VRIJE UNIVERSITEIT BRUSSEL

UNIVERSITEIT GENT

Automatic segmentation of bone metastases using deep learning

Marina Souto Pastor

Promotor:
Prof. Dr. Ir. Jef Vandemeulebroucke
Supervisor:
Ir. Jakub Ceranka

Master thesis submitted in order to be awarded the Master's Degree in
Master of Science in Biomedical Engineering

Academic year 2018-2019

Abstract

Bone metastases are a common complication in some high incidence types of cancer, like prostate or breast cancer. The complications associated with bone metastases include bone pain, fractures and spinal cord compression. Most part of bone metastases are irreversible and treatments are focused on slowing the growth of the lesions. In the United States, 17% of the total direct medical cost was employed treating bone metastases. In order to improve the health of the patients and cut down medical costs, early detection is crucial. Some studies have shown that Whole-Body MRI has the potential to become the best method for diagnosis but there are still some difficulties left. One patient can have multiple bone metastases all over the skeleton in different sizes. This makes diagnosing bone metastases a tough task for the radiologists and because of the irregular shapes of the bone metastases, changes in size are also difficult to measure. The goal of this project is to provide an automatic tool for the segmentation of bone metastasis, making it easier for the clinicians to diagnose and to control the size of the present metastases. Using different modalities of MRI (T1 and B1000) and different patch sizes (16x16x16 and 32x32x32) a convolutional neural network (U-Net) was trained. The segmentations predicted by each U-Net employing one modality and size, were later combined into one final segmentation. The best results achieved with this approach are the following: a correct detection of 37 bone metastases out a total of 100 with 67 false positives using k fold cross-validation and a dataset of 6 different patients with multiple acquisitions making a total of 100 lesions.

Contents

List of Figures

List of Tables

List of Abbreviations

ADC	Apparent diffusion coefficient
BS	Bone Scintigraphy
CNN	Convolutional neural network
CT	Computed Tomography
DWI	Diffusion-weighted imaging
FN	False negative
FP	False positive
MRI	Magnectic Resonance Imaging
PET	Positron Emission Tomography
RF	Radio Frequency
TE	Time to echo
TN	True negative
TP	True positive
TR	Repetition Time
WB-MRI	Whole-Body Magnetic Resonance Imaging

Chapter 1

Introduction

1.1 Clinical relevance of bone metastases

The skeleton is the third most common location of metastases in the body and bone metastases are a common complication in breast and prostate cancer. Although other types of cancer such as thyroid cancer, lung cancer, bladder cancer or multiple myeloma can also increase the risk of having bone metastases [1]. Breast cancer and prostate cancer are the main cause of 80% of the cases of metastasis bone disease and at the same time are among the types a cancer with most incidence in 2018 worldwide [2]. Table 1.1 shows the incidence of the four most common types of cancer worldwide in the year 2018.

Some of the complications associated with bone metastases include bone pain, fractures and spinal cord compression [3]. In 2007 the American Cancer Society estimated the cost burden for bone metastasis at $12.6 billion, which corresponds to the 17% of the $74 billion total direct medical cost estimated by the National Institutes of Health in the United States of America [4]. Therefore, due to its high incidence and cost, strategies to reduce the development and complications of bone metastases are clinically relevant and could potentially help a huge number of patients [5].

Cancer Site	No. of new cases (% of all sites)	No of deaths (% of all sites)
Lung	2,093,876 (11.6)	1,761,007 (18.4)
Breast	2,088,849 (11.6)	626,679 (6.6)
Prostate	1,276,106 (7.1)	358,989 (3.8)
Colon	1,096,601 (6.1)	551,269 (5.8)

Table 1.1: New cases and deaths for 36 Cancers and All Cancers Combined in 2018 based on the GLOBOCAN 2018. Image adapted from [2].

1.2 Biological background of bone metastases

The human body is made up of cells, all of our organs and tissues are made of cells. Cells have a predetermined behaviour transmitted through genetic information (DNA). This behaviour regulates reproduction, death and particular cell function. Cells have an organization between them that allows the formation of the complex structures of the human body, but sometimes abnormal cells do not follow this behaviour and grew and reproduce without necessity and without following the patterns inside the tissue. This groups of abnormal cells are called a tumour. Tumours can be either benign or malignant. Benign tumours are the ones that only push healthy tissue aside and do not have the ability to invade other healthy tissue. This kind of tumour is not considered cancer. Malignant tumours are the kind that has the ability to invade healthy tissue and spread to the body, and the kind of tumour that it is considered cancer [6].

There are two main types of tumours: primary and secondary. Primary cancers or tumours refer to cancer in its original location. Most types of primary cancers are named for the part of the body from where they started or for the type of cell. Such as breast cancer, prostate cancer or small cell lung cancer. When cancer cells migrate away from the original place and form other tumours away from the original one, those new tumours are called secondary tumours. The malignant cells travel to distant parts of the body through the circulatory system or the lymphatic system and form secondary tumours or metastases in places such as the bone, brain or lungs. The process of cancer cells migrating to other parts of the human body away from the primary cancer site is called metastasis [6].

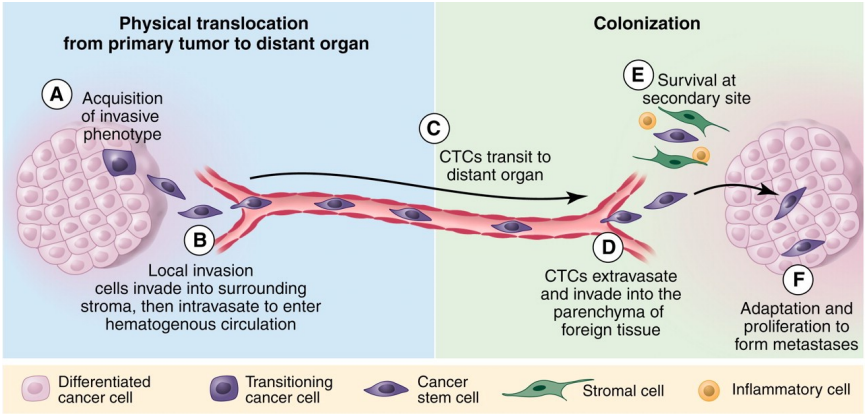


Figure 1.1: The metastatic cascade. Outline of the processes involved in the formation of a metastasis. Image adapted from [6].

Metastasis is a process that involves loss of intercellular cohesion, cell migration, angiogenesis, access to the systemic circulation, survival in circulation, evasion of local immune responses, and growth at distant organs. Metastasis typically involves the following process:

1. Cancer cells invade normal tissue nearby, then move through the walls of nearby lymph or blood vessels and begin circulating through the lymphatic system and bloodstream to reach other parts of the body. After stopping in small blood vessels at a further location, they invade the blood vessel walls and migrate into surrounding tissue where they multiply and form smaller tumours. Those new tumours need a blood supply for continued growth, so they stimulate the growth of new blood vessels.

2. Once they have reached the bone, cancer cells must avoid attacks from the body's immune system. So, they may go through more changes. This means the new tumour may be somewhat different from the primary tumour. This can make it more difficult to treat [6].

In most cases, our immune system will destroy the majority of cancer cells circulating throughout the blood or lymph. Besides, cancer cells which do not find a place to grow and reproduce will die quickly. Cancerous cells need a favourable environment in order to grow or metastasize. While cancer cells can metastasize in many areas of the body, some places are more common than others. Bone is the third most frequent site of metastases, after lung and liver.

Bone metastases are classified as osteolytic, osteoblastic or mixed, according to the primary mechanism of interference with normal bone remodelling:

- Osteolytic: characterized by destruction of normal bone, present in multiple myeloma (MM), renal cell carcinoma, melanoma, non-small cell lung cancer, non-Hodgkin lymphoma, thyroid cancer or Langerhans-cell histiocytosis. The great majority of breast cancer produces osteolytic metastases. This bone destruction is primarily mediated by osteoclasts and not a direct effect of tumour cells. Another way, with lesser importance, is the compression of the vasculature and consequent ischaemia in the late stages of cancer.

- Osteoblastic (or sclerotic): characterized by deposition of new bone, present in prostate cancer, carcinoid, small cell lung cancer, Hodgkin lymphoma or medulloblastoma. The mechanisms of osteoblastic metastases are still poorly understood. In some instances, the new bone formation is not necessarily preceded by bone resorption.

- Mixed: if a patient has both osteolytic and osteoblastic lesions, or if an individual metastasis has both osteolytic and osteoblastic components, present in breast cancer, gastrointestinal cancers and squamous cancers.

Regardless of the type of metastases, they can appear as a focal tumour with a spherical shape and clear limits (focal metastasis) or as a diffuse area of tumour in a bone that has no clear frontier. The number of metastases could oscillate between just one or dozens (multifocal) [3].

1.3 Diagnosis of bone metastases

The symptoms of bone metastases vary depending on which and how many bones are affected. Other health conditions can cause the same symptoms as bone metastases. The most common, and usually first, symptom of bone metastases is a pain in the bone. Bone pain can be intermittent, or it can be constant. Often the pain is worse at night. The pain may be only in one area or it may spread throughout the body. It may be a dull ache or a sharp pain. There may also be swelling along with bone pain.

Other signs and symptoms of bone metastases include:

• Constipation, loss of appetite, nausea, the need to urinate often, extreme thirst and confusion caused by high levels of calcium in the blood (called hypercalcemia).

• Broken bones, or fractures (most often the ribs, vertebrae and long bones of the legs).

• Loss of balance, weakness or numbness in the legs and sometimes arms, and problems with the bladder or intestine caused by pressure on the nerves of the spinal cord (called spinal cord compression) [3].

In order to confirm a diagnosis of bone metastases a imaging tests are the standard procedure.The different imaging methods used for the diagnosis of metastases include X-rays, bone scintigraphy (BS), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). Although the most commonly used in the clinical practice are a combination of bone scintigraphy and computed tomography or magnetic resonance imaging [7].

X-rays

X-rays are a form of electromagnetic ionizing radiation that can be used to take images of the human body in different shades of white and black. The different shade depends on the absorption of radiation, calcium absorbs the most amount of radiation and causes the bones to appear white. Like CT, this technique delivers a small radiation dose to the patient that can accumulate in the body when multiple images are taken.

Using x-rays, clinicians can obtain information on the size and location of the bone tumour, but and also assess the risk of pathological fracture. X-ray allows us to determine whether the tumour is osteolytic, osteoblastic, mixed. Due to its accessibility and popularity, the X-ray examination is quick and cheap, giving it a crucial advantage [7]. On the other hand the diagnostic capability of plain films of X-rays in the skull, spine, and pelvis is very limited by superposition effects. In these areas, the sensitivity of plain films for bone metastases is only in the range of 44– 50% [8].Furthermore, it has a limited field of view so to check different parts multiple images would be required.

Figure 1.2: Example of a bone metastasis obtained with plain film X-Rays. Image adapted from [8].

Bone Scintigraphy (BS)

Skeletal scintigraphy or bone scintigraphy is a type of nuclear medicine procedure that uses small amounts of radioactive material injected in the bloodstream. The radioactive material is usually technetium-99m, which emits gamma rays that are detected by a gamma camera and allows the clinician to evaluate the distribution of active bone formation.

This technique is able to diagnose and assess the severity of a variety of bone diseases and conditions, including fractures, infection, and cancer. This method of imaging has the highest sensitivity (95%) regarding metastases, and it is the basis of screening. Although observing an abnormality of bone formation could also be related to leukaemia or bone fracture causing a low specificity in this type of imaging procedure [7] and has a low resolution.

Computed tomography (CT)

The imaging technique computed tomography (CT) refers to a computerized x-ray imaging procedure in which a narrow beam of x-rays is aimed at a patient and quickly rotated around the body, producing signals that are processed by the machine's computer to generate cross-sectional images of the body. These slices are called tomographic images and can be digitally "stacked" together to form a three-dimensional image of the patient.

CT scans provide enough information to determine the spatial structure and volume of the bone metastases which is really useful in preoperative planning. CT is highly sensitive for bone lesions involving cortical bone (the external part of the bone more dense and stiff), but less so for tumors restricted to the marrow region of the bones (found in the internal cancellous bone also known as spongy). The lesions on the marrow regions must be very extensive to be detectable in CT. As a result, this technique is not the most suitable to use as a screening test for bone metastases, despite its high specificity [8]. As X-rays, CT administer the patient radiation but in the case of CT the dose is higher.

Positron emission tomography (PET)

PET is a nuclear medicine imaging technique that produces a three-dimensional image of functional processes in the body. The image is obtained when the system detects pairs of gamma rays emitted indirectly through an annihilation reaction by a positron-emitting radionuclide which has been injected into the body through a biologically active molecule as a carrier.

The biologically active molecule most commonly used for PET is 2-deoxy-2-18F-fluoro-β-D glucose (18F-FDG), an analogue of glucose, which is used for early detection of tumours and assessment of response to cancer therapy. The concentration of tracer accumulation provides information about tissue metabolism, which is known to be increased in cancer cells compared with non cancerous cells. The disadvantage of PET is the low resolution of the images and some studies suggest that the diagnosis value of the technique is tumour specific [9]. By combining PET and CT, high-resolution images can be obtained. Their diagnostic value is comparable to that obtained with the use of MRI whole-body mode [7].

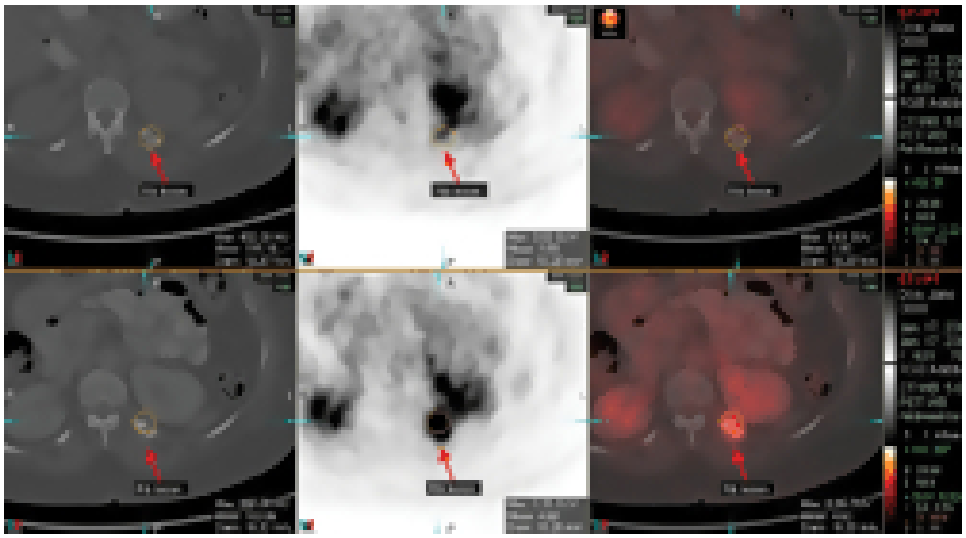


Figure 1.3: Example of a bone metastasis obtained combining CT and PET . Image adapted from [10].

Magnetic resonance imaging (MRI)

MRI is a non-invasive imaging technology that produces three-dimensional anatomical and functional images using powerful magnets.The values displayed on MRI images are based on the magnetization properties of the water molecules within the human body. First, the protons of the water molecules within the tissues, which normally are randomly oriented, are aligned using a uniform magnetic field. Secondly, the alignment is disrupted with an external Radio Frequency (RF) pulse. After the pulse, those molecules loose the alignment and go back to a resting random orientation. In order to reach that resting point, the molecules emit radio frequency energy. That radio frequency energy emitted is measured and displayed in the final image using Fourier transformations. Fourier transformation allow the conversion of the radio frequency energy emitted by the molecules in a exact location on the body to an intensity level (in gray scale) on a voxel in the final image [11]. The time it takes for the protons to realign with the magnetic field, as well as the amount of energy released, changes depending the tissue composition.

In order to obtain the different images some parameter can be modified, such as the Repetition Time (TR), the amount of time between successive pulse sequences applied, or Time to Echo (TE), which is the time between the delivery of the RF pulse and the time when the energy emitted is measured (the echo signal) [11].

The diagnostic of bone metastases is based on the fact that cancer cells contain more water molecules than healthy bone marrow or cortical bone. Diagnostic images determine the degree and type of damage to the bone and also give the opportunity to assess the soft tissues adjacent to the tumour. During whole-body MRI examination there is a greater chance of diagnosing metastases, than with the use of scintigraphy. Nevertheless, this method can be less available due to equipment limitations [7]. Some articles suggest that WB-MRI can replace other imaging techniques such as the combination of BS/TXR and CT for a single-step diagnosis [12].

The most used modalities for the diagnosis of bone metastases are T1 (Figure 1.4 image B) which provides more anatomical information and diffusion weighted images (Figure 1.4 image C) which provide functional information. DW MRI is based on the detection of random (Brownian) motion of water over very small distances, and allows a better diagnostic of tumours due to the increase amount of water in cancerous cells [13].

The advantages of the WB-MRI include the ability to quantify tumour load and response to therapy in soft-tissue metastases and bone. Although a WB-MRI examination takes 30- to 45-min, it does not require any contrast injection and/or irradiation, which is a significant advantage considering modern radioprotection concerns [12]. The cumulative irradiation of BS, X-Rays, and CT generates a dose effectively representing more that than several years of natural irradiation at each staging procedure [12]. Since multiple sequences can be performed in any patient examination, only MR can be considered a truly multifunctional single (non-hybrid) evaluation method for evaluating bony metastatic disease [14].

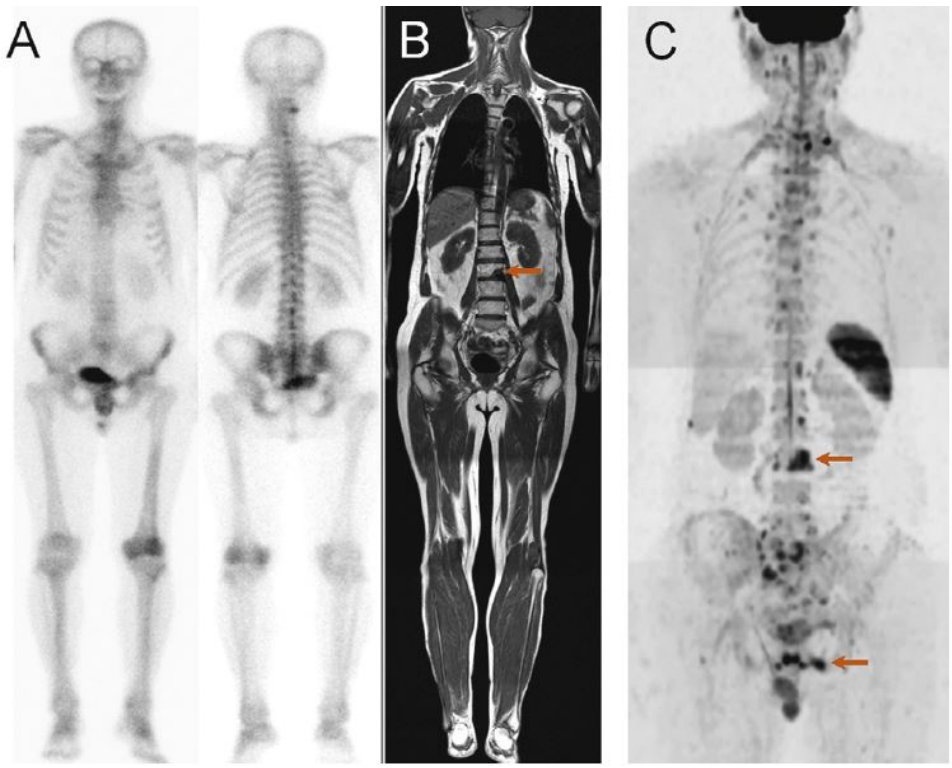


Figure 1.4: Whole-body magnetic resonance imaging (MRI) versus false-negative bone scintigraphy (BS) for bone metastasis detection in a 65-yr-old patient with newly diagnosed prostate cancer. (A) BS (anterior-posterior and posterior-anterior views) shows no significant lesion. (B) Coronal T1 and (C) diffusion-weighted MRI images of the whole body confirm bone metastases within L3 and the left iliac bone (arrows).Image adapted from [12].

1.4 Current Challenges

Nowadays, the life expectancy after a diagnosis of bone metastasis is very low. The median survival from diagnosis of bone metastases is between 12-53 months for prostate cancer and between 19-25 for breast cancer patients [15]. Treatment decisions depend on several parameters, for example, if the bone disease is localized or widespread, if there is evidence of extraskeletal metastases, the kind of cancer and his features, prior treatment history and disease response, the symptoms and the general state of health. Treatments can often shrink or slow the growth of bone metastases and can help with the symptoms they are causing but, they are not curative. Some of the most common treatments are: bisphosphonate drugs, external radiation therapy, ablation or surgery [3].

The current golden standard is screening the patients more prone to develop these metastases and treat them to control the size and the possible complications. Clinical evaluation of the progression of tumours is usually done following the RECIST criteria, Response Evaluation Criteria in Solid Tumor. This criteria measures the longest diameter of the tumour as a way to asses the growing of shrinking of it. RECIST 1.1 designates numerous lesions as unmeasurables including small tumors and bone metastases without soft tissue masses measuring 10 mm (the large majority of bone metastases) [16].

Bone metastases differ in size and can be difficult to measure using only the longest diameter due to its irregular shapes .A lot of research has been conducted trying to obtain an automatic tool that could help the clinicians to measure the volume of a lesion in a easy way that could be reproducible. Due to the lack of an uniform criteria to measure Bone metastases is difficult to asses the efficacy of experimental treatments in clinical investigations for developing new treatments and also to decide if the treatment for a patient is working as expected or needs another treatment.

On the other hand, bone metastases only appear in the late stages of some types of cancer and can be difficult to diagnosis because they can differ in size and location all over the skeleton. Finding all the lesions in a patient can be time consuming and a labour intensive task for a radiologist that could be easily missing some lesions.

1.5 Goal of the thesis

The goal of this thesis is to propose an fully automatic tool for the segmentation of bone metastases from multi-modal whole-body MRI. The tool will use WB-MRI images due to its high diagnosis value [12, 14] without irradiation or a contrast. Although some years ago this technology was not very available, in the present day most hospitals have access to it and the benefits of this diagnostic technique have been explained above.

The advantages of this automatic tool versus traditional methods will be the automatic volume measurements in a reproducible way and assurance that all the lesions are detected, both key steps in automatic treatment response assessment. This work is focused on the development of a software that is a diagnostic tool and a treatment assessment tool (response evaluation criteria) for bone metastases at the same time. The ultimate purpose of this work is to provide an useful tool for the clinicians saving them time diagnosing and decreasing the doctors work load significantly because of the extensive work that is checking the whole skeleton of a patient. Along with equipping them with an uniform method for evaluating the growth or shrinking of the lesions that could allow for automatic cohort comparison studies.

Chapter 2

State-of-the-art

2.1 Automatic methods for detection

Before the development of automatic tools for segmentation of the volume of the tumours, most studies were focused on the detection of the lesions. One of the first fully automatic automatic tools was created in 2004 by Yin et al. They developed a computer-aided diagnosis (CAD) system in order to assist radiologists in the diagnosis of bone metastasis in bone scintigraphy images. The system provided warning marks and abnormal scores on some locations of the images to direct radiologists' attention toward these locations as shown in Figure 2.1 [17].

The detection method is based on fuzzy system called characteristic-point-based fuzzy inference system (CPFIS) is employed to implement the diagnosis system and three minimization are used to systematically train the CPFIS. Asymmetry and brightness were chosen as the two inputs to the algorithm based on radiologists' knowledge [17].

The sensitivity achieved by this system is 91.5% (227 of 248 lesions detected) and the mean number of False Positives was 37.3 per image. The objective of this project was provide an effective second-reader information to radiologists in the diagnosis of bone scintigraphy. Although this system can be useful for diagnosis purposes, Bone scintigraphy is known for its poor sensibility [7] and other, more modern, diagnosis techniques provide better results.

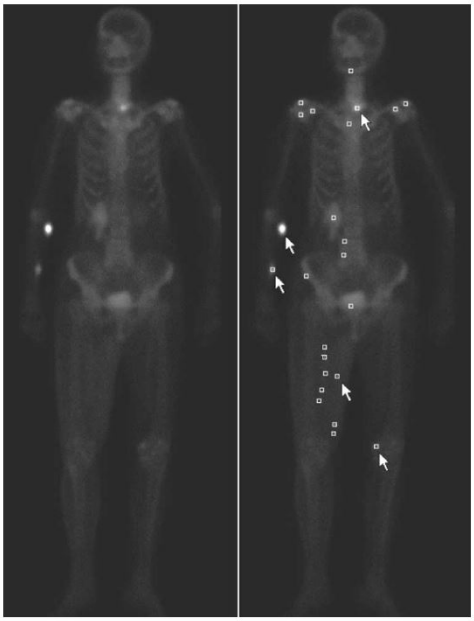


Figure 2.1: The left is the original image, and the right is its CAD. Five of these marks at the arrow-pointed locations are true abnormalities. The other 19 marks are False Positives. Image adapted from [17].

In 2012, Wels et al. developed a fully automatic algorithm for osteolytic spinal bone lesion detection from 3D CT data. The method proposed is a multi-stage approach subsequently applying multiple discriminative models, i.e., multiple random forests, for lesion candidate detection and rejection to an input volume. In their study, it is detected the center of the lesion and a cube is defined surrounding the lesion to approximate its spatial extension.

This method achieved a cross-validated sensitivity score of 75% and a mean false positive rate of 3.0 per volume on a data collection consisting of 34 patients with 105 osteolytic spinal bone lesions. The median sensitivity score is 86% at 2.0 false positives per volume [18].

Some of the limitations of this study include that fact that only detects osteolytic lesions on the spine, so osteoblastic or sclerotic lesions will not be detected apart from all the tumours that are outside the spine. Furthermore, computed tomography is a diagnosis method that uses ionizing radiation, which makes it dangerous to use as a follow up method of the progression of the illness because of the amount of radiation received by the patient.

Regarding the bounding box method for assessing the extension of the tumour, as seen in Figure 2.2, the measures taken are not very precise. Mainly because the shape of the lesions is quite irregular and a cube makes a poor approximation.

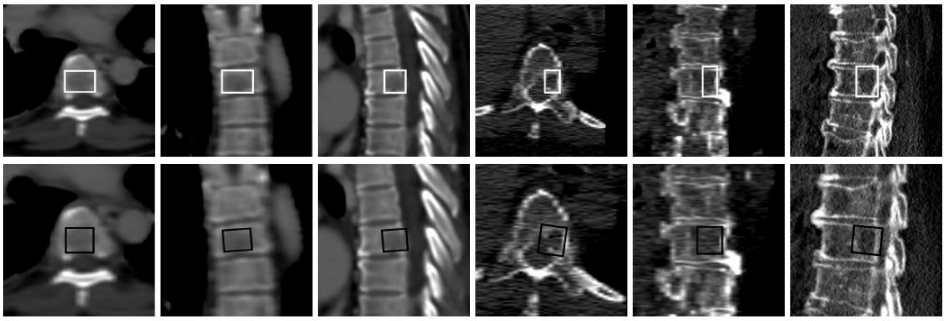


Figure 2.2: Detection results in axial (a, d), coronal (b, e), and sagittal views (c, f) of the automatic method developed by Wels et al . The first row shows the ground-truth annotations in white. The second row shows the detection results in black.Image adapted from [18].

Although the detection tools are useful for helping the radiologist find all the lesions, the other problem associated with bone metastases is the difficulty of measuring changes in the size of the tumours. That problem causes poor results in clinical investigations when assessing the effectiveness of a new treatment and poor decision making when deciding the treatment of the patient after a follow up.

2.2 Automatic methods for segmentation

One project with a similar goal to this thesis was proposed in 2013, when Burns et al. designed a similar tool for the automatic segmentation of sclerotic metastases of the thoracolumbar spine on computed tomography (CT) images. A watershed algorithm was used as a lesion detection method and later a Super Vector Machine (SVM) classifier detected which 3D regions are tumours and which not. The results obtained for images in the testing set, sensitivity was 79%, with an False Positive Rate of 10.9 per patient. The limitations of this study were the are of detection, that was limited to the thoracolumbar spine and the use of CT images. Because, as mentioned before, the radiation of CT is dangerous in a tool that is created to manage the changes in size of tumours [19].

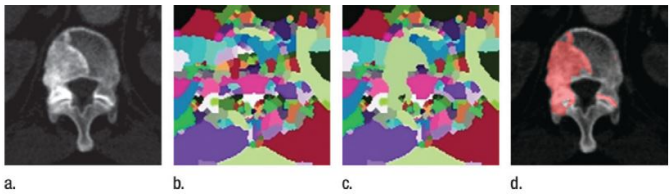


Figure 2.3: Lesion segmentation results by Burns et al.(a) Original axial CT image, with heterogeneous sclerotic lesion (b) Watershed segmentation of axial CT cross section (c) Lesion segmentation after 2D merging (d) Axial CT image with lesion detection in red. Image adapted from [19].

In 2015, Franzle et al. designed a similar tool for the automatic segmentation of multiple myeloma. Multiple myeloma is a systemic illness that grows tumours all over the skeleton. This study was focused on segmenting the focal tumours in the spine using MRI images. Specifically they used T1 and T2 weighted images. This project was focused on segmenting focal lesions in vertebrae, with the bony regions previously delimited manually on CT images to transfer the region to the MRI images [20].

The classification method used in this study was random forest classifier and the result obtained was an accuracy of 95% in a sample of lesions in 8 lumbar vertebrae from 4 multiple myeloma patients [20]. The results of this study can be seen in Figure 2.4.

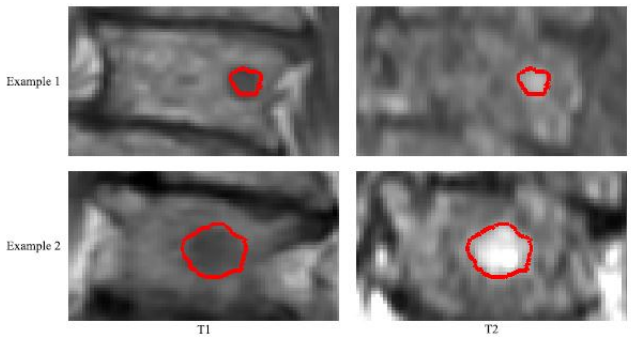


Figure 2.4: Lesion segmentation results by Fränzle et al. Image adapted from [20].

Although this method achieves a good segmentation of the lesions, the results are limited to the vertebrae and those still have to be manually segmented.

In 2014, Blackledge et al. proposed a semi-automatic segmentation of Bone Metastases using a combination whole-body diffusion-weighted MRI (WB- DWI) with anatomic T1-weighted MRI images and a Markov random field (MRF) model. The goal of this study was to develop a tool for asses treatment effectiveness and the volume of the lesions was measures along with the associated global apparent diffusion coefficient (gADC). The ADC map measures the magnitude of diffusion (of water molecules) within tissue and those values are usually higher in tumours. As shown in Figure 2.5, this method works first detecting the areas with a higher probability of being tumours and after a manual segmentation, eliminating all the false positives, a Markov random field is applied. Although this tool can not automatic segment lesions, it has a great applicability for treatment assessment and also proves the good results obtained combining different modalities of MRI [21].

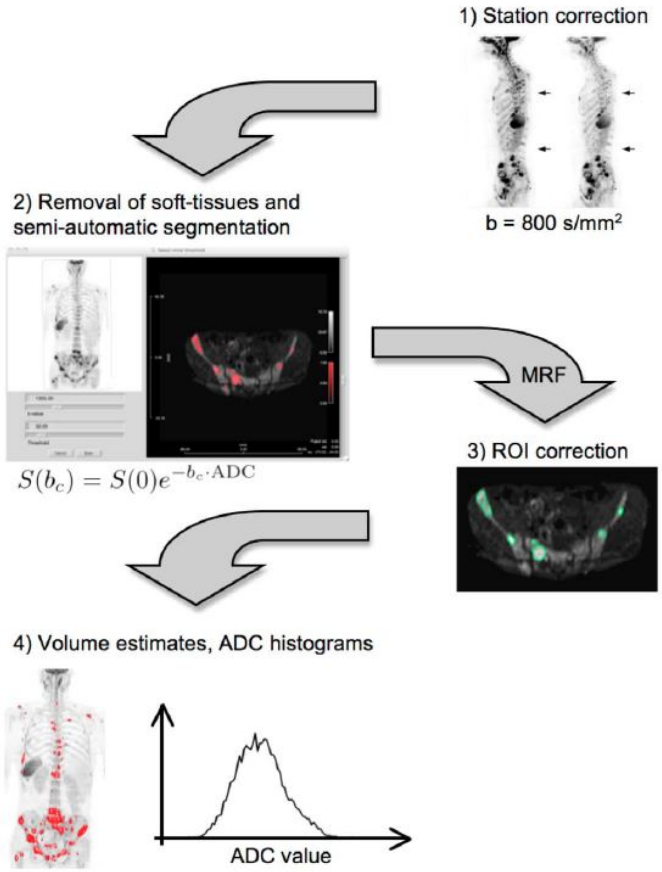


Figure 2.5: Lesion segmentation pipeline and results obtained by Blackledge et al. Image adapted from [21].

2.3 Automatic segmentation using CNN

In recent years, neural networks and deep learning methods have been employed to solve a variety of problems in different fields such as engineering, from computer vision to speech recognition, as well as in the natural sciences, in areas like high energy physics, to chemistry and to biology. Hence it is logical to consider applying them to biomedical images analysis applications [22].

Convolutional neural networks (CNNs) emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s. In the last few years, due to to the increase in computational power, low prices of technological equipment, cloud computation and the amount of available training data; CNNs have achieved exceptional performance on some complex visual tasks. Nowadays, CNN are behind image search services, self-driving cars or automatic video classification systems. Moreover, CNNs are not only useful for visual perception, they can succeed at other tasks, such as voice recognition or natural language processing (NLP) [23].

Wang et al developed an automated spinal metastasis segmentation tool in magnetic resonance imaging (MRI) by using Convolutional Neural Networks. In order to adjust to the large variability in metastatic lesion sizes, they created a Siamese deep neural network approach comprising three identical subnetworks for multi-resolution analysis. At each location of interest, three image patches at three different resolutions but same size are extracted and used as the input to the Siamese network. The Siamese neural network is composed of several identical multilayer convolutional neural networks, each with an input patch of a different resolution. The results show that the proposed approach correctly detects all the spinal metastatic lesions while producing only 0.40 FPs per case. Nevertheless, as all the approaches reviewed, they only detect lesions in the spine and bone metastases can be found in many different parts of the skeleton [22].

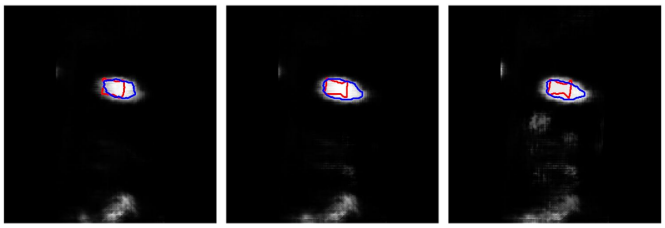


Figure 2.6: Examples of the predicted likelihood maps obtained by Wang et al. The spinal metastasis boundaries provided by the radiologist are marked by red contours, while the boundaries of the detections obtained are marked by blue contours. Image adapted from [22].

In 2018, Chmelik et al developed a similar tool using Convolutional Neural Networks to classify, and automatically segment, osteoblastic and osteolytic spinal bone metastases. The tool uses CT images as an input to a CNN and employs 3D vertebra and spinal cord masks obtained automatically. After the CNN classification a medial axis transform post-processing and a random forest object-wise meta-analysis are applied to eliminate false positives. The results obtained are a high sensitivity for both types of lesions (osteolytic and osteoblastic) of 0.80 and 0.92 respectively with 1.59 and 3.40 False Positives detections per vertebrae [24].

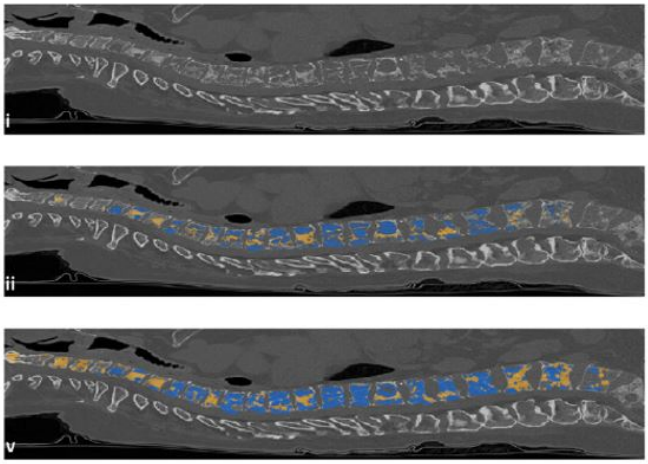


Figure 2.7: Results obtained by Chmelik et al. Image i) is the original CT acquisition, ii) is the segmentation done by a radiologist and v) the final result obtained by the automatic algorithm. Image adapted from [24].

Although this two methods are only focused on the segmentation of lesions on the spine, and ignore possible lesions in the rest of the skeleton, they obtain the segmentation fully automatically using convolutional neural networks with good results.

Chapter 3

Data Pre-Processing

3.1 Experimental Dataset

All the MRI images used in this project came from male patients with Prostate cancer that has metastasize in the skeleton. The experimental dataset is composed of 6 different patients and will be named with the numbers: 24, 27, 33, 34, 36 and 45. Although there are only 6 patients some of them had different acquisitions, meaning that the dataset contains images of the same patient taken in different dates as a follow up during treatment. In order to classify those images patients will be named as Patient-24.2, which refers to patient number 24 the second acquisition. It is important to keep track of the different patients to avoid problems in the separation of a training and a validation set in the neural network. In total the dataset contains 12 WB-MRI images.

Patient	Acquisitions
Patient 24	24.3, 24.4, 24.5, 24.7
Patient 27	27.4
Patient 33	33.1, 33.2
Patient 34	34.1, 34.2
Patient 36	36.1
Patient 45	45.1, 45.2

Table 3.1: Different patients and acquisitions included in the dataset of this project.

MRI images: types and characteristics

The images used in this project are whole body MRI (WB-MRI). WB-MRI employs high spatial resolution images of parts of the body and aligns them together to show a complete view of the whole body. To obtain a WB-MRI image, each part of the patient's body is brought to the magnet isocenter and then the partial images are stitched together using different image processing algorithms. Thanks to recent technological advances such as more powerful and faster gradients, phased array coils and the use of continuously moving-table, WB-MRI images can be obtained in a relatively short time with good resolution. Nowadays this type of imaging equipment can be found in most hospitals and it is common type of examination.

MRI images can be obtained using different sequences that obtain very different images.Taking into account the results obtained combining different MRI modalities for the automatic segmentation of brain tumours [25], in this thesis we will combine different sequences of MRI images. The modalities that have shown more diagnosis relevance, and therefore the ones that will be used are T1 and B1000.

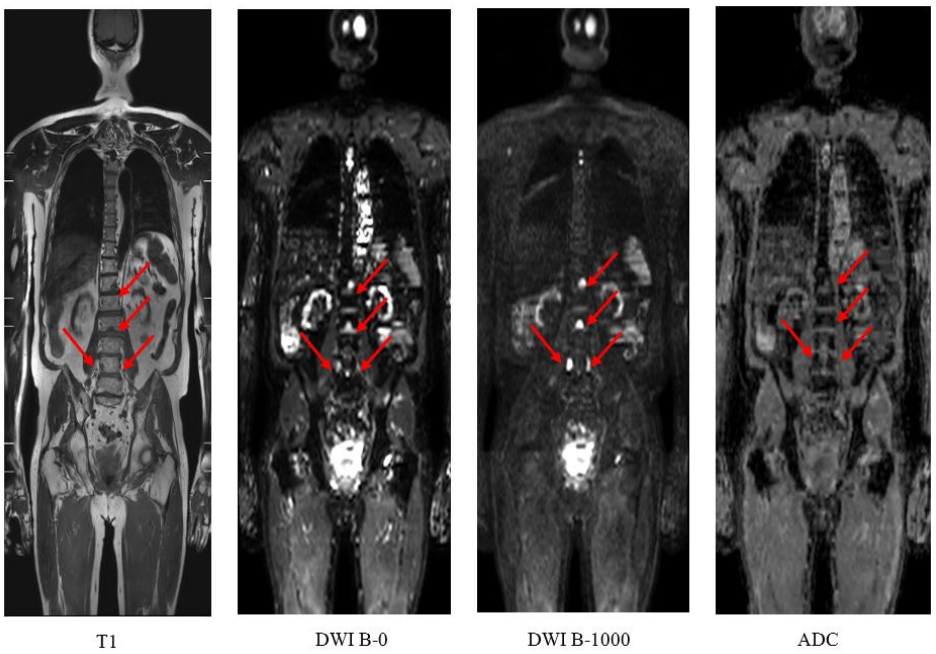


Figure 3.1: Different sequences of WB-MRI with the best diagnostic outcomes. Red arrows pointed to the bone metastases

In the Figure 3.1, the first image from the left is an WB-MRI T1 sequence. T1 are part of almost all MRI protocols because of the anatomical detail they provide. In the T1 images, bone metastases appear as part of the skeleton with a darker colour. T1-weighted images are obtained using short TE and short TR times. The parameters employed for the T1 images in this dataset are the following: repetition time (TR) was 382–546 ms, echo time (TE) 8–20 ms, slice thickness 1.19–4.40 mm, matrix size of 480X480–500x500, pixel spacing 0.65–0.93 mm and field of view (FOV) 480x480–768x768 mm^2 [26].

Diffusion weighted images (DWI) show the random movements of the water molecules in the body (also known as Brownian movement) propelled by the thermal energy of the body [27]. Furthermore, DWI is sensitive to changes in the micro-diffusion of water within the intra and extracellular spaces [11]. In the Figure 3.1, second and third image from the left, it can be seen how the tumours have an hyper intense appearance due to their high water content along with the kidneys and prostate. The b-value identifies the measurement's sensitivity to diffusion. It combines the following physical factors in the equation below into one b-value that is measured in s/mm^2.γ is the gyromanegtic ratio, G equals the amplitude of the two diffusion gradient pulses, δ is the duration of the pulses and Δ is the time between the two pulses [27].

$$b = \gamma^2 G^2 \delta - (\Delta - \delta/3)$$

For most DWI MRI imaging studies, the b values employed for diagnostic purposes vary between 0 and 1000 sec/mm^2. Some studies suggest that using higher b values (1000 sec/mm^2) may improve the diagnostic accuracy of disease detection by increasing suppression of the background signal, thus increasing malignant lesion detection. Higher b values of up to 2000 sec/mm^2 have been applied in specific anatomic sites, besides the use of b values higher than 1000 sec/mm^2 in the breast and prostate has improved disease visualization and detection. However, even though higher (> 1000 sec/mm^2) b value images are clinically desirable, obtaining images with a high b value by direct measurement is challenging because such images have an inherently low signal-to-noise ratio (SNR) and are prone to severe eddy current distortions from the large diffusion-sensitizing gradients used [28].

Diffusion weighted-images in this project were acquired with axial free breathing echo-planar diffusion-weighted sequence with 1 average across 3 directions (x, y, z). Eight scans with two (b0, 1000 s/mm^2), four scans with three (b0, 150, 1000 s/mm^2) and eights scans with four (b0, 50, 150, 1000 s/mm^2) b-values were obtained [26].

Other type of modality employed in diagnosis is the ADC map shown in Figure 3.1. In this type of image the values in grey scale correspond to the Apparent Diffusion Coefficient (ADC). ADC is a measurement of the movement of water molecules within the different tissues and provides a quantitative value of the flow and distance a water molecule has moved. ADC quantifies the combined effects of both diffusion and capillary perfusion. ADC maps are obtained through linear regression between at least two acquisitions of different b values [11]. Although this type of image can be useful diagnosing bone metastases, at the end was not employed in this work. It was decided to implement a more basic approach using T1 and B1000 images and from that point seek further improvement.

MRI images: normalization and resampling

One of the main problems working with any kind if MRI images, is the lack of standardization in the intensity scale. Unlike other image techniques such as CT, that use X-rays and have a Hounsfield Scale that links an attenuation value range with a type of tissue giving physical meaning to the value. For MRI images values, the intensity value has an arbitrary meaning. MRI intensity values do not have a fixed meaning, not even within the same protocol, for the same body region, for images obtained on the same scanner, for the same patient. This absence of standardization protocols is a huge problem for applications such as image segmentation and quantification.

Although there are some strategies for MRI standardization using phantoms to calibrate MR signal characteristics at the time of acquisition, approaches for post-processing standardization are preferred. The intuitive approach of simple scaling of the minimum to maximum intensity range of the given image to a fixed standard range does not provide good results, so a more complicate method is needed. Nyul et al developed a simple method based on transforming the intensity histogram of each given volume image into a “standard” histogram [29]. The comparison of both strategies is shown in Figure 3.2.

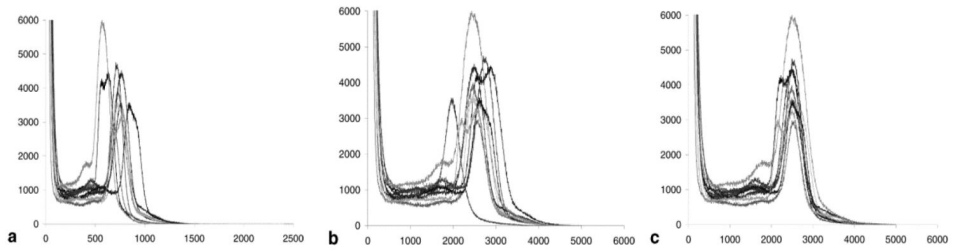


Figure 3.2: Histogram transformation method developed by Nyul et al. Image a) corresponds to the original histograms, b) is the result of a standardization scaling with the maximum and minimum and c) is the result of the method proposed by Nyul et al. Image adapted from [29].

Their method applies nonlinear transformation to images to achieve a significant gain in similarity of the resulting images. The histogram transformation is achieved in two steps, first a training step that is applied only once for a given protocol and body region and secondly, a transformation step that is executed for each given volume image. During the training step, specific landmarks of a standard histogram (for the body region and protocol under consideration) are estimated from a given set of volume images. Later on, in the transformation step, the actual intensity transformation from the intensity scale of the input volume image to the standard scale is computed by mapping the landmarks determined from the histogram of the given volume image to those of the standard histogram [29]. The results of this method can be seen in Figure 3.3.

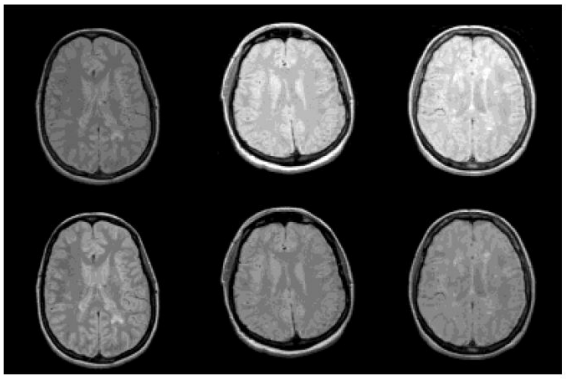


Figure 3.3: The first row of images correspond to the original MRI images and the second row is the result of standardization method proposed by Nyul et al. Image adapted from [29].

Another problem faced in the dataset was the different spacing between MRI images. Spacing is the distance between two consecutive points in an image. Usually in MRI, that distance is the same in the axis x and axis y, but in the axis z is a bit greater. Due to this, voxels in MRI are not square. They have a square base (x and y) but the high (z) is greater. This is a common practice in MRI because in most cases is convenient to have more resolution in an axial cut and have less axial cuts to stack to save time during the acquisition process. When working in resolution dependent problems, like automatic segmentation, the best approach is to resample all images to a fixed spacing. In this case, all images were resampled isotropically to a spacing of (1,1,1) using a linear interpolator.

Binary masks

Apart from MRI images, a binary skeleton mask and a manual segmentation of the bone metastases were used in this project. The skeleton mask was obtained through an automatic atlas-based segmentation with a manual atlas based on the work of Sabrina Verga [30]. The segmentation of the bone metastases was done manually under supervision of a radiologist using the software ITK-SNAP [31] to obtain a binary mask. The Figure 3.4 shows all the modalities and masks employed in this project.

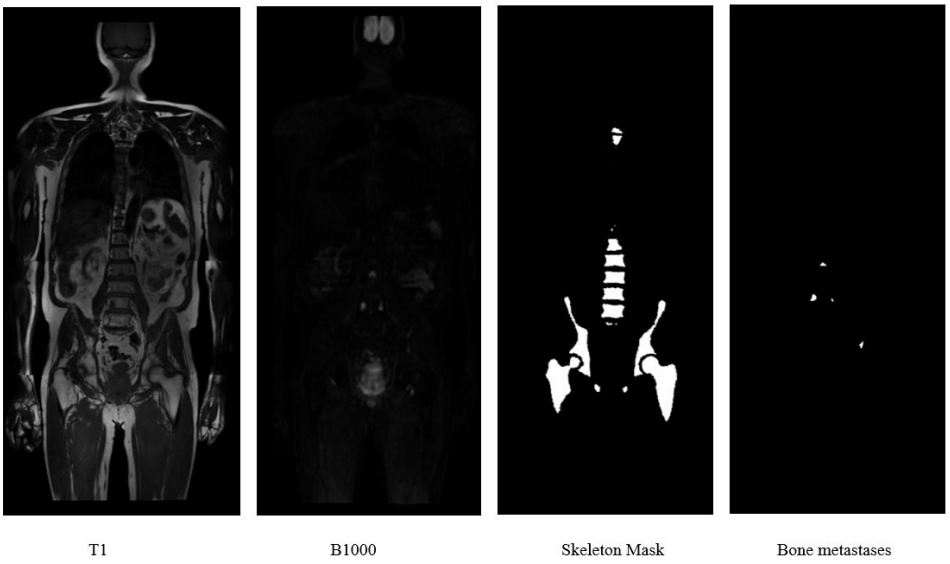


Figure 3.4: Different types of medical images used in this project for the automatic segmentation of bone metastases.

3.2 Patch extraction

Patches were extracted and used as an input to the neural network. In order to incorporate both local and larger contextual information, two sizes of patches were selected. The reason behind using patches instead of the whole image is to focus the filters of the neural network in small pieces of the images and also to estimate a smaller number of parameters. The patches were extracted in the sizes 16x16x16 and 32x32x32 mm. These patches were extracted around the centroids of the lesions and only one patch was extracted per lesion. Those centroids of the lesions were calculated from the manual segmentation of bone metastasis. Secondly, the exact patches were extracted in T1 images, B-1000 and the segmentation of the bone metastases for each one of the different dimensions. An example of the patches extracted can be seen in Figure 3.5.

To obtain a balanced dataset random samples of healthy skeleton were also extracted following the same process. Using the segmentation of the skeleton random points were selected all over the skeleton. All those points were further than 50 mm of the centroids of the lesions to ensure no lesions were included (totally or partially) in the healthy samples all the dataset. After the extraction, healthy samples and lesions samples followed the same processes and were shuffled in the dataset.

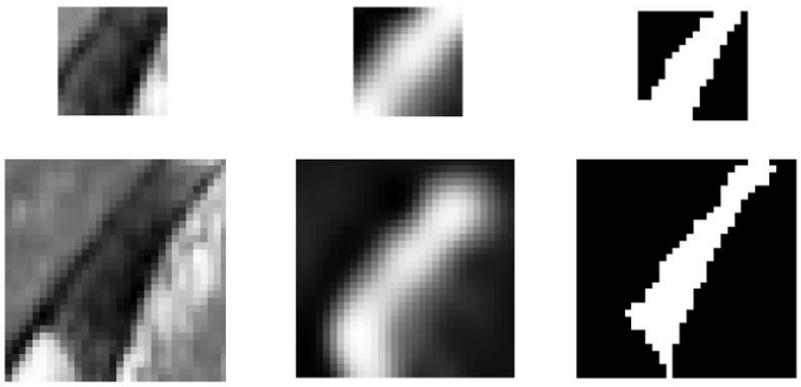


Figure 3.5: Example of patches in different sizes. In the firs row there are the smaller patches (size 16x16x16) and in the second row the bigger patches (size 32x32x32). T1 images are the first column, in the second column B1000 and the last column corresponds to the segmentation of the bone metastases.

The reason behind using different sizes of patches is because larger patches provide higher level features such as the location within a part of the body while smaller patches provide detail and lower lever features. Some studies using CNN for segmentation of lesions in brain with MRI images show a increased performance when combining different sizes in the input [32].

Machine learning methods, such as neural networks, improve the performance with larger datasets. Although the dataset employed in this project contains a total of 110 bone metastases, data augmentations methods were used to artificially augment the samples. First, translation methods were applied moving the centre of the patches a 10% of the measure of the side in all of three axis and combinations. After, a rotation of 90, 180 and 270 degrees were applied. Lastly flipping methods, to flip the images left-right, up-down and transpose. After the data augmentation per one original sample other 431 samples were artificially obtained. The total dataset contains 113616 patches per modality and per size. All the process from begining to end can be seen in Figure 3.6.

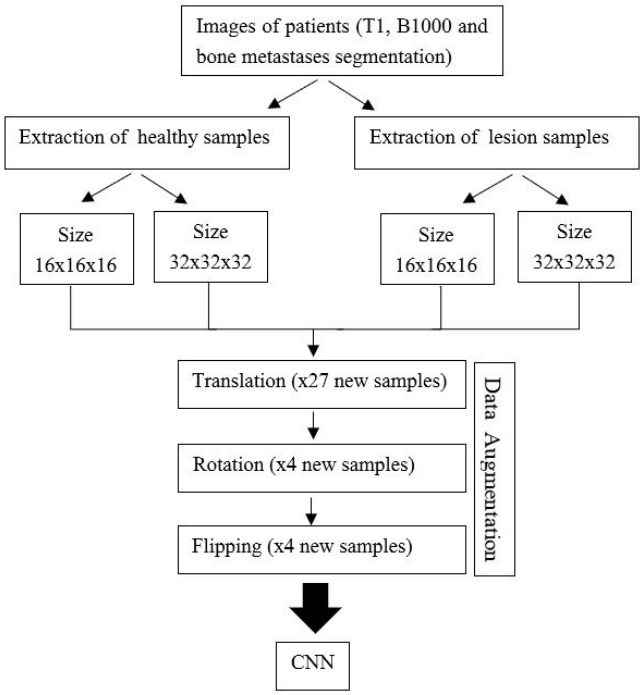


Figure 3.6: Detailed review of all the procedures involved in the extraction of a patch for the CNN.

After the data data augmentation, the images were label according to the data augmentation procedure, if they were a healthy sample or a lesion and the patient number. With the corresponding label, all the patches were stored to be later divided into a training set, validation set and test set for the convolutional neural network.

Chapter 4

Deep learning method

4.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s. In the last few years, CNNs have achieved exceptional performance on some complex visual tasks such as segmentation on medical images [23]. Although there are other types of neural networks architectures, in the last two years, deep convolutional networks have outperformed other state-of-the-art neural networks in many visual recognition tasks [33].

For all neural networks, the most basic unit of computation is the neuron, also called a node or unit. A neuron receives input, which can come from another neuron or from an external source, and computes an output. Each input has an associated weight, which is assigned on the basis of its relative importance to other inputs. The neurons are organized in layers, each neural network must contain at least an input layer, one or more hidden layers and an output layer. The type of hidden layers an the number can differ in each type of neural network.

In the specific case of CNN, it can contain different hidden layers but the essential one is the convolutional layer. As seen in Figure 4.1, neurons in the first convolutional layer are only connected to pixels in their receptive fields and not to every single pixel in the input image (or input layer). As the number of convolutional layers increases, each neuron in the next convolutional layer is connected only to neurons located within a small rectangle in the previous layer. The advantage of this type of neural networks is that it focus on low-level features in the first hidden layer, then assemble them into higher-level features in the next hidden layer, and so on. This type of hierarchical structure is similar to how the brain neurons' process images, which could be one of the reasons why CNNs work so well for image recognition [23].

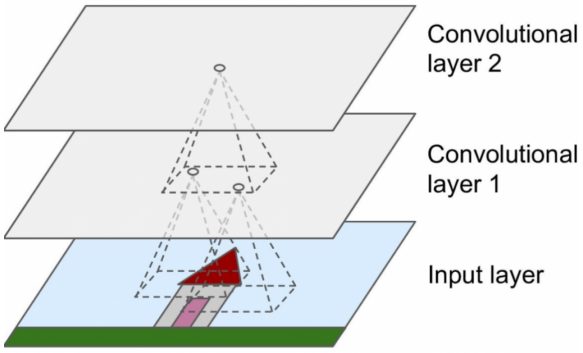


Figure 4.1: CNN layers with rectangular local receptive fields. Image adapted from [23].

The benefit of using convolutional networks for image recognition tasks, instead of regular deep neural networks with fully connected layers, is the reduction in the number of parameters. Fully connected layers can work fine for small images but it will crash for larger inputs due to the huge number of parameters required. For example, a 100×100 image has 10,000 pixels, and if the first layer has just 1,000 neurons (which already severely restricts the amount of information transmitted to the next layer), this means a total of 10 million connections. And that's just the first layer [23]. CNNs solve this problem using partially connected layers as seen in Figure 4.1.

Nevertheless, although convolutional networks have been around for a long time, their success was limited to the size of the applicable training sets and the size and architecture of the convolutional networks proposed. The typical application of convolutional networks is mainly classification tasks on images, in which the output to an image is a single class label. However, in more complex visual tasks, such as applications in medical image processing, the desirable output should include a localization, particularly each pixel should have assigned a specific class label [33].

In 2015, Ronnenberger et al. proposed a new structure of fully convolutional network called U-Net [33]. The main advantage of this architecture is the efficient use of training samples. Although, there is an extended agreement that successful training of deep networks needs thousand of annotated training samples, the training strategy and strong implementation of data augmentation of this works make the CNN outperform the prior best method while working with few samples.The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization.

The network architecture of U-Net is shown in Figure 4.2. It contains first a contracting path (left side) and later an expansive path (right side). The contracting path follows the usual architecture of a CNN. The U-Net is based on the repeated utilization of two 3x3 convolutions (unpadded convolutions), each continued by a rectified linear unit (ReLU) and then a 2x2 max pooling operation using a stride 2 for down-sampling. At each and every down-sampling step, the number of feature channels is doubled. All the steps contained in the expansive path consist of an up-sampling of the feature map followed by a 2x2 convolution (up-convolution) that divides by two the number of feature channels, after that a concatenation with the correspondingly cropped feature map from the contracting path, and lastly two 3x3 convolutions, each of them followed by a ReLU. In the last layer a 1x1 convolution is employed to map each of the 64 component feature vector to the selected number of classes. In total, the U-Net architecture has 23 convolutional layers. In order to enable a stable tiling of the output segmentation map, it is crucial to select an input size in such a way that all 2x2 max-pooling operations are implemented in a layer with an even x and y size [33].

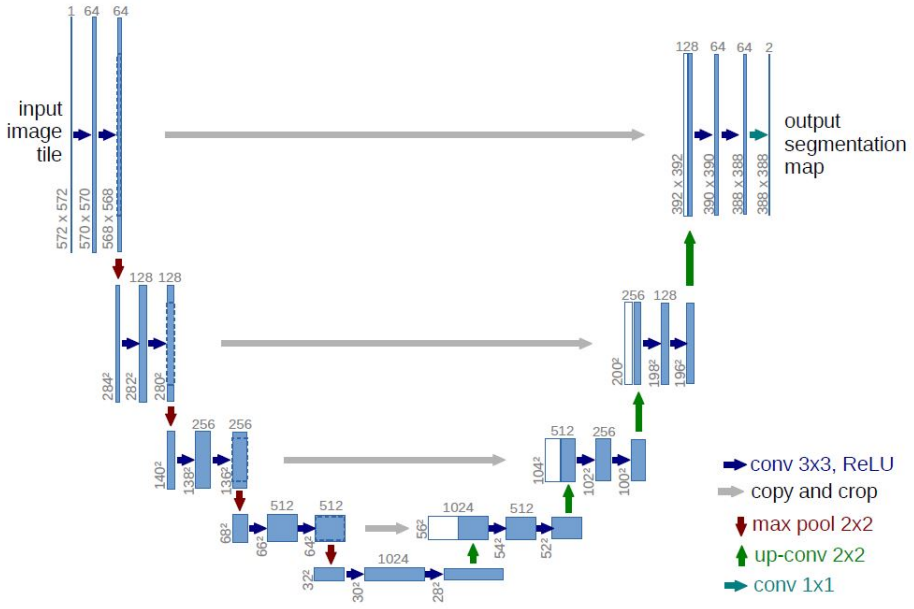


Figure 4.2: U-net architecture. The blue boxes correspond to each the multichannel feature maps. The number of channels is annotated on top of the box. The x-y-size dimensions are provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows indicate the operations. Image adapted from [33].

The results obtained with U-Net in cell segmentation outperformed every other state-of-the-art network proposed for the same application. Using only 35 partially annotated images in the PhC-U373 dataset and 20 partially annotated images in the DIC-HeLa dataset, the U-Net achieved an Intersection over Union (IOU) of 0.9203 and 0.7756 respectively [33]. Some of those results are displayed in Figure4.3.

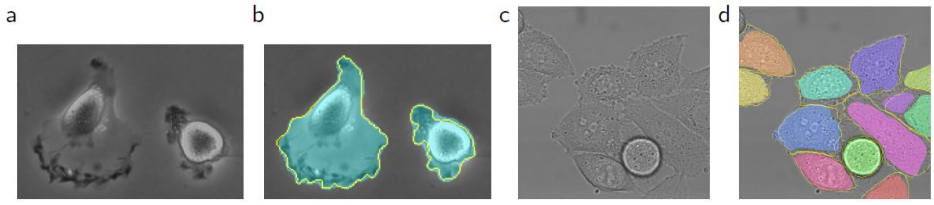
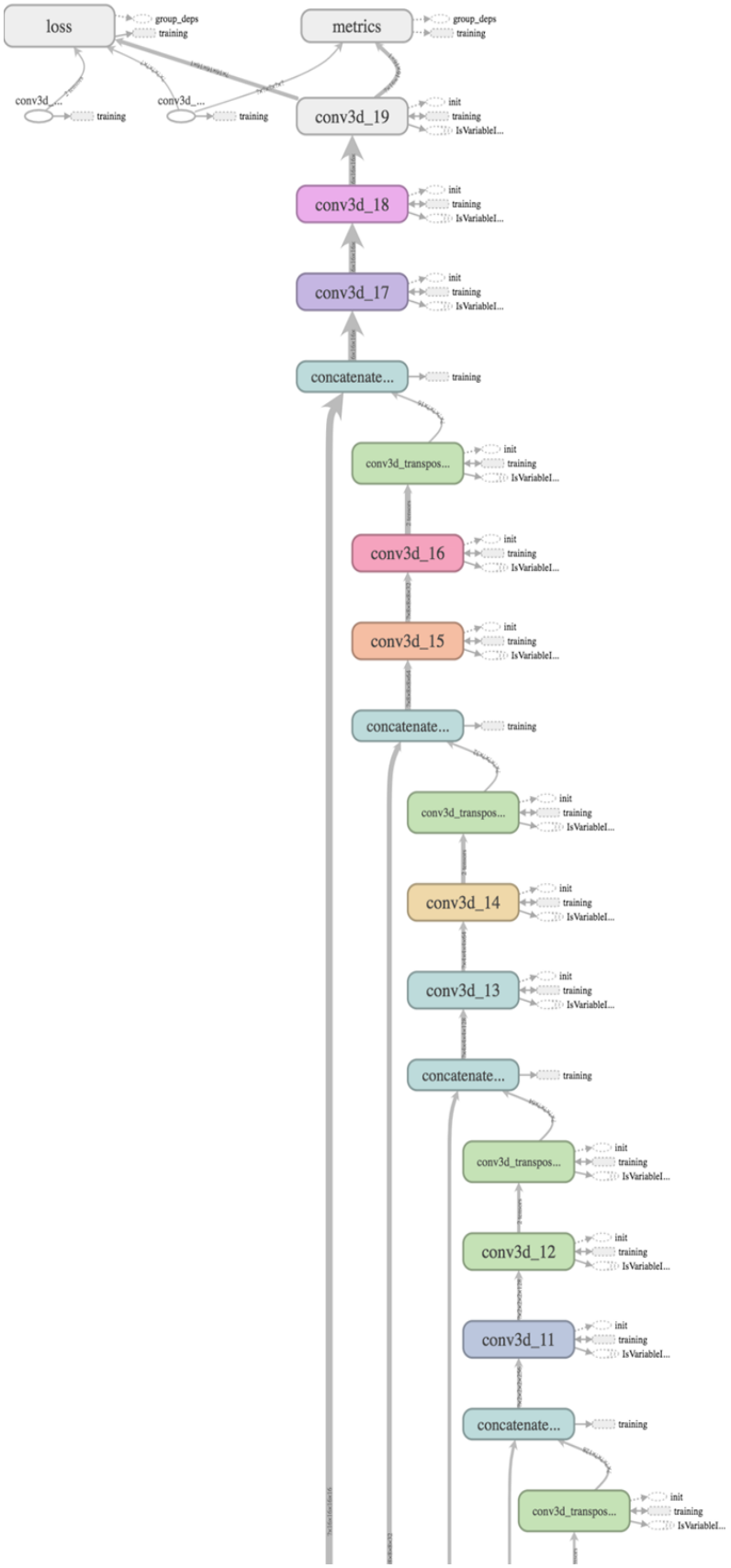


Figure 4.3: Segmentation results obtained with U-net architecture. (a) part of an input image of the PhC-U373 data set. (b) Segmentation result (cyan mask) with manual ground truth (yellow border) (c) input image of the DIC-HeLa data set. (d) Segmentation result (random colored masks) with manual ground truth (yellow border). Image adapted from [33].

4.2 Building a 3D U-Net

Due to the good results obtained by U-Net, a 3-D application of that network was chosen in this thesis. The main difference between the original U-Net and the network created in this work, is that the input image is a volume. So instead of being an 572x572 image as the input, it is a volume of 16x16x16 or 32x32x32 pixels. In order to adapt the network, the convolutions are extended of kernels of size 3x3x3 and the max pooling layer to 2x2x2 kernels. The number of layers and its order is the same as in the original U-Net.

The structure of the 3D U-Net is shown in the Figure 4.4 for an input size of 16x16x16. For the biggest size (32x32x32) the only change required is the modification of the input size. The training, test and validation sets where created taking into account the different patients. To obtain a more representative result, a k-fold cross validation was employed. From the total of 6 patients, 1 was used for test, 1 for validation and 4 for training and all the combinations were calculated. In total 30 models were created, 5 per each test patient.


loss
group_deps
training
metrics
group_deps
training
conv3d_...
training
conv3d_...
training
conv3d_19
init
training
IsVariableI...
conv3d_18
init
training
IsVariableI...
conv3d_17
init
training
IsVariableI...
concatenate...
training
conv3d_transpos...
init
training
IsVariableI...
conv3d_16
init
training
IsVariableI...
conv3d_15
init
training
IsVariableI...
concatenate...
training
conv3d_transpos...
init
training
IsVariableI...
conv3d_14
init
training
IsVariableI...
conv3d_13
init
training
IsVariableI...
concatenate...
training
conv3d_transpos...
init
training
IsVariableI...
conv3d_12
init
training
IsVariableI...
conv3d_11
init
training
IsVariableI...
concatenate...
training
conv3d_transpos...
init
training
IsVariableI...

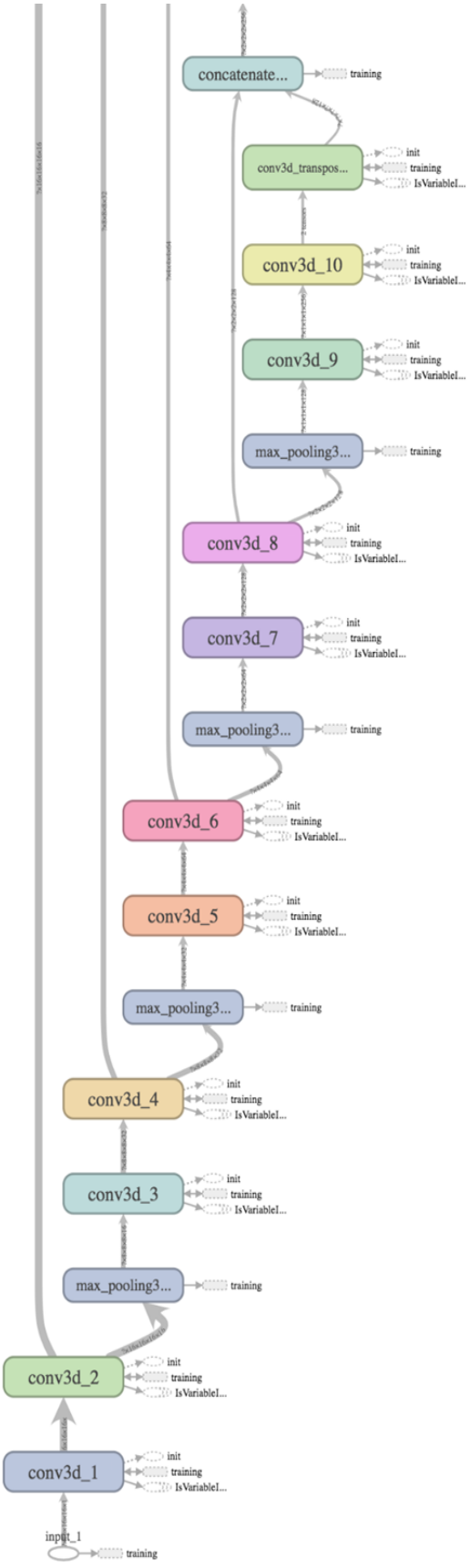


Figure 4.4: Structure of the 3D U-Net adapted for an input of 16x16x16.

There are some works that employ more than one MRI modality as an input to a CNN, like the application for automatic infant brain segmentation developed by Wang et al in 2018 [34]. Nevertheless, when comparing other studies single modalities can achieve as good results as a multi modality approach [25]. For this project, it was decided to make a single network per modality and size. With this approach is easy to combine the results once obtained individually. This approach was chosen because of the simplicity of combining the segmentation masks instead of modifying the neural network every time a new input was chosen to be included. This configuration is shown in Figure 4.5.

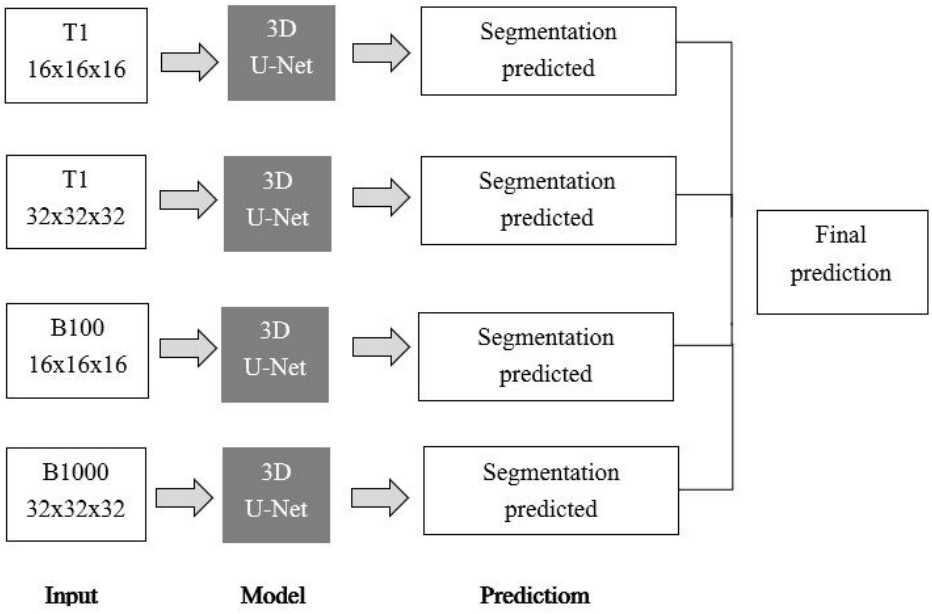


Figure 4.5: Structure of the automatic approach developed.

While training neural networks it is important to adjust parameters such as the number of epochs or iterations, which is the number of times the whole training set will be read, the batch sizes, that regulates the number of training samples used to adjust the weights of the neural network, and the loss and accuracy of the model (training and validation set). Once the model is trained the results will be obtained as patches of the same size as the input with values that correlate to the probability of each voxel of being a bone metastases.

4.3 Prediction of bone metastases

As previously mentioned, the prediction obtained from a neural network is a probability map the exact same size as the input. In this case, cubes of 16 mm and 32 mm of side were obtained from the different U-Nets. The first step towards a final result was to put all those patches together in order to obtain a probability map of the whole body. The approach chosen was to input the model patches of each image (T1 and B1000) with a sliding window that had an overlap of the 50% of the volume. Once the model predicted a result, all those patches were received as probability maps. Taking into account the overlap, the predicted patches were stitched together to form the complete image. The overlap of 50% was chosen because it makes unnecessary to average overlapping voxels, all the image will be overlapped except from the first patches. Because the sliding window takes the full size of the images, it is safe to assume that the first patch will mainly contain the blank space of the margin in a medical image.

When the patches were transformed in a single probability mask the size of a whole body image, the next step was to normalize the probabilities between 0 and 1. So all the values in the probability map were divided by the highest value of them. The segmentation masks are commonly binary masks with values of 1 designated for the bone metastases and values of 0 designated for the healthy tissue. To go from a probability map to a binary mask a threshold was employed. In this work, it was employed a threshold of 0.5 and a threshold of 0.25. The reason behind this low values is because in this kind of problematic, bone metastases, is better to detect all the lesions even though it will increase the number of false positives.

After all the binary segmentation masks of each size and modality were obtained, they were added on top of the another. With the superposition of all the binary masks, the maximum value was where most single masks predicted a bone metastases and the minimum where no mask predicted a bone metastases. The maximum possible value is equal to the number of masks added, because the values of each voxel are added. When the four masks predict a bone metastases in a voxel, the value of the binary mask is one four times. The goal behind adding all the binary masks together is to cut down false positives. All the binary segmentation masks have a huge number of false positives but each modality of image predicts bone metastasis according to the characteristics of the image. Therefore, it is highly likely that the false positives of each type of prediction are in different locations while the bone metastasis remain in the same place. Briefly, the idea of adding modalities is because the overlapping of binary masks is more likely to be a bone metastases.

Ultimately, following the superposition of the masks a final binarization is required. In this work, the values employed started from 1, so all the possible bone metastases of each masks were kept, to the total number of masks, so only when an area was detected as a bone metastases by all the masks was kept.

4.4 Validation of the prediction obtained

As outlined in the introduction of this work, this project aims to be able to detect bone metastases and segment them. For that reason it is evaluated it diagnostic ability through the detection of all the lesions and its value as a treatment assessment tool through the evaluation of the segmentation obtained.

In order to check that all the bone metastases were detected in the predicted segmentation, a confusion matrix was used for the detection results. All the metrics of detection are measured at lesion level without taking into account if all the voxels were correctly labeled. In other terms, it does not matter how well the bone metastases is segmented as long as it is detected partially at least. True positives correspond to the number of bone metastases correctly detected, False positives to the areas detected that were not a metastases and False negatives to the lesions that were not detected. True negatives were not calculated due to the fact that is impossible to calculate the number of no bone metastases inside the body. Sensitivity was also calculated following the next equation:

$$Sensitivity = \frac{True\,Positives}{False\,Positives + False\,Negatives}$$

Due to the inability to calculate true negatives at a lesion level, it was not possible to calculate specificity. But the number of false positives detected could give a good approximation of how good this system is at selecting only the true bone metastases.

In most treatments the doctor has to asses the change of size of the lesions, with an automatic segmentation of the bone metastases this could be done by the computer itself. To check the results obtained, if all the area labeled as bone metastases was predicted, the dice coefficient, volume similarity and Hausdorff distance were employed.

The Dice coefficient (DICE), is the most popular metric in validating medical volume segmentations.The Dice coefficient is a the similarity measure, which compares two binary objects (V_M the manual segmentation and V_P the predicted volume) and it is calculated according to the formula:

$$DICE(V_M, V_P) = \frac{2|V_M \bigcap V_P|}{|V_M| + |V_P|}$$

Other measure for evaluating a segmentation is the volume similarity (VS) that is calculated according to the formula:

$$VS = 1 - \frac{||V_M^1| - |V_P^1||}{|V_M^1| + |V_P^1|}$$

The Hausdorff distance (in mm) is an indicator of the largest segmentation error. This distance was computed from the Euclidean distance map of the ground truth manual segmentation and the surface of the segmentation predicted according to the formula:

$$H(V_M, V_P) = max(h(V_M, V_P), h(V_M, V_P))$$

It is important to take into account that the segmentation predicted is compared to a manual segmentation, therefore some disagreements are possible due to the fact that even different radiologist would segment different the lesion's volume. This could make the predictions obtained a bit more or less accurate between medical specialists because the perimeters can differ.

The results obtained are displayed in the next chapter.

Chapter 5

Results

5.1 Training of the U-Net

While training the neural networks, accuracy and loss for the training and validation sets were obtained and are shown in the Figures 5.1 - 5.6. The loss function is always supposed to decrease as the number of iterations or epochs augment, it means that the neural network is learning and changing parameters to adapt to the input data. When the training loss keeps decreasing and the validation loss is stable, probably the network is overfitting. Overfitting is when the neural network parameters are so closely related to the training set that are unable to predict accurately other samples.

In the case of neural networks, overfitting is a consequence of overtraining or overparametrized. Overtraining happens when an excessive number of iterations causes the system to memorize the data and predict poorly outside the training set. Overparametrized happens when the number of parameters employed by a neural network is so high that they predict perfectly the training set but poorly the validation samples. So it is important to always check if the number of epochs are increasing the efficiency of the neural network.

All the graphs were obtained using Tensorboard (develovep by Keras) and each of the lines displayed in the graphs correspond to the each model with a different test patient and/or a different validation patient. In total for the 6 patients, 30 models were created for each modality and size and displayed in each graph with different colors.

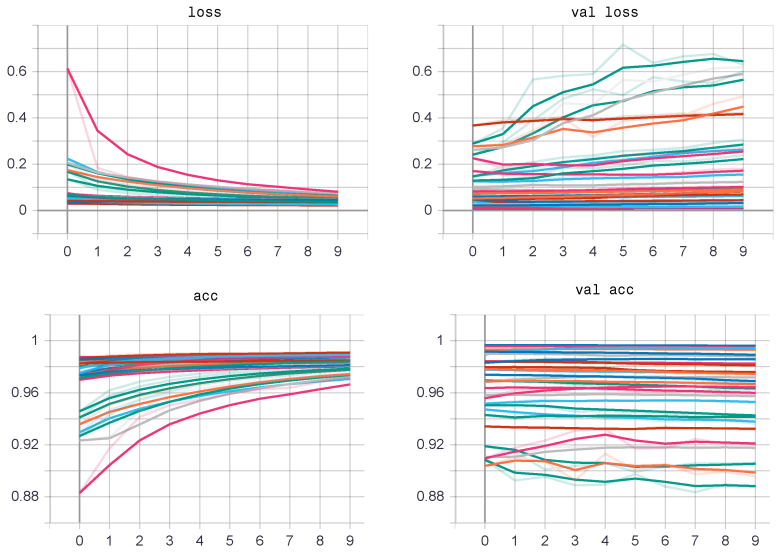


Figure 5.1: Loss and accuracy of training and validation sets for U-Net T1 size 16 and Batch size=100

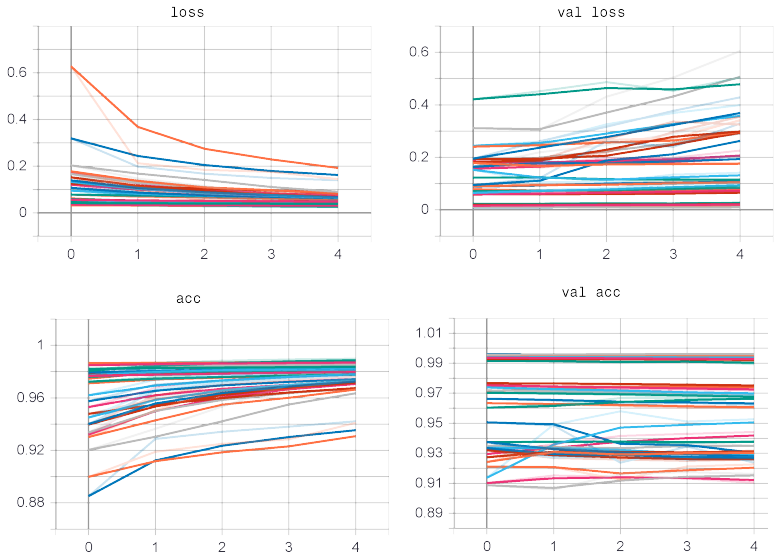


Figure 5.2: Loss and accuracy of training and validation sets for U-Net T1 size 16 and Batch size=1000.

Other important parameter while training is the batch size. In Figures 5.1 and 5.2, it can be seen that increasing the batch size increases the accuracy in both the training set (acc) and the validation set (val_acc) and also reduces the loss slightly in the training set (loss) and in the validation set (loss_val) making the second model slighlty better.

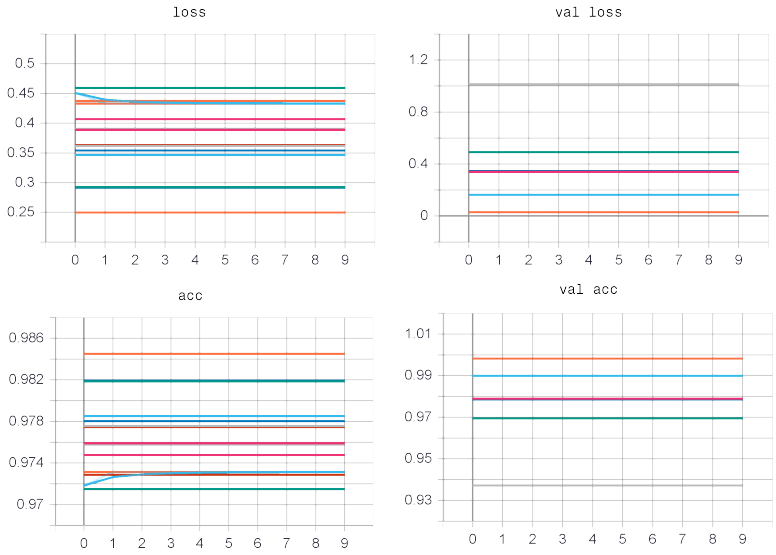


Figure 5.3: Loss and accuracy of training and validation sets for U-Net T1 size 32 and Batch size=100.

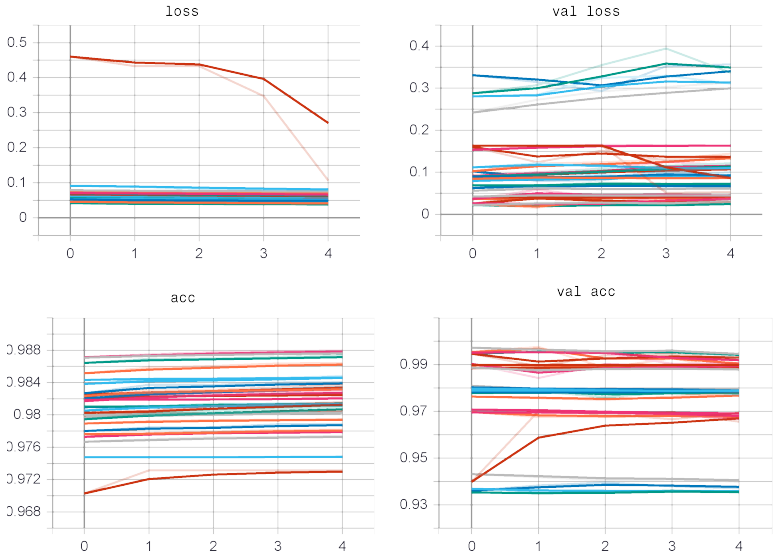


Figure 5.4: Loss and accuracy of training and validation sets for U-Net T1 size 32 and Batch size=300.

In Figures 5.1 and 5.2 the increasing in accuracy and decreasing in loss is bit more noticeable when augmenting the batch size than in the previous case. Other advantage of bigger batches is the reduction in the time needed to train a model.

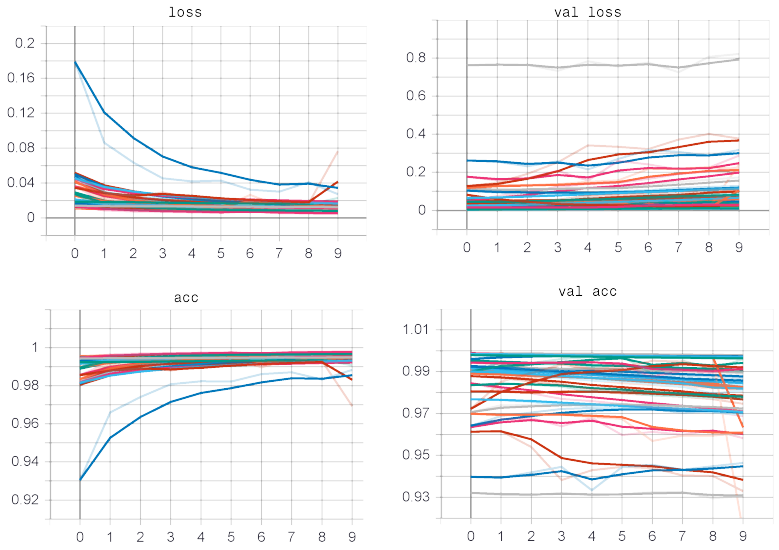


Figure 5.5: Loss and accuracy of training and validation sets for U-Net B1000 size 16 and Batch size=1000.

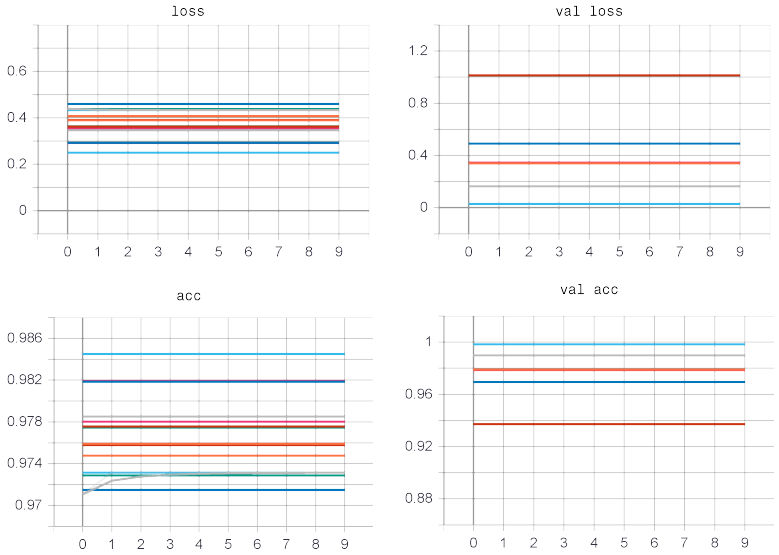


Figure 5.6: Loss and accuracy of training and validation sets for U-Net B1000 size 32 and Batch size=300

In Figures 5.1, 5.3, 5.5 and 5.6, the loss and accuracy show no significant improvement after epoch 5 (in the x axis of all graphs). For this reason it was chosen to predict all results with 5 iterations, which also makes the training process faster.

In general, training networks with that many inputs (113616 samples in total) can take a lot of time. For example, a U-Nets with an input size of 16, batch size of 100 and 10 epochs required around 15 hours of training time. Instead, when using a batch size of 1000 and decreasing the number of epochs to 5, the total time of training went down to 6 hours. In the case of the input size 32, for 10 epochs and batch size of 100, the time needed to train the U-Net was around 30 hours and with a batch size of 300 and 5 epochs it became 10 hours.

The training time corresponds to the model of one modality, so to obtain both T1 and B1000 it should be doubled. Other important factor is that because of the cross validation k-fold, for each model 30 different files were created each one with a different validation and test patient dataset and this increases the time required saving those files. Also the fact that the samples were read from an external drive increased the time as well.

5.2 Results of prediction with single modalities and sizes

The next results were obtained using only the predictions obtained with one modality and one size of input. The table 5.1 is a summary of the detection results of the tables in this section. The bone metastases detected and number of false positive are calculated at lesion levels. In other words, it is only evaluated if the bone metastases is detected regardless of the precision of the segmentation.

Type of input and threshold	No. of bone metastases detected	No. of false positives
T1 16 t=0.5	19	96
T1 16 t=0.25	46	360
T1 32 t=0.25	32	141
B1000 16 t=0.25	44	172
B100 32 t=0.25	0	2
B100 32 t=0.15	0	4

Table 5.1: Comparison of results achieved with single modalities and single sizes.

Patient	Total Lesions	True Positives	False Positives	False Negatives	Sensitivity	Dice
Patient 24.3	1	0	3	1	0.0	0.0
Patient 24.4	4	1	5	3	0.25	0.0
Patient 24.5	4	1	10	3	0.25	0.14197
Patient 24.7	2	1	10	1	0.5	0.0
Patient 27.4	12	6	12	6	0.5	0.0
Patient 33.1	3	2	5	1	0.66666	0.0
Patient 33.2	19	2	11	17	0.10526	0.0
Patient 34.1	9	2	9	7	0.22222	0.0
Patient 34.2	39	3	7	36	0.07692	0.07817
Patient 36.1	3	1	6	2	0.33333	0.0
Patient 45.1	3	0	11	3	0.0	0.0
Patient 45.2	1	0	7	1	0.0	0.0

Table 5.2: Results obtained from the U-Nets of size 16x16x16 with T1 MRI and a threshold for binarization of 0.5.

Patient	Total Lesions	True Positives	False Positives	False Negatives	Sensitivity	Dice
Patient 24.3	1	1	28	0	1	0.0
Patient 24.4	4	2	26	2	0.5	0.0
Patient 24.5	4	3	39	1	0.75	0.0
Patient 24.7	2	1	32	1	0.5	0.0
Patient 27.4	12	10	27	2	0.83333	0.0
Patient 33.1	3	2	16	1	0.66666	0.0
Patient 33.2	19	8	20	11	0.42105	0.0
Patient 34.1	9	3	49	6	0.33333	0.0
Patient 34.2	39	14	39	25	0.35897	0.20412
Patient 36.1	3	2	14	1	0.66666	0.0
Patient 45.1	3	0	43	3	0.0	0.0
Patient 45.2	1	0	27	1	0.0	0.0

Table 5.3: Results obtained from the U-Nets of size 16x16x16 with T1 MRI and a threshold for binarization of 0.25.

Patient	Total Lesions	True Positives	False Positives	False Negatives	Sensitivity	Dice
Patient 24.3	1	0	9	1	0.0	0.0
Patient 24.4	4	3	10	1	0.75	0.09647
Patient 24.5	4	3	15	1	0.75	0.153409
Patient 24.7	2	1	18	1	0.5	0.16847
Patient 27.4	12	7	15	5	0.83333	0.0
Patient 33.1	3	1	7	2	0.33333	0.0
Patient 33.2	19	0	9	19	0.0	0.0
Patient 34.1	9	4	11	5	0.44444	0.0
Patient 34.2	39	12	11	27	0.30769	0.00358
Patient 36.1	3	1	2	2	0.33333	0.034007
Patient 45.1	3	0	17	3	0.0	0.0
Patient 45.2	1	0	17	1	0.0	0.0

Table 5.4: Results obtained from the U-Nets of size 32x32x32 with T1 MRI and a threshold for binarization of 0.25.

Patient	Total Lesions	True Positives	False Positives	False Negatives	Sensitivity	Dice
Patient 24.3	1	1	27	0	1.0	0.10199
Patient 24.4	4	4	14	0	1.0	0.20839
Patient 24.5	4	2	16	2	0.5	0.0
Patient 24.7	2	1	7	1	0.5	0.7357
Patient 27.4	12	11	13	1	0.9166	0.15856
Patient 33.1	3	2	6	1	0.66666	0.30326
Patient 33.2	19	10	18	9	0.52631	0.0
Patient 34.1	9	1	11	8	0.11111	0.0
Patient 34.2	39	7	6	32	0.17948	0.35620
Patient 36.1	3	2	20	1	0.66666	0.434607
Patient 45.1	3	3	24	0	1.0	0.0
Patient 45.2	1	0	10	1	0.0	0.0

Table 5.5: Results obtained from the U-Nets of size 16x16x16 with B1000 MRI and a threshold for binarization of 0.25.

Patient	Total Lesions	True Positives	False Positives	False Negatives	Sensitivity	Dice
Patient 24.3	1	0	0	1	0.0	0.0
Patient 24.4	4	0	2	4	0.0	0.0
Patient 24.5	4	0	0	2	0.0	0.0
Patient 24.7	2	0	0	2	0.0	0.0
Patient 27.4	12	0	0	12	0.0	0.0
Patient 33.1	3	0	0	3	0.0	0.0
Patient 33.2	19	0	0	19	0.0	0.0
Patient 34.1	9	0	0	9	0.0	0.0
Patient 34.2	39	0	0	39	0.0	0.0
Patient 36.1	3	0	0	3	0.0	0.0
Patient 45.1	3	0	0	3	0.0	0.0
Patient 45.2	1	0	0	1	0.0	0.0

Table 5.6: Results obtained from the U-Nets of size 32x32x32 with B1000 MRI and a threshold for binarization of 0.25.

Patient	Total Lesions	True Positives	False Positives	False Negatives	Sensitivity	Dice
Patient 24.3	1	0	0	1	0.0	0.0
Patient 24.4	4	0	4	4	0.0	0.0
Patient 24.5	4	0	0	2	0.0	0.0
Patient 24.7	2	0	0	2	0.0	0.0
Patient 27.4	12	0	0	12	0.0	0.0
Patient 33.1	3	0	0	3	0.0	0.0
Patient 33.2	19	0	0	19	0.0	0.0
Patient 34.1	9	0	0	9	0.0	0.0
Patient 34.2	39	0	0	39	0.0	0.0
Patient 36.1	3	0	0	3	0.0	0.0
Patient 45.1	3	0	0	3	0.0	0.0
Patient 45.2	1	0	0	1	0.0	0.0

Table 5.7: Results obtained from the U-Nets of size 32x32x32 with B1000 MRI and a threshold for binarization of 0.15.

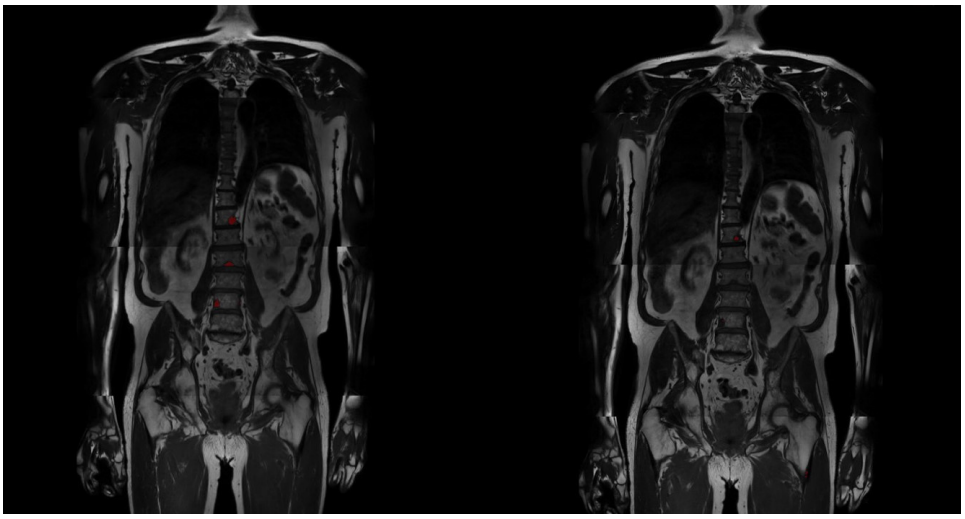


Figure 5.7: Segmentation results in Patient 27 with U-Net T1 input size 16 threshold=0.5. On the left the manual segmentation done under the supervision of a radiologist and on the right the segmentation achieved with the U-Net. The areas segmented as bone metastases are shown in red.

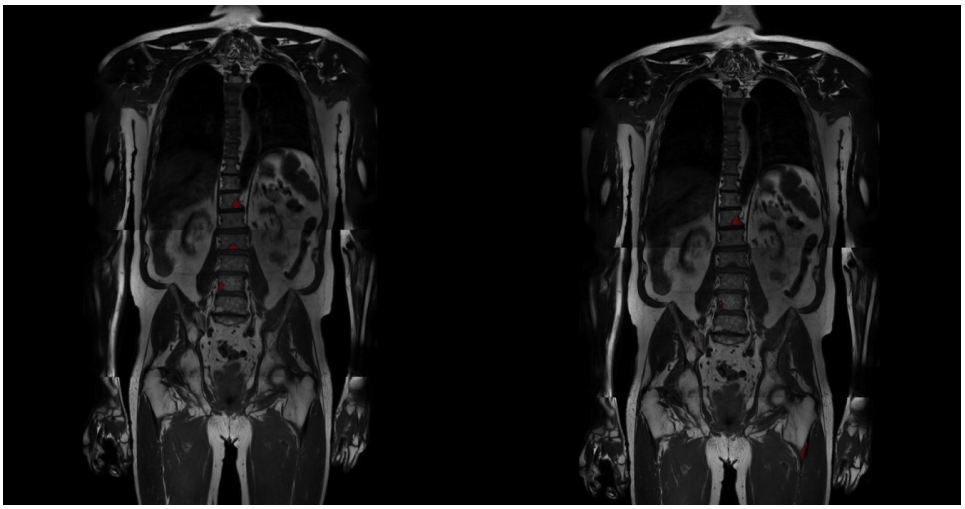


Figure 5.8: Segmentation results in Patient 27 with U-Net T1 input size 16 threshold=0.25. On the left the manual segmentation done under the supervision of a radiologist and on the right the segmentation achieved with the U-Net. The areas segmented as bone metastases are shown in red.

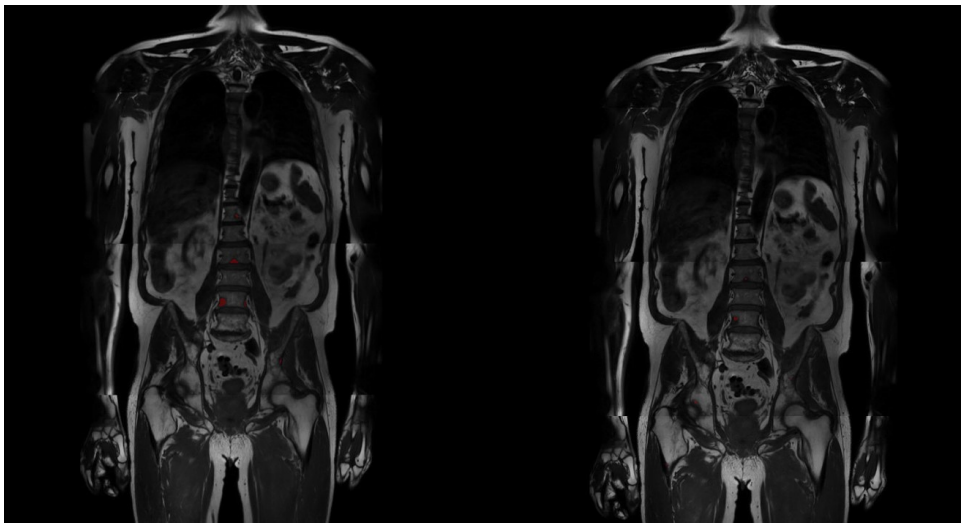


Figure 5.9: Segmentation results in Patient 27 with U-Net T1 input size 32 threshold=0.25. On the left the manual segmentation done under the supervision of a radiologist and on the right the segmentation achieved with the U-Net. The areas segmented as bone metastases are shown in red.

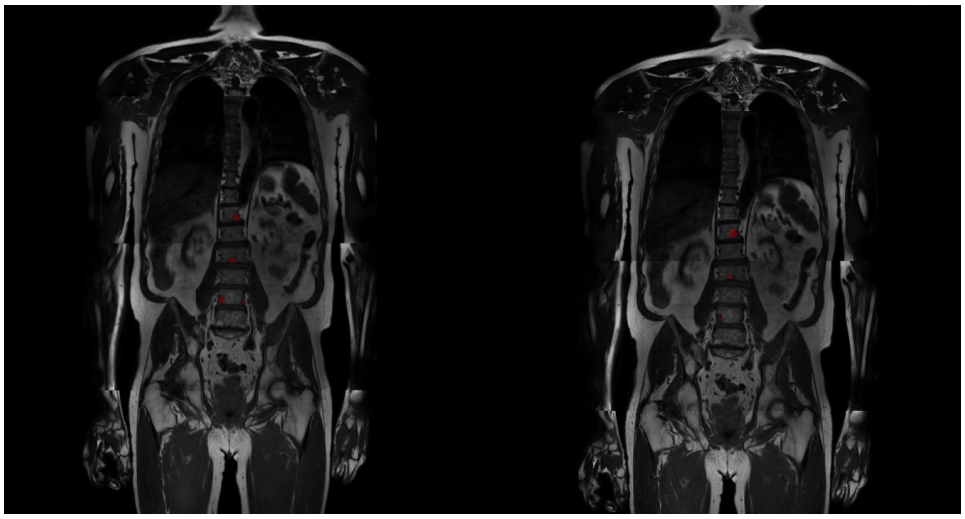


Figure 5.10: Segmentation results in Patient 27 with U-Net B1000 input size 16 threshold=0.25. On the left the manual segmentation done under the supervision of a radiologist and on the right the segmentation achieved with the U-Net. The areas segmented as bone metastases are shown in red.

5.3 Results of prediction with superposition of masks

The results that are shown in tables 5.9, 5.10 and 5.11 are obtained from the combination of the binary masks achieved from U-Nets T1 size 16x16x16, U-Nets T1 size 32 and U-Nets B1000 size 16 all of them with a threshold of 0.25 for the binarization. The binary masks from U-Net B1000 size 32 were not used because of the bad results obtained, it did not achieve any true positive in any of the test patients. So it was decided to be left in the final results.

The reason behind using more than two masks is because the range of thresholds. With three different masks superposed, the values go from a minimum of 0 to a maximum of 3, and in order to achieve a binary segmentation a threshold is needed. When setting a threshold of one, the final mask keeps all the detected areas in all the three binary masks and will include all the false positives found in them. With a threshold of two, the areas kept are the ones that,at least, two different U-Nets detected as bone metastases. Lastly, with a threshold of three the only areas shown in the final segmentation are the overlapping areas of the three binary masks, so the results can only be as good as the worst prediction all the binary masks.

The table 5.8 shows the detection results with the different thresholds. All the results are at lesion level, so the evaluation focuses on if the bone metastases is detected regardless of the quality of the segmentation. The total number of bone metastases is 100 taking into account all the patients. It can be seen that the number of detected bone metastases decreases as the thresholds augments and the opposite issue happens to the number of false positives.

Threshold selected	No. of bone metastases detected	No. of false positives
Threshold=1	64	440
Threshold=2	37	64
Threshold=3	9	2

Table 5.8: Comparison of results achieved with the superposition of masks using different thresholds.

In the following tables, the segmentation was also evaluated with volume similarity and Hausdorff distance, which was measured in mm.

Patient	Total Lesions	True Positive	False Positves	False Negatives	Sensitivity	Dice	Volume Similarity	Hausdorff distance
Patient 24.3	1	1	51	0	1.0	0.07441	-1.81529	560.62911
Patient 24.4	4	4	32	0	1.0	0.16372	-1.54669	347.31253
Patient 24.5	4	3	43	1	0.75	0.0	-1.50495	536.76624
Patient 24.7	2	1	43	1	0.5	0.0	-1.30550	496.23079
Patient 27.4	12	12	35	0	1.0	0.00132	-1.50652	253.95275
Patient 33.1	3	2	23	1	0.66666	0.20426	-1.27452	460.63651
Patient 33.2	19	11	23	8	0.57894	0.0	-1.70478	172.72811
Patient 34.1	9	6	52	3	0.66666	0.0	-1.37088	261.58937
Patient 34.2	39	19	38	20	0.48717	0.34400	-0.364391	137.11673
Patient 36.1	3	2	24	1	0.66666	0.36547	-0.88390	445.04606
Patient 45.1	3	3	45	0	1.0	0.0	-1.97175	470.90338
Patient 45.2	1	0	31	1	0.0	0.0	-1.99343	638.46456

Table 5.9: Results obtained from the combination of masks from T1 size 16, T1 size 32 and B1000 size 16. Threshold=1 for binarization.

Patient	Total Lesions	True Positive	False Positves	False Negatives	Sensitivity	Dice	Volume Similarity	Hausdorff distance
Patient 24.3	1	1	0	0	1.0	0.02305	1.91642	20.83267
Patient 24.4	4	3	5	1	0.75	0.15204	1.0127	184.02445
Patient 24.5	4	3	9	1	0.75	0.0	0.85549	205.27055
Patient 24.7	2	1	5	1	0.5	0.25678	1.09809	309.82899
Patient 27.4	12	9	3	3	0.75	0.18515	0.47347	217.87611
Patient 33.1	3	2	3	1	0.66666	0.40077	0.96256	142.64992
Patient 33.2	19	6	14	13	0.31578	0.0	0.61369	415.54061
Patient 34.1	9	1	4	8	0.11111	0.0	1.42873	242.25193
Patient 34.2	39	9	7	30	0.23076	0.29190	1.25956	155.84929
Patient 36.1	3	2	1	1	0.66666	0.032205	1.83444	194.63556
Patient 45.1	3	0	6	3	0.0	0.0	-0.87906	189.21152
Patient 45.2	1	0	7	1	0.0	0.0	-1.96153	605.32470

Table 5.10: Results obtained from the combination of masks from T1 size 16, T1 size 32 and B1000 size 16. Threshold=2 for binarization.

Patient	Total Lesions	True Positive	False Positves	False Negatives	Sensitivity	Dice	Volume Similarity	Hausdorff distance
Patient 24.3	1	0	0	1	0.0	0	0	0
Patient 24.4	4	1	0	3	0.25	0.0	1.96181	108.89444
Patient 24.5	4	1	0	3	0.25	0.0	1.80322	112.18289
Patient 24.7	2	1	1	1	0.5	0.05543	1.84049	231.06059
Patient 27.4	12	5	0	7	0.41666	0.0	1.95177	153.90607
Patient 33.1	3	0	0	3	0.0	0.0	0.0	0.0
Patient 33.2	19	0	1	19	0.0	0.0	1.99559	494.11323
Patient 34.1	9	0	0	9	0.0	0.0	0.0	0.0
Patient 34.2	39	1	0	38	0.02564	0.03383	1.88	572.54257
Patient 36.1	3	0	0	3	0.0	0.0	0.0	0.0
Patient 45.1	3	0	0	3	0.0	0.0	0.0	0.0
Patient 45.2	1	0	0	1	0.0	0.0	0.0	0.0

Table 5.11: Results obtained from the combination of masks from T1 size 16, T1 size 32 and B1000 size 16. Threshold=3 for binarization.

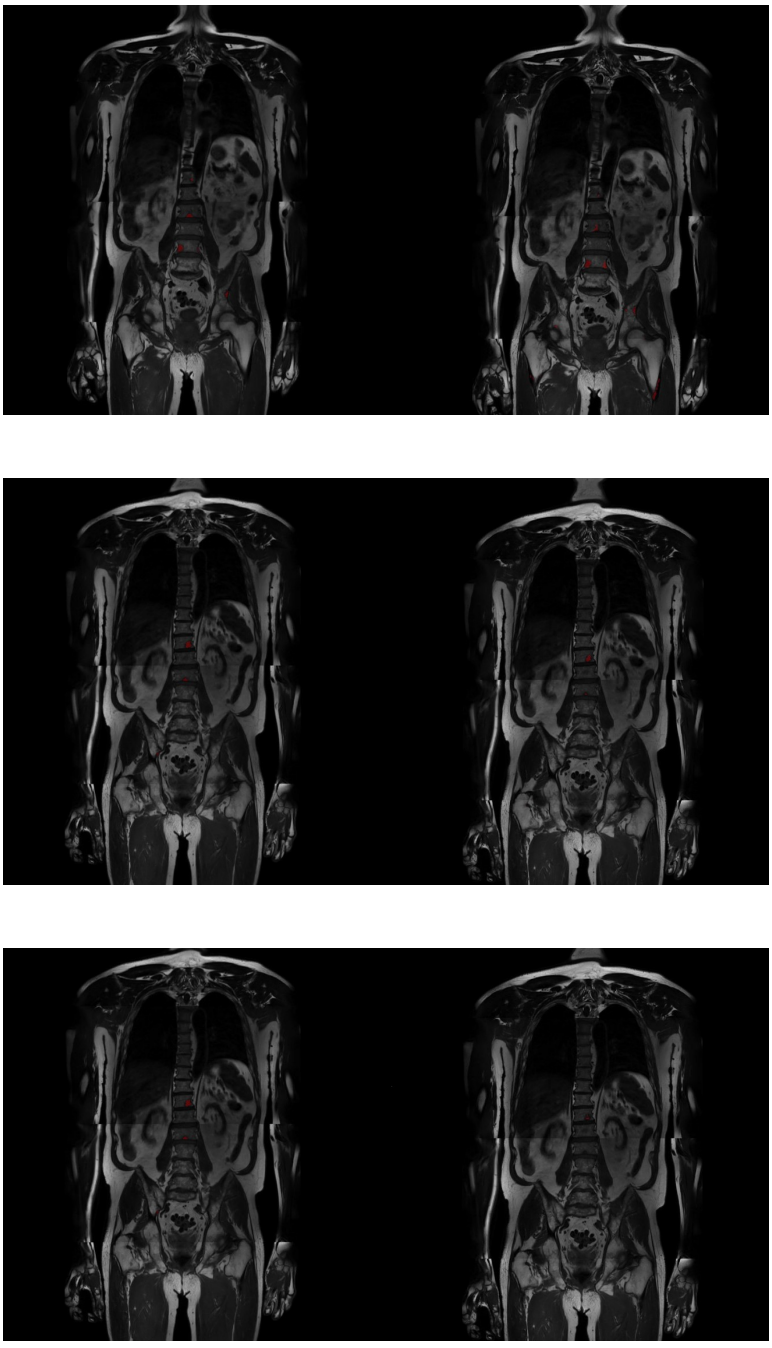


Figure 5.11: Segmentation results in Patient 27 with the combination of different predictions. On the left the manual segmentation and on the right the predicted segmentation. The areas segmented as bone metastases are shown in red. On the first row the results obtained with a threshold=1, in the second row threshold=2 and in the third row threshold=3

Chapter 6

Discussion

Bone metastases have been a problem to treat and diagnose in recent years and although there have been numerous tries to develop an automatic tool, some studies limit the are of detection or rely in CT images that radiate the patients for each acquisition. The aim of this project was to provide the clinicians with a tool that allowed a fully automatic segmentation and detection of bone metastases using MRI images. Although the results achieved with the combination of different U-Nets with different input sizes and types of images were not comparable to the state-of-the-art methods, they are a promising approach. This work proposes a innovative approach to the problem that employs MRI images as an input, which is the only type of radiation and contrast free imaging method for diagnosing bone metastases.

Regarding all the results, it is important to consider the small dataset employed. In this project only 6 different prostate cancer patients were used in the dataset. Even tough data augmentation techniques applied increased drastically the number of samples, it is still a small dataset. Samples obtained through data augmentation do not provide the same amount of information as new samples and therefore, can not achieve the same results. To enhance the performance of this work, a bigger dataset will be needed.

Beyond the small dataset, in this project a k-fold cross validation was used, so the results obtained differ between patients. Some of them reach high levels of sensitivity while others perform poorly. This could be caused by the different quality of the WB-MRI images and also the number of metastases. In the case of patient 34, the total number of bone metastases is around half the total dataset. Which means that when training the model for that patient, only half the dataset is available. This explains way this patient always performs worse than the rest. The number of lesions in a patient affects the number of lesions employed in the training set and the results show that the patients with better results are the ones with less metastases.

6.1 Predictions with single modalities and sizes

In this project, it was decided to start simple and keep adding more complexity to improve the system. First, the predictions obtained with each size of input and type of image were tested. The results obtained with single sizes and modalities show that is possible to predict bone metastases using only one type of image. Although the results are not great, they still reached a detection of almost half the dataset of bone metastases.

Testing the single modalities and sizes, showed which combinations provide the best results. Regarding the size of the input, the best detection was found using the smaller size (16x16x16) that completely outperforms the bigger size of input (32x32x32) with a higher rate of detection. Nevertheless, bigger sizes show a decreased number of false positives. When comparing the results between different types of images, with T1 images more bone metastases are detected but it also provides the highest number of false positives. Also, it is important to notice that the bone metastases detected in each type or image are not the same ones, which suggests that combining modalities could increase the sensibility.

The reason behind keeping the thresholds for binarization so low is because in this type of problem is better to increase the detection even if it causes to augment the number of false positives. It is better to detect more than necessary and rely on the clinician to detect which are the true metastases. In that case, we ensure that the system will show all the bone metastases and it is not leaving out any of them. Lowering more the threshold does not increase the number of bone metastases detected but drastically affects the number of false positives.

Although most predictions reached a certain degree of precision, the U-Net B1000 size 32 was a complete failure. This model did not detect any of the lesions with a threshold of 0.25 or with a even lower threshold of 0.15. When examining the predictions, it was evident that the predictions for most patients were completely empty. Probably this model is overfitted. It could be due to the fact that most part of the images are healthy tissue. Perhaps this model overestimates the probability of healthy tissue and stills gets a high accuracy overall because in most part, the prediction is true. Furthermore, in the case of patches of size 32x32x32 even more percentage of the samples is healthy tissue.

In all the segmentations obtained with one type of image and one type of input, the number of false positives is so high that the dice coefficient is almost 0. This is caused because in all the segmentations, there is more volume wrongly detected as bone metastases than the actually volume of the bone metastases.

6.2 Predictions combining modalities and sizes

In this work it was chosen to add more than two modalities because of the range of thresholds available. When only two masks are superposed, the only two possibilities is to put a threshold of one and keep all the areas detected, keeping all the false positives, or to put a threshold of two and obtain a result as good as the worst prediction between the two masks. The original idea in this work was to superpose the four U-Nets trained. After the bad results obtained with the U-Net trained with B1000 input size 32, it was decided to continue without that predictions because it did not detected any bone metastases in the whole dataset.

After the superposition of the other three predictions: T1 size 16, T1 size 32 and B1000 size 16, only three different thresholds were available. With a threshold of 1 all the areas detected as bone metastases are kept, this increases of bone metastases detected as well as the number of false positives. With a threshold of 3 only the overlapping areas of the three predicitons are kept, this reduces drastically the number of false positives and the number of true positives as well. Lastly, with a threshold of 2 an equilibrium can be kept because it only needs two masks selecting an area to be kept as a bone metastases in the final segmentation.

The thresholds of two reached the best ratio of true positives for each false positive, confirming that approaches that focus on keeping the areas selected by the majority of the masks can improve the performance of this kind of methods. It can be seen in the results how the addition of different modalities and sizes increase the sensitivity of the system as well as reducing the number of false positives. Due to the high number different modalities that can be obtained with MRI, there is a high chance that adding more of them could drastically change the results obtained. So even this work reached some success, the approach of adding more modalities and different sizes has been proven to be effective.

Like in the case of predictions using single sizes and modalities, it is important to remember that patients with a high number of metastases (like patient 34) will be predicted with models that employ smaller training sets. This definitely affects the results achivieded.

During the evaluation of the segmentation all the measurements are very poor (DICE, Volume similarity and Hausdorff distance). This could be caused because when comparing the volume of the WB-MRI to the volume of the area segmented the difference is huge. The predictions have to segment very small zones in a very big volume. So the results have to be really precise to achieve a high score. Usually in this project, more false positives than true positives are detected which affects the score of the segmentation measurements because more segmentation volume is mistakenly labeled than correctly labeled. But, in this case is better to increase the number of false positives than to miss a metastases. An example of the segmentation obtained can be found in Figure 6.1.

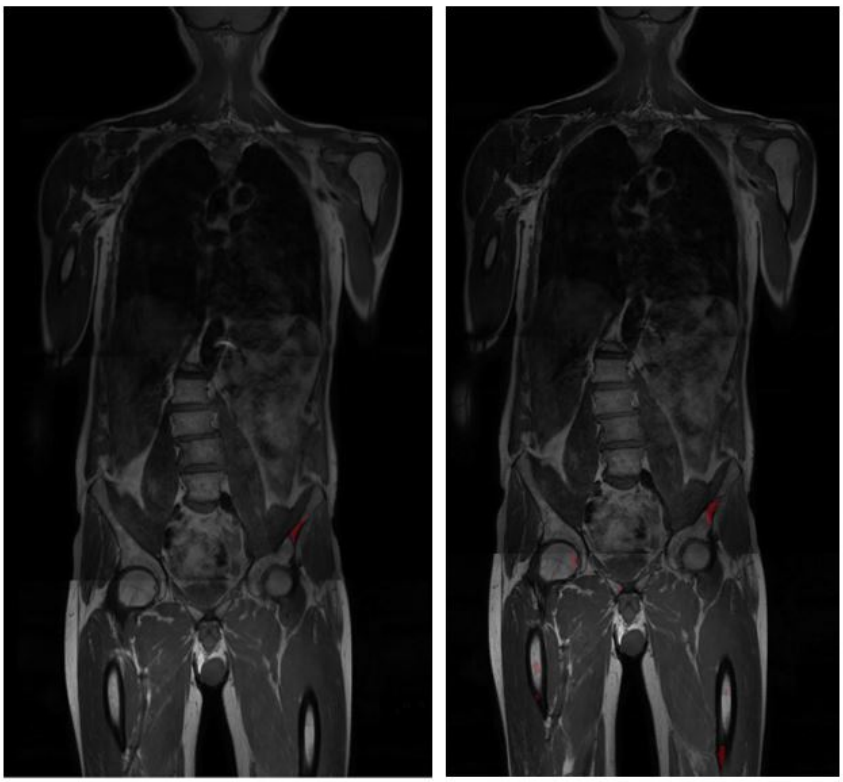


Figure 6.1: Segmentation results in Patient 23.4. On the left the manual segmentation done under the supervision of a radiologist and on the right the segmentation achieved with the combination of masks and a threshold of 2. The areas segmented as bone metastases are shown in red.

6.3 Limitations of the study

One of the biggest limitations of this study is the dataset. For this project it was required WB-MRI images that include T1 and B100 modalities. Although this kind of procedure is increasing its popularity it still is not the gold standard. This caused a limitation in the number of images available and also reduced the number of patients. Other problem with the dataset is the lack of different types of lesions shown. As mentioned in the introduction, there are different types of bone metastases: osteolytic and osteoblastic. Prostate cancer patients mainly show osteoblastic lesions which makes unclear if the tool developed could detect also osteolytic lesions. Even though some studies state that is hard to differentiate between the two kinds in MRI images [35], it is impossible to asses it this automatic approach can achieve good results in their segmentation.

Other limitation that should be considered are the segmentation masks employed (bone metastases mask and skeleton mask). The segmentation mask of the bone metastases was done manually. Due to this, some of the perimeters selected could be inexact and could be variability between the criteria of different radiologist. Because of that, although the segmentation is considered as the ground truth, this could be argued by other medical professionals. Which ultimately could make the results obtained more or less accurate.

Furthermore, the segmentation of the skeleton used only contains the spine, collarbones, pelvis and the upper half of the femur. This causes the automatic tool developed unable to detect lesions in the ribs, sternum or arms. Although those are less likely places to find bone metastases, this problem could be easily overcame using a complete version of the skeleton.

Lastly, there are some computational limits. Initially, the chosen sizes of the input included a bigger patch of 48x48x48mm but the memory capability of the computer employed could not meet the requirements of a CNN with that size as an input. This also caused limitations on the size of the batch chosen to train the different U-Nets. All the sizes of batches chosen were the bigger possible in each case for the computer. Other complication found along the project was the time required to train a neural network. With times always above 10 hours for the training of each U-Net, the number of tries changing parameters was limited in order to be able to finish the project.

6.4 Suggestions for future work

Although the results are promising, there are some steps that can be taken in order to improve the results obtained. First of all, including more modalities could benefit the segmentation masks. For example, T2 images and different DWI could add more information and enhance the segmentation. Furthermore, it could be interesting to add computed diffusion weighted MR images, which are obtained using a mathematical computation technique that accomplish high b-value images from DW MRI performed at lower b values. Some studies have proved that computed DW MR imaging in the body allows higher b-value images to be obtained with a good sound to noise ratio (SNR) and improve signal suppression from normal tissues that may mimic disease. As a result, computed DW MR imaging may improve disease detection [28].

Other approach to enhance the performance of the method could be to extend the neural network to incorporate, into only one network, all the different modalities and sizes of the input. There are some studies that employ two modalities as an input to a U-Net with good results [34] for segmentation problems and the same method could be applied to the segmentation of bone metastasis. Other studies have shown that is possible to incorporate different sizes as the input of deep convolutional neural network and good results are also achieved [32]. So overall, there are some strategies that could be employed to modify the neural network proposed and achieve better results.

Another factor that could improve the results is the dataset employed. In this project only six different patients were used, which is a small dataset when comparing with other projects. Although there were more acquisitions for the same patient, there is not the same variability between different patients and different acquisitions. Apart from that, all the images in this project came from prostate cancer patients. In order to make the tool more useful, images of breast cancer patients should be included because breast cancer patients are also at a high risk of developing bone metastases as the illness progresses.

Lastly, even though, using a U-Net has given good results, there are other options of convolutional neural networks that have proven a good performance in similar applications. One of the alternatives is ResNet, a deep residual network that also employs convolutional layers and that has been successfully applied to the automatic segmentation of organs [36]. Other candidate could be the CNN model of GoogleNet, which automatically obtains the delineation of tumours in histological samples of breast tissue with high accuracy [37].

Chapter 7

Conclusion

Although the results achieved in this work can not compete with the state-of-the-art methods reviewed, it is a promising work. Most methods reviewed in the literature achieve an automatic segmentation from CT images and/or locate the bone metastases only in this spine. This work proves that using MRI images an automatic segmentation could also be achieved. The main advantage of MRI images is that is the only imaging method radiation and contrast free able to diagnose bone metastases. Considering the fact that the segmentation of bone metastases could be used as a tool to control the progression of the illness, it is really important to minimize the dose when regular images will be needed. Besides that, the detection of bone metastases is not limited to the spine in this work.

Regarding the machine learning method applied, the different U-Nets trained have reached a good level of detection, specially the ones with the smaller input size of 16x16x16. But the better results are achieved with the combination of the binary masks of different MRI images and input sizes. The combination of different predictions augmented the number of bone metastases detected while lowered the number of false positives. This proves that this approach (superposition of binary masks) could be further improved adding more images and there are plenty of other modalities of MRI that could improve the results, such as T2, ADC map, higher b values ...

The best results achieved in this work reached a correct detection of 37 bone metastases out a 100 with 67 false positives using k fold cross-validation and a dataset of 6 different patients with multiple acquisitions.

Bibliography

[1] R.E. Coleman. "Metastatic bone disease: clinical features, pathophysiology and treatment strategies". In: *Cancer Treatment Reviews* 27.3 (June 2001), pp. 165–176. ISSN: 03057372. DOI: 10.1053/ctrv.2000.0210. URL: https://linkinghub.elsevier.com/retrieve/pii/S030573720090210X.

[2] Freddie Bray et al. "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". In: *CA CANCER J CLIN* 68 (2018), pp. 394–424. DOI: 10.3322/caac.21492. URL: https://onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21492.

[3] Filipa Macedo et al. "Bone metastases: An overview". In: *Oncology Reviews* 11.1 (2017). ISSN: 19705565. DOI: 10.4081/oncol.2017.321.

[4] Kathy L. Schulman and Joseph Kohles. "Economic burden of metastatic bone disease in the U.S." In: *Cancer* 109.11 (2007), pp. 2334–2342. ISSN: 0008543X. DOI: 10.1002/cncr.22678.

[5] Robert E Coleman. "Clinical Features of Metastatic Bone Disease and Risk of Skeletal Morbidity". In: (2006). DOI: 10.1158/1078-0432.CCR-06-0931. URL: www.aacrjournals.org.

[6] Robert a Weinberg Christine L Chaffer. "A Perspective on Cancer". In: *Nature medicine* 19.2 (2013), pp. 179–92. DOI: 10.1126/science.1203543. arXiv: 0011002 [physics]. URL: http://www.ncbi.nlm.nih.gov/pubmed/23389618.

[7] Bartosz Łukaszewski et al. "Diagnostic methods for detection of bone metastases". In: *Wspolczesna Onkologia* 21.2 (2017), pp. 98–103. ISSN: 14282526. DOI: 10.5114/wo.2017.68617.

[8] Walter Heindel et al. "The Diagnostic Imaging of Bone Metastases". In: (2014). DOI: 10.3238/arztebl.2014.0741. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4239579/pdf/Dtsch%7B%5C_%7DArztebl%7B%5C_%7DInt-111-0741.pdf.

[9] Ning-Bo Liu et al. "Diagnostic value of 18F-FDG PET/CT in comparison to bone scintigraphy, CT and 18F-FDG PET for the detection of bone metastasis." In: *Asian Pacific journal of cancer prevention : APJCP* 14.6 (2013), pp. 3647–52. ISSN: 2476-762X. DOI: 10.7314/apjcp.2013.14.6.3647. URL: http://www.ncbi.nlm.nih.gov/pubmed/23886160.

[10] Iwao Tanaka, Munir Ghesani, and Vamsee Torri. "Bone scan vs. PET/CT in the assessment of skeletal metastatic disease progression". In: *HemOnc Today* (2008). URL: https://www.healio.com/hematology-oncology/breast-cancer/news/print/hemonc-today/%7B%5C%%7D7B5e51f6da-4ef6-4a6d-a5be-6f9429a2822a%7B%5C%%7D7D/bone-scan-vs-petct-in-the-assessment-of-skeletal-metastatic-disease-progression.

[11] Michael A Jacobs et al. "Diffusion Weighted Imaging with ADC Mapping and Spectroscopy in Prostate Cancer". In: (2008). DOI: 10.1097/RMR.0b013e3181aa6b50. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3110834/pdf/nihms118042.pdf.

[12] E Lecouvet et al. "Can Whole-body Magnetic Resonance Imaging with Diffusion-weighted Imaging Replace Tc 99m Bone Scanning and Computed Tomography for Single-step Detection of Metastases in Patients with High-risk Prostate Cancer ?" In: 62 (2012), pp. 68–75. DOI: 10.1016/j.eururo.2012.02.020.

[13] T Takahara, Isehara City, and Kangawa Prefecture. "DWIBS : Diffusion-weighted whole-body imaging with background body signal suppression". In: (2005), pp. 38–41.

[14] Anwar R Padhani et al. "Therapy Monitoring of Skeletal Metastases With Whole-Body Diffusion MRI". In: 1078 (2014), pp. 1049–1078. DOI: 10.1002/jmri.24548.

[15] Christian Rolfo et al. "Molecular target therapy for bone metastasis: starting a new era with denosumab, a RANKL inhibitor". In: *Expert Opinion on Biological Therapy* 14.1 (2013), pp. 15–26. ISSN: 1471-2598. DOI: 10.1517/14712598.2013.843667.

[16] Colleen M Costelloe et al. *Cancer Response Criteria and Bone Metastases: RECIST 1.1, MDA and PERCIST*. Tech. rep. 2010, pp. 80–92. URL: http://www.jcancer.org.

[17] Tang Kai Yin and Nan Tsing Chiu. "A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach". In: *IEEE Transactions on Medical Imaging* 23.5 (2004), pp. 639–654. ISSN: 02780062. DOI: 10.1109/TMI.2004.826355.

[18] M Wels et al. *Multi-Stage Osteolytic Spinal Bone Lesion Detection from CT Data with Internal Sensitivity Control*. Tech. rep. 2012. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.448.3799%7B%5C&%7Drep=rep1%7B%5C&%7Dtype=pdf.

[19] Joseph E Burns et al. "automated Detection of sclerotic Metastases in the Thoracolumbar spine at CT". In: *Original research n Computer AppliCAtions Radiology* 268 (2013). DOI: 10.1148/radiol.13121351. URL: www.rsna.org/rsnarights..

[20] Andrea Fränzle, Jens Hillengass, and Rolf Bendl. "Spinal focal lesion detection in multiple myeloma using multimodal image features". In: *Medical Imaging 2015: Computer-Aided Diagnosis* 9414 (2015), 94143B. DOI: 10.1117/12.2081990.

[21] Matthew D Blackledge et al. "Assessment of Treatment Response by Total Tumor Volume and Global Apparent Diffusion Coefficient Using Diffusion-Weighted MRI in Patients with Metastatic Bone Disease : A Feasibility Study". In: 9.4 (2014), pp. 1–8. DOI: 10.1371/journal.pone.0091779.

[22] Juan Wang et al. "A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks". In: *Computers in Biology and Medicine* 84 (May 2017), pp. 137–146. DOI: 10.1016/j.compbiomed.2017.03.024. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010482517300793.

[23] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn*. 2017. ISBN: 9781491962299.

[24] Jiri Chmelik et al. "Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data". In: *Medical Image Analysis* 49 (Oct. 2018), pp. 76–88. ISSN: 13618415. DOI: 10.1016/j.media.2018.07.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S1361841518305528.

[25] Bjoern H Menze et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". In: 34.10 (2015), pp. 1993–2024. DOI: 10.1109/TMI.2014.2377694.

[26] Jakub Ceranka et al. "Registration strategies for multi-modal whole-body MRI mosaicing". In: *Magnetic Resonance in Medicine* 79.3 (2018), pp. 1684–1695. ISSN: 15222594. DOI: 10.1002/mrm.26787.

[27] Joachim Graessner and Dipl Ing. "Frequently Asked Questions : Diffusion-Weighted Imaging (DWI)". In: (2011), pp. 84–87.

[28] Matthew D Blackledge et al. "Computed Diffusion-weighted Purpose : Methods : Results :" in: *Radiology* 261.2 (2011), pp. 573–81. ISSN: 1527-1315. DOI: 10.1148/radiol.11101919/-/DC1. URL: http://www.ncbi.nlm.nih.gov/pubmed/21852566.

[29] Laszlo G. Nyul and Jayaram K. Udupa. "On Standardizing the MR Image Intensity Scale". In: *Medical Imaging 1999: Image Display* 3658 (1999), p. 595. DOI: 10.1117/12.349472.

[30] Sabrina Verga. "Atlas-based segmentation of skeleton in whole-body MRI". PhD thesis. 2018.

[31] Paul A. Yushkevich, Yang Gao, and Guido Gerig. "ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Aug. 2016, pp. 3342–3345. ISBN: 978-1-4577-0220-4. DOI: 10.1109/EMBC.2016.7591443. URL: http://ieeexplore.ieee.org/document/7591443/.

[32] Konstantinos Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical Image Analysis* 36 (2017), pp. 61–78. ISSN: 13618423. DOI: 10.1016/j.media.2016.10.004. arXiv: arXiv:1603.05959v2.

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351 (2015), pp. 234–241. ISSN: 16113349. DOI: 10.1007/978-3-319-24574-4_28. arXiv: arXiv:1505.04597v1.

[34] Zhengyang Wang et al. *Global Deep Learning Methods for Multimodality Isointense Infant Brain Image Segmentation*. Tech. rep. arXiv: 1812.04103v1. URL: https://arxiv.org/pdf/1812.04103.pdf.

[35] Sarah M Böker et al. "Differentiation of Predominantly Osteoblastic and Osteolytic Spine Metastases by Using". In: 6 (2019), pp. 1–9.

[36] Sihong Chen, Kai Ma, and Yefeng Zheng. "Med3D: Transfer Learning for 3D Medical Image Analysis". In: (2019), pp. 1–12. arXiv: 1904.00625. URL: http://arxiv.org/abs/1904.00625.

[37] Yibao Sun et al. "Detection of Breast Tumour Tissue Regions in Histopathological Images using Convolutional Neural Networks". In: *IEEE 3rd International Conference on Image Processing, Applications and Systems, IPAS 2018* (2019), pp. 98–103. DOI: 10.1109/IPAS.2018.8708869.