FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Learning models for bone segmentation in radiological images

Daniela Filipa Oliveira Faria

Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Tânia Maria Pereira Lopes, PhD

Second Supervisor: Francisco Carvalho Moreira da Silva, MSc

Third Supervisor: Sílvia Costa Dias, MD

October 14, 2022

Abstract

Bone Marrow Edema (BME), and more recently known as Edema Like Marrow Signal Intensity (ELMSI), is an observed change in the Bone Marrow (BM) in Magnetic Resonance Imaging (MRI), described as areas of intermediate to low signal in the T1-weighted sequences and areas of high signal in the fluid-sensitive sequence. Aiding health care professionals in identifying ELMSI more efficiently is the primary motivation for developing a Computer-aided diagnosis system. The first step in the its development is efficiently segmenting the bone in the MRI. For the past few years, occurred an increase in the usage of Deep Learning (DL) architectures in medical imaging problems, especially the usage of Convolutional Neural Networks (CNN) to solve increasingly more complex problems. The dissertation proposes the implementation of DL architectures and later Transfer Learning (TL) techniques to achieve the goal of bone segmentation of the dataset provided by the University Hospital Center of São João. With the data from 72 patients, implementing a U-Net, a slightly modified U-Net and an Attention U-Net and TL techniques, using a pre-trained U-Net on the OAI ZIB dataset and a pre-trained VGG-16 encoder for a U-Net model pre-trained the ImageNet dataset. For T1-weighted images, segmentation results obtained values from 88.38 to 93.69% for DICE Similarity Coefficient and from 78.68% to 88.46% for Intersection over Union coefficient. According to the segmentation results, the DICE Similarity Coefficient coefficient for the fluid-sensitive images ranged from 86.20 to 92.02%, while the Intersection over Union coefficient ranged from 76.51% to 85.67%. The best results were achieved with the Attention U-Net and the VGG-16 encoder U-Net model pre-trained in the ImageNet dataset.

Keywords: Bone Marrow Edema, Edema Like Marrow Signal Intensity, Magnetic Resonance Imaging, Image Segmentation, Machine Learning, Deep Learning, Bone Segmentation, Transfer Learning.

Resumo

O Edema da Medula Óssea é uma alteração observada na Medula Óssea na Ressonância Magnética (RM), descrita como sinal intermédio e hipossinal na sequência ponderada em T1 e como hipersinal na sequência sensível ao líquido. Auxiliar os profissionais de saúde a identificar o Edema da Medula Óssea de forma mais eficiente é a principal motivação para o desenvolvimento de um sistema de diagnóstico auxiliado por computador. O primeiro passo no seu desenvolvimento é a segmentação eficiente do osso na RM. Nos últimos anos, ocorreu um aumento no uso de arquiteturas de Deep Learning (DL) em problemas de imagem médica, especialmente o uso de Redes Neurais Convolucionais (CNN) para resolver problemas cada vez mais complexos. Esta dissertação propõe a implementação de arquiteturas DL e posteriormente técnicas de Transfer Learning (TL) para atingir o objetivo de segmentação óssea do conjunto de dados fornecido pelo Centro Hospitalar Universitário do São João. Com os dados de 72 pacientes, a implementação de uma U-Net, uma U-Net ligeiramente modificada e uma Attention U-Net, acrescida de técnicas de TL, usando uma U-Net pré-treinada no conjunto de dados OAI ZIB e um codificador, pré-treinado no dataset ImageNet, de uma VGG-16 num modelo U-Net. Os resultados da segmentação alcançaram valores para a sequência ponderada de T1 para o coeficiente de similaridade DICE entre 88.38% e 93.69% e para o coeficiente de Cruzamento sobre União entre 78.68% e 88.46%. Os resultados da segmentação alcançaram valores para a sequência sensivel ao líquido no que diz respeito ao coeficiente de similaridade DICE entre 86.20% e 92.02% e ao coeficiente de Cruzamento sobre União entre 76.51% e 85.67%. Os melhores resultados foram alcançados com o modelo Attention U-Net e com o modelo que usa o codificador da VGG-16 numa U-Net, pré-treinado no conjunto de dados ImageNet.

Palavras-chave: Edema da Medula Óssea, Intensidade do Sinal de Edema Semelhante à Medula, por Ressonância Magnética, Segmentação de Imagem, Aprendizagem Computacional, Aprendizagem Profunda, Segmentação Óssea, Conhecimento por Transferência.

Agradecimentos

Começo por agradecer à equipa que me orientou, sempre prestativos, sempre compreensivos e sempre incansáveis na procura do sucesso, Helder P. Oliveira, Tânia Pereira, Francisco Silva, Sílvia Costa Dias e Diogo Costa Carvalho. Não esquecendo, um agradecimento ao Gonçalo Ribeiro que partilhou comigo esta aventura. Um obrigada também a todos os autores de artigos, tutoriais, livros e aulas que ajudam outros através das interpretações dos seus trabalhos e que se apoiam nisso para resolver os próprios desafios.

Naturalmente, o segundo agradecimento, vai para a minha família que sempre me apoiou nos momentos mais difíceis e sempre comigo partilhou os momentos de alegria. Um especial agradecimento para a minha mãe que nunca me faltou, outro especial agradecimento para a minha avó Luísa que sempre me trouxe alegria e motivação para continuar e um obrigado ao Daniel que sempre me ajudou quando precisei. Agradeço ao meu pai que sempre me incentivou no curso e nas minhas escolhas e ao meu avô que me incentivou a ser o melhor possível na minha vida académica.

Para os de 16, que comigo abraçaram a aventura do curso e que com eles partilhei alguns dos melhores momentos da minha vida, assim como dificuldades e resolução de vários problemas durante todos estes anos, um obrigada. No mesmo tom, agradeço aos Besties que comigo evoluíram a associação e evoluíram a nível pessoal. Vivi convosco os momentos mais desafiantes e recompensadores que poderia pedir, enquanto vida académica e não só.

Para terminar, obrigada aos Rufianos, que foram mais que colegas de casa, foram aqueles que vivenciaram todas as alegrias, todos os choros, todo o sofrimento que se vai sentindo na vida académica comigo. São a família emprestada do dia-a-dia. Obrigada pelas noitadas de diversão, pelas noitadas de estudo intensivo e por todas as noitadas de partilha. Obrigada por todos os ombros amigos que encontrei, todo o amor que recebi e espero ter retribuído, todo o carinho que foram mostrando. Obrigada por me proporcionarem a oportunidade de gostar e de ser gostada. Levarei comigo o que é viver com alegria e tirar o sempre o melhor partido da situação em que nos encontramos graças aos ensinamentos que aprendi convosco. Espero que não seja um adeus mas um até já. Obrigada por um dos melhores anos da minha vida.

Daniela Filipa Oliveira Faria

"Do not let anyone rob you of your imagination, your creativity, or your curiosity. It is your place in the world; it is your life. Go on and do all you can with it, and make it the life you want to live."

Mae Jemison

Contents

List of Figures

List of Tables

Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AAM	Active Appearance Model
ASM	Active Shape Model
Adam	Adaptive Moment Estimation
API	Application Programming Interface
ANN	Artificial Neural Networks
AG	Attention Gates
BM	Bone Marrow
BME	Bone Marrow Edema
BMES	Bone Marrow Edema Syndrome
BS	Bone Segmentation
CAD	Computer Aided Diagnosis
CED	Convolutional Encoder-Decoder
CNN	Convolutional Neural Network
CPU	Central Process Unit
CRF	Conditional Random Field
CT	Computed Tomography
DL	Deep Learning
DICOM	Digital Imaging and Communications in Medicine
DSC	DICE Similarity Coefficient
DESS	Double Echo Steady State
DICE	Sørensen–Dice
ELMSI	Edema Like Marrow Signal Intensity
FN	False Negatives
FP	False Positives
FCN	Fully Convolutional Network
GPU	Graphics Processing Unit
HNN	Holistically Nested Network
IS	Image Segmentation
IOD	Information Object Definitions
IoU	Intersection over Union
JPEG	Joint Photographic Experts Group
ML	Machine Learning
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MPEG	Moving Picture Experts Group
OAI	The Osteoarthritis Initiative

PD	Proton-density
RF	Random Forest
ReLU	Rectifier Linear Unit
STIR	Short Tau Inversion Recovery
SSM	Statistical Shape Model
SGD	Stochastic Gradient Descent
TBMES	Transient Bone Marrow Edema Syndrome
TE	Echo Time
TL	Transfer Learning
TR	Repetition Time
UHCSJ	University Hospital Center of São João
US	Ultrasound
XML	eXtensible Markup Language

Chapter 1

Introduction

1.1 Context

Bone marrow (BM) is one of the most extensive human tissues found in the centre of long and axial bones. Among its composition are "red marrow" and "yellow marrow". While "red marrow" is composed of 40% water, 40% fat cells, and the remaining 20% hematopoietic cells, "yellow marrow" consists of 80% fat cells, 15% water, and 5% hematopoietic cells. The bone marrow dynamically changes during growth, and there is a decrease in the amount of "red marrow" throughout the years [1, 2].

The term bone marrow edema (BME) was firstly introduced by Wilson *et al.* [3] in 1988 to describe the fluid content alteration in the bone marrow, is currently being replaced by the term Edema Like Marrow Signal Intensity (ELMSI), representing more properly a histopathological diagnosis [4]. It is observed only in Magnetic Resonance Imaging (MRI), described as areas of intermediate to low signal in the T1-weight imaging and areas of high signal in the fluid-sensitive sequence. The fluid-sensitive sequences have a large range of types: T2 or proton-density (PD) weighted images with fat suppression. Those alterations indicate an increase in water content, replacing normal bone marrow tissue with more vascular tissue [1]. Although the presence of ELMSI in MRI is not a pattern to a specific diagnosis, it can be caused by trauma, tumour, some illnesses such as osteoarthritis, human growth or high-performance sports practice [5]. Being present in multiple parts of the human body, such as the foot, ankle, knee, hip and shoulder, can be associated with a wide range of diseases, sometimes associated with pain [6]. A Computed Tomography (CT) scan or an X-ray cannot detect ELMSI lesions, so MRI is the primary method of searching for them.

1.2 Motivation

Due to the pain that ELMSI is most of the time associated with, adding to the need to study the progressiveness of some diseases that have ELMSI as one of its symptoms, and the association with sports medicine, the early screening of the ELMSI could help physicians and healthcare

professionals on increasing the accuracy of the diagnosis, the efficiency of the treatments and maybe delay or prevent the progress of some diagnostics [7]. Therefore, if there is a possibility to make the automatic ELMSI identification, the impact would be widely positive on the life of the patients and the treatments applied.

Easing the diagnosis process can save time and increase the efficiency of the job done by physicians and health professionals. Since the 1960s, computer-aided diagnosis (CAD) has become one of the most relevant research topics, where Machine Learning (ML) techniques and later Deep Learning (DL) techniques have been applied to medical imaging problems to provide improved diagnostics and disease detection resources [8]. To run an efficient CAD, one of the most important steps is image segmentation, which is the main goal of the dissertation. The importance of this step in a CAD system development relies on separating important regions in the image, particularly the "separation" between the bone and the rest of the MRI scans.

The diagnosis by medical imaging is vital to health professionals to identify which is the illness and its correspondent treatment in a noninvasive approach. Although essential, usually, the raw analysis of the resultant image takes a considerable amount of time (depending on the reason and the specifics of the case and illness), and it is also dependent on the visual interpretation and experience of the health professionals when analysing. Thus, using CAD systems can reduce the impact of the mentioned factors when trying to make a diagnosis, tumours localisation or other types of structures, assist the study of the state of the anatomical structure, the adaptation of treatments and the evaluation of the progressiveness of diseases [9].

In order to have a CAD system that can reliably identify ELMSI in MRI scans, the first step has to proceed with the image segmentation, especially identifying where the bone is. There is already some work done on similar problems regarding bone segmentation in MRI, but the variability and complexity of the human anatomy are crucial factors that influence the segmentation problem. MRI is a noninvasive approach to the study of soft-tissue structures of the human body. MRI scans usually contain dozens of 2D images, so manual segmentation is time and resource consuming. The healthcare and medical field have more relevant and time-sensitive tasks to address. Therefore, automatic bone segmentation is necessary when analysing the presence or absence of ELMSI [10].

1.3 Goal

This dissertation is focused on the implementation of segmentation models for the problem of bone segmentation in MRI.

1.4 Contributions

This dissertation presents the following contributions:

- The study and implementation of deep learning architectures to segment the bone in the MRI image dataset;

- The presentation of different approaches regarding the U-Net, some of its variants and pre-trained models applied to the segmentation task;
- Finishing with an discussion of the results achieved.

1.5 Document Structure

The document is divided into six chapters with the following structure:

- In Chapter 1, are presented the context of the ELMSI, motivation for the work done, the goals of the dissertation and the contributions achieved;
- Followed by Chapter 2, where is presented the clinical background concerning ELMSI and medical imaging;
- Then, Chapter 3, are presented all the literature research done concerning DL architectures, as well as some already implemented approaches in similar topics of medical imaging segmentation;
- Later on Chapter 4, the implementation method, data and environment of the study will be described;
- The Chapter 5, presents the results and their evaluation regarding all the implementation approaches;
- Finally, Chapter 6 gives a global conclusion about the work done and the results obtained, giving some insights regarding possible work for the future.

Chapter 2

Background

In this chapter, it is presented in Section 2.1, the topic Medical Imaging, more specifically Magnetic Resonance Imaging. In Section 2.2, it is possible to contextualize clinically Bone Marrow and Bone Marrow Edema.

2.1 Medical Imaging

Medical imaging is an essential field and asset for physicians and healthcare providers to be able to analyze and inspect inside of the human body, evaluate the existence of conditions and study their progression in a non-invasive way. X-Ray, Magnetic Resonance Imaging (MRI), Ultrasound, and Computed Tomography (CT) are examples of modalities currently used by health care professionals are [11].

MRI is a diagnostic modality and imaging technique that uses no ionizing radiation, so it impacts the patient's health less than other imaging modalities with ionizing radiation. Additionally, it has great soft tissue contrast and can produce three-dimensional images with high resolution. The first mention of the technique, where the current MRI is based, is from the 1940s, but only in the 1970s was used to do imaging of the human body [11]. Some atoms, especially hydrogen atoms present in the human body in water and fat tissue, have a fundamental angular momentum, the spin, making them behave like spinning magnets when in the presence of an external uniform magnetic field. Under those circumstances, the phenomenon happens because the hydrogen atoms only contain a proton in their nucleus. The atoms are excited by a radio pulse in a resonance frequency, generating an oscillation, and a receiver coil picks up the potential difference generated by that oscillation. When the radio pulse is shut off, the nuclei realign, but at different rates, generation differential relaxation rates and signals. Those signals can be detected, and the different tissues react with different timings generating different signals in the MR image creation. In order to obtain greater contrast, the proton density can also be weighted by the relaxation times generating different modalities of the image, the T1-weighted and the T2-weighted images. Those modalities are obtained by the radio-frequency pulse variation and spacings of pulse sequences, for example, the spin-echo pulse sequence [11]. The T1-weighted images are MR images that

display signal intensity based on the longitudinal relaxation time and are acquired using short TR (repetition time, the time to complete a full iteration of a pulse) and TE (echo time) values. The fatty tissues appear brighter than other anatomical tissues. T2-weighted images display signal intensity based on the transverse relaxation time and are acquired using long TR and TE values [12, 13].

2.2 Bone Marrow

Bone Marrow (BM) is found in the centre of axial and long bones, accounting for up to 4 to 5% of the total body weight and is one of the largest tissues in the human body. Coupled with being a very dynamic organ, BM is always changing with the increase of age and due to its main purpose of supplying red cells, platelets and white cells to the blood. In Figure 2.1, it is possible to observe a representation of the bone anatomy where bone marrow is present. Although being a tissue with a varied constitution, as already mentioned, the main parts are the "Red marrow" and "Yellow marrow". These two types of marrow differ significantly in their structure. The first is hematopoietically active marrow, being responsible for the production of the blood and blood plasma cells such as platelets, and white and red blood cells, which are found in adult joints at the end of long bones, in the cavities of the skull, sternum, scapulae, vertebrae, ribs and pelvic bones. Contrasting, the "yellow marrow" is hematopoietically inactive, meaning that it does not produce the components of the blood and plasma, and is composed mainly of fat cells containing only a small amount of capillaries making its vasculature scant [2, 14]. Whereas both types of marrow have fat cells, they differ in the concentration of unsaturated acids, being the "red marrow" the one with greater concentration. The hematopoietic activeness of the "red marrow" is reduced with the human growth and bone development until more or less the 25 years old mark, and that is responsible for the conversion of the "red marrow" to the "yellow marrow". In Figure 2.2, it is possible to see the amount of "red marrow" through the growing years. Notably, the conversion starts right after birth, first evident in the phalanges of the hands and feet. The conversion is not homogeneous to all the bones and occurs at different rates until a good distribution of both marrows is achieved for each individual. The balance is different from person to person, sex, age, health state and bone shape are parameters that will affect the ratio between "red" and "yellow" marrow. MRI is the ideal monitor of those changes thanks to its rich soft-tissue contrast. [2, 14].

2.3 Bone Marrow Edema

BME, as introduced by Wilson *et al.* [3] in 1988, is currently being replaced by the term Edema Like Marrow Signal Intensity (ELMSI), representing more properly a histopathological diagnosis [4]. It is used to express the existence of interstitial fluid within marrow extracellular spaces, which demonstrates areas of intermediate to low signal in the T1-weight imaging and areas of high signal in the fluid-sensitive sequence, when compared with the normal bone marrow [17]. The different sensitivity can be seen in the T1-weighted images as darker regions, and fluid-sensitive sequences

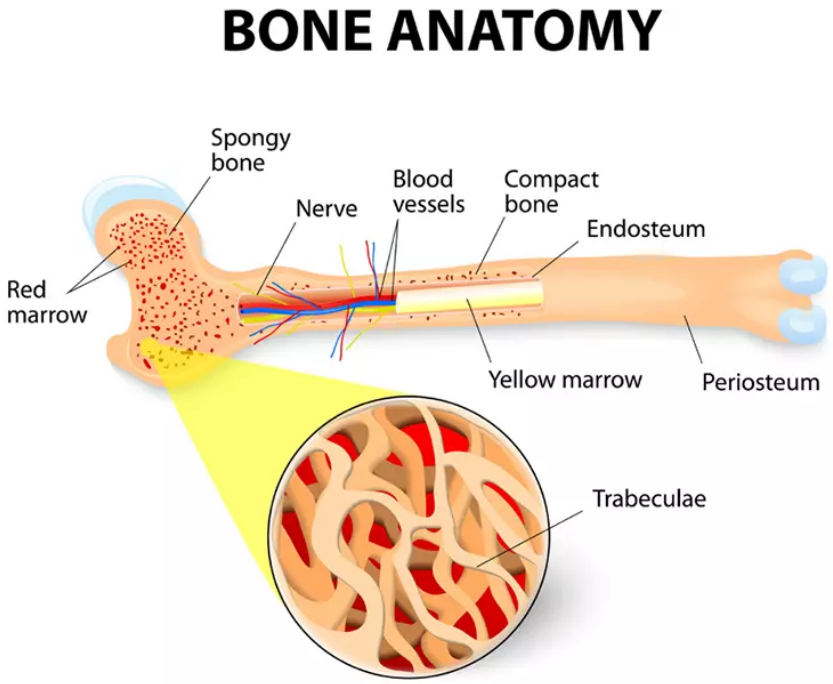


Figure 2.1: Representation of bone anatomy. From [15].

as brighter zones, [18] as shown in Figure 2.3. The change in the sensitivity of T1-weighted and fluid-sensitive sequences can be associated with the replacement of bone marrow fat by water [19], and are significantly different from the changes in case of infection, stress or other prognoses.

As previously mentioned, ELMSI can be associated with multiple diagnoses and different stages of the disease and can be found in multiple parts of the human body such as the hip, knee, ankle, foot, shoulder or spine. [20]. Currently, MRI is the radiographic modality of choice to visualize bone marrow lesions since they can not be visualized by X-Ray or CT scanning. ELMSI is not specific to any pathology type, but the context in which it is found should be considered when diagnosing and prescribing treatment and recovering analysis [18]. The main influencing factors are the patient's age, sex, medical history, and the presence or absence of symptoms, for example, pain. It is common to find appearances of ELMSI in growing children since the conversion of the "red marrow" to "yellow marrow" is still taking place [21]. Similarly, there are reported cases of ELMSI in people with intensive regular exercise, such as athletes, as seen in runners that show foot and ankle marrow edema [22].

Although common to growing children and to regular intensive athletes and most of the time asymptomatic, ELMSI is associated with some painful diagnoses. In the case of athletes, the appearance of ELMSI is associated with damage in their articular cartilage, and early screening can avoid some possible injuries [23]. The list of associated diagnoses is extensive, including trauma, Osteomyelitis, Bone Marrow Edema Syndrome (BMES), Osteochondritis Dissecans, Neoplasm, Osteonecrosis, Degenerative Arthritis, Transient Osteoporosis, Transient Bone Marrow Edema

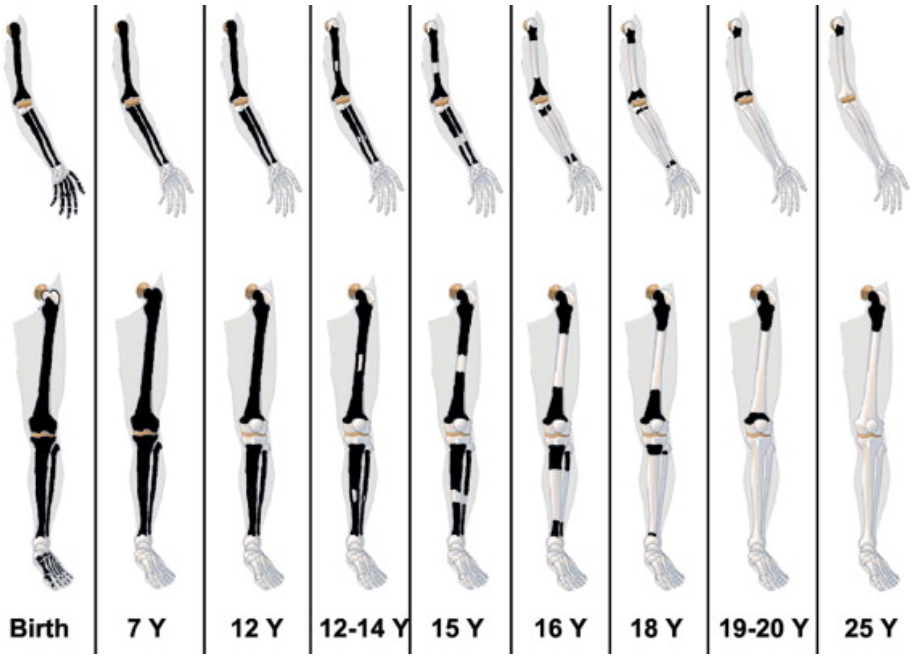


Figure 2.2: Representation of the amount of "red marrow" in tubular bones from birth until 25 years old, expressed as black colour. From [16].

Syndrome (TBMES) and tumours [4]. Bone marrow edema is commonly (up to 82% cases) seen in degenerative joint diseases such as Osteoarthrosis and Rheumatoid Arthritis [18]. Concerning the Osteoarthrosis diagnosis, ELMSI is limited to areas where the cartilage is defective, where the patients are very likely to have pain when ELMSI is found in the MRI scans. The presence and growth of the ELMSI can be a prognostic indicator of the progression of Osteoarthrosis. For Rheumatoid Arthritis, which is a chronic condition that usually leads to joint destruction, ELMSI is frequently sighted in MRIs of patients and is also an important indicator of the progression of the chronic disease. However, it is most common to associate ELMSI with traumatic cases. After the contusion or bruise, leakage of the interstitial fluid and haemorrhage within the marrow occurs, achieving fracture when the injury reaches the bony cortex. In the presence of tumours, we can also encounter regions with ELMSI [18], admitting that a large amount of ELMSI is associated with a small-sized lesion, we are usually in the presence of a benign tumour.

In terms of treatments, when trying to avoid surgical options, the physicians apply multiple medication combinations according to the patient's symptoms and reduce the progression of the associated diseases. However, the treatment depends on the cause of the ELMSI, having different medications for each condition, but with the intent of pain relief and anti-inflammatory action. One common approach is the use of Iloprost (Ilomedin, Schering, Berlin, Germany), applied as a pain reliever and to regress the ELMSI in patients with a bone bruise, stress-related ELMSI, and reactive ELMSI with osteoarthritis. Regarding the non-invasive therapies, the patient could be subject to extracorporeal shock wave therapy [24]. In contrast, the invasive approaches are counting with surgical intervention, associated with core decompression when ELMSI is located in the hip, and possible injections of hydroxyapatite cement in the ELMSI area [25].

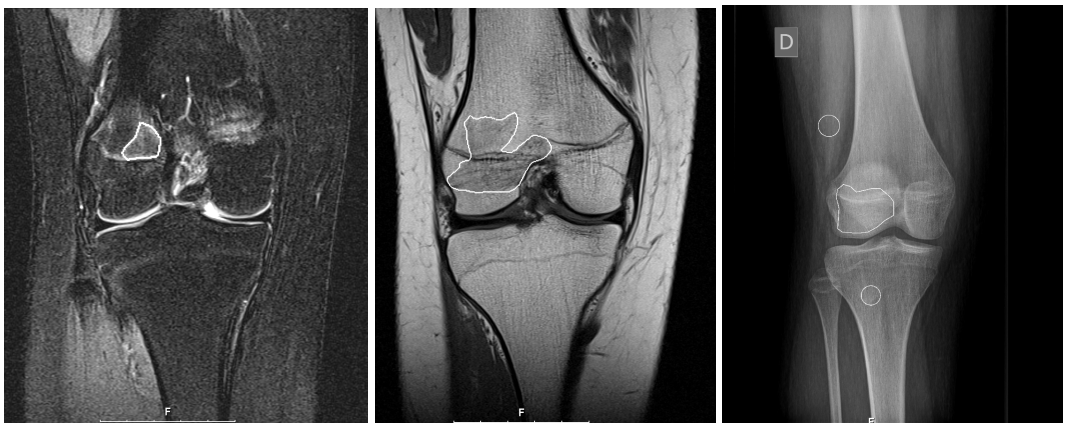


Figure 2.3: From left to right, fluid-sensitive sequence, T1-weighted sequence and X-Ray. In all the images, we can see noted ELMSI and the correspondent area in the X-Ray. Scans from the dataset of the problem associated with this dissertation.

Diagnosing ELMSI faster, with more accuracy and less resource consuming, especially when ELMSI is a sign of chronic degenerative diseases, is essential to getting a faster and more efficient treatment that accesses each need of the patients. Additionally, treatment is helpful for retarding the progression of those diseases as well as preventing further cartilage damage by applying preventive and conservative measures [23].

2.4 Summary

Medical imaging is a crucial asset to health professionals because it allows the examination of the human body in a non-invasive way and without using surgical approaches. The development of non-impact methods has made it possible to examine conditions, such as ELMSI in bones, with less impact on the human body. ELMSI is only observed using some MRI techniques and is associated with multiple medical diagnoses. Its early screening and turning it easier to diagnose will greatly impact the approach of the health professionals regarding treatment and increase the quality of living of the patients.

Chapter 3

Literature Review

The chapter is divided into two sections. The first Section, 3.1 has the purpose of introducing the Image Segmentation (IS) problem, its basic algorithms related to Machine Learning (ML), such as thresholding, regional growth, random forests and clustering methods, and the usage of approaches related to Deep Learning (DL) for problems related with brain and tumour segmentation. After that, Section 3.2 presents the methods found relevant regarding the bone, cartilage segmentation of the knee and bone segmentation of the shoulder joint and the models used to achieve the goal of bone segmentation.

3.1 Image Segmentation

Using Image Segmentation (IS), a picture can be broken up into several segments, such as sets of pixels, intensities, or textures, to identify objects and their boundaries. In ML and DL applied to the image data type, being able to understand and extract information from images is the main goal, and thus image segmentation is the first step in image analysis [26]. The problems in ML can be divided into supervised learning, where the model takes advantage of a provided label dataset and uses it to train; unsupervised learning, where a dataset in which no labelled data is provided, and the goal is to discover patterns within the data and create clusters based on the similarity between the data; and reinforced learning where the goal is to make a sequence of decisions, the training stage is responsible for learning how to detect the external environment and decide which interpretation will yield the best result [27].

Medical Imaging Segmentation is a difficult task due to the variety of the anatomy of the human body, the complexity of the images because of the presence of different artifacts, and the lack of homogeneity in intensity. It requests that the process of analysing the images be done by experienced experts [28]. Image segmentation algorithms play a vital role in biomedical-imaging applications, such as quantification of tissue volumes, localisation of pathology and study of anatomical structure [29]. IS can be divided by the goal of the actual task. If the idea is to allocate classes to each image pixel with a semantic label, it is called semantic segmentation [30]. If the primary purpose of the technique is to partition individual objects in the image, it is called

instance segmentation and extends the scope of the semantic segmentation because it not only assigns a class to each pixel but also can delineate each object of interest in the image [31, 32].

As previously mentioned, IS is the first step to the image processing problem. There is a wide range of ML approaches. Regarding methods that are under classic ML algorithms and that are common to be referenced in literature are the techniques that use intensity analyses and shape modellings, such as edge-based detection, threshold techniques, regional growth segmentation, as well as non-deep learning techniques such as random forests and clustering methods [32, 33].

Starting with the simplest method of IS, the threshold method takes the grayscale information processing by evaluating the grayscale value of the different encountered elements. The algorithm can be divided into global and local threshold methods, where the first divide the background and target by a threshold, and the second selects multiple thresholds to divide the image into multiple regions and backgrounds depending on those thresholds. The approach is helpful because it does not require an extensive calculation and obtains results faster, but it has struggled with problems where there is not a significant grayscale difference and an excellent overlap of elements [34].

Another example of an algorithm is the regional growth segmentation, where its main characteristic is to have similar properties to the pixels and thus form small regions, making them expand to include all the homogeneous neighbours [35]. It requires the selection of the seed pixel, checking its surroundings to merge the similar pixels. It has the advantage of providing a reasonable boundary, but it has a high computational cost compared to other ML algorithms, and it is influenced by noise and grayscale unevenness [34]. Segmentation by edge detection is also used in IS problems to separate elements within the image into distinct regions. It is still affected by noise and can identify fake and weak edges [36].

Random Forests (RF) is an algorithm that introduced a new concept to image segmentation: "contextual information". They are a good algorithm for detection, localization, segmentation and image-based prediction. Combining the idea of bagging (training each tree with the dataset) with random feature selection where, at each node, is selected a subset of features, to assemble randomized decision-based trees (set of binary tests of the features), that they are fed with the input data to make a prediction [37, 38]. They help remove the irrelevant features and have good accuracy with small amounts of data, but they are very computational expensive [39].

The clustering approach is the grouping of objects regarding their similar attributes, and those groups are called clusters. Clustering is a general term that includes multiple algorithms that have different "main" ideas concerning how the clusters are formed. In the K-Means algorithm, for example, data vectors are grouped into predefined clusters, the centroids are initialized randomly, and each pixel is assigned to a cluster by Euclidean distance. This process is repeated until no significant changes are observed in the arrangement of the clusters [35]. The K-Means clustering is very fast and straightforward, being a promising approach for large data sets, but the number of clusters has no explicit criteria for its selection, and each interaction has to go by all the samples [34].

Although the success of the application of the ML methods, the usage of Deep Learning (DL) approaches to Medical Image Segmentation has grown throughout the years, especially for MRI

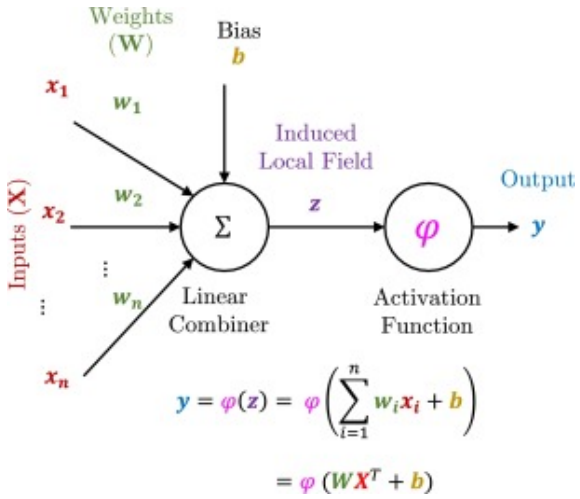


Figure 3.1: A graphical representation of the artificial neuron. From [42].

and ultrasound (US) images [40]. Deep Learning (DL) is a sub-part or the "evolution" of ML concerning the usage of multi-layer structures to automatically learn and extract features from a raw set of data. It takes advantage of the Artificial Neural Networks (ANN), which are inspired by the human brain, to create models composed of those structures and combine them in multiple layers. Usually, the ANN is a set of interconnected, hierarchically organized neurons, where the different layers perform different transformations on the data input to obtain different levels of abstraction to extract features, and there are three types of layers: the input layer, hidden layers and the output layer[41]. Its main structure, the neuron demonstrated in Figure 3.1, has a set of weights that grant the ability to control the flow of information from each input, and a decision is taken using an activation function. The output of the neuron y can be defined as represented in the equation 3.1, where x values are the inputs, the w are the weights and ϕ is the activation function.

$$y = \sum_{i=1}^{n} w_i x_i + b \tag{3.1}$$

For the network to learn, it occurs, during training, modification of the weights that control the inputs into the neuron, where the goal is to minimize a Loss function. In trying to solve the problem, this function is a relationship between the values obtained and the reference, and its primary purpose is to evaluate the precision of the network. There are also a large number of loss functions, and the selection of the appropriate one to solve the problem can be very crucial [42]. Adding to that, the neural networks take advantage of a method called *backpropagation*, where the difference between the output result and the reference is calculated and then propagated from the last layer to the first one. The network also needs an optimizer responsible for, interactively, finding a local minimum to find the optimal weights and bias.

The advantage of using DL techniques is that instead of extracting the features manually, it only needs a small prepossessing and the feature engineering that was needed before is now on the side of the computer [43]. There are multiple approaches in the DL scope, depending on

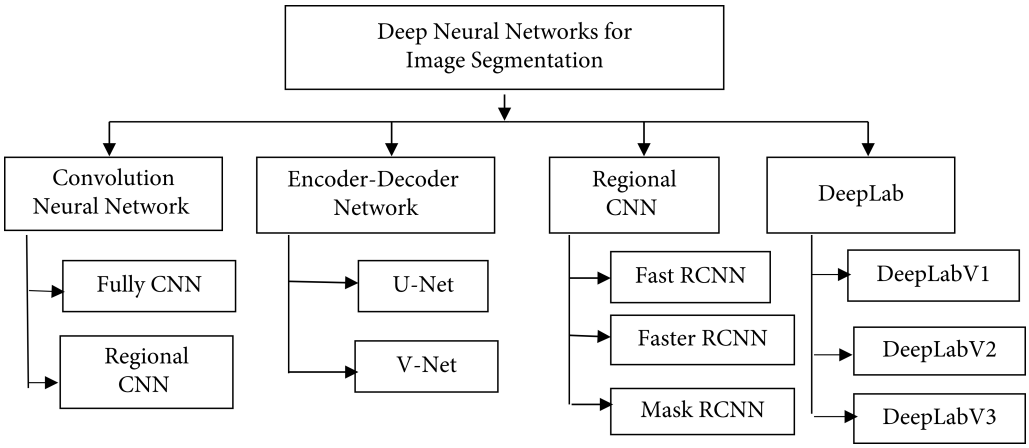


Figure 3.2: For the IS problem the most common network architectures applied. From [28].

the actual dataset. Convolution Neural Networks (CNNs), for example, have been applied for automatic brain tumour segmentation [44] and for segmentation of breast tumour [45], and the reported results surpass the ones using classical ML approaches. In Figure 3.2 can be seen as the most common method of DL for image segmentation.

Sun *et al.* [46] applied a Convolutional Neural Networks (CNN) to segment brain tumors with low computational requirement. To achieve the goals, it was implemented an Application Specific CNN, which focused on the specificity of the task, in this case, the brain tumour segmentation task. The implementation had three steps: pre-CNN, the CNN and pos-CNN. The first had the goal of reducing the volume of the input data by removing tumour-free slices by evaluating the symmetry of the brain (slices with tumours are asymmetrical) and slices with an incomplete appearance in the brain. Afterwards, the data was fed to the developed CNN.

In a nutshell, CNNs are models for processing data that have a grid pattern (example: images), inspired by the hierarchical organisation of the human visual cortex, and designed to automatically and adaptively learn features from low to high-level patterns. These structures are composed of three types of layers, convolution layers and pooling layers, both responsible for feature extraction and fully connected layers responsible for mapping the extracted features into a final output. The convolution layers imply the stack of mathematical operations like convolutions, and it plays the most critical role in this approach [47]. Each convolutional layer applies a filter, called the kernel, convoluting each filter across the spatial dimensionality of the input to produce a 2D activation map. Each kernel will have a corresponding activation map, and at the end, all maps will be stacked along the depth dimension to form the total output volume [48]. Each feature map is used to extract local characteristics of the same positions in former different feature maps [49]. These layers can significantly reduce the complexity of the model by the optimisation of its output [48]. Pooling layers reduce the number of parameters and complexity of the representation by reducing the dimensionality. Finally, the fully connected layers contain neurons that are directly connected to the neurons in the previous layer and connect them to every single neuron of the current layers, and the last layer is followed by an output layer [50], as shown in Figure 3.3.

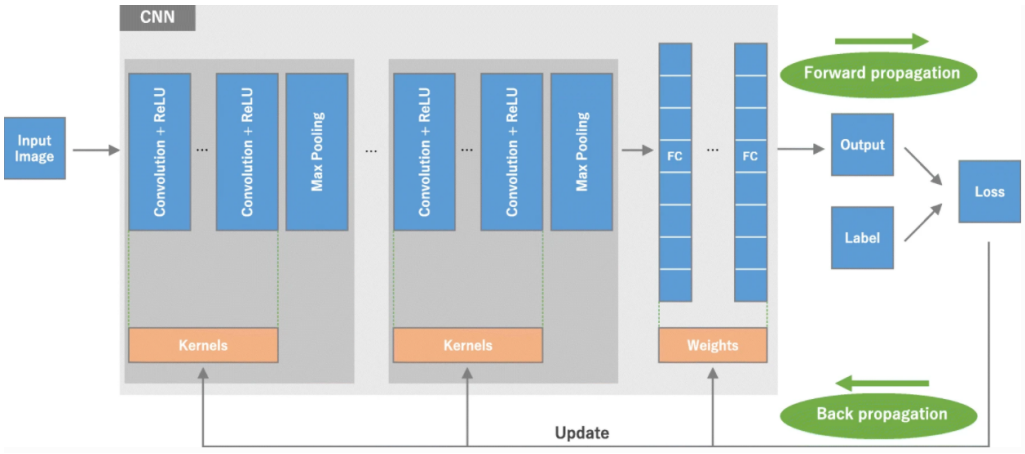


Figure 3.3: Overview of a CNN and the training process. From [50].

Regarding Sun *et al.* [46] implementation of the CNN, it was specifically designed to receive as input 2D brain image slices of the four modalities present in the dataset. The convolution layers are used to feature extraction from the four modalities input, as well as tumour localization and pixel classification. It only uses 108 convolution kernels used in the seven convolution layers. Due to the simplicity of the model developed, were used three different modes of convolution (standard, depthwise and group convolution), as well as normalization to uniform the data distributions were performed, the kernel applied in each convolution was carefully tailored, and the usage of Full-ReLu activation function to cope with information loss of the normal ReLu function. The pos-CNN is used to identify the pixels that were wrongly classified as positive, using the idea that tumours are a 3D volume and the area of each one of them should be found in consecutive slices. The model performed well compared with the baseline, achieving DICE Similarity Coefficient (DSC) scores for enhancing tumour of 77.2%, 89.2% for whole tumour and 76.3% for tumour core. However, the high performance the model presented is specific for tumour segmentation and allowed the reduction of randomness and redundancy in computation, as well as the dependency on training samples, increasing efficiency.

Yongchao *et al.* [51] proposed an approach to segment brain MRIs utilizing a Fully Convolutional Network (FCN) and Transfer Learning (TL). Following 2D slides of the 3D volume was the input to the FCN that used the VGG network [52] (composed of thirteen convolutional layers followed by three fully connected layers), that was pre-trained on the ImageNet dataset [53] (large dataset containing annotated images for computer vision projects). The VGG network discarded the fully connected layers, being substituted by specialized convolutional layers, and the max-pooling layers divide the base network into five stages, as represented in Figure 3.4. The main advantage reported was the need for a low number of training images regarding the main problem of 3D brain segmentation. It is essential to mention that the main goal was to segment brain MRI images from neonates to ageing adults with core structure differences. It was reported significantly reduce segmentation running time and promising results with a lower amount of training images

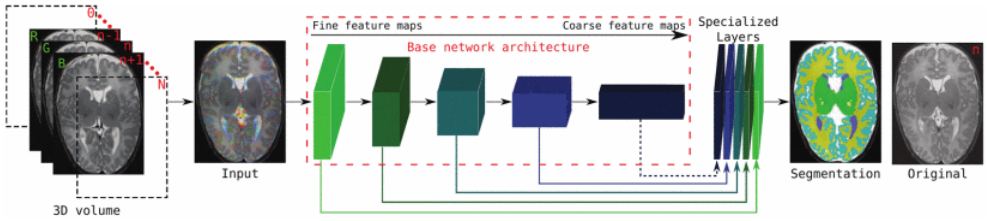


Figure 3.4: Proposed network for the brain segmentation, where the last VGG Net layers were substituted by the specialized layers. From [51].

while comparing with the baseline.

3.2 Bone Segmentation

3.2.1 Knee Bones and Cartilage segmentation in MRI

The evaluation of the Osteoarthritis progressiveness in knee joints motivated multiple approaches for the segmentation of knee cartilage. Older studies apply classical approaches such as "region growing" [54] and "edge-based" [55], semi-automated approaches. The more recent approaches mainly work with DL techniques due to the upgrade in the accuracy and better results [56].

3.2.1.1 Model and Atlas Based approaches

Graham Vincent *et al.* [57] proposed a statistical model based on Active Appearance Model (AAM) (introduced by Cootes *et al.* [58]), that uses the training set of manual segmented images by an expert, to develop statistics of the shape and image features information. The model performed reasonably well, presenting a maximum mean distance (evaluates the similarity of two images) of 1.4 mm. Fripp *et al.* [59] proposed the utilization of a hybrid 3D Active Shape Model (ASM) (also introduced by Cootes *et al.* [60]) for the BS, which is also a statistical model to generate a suggestion of the shape of the bone and it is used an atlas as a base for the development of the statistical model. The results presented a high accuracy on an inhomogeneous database, where the ASM obtained close enough solutions to boundaries. Both approaches were marked as sensible to the initialization parameters. Besides statistical approaches, Yin Yin *et al.* [61] proposed the utilization of a layered optimal graph image segmentation of multiple objects and surfaces (LOGISMOS), which is an n-dimensional graph approach based on "algorithmic incorporation of multiple spatial inter-relationships" in a single graph [61]. The proposed method envisages the segmentation of the femoral bone, femoral cartilage, the tibial bone, tibial cartilage, the patellar bone and the patellar cartilage, constituting an issue of simultaneous segmentation. It starts with a pre-segmentation step, followed by the construction of a graph that contains all relationships and surface cost elements, where the segmentation of all the surfaces coincides with the optimization [61]. Regarding the atlas-based methods, the approach seeks to label structures

by mapping the image to an anatomical pre-constructed atlas (introduced by Rohlfing *et al.* [62], and the atlas distinguishes the spatial relationship of the anatomical structures) and each image voxel is assigned by using the label of the correspondent structure in the atlas [63]. Tamez-Peña *et al.* [64] implemented a multi-atlas approach by enhancing the atlas-based method with outlier detection, developing multiple anatomical knee atlas defined by experts from MRI datasets. The results presented confirm that it was possible and reliable to automatically segment the bone and cartilage in the knee, presenting a DSC Similarity Coefficient (DSC) (measures the similarity of two images) of 88% for the femur bone. Liang Shan *et al.* [65] also applied multi-atlas to achieve a fully-automatic segmentation of the femoral and tibial cartilage with a non-local patch-based label function to allow spatial separation of the cartilage. It also required an expert segmentation of the femur, tibial and the corresponding cartilage. It was demonstrated that the multi-atlas segmentation is appropriate for cartilage segmentation, being robust, overcoming occasional failures [65].

3.2.1.2 Deep Learning approaches

The most recent approaches use CNNs and their variations and improvements for segmenting knee cartilage and bones. Although the CNNs can be extended from 2D to 3D, the 3D version has large memory and training time requirements [66]. Liu *et al.* [56] proposed to implement musculoskeletal cartilage and bone segmentation using an automated method that combines a semantic segmentation Convolutional Encoder-Decoder network (CED), which is a paired encoder and decoder network and a 3D simplex deformable modelling. As the core of the CED segmentation engine, the team chose a SegNet architecture composed of the encoder and decoder networks, designed to be an efficient pixel-wise semantic segmentation [67]. The SegNet is similar to the VGG-Net network type because it uses the same thirteen convolutional layers, and each encoder layer group is repeated five times. To recover the feature maps using up-sampling, the proposed approach removes the same structure as the VGG-Net. The big difference between the SegNet with the VGG-Net is that the decoder network is a reverse process of the encoder and mimics the same number of five decoder layer groups, consisting of an up-sampling layer followed by a trainable convolutional layer and a ReLU activation. Finally is followed by the last layer that is responsible for the production of class probabilities for each pixel [56]. The 3D dimensional deformable model is based on the class probability output by the SegNet. In Figure 3.5, it is possible to see a representation of the SegNet utilized. For training the SegNet the 3D volume was dissembled into 2D images and labels. The SegNet uses batches of the 2D image slices and compares them with the pixel-wise class labels. Ally to that was applied an objective function to estimate the training loss and to update the parameters of the SegNet until a pre-defined maximum of interaction steps. In the test phase, the fully trained SegNet is utilized to segment the 2D testing images. A 3D fully connected Conditional Random Field (CRF) is applied to refine the segmentation by improving the label assignment for the voxels with similar contrast. The 3D simplex deformable process takes the individual segmented object to be refined to preserve tissue boundary and maintain anatomical geometry. After that, the 3D segmentation is obtained by merging all

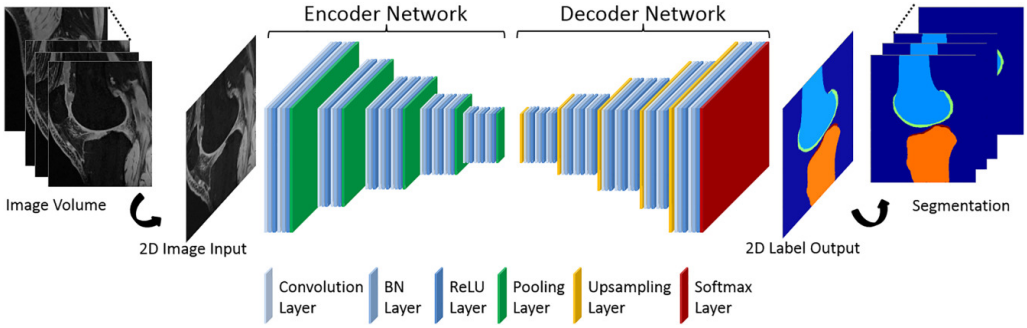


Figure 3.5: SegNet used to perform the cartilage segmentation. From [56].

the segmentation objects [56]. It is mentioned that the CED network implementation was crucial because it is better suited to perform complex transformations and is less sensitive to overfitting. There is also a suggestion of the methodology to implement concerning the ELMSI problem by applying a different image intensity threshold technique to isolate and measure the volume of the ELMSI [56]. Despite the success of the approach, and the top performance for segmenting the femur mentioned, some aspects need to be considered when trying the same approach in the future. The CED process requires computational resources and many pixel-wise annotated training sets for evaluating each new tissue contrast.

Almajalid *et al.* [10] addressed the problem of segmenting soft tissues of the knee by starting to segment the bone first, considering it as the first step to cartilage segmentation, as mentioned before. Therefore, it was developed a fully automatic segmentation for the knee bone by taking advantage of a U-Net and applying some changes to the basic U-Net structure. U-Net was firstly introduced by Ronneberger *et al.* [68] as a network that would be able to work with smaller training sets and achieve more precise segmentation. The U-Net basic structure is represented in Figure 3.6 and has two paths. The first one is the contracting path, also known as the encoder, which is similar to the regular convolution network and provides some classification information. The second is the expansion path, known as the decoder, and consists of up-convolutions and concatenations with feature extraction. The increase in the output's resolution is passed to the final convolution layer to achieve the fully segmented image [69]. The contracting path consists of two successive 3×3 convolutions followed by a ReLU (Rectified linear activation unit, which transforms the negative input values into zero) and a max-pooling layer (which calculates the maximum value for each path of the feature maps), repeating multiple times. The second path is the expansion path, where each stage up-samples the feature map using 2×2 up-convolution. After that, the feature map is cropped and concatenated onto the up-sampled map. This process is followed by two successive 3×3 convolutions and ReLU activation. Moreover, finally, to produce the segmented image, an additional 1×1 convolution is done to reduce the feature map to the required number of channels [69]. The U-Net has proven to efficiently perform the segmentation of images with a limited number of labelled training images. It combines the information from the downsampling path with the contextual information of the upsampling path. The limitations

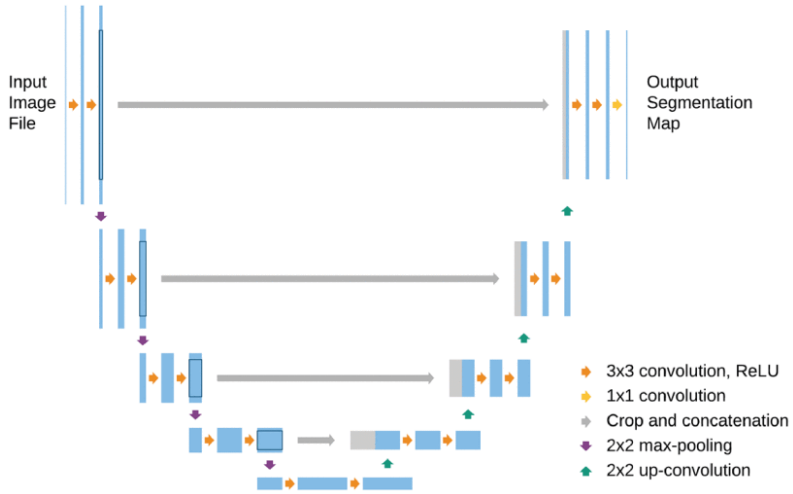


Figure 3.6: Basic U-net architecture. From [69].

are regarding the size of the image being limited and the skip connections imposing a restrictive fusion, which causes accumulation of the same scale feature maps [28].

The slight modified U-Net used Almajalid *et al.* [10] takes the 3D image chunk to tune any parameter of the dataset due to the capability of the method to self-adjust to a varying dataset. Before the implementation, the dataset was optimized and also divided into three groups train, validation and test. The preparation of the dataset also included the manual marking of all the femur bones as ground truth. The U-Net basic structure was modified to accept multiple channels as inputs instead of only one channel, as well as the addition of padding to control the shrinkage of the image size while applying each convolution to avoid losing the pixels near the borders. Adding to that, instead of the original stochastic gradient descent optimizer of the original version, it was used the Adam optimization algorithm to update the network weights based on the training data and take advantage of the DSC to verify the accuracy [10]. In total, the modified network has twenty three convolutional layers. It was reported favourable results for the segmentation of 3D knee MRI, for the segmentation achieving a DSC of 96.73% to original U-Net and 97.22% using the modified U-Net for the image size of 352×352 [10].

Cheng *et al.* [70] presented an alternative to patellofemoral bone segmentation, simplifying the CNN architecture with the implementation of a Holistically Nested Network (HNN), firstly introduced by Xie *et al.* [71]. The HNN is a simplification of the CNNs, deleting the decoding path. It produces multiscale outputs in a single path using feed-forward neural networks. It can reduce the hyperparameter space, avoid overfitting, sometimes associated with the U-Net, and reduce the computational cost. Due to its capability of training and prediction using the whole image end-to-end, the segmentation uses local and global contextual information [70]. It is mentioned that the HNN is a appropriate candidate for the knee segmentation task because it does not need manual pre-processing of the data, as well as being a good approach for both young and older patients and is not subject to the variability of the structures between different patients. Before the data

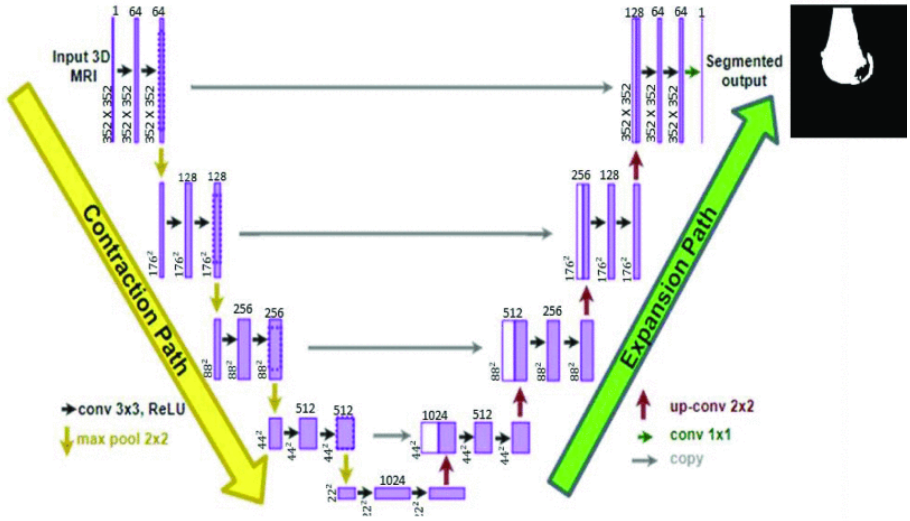


Figure 3.7: Modified U-net architecture used by Almajalid et [69]

entered the HNN, all images were pre-processed using different techniques for uniformity of the images. After that, the images went through the HNN, which was divided into five staged convolution layers, where each layer used different size and number of filters, feeding the feature maps to be used in the layer that follows, with its side outputs, that produce fine-to-coarse probability maps. From left to right, the most shallow layer generated a probability map with finer details and the deep layer generated the probability map containing global contextual information. With the production of a probability map in mind, each output layer was inserted following the convolutional layers by merging the feature maps at each stage. To merge every resultant probability map, a weighted fusion layer was used to compute the final prediction. The representation of the HNN is shown in Figure 3.8. The implementation was fully automatic, and it successfully segmented the patella and the distal femur, with a high DSC similarity coefficient (which validation metric that gives the similarity of two samples) results (97% for the femur and 94% for the patella). It was no mentioned significant differences in the accuracy of automatically segmenting immature bones compared to the mature bones [70]. It is mentioned that the size of the training dataset and variability should be considered when applying the model to other similar problems.

Ambellan *et al.* [72] applied, for the segmentation of bone and cartilage of the knee, an approach that combined Statistical Shape Model (SSM) and a CNN. The combination is represented in Figure 3.9. SSM are sets of models that can describe a collection of similar structures and the possible anatomic variations, in this case, the variability for the knee. The model applied 2D CNNs, sub-regional 3D CNNs and SSM-based techniques in a segmentation pipeline. In the first step, the 2D CNN created the rudimentary segmentations of the femur and tibial bone. The regularization of the segmentation masks was followed by fitting the SSM. After that, the output of the SSM was fed to a 3D CNN to segment small MRI sub-volumes. The next step concerns the application of SSM post-processing, which uses pre-defined regions of the SSMs to enhance

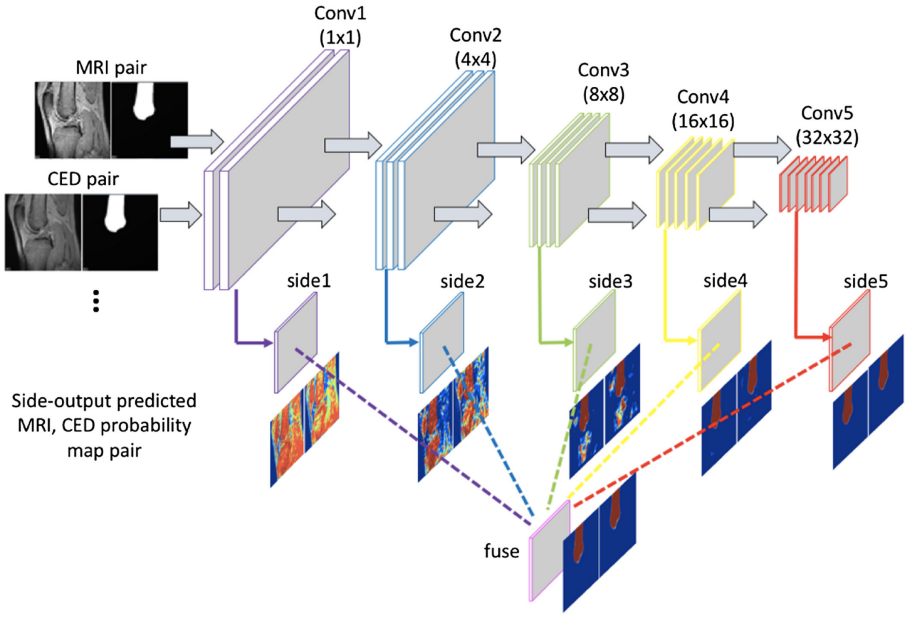


Figure 3.8: HNN architecture representation. From [70].

the output results of the 3D CNN. To finalize, the cartilage was segmented using 3D CNNs. The process segments the bone and the tibia independently and individually. For the 2D CNNs, it was applied a variant of the U-Net architecture, where it was possible to segment the tibia and femur individually by the increase of the number of input channels compared with the original model to enhance the spatial consistency of the segmentation results between individual slices. Regarding the SSM step, it was implemented to face inaccuracies in areas with low-intensity contrast. SSMs cannot express osteophytic details since deformations are patient-specific. The 3D CNN, mainly 3D U-Nets, were applied to face this issue, trained on small sub-volumes of the MRI scans. The SSM post-processing is used to finalize the femur and tibia bones segmentation by correcting the false positive voxels that are located outside the typical range of osteophytic growth. The model achieved high segmentation accuracy rounding 98% of DSC in all the tests performed. Several compromises were made regarding the sizes of the sub-volumes in order to reduce the computational costs [72].

3.2.2 Shoulder Joint Segmentation in MRI

Wang *et al.* [73] combined a U-Net segmentation model with an AlexNet (proposed by Krizhevsky *et al.* [74]). The U-Net received an input image of the labelled training set, including the humeral head and articular bone. The AlexNet is used to determine the edges of the bones accurately. AlexNet was adapted by increasing the number of channels to segment the MRI image and adjusting the convolutional kernel and pooling kernels. For the dataset, the images were filtered for noise reduction and radiologists labelled the joint bones and humeral head parts that mattered. In order to increase the data seen by the architecture, were performed data augmentation techniques,

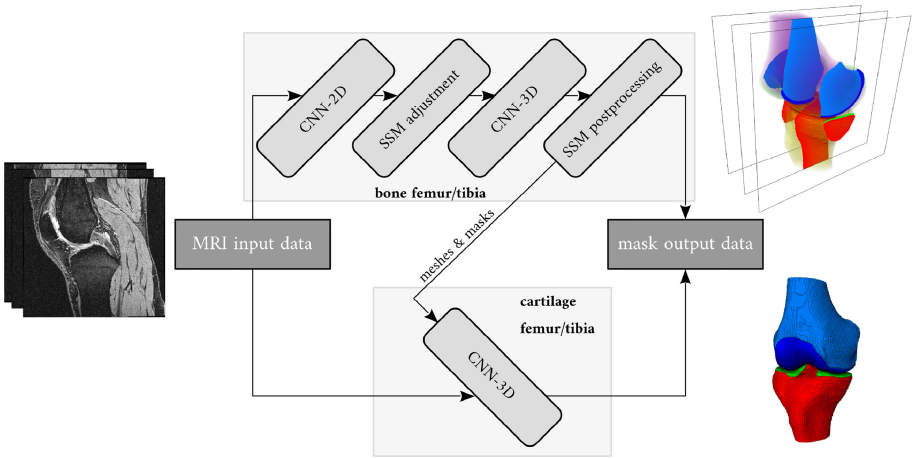


Figure 3.9: Proposed cascade of CNN and SSM steps. From [72].

such as random 90-degree rotation, random vertical flip, horizontal random flip and others. The workflow, represented in Figure 3.10, starts with generating candidate bone regions with the resultant partial segmentation using the U-Net and the AlexNet. The next step was extracting the effective bone from the rectangle candidates, predicted by the average value of the rectangle area in the forward and backward prediction frames. Finally, the edge refining segmentation occurs, after which the soft tissues and noisy areas that were not the bone and cartilage parts are filtered. It was reported that high values for accurate measurement were dependent on the size of the block used in the first steps, but rounding accuracy of 97% [73].

3.3 Summary

This Chapter presented the problem of IS and its importance when analysing Medical Images because IS is the first step in multiple biomedical-imaging applications. For the past few years, DL methods have surpassed the usage of ML approaches to the medial imaging world due to less feature engineering required by DL methods, and it has shown better results in multiple different problems. Multiple networks were developed and optimised to be focused on biomedical imaging problems. Multiple CNNs were developed for the brain, organs and tumour segmentation, as well as the implementation of pre-trained models in the segmentation problem. Concerning BS, the DL models have been gaining more relevance than the latter ML approaches, the U-Net architecture and some of its variants are the most relevant and used architectures for the BS problem.

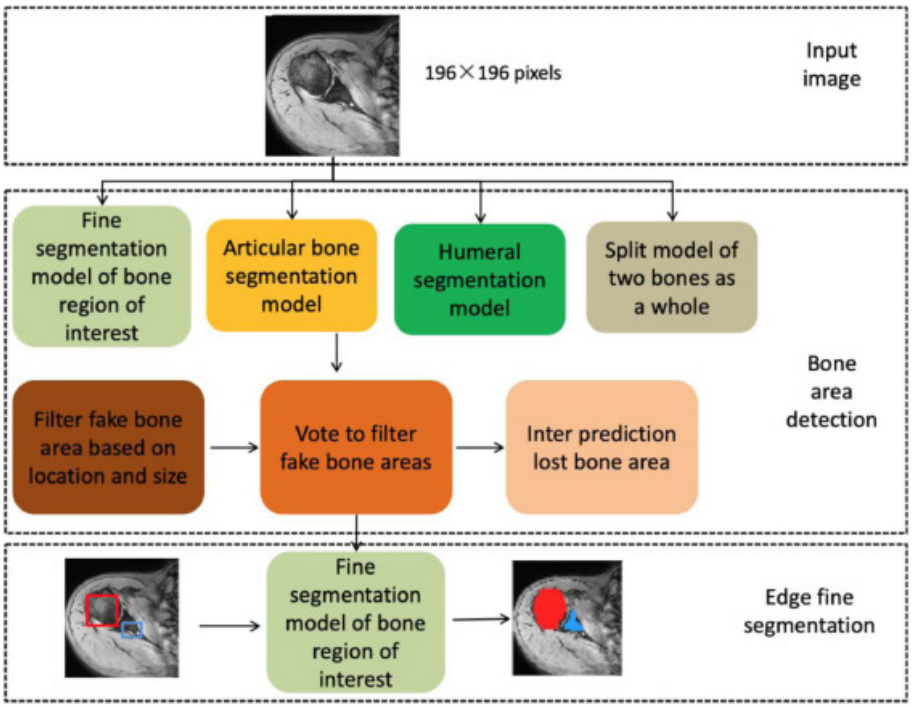


Figure 3.10: Segmentation framework representation. From [73].

Chapter 4

Methodology and Experimental Setup

This Chapter gives a data contextualisation, its analyses and the steps taken to its presentation (Section 4.1). All the implementation procedures taken to apply the segmentation models are presented in Section 4.2.

4.1 Data

4.1.1 Data from University Hospital Center of São João

The dataset provided by the University Hospital Center of São João (UHCSJ) contained images from 72 patients until the age of 18 that had MRI scans and X-ray scans of each subject. In Figure 4.1 is represented the sex distribution of the dataset and in Figure 4.2 is presented the age distribution between cases. Each folder of each patient has an XML file that contains patient details, exam details and scan file distribution, an executable to see the scans, and the Digital Imaging and Communications in Medicine (DICOM) binaries. Most medical imaging methods rely on computer processing, where the computer collects, stores, and utilises the data, for example, constructing 3D models from the input data. Different methods and technologies that acquire medical imaging created the need to standardise the way that even if we used different machines and computers, one could get to process the outputs of the acquisition. Therefore, it was developed a standard communication protocol where different manufactured equipment could output information to be processed in the same computer, the protocol DICOM. Nowadays, the DICOM protocol allows effective medical imaging storage and transfer over large geographical areas, a great development from its first version, developed in the 1970s [75]. As said, DICOM was developed to ease the interchange and standardisation of medical images. It also defines network-oriented services that provide transfer and printing capabilities and media formats for the exchange of data successfully. The DICOM documents use Information Object Definitions (IODs), which are attributes to describe certain aspects of the image and are the central components of the data structures [76], and are of great importance. For example, "patient" is an information object that has attributes such as "patient name" and "patient ID number". Those attributes describe the type of the object, data of the patient, performed procedures or reports and technical information about the medical imaging

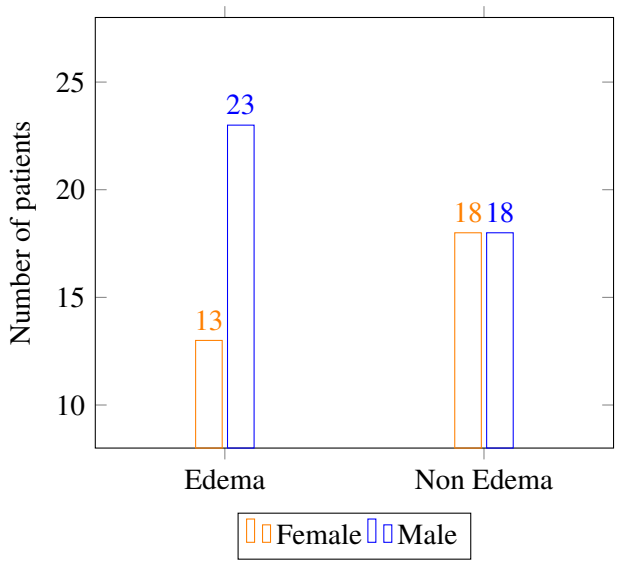


Figure 4.1: Representation of the sex of patients in the dataset.

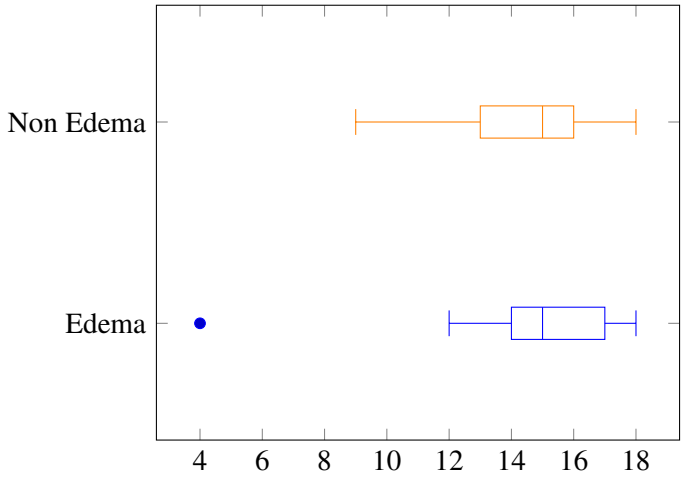


Figure 4.2: Representation of distribution of patients' ages.

device used that usually differ from manufacturer to manufacturer, such as device manufacturer, device serial number and others. It is accepted a wide range of image types, for example, X-ray intensity images are used, and it supports multi-dimensional multi-frame images. The compression method accepts the standards such as JPEG, JPEG Lossless, or MPEG-2 for video [75].

The dataset has 72 patients, that had all the personal information that can identify the patient was omitted for motives of confidentiality, where in 36 patients ELMSI was identified in the MRI scans, and manually annotated by two health professionals as demonstrated in Figure 4.3, also locating the normal signal muscle and normal signal bone marrow, as well as the annotation of the resultant ELMSI area in the X-Ray scan latter on, there were added slices with bone annotated. The other 36 patient scans with no ELMSI found had the healthy bone annotated, as seen in Figure 4.4. For the dissertation problem, the only scans used to perform training, the validation and testing of the models, were the ones with the bone annotation without ELMSI. Later on, more

slices with bone annotated were added concerning the scans of the patients with ELMSI. The T1-weighted images had 356 slices, while the fluid-sensitive images had 349 slices, where the bone was annotated. All the scans were from the coronal plane.

4.1.1.1 Preparing the data

For each case, there is available an XML file, where it is possible to find some information about the anonymized patient, for example, age and sex, as well as to access the correspondent images of the MRI scans and the medical annotations. Some slices had irrelevant annotations, such as arrows and measurements not concerning the bone segmentation, as shown in Figure 4.5. Those irrelevant annotations were removed, only remaining the annotations of the bone. Some other slices had annotations of bone with open lines that needed to be closed in order to produce the pair MRI slice and the ground truth to feed the models for training. Adding to that, some pairs needed the correction of the annotation positioning in order to coincide with the location of the bone. After getting the pair MRI and ground-truth, as presented in Figure 4.6, the images go through a pipeline to feed the models, where they are normalised to have values between $[0, 1]$ by the min-max normalisation method [77]. Then they are rescaled to the target input size, depending on the model to be applied. The dataset was divided in order to obtain the training, validation and testing set according to the 60% for training, 20% for validation and 20% for testing ratio, as represented in Table 4.1.

Table 4.1: Dataset split into training, validation and testing sets

	T1-weighted sequence	**Fluid-sensitive sequence**
Training	213	209
Validation	72	70
Testing	71	70

4.1.1.2 Data augmentation

The dataset is considered very small compared with other problems in DL. In order to provide more cases for the model to learn, some techniques of data augmentation were applied: horizontal flip, rotation of 20 and -20 degrees, a small vertical shift and a Gaussian blur. In Figure 4.7, are represented examples of the applied techniques. All the augmentation techniques were applied in the training set, increasing the number of training images as presented in Table 4.2.

Table 4.2: Data augmentation numbers

	T1-weighted sequence	**Fluid-sensitive sequence**
Before data augmentation	213	209
After data augmentation	1278	1254

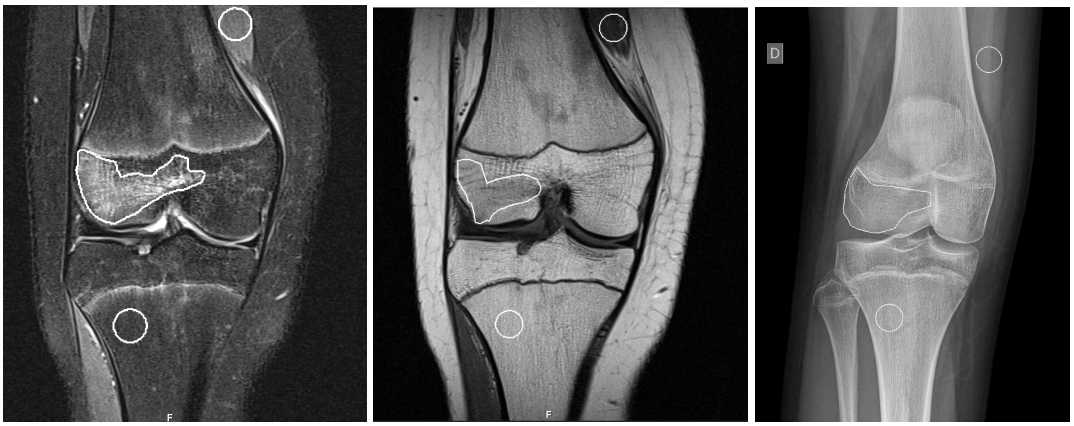


Figure 4.3: From the dataset, scans where ELMSI was annotated, as well as the identification of the tibia and the quadriceps muscle. From left to right, fluid-sensitive sequence image, T1-weighted sequence and X-Ray.

Figure 4.4: From the dataset, scans where no ELMSI was found, and the healthy bone was annotated, as well as the identification of the tibia and the quadriceps muscle. From left to right, fluid-sensitive sequence image, T1-weighted sequence MRI and X-Ray.

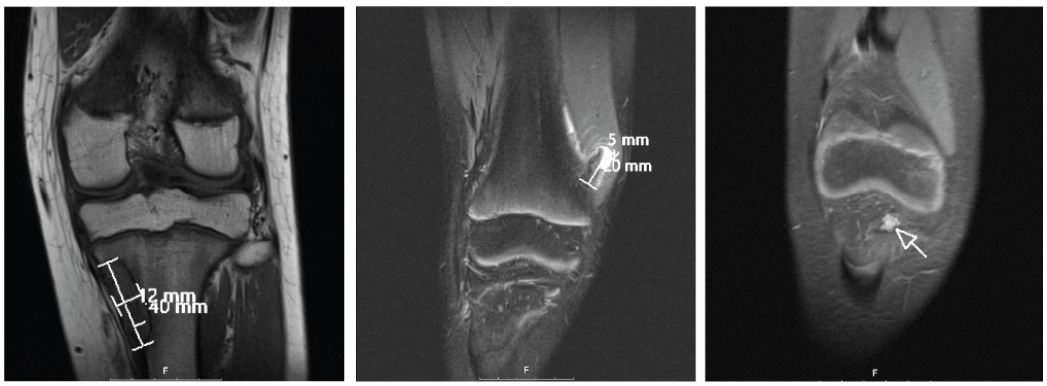
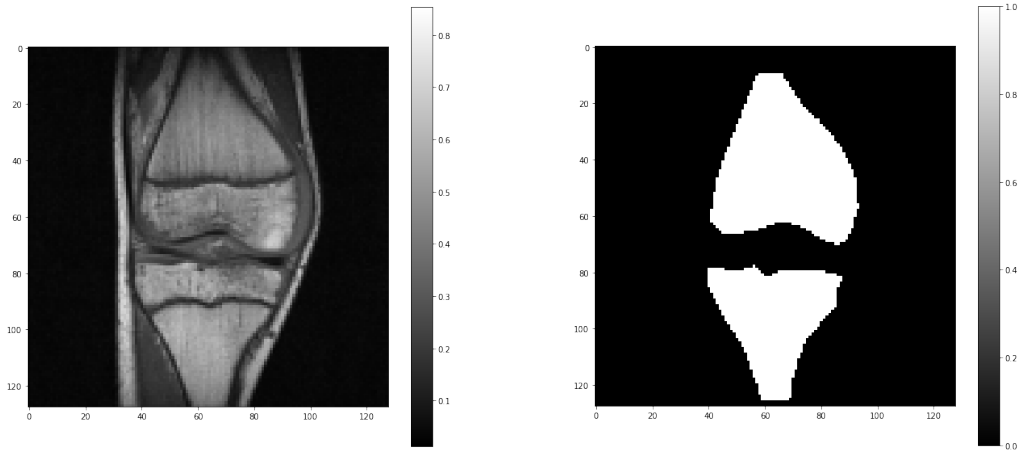


Figure 4.5: Examples of excluded annotations.



(a) MRI slice

(b) Ground truth

Figure 4.6: Slice from the University Hospital Center of São João and the correspondent ground truth.

4.1.2 OAI ZIB dataset

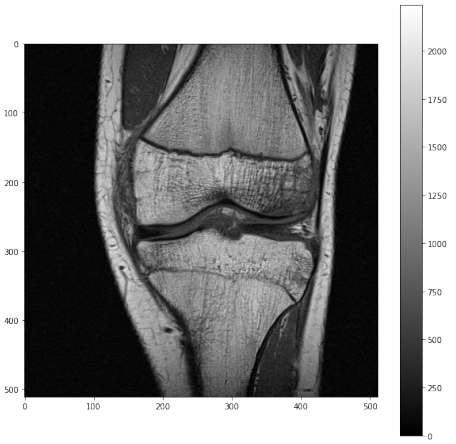
This dataset was produced by experts in Zuse Institute Berlin[1], and is a publicly available dataset [78]. It contains manual segmentations for four anatomical structures: femoral, femoral cartilage, tibial, and tibial cartilage. The scans correspondent of the manual segmentations are from the dataset of the Osteoarthritis Initiative (OAI) [2]. A pair of slice and correspondent annotation is represented in Figure 4.8. There are 507 different Double Echo Steady State (DESS) sequences of the sagittal view of knees corresponding a 507 different patients. The DESS sequences are 3D coherent steady-state sequences, that combine features from different types of signals, granting the capability of signalising the fluid brightly, and the bone appears extremely dark. The DESS sequence is great for combining morphological and functional analysis with high resolution, but it requires a long scan time and presents metallic susceptibility artefacts. In general, 3D DESS allows great assessment of cartilage thickness and volume [79, 80]. The ages of the patients have an average of 61.87 ± 9.33, and 48.32% are women [78].

4.1.2.1 Preparing the data

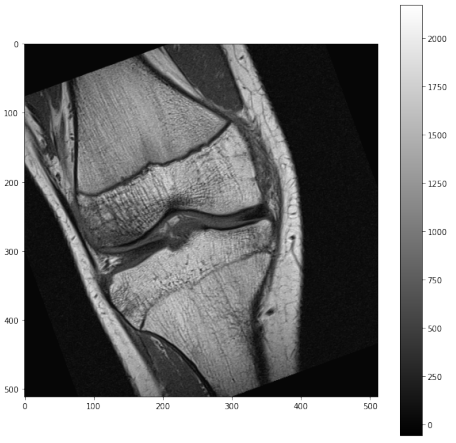
The OAI ZIB dataset comes with the location of the MRI slices on the OAI database that corresponds to the annotations. When the MRIs were extracted from the OAI database, they had to be pre-processed, so all unwanted cartilage labels were converted into the background, and all bones labels were converted to the same label, 1, and were flipped so that the bone areas coincided with the bone areas in the MRI. In addition, the slices where there was no bone annotated were removed. Motivated by the need to feed the pair to the model, there was also a process of normalisation, as mentioned before, where the values of the slice were normalised in order to belong to

[1] https://pubdata.zib.de/
[2] https://data-archive.nimh.nih.gov/oai/

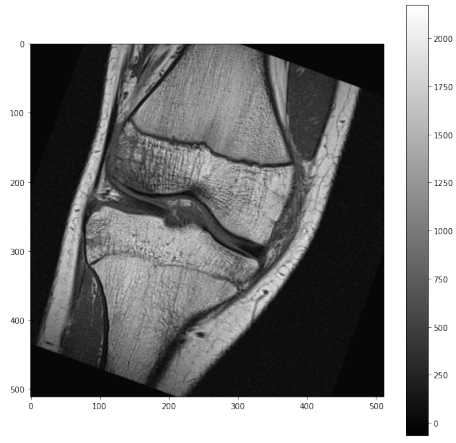
(a) Original MRI slice



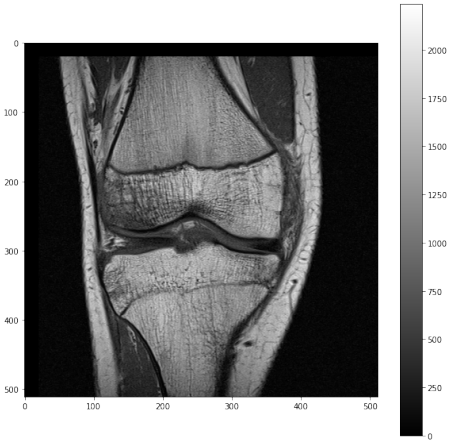
(b) Horizontally flipped slice



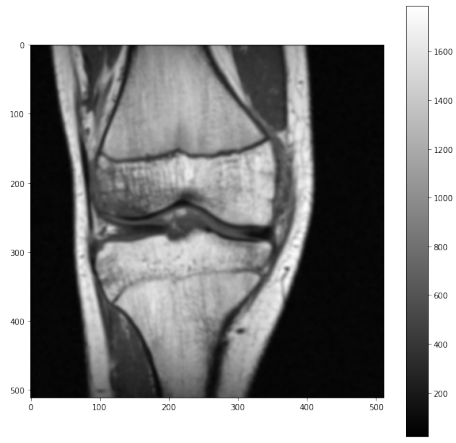
(c) 20º rotated slice



(d) -20º rotated slice

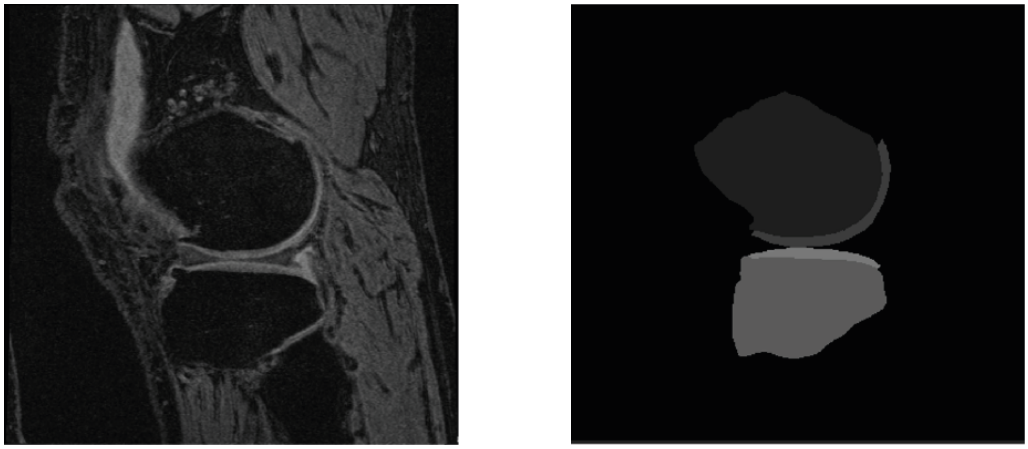


(e) Vertical shifted slice



(f) Gaussian blurred slice

Figure 4.7: Data augmentation techniques applied to a T1-weighted image slice.



(a) MRI slice (b) Annotation

Figure 4.8: Slice from the OAI dataset and the correspondent annotation from the OAI ZIB dataset [78].

the range of $[0, 1]$ by the same min-max method [77]. In total, there were 62708 slices of the knee with annotated bone in it. Figure 4.9 represents a final pair of slice and ground truth.

4.2 Models implementation

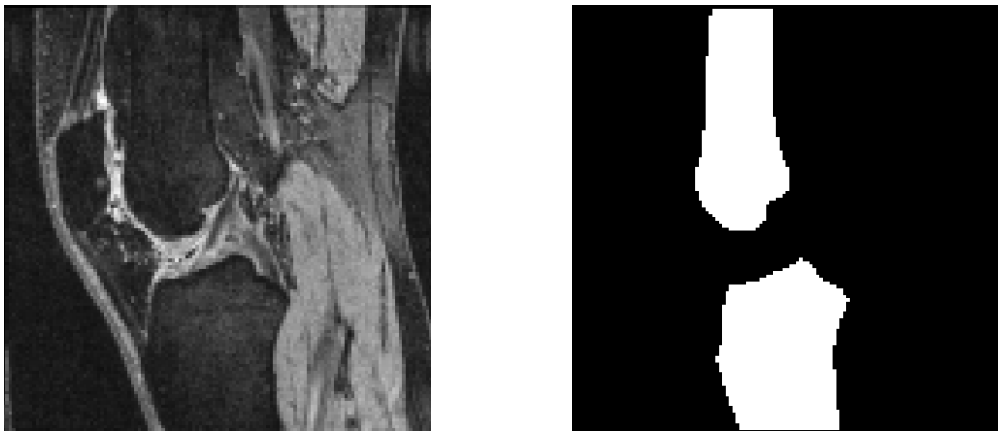
The main goal of this study was to implement DL approaches to the problem of bone segmentation. According to [28] there are multiple implemented approaches regarding Medical Imaging Segmentation, especially for segmentation of internal organs, such as the chest organs, eyes and brain. It implemented the following models: U-Net and a slightly modified version of the U-Net, and an Attention U-Net. Furthermore, pre-trained models were implemented in the OAI ZIB dataset and a U-Net that used the pre-trained encoder of a VGG-16 model in the ImageNet dataset.

4.2.1 Environment and tools

The implementation of the proposed models took place in the *Google Colab*[3] environment. It is an environment facilitated by *Google Research*, free to use, that is a hosted Jypyter notebook service, with no setup required, providing free access to computing resources and GPUs. The used GPU for the training of the models fluctuated between NVIDIA Tesla T4 and NVIDIA Tesla K80 and disk space of 36GB. The CPU was Intel(R) Xeon(R) CPU @ 2.30GHz.

All the code was developed in *Python* 3.7 [81] due to its known applicability in DL projects, as well as its available libraries with free access for image processing and DL development. Regarding image processing, were used some libraries: *OpenCV* [82], *Scikit-image* [83], and *Pydicom* [84]. For the model implementation and all the processes involving training, validating and testing the libraries or APIs used were *TensorFlow* [85] and *Keras* [86].

[3] https://colab.research.google.com/



(a) MRI slice (b) Annotation

Figure 4.9: Slice from the OAI dataset and the correspondent pre-processed annotation from the OAI ZIB dataset.

4.2.2 U-Net

The network implemented is referenced in subsection 3.2.1.2, which was mainly developed for solving problems of medical imaging segmentation [68]. It is a model that applies a contracting path, responsible for the down-sampling, and an expansive path, responsible for the up-sampling part. The network is composed of nineteen 2D convolution layers. A Rectifier Linear Unit follows the first eighteen convolution layers (ReLU), represented in equation 4.1 as an activation function, responsible for outputting a positive number in case of positive input and zero in case of negative input. The last convolution layer is followed by a Sigmoid activation function, represented in equation 4.2, that outputs values between $[0,1]$ in order to obtain a prediction for the pixels. The first eighteen convolution layers are followed by a Batch Normalisation layer responsible for standardising the inputs to a layer for each mini-batch, stabilising the learning process. In the contracting path, are used four layers of Max Pooling 2×2, with the responsibility of reducing the spatial size of the representation, in order also to reduce the number of parameters to be computed for each feature map. It is an efficient way to detect features after down-sampling in a more efficient way, avoiding overfitting, and it does that by picking the max value of a determined region of the feature map. In the expanding path, are applied four transposed 2D convolution layers to the connections, which were obtained by an image cut. These steps are important because non of the convolutions applies padding to the input images, giving masks with less spatial dimensions. In Figure 4.10 is a representation of the model used.

$$ReLU = max(0,x) \tag{4.1}$$

$$Sigmoid = \sigma(x) = \frac{1}{1+e^{-x}} \tag{4.2}$$

Were tested multiple hyperparameters, such as different loss function, learning rate, batch size, and optimiser. In Table 4.3 is presented, all the options considered while searching for the most

suitable combination of hyperparameters. Concerning loss functions, were tested the DICE loss function, the Binary Cross Entropy loss function and the Tversky loss function. In the following expressions, A represents the true value and B the prediction.

- **DICE loss function**: it is inspired by the DSC, which measures the similarity of the perdition and the ground truth. It is twice the amount of the intersection area between the segment predicted and the ground truth, divided by the total number of pixels, as represented in equation 4.3 and the DICE loss function is represented in equation 4.4 [28, 87].

$$DSC = \frac{2|A \cap B|}{|A \cup B|} \tag{4.3}$$

$$DICELoss = 1 - \frac{2|A \cap B|}{|A \cup B|} \tag{4.4}$$

- **Binary Cross Entropy loss:** it measures the difference between two probability distributions. Usually used in classification problems, image segmentation is applied at pixel level classification [87]. It is defined by the expression on equation 4.5.

$$BCE = -(A \times log(B) + (1-A) \times log(1-B)) \tag{4.5}$$

- **Tversky loss:** is associated with Tversky index (TI) [88] (presented in equation 4.6), is a generalization of the DSC. It adds a weight on errors by the help of a β coefficient (default being 0.2) [87], represented in equation 4.7.

$$TI = \frac{AB}{AB + \beta(1-A)b + (1-\beta)A(1-B)} \tag{4.6}$$

$$TverskyLoss = 1 - \frac{AB}{AB + \beta(1-A)b + (1-\beta)A(1-B)} \tag{4.7}$$

Concerning optimisers, two were tested, Stochastic Gradient Descent (SGD), the basic algorithm responsible for the convergent of the networks, by updating the parameters of the network, using *backpropagation* and the Adaptive Moment Estimation (Adam) optimiser, which is a stochastic gradient descent method as well, but based on adaptative estimation of moments of first and second order [89]. The learning rate and the batch size were selected according to the most efficient pair. The method used for the initialisation of the weights is called He initialisation, which finds values from a truncated normal distribution on zero and keeps the size of the previous layer in mind. The weights are still random but have different ranges according to the size of the previous layer of neurons.

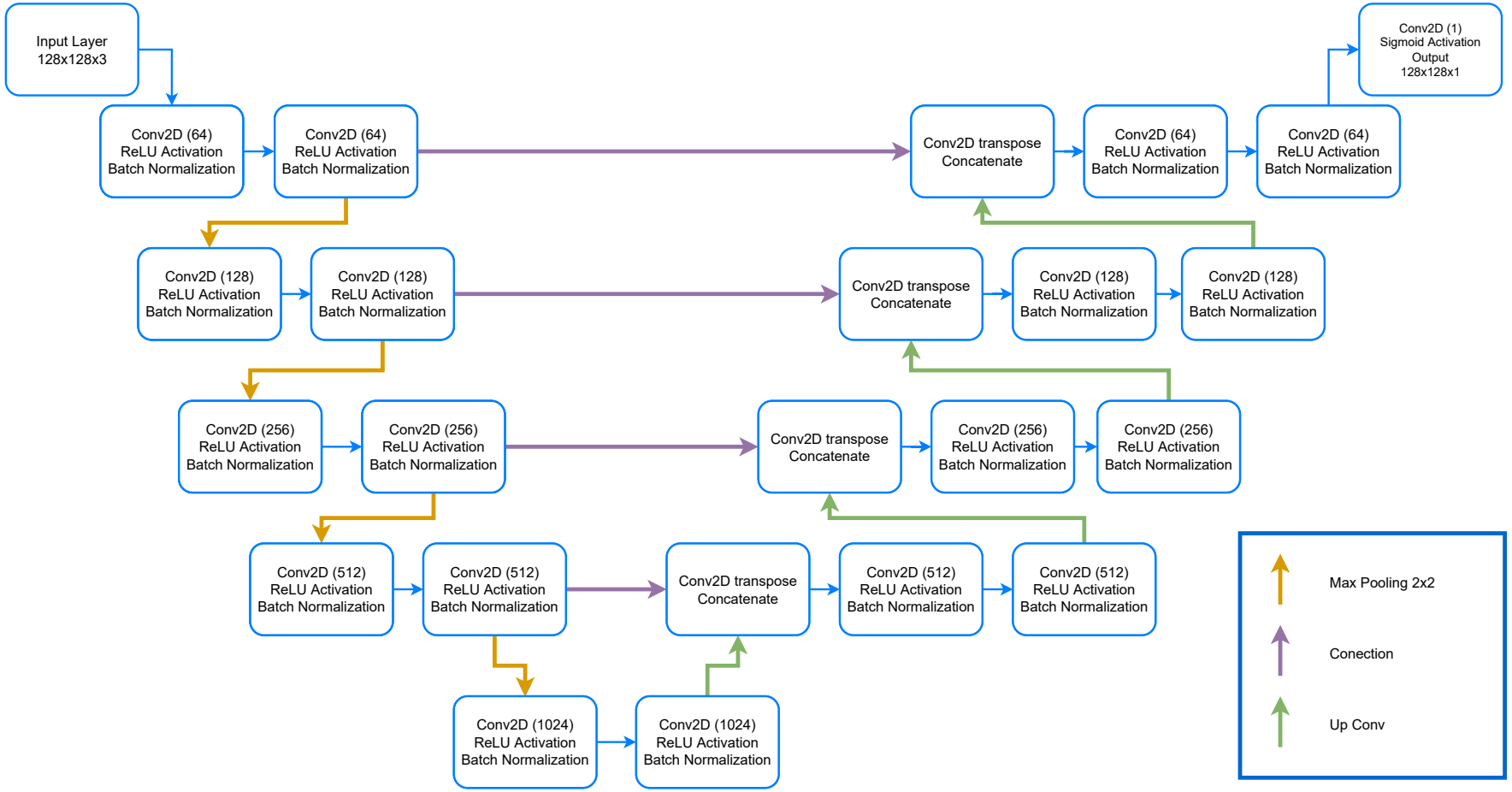


Figure 4.10: U-NET model representation used.

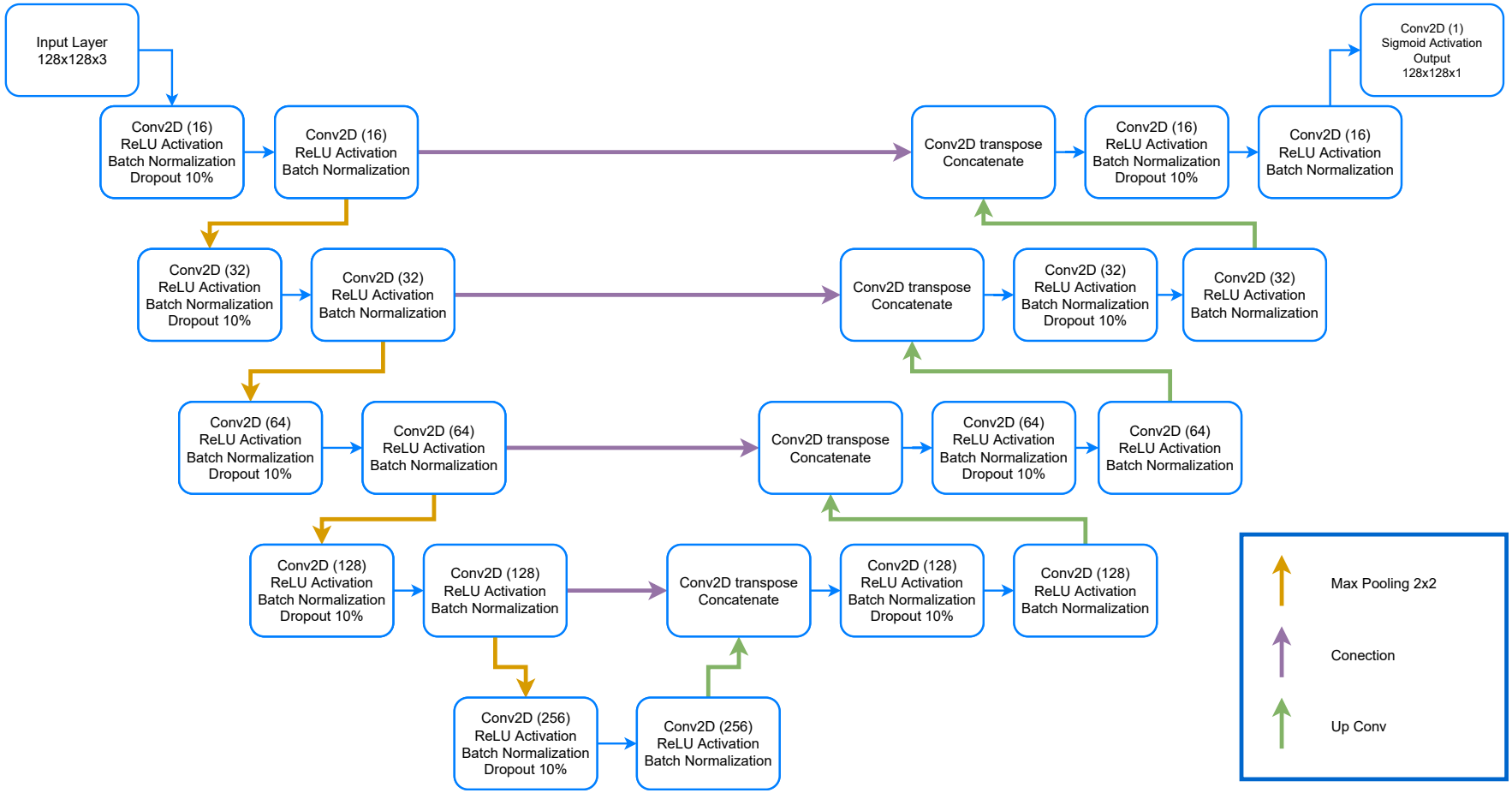


Figure 4.11: Slightly modified U-NET model representation used.

Table 4.3: Hyper parameters tested

	Optimizer	Learning Rate	Epochs	Batch size	Loss
1	Adam	0.1	100	8	DICE loss
2	Adam	0.01	100	8	DICE loss
3	Adam	0.001	100	8	DICE loss
4	Adam	0.0001	100	8	DICE loss
5	SGD	0.1	500	8	DICE loss
6	SGD	0.01	500	8	DICE loss
7	Adam	0.001	100	16	DICE loss
8	Adam	0.001	100	32	DICE loss
9	Adam	0.001	100	16	BCE loss
10	Adam	0.001	100	16	Tversky loss

4.2.3 Attention U-Net

Furthermore, another model was applied, in the same line as U-Net. The Attention U-Net is a variant of the original model developed by Oktay *et al.* [90] for the segmentation of the pancreas. Adding Attention Gates (AG) to the U-Net standard model helped it focus on specific objects and improve sensitivity to foreground pixels [90]. The AG is applied in the expansive path through which the features extracted in the contracting path must pass before the concatenation. For the segmentation problem, the AG allow the network to focus on the segmenting objects [91]. The AG weights the feature map according to each class, and the network can then focus on a particular class. Figure 4.12 represents the AG, where the input features x^l are scaled with the attention coefficient α. Both input signals suffer a $1\times1\times1$ convolution and its result goes through a pipeline of transformations that include a ReLU activation (σ_1), followed by a $1\times1\times1$ convolution, then Sigmoid activation (σ_2), and finally a grid resampler [90, 91]. In the end, there is a concatenation of the pipeline signal with the original input signal. The representation of the model used [4] is in 4.13.

4.2.4 Pre-trained models

Intending to increase the performance of the bone segmentation task, TL was applied. Firstly, research was done on public datasets concerning the problem of bone segmentation. Of the two datasets found, one of them was already unavailable, remaining the OAI ZIB dataset. The goal was to train the U-Net with the OAI ZIB and then use the training to transfer the features learned and used in the dataset from the University Hospital Center of São João. Minor alterations were needed in the U-Net model, mainly due to the efficiency of training, the reduction of the time needed to train with this large dataset and due to the limitations of the environment used to train the network. For example, the environment was not able to support batch sizes bigger than one hundred and twenty eight. Besides Batch Normalization, Dropout layers were used in order to prevent the possibility of overfitting. The number of filters for every convolution was reduced as

[4] https://github.com/bnsreenu/python_for_microscopists

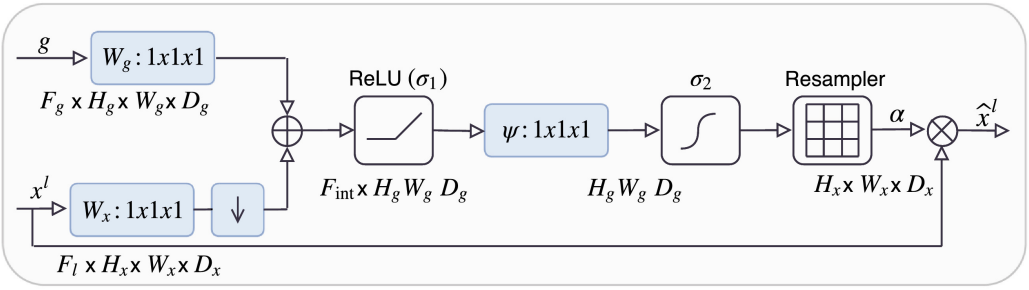


Figure 4.12: Attention Gate and all its constituents. From [90].

presented in Figure 4.11. The network was trained using 10^{-3} as a learning rate, for 50 epochs, with the batch size of 128 (for training time optimisation and due to the larger amount of data available), using DICE loss and the Adam optimiser. After having the outcomes of the modified U-Net implementation, meaning after the model was trained on the OAI ZIB dataset and obtaining the correspondent weights, in a first attempt, they were used to initialise the weights to perform the model training for the dataset from the University Hospital Center of São João, instead of the previous initialisation method, to understand the impact of weights initialisation in the problem resolution. The second experiment was done, taking only the encoder part of this pre-trained model corresponding to the contraction path of the model, freezing its layers, and training the rest of the model with the dataset provided by the University Hospital Center of São João. After that occurred a fine-tuning by de-freezing those layers and applying some more training with a very small learning rate of 10^{-7}. For the last tests, the encoder of one VGG-16 network [52] trained in the large ImageNet dataset [53] and used as the base encoder in the contracting path, with its layers frozen while training with the University Hospital Center of São João's dataset. After that training, fine-tuning was applied with a very small learning rate of 10^{-7}.

4.3 Summary

Before the implementation of the DL architectures, the data was analysed and needed a small pre-processing before being fed to the models, as well as the implementation of techniques of data augmentation. Another dataset was also pre-processed to prepare for the implementation of TL techniques. The implementation of the architectures followed the pre-processing: U-Net and a slightly modified version of the U-Net, an Attention U-Net and pre-trained models in the OAI ZIB dataset and a U-Net that used the pre-trained encoder of a VGG-16 model in the Image Net dataset.

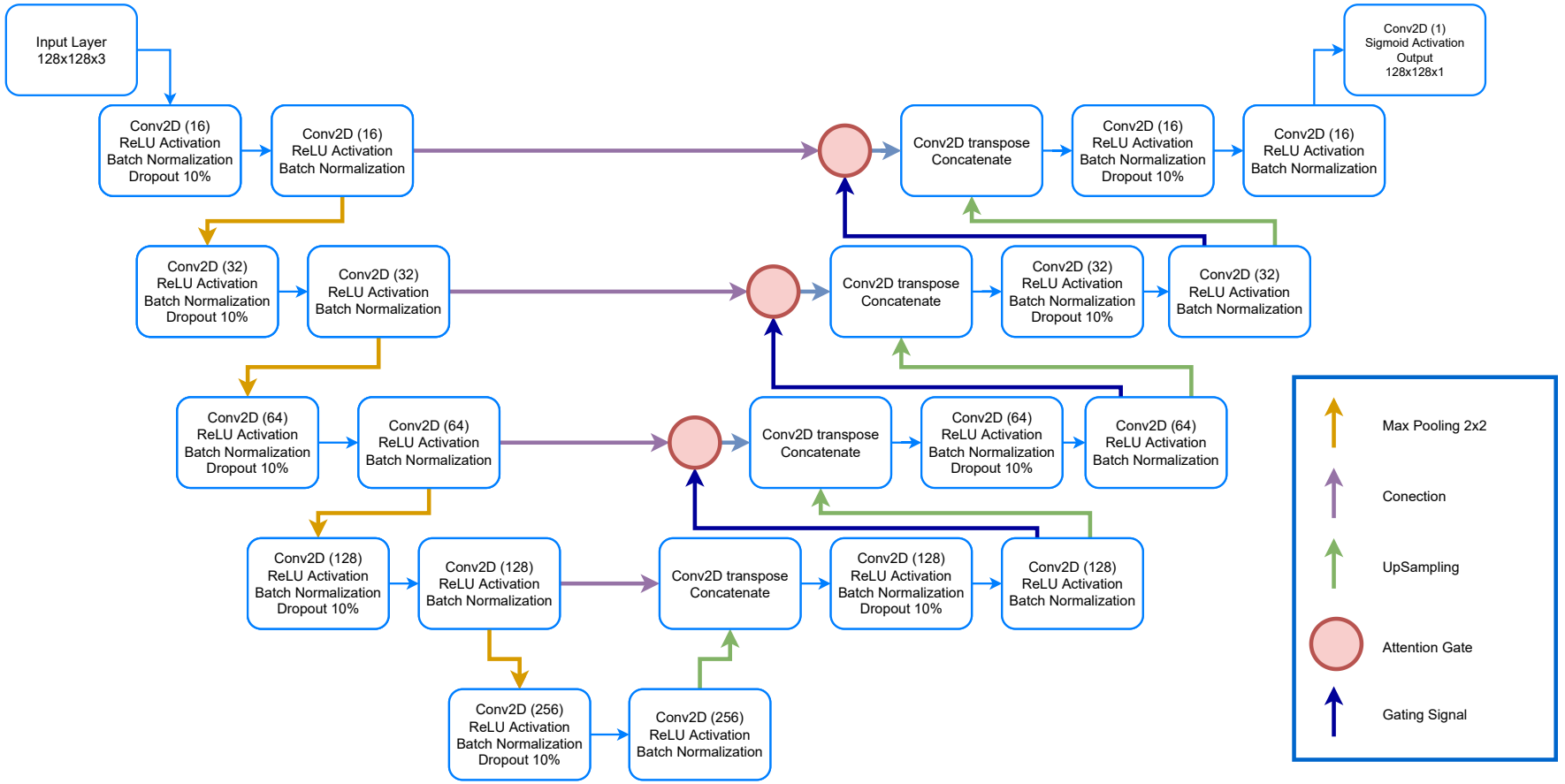


Figure 4.13: Attention U-NET representation.

Chapter 5

Results and Discussion

The current chapter presents the metrics in Section 5.1 that will be used in Section 5.2 to evaluate the results obtained regarding the segmentation of bone with methods referenced earlier in this dissertation.

5.1 Metrics

The metrics are responsible for evaluating the performance of the models in the problem context. The annotations done by the health professionals are taken as ground truth, and with those as references, it is possible to evaluate the segmentation result. The following metrics were selected to measure the performance of the models implemented [28]:

- **Precision:** measures the proportion of input cases that are reported as true as represented in equation 5.1.

$$Precision = \frac{TP}{TP + FP} \tag{5.1}$$

- **Recall:** measures the proportion of the relevant results correctly classified as represented in equation 5.2.

$$Recall = \frac{TP}{TP + FN} \tag{5.2}$$

- **F1- Score:** combines the Recall and the Precision in an harmonic mean as represented in equation 5.3.

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5.3}$$

- **DICE Similarity Coefficient (DSC):** measures the similarity of the perdition and the ground truth. It is twice the amount of the intersection area between the segment predicted and the ground truth, divided by the total number of pixels, as represented in equation 5.4.

$$DSC = \frac{2|A \cap B|}{|A \cup B|} \tag{5.4}$$

- **Jacquard Coefficient:** also known as Intersection over Union (IoU), is a widespread metric regarding image segmentation, and it is the amount of intersecting area of the prediction and the ground truth mask, divided by the area of the union, and it is represented in the equation 5.5.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{5.5}$$

5.2 Results and Discussion

The present thesis implemented the U-Net model and the Attention U-Net either trained from scratch or by transferring knowledge from pre-trained models from a similar dataset [78] or in the ImageNet [53]. For the identification of the adequate hyperparameters, the final combination that revealed the most successful in terms of segmentation and the available environment resources management is represented in Table 5.1. The results obtained in every test done were decisive for the selection of the hyperparameters and for the guidance of the following steps to be taken. All the tests were done in both T1-weighted and fluid-sensitive image sequences. All the models were trained and tested three times, with different combinations of data for training, validation and testing.

Table 5.1: Hyperparameters selected.

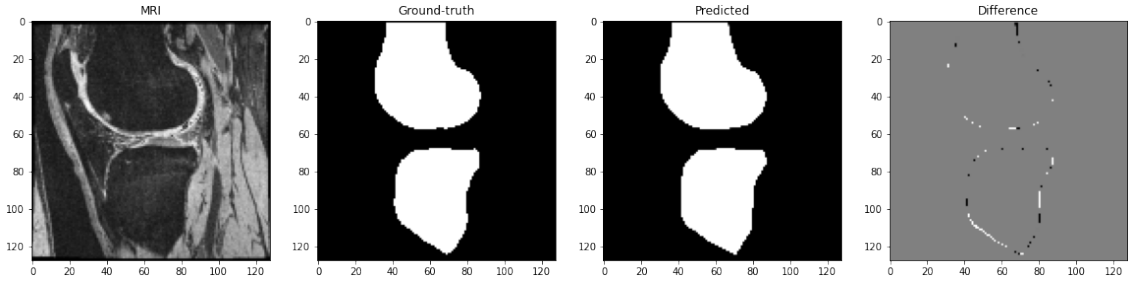
Optimizer	Learning Rate	Batch size	Loss
Adam	0.001	16	DICE loss

Starting with the implementation of the U-Net model, the slightly modified U-Net model, the Attention U-Net and the usage of data from the University Hospital Center of São João (UHCSJ), the difference in the inputs of the models and architecture are responsible for the differences in the segmentation results. In the Table 5.2 are presented the results obtained for each model implementation, concerning the T1-weighted sequences, and in Table 5.3 concerning the fluid-sensitive images. It is possible to observe that the original U-Net, where the input has the size of 224×224 pixels, performs better than using the size of 128×128 pixels. Regarding this difference, it is due to the compression of the input image to 128×128. The Attention U-Net has the best results in almost every metric using the size of 224×224 pixels and presented better results when compared with the traditional U-Net. Although the improvements in the segmentation results are not far apart, and the model was originally developed for the segmentation of the pancreas [90], the model performs well. Regarding each metric, the best results are pointed between the different models. It is possible to observe that the results regarding all metrics, except Recall, are better concerning the T1-weighted sequence than the fluid-sensitive sequence. The values for the Recall are possibly influenced by the difficulty of the model in determining the exact contours of the bone. In the case of the fluid-sensitive sequence, there is a larger number of places where there is no bone, but the model identified it (FP), and most of them are encountered within the limits of the bone. Consequently, there is a decrease in the places where there was not bone predicted, but it is bone (FN), preventing the onset of FP in the contours, as it happens in the T1-weighted. Therefore

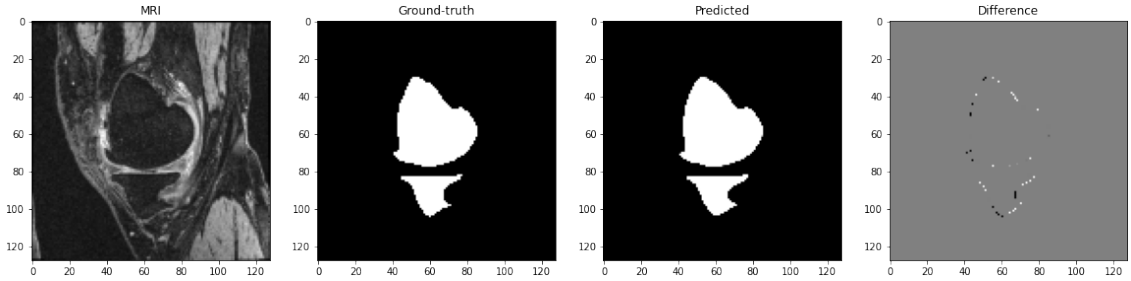
the recall is slightly higher. Concerning the metrics besides Recall, the better results are possibly due to the slightly bigger dataset for the T1-weighted images and the fact that T1-weighted images have a better defined transition/contrast between the bone and adjacent structures, making it easier for the model to distinguish the difference between the anatomical structures.

Although there was no visible overfitting of the data, the models converged in the first few epochs of the training,there was the will to increase the performance of the models motivated the search for the already mentioned approaches of TL. All the results for the segmentation using TL approaches are presented in Table 5.4 concerning the T1-weighted images and in Table 5.5 concerning the fluid-sensitive images. The first step was the implementation of the U-Net slightly altered, as well as the training of the model in the OAI ZIB dataset. Being a large and diverse dataset, the segmentation results proved the robustness of the model, as shown in Table 5.6, achieving high values for all the metrics used. In Figure 5.1, it is possible to visualize three segmentations of the test set for the OAI ZIB dataset and the resultant metrics. First, the usage of the weights obtained from the training of the OAI ZIB dataset in the initialization of the slightly modified U-Net did not show any impact on the results of the usage of this model, showing that the low error values are not dependent on the initialization of the weights. Then, the encoder pre-trained with the OAI ZIB dataset did not perform better than the model only trained with the data from the University Hospital Center of São João. This might be explained by: (i) differences in the image modalities between datasets (the OAI ZIB uses DESS and the data from the University Hospital Center of São João uses T1-weighted and fluid-sensitive images); (ii) the view used in the MRIs (in OAI ZIB have slices in sagittal view and the data from University Hospital Center of São João have slices in coronal view); (iii) adding the difference in the anatomy of the bone between the group of subjects (OAI ZIB have scans from full grown patients and the data from University Hospital Center of São João has scans from children, which presents very different anatomical structures when compared with adults). As a result of using the encoder of a VGG-16 model, pre-trained with the ImageNet dataset, and then transferring the knowledge to a U-Net, were able to obtain results that were close to the best used model for T1-weighted images, the Attention U-Net. The best results were obtained using this method, compared to all others, for fluid-sensitive images. In Appendix A is possible to observe the DSC and IoU coefficient results obtained for all the models implemented for the T1-weighted images and in Appendix B for the fluid-sensitive images.

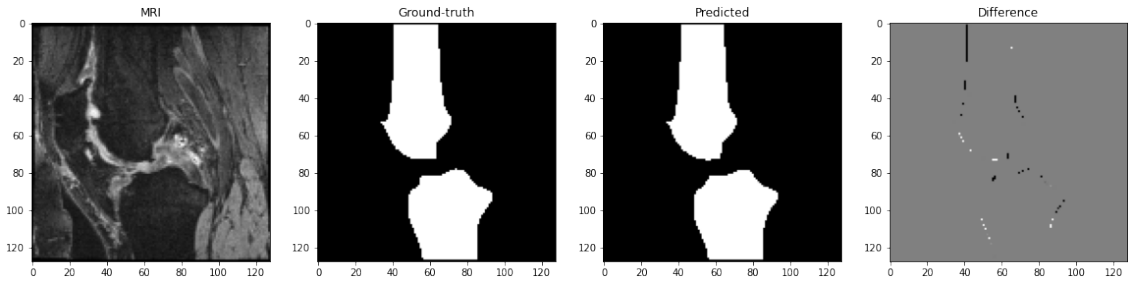
In brief, the best results concerning the segmentation of the dataset from the University Hospital Center of São João are for the similarity coefficients DSC and IoU are, respectively, $93.69 \pm 0.81\%$ and $88.46 \pm 1.23\%$ produced by the Attention U-Net model for the T1-weighted images. For the fluid-sensitive images, the similarity coefficients were $92.02 \pm 0.49\%$ and $85.67 \pm 0.78\%$, produced by the U-Net where the encoder uses the pre-trained encoder from a VGG-16 trained in the ImageNet. Despite the high performance of the models, some aspects impact the achievement of better results. First, the dataset from the University Hospital Center of São João is still considered a small dataset, even after applying some data augmentation techniques. The usage of a bigger dataset will probably make the training more robust, as shown by the segmentation of the



(a) Recall = 99.12%, Precision = 98.81%, F1 Score = 98.96%, DSC = 98.96%, IoU = 97.97%.



(b) Recall = 99.12%, Precision = 98.85%, F1 Score = 98.98%, DSC = 98.98%, IoU = 97.96%.

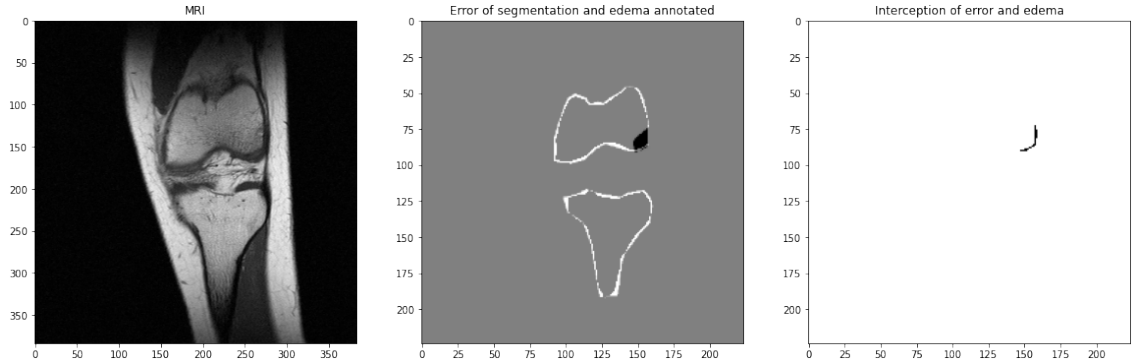


(c) Recall = 98.62%, Precision = 99.57%, F1 Score = 99.10%, DSC = 99.09%, IoU = 98.19%.

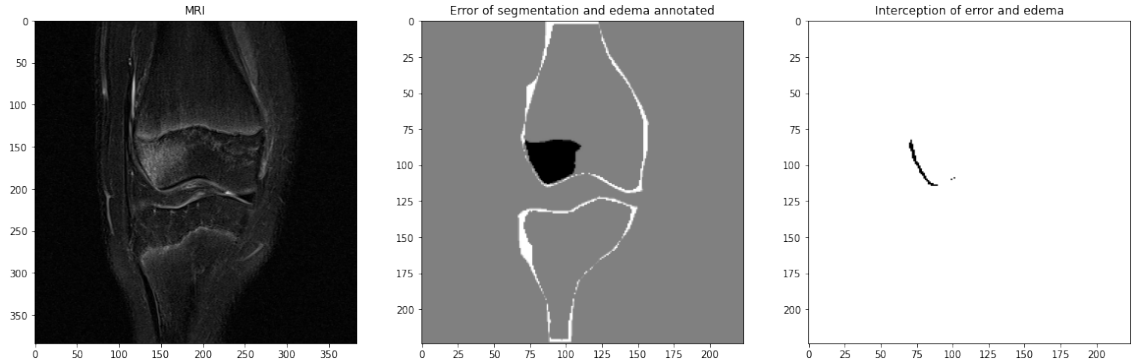
Figure 5.1: Results of three segmentations of the OAI ZIB dataset using the U-Net-S. The results are presented below each image segmentation result.

dataset from OAI ZIB. The second aspect worth mentioning is the variability of the dataset, which is usually a good thing for the model training, but there is an unbalance between the number of examples of each bone anatomical shape. The third aspect that should be mentioned is the need to increase the quality of the images and correspondent annotations to be fed to the model's training.

In Figure 5.3 is possible to observe the results of the segmentation of five images of the test set concerning T1-weighted images, and in Figure 5.4 are presented, the segmentation results of five images of the fluid-sensitive images, regarding the application of the Attention U-Net. It is possible to observe that the model has more difficulty hitting the exact contours of the bones, but they can segment the images with considerable high performance. Concerning the T1-weighted sequence segmentation results, besides the difficulty of segmenting the contours more or less accurately, the model also has difficulty segmenting slices where the anatomical shape has less frequency, as seen in 5.3b. For the fluid-sensitive images, is more flagrant the difficulty of segmenting an image that has less contrast between the structures, with more "homogeneous signal pattern" for the global image and the cortical bone that delimitates the bone/bone marrow is less evident in



(a) From left to right, MRI scan to be segmented, followed by the error resultant of the segmentation of the MRI T1-weighted sequence (represented in white) and the edema findings (represented in black) and the annotation of edema that coincides with the error in segmentation (represented in black).



(b) From left to right, MRI scan to be segmented, followed by the error resultant of the segmentation of the MRI fluid-sensitive sequence (represented in white) and the edema findings (represented in black) and the annotation of edema that coincides with the error in segmentation (represented in black).

Figure 5.2: Evaluation of the influence of edema findings in comparison with the segmentation results of the Attention U-Net for T1-weighted sequence and fluid-sensitive sequence.

this sequence type compared to the T-weighted sequence type 5.4b. Adding to that is the prediction of bone that is not annotated but is expectable, as shown in 5.4c and 5.4e. These difficulties are transversal to the other models.

Finally, after obtaining the segmentation results, it was intended to evaluate whether the ELMSI would have any influence on the segmentation performed by the model, that is, if the edema findings annotated in the slices of the test set were an integral part or not of the error resulting from segmentation. As previously mentioned, the models had more difficulty in segmenting the contours of the bone, where contours of ELMSI can also be found, but as observed in Figure 5.2a for the T1-weighted images and Figure 5.2b for the fluid sensitive sequences, it makes a small part of the total error of the segmentation. It was observed that the edema does not significantly influence the bone segmentation by the models, making on average 7.20% of the error in segmentation for the T1-weighted sequence and 8.01% for the fluid-sensitive sequences.

5.3 Summary

This chapter presents the evaluation of the methods implemented and the discussion of the results obtained. Besides comparison, selects the best model according to the MRI modality. The best model for the segmentation of the T1-weighted images was the Attention U-Net, and the best model for fluid-sensitive images segmentation was the U-Net that had a VGG-16 model encoder, pre-trained with the ImageNet dataset. The results showed high performance in the segmentation task, presenting some difficulties in the segmentation of the exact contours of the bone and in the segmentation of slices where the anatomical shape has less frequency.

Table 5.2: Results for the T1-weighted images for the U-Net and Attention U-Net models. All the metrics are presented using the resultant average ± its standard deviation (σ).

Model	Image shape	Loss	Recall (%)	Precision (%)	F1 Score (%)	DSC (%)	IoU (%)
U-NET	128×128	0.113 ± 0.007	66.68 ± 0.826	**99.75 ± 0.42**	80.00 ± 0.72	88.74 ± 0.71	79.93 ± 0.94
U-NET	224×224	0.066 ± 0.005	81.44 ± 1.22	99.20 ± 0.51	89.45 ± 0.56	93.40 ± 0.48	87.83 ± 0.77
U-NET-S	128×128	0.116 ± 0.010	66.16 ± 1.02	99.60 ± 0.84	79.51 ± 1.01	88.38 ± 1.02	79.43 ± 1.31
Attention U-Net	224×224	**0.063 ± 0.008**	**82.60 ± 0.50**	98.17 ± 1.22	**89.71 ± 0.80**	**93.69 ± 0.81**	**88.46 ± 1.23**

Table 5.3: Results for the fluid-sensitive images for the U-Net and Attention U-Net models. All the metrics are presented using the resultant average ± its standard deviation (σ).

Model	Image shape	Loss	Recall (%)	Precision (%)	F1 Score (%)	DSC (%)	IoU (%)
U-NET	128×128	0.120 ± 0.011	68.83 ± 2.19	97.67 ± 0.70	80.53 ± 1.38	88.04 ± 1.07	79.22 ± 1.60
U-NET 224×224	0.091 ± 0.003	83.00 ± 0.21	95.16 ± 0.09	88.67 ± 0.16	90.86 ± 0.34	**84.12 ± 0.29**	
U-NET-S	128×128	0.127 ± 0.012	67.86 ± 2.00	**97.80 ± 0.83**	79.83 ± 1.38	87.30 ± 1.18	78.13 ± 1.63
Attention U-Net	224×224	**0.090 ± 0.011**	**84.08 ± 1.67**	94.40 ± 2.13	**88.93 ± 0.94**	**90.97 ± 1.08**	83.99 ± 1.73

Table 5.4: Results for the transfer learning approach for the T1-weighted images. All the metrics are presented using the resultant average ± its standard deviation (σ).

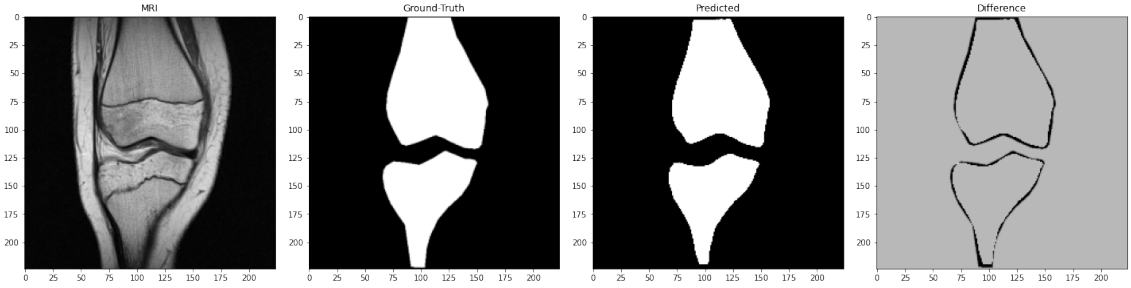
Model	Image shape	Loss	Recall (%)	Precision (%)	F1 Score (%)	DSC (%)	IoU (%)
U-NET-S (Weights)	128×128	**0.113 ± 0.004**	66.38 ± 0.52	**99.66 ± 0.61**	79.77 ± 0.51	88.73 ± 0.45	79.98 ± 0.54
U-NET-S Encoder	128×128	0.121 ± 0.009	66.25 ± 1.26	99.15 ± 0.40	79.44 ± 0.97	87.91 ± 0.89	78.68 ± 1.15
U-NET VGG-16 Encoder	224×224	0.064 ± 0.007	**83.54 ± 0.51**	99.12 ± 0.61	**89.48 ± 0.55**	**93.64 ± 0.66**	**88.31 ± 0.97**

Table 5.5: Results for the transfer learning approach for the fluid-sensitive images. All the metrics are presented using the resultant average ± its standard deviation (σ).

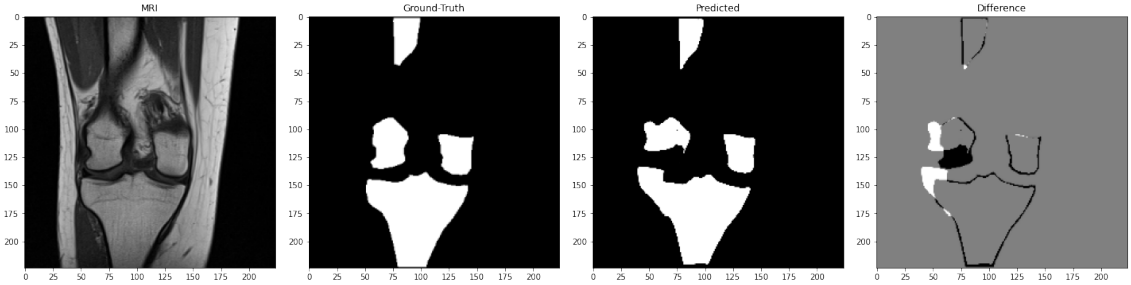
Model	Image shape	Loss	Recall (%)	Precision (%)	F1 Score (%)	DSC (%)	IoU (%)
U-NET-S (Weights)	128×128	0.113 ± 0.014	68.95 ± 1.14	**97.61 ± 1.28**	80.46 ± 1.14	87.73 ± 1.36	78.57 ± 0.33
U-NET-S Encoder	224×224	0.138 ± 0.003	67.51 ± 0.07	96.30 ± 0.12	79.37 ± 0.09	86.20 ± 0.21	76.51 ± 0.45
U-NET VGG-16 Encoder	224×224	**0.080 ± 0.005**	**83.22 ± 1.66**	96.83 ± 0.84	**89.50 ± 0.60**	**92.02 ± 0.49**	**85.67 ± 0.78**

Table 5.6: Results for the OAI ZIB dataset for the slightly modified U-Net. All the metrics are presented using the resultant average ± its standard deviation (σ).

Model	Image shape	Loss	Recall (%)	Precision (%)	F1 Score (%)	DSC (%)	IoU (%)
U-NET-S	128×128	0.0116	98.81	98.87	98.84	98.84	97.70

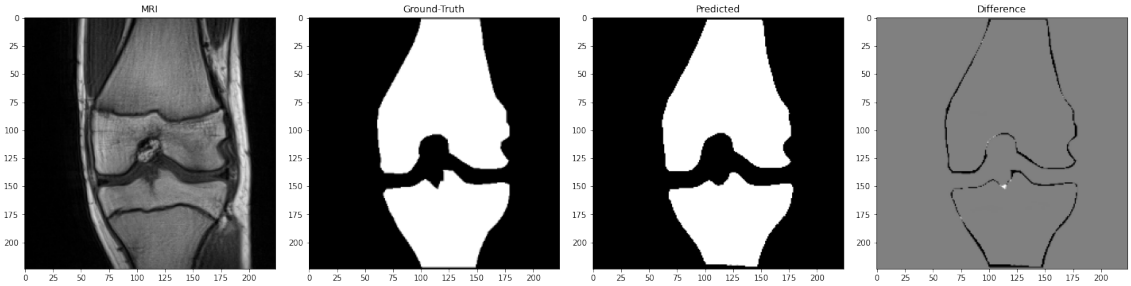


(a) Results of the segmentation of MRI slice T1-weighted sequence number 27: DSC = 97.29%, IoU = 94.71%.

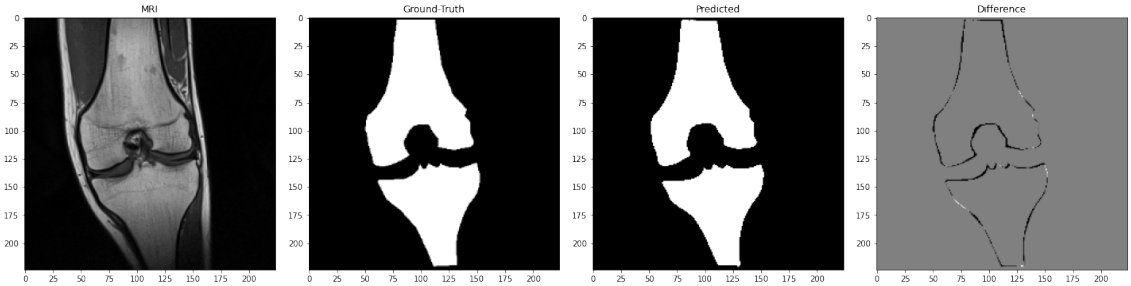


(b) Results of the segmentation of MRI slice T1-weighted sequence number 20: DSC = 75.91%, IoU = 61.20%.

(c) Results of the segmentation of MRI slice T1-weighted sequence number 15: DSC = 95.17%, IoU = 90.80%.

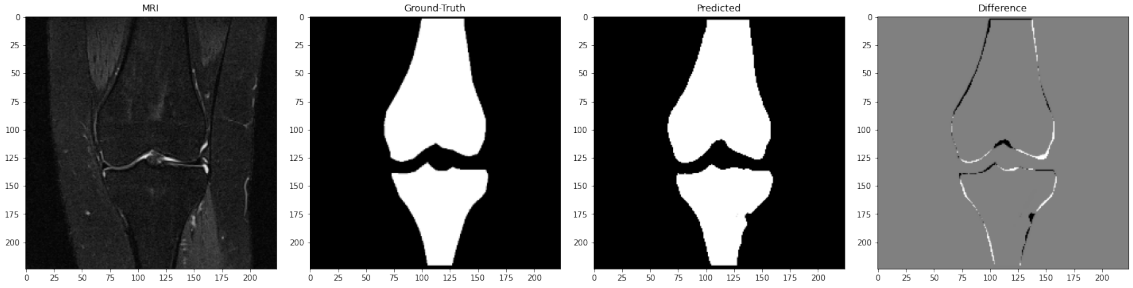


(d) Results of the segmentation of MRI slice T1-weighted sequence number 45: DSC = 85.42%, IoU = 74.55%.

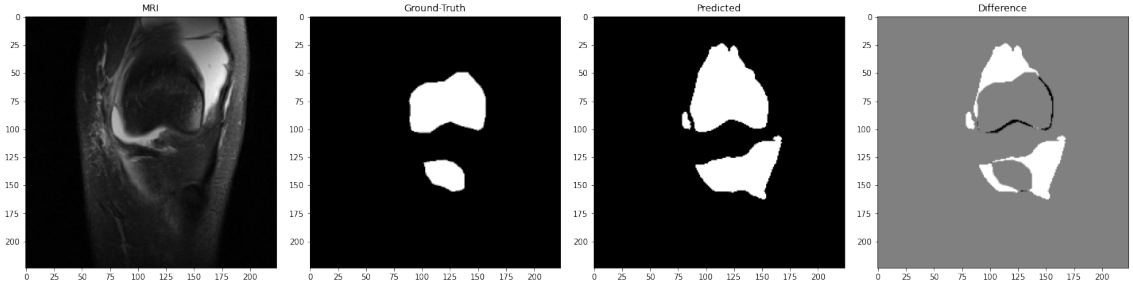


(e) Results of the segmentation of MRI slice T1-weighted sequence number 31: DSC = 84.03%, IoU = 72.50%.

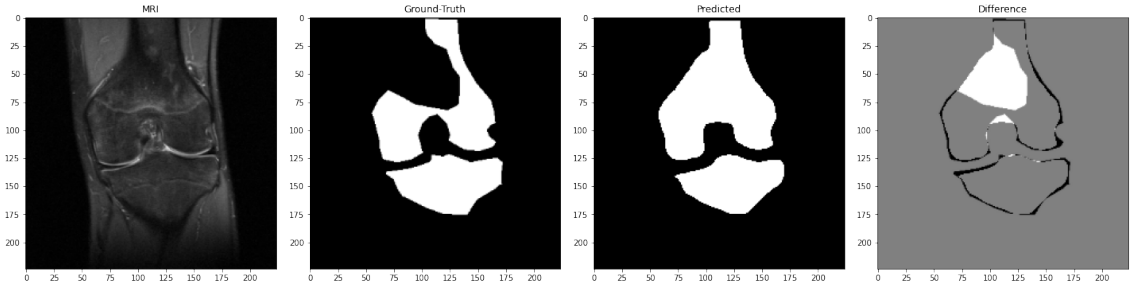
Figure 5.3: Segmentation results using the Attention U-Net concerning five T1-weighted images from the data provided by University Hospital Center of São João.



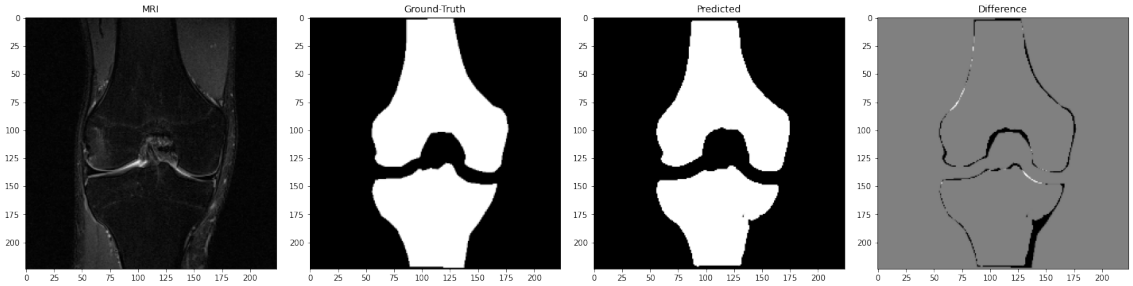
(a) Results of the segmentation of MRI slice fluid-sensitive sequence number 11: DSC = 97.22%, IoU = 94.60%.



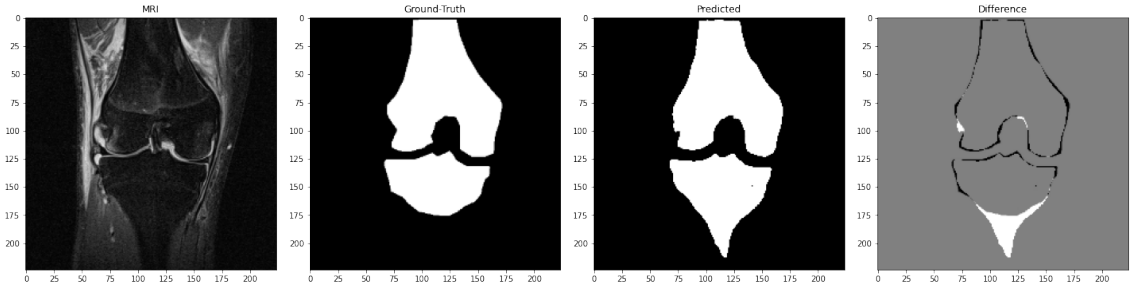
(b) Results of the segmentation of MRI slice fluid-sensitive sequence number number 24: DSC = 69.79%, IoU = 53.59%.



(c) Results of the segmentation of MRI slice fluid-sensitive sequence number number 37: DSC = 82.64%, IoU = 70.40%.



(d) Results of the segmentation of MRI slice fluid-sensitive sequence number number 47: DSC = 96.16%, IoU = 92.65%.



(e) Results of the segmentation of MRI slice fluid-sensitive sequence number 61: DSC = 93.41%, IoU = 87.62%.

Figure 5.4: Segmentation results using the Attention U-Net concerning five fluid-sensitive sequences from the data provided by University Hospital Center of São João.

Chapter 6

Conclusion and Future work

Bone marrow is one of the largest human tissues with different constitutions during human growth and is afflicted by multiple pathologies, EMLSI one of them. EMLSI describes an observed alteration of the signal produced in the MRI modalities of T1-weighted and fluid-sensitive sequences images and is a condition associated with several diagnoses. It is an important aspect for the evaluation of the progressiveness of some degenerative diseases and is most of the times associated with pain. With the easement of the task of health professionals of identifying the presence of EMLSI in MRIs, the implementation of a Computer-aided diagnosis system has the most significant importance. For the development of this system, the first step is to segment performance and accurately the bone in the MRI slices.

This thesis proposed the implementation of DL architectures and TL techniques to achieve the Bone Segmentation goal in the dataset provided by the University Hospital Center of São João. The models implemented were a U-Net, a slightly modified U-Net and an Attention U-Net. For the techniques of Transfer Learning, first, only the weights from a pre-trained U-Net in another similar dataset were used, followed by the usage of only the encoder part of that pre-trained U-Net and finally, the usage of the encoder of a VGG-16 pre-trained in the Image Net dataset in a U-Net model.

The best results for the similarity coefficients DSC and IoU are, respectively, $93.69 \pm 0.81\%$ and $88.46 \pm 1.23\%$ produced by the Attention U-Net model for the T1-weighted images and $92.02 \pm 0.49\%$ and for the fluid-sensitive images $85.67 \pm 0.78\%$, produced by the U-Net where the encoder uses the pre-trained encoder from a VGG-16 trained in the ImageNet. The difference in the results concerning the two modalities is possibly associated with the slightly better identified transition between the bone and the rest of the knee surrounding structures.

Even though the relevant results obtained for the bone segmentation problem have shown that it is possible to segment the bone of a small dataset with the characteristics presented and its usage for training the models, some aspects can increase the efficiency of the models and increase the segmentation metrics. Increasing the dataset size will benefit the model's training, especially by decreasing the discrepancy of the bone anatomical shape to help increase the performance of the models. The thesis performed the application of models that accept 2D inputs, but some

available models take into consideration the 3D nature of the bone. Another approach to be used in future can be the development of a methodology that uses both image modalities to perform the segmentation, instead of one for each, due to its interest in the future CAD system.

Appendix A

Segmentation Results - T1-weighted images

Table A.1: Results of the segmentation of the T1-weighted images test set to all the implemented models.

	U-Net (128×128)		**U-Net (224×224)**		**U-Net-S**		**Attention U-Net**	
	DSC %	**IoU %**	**DSC %**	**IoU %**	**DSC %**	**IoU %**	**DSC %**	**IoU %**
1	78.80	65.04	86.94	76.94	87.92	78.43	96.17	92.62
2	90.63	82.85	86.68	76.48	89.14	80.47	86.97	76.95
3	84.91	73.78	92.36	85.82	83.18	71.24	94.14	88.90
4	83.85	72.15	91.73	84.77	84.27	72.88	93.69	88.13
5	93.31	87.44	96.51	93.26	90.38	82.47	96.00	92.31
6	90.97	83.46	93.88	88.46	90.67	82.93	95.14	90.74
7	74.69	59.62	87.97	78.58	72.74	57.19	86.66	76.54
8	90.53	82.68	95.68	91.71	90.58	82.83	96.39	93.03
9	87.33	77.51	90.14	82.06	87.05	77.10	93.42	87.65
10	54.19	37.01	79.85	66.42	36.74	22.68	60.67	43.75
11	90.39	82.45	94.46	89.49	89.42	80.81	95.39	91.21
12	92.92	86.79	96.29	92.85	91.27	83.98	96.55	93.33
13	92.27	85.66	92.71	86.36	92.60	86.23	94.99	90.45
14	84.01	72.52	88.94	80.09	83.27	71.39	92.33	85.77
15	78.08	64.05	87.52	77.82	77.54	63.40	89.15	80.44
16	90.82	83.19	94.85	90.23	90.24	82.21	95.73	91.82
17	91.02	83.52	95.31	91.05	89.63	81.18	96.71	93.62
18	89.61	81.18	95.38	91.20	89.37	80.81	96.23	92.73
19	82.44	70.09	88.49	79.37	81.98	69.44	90.28	82.26
20	70.18	54.01	90.47	82.55	63.90	46.92	75.91	61.20

21	90.29	82.32	95.63	91.62	88.81	79.89	96.32	92.92
22	87.63	78.03	77.33	63.05	87.04	77.07	90.99	83.51
23	90.02	81.88	94.88	90.26	89.12	80.41	94.94	90.36
24	90.00	81.84	91.33	84.10	89.67	81.28	93.27	87.39
25	90.86	83.26	96.12	92.51	89.75	81.41	94.81	90.13
26	90.36	82.42	94.23	89.07	89.56	81.04	94.94	90.37
27	92.76	86.50	96.67	93.54	92.64	86.34	97.29	94.71
28	90.71	83.01	95.24	90.94	88.65	79.58	95.17	90.80
29	86.56	76.30	91.41	84.20	85.56	74.81	94.45	89.51
30	90.75	83.05	95.39	91.19	90.78	83.08	94.90	90.30
31	80.62	67.47	84.56	73.25	68.75	52.45	84.03	72.50
32	89.26	80.59	93.65	88.08	88.06	78.67	95.32	91.06
33	90.58	82.78	94.55	89.64	90.43	82.49	95.54	91.47
34	90.57	82.80	95.30	91.02	90.46	82.57	95.61	91.59
35	89.80	81.51	93.60	87.98	89.42	80.87	95.78	91.91
36	92.58	86.18	96.17	92.61	90.89	83.30	96.38	93.02
37	90.01	81.84	95.08	90.63	89.00	80.23	95.70	91.76
38	91.22	83.88	94.17	88.96	90.90	83.28	94.86	90.24
39	93.49	87.79	97.08	94.34	93.11	87.06	97.08	94.32
40	82.61	70.32	91.04	83.56	78.30	64.45	90.20	82.13
41	93.06	87.00	95.50	91.35	93.10	87.06	94.80	90.13
42	90.32	82.33	94.19	89.01	89.93	81.75	95.32	91.06
43	91.02	83.55	95.09	90.67	90.94	83.38	96.25	92.79
44	88.43	79.27	92.48	85.99	85.80	75.14	89.83	81.54
45	89.80	81.48	92.42	85.92	89.52	81.04	85.42	74.55
46	92.26	85.61	96.14	92.59	89.58	81.11	93.59	87.97
47	87.15	77.24	91.40	84.16	86.62	76.39	91.65	84.60
48	69.95	53.79	74.56	59.47	77.77	63.61	89.22	80.61
49	92.82	86.61	96.90	93.99	92.25	85.60	97.07	94.30
50	90.67	82.94	95.62	91.60	88.23	78.94	95.82	91.98
51	90.75	83.07	94.49	89.56	90.32	82.37	95.48	91.34
52	89.68	81.28	95.51	91.40	89.22	80.55	96.06	92.41
53	84.17	72.68	91.26	83.97	83.23	71.31	91.43	84.28
54	89.95	81.75	94.17	88.99	90.68	82.93	95.88	92.08
55	92.36	85.82	95.57	91.51	92.38	85.82	96.65	93.53
56	91.35	84.05	95.08	90.62	90.70	82.99	96.09	92.47
57	90.30	82.34	94.78	90.08	89.86	81.58	95.84	92.00
58	93.12	87.15	96.26	92.79	91.33	84.03	96.46	93.17
59	89.26	80.61	94.18	88.99	88.77	79.87	95.59	91.57
60	81.34	68.56	77.68	63.54	82.77	70.58	87.11	77.14

61	89.32	80.66	92.05	85.29	82.42	70.11	94.28	89.16
62	91.45	84.28	95.25	90.95	89.96	81.80	95.58	91.54
63	91.72	84.68	96.67	93.56	92.06	85.29	97.06	94.29
64	81.39	68.69	91.38	84.12	80.23	67.06	91.66	84.58
65	89.79	81.47	95.11	90.68	90.01	81.80	95.69	91.74
66	91.20	83.84	96.27	92.82	90.49	82.64	94.84	90.19
67	91.48	84.32	95.31	91.04	91.01	83.50	95.45	91.30
68	86.35	75.97	92.78	86.53	86.77	76.62	92.66	86.29
69	90.41	82.50	94.52	89.62	89.14	80.42	94.83	90.18
70	75.22	60.31	90.24	82.24	73.31	57.96	73.65	58.29
71	89.95	81.74	95.40	91.24	89.78	81.52	96.44	93.12
72	90.67	82.97	95.92	92.14	89.94	81.73	96.46	93.16
AVG	87.75	78.71	92.69	86.70	86.68	77.24	93.11	87.59

Table A.2: Results of the segmentation of the T1-weighted images test set to all the transfer learning models implemented.

	U-Net-S (Weights)		**U-Net-S (Encoder)**		**U-Net VGG16 (Encoder)**	
	DSC %	**IoU %**	**DSC %**	**IoU %**	**DSC %**	**IoU %**
1	88.93	80.11	84.06	72.52	84.61	73.34
2	90.62	82.86	91.37	84.10	96.04	92.38
3	85.91	75.34	84.63	73.33	94.01	88.70
4	85.23	74.35	67.87	51.27	93.33	87.51
5	92.53	86.11	93.17	87.20	96.71	93.64
6	91.12	83.68	84.55	73.24	94.86	90.24
7	75.92	61.18	74.91	59.76	89.19	80.56
8	91.37	84.11	91.15	83.75	95.71	91.79
9	88.18	78.87	88.63	79.62	93.16	87.18
10	49.81	33.53	46.80	30.91	71.36	55.54
11	90.39	82.47	88.95	80.05	95.30	91.04
12	92.69	86.39	92.69	86.34	96.84	93.87
13	92.80	86.58	90.41	82.58	93.09	87.05
14	83.47	71.64	79.31	65.65	90.16	82.10
15	76.85	62.36	75.86	61.12	86.33	75.94
16	90.51	82.66	87.41	77.64	94.70	89.96
17	90.93	83.39	90.81	83.16	95.51	91.40
18	90.07	81.96	89.40	80.85	95.41	91.24
19	83.89	72.24	82.61	70.37	91.54	84.40
20	73.02	57.51	69.18	52.89	86.72	76.56

21	90.48	82.60	91.00	83.46	96.51	93.27
22	89.07	80.31	80.38	67.20	95.32	91.05
23	90.31	82.34	90.06	81.93	94.50	89.58
24	90.99	83.45	91.67	84.61	95.35	91.11
25	90.23	82.22	89.82	81.52	95.68	91.75
26	90.58	82.78	89.19	80.45	94.41	89.45
27	92.91	86.75	91.61	84.51	96.70	93.60
28	88.50	79.34	88.03	78.58	95.38	91.17
29	87.39	77.64	87.16	77.26	93.13	87.15
30	90.49	82.61	88.05	78.62	94.17	89.00
31	74.60	59.57	61.25	44.18	67.66	51.13
32	88.98	80.16	88.57	79.48	94.52	89.60
33	90.88	83.28	90.75	83.07	94.90	90.30
34	90.61	82.85	90.54	82.70	96.05	92.41
35	90.18	82.14	88.48	79.34	94.94	90.38
36	92.15	85.43	92.25	85.60	96.44	93.12
37	89.60	81.17	90.73	83.04	95.92	92.16
38	90.54	82.75	90.76	83.13	95.90	92.14
39	93.84	88.40	93.64	88.05	97.17	94.49
40	80.12	66.94	82.51	70.21	90.18	82.14
41	93.02	86.95	92.34	85.77	95.35	91.12
42	89.62	81.21	89.16	80.39	95.29	91.00
43	90.19	82.12	90.58	82.82	92.58	86.23
44	85.22	74.26	80.29	67.10	92.33	85.73
45	89.99	81.77	85.95	75.40	89.74	81.37
46	92.70	86.40	92.79	86.62	96.99	94.14
47	88.02	78.58	86.23	75.79	91.69	84.66
48	77.47	63.11	81.18	68.31	80.03	66.76
49	92.55	86.09	91.81	84.85	96.80	93.77
50	91.06	83.58	89.10	80.35	95.71	91.77
51	90.50	82.67	84.62	73.31	94.69	89.91
52	90.57	82.78	88.80	79.84	95.33	91.10
53	82.73	70.60	81.46	68.73	92.56	86.17
54	89.70	81.32	89.32	80.72	93.97	88.63
55	93.25	87.33	92.13	85.41	96.53	93.29
56	91.39	84.18	90.91	83.31	96.06	92.40
57	90.13	82.05	89.58	81.15	95.86	92.03
58	92.68	86.35	88.61	79.56	96.62	93.48
59	88.73	79.75	88.60	79.54	94.65	89.86
60	88.32	79.15	86.37	76.06	83.40	71.53

61	87.82	78.30	87.08	77.13	93.27	87.41
62	91.51	84.38	90.14	82.09	95.01	90.48
63	91.82	84.89	90.83	83.21	96.95	94.07
64	81.48	68.77	81.03	68.11	92.05	85.31
65	89.16	80.47	88.54	79.40	94.98	90.42
66	91.49	84.34	91.99	85.20	94.52	89.60
67	91.49	84.31	91.16	83.75	95.22	90.88
68	88.67	79.62	85.63	74.94	94.02	88.70
69	90.14	82.06	90.20	82.21	94.55	89.67
70	77.37	63.12	74.61	59.64	90.58	82.84
71	90.46	82.61	89.59	81.14	96.20	92.68
72	91.81	84.85	91.20	83.85	95.86	92.05
AVG	88.02	79.14	86.42	76.79	93.18	87.62

Appendix B

Segmentation Results - fluid-sensitive images

Table B.1: Results of the segmentation of the fluid-sensitive images test set to all the implemented models.

	U-Net (128x128)		**U-Net (224x224)**		**U-Net-S**		**Attention U-Net**	
	DSC %	**IoU %**	**DSC %**	**IoU %**	**DSC %**	**IoU %**	**DSC %**	**IoU %**
1	91.28	83.97	95.08	90.59	90.63	82.91	94.93	90.33
2	89.58	81.14	92.57	86.17	90.51	82.66	81.90	69.31
3	90.48	82.63	94.68	89.89	91.02	83.49	94.75	90.02
4	92.04	85.28	96.29	92.85	92.00	85.19	91.91	85.04
5	93.41	87.61	96.27	92.82	92.80	86.58	96.25	92.78
6	76.09	61.49	84.32	72.87	76.70	62.12	85.70	74.99
7	83.88	72.24	87.63	77.96	85.65	74.94	87.58	77.89
8	94.34	89.33	96.45	93.12	95.10	90.64	92.63	86.26
9	89.43	80.91	90.12	82.02	91.82	84.88	95.25	90.92
10	94.81	90.15	96.95	94.07	94.38	89.36	93.84	88.39
11	95.33	91.07	97.41	94.94	93.78	88.29	97.22	94.60
12	89.25	80.57	92.20	85.54	90.43	82.53	79.60	66.13
13	92.43	85.91	88.19	78.88	89.94	81.72	91.34	84.05
14	58.94	41.83	62.31	45.27	55.11	38.09	66.50	49.78
15	94.13	88.92	88.88	79.99	93.81	88.30	96.02	92.36
16	85.50	74.70	89.08	80.31	86.28	75.90	88.94	80.07
17	88.63	79.63	92.33	85.73	85.60	74.87	92.93	86.79
18	91.64	84.60	94.48	89.53	89.26	80.61	95.25	90.95
19	71.55	55.72	83.74	72.04	78.18	64.24	83.04	71.02
20	89.62	81.20	93.04	87.02	87.39	77.63	94.12	88.89

21	83.51	71.64	89.67	81.27	82.23	69.83	80.33	67.13
22	90.46	82.60	94.73	89.98	90.74	83.09	94.43	89.44
23	79.77	66.44	68.27	51.82	71.23	55.35	73.68	58.32
24	78.45	64.53	71.84	56.04	74.58	59.59	69.79	53.59
25	93.20	87.28	96.78	93.75	87.57	77.86	92.29	85.68
26	84.57	73.28	90.27	82.24	84.29	72.83	84.26	72.80
27	94.81	90.14	95.15	90.75	94.82	90.12	95.04	90.53
28	89.85	81.55	92.56	86.16	88.35	79.20	93.12	87.13
29	92.30	85.73	96.62	93.44	92.23	85.58	95.95	92.21
30	92.34	85.79	93.17	87.18	92.82	86.57	94.88	90.26
31	91.42	84.23	95.55	91.52	89.68	81.30	92.74	86.45
32	90.65	82.93	92.04	85.27	89.71	81.37	93.71	88.20
33	92.77	86.51	94.68	89.89	93.01	86.96	93.12	87.13
34	72.64	57.01	74.88	59.83	69.32	53.11	87.58	77.91
35	91.25	83.92	94.94	90.35	91.23	83.91	96.48	93.19
36	81.02	68.08	82.67	70.46	88.17	78.77	85.17	74.17
37	84.82	73.68	84.60	73.30	78.96	65.20	82.64	70.40
38	90.80	83.15	94.53	89.63	89.92	81.72	94.92	90.33
39	92.41	85.90	95.26	90.94	91.36	84.10	95.90	92.11
40	91.40	84.20	94.95	90.36	92.11	85.38	95.20	90.86
41	89.37	80.76	94.79	90.10	89.59	81.20	95.70	91.77
42	92.10	85.35	95.31	91.02	92.83	86.63	94.89	90.29
43	69.24	52.93	70.51	54.39	65.88	49.15	81.74	69.08
44	85.82	75.19	90.18	82.13	86.00	75.44	91.27	83.95
45	92.02	85.20	95.55	91.49	91.72	84.72	88.49	79.35
46	87.28	77.43	92.36	85.82	80.90	67.87	83.40	71.52
47	92.47	86.00	95.32	91.07	92.10	85.38	96.19	92.65
48	92.40	85.87	96.13	92.55	91.40	84.15	96.15	92.59
49	85.95	75.45	87.98	78.52	84.39	73.06	79.74	66.33
50	82.67	70.51	89.30	80.63	78.58	64.79	78.92	65.15
51	91.44	84.24	94.83	90.17	90.50	82.61	92.33	85.75
52	86.98	77.02	90.35	82.39	83.54	71.84	91.08	83.62
53	93.37	87.58	94.80	90.12	92.98	86.90	95.85	92.05
54	92.64	86.30	96.56	93.35	92.18	85,51	96.50	93.23
55	93.15	87.15	94.83	90.17	92.04	85.25	96.65	93.53
56	82.05	69.58	88.62	79.57	82.80	70.65	85.31	74.42
57	93.41	87.63	96.03	92.33	93.09	87.09	96.33	92.90
58	90.28	82.28	93.79	88.31	88.91	80.03	90.76	83.12
59	92.58	86.17	96.62	93.47	91.49	84.32	95.00	90.46
60	89.36	80.73	92.50	86.02	85.08	74.00	85.88	75.25

61	90.30	82.35	94.21	89.08	92.38	85.84	93.41	87.62
62	81.59	69.02	88.66	79.64	80.85	67.98	89.82	81.56
63	84.82	73.68	91.78	84.83	82.55	70.35	91.60	84.52
64	95.36	91.09	97.25	94.64	93.89	88.49	92.44	85.96
65	88.27	78.99	92.86	86.69	88.31	79.00	89.00	80.17
66	81.16	68.32	87.61	77.98	84.33	72.93	88.53	79.48
67	83.78	72.08	93.35	87.55	91.29	84.02	93.03	86.97
68	92.12	85.40	92.94	86.82	90.71	82.97	93.03	86.96
69	91.68	84.60	95.20	90.84	90.20	82.14	94.81	90.14
70	88.27	79.07	92.14	85.39	87.50	77.77	85.80	75.12
AVG	88.04	79.22	91.11	84.33	87.30	78.13	90.21	82.77

Table B.2: Results of the segmentation of the T2 fluid-sensitive images test set to all the transfer learning models implemented.

	U-Net-S (Weights)		**U-Net-S (Encoder)**		**U-Net VGG16 (Encoder)**	
	DSC %	**IoU %**	**DSC %**	**IoU %**	**DSC %**	**IoU %**
1	91.79	84.83	91.60	84.51	94.54	89.64
2	91.15	83.73	88.47	79.32	91.43	84.20
3	91.79	84.80	88.97	80.20	95.98	92.28
4	91.85	84.96	92.05	85.26	96.09	92.47
5	93.34	87.52	93.06	87.03	96.50	93.24
6	77.28	62.96	75.78	61.18	87.22	77.34
7	84.24	72.81	83.00	70.94	87.48	77.74
8	94.83	90.19	93.89	88.50	96.92	94.04
9	91.61	84.50	88.84	79.87	92.96	86.85
10	94.56	89.68	94.57	89.71	97.10	94.37
11	94.37	89.35	94.83	90.20	97.60	95.32
12	89.48	80.91	86.75	76.61	93.61	87.97
13	89.99	81.79	86.36	76.02	88.53	79.43
14	58.49	41.37	50.83	34.19	65.36	48.56
15	94.18	88.99	93.43	87.70	90.67	82.94
16	86.78	76.67	83.77	72.11	91.37	84.17
17	83.23	71.27	87.63	77.95	88.53	79.39
18	91.09	83.62	89.64	81.22	94.91	90.31
19	74.97	59.93	67.49	50.96	86.16	75.64
20	89.62	81.21	89.48	80.89	94.38	89.35
21	83.44	71.66	83.33	71.36	89.07	80.31
22	90.95	83.42	89.76	81.45	95.38	91.17

23	79.13	65.62	77.28	62.94	80.48	67.32
24	74.40	59.25	73.00	57.54	81.59	68.88
25	88.87	79.97	89.42	80.84	96.42	93.08
26	81.58	68.96	82.74	70.65	86.96	76.95
27	94.07	88.83	94.55	89.65	95.70	91.74
28	88.58	79.51	88.30	79.06	93.39	87.59
29	92.16	85.40	91.31	84.06	96.12	92.51
30	92.88	86.70	91.73	84.80	96.60	93.43
31	91.04	83.56	90.33	82.43	95.75	91.85
32	88.74	79.76	87.88	78.43	95.51	91.41
33	92.34	85.77	92.63	86.28	94.32	89.25
34	70.81	54.81	64.27	47.47	87.92	78.40
35	90.98	83.47	91.68	84.55	95.98	92.27
36	86.23	75.82	83.28	71.38	88.25	79.01
37	82.87	70.82	79.40	65.88	88.34	79.11
38	90.33	82.36	89.02	80.24	95.53	91.48
39	93.35	87.54	93.07	87.04	95.85	92.02
40	91.66	84.61	91.68	84.64	95.57	91.52
41	90.57	82.77	87.06	77.12	94.37	89.32
42	92.50	86.06	92.27	85.63	95.87	92.08
43	67.00	50.44	55.86	38.90	80.41	67.19
44	86.35	76.04	86.24	75.76	91.55	84.44
45	91.94	85.07	91.36	84.11	96.11	92.51
46	84.77	73.64	81.69	68.97	92.64	86.31
47	92.28	85.68	91.91	85.02	96.27	92.81
48	91.18	83.78	91.68	84.64	96.79	93.80
49	80.29	67.25	82.68	70.41	85.96	75.42
50	82.13	69.65	79.72	66.24	85.51	74.73
51	90.85	83.25	91.05	83.51	95.84	92.01
52	84.89	73.77	84.78	73.55	92.76	86.49
53	93.52	87.83	93.79	88.33	97.06	94.30
54	92.53	86.09	91.79	84.87	96.47	93.16
55	92.40	85.89	92.10	85.33	96.93	94.04
56	85.05	74.09	85.36	74.39	90.88	83.27
57	93.61	88.00	92.94	86.86	95.75	91.85
58	90.60	82.81	90.11	82.03	87.52	77.82
59	90.84	83.23	92.17	85.53	96.68	93.57
60	87.84	78.31	87.52	77.85	92.71	86.42
61	91.57	84.45	89.18	80.42	93.30	87.42
62	82.89	70.83	83.55	71.83	91.34	84.07

63	88.62	79.57	75.26	60.17	92.07	85.36
64	94.27	89.17	94.33	89.29	97.43	94.99
65	87.28	77.40	86.38	75.99	94.57	89.72
66	83.15	71.19	82.73	70.57	88.96	80.09
67	90.03	81.89	78.71	64.92	92.88	86.70
68	91.51	84.38	90.14	82.06	94.67	89.85
69	89.70	81.30	91.73	84.67	96.28	92.82
70	84.93	73.75	85.02	73.95	93.72	88.20
AVG	87.77	78.81	86.35	76.83	92.36	86.22

References

[1] Arya Minaie, Katherine Woolley, Ryan Smith, Olohirere Ezomo, Richard Feinn, Bernadette Mele, Thomas Martin, Juan Garbolosa, and Karen M Myrick. Detecting bone marrow edema with magnetic resonance spectroscopy: A brief report. *The Journal for Nurse Practitioners*, 16(8):e129–e132, 2020.

[2] Dimitrios C Karampinos, Stefan Ruschke, Michael Dieckmeyer, Maximilian Diefenbach, Daniela Franz, Alexandra S Gersing, Roland Krug, and Thomas Baum. Quantitative mri and spectroscopy of bone marrow. *Journal of Magnetic Resonance Imaging*, 47(2):332–353, 2018.

[3] AJ Wilson, WA Murphy, DC Hardy, and WG Totty. Transient osteoporosis: transient bone marrow edema? *Radiology*, 167(3):757–760, 1988.

[4] Davide Maraghelli, Maria Luisa Brandi, Marco Matucci Cerinic, Anna Julie Peired, and Stefano Colagrande. Edema-like marrow signal intensity: a narrative review with a pictorial essay. *Skeletal Radiology*, 50(4):645–663, 2021.

[5] Lucy A Fowkes and Andoni P Toms. Bone marrow oedema of the knee. *The knee*, 17(1):1–6, 2010.

[6] Anastasios V Korompilias, Apostolos H Karantanas, Marios G Lykissas, and Alexandros E Beris. Bone marrow edema syndrome. *Skeletal radiology*, 38(5):425–436, 2009.

[7] Kyung-Hoi Koo, In-Oak Ahn, Rokho Kim, Hae-Ryong Song, Soon-Taek Jeong, Jae-Boem Na, Yong-Sik Kim, and Se-Hyun Cho. Bone marrow edema and associated pain in early stage osteonecrosis of the femoral head: prospective study with serial mr images. *Radiology*, 213(3):715–722, 1999.

[8] Hiroshi Fujita. Ai-based computer-aided diagnosis (ai-cad): the latest review to read first. *Radiological physics and technology*, 13(1):6–19, 2020.

[9] Dinesh D Patil and Sonal G Deore. Medical image segmentation: a review. *International Journal of Computer Science and Mobile Computing*, 2(1):22–27, 2013.

[10] Rania Almajalid, Juan Shan, Maolin Zhang, Garrett Stonis, and Ming Zhang. Knee bone segmentation on three-dimensional mri. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1725–1730, 2019. doi:10.1109/ICMLA.2019.00280.

[11] Michael YM Chen, Thomas Lee Pope, and David J Ott. *Basic radiology*. McGraw-Hill Medical, 2011.

[12] Drew A Torigian and Parvati Ramchandani. *Radiology Secrets Plus E-Book*. Elsevier Health Sciences, 2016.

[13] Reza Azad, Nika Khosravi, Mohammad Dehghanmanshadi, Julien Cohen-Adad, and Dorit Merhof. Medical image segmentation on mri images with missing modalities: A review, 2022. URL: https://arxiv.org/abs/2203.06217, doi:10.48550/ARXIV.2203.06217.

[14] JB Vogler 3rd and Wil1iam A Murphy. Bone marrow imaging. *Radiology*, 168(3):679–693, 1988.

[15] B. Koffman. "what is bone marrow?". Accessed on 23- Jul- 2022. URL: https://cllsociety.org/2016/09/what-is-bone-marrow/.

[16] Judy S. Blebea, Mohamed Houseni, Drew A. Torigian, Chengzhong Fan, Ayse Mavi, Ying Zhuge, Tad Iwanaga, Shipra Mishra, Jay Udupa, Jiyuan Zhuang, Rohit Gopal, and Abass Alavi. Structural and functional imaging of normal bone marrow and evaluation of its age-related changes. *Seminars in Nuclear Medicine*, 37(3):185–194, 2007. Normal Functional and Structural Variations: PET, CT, and MR Imaging (Part II). URL: https://www.sciencedirect.com/science/article/pii/S0001299807000189, doi:https://doi.org/10.1053/j.semnuclmed.2007.01.002.

[17] S. EUSTACE, C. KEOGH, M. BLAKE, R.J. WARD, P.D. ODER, and M. DIMASI. Mr imaging of bone oedema: Mechanisms and interpretation: Pictorial review. *Clinical Radiology*, 56(1):4–12, 2001. URL: https://www.sciencedirect.com/science/article/pii/S0009926000905853, doi:https://doi.org/10.1053/crad.2000.0585.

[18] Lucy A. Fowkes and Andoni P. Toms. Bone marrow oedema of the knee. *The Knee*, 17(1):1–6, 2010. URL: https://www.sciencedirect.com/science/article/pii/S0968016009001100, doi:https://doi.org/10.1016/j.knee.2009.06.002.

[19] FM McQueen. A vital clue to deciphering bone pathology: Mri bone oedema in rheumatoid arthritis and osteoarthritis, 2007.

[20] W.A. Thiryayi, S.A. Thiryayi, and A.J. Freemont. Histopathological perspective on bone marrow oedema, reactive bone change and haemorrhage. *European Journal of Radiology*, 67(1):62–67, 2008. Bone Marrow Edema. URL: https://www.sciencedirect.com/science/article/pii/S0720048X08000971, doi:https://doi.org/10.1016/j.ejrad.2008.01.056.

[21] CR Pal, AD Tasker, SJ Ostlere, and MS Watson. Heterogeneous signal in bone marrow on mri of children's feet: a normal finding? *Skeletal radiology*, 28(5):274–278, 1999.

[22] Kathleen M Lazzarini, Robert N Troiano, and Robert C Smith. Can running cause the appearance of marrow edema on mr images of the foot and ankle? *Radiology*, 202(2):540–542, 1997.

[23] Arya Minaie, Katherine Woolley, Ryan Smith, Olohirere Ezomo, Richard Feinn, Bernadette Mele, Thomas Martin, Juan Garbolosa, and Karen M. Myrick. Detecting bone marrow edema with magnetic resonance spectroscopy: A brief report. *The Journal for Nurse Practitioners*, 16(8):e129–e132, 2020. URL: https://

www.sciencedirect.com/science/article/pii/S1555415520302610, doi:https://doi.org/10.1016/j.nurpra.2020.05.004.

[24] Siegfried Hofmann, Josef Kramer, Anosheh Vakil-Adli, Nicolas Aigner, and Martin Breitenseher. Painful bone marrow edema of the knee: differential diagnosis and therapeutic concepts. *Orthopedic Clinics*, 35(3):321–333, 2004.

[25] Pei Yang, Chuanbao Bian, Xin Huang, Anli Shi, Chunsheng Wang, and Kunzheng Wang. Core decompression in combination with nano-hydroxyapatite/polyamide 66 rod for the treatment of osteonecrosis of the femoral head. *Archives of orthopaedic and trauma surgery*, 134(1):103–112, 2014.

[26] Rajeshwar Dass and Swapna Devi. Image segmentation techniques 1. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[27] Gopinath Rebala, Ajay Ravi, and Sanjay Churiwala. Machine learning definition and basics. In *An Introduction to Machine Learning*, pages 1–17. Springer, 2019.

[28] Priyanka Malhotra, Sheifali Gupta, Deepika Koundal, Atef Zaguia, and Wegayehu Enbeyle. Deep neural networks for medical image segmentation. *Journal of Healthcare Engineering*, 2022, 2022.

[29] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(1):315–337, 2000.

[30] Swarnendu Ghosh, Nibaran Das, Ishita Das, and Ujjwal Maulik. Understanding deep learning techniques for image segmentation. *ACM Comput. Surv.*, 52(4), aug 2019. URL: https://doi.org/10.1145/3329784, doi:10.1145/3329784.

[31] Shervin Minaee, Yuri Y. Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021. doi:10.1109/TPAMI.2021.3059968.

[32] Om Prakash Verma, Sudipta Roy, Subhash Chandra Pandey, and Mamta Mittal. *Advancement of machine intelligence in interactive medical image analysis*. Springer, 2019.

[33] K. Harrison, H. Pullen, C. Welsh, O. Oktay, J. Alvarez-Valle, and R. Jena. Machine learning for auto-segmentation in radiotherapy planning. *Clinical Oncology*, 34(2):74–88, 2022. Artificial Intelligence in Radiation Therapy. URL: https://www.sciencedirect.com/science/article/pii/S0936655521004635, doi:https://doi.org/10.1016/j.clon.2021.12.003.

[34] Song Yuheng and Yan Hao. Image segmentation algorithms overview. *arXiv preprint arXiv:1707.02051*, 2017.

[35] M Lalitha, M Kiruthiga, and C Loganathan. A survey on image segmentation through clustering algorithm. *International Journal of Science and Research*, 2(2):348–358, 2013.

[36] Neeraj Sharma and Lalit M Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3, 2010.

[37] Ender Konukoglu and Ben Glocker. Random forests in medical image computing. In *Handbook of Medical Image Computing and Computer Assisted Intervention*, pages 457–480. Elsevier, 2020.

[38] Chesner Désir, Simon Bernard, Caroline Petitjean, and Laurent Heutte. A random forest based approach for one class classification in medical imaging. In Fei Wang, Dinggang Shen, Pingkun Yan, and Kenji Suzuki, editors, *Machine Learning in Medical Imaging*, pages 250–257, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[39] Deepshikha Shrivastava, Sugata Sanyal, Arnab Kumar Maji, and Debdatta Kandar. Chapter 17 - bone cancer detection using machine learning techniques. In Sudip Paul and Dinesh Bhatia, editors, *Smart Healthcare for Disease Diagnosis and Prevention*, pages 175–183. Academic Press, 2020. URL: https://www.sciencedirect.com/science/article/pii/B9780128179130000171, doi:https://doi.org/10.1016/B978-0-12-817913-0.00017-1.

[40] Yahya Alzahrani and Boubakeur Boufama. Biomedical image segmentation: a survey. *SN Computer Science*, 2(4):1–22, 2021.

[41] Chen Lei. *Deep Learning Basics*, pages 17–28. Springer Singapore, Singapore, 2021. URL: https://doi.org/10.1007/978-981-16-2233-5_2, doi:10.1007/978-981-16-2233-5_2.

[42] KC Santosh, Nibaran Das, and Swarnendu Ghosh. Chapter 2 - deep learning: a review. In KC Santosh, Nibaran Das, and Swarnendu Ghosh, editors, *Deep Learning Models for Medical Imaging*, Primers in Biomedical Imaging Devices and Systems, pages 29–63. Academic Press, 2022. URL: https://www.sciencedirect.com/science/article/pii/B978012823504100012X, doi:https://doi.org/10.1016/B978-0-12-823504-1.00012-X.

[43] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[44] Yanming Sun and Chunyan Wang. A computation-efficient cnn system for high-quality brain tumor segmentation. *Biomedical Signal Processing and Control*, 74:103475, 2022. URL: https://www.sciencedirect.com/science/article/pii/S1746809421010727, doi:https://doi.org/10.1016/j.bspc.2021.103475.

[45] Rahimeh Rouhi, Mehdi Jafari, Shohreh Kasaei, and Peiman Keshavarzian. Benign and malignant breast tumors classification based on region growing and cnn segmentation. *Expert Systems with Applications*, 42(3):990–1002, 2015. URL: https://www.sciencedirect.com/science/article/pii/S0957417414005594, doi:https://doi.org/10.1016/j.eswa.2014.09.020.

[46] Yanming Sun and Chunyan Wang. A computation-efficient cnn system for high-quality brain tumor segmentation. *Biomedical Signal Processing and Control*, 74:103475, 2022.

[47] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.

[48] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[49] Tianmei Guo, Jiwen Dong, Henjian Li, and Yunxing Gao. Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pages 721–724, 2017. doi:10.1109/ICBDA.2017.8078730.

[50] Tianmei Guo, Jiwen Dong, Henjian Li, and Yunxing Gao. Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pages 721–724. IEEE, 2017.

[51] Yongchao Xu, Thierry Géraud, and Isabelle Bloch. From neonatal to adult brain mr image segmentation in a few seconds using 3d-like fully convolutional network and transfer learning. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 4417–4421, 2017. doi:10.1109/ICIP.2017.8297117.

[52] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[53] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[54] CG Peterfy, CF Van Dijke, DL Janzen, CC Glüer, R Namba, SHARMILA Majumdar, P Lang, and HK Genant. Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed mr imaging: optimization and validation. *Radiology*, 192(2):485–491, 1994.

[55] Ashwini A Kshirsagar, Paul J Watson, Jenny A Tyler, and Laurance D Hall. Measurement of localized cartilage volume and thickness of human knee joints by computer analysis of three-dimensional magnetic resonance images. *Investigative radiology*, 33(5):289–299, 1998.

[56] Fang Liu, Zhaoye Zhou, Hyungseok Jang, Alexey Samsonov, Gengyan Zhao, and Richard Kijowski. Deep convolutional neural network and 3d deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magnetic resonance in medicine*, 79(4):2379–2391, 2018.

[57] Graham Vincent, Chris Wolstenholme, Ian Scott, and Mike Bowes. Fully automatic segmentation of the knee joint using active appearance models. *Medical Image Analysis for the Clinic: A Grand Challenge*, 1:224, 2010.

[58] Timothy F Cootes, Carole J Twining, Vladimir S Petrovic, Roy Schestowitz, and Christopher J Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *BMVC*, volume 5, pages 879–888, 2005.

[59] Jurgen Fripp, Stuart Crozier, Simon K Warfield, and Sebastien Ourselin. Automatic segmentation of the bone and extraction of the bone–cartilage interface from magnetic resonance images of the knee. *Physics in Medicine & Biology*, 52(6):1617, 2007.

[60] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995. URL: https://www.sciencedirect.com/science/article/pii/S1077314285710041, doi:https://doi.org/10.1006/cviu.1995.1004.

[61] Yin Yin, Xiangmin Zhang, Rachel Williams, Xiaodong Wu, Donald D. Anderson, and Milan Sonka. Logismos—layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint. *IEEE Transactions on Medical Imaging*, 29(12):2023–2037, 2010. doi:10.1109/TMI.2010.2058861.

[62] Torsten Rohlfing, Robert Brandt, Randolf Menzel, Daniel B Russakoff, and Calvin R Maurer. Quo vadis, atlas-based segmentation? In *Handbook of biomedical image analysis*, pages 435–486. Springer, 2005.

[63] Somayeh Ebrahimkhani, Mohamed Hisham Jaward, Flavia M. Cicuttini, Anuja Dharmaratne, Yuanyuan Wang, and Alba G. Seco de Herrera. A review on segmentation of knee articular cartilage: from conventional methods towards deep learning. *Artificial Intelligence in Medicine*, 106:101851, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0933365719300144, doi:https://doi.org/10.1016/j.artmed.2020.101851.

[64] José G. Tamez-Peña, Joshua Farber, Patricia C. González, Edward Schreyer, Erika Schneider, and Saara Totterman. Unsupervised segmentation and quantification of anatomical knee features: Data from the osteoarthritis initiative. *IEEE Transactions on Biomedical Engineering*, 59(4):1177–1186, 2012. doi:10.1109/TBME.2012.2186612.

[65] Liang Shan, Christopher Zach, Cecil Charles, and Marc Niethammer. Automatic atlas-based three-label cartilage segmentation from mr knee images. *Medical image analysis*, 18(7):1233–1246, 2014.

[66] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *International conference on medical image computing and computer-assisted intervention*, pages 246–253. Springer, 2013.

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

[68] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[69] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. doi:10.1109/ACCESS.2021.3086020.

[70] Ruida Cheng, Natalia A Alexandridi, Richard M Smith, Aricia Shen, William Gandler, Evan McCreedy, Matthew J McAuliffe, and Frances T Sheehan. Fully automated patellofemoral mri segmentation using holistically nested networks: implications for evaluating patellofemoral osteoarthritis, pain, injury, pathology, and adolescent development. *Magnetic resonance in medicine*, 83(1):139–153, 2020.

[71] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[72] Felix Ambellan, Alexander Tack, Moritz Ehlke, and Stefan Zachow. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. *Medical image analysis*, 52:109–118, 2019.

[73] Guangbin Wang and Yaxin Han. Convolutional neural network for automatically segmenting magnetic resonance images of the shoulder joint. *Computer Methods and Programs in Biomedicine*, 200:105862, 2021.

[74] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[75] Mario Mustra, Kresimir Delac, and Mislav Grgic. Overview of the dicom standard. In *2008 50th International Symposium ELMAR*, volume 1, pages 39–44. IEEE, 2008.

[76] Peter Mildenberger, Marco Eichelberg, and Eric Martin. Introduction to the dicom standard. *European radiology*, 12(4):920–927, 2002.

[77] S Gopal Krishna Patro and Kishore Kumar Sahu. Normalization: A preprocessing stage. *arXiv e-prints*, pages arXiv–1503, 2015.

[78] Felix Ambellan, Alexander Tack, Moritz Ehlke, and Stefan Zachow. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. *Medical Image Analysis*, 52(2):109 – 118, 2019. doi:10.1016/j.media.2018.11.009.

[79] RT Abraham Padua and John A Carrino. 3t mr imaging of cartilage using 3d dual echo steady state (dess). *MAGNETOM Flash*, pages 33–36, 2011.

[80] Allen D. Elster. Questions and answers in mri - what is dess?, 2021. URL: https://www.mriquestions.com/dess.html.

[81] Python Software Foundation. Python. URL: https://www.python.org/.

[82] OpenCV. Open source computer vision library, 2015. URL: https://opencv.org/.

[83] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014. URL: https://scikit-image.org/.

[84] Mason, d. l., et al, pydicom: An open source dicom library. URL: https://github.com/pydicom/pydicom.

[85] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang

Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. URL: https://www.tensorflow.org/.

[86] François Chollet et al. Keras. https://keras.io, 2015.

[87] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7. IEEE, 2020.

[88] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging*, pages 379–387. Springer, 2017.

[89] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL: https://arxiv.org/abs/1412.6980, doi:10.48550/ARXIV.1412.6980.

[90] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[91] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. URL: https://doi.org/10.1109%2Faccess.2021.3086020, doi:10.1109/access.2021.3086020.