the dataset for training a classification algorithm. The pipeline started with a dual unsupervised k-means clustering, whose results suffered an intersection operation, a connected component analysis and a patch expansion, that was

subsequently used to train an AlexNet network, applying cross-validation. Firstly, CT slices were split into image patches and next, a k-means clustering algorithm was performed twice in order to segment the lungs, using the mean and minimum intensity of the image patch, respectively. Hence the intersection of the volumes was obtained and submitted to a connected component analysis algorithm to extract the lung parenchyma. In this way, Xu proposed a faster, automatic, semi-supervised method to generate a labelled dataset of lung parenchyma and non-lung parenchyma. Then, a simplified AlexNet structure of a CNN model was trained with the resulting patches, together with non-lung parenchyma patches randomly selected. The results were evaluated using the average value of F-score (F_avg) and the computational time of training, achieving a 0.9917 F_avg on 1609.34 s. A dataset of 121.728 image patches was divided into training and validation, and cross-validation was performed (with a ratio of 7:1, so that the 8-fold is carried out). The results showed a training accuracy and loss of 99.08% and 0.0294, respectively. Also, they tested the performance using a separate dataset of 201 patients with different diseases (COPD or lung cancer), acquired by PET/CT scanner or CT scanner only. The results achieved an DICE of 0.968, HD 1.40, TPr 0.909 and Sp 0.999.

3.1 Other segmented areas

Occasionally, studies made on other segmentation areas may prove to be of interest due to the novelty of the techniques and the relevance of their results in the context they are inserted to. Such studies can inspire and be applied in other areas of study, leading to a better understanding and to the evolution of the state of the art. One of the major difficulties of DL algorithms regarding medical data is the number of training samples required [37, 32, 30]. However, large labelled databases may prove sparse in medical research, making it harder for the network to learn enough and perform well, leading to overfitting and hard parametrization [37, 32, 30]. It may, therefore, be advantageous to make use of current solutions as data augmentation techniques, drop-out layers, transfer learning and fine-tuning.

In 2015, on the University of Freiburg, in Germany, Ronneberger *et al.* [37] proposed an end-to-end network for the segmentation of neuronal structures in electron microscopic stacks. The network had a convolutional network architecture consisting of a contracting path, responsible to capture the context and increasingly abstract representations of the input, followed by a symmetric expanding path enabling the capture of the precise location. Concatenating several features (high-resolution feature maps) from the contracting path to the expanding path makes possible to obtain a pixel-classification. Data augmentation techniques were also performed. In the case of their dataset, training shift and rotation invariance, along with robustness to deformations and grey value variations were fundamental. For this purpose, random elastic deformations seemed to be the key proce-

dure to train such a network with such a small dataset. Despite the reduced number, the dataset had images of touching objects of the same class, hindering the separation between segments. To overcome this issue, they proposed to adopt a weighted loss, giving more weight in the loss function to the separating background labels between touching cells. This lead to a bigger penalization when the network was not able to correctly separate the segments. To test the experiment, three different segmentation tasks were performed. First, the segmentation of neuronal structures in electron microscopic recordings, on a dataset composed of 30 images (512x512), achieving a warping error of 0.0003529 and a rand-error of 0.0382. Second, a cell segmentation task in light microscopic images on two different datasets. The first set was composed of 35 partially annotated training images, for which was achieved an average IOU ("intersection over union") of 92% (being significantly better than the second-best algorithm). Lastly, they tested on a second dataset composed of 20 partially annotated training images of HeLa cells recorded by differential interference contrast microscopy. In this dataset, they achieved an average IOU of 77.5%, also being significantly better than the second-best solution.

Brain Tumour Segmentation (BraTS) challenge aims at the evolution of state-of-art methods for the segmentation of brain tumours in multimodal Magnetic Ressonance Imaging (MRI) scans.The 2017 edition's dataset was composed of annotated Low-Grade Gliomas (LGG) and High-Grade Glioblastomas (HGG), in a total of 210 HGG and 75 LGG cases. All the data was manually annotated, each one containing a T1 weighted, a post-contrast T1-weighted, a T2-weighted and a FLAIR MRI. The classification task had 3 labels, correspondent to each part of the tumour: edema (label 2), necrosis and non-enhancing tumour (label 1) and active/enhancing tumour (label 4). Besides the segmentation challenge, the participants were also asked to develop an algorithm for survival prediction, for which an additional dataset of 163 training cases was provided. Just like Ronneberger [37] in 2015, Isensee *et al.* [32] proposed in 2017 a U-Net architecture. However, the team adopted different design choices, such as the exact architecture of the context pathway, the normalization schemes, number of feature maps throughout the network, nonlinearity and the structure of the upsampling pathway. For the survival prediction challenge, several features were computed from the segmentation mask of the tumour subregions, such as intensity, shape and texture features, and added to other handcrafted measures such as the distance of the tumour to the ventricles and critical structures in the brain. The authors based their resolution on a simple radiomics based approach, combined with a random forest regressor and a multilayer perceptron ensemble for survival prediction. In order to reduce the patient and modalities variability of the data, it was needed a normalization, ensuring that all the value ranges match and avoiding initial biases of the network. For this purpose, they normalized each modality of each patient independently. The process consisted on the following steps: subtraction of the mean and division by the standard deviation of the brain region; clipping of the resulting images at [-5,5] to remove outliers; rescaling to [0,1], with the non-brain region being set to 0. To deal with class imbalances,

a multiclass adaptation of the DICE loss function was used. Just like Ronneberger, to avoid overfitting, Isensee used data augmentation techniques such as random rotations, random scaling, random elastic deformations, gamma correction augmentation and mirroring. The network has shown to be able to accurately segment large tumour regions as well as fine-grained details, obtaining results that had beaten the current state of the art on BraTS 2015 and was one of the leading methods on the BraTS 2017 validation set. The results for the validation set were DICE scores of 0.896, 0.797 and 0.732 for the whole tumour, tumour core and enhancing tumour, respectively. For the test set, they achieved DICE scores of 0.858 for whole, 0.775 for core and 0.647 for enhancing tumour. However, the survival prediction proposed method obtained an accuracy of 52.6% and a Spearman correlation coefficient of 0.496.

3.2 Conclusions

Table 3.1 resumes the main characteristics of the analyzed works on segmentation approaches.

Several works [20, 32, 37, 46, 51] use classifiers in their pipelines. This has, however, the disadvantage of the need of a training phase.

In the papers [2, 48], segmentation is performed in 2D and 3D connectivity is performed afterwards. This may lead to “jagged” and inconsistent final results.

Other proposals like [32, 37] have shown success in other segmentation objects, even in small datasets, representing a possible approach.

Perhaps the most similar works to the 3D Iterative Region Growing method here described are the ones presented in [21, 47], where 3D RG models were also used. Their works, however, use “of the shelf tools” belonging to ITK.

As can be seen, several different algorithms have already been proposed. They are characterized by their complexity, high running time or segmentation in 2D. This complexity is, in some cases, justified by the application on pathological lungs.

A new, tridimensional, simple method, based solely on image processing, 3D Iterative Region Growing Algorithm, is described in Section 4.2. Not recurring to classification algorithms makes our proposal simpler, faster and does not require a large training set, requiring only some basic anatomic knowledge. ´ On the other hand, besides this method, a CNN 3D U-Net segmentation algorithm is also analyzed on Section 4.3, using the work of [32] as a basis. As a classifier, it is robust, designed to process large 3D input blocks, does not need a large dataset to train and is able to process simultaneously high dimensional feature representations with high spatial resolution, achieving good results in

pixel-classification tasks. A comparison of both algorithms will be described in the next Chapters.

Table 3.1: Overview of the works presented.

Author	Dataset	Preprocess	Methods	Post-Process	Results
Zhou *et al.* [48]	25 CT scans, 2 342 slices	Histogram equalization	Otsu thresholding, Contour analysis	Texture-aware active contour model	DICE = 90.59%
Nobrega *et al.* [47]	30 CT scans	Gaussian Filter, Threshold	2D RG + area filter	Subtraction of 3D Region Growing for trachea	TPr = 98.09 Sp =99.87 F-score = 95.98 Matthrews Correlation Coefficient = 95.65
Sun *et al.* [21]	20 CT scans	Anisotropic filter	3D RG with 3 criteria and 2 methods for seed initialization	3D closing	PVO = 88.5% PVD = 11.5%
Mingjie Xu *et al.* [20]	121 728 patches + 201 CT scans of pathological lungs		AlexNet CNN		F-score = 99.17 Time (training) = 1609.34 s DICE = 96.8 HD = 1.40 mm TPr = 90.9 Sp = 99.9
Brahim Skourt *et al.* [51]	1018 CT scans of pathological lungs (LIDC-IDRI)	Cropping	U-Net CNN		DICE = 95.02%
Astaraki *et al.* [2]	20 CT scans of pathological lungs (1340 slices)		3 semi-automatic active contour models, 2D		LGDF: DICE = 94 LCV: DICE = 83% LRBAC: DICE = 76%
Nithila *et al.* [24]	1018 CT scans of pathological lungs (LIDC-IDRI)		Active contour models		Convergency bet. 100-1018 iter. and 17s-811s Accuracy = 98.95%
Birkbeck *et al.* [46]	CT scans with 218 right lung and 213 left lung annotated g.t.	Organ detection, rib points extraction, contrast adjustment	Statistical Shape model using statistical classifier + Area filter	Area filter + Lung pathology detection + Fine-scale refinement	Basic pipeline: DICE = 84.5% Lung pathological pipeline: DICE = 92.5%
Filho *et al.* [50]	40 CT scans, healthy and pathological		3D active contour model (2D and 3D)		F-Score = 96.33%-99.14%
Shi *et al.* [19]	23 CT scans, 883 slices, healthy and pathological	Anisotropic Diffusion Filter (guided filter) + Otsu thresholding	Region growing + Random Walk	Rolling-ball + curvature-based correction	OR= 1.87% UR = 2.36% PVO = 98.4% ABD= 0.62%
Ronneberger *et al.* [37]	30 EMR 55 part. annotated training images		U-Net CNN		warping error = 0.0003529 rand-error = 0.0382 IOU = 92% / 77.5%
Isensee *et al.* [32]	285 MRI (210 glioblastomas and 75 gliomas)	Normalization, rescaling and cropping	U-Net CNN		Multiclass DICE = 85.8%, 77.5% and 64.7% for whole tumour, tumour core and enhancing tumour

4

Material and Methods

The main goal is to help specialists in the task of radiotherapy planning, by providing algorithms that help in the delimitation of the lungs. Such delimitation is crucial to minimize their exposure to radiation. With this goal in mind, two algorithms were evaluated, an Iterative Region Growing (IRG) algorithm and an U-Net Neural Network.

Therefore, this section will be divided in two approaches, with a brief description of the dataset used to develop and train both algorithms.

4.1 Dataset Description

The private dataset of 145 patients with Hodgkin Lymphoma used for this project was provided by the Institute of Oncology of Porto. For each patient, pre-treatment PET and CT scans, post-treatment PET scans, radiotherapy planning CT (acquired after the front-line chemotherapy treatment) and contours manually delimited by experts were available. However, from these 145 cases, 2 did not include CT and 3 were corrupted. The resulting number of manually annotated Regions of Interests (ROIs) are given in Table 4.1.

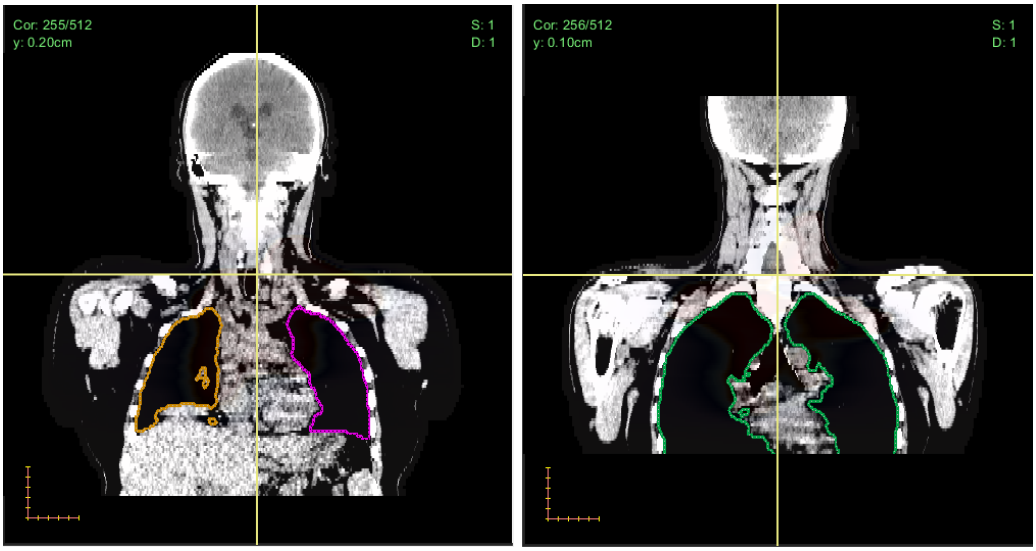
Table 4.1: Number of manually annotated regions of interest of the dataset.

Name	Quantity
Patients	145
CTV	198
PTV	248
GTV	42
Lungs	93
Esophagus	121
Heart	97
Liver	20

The CT volumes were acquired with a patient position between adjacent slices of 2.0, 2.5 or 5.0 mm and a pixel spacing of 1.0, 1.1, 1.2 or 1.3 mm, all in the DICOM format[1]

[1] For further details on the dataset, see [11]. Other projects done with the same dataset can be consulted

From the 93 patients with lungs contours, 63 had a separate ground truth for each of the lungs and 7 for only one of the lungs. These 70 patients will be called group A for easy understanding of the following steps. The remaining 23 patients had a ground truth involving both lungs. Some patients had both ground truths identified simultaneously. Group B will comprise patients with only “both lungs” ground truth.



(a) **(b)**

Figure 4.1: Example of variety of lung ground truths: (a) individual ground truths and (b) “both lungs” ground truth

To compare and evaluate the results, the manual contours of the lung’s ROIs were extracted and used to create binary masks, using the software computation environment for radiotherapy research (CERR) [55] for MATLAB.

4.2 Region Growing

The RG method can be divided in three major steps, as represented in Figure 4.2: in “Laterality separation”, two copies of the original volume are made, each of them containing only one of the lungs; in “Seed definition”, three automatic methods for the placement of the seed are proposed, needed to initialize the segmentation method; in “Segmentation”, the lung volumes are defined, with a new and iterative region-growing-based method.

Figure 4.2: Lung Segmentation pipeline - Region Growing technique.

on [52, 53, 54].

By working with 3D volumes, it is intended to take advantage of volumetric features, typically not used in other works, that allow physicians to see and understand more easily the shapes and connections of the tissues represented in the volume and can, therefore, be relevant to the segmentation itself.

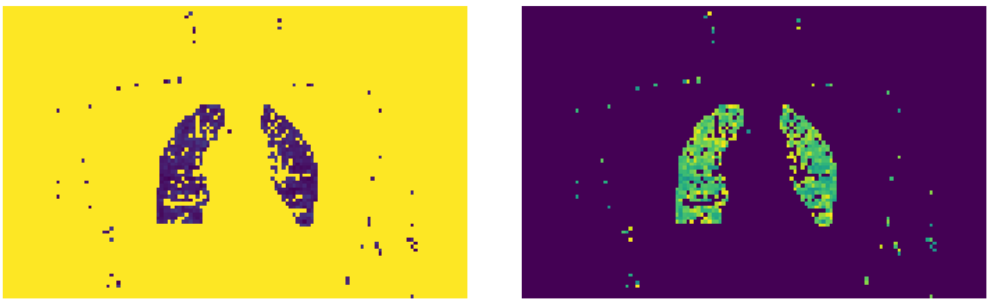
Firstly, gray intensity levels are converted to HU, a measure of radiodensity. This transformation is possible due to the information contained in the DICOM header, namely the Rescale Slope and Intercept.

As the pixel size variates from scan to scan, it is also important to reduce the variance in scanner resolution between systems. For this purpose, voxels are resampled to an isomorphic resolution of [5,5,5] mm, making use of the Slice Thickness and Pixel Spacing information for the transformation.

After reducing the invariance of the intensity values of each scan, it is now possible to find the seed from which the segmentation will grow.

4.2.1 Separation between right and left lungs

For the division of the lungs, the pipeline starts by thresholding the volume in such a way that only the lungs are present. However, not all authors agree on the HU values of the lung: in studies [47, 48, 56], the intervals suggested are $[-1000, -500]$, $[-1000, -400]$ and $[-700, -400]$, respectively. Although the method is quite robust to this selection,the interval $[-800, -500]$ seemed to produce the best results in this dataset. Using this interval, a mask M_{HU} is generated with ones in the voxels whose intensity belong inside this range, and zeros in the remaining parts.



(a) **(b)**

Figure 4.3: Exemplification of HU Mask and corresponding application: plot (a) is the mask HU, obtained with the values belonging to the interval $[-800, -500]$; plot (b) is the result of the application of the same mask to the CT volume

Subsequently, a sum projection of this mask is made, creating a “cumulative transverse plane”, as illustrated in the left part of Figure 4.4. Then, a new sum projection of this plane is performed, originating a line profile (right part of Figure 4.4). Using the local minimum,

a plane can be defined in the original image that separates both lungs and subsequently create two different volumes with only one lung each.

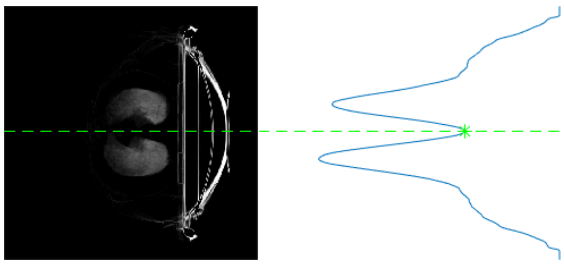


Figure 4.4: Separation between right and left lungs. On the left, a cumulative transverse plane with separation line superimposed in dashed green; On the right, cumulative profile, with local minimum as a green star and the separation line in dashed green.

Figure 4.5: Resulting volumes after the separation

4.2.2 Placement of the seed

Three different techniques were developed for the choice of the seed. These are described next.

4.2.2.1 Method 1

This technique takes advantage of anatomic and image acquisition knowledge.When considering the lungs, one can identify them as being on the top of the thorax. As most of the scans showed images from the thoracic cavity, the methods starts by placing the seed in the position $[\frac{1}{3}, \frac{2}{3}]$ for the right lung and $[\frac{2}{3}, \frac{2}{3}]$ for the left lung, of the central coronal plane. On the next step, a vertical search is performed, by looking for an intensity inside the theoretical range of HU values of the lungs considered ($[-800, -500]$),

as shown on the left side of Figure 4.6. If by the end of the search, no seed is found, then a tolerance is added to the interval, iteratively, until a value belonging to the range $[-800 - tolerance, -500 + tolerance]$ is detected.

4.2.2.2 Method 2

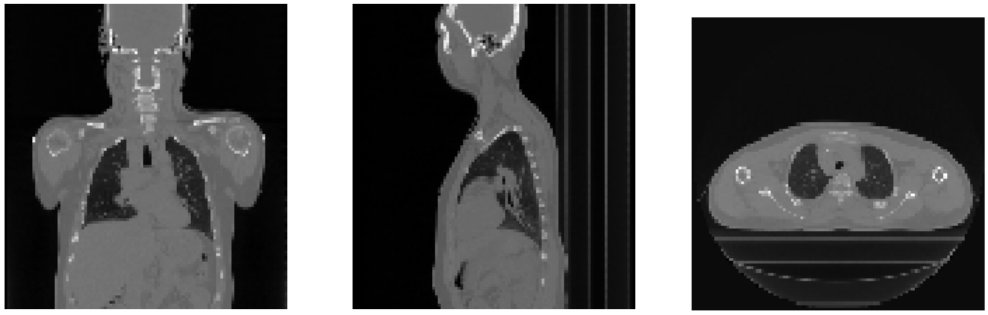
This method represents a variation from the previous method. As before, the seeds are initialized in the positions $[\frac{1}{3}, \frac{2}{3}]$ and $[\frac{2}{3}, \frac{2}{3}]$, for left and right lungs, respectively. However, the search is made radially, based on the distance from the initial point. In this way, the closest point to each of the initial seeds in the mask M_{HU} is chosen as the new seed, as shown on the right side of Figure 4.6.

(a) **(b)**

Figure 4.6: Seed location by Methods 1 (a) and 2 (b). The first method consists on the search of an intensity value in the interval $[-800 - tolerance, -500 + tolerance]$ on the vertical axis, whereas the second method searches for the voxel in the interval that corresponds to the minimal distance to the initial seed.

4.2.2.3 Method 3

The third method is based upon the reasoning applied on Section 4.2.1, and is performed for each lung separately. Considering the coronal plane (plot (a) from Figure 4.7), the y position of the seed is determined according to the local maximum closest to the plane defined to separate the lungs (see right part of Figure 4.4). Next, the sagittal plane (plot (b) from Figure 4.7) of M_{HU} is obtained, represented on the top-left plot of Figure 4.8. The z coordinate will correspond to the local maximum of the sum projection of the biggest connected component of M_{HU} (yellow structure in the top-left plot of Figure 4.8), as visible on the right plot of Figure 4.8. Lastly, the x coordinate is chosen in a similar way: the sum profile of this direction is computed and the local maximum position is chosen, as shown in the bottom plot of Figure 4.8.



(a) (b) (c)

Figure 4.7: Planes of the volume: coronal on the left (a), sagital on the middle (b) and transversal on the right (c)

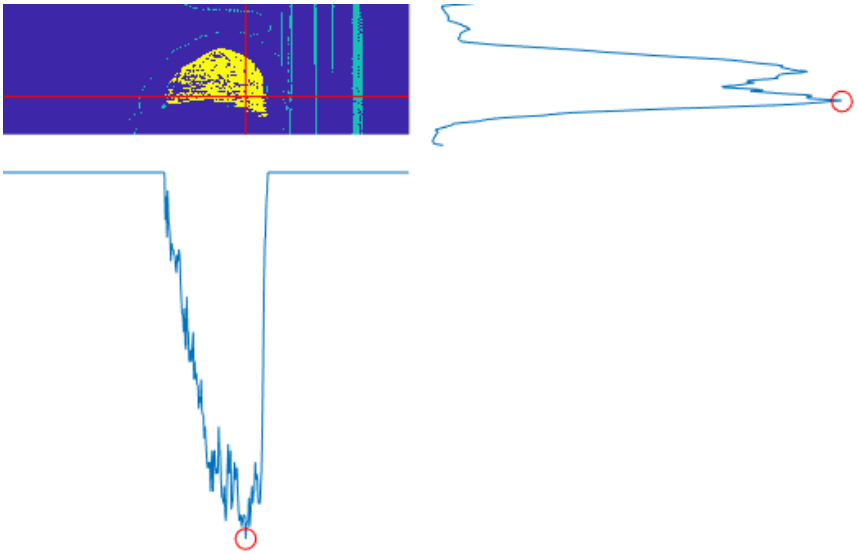


Figure 4.8: Seed location by Method 3 (illustration for one of the lungs, only).

4.2.3 Iterative Region Growing

As mentioned in the Background Chapter (Section 2), in the standard RG algorithm, a threshold is manually chosen, that determines which voxels belong to the segmented region. Moreover, this parameter depends not only on the problem but also on the CT scan in question, as it depends on either the pathological condition of the lungs or the machine in which the patient was scanned. The method here proposed allows to update this threshold, Th, automatically and iteratively, as represented on the Algorithm 2. This procedure uses three parameters: the Tolerance Threshold Initialization Th_0, Maximum Area Threshold Th_H, and Minimum Area Threshold Th_L. The Th_H and Th_L parameters are defined according to previous biomedical knowledge on maximum and minimum lung volumes. Firstly, Th is initialized (Th_0). Then, just like the standard Algorithm 1, the segmented region expands from the seed, following a connectivity criteria, retrieving the

resulting segmented volume. If this volume is superior/inferior to Th_H/Th_L, then Th is decreased/increased, respectively. This step allows to adjust the threshold parameter iteratively and to make the algorithm more robust to the initial parameter Th_0.

Algorithm 2 Iterative Region Growing Algorithm.

Inputs:
Seed vector, $s = [s_x,s_y,s_z]$
Volume to be segmented, CT
Tolerance threshold initialisation, Th_0
Maximum area threshold, Th_H
Minimum area threshold, Th_L
Output:
Volumetric mask, $Mask$, with the same size as CT

```
Th initialisation, Th = Th_0
while Algorithm 1 is iterating do
    if Count(Mask == 1) > Th_H then
        Stop Algorithm 1
        Update Th, Th = Th − 10
        Restart Algorithm 1
    end if
    if Count(Mask == 1) < Th_L then
        Update Th, Th = Th + 10
        Restart Algorithm 1
    end if
end while
```

In this version of the algorithm, there are more parameters to define (Th_0, Th_H and Th_L) when comparing with the standard version (Th,Algorithm 1), which could suggest that it would be even more difficult to establish all these values. The original algorithm can be quite sensitive to the value of Th, which usually is manually introduced and not generalizable to all patients. In this version, the values of the parameters were predefined as follows: $Th_0 = 225$ HU, $Th_L = 3000$ voxels and $Th_H = 40000$ voxels. These limits were empirically found, taking into account the maximum and minimum lung sizes in voxels.

A resume of all the processes can be observed on Figure 4.9, representing all the steps explained above. In the end, the results were evaluated by comparison with the ground truth manually delineated.

A comparison method was also implemented, using a DL approach. Such models have been increasing popularity and are considered the future of image processing. They do not require (a manual) selection and extraction of features, they are able to identify hidden and complex patterns and to learn the optimal attributes from the available images. A U-Net method was chosen due to its capacity to deal with coarse inference and low-resolution problems present in other CNN architectures (see Chapter 2).

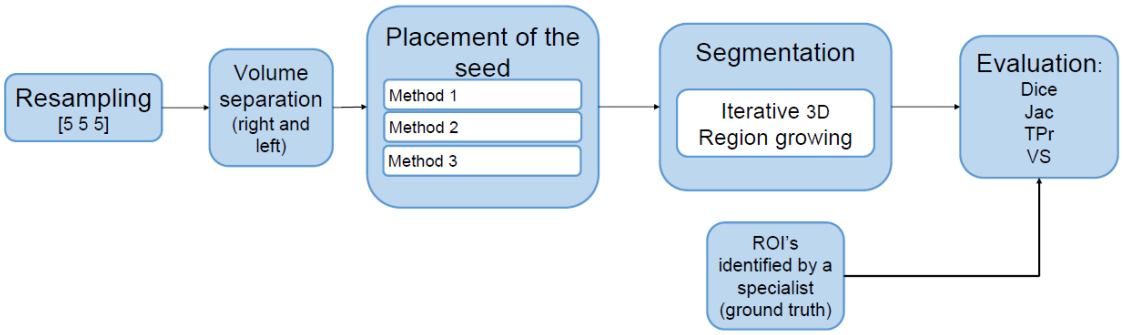


Figure 4.9: Lung Segmentation methodology - Region Growing technique.

4.3 U-Net

This model was adapted from the original work of Fabian Isensee *et al.* [32] for BraTS, a Brain Tumor segmentation challenge. The team achieved DICE scores of 0.858 for whole brain tumour, 0.775 for core tumour and 0.647 for enhancing tumour. Their work, inspired by [37], presents some novelty aspects: normalization techniques, nonlinearity of the model, the number of feature maps throughout the network, the architecture of the contracting pathway and the structure of the upsampling pathway [32]. Besides, to deal with a small and imbalanced dataset, a variety of data augmentation techniques and multiclass adaptation of the dice loss was implemented by the team.

Due to the use of different machines and non-standardized intensity values, a normalization process is crucial to allow data from different patients to be processed by one single algorithm, especially in the implementation of neural networks. This could lead to initial biases of the network. With this goal, the following pre-processing steps were implemented, as in the original work, for each patient:

1. Normalization - subtraction of the mean and division by the standard deviation;
2. Clipping - clip all the values inside the range [-5,5];
3. Rescaling - rescale the remaining values to [0,1], with background defined as 0.

Next, the dataset A was split into train and validation sets. It was defined 80% as the percentage of data that should be held over to training, remaining 20% to be used for validation, a common ratio in machine learning models [57]. In this way, it is maximized the number of samples, giving the model more variance to learn, while leaving a fair part for validation of the models, allowing the performance statistics to have more variance.

To prevent overfitting due to the limited size of the dataset, and like the original model, a large variety of data augmentation techniques was performed on the fly during training: random scaling, random elastic deformations, random rotations, gamma correction augmentation and mirroring.

Just like Ronneberger *et al.* [37], Isensee architecture comprises a context/contracting path, followed by a localization/expanding path with multiple connections between both pathways. This allows the network to encode higher dimensional feature representations and associate them with features from shallower layers to precisely locate the structures of interest. Unlike [37], the author uses context modules following the convolution blocks as a pre-activation residual block. Each block is composed of two convolutional layers interspersed by a dropout layer. Each context block is connected to the following blocks through a convolution layer, allowing the network to decrease the resolution of the feature maps. On the other side, to achieve a higher spatial resolution, upsampling blocks are employed. Each of these blocks is composed of an upscale layer, which repeats feature voxels twice in each spatial direction, succeeded by a convolution layer, responsible for halving the number of feature maps. Then, a concatenation recombines the resulting features with features from the corresponding level of the contraction pathway. This concatenation is followed by a localization module, responsible for recombining these feature maps and for shortening their number, reducing the memory consumption. This localization module comprises two convolution layers (3x3x3 and 1x1x1), reducing once again to half the number of feature maps. Lastly, each convolution layer was followed by a non-linear leaky ReLU with a negative slope of 10^{-2}, defined by $y = max(0.01x,x)$. The architecture of the network was not changed in this project[2].

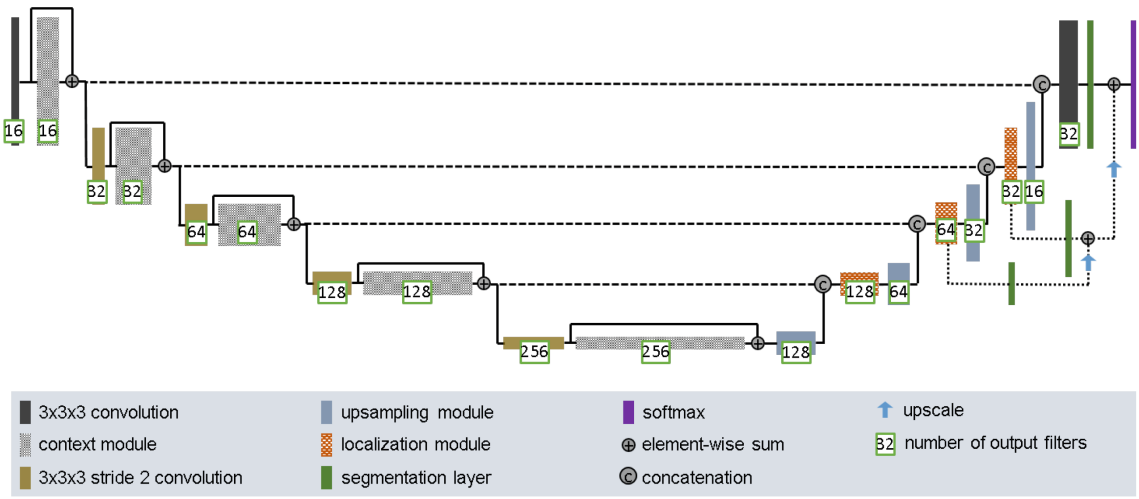


Figure 4.10: U Net architecture, aggregating high level information from the context pathway (left) with the precise location from the localization pathway (right). Figure adapted from [32].

A common problem in medical image segmentation is the class imbalance in each image, where, usually, the background is composed of much more voxels than the class it is pretended to identify. Such imbalance can interfere with the training of the network, depending on the used loss function [32]. In this dataset, lung voxels represent only 0.5% to 3.5% of all the voxels in the image, depending on the patient. To properly deal with

[2]For more information on the network architecture, please check [32].

class imbalance, a multiclass (weighted) DICE loss function is applied:

$$L_{dc} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_{i,k} v_{i,k}}{\sum_i u_{i,k} + \sum_i v_{i,k}} \tag{4.1}$$

where u is the prediction of the network and v the true value, from the ground truth segmentation map. Both u and v have the same shape, I by K, with $i \in I$ representing the voxels in the training patch and $k \in K$ the classes. $u_{i,k}$ and $v_{i,k}$ represent the output of the network and ground truth for class k at voxel i, respectively [32].

All results obtained for both approaches are presented in the next section - Chapter 5, as well as a comparison of both algorithms in terms of performance.

5

Results and Discussion

In this chapter, it will be presented the results obtained thought the two approaches described in Chapter 4. The purpose is to analyze each of these techniques, and benchmark the results. The explanation is divided into IRG (Section 5.1), U-Net network (Section 5.2), and comparison of both (Section 5.3).

5.1 Region Growing

For the IRG algorithm, three different methods for seed placement were implemented on CT scans of dataset A. To evaluate the performance of these methods, it was verified if the resulting seeds fell inside the ground truth mask, and the distance to the centroid of these masks was computed[1]. Although the seed being close to the centroid of the lung is not mandatory, it is a good indicator of the proximity to the lung.

The results are illustrated in Table 5.1. It is possible to notice that all seeds defined by method 3 fall inside the lung region and additionally, are the ones that are significantly closer to the centroid when compared with the seeds resultant from the other techniques, according to the Wilcoxon Signed-Rank test at the 5% level.

Table 5.1: Seed placement performance. Best results in bold and signalled with "*" if statistically significant, according to the Wilcoxon Signed-Rank test at the 5% level.

	Method 1	**Method 2**	**Method 3**
Percentage of valid seeds	78.03	78.03	**100.00***
Distance to centroid	21.47	18.61	**6.41***

To evaluate the segmentation performance, an HU threshold baseline technique was computed, by selecting the biggest connected component of M_{HU} (the yellow region on Figure 4.8). The results for each automatic seed placement method and for the centroid of the ground truth mask are shown in Table 5.2, for all segmentation methods.

[1] However, it is important to notice that there are cases in which the centroid does not fall inside the ground truth mask.

Table 5.2: Segmentation performance. Best results are presented in bold.

Segmentation	Seed	DICE	Jac	TPr	VS
HU threshold	Method 1	0.661	0.576	0.605	0.703
HU threshold	Method 2	0.656	0.571	0.600	0.691
HU threshold	Method 3	0.812	0.707	0.740	0.858
HU threshold	GT centroid	0.812	0.707	0.740	0.861
Standard Region Growing	Method 1	0.736	0.703	0.716	0.779
Standard Region Growing	Method 2	0.720	0.687	0.700	0.752
Standard Region Growing	Method 3	0.894	0.853	0.871	0.926
Standard Region Growing	GT centroid	0.872	0.833	0.850	0.916
Iterative Region Growing	Method 1	0.736	0.703	0.716	0.836
Iterative Region Growing	Method 2	0.741	0.707	0.721	0.841
Iterative Region Growing	Method 3	**0.923**	**0.882**	**0.900**	**0.956**
Iterative Region Growing	GT centroid	0.886	0.846	0.863	0.930

From Table 5.2, it can be confirmed that the proposed IRG outperforms the Standard RG. As expected from the results on Table 5.1, seeds placed by method 3 generate the best segmentation results, and the corresponding segmentation results with the IRG method are statistically significant when compared to all other results, except for the one of Standard RG with Method 3, according to the Wilcoxon test with Bonferroni correction at the 5% level. Using the same test, the results shown a significant difference of Volume Similarity between the Standard RG and the IRG results for methods 1 and 2.

In order to evaluate the performance of the segmentation algorithm, and minimizing the effect of the placement of the seed on the results, it was studied the efficiency of the technique solely on the cases where the seeds were placed inside the ground truth lungs. The results are shown on Table 5.3. It can be seen that, when the seed is valid, method 1 originates seeds whose segmentation results are better, according to DICE, Jaccard Index and True Positive Rate metrics. However, when looking at Volumetric Similarity, method 3 still leads to better segmentations. Such results may indicate that the design of an algorithm combining the ideas behind method 1 and method 3 is a possible future direction.

The significance of these results was tested according to the Wilcoxon Signed-Rank test at the 5% level and revealed that the difference between iterative and non-iterative versions are not significant, except against Iterative and Standard RG using the centroid (in terms of TPr and VS for method 3, DICE and VS for method 2, and DICE, JAC and VS for method 1). Both are significantly better than HU threshold, according to the same test.

The results also suggest that the IRG version leads the algorithm to be more robust to the placement of the seed since the results are not significantly different between different seed placement methods. In fact, as long as the seed is correctly placed inside the lung surrounded by HU values belonging to the interval $[-800, -500]$, the segmentation results

Table 5.3: Segmentation performance, considering valid seeds only. Best results are presented in bold.

Segmentation	Seed	DICE	Jac	TPr	VS
HU threshold	Method 1	0.639	0.559	0.587	0.671
HU threshold	Method 2	0.634	0.554	0.580	0.663
HU threshold	Method 3	0.812	0.707	0.740	0.858
HU threshold	GT centroid	0.786	0.685	0.717	0.834
Region Growing	Method 1	**0.925**	**0.883**	**0.900**	0.952
Region Growing	Method 2	0.895	0.854	0.870	0.922
Region Growing	Method 3	0.894	0.853	0.871	0.926
Region Growing	GT centroid	0.892	0.852	0.870	0.931
Iterative Region Growing	Method 1	**0.925**	**0.883**	**0.900**	0.952
Iterative Region Growing	Method 2	0.914	0.872	0.888	0.945
Iterative Region Growing	Method 3	0.923	0.882	**0.900**	**0.956**
Iterative Region Growing	GT centroid	0.900	0.859	0.877	0.939

are quite similar, as it can seen from Figure 5.1, which demonstrates one of the patients of this dataset. The first row represents the results from method 1, where the seed is located near the bottom frontier of the lung, while in the second row are the results of method 3, where the seed is close to the centre of the lung. As expected from the results of Table 5.3, the final segmentations are alike.

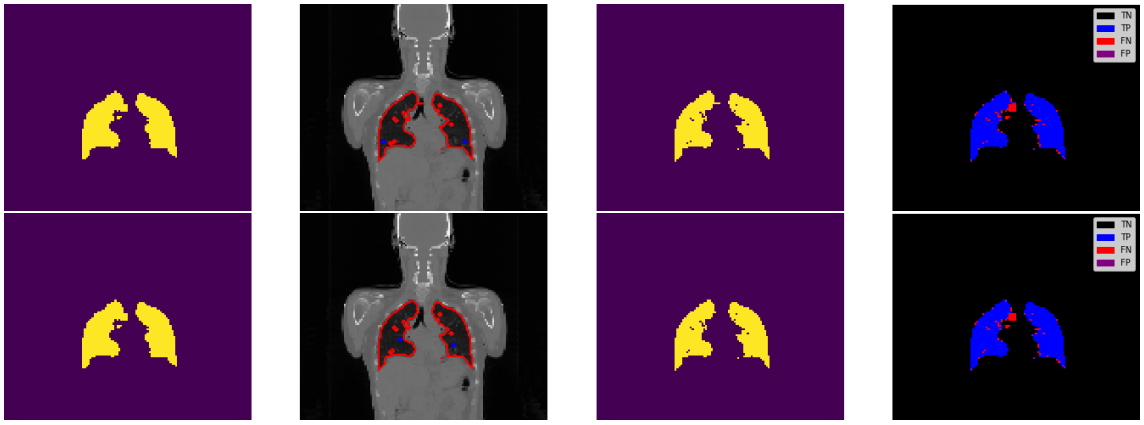


Figure 5.1: Example of the results of lung segmentation for method 1 (top) and 3 (bottom). Ground truth on the far left; CT with superimposed results (red) and seeds (blue) on the middle left; segmentation results on the middle right; and representation of the four cardinalities of the confusion matrix on the far right.

It is important to note that there were multiple cases in which the seeds were correctly placed inside the lung, but too close to bronchi or vessels, and therefore close to HU values outside the desired interval. In such cases, the algorithm cannot segment properly.

5.2 U-Net

For the U-Net approach, multiple parameters were experimented (Table 5.4), mostly concerning data structure, data augmentation techniques and learning parameters of the network. Volume size corresponds to the size of the input and output volume (image). The up-convolution operation is, as the name suggests, the operation used to augment the volume size through layers. The batch size and validation size correspond to the number of volumes that passes through the network at each time for each epoch for training and validation sets, respectively. The number of epochs corresponds to the number of times the dataset will pass through the network. Patience corresponds to the number of epochs that the network will wait for to reduce the learning rate to half after the validation loss is not improving. Early stop corresponds to the number of epochs that the network will wait to stop the training after no improvement is made in validation loss. Flipping, Permutation and Distortion are data augmentation techniques.

Table 5.4: Parameters tested in different configurations.

Parameter	Experimented Values
Volume Size	64^3,112^3
Up-convolution operation	Deconvolution, Upsampling
Batch Size	1,5
Validation Size	2,5
Number of Epochs	250,300
Patience	6,70,107
Early Stop	30,129,150
Flip	False, True
Permute	False, True
Distort	False, True

Due to limitations of computational resources and time required for each configuration of the model, most of them were experimented using an input volume of 64x64x64 voxels. However, using an additional computer, it was possible to verify the influence of the input size on one of the configurations (Configuration 2, Table 5.5). The input size tested corresponded to a volume of 112x112x112 voxels. To perform this decreases in shape, it was required a resampling operation, which lead to a decrease on the input resolution and perhaps the inability of the model to learn certain inherent features. To reduce the time and memory required for each model to train, it was taken advantage of GPU capabilities.

From the 70 patients with ground truth for each lung (Group A), 1 had only ground truth for one of the lungs. Therefore, only 69 CT scans were used, 80% of those were used for training of the network (corresponding to 55 patients) and 20% for validation (14 patients). Additionally, 22 volumes from Group B (with ground truth of “both lungs”) were used to test overfitting.

Table 5.5: U-Net Configurations. Parameters in italic correspond to parameter differences from the first model. From model 10, only the parameters that led to better results were changed.

Model	Shape	Labels	Deconv.	Batch size	Validation Batch Size	# Epochs	Patience	Early Stop	Flip	Permute	Distort
1	64	1,0	True	1	2	300	70	150	False	True	True
2	*112*	1,0	True	1	2	*250*	70	150	False	True	True
3	64	*1*	True	1	2	300	70	150	False	True	True
4	64	1,0	*False*	1	2	300	70	150	False	True	True
5	64	1,0	True	*5*	2	300	70	150	False	True	True
6	64	1,0	True	1	*5*	300	70	150	False	True	True
7	64	1,0	True	1	2	300	*6*	*30*	False	True	True
8	64	1,0	True	1	2	300	*107*	*129*	False	True	True
9	64	1,0	True	1	2	300	70	150	*True*	True	True
10	64	1,0	True	1	2	300	70	150	False	*False*	True
11	64	1,0	True	1	2	300	70	150	False	True	*False*
12	64	1,0	True	1	2	300	70	150	False	*False*	*False*
13	64	1,0	True	1	2	300	70	150	*True*	*False*	*False*
14	64	1,0	*False*	1	2	300	70	150	False	*False*	*False*

Table 5.6: Results for each model configuration. The multiclass DICE mean corresponds to the symmetric of the loss function. Metrics mean corresponds to the average of the metrics used in the last section, respectively DICE, JAC, TPr and VS. The best results are indicated in bold.

Model	Multiclass Dice mean	Metrics mean	Multiclass Dice median	Metrics median
1	96.6 ± 0.6	96.5 ± 0.7	96.8 ± 0.6	96.7 ± 0.7
2	96.3 ± 1.9	96.1 ± 1.9	96.6 ± 1.9	96.6 ± 1.9
3	96.5 ± 0.6	96.4 ± 0.8	96.6 ± 0.6	96.6 ± 0.8
4	96.4 ± 0.6	96.2 ± 0.9	96.5 ± 0.6	96.3 ± 0.9
5	96.2 ± 0.8	96.1 ± 1.0	96.2 ± 0.8	96.0 ± 1.0
6	96.2 ± 0.9	95.9 ± 1.2	96.5 ± 0.9	96.2 ± 1.2
7	88.0 ± 17	88.5 ± 13.2	93.0 ± 17	91.3 ± 13.2
8	96.3 ± 0.6	96.2 ± 0.7	96.5 ± 0.6	96.5 ± 0.7
9	95.8 ± 1.3	95.3 ± 1.5	96.0 ± 1.3	95.2 ± 1.5
10	96.9 ± 0.7	96.7 ± 0.9	97.1 ± 0.9	97.0 ± 0.9
11	97.3 ± 0.5	97.1 ± 0.7	97.4 ± 0.5	97.2 ± 0.7
12	97.4 ± 0.7	97.3 ± 0.9	97.7 ± 0.7	97.6 ± 0.9
13	97.4 ± 0.6	97.2 ± 0.9	97.8 ± 0.6	97.4 ± 0.9
14	**97.5±0.7**	**97.3±0.9**	**97.7±0.7**	**97.7±0.9**

From Table 5.6, the results suggest that the effect of data augmentation may not always be advantageous. Flipping, distorting and permutation are artificial ways to create data and therefore, if not realistic, they may induce error in the model. However, another way of explaining these results may be that the network has not had time to learn to ignore such transformations. On Figure 5.2, it can be verified that there is still room for improvement: the standard model, like others using data augmentation, has not yet converged, despite all the spikes caused by a high learning rate. However, it was required more time and memory to test these models through more epochs. In these cases, where no data augmentation techniques are applied and there is less data for the model to learn, it can start to memorize the training samples, and therefore it may not be able to generalize to test samples (overfitting). To check if the model was overfitting, it was tested the model in a test dataset with 22 samples. The results can be seen on Table 5.7 and Figure 5.3. From the achieved results, it can be concluded that the model is able to generalize to new samples.

Table 5.7: Results for the testing set.

Model	Multiclass Dice mean	Metrics mean	Multiclass Dice median	Metrics median
1	91.0±22.1	91.8±17.4	96.9±22.1	96.5±17.4

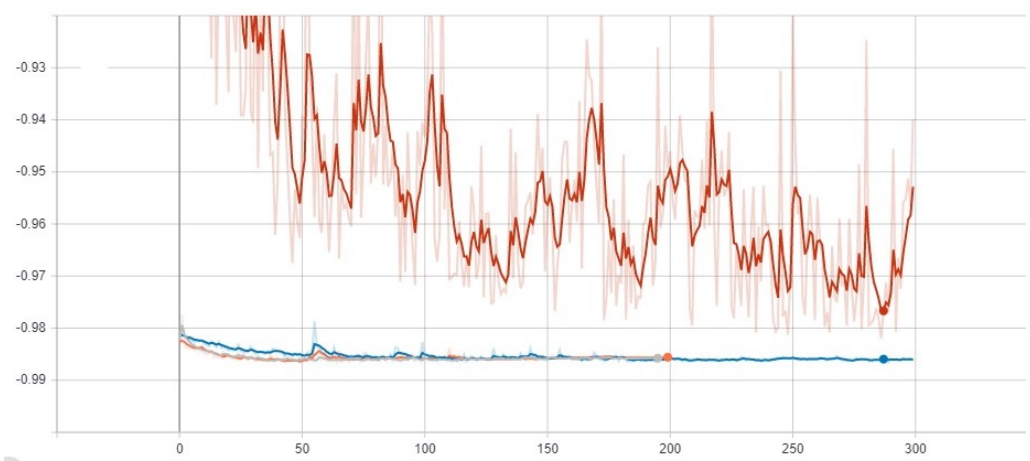


Figure 5.2: Comparison of the convergence speed between three models (validation loss). In red, it can be seen the standard model, configuration 1, with data augmentation techniques. In grey, the model 12, without permutation and distortion. In blue, the model 13, with flipping but without permutation and distortion. In orange, the model 14, without permutation and distortion and upsampling (instead of deconvolution).

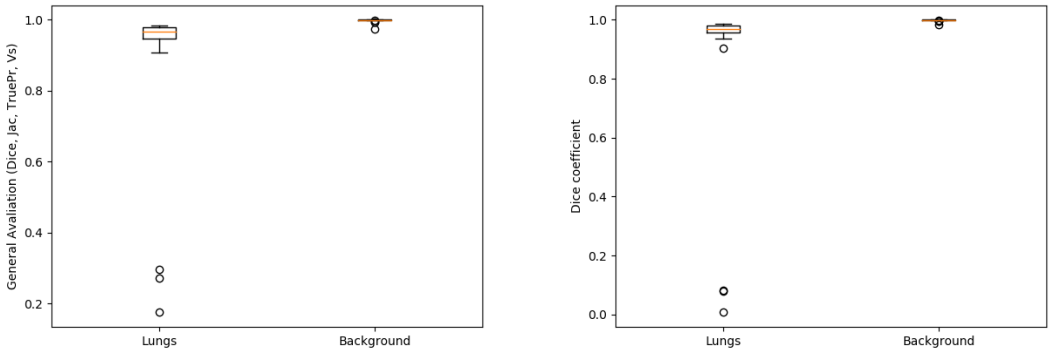


Figure 5.3: Results quantification with the Mean of the Metrics previously chosen (Dice, Jac, TPr, VS) and with the multiclass Dice function

5.3 Comparison

The two different approaches were compared according to the metrics previously defined DICE, JAC, TPr and VS.

Using the dataset that included ground truth for each of the lungs (69 patients), 55 were used to train the network (80%) and 14 for validation (20%) and comparison. The results given by the IRG algorithm using the seed placement method (which demonstrated to perform better, method 3) and by the U-Net model (which also performed better, model 14) were compared.

To compare in an unbiased way, the results of the latter model were resampled to the orig-

inal shape and then resampled to an isometric resolution of $[5,5,5]$ mm, like performed for the first algorithm. This upsampling creates more resolution, using interpolation methods, and therefore may negatively affect the results. The results are shown on Table 5.8 and on Figure 5.5.

Table 5.8: Means, medians and respective standard deviation of each method. Statistically significant results are marked with "$*$", according to Wilcoxon-Signed Rank test at the 5% level.

Metric	Iterative Region Growing mean	Iterative Region Growing median	U-Net CNN mean	U-Net CNN median
Dice	0.872±0.247	**0.958±0.247**	**0.938±0.013**	0.944±0.013
Jac	0.826±0.242	**0.920±0.242**	**0.884±0.023**	0.894+0.023
TPR	0.844±0.243	0.928±0.243	**0.955±0.015***	**0.958±0.015***
VS	0.933±0.137	0.968±0.137	**0.981±0.008***	**0.981±0.008***

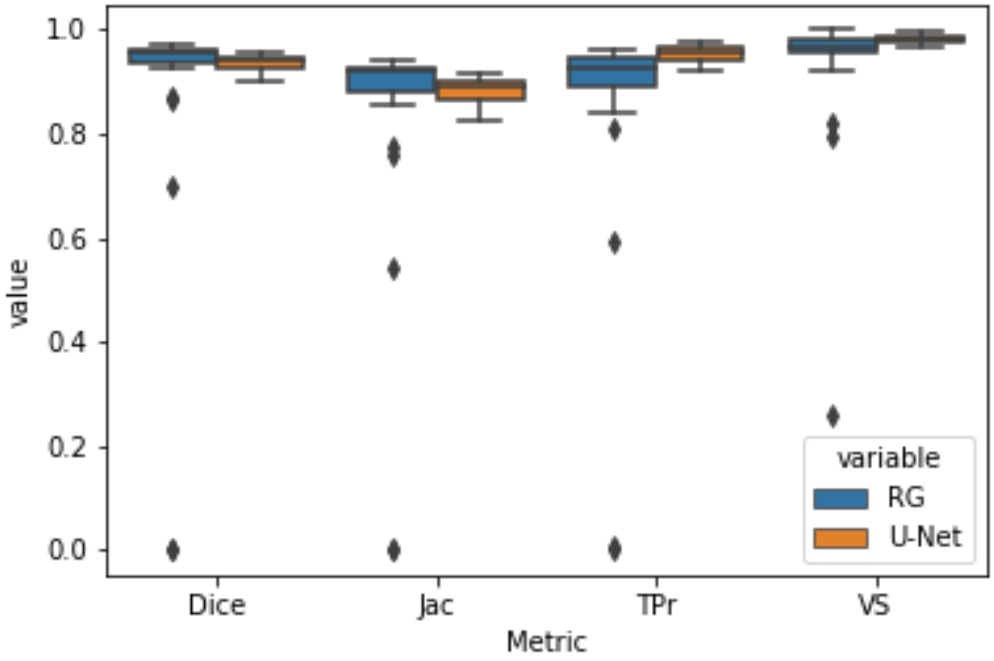


Figure 5.4: Results quantification with the different metrics

The results suggest that the DL model outperforms significantly the IRG method in terms of VS and TPr. For DICE and JAC, the results seem to suggest that the IRG method is better, however there is no statistical difference between the two distributions according to the Wilcoxon Signed-Rank test at the 5% level.

Figure 5.5: Example of the results of lung segmentation for the Iterative Region Growing algorithm (top) and U-Net CNN (bottom). Ground truth on the far left; CT with superimposed results (red) and seed (yellow) on the middle left; segmentation results on the middle; representation of the four cardinalities of the confusion matrix on the middle right; CT after application of the segmentation mask on the far right.

5.4 Overview

Overall, both algorithms achieved good segmentation results. All lungs were identified by the U-Net model. However, the Iterative Region Growing method was not able to segment properly 3 lungs, despite correct seed placement (near the bronchi, patient 2, 9 and 10 represented on Appendix C). Both models were able to accurately delineate the overall shape of this organ, as well as fine-grained details such as the limitations with the trachea and the bronchi. It is noticeable in Figure 5.5 that Iterative Region Growing creates tiny holes, but in future work, these can easily be filled with a closing operation[2].

Due to differences in the used databases, a fair comparison with other reported studies is not possible. This database is composed of CT scans from both adults and infants, and CT's of full-body, abdominal and thoracic, increasing the difficulty of the segmentation. It is also important to note that there are cases where the ground truth had small holes or included part of other structures (bronchi). Generally, lung segmentation can reveal itself as a challenging task because of the similar intensity values present in the pulmonary structures and inhomogeneity of the lung region [58].

From Table 5.8 it is possible to conclude that the results from U-Net methods are significantly better when comparing True Positive Rate and Volume Similarity, an overlap and a volume-based metric, respectively. The results also suggest that the Iterative Region Growing model performs better according to DICE and the Jaccard Index, however there is no significant differences.

Different metrics were used to evaluate different quality aspects of the segmentation. DICE, Jaccard Index and True Positive Rate are overlapping metrics. Dice is the most used metric in this type of task, usually to perform direct comparisons between ground

[2]All the results can be seen in Appendix C

truth and automatic segmentations, but also as a measure of reproducibility [43]. Jaccard index evaluates the intersection between two segmentations relatively to their union. These measures tend to penalize bigger segmentations, that is, tend to penalize more False Positives (purple on the Figure 5.5 than False Negatives (in red)) [43]. This may be the case of these results.

True Positive Rate measures, as the name indicates, the portion of positive voxels that are identified as true by both segmentations. Recall errors have two causes: missing regions on the segmentation results or added regions that do not have correspondence in the ground truth. In some medical imaging tasks, it is preferable added regions in detriment of missing regions, such as the delimitation of tumours. In this type of scenario, an algorithm that maximizes recall instead of precision, avoiding missing regions, is preferable, even if it requires to add false regions [43]. In such cases, the U-Net CNN algorithm will have significantly better results than the Iterative Region Growing algorithm.

Volume Similarity compares the absolute volume of the ground truth with the corresponding volume of the segmentation, not taking in count at all the overlap of the two sets. This measure is recommended when the magnitude of the predicted volume is more important than its boundary or alignment and therefore, in this cases, algorithms should perform segmentations with volumes as more similar to the ground truth as possible. In such cases, the U-Net CNN algorithm will have significantly better results than the Iterative Region Growing algorithm.

Each of these metrics penalizes low density, and therefore, the presence of tiny holes in the segmentations performed by the Iterative Region Growing algorithm may have detracted the results.

It is important to refer that each case is different, and these results may not replicate in other datasets. In fact, the dataset used to test this results has only 14 samples. Further work would include tests in more samples to better generalize the results.

In conclusion, two algorithms were developed to segment the lungs using two very different approaches. The Iterative Region Growing algorithm is able to achieve good performances, and even surpass some of the recently proposed algorithms such as the ones described in [46, 48]. Besides the advantages described in Chapter 3, is characterized by its simplicity, no need of training phase, and low running time. Nevertheless, when the placement of the seed occurs near other structures such as the bronchi, its expansion may not satisfy the expectations. Deep learning models had been tested in many approaches as it has seen in Chapter 3, achieving state-of-art results in many medical imaging tasks. This model is a data-driven, end-to-end model trained from scratch using only 55 CT scans and it does not require feature extraction and selection. However, interpreting the model is not easy, since it is a complex system and it requires time and high computing power. A resume of the advantages and disadvantages of each model is represented on Table 5.9.

Table 5.9: Advantages and disadvantages of Iterative Region Growing and U-Net CNN algorithms

	Iterative Region Growing	**U-Net CNN**
Complexity	Simple	Complex
Training	No training	Hours of training
Knowledge	Uses previous knowledge of the problem	Does not require previous knowledge
Running time	Low	Low once it is trained
Generalization	Difficulties with inhomogeneity	Good
Computing power	Low	High
Performance	Good	Good

Therefore, we conclude that both methods have good performances. In cases where the alignment and position of the segmentations are important, both methods have similar results, so perhaps the Iterative Region Growing would be a better choice, due to its simplicity and low memory requirements. In cases where is preferable to keep track of the true limits of the segmentation, perhaps in detriment of added false positives, or the magnitude of the predicted volume is of major importance, then the U-Net CNN algorithm outperforms the Iterative Region Growing.

6

Conclusions and Future work

Segmentation is a problem present in many different domains, and of great importance in the medical image field, such as in Radiotherapy Planning. This is a crucial task for the management of cancer patients therapy, but it is still very time consuming, prone to human errors and high intra and inter-subject variance. The identification and delimitation of organs at risk is crucial to minimize its exposure to radiation, and therefore, an important topic in treatment.

In this work, two different algorithms for lung segmentation in CT are presented and evaluated. While the Iterative Region Growing is simple, fast, does not need a training phase and it is based on expert knowledge of the problem, the U-Net CNN is a data-driven, end-to-end model, who does not require feature selection.

These two completely automated models may help to minimize the time spent by clinicians in the manual analysis of each CT scan, reducing the variability, and ultimately assist in the clinical decisions making of the patient's treatment plan and/or evaluation of its effectiveness.

At the end, three intensity-based seed placement methods were tested, an improvement of the Standard Region Growing model was proposed and a U-Net CNN model was adapted to this problem. Seed placement methods reached 84%, 84% and 96% of Volume Similarity for method 1, 2 and 3 with Iterative Region Growing, respectively. When considering only the valid seeds, the Iterative Region Growing method registered results of 95%, 94% and 96%, respectively. On the other hand, U-Net model achieved 91% of Weighted Dice mean and a mean of the other metrics (DICE, JAC, TPr and VS) of 91.8%. After the comparison of both algorithms, our results suggested that the Deep Learning approach leads to better segmentations in terms of recall and volume similarity. The Iterative Region Growing model seems to perform better in terms of Dice and Jaccard Index, but there is no statistical differences that confirm these results. It is of importance to note how similar the overall results are and how far a standard model can be pushed. It is also relevant to note that the size sample is small and therefore more tests should be performed on more CT scans.

There is always room for improvement. Iterative Region Growing algorithm can be refined

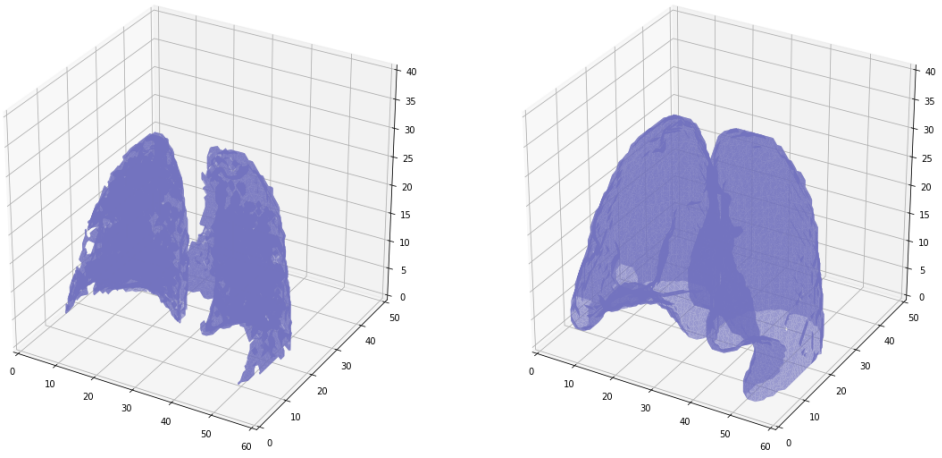
by the application of morphology operations, such as closing, or any other filter in order to eliminate small holes present in the segmentation. Another type of threshold could be also explored, as well as limitations to the placement of the seed outside the thorax. On the U-Net algorithm, would also be worth to see the effect of data augmentation techniques over an increased number of epochs, the implementation of hyperparameters optimization techniques, therefore studying in more detail the non-linearity of the model. It would also be interesting to verify the performance of a model who integrated the reasoning behind both methods, reducing the computational requirements of the U-Net and improving the seed placement and limitations regarding other respiratory structures of the Iterative Region Growing method. Future work should also include the validation of these techniques on bigger and more diverse datasets that could include lung abnormalities, such as pleural effusions, consolidations, and masses, considering that many image segmentation approaches perform well only in lungs with normal or minimal pathological conditions [59]. Another line of interest would be the development of interactive tools for better visualization and manual adjustment of the results.

Appendices

A

Morphology operations

In this chapter, a brief demonstration of how much the morphological operations could benefit the results from the Region Growing method. The images demonstrated correspond to the results of patient 10, obtained with the third method of IRG before and after dilation.



(a) **(b)**

Figure A.1: Example of the potencial of the morphological operations in the results. In (a), the results obtained throught the IRG model. In (b), the same results after dilation.

B

Equalization studies

In the first stages of this project, some pre-processing studies were performed to understand how these could benefit the segmentation procedure.

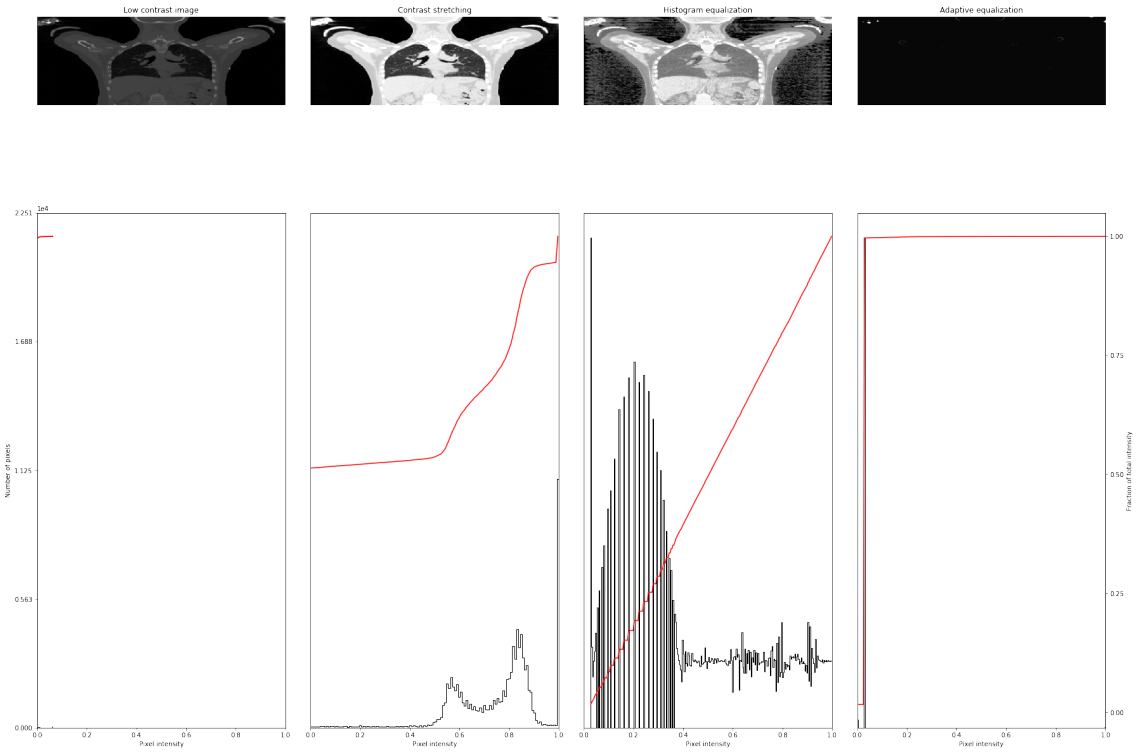


Figure B.1: Results of lung segmentation for patient 2.

C

Results of Comparison

In this chapter, it is possible to see the results of the comparison of the models for all the patients. For all the images, the lung segmentation for the IRG algorithm is represented on the top, and U-Net CNN at the bottom. On the far left, the ground truth; on the middle left, the CT with superimposed results (red); on the middle, segmentation results; on the middle right, the representation of the four cardinalities of the confusion matrix and finally, on the far right, the CT after application of the segmentation mask.

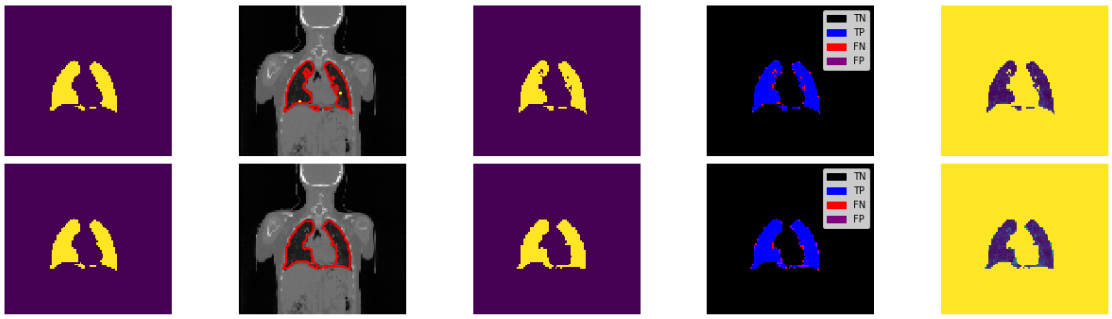


Figure C.1: Results of lung segmentation for patient 1.

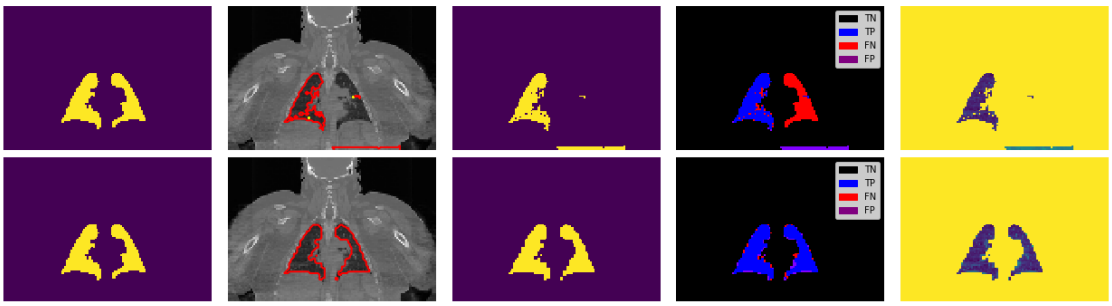


Figure C.2: Results of lung segmentation for patient 2.

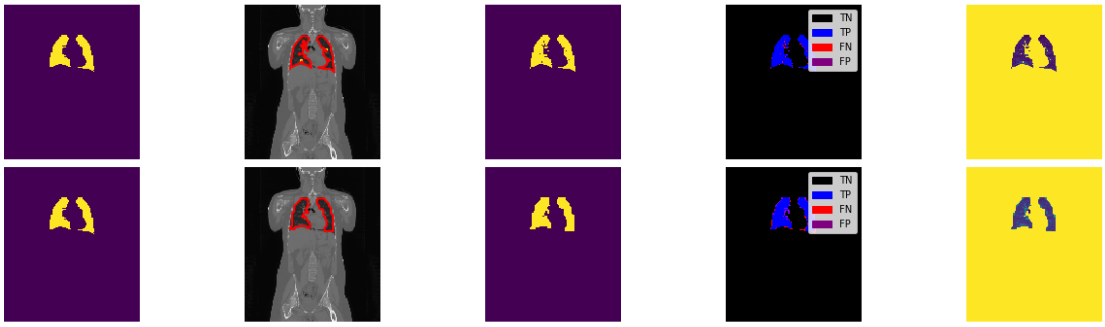


Figure C.3: Results of lung segmentation for patient 3.

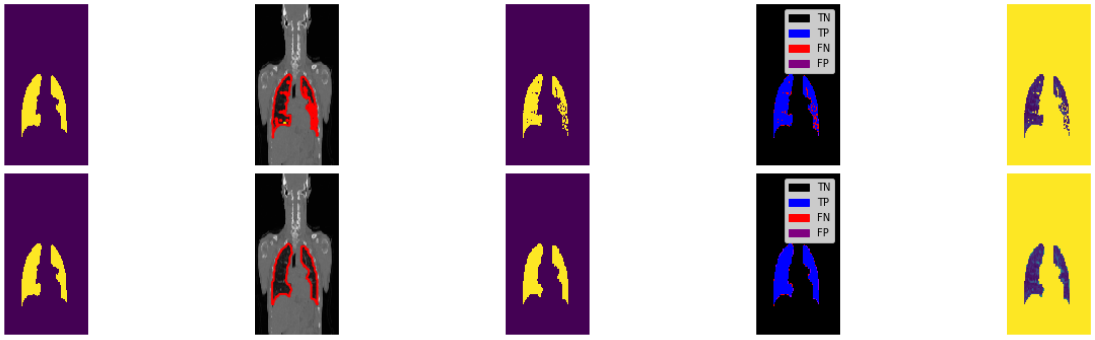


Figure C.4: Results of lung segmentation for patient 4.

Figure C.5: Results of lung segmentation for patient 5.

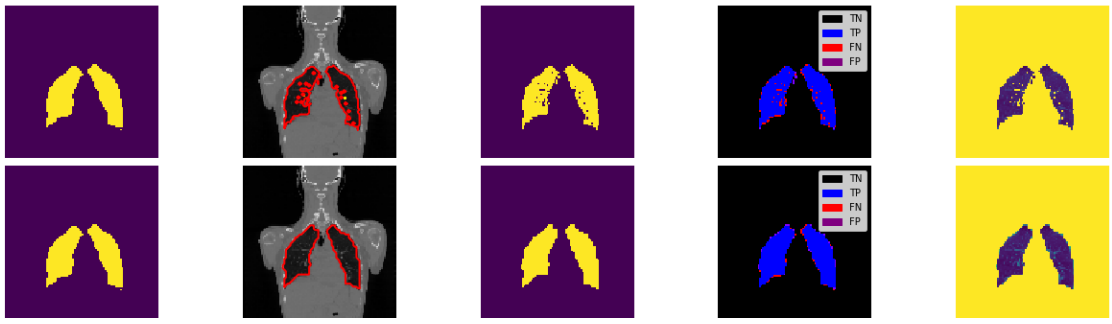


Figure C.6: Results of lung segmentation for patient 6.

Figure C.7: Results of lung segmentation for patient 7.

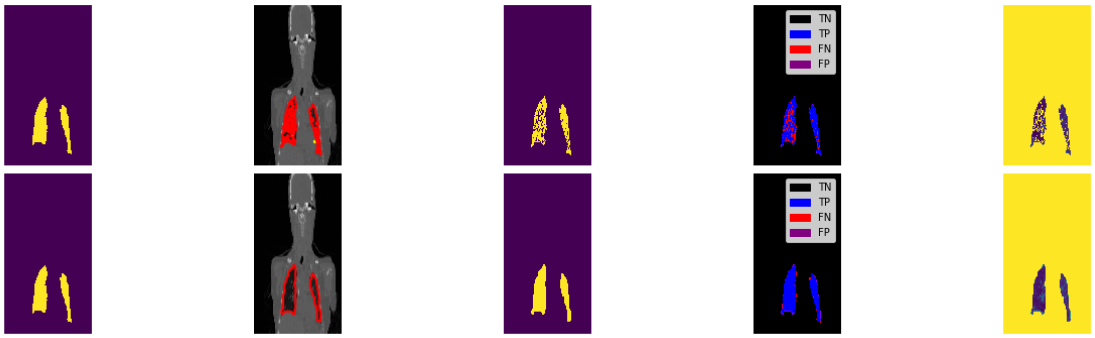


Figure C.8: Results of lung segmentation for patient 8.

Figure C.9: Results of lung segmentation for patient 9.

Figure C.10: Results of lung segmentation for patient 10.

Figure C.11: Results of lung segmentation for patient 11.

Figure C.12: Results of lung segmentation for patient 12.

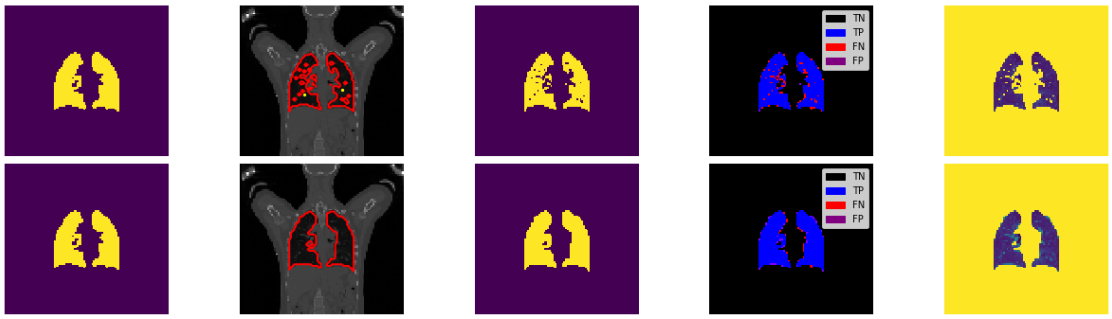


Figure C.13: Results of lung segmentation for patient 13.

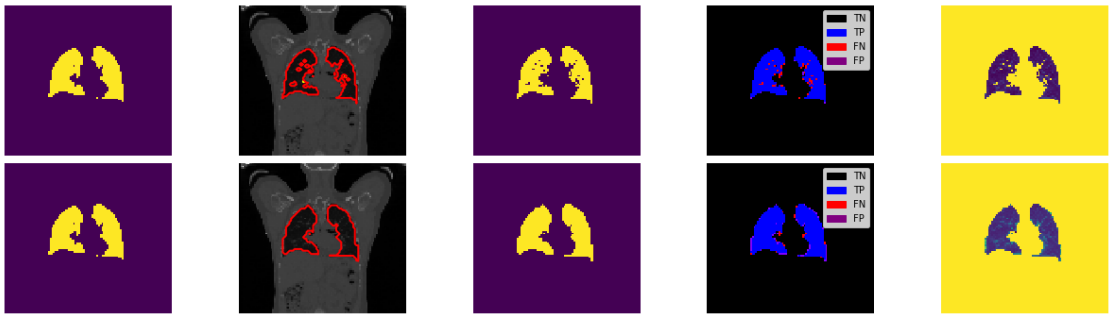


Figure C.14: Results of lung segmentation for patient 14.

D

Scientific Paper

Going back to basics on volumetric segmentation of the lungs in CT: a fully image processing based technique*

Ana Catarina Oliveira[1], Inês Domingues[2], Hugo Duarte[3,2], João Santos[4,2], and Pedro H. Abreu[5]

[1] University of Coimbra, Portugal
anaf.oliveira95@gmail.com
[2] Medical Physics, Radiobiology and Radiation Protection Group, Portuguese Institute of Oncology of Porto (IPO-Porto) Research Center
ines.domingues@isec.pt
[3] Nuclear Medicine Department, IPO-Porto
hugo.duarte@ipoporto.min-saude.pt
[4] Medical Physics Department, IPO-Porto
Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Portugal
joao.santos@ipoporto.min-saude.pt
[5] CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
pha@dei.uc.pt

Abstract. Radiotherapy planning is a crucial task in cancer patients management. This task is, however, very time consuming and prone to a high intra and inter subject variance and human errors.
In this way, the present line of work aims at developing a tool to help the specialists in this task. The developed tool will consider the delimitation of anatomical regions of interest, since it is crucial to identify the organs at risk and minimize the exposure of these organs to the radiation.
This paper, in particular, presents a lung segmentation algorithm, based on image processing techniques, such as intensity projection and region growing, for Computed Tomography volumes. Our pipeline consists in first separating two halves of the volume to isolate each lung. Then, three techniques for seed placement are developed. Finally, a traditional region growing algorithm has been changed in order to automatically derive the value of the threshold parameter.
The results obtained for the three different techniques for seed placement were, respectively, 74%, 74% and 92% of DICE with the Iterative Region Growing algorithm.
Although the presented results have as use case the Hodgkin Lymphoma, we believe that the developed method is generalizable to any other pathology.

Keywords: Lung Segmentation · Computed Tomography (CT) · 3D.

* This article is a result of the project NORTE-01-0145-FEDER-000027, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

Bibliography

[1] M. Dabass, S. Vashisth, and R. Vig, "Effectiveness of Region Growing Based Segmentation Technique for Various Medical Images - A Study," in *Data Science and Analytics* (N. Panda, B and Sharma, S and Roy, ed.), vol. 799 of *Communications in Computer and Information Science*, pp. 234–259, Springer-Verlag Berlin, 2018.

[2] M. Astaraki, M. Severgnini, V. Milan, A. Schiattarella, F. Ciriello, M. de Denaro, A. Beorchia, and H. Aslian, "Evaluation of localized region-based segmentation algorithms for CT-based delineation of organs at risk in radiotherapy," *Physics and Imaging in Radiation Oncology*, vol. 5, pp. 52–57, 2018.

[3] N. G. Burnet, "Defining the tumour and target volumes for radiotherapy," *Cancer Imaging*, vol. 4, no. 2, pp. 153–161, 2004.

[4] T. Lustberg, J. van Soest, M. Gooding, D. Peressutti, P. Aljabar, J. van der Stoep, W. van Elmpt, and A. Dekker, "Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer," *Radiotherapy and Oncology*, vol. 126, no. 2, pp. 312–317, 2018.

[5] International Early Lung Cancer Action Program Investigators, C. I. Henschke, D. F. Yankelevitz, D. M. Libby, M. W. Pasmantier, J. P. Smith, and O. S. Miettinen, "Survival of Patients with Stage I Lung Cancer Detected on CT Screening," *New England Journal of Medicine*, vol. 355, no. 17, pp. 1763–1771, 2006. PMID: 17065637.

[6] "Cancer statistics - Statistics Explained." `https://ec.europa.eu/eurostat/statistics-explained/index.php/Cancer_statistics#Deaths_from_cancer`. Accessed on 28/06/2019.

[7] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[8] N. P. Nguyen, S. P. Krafft, V. Vinh-Hung, P. Vos, F. Almeida, S. Jang, Ceizyk, A. Desai, R. Davis, M. Russ Hamilton, H. Modarresifar, D. Abraham, and L. Smith-Raymond, "Feasibility of tomotherapy to reduce normal lung and cardiac toxicity for

distal esophageal cancer compared to three-dimensional radiotherapy," *Radiotherapy and Oncology*, vol. 101, no. 3, pp. 438 – 442, 2011.

[9] F.-M. S. Kong, T. Ritter, D. J. Quint, S. Senan, L. E. Gaspar, R. U. Komaki, C. W. Hurkmans, R. Timmerman, A. Bezjak, J. D. Bradley, B. Movsas, L. Marsh, P. Okunieff, H. Choy, and W. J. Curran, "Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus," *International Journal of Radiation Oncology*Biology*Physics*, vol. 81, no. 5, pp. 1442 – 1457, 2011.

[10] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76–96, 2014.

[11] G. Pereira, *Deep Learning techniques for the evaluation of response to treatment in Hodgkin Lymphoma*. Master in biomedical engineering, University of Coimbra, 2018.

[12] OpenCV, "Histograms - 2: Histogram Equalization." `https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_histograms/py_histogram_equalization/py_histogram_equalization.html`. Accessed on 30/06/2019.

[13] W. M. Morrow, R. B. Paranjape, R. M. Rangayyan, and J. E. L. Desautels, "Region-based contrast enhancement of mammograms," *IEEE Transactions on Medical Imaging*, vol. 11, pp. 392–406, Sep. 1992.

[14] S. v0.16.dev0 Docs, "Image Adjustment: Transforming Image Content." `https://scikit-image.org/docs/dev/user_guide/transforming_image_data.html`. Accessed on 30/06/2019.

[15] Pythontic.com, "Contrast Stretching Using Python And Pillow." `https://pythontic.com/image-processing/pillow/contrast%20stretching`. Accessed on 30/06/2019.

[16] G. Bradski and A. Kaehler, *Image Processing*, ch. 5, pp. 109–139. O'Reilly Media, 2015.

[17] N. A. Dodgson, "Image resampling," Tech. Rep. 261, University of Cambridge Computer Laboratory, 1992.

[18] J. A. Parker, R. V. Kenyon, and D. E. Troxel, "Comparison of Interpolating Methods for Image Resampling," *IEEE Transactions on Medical Imaging*, vol. 2, no. 1, pp. 31–39, 1983.

[19] Z. Shi, J. Ma, M. Zhao, Y. Liu, Y. Feng, M. Zhang, L. He, and K. Suzuki, "Many Is Better Than One: An Integration of Multiple Simple Strategies for Accurate Lung

Segmentation in CT Images," *BioMed Research International*, vol. 2016, pp. 1–13, 2016.

[20] M. Xu, S. Qi, Y. Yue, Y. Teng, L. Xu, Y. Yao, and W. Qian, "Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset," *BioMedical Engineering OnLine*, vol. 18, no. 1, p. 2, 2019.

[21] X. Sun, H. Zhang, and H. Duan, "3D Computerized Segmentation of Lung Volume With Computed Tomography," *Academic Radiology*, vol. 13, no. 6, pp. 670–677, 2006.

[22] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.

[23] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.

[24] E. E. Nithila and S. Kumar, "Segmentation of lung from CT using various active contour models," *Biomedical Signal Processing and Control*, vol. 47, pp. 57–62, 2019.

[25] Q. Guo, S. Sun, X. Ren, F. Dong, B. Z. Gao, and W. Feng, "Frequency-tuned active contour model," *Neurocomputing*, vol. 275, pp. 2307–2316, 2018.

[26] N. Sharma, A. Ray, K. Shukla, S. Sharma, S. Pradhan, A. Srivastva, and L. Aggarwal, "Automated medical image segmentation techniques," *Journal of Medical Physics*, vol. 35, no. 1, p. 3, 2010.

[27] Z. Kato, "Markov Random Fields in Image Segmentation." `https://inf.u-szeged.hu/~ssip/2008/presentations2/Kato_ssip2008.pdf`. Accessed on 06/09/2019.

[28] A. C. Oliveira, I. Domingues, H. Duarte, J. Santos, and P. H. Abreu, "Going back to basics on volumetric segmentation of the lungs in CT: a fully image processing based technique," in *Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, 2019.

[29] J. Kim, J. Hong, and H. Park, "Prospects of deep learning for medical imaging," *Precision and Future Medicine*, vol. 2, no. 2, pp. 37–52, 2018.

[30] D. Bacciu, P. J. G. Lisboa, J. D. Martin, R. Stoean, and A. Vellido, "Bioinformatics and Medicine in the Era of Deep Learning," *arXiv e-prints*, p. arXiv:1802.09791, 2018.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[32] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge," in *Lecture Notes in Computer Science (including subseries Lecture*

Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 287–297, Springer International Publishing, 2018.

[33] 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, no. 1995, pp. 60–88, 2017.

[34] G. L. Oliveira, C. Bollen, W. Burgard, and T. Brox, "Efficient and robust deep networks for semantic segmentation," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 472–491, 2018.

[35] K. Fukushima and S. Miyake, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition," in *Competition and cooperation in neural nets*, pp. 267–285, Springer, 1982.

[36] 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.

[37] 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 (MICCAI)*, pp. 234–241, Springer International Publishing, 2015.

[38] A. Deshpande, "A beginner's guide to understanding convolutional neural networks." `https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/`. Accessed on 06/09/2019.

[39] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 424–432, Springer International Publishing, 2016.

[40] A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers, and D. J. Mollura, "Progressive and Multi-path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 621–629, Springer International Publishing, 2017.

[41] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv e-prints*, 2018.

[42] X. Bouthillier, K. Konda, P. Vincent, and R. Memisevic, "Dropout as data augmentation," *arXiv e-prints*, 2015.

[43] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool,” *BMC Medical Imaging*, vol. 15, no. 1, p. 29, 2015.

[44] I. Arganda-Carreras, S. C. Turaga, D. R. Berger, D. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, D. Laptev, S. Dwivedi, J. M. Buhmann, T. Liu, M. Seyedhosseini, T. Tasdizen, L. Kamentsky, R. Burget, V. Uher, X. Tan, C. Sun, T. D. Pham, E. Bas, M. G. Uzunbas, A. Cardona, J. Schindelin, and H. S. Seung, “Crowdsourcing the creation of image segmentation algorithms for connectomics,” *Frontiers in Neuroanatomy*, vol. 9, p. 142, 2015.

[45] T. Dietenbeck, M. Alessandrini, D. Barbosa, J. D’hooge, D. Friboulet, and O. Bernard, “Detection of the whole myocardium in 2D-echocardiography for multiple orientations using a geometrically constrained level-set,” *Medical Image Analysis*, vol. 16, no. 2, pp. 386–401, 2012.

[46] N. Birkbeck, T. Kohlberger, J. Zhang, M. Sofka, J. Kaftan, D. Comaniciu, and S. K. Zhou, “Lung Segmentation from CT with Severe Pathologies Using Anatomical Constraints,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 17, pp. 804–811, Springer International Publishing, 2014.

[47] R. V. M. da Nobrega, M. B. Rodrigues, and P. P. R. Filho, “Segmentation and Visualization of the Lungs in Three Dimensions Using 3D Region Growing and Visualization Toolkit in CT Examinations of the Chest,” in *IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)* (P. P. Bamidis, PD and Konstantinidis, ST and Rodrigues, ed.), vol. 2017, pp. 397–402, 2017.

[48] H. Zhou, D. B. Goldgof, S. Hawkins, L. Wei, Y. Liu, D. Creighton, R. J. Gillies, L. O. Hall, and S. Nahavandi, “A Robust Approach for Automated Lung Segmentation in Thoracic CT,” in *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2267–2272, 2015.

[49] P. P. Rebouças Filho, P. C. Cortez, A. C. da Silva Barros, V. H. C. Albuquerque, and J. M. R. S. Tavares, “Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images,” *Medical Image Analysis*, vol. 35, pp. 503–516, 2017.

[50] P. P. Filho, P. Cortez, J. H S Felix, T. Cavalcante, and M. Holanda, “Adaptive 2D Crisp Active Contour Model applied to lung segmentation in CT images of the thorax of healthy volunteers and patients with pulmonary emphysema,” *Brazilian Journal of Biomedical Engineering*, vol. 29, pp. 363–376, 2013.

[51] B. Ait Skourt, A. El Hassani, and A. Majda, “Lung CT Image Segmentation Using Deep Neural Networks,” *Procedia Computer Science*, vol. 127, pp. 109–113, 2018.

[52] M. A. Nogueira, P. H. Abreu, P. Martins, P. Machado, H. Duarte, and J. Santos, “An artificial neural networks approach for assessment treatment response in oncological patients using pet/ct images,” *BMC Medical Imaging*, vol. 17, p. 13, Feb 2017.

[53] M. A. Nogueira, P. H. Abreu, P. Martins, P. Machado, H. Duarte, and J. Santos, “Image descriptors in radiology images: a systematic review,” *Artificial Intelligence Review*, vol. 47, pp. 531–559, Apr 2017.

[54] G. Pereira, I. Domingues, P. Martins, P. H. Abreu, H. Duarte, and J. Santos, “Registration of CT with PET: a comparison of intensity-based approaches,” in *International Workshop on Combinatorial Image Analysis (IWCIA)*, 2018.

[55] J. O. Deasy, A. I. Blanco, and V. H. Clark, “CERR: A computational environment for radiotherapy research,” *Medical Physics*, vol. 30, no. 5, pp. 979–985, 2003.

[56] S. Lan, X. Liu, L. Wang, and C. Cui, “A Visually Guided Framework for Lung Segmentation and Visualization in Chest CT Images,” *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 3, pp. 485–493, 2018.

[57] A. Géron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, 2017. Chapter “End-to-End Machine Learning Project.”.

[58] A. Mansoor, U. Bagci, Z. Xu, B. Foster, K. N. Olivier, J. M. Elinoff, A. F. Suffredini, J. K. Udupa, and D. J. Mollura, “A Generic Approach to Pathological Lung Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2293–2310, 2014.

[59] A. Mansoor, U. Bagci, B. Foster, Z. Xu, G. Z. Papadakis, L. R. Folio, J. K. Udupa, and D. J. Mollura, “Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends,” *RadioGraphics*, vol. 35, no. 4, pp. 1056–1076, 2015.