

New radiomics approaches for hepatic tumor characterization by
imaging analysis

OUHMICH Farid

Supervised by: AGNUS Vincent, HEITZ Fabrice & NOBLET Vincent

November 30, 2020

Contents

Contents	3
1 Introduction	11
1.1 Context	11
1.2 Contributions	11
1.3 Outline of the thesis	12
2 Liver cancer	15
2.1 Liver cancers description	15
2.2 Hepatocellular Carcinoma	17
2.2.1 Risk factors	17
2.2.2 HCC pathophysiology	18
2.2.2.1 Liver cancer cells of origin	18
2.2.2.2 Hepatocarcinogenesis	18
2.2.2.2.1 The different stages of the hepatocarcinogenesis	19
2.2.2.2.2 Physiological changes brought by the hepatocarcinogenesis . .	21
2.2.3 Diagnosis	24
2.2.3.1 Biopsy	24
2.2.3.1.1 Histological HCC grading systems	24
2.2.3.1.2 Limitations	27
2.2.3.2 CT and MR imaging	29
2.2.4 Staging & Treatment	32
3 Radiomics	37
3.1 Introduction	37
3.2 Handcrafted Radiomics	38
3.2.1 HCR workflow	39
3.2.1.1 Image acquisition and reconstruction	39

3.2.1.2	Segmentation	40
3.2.1.3	Features	41
3.2.1.4	Features Selection	42
3.2.2	HCR applied to the HCC	43
3.2.2.1	Experimental setup	48
3.2.2.2	ROI selection and features extraction	49
3.2.2.3	Features selection	50
3.2.2.4	Study reproducibility	54
3.3	Deep Learning Radiomics	57
3.3.1	DL applied to the medical imaging	57
3.3.2	Difference between HCR and DLR	57
3.3.3	DLR workflow	58
3.3.4	DLR applied to the liver	59
3.3.4.1	Study Endpoint	60
3.3.4.2	Image processing pipeline	60
3.3.4.3	Architecture choice	62
3.3.4.4	Performances	63
4	Segmentation of Liver Tissue in computed tomography	65
4.1	Introduction	65
4.2	Publicly available datasets & open challenges	66
4.3	Hand-crafted feature based semantic segmentation methods	68
4.3.1	Image Intensity based methods	69
4.3.2	Region Growing	71
4.3.3	Geometric deformable models	73
4.3.4	Graph-theory based methods	77
4.3.5	Probabilistic Atlases	79
4.3.6	Statistical Shape Models	81
4.3.7	Methods integrating machine learning based steps	85
4.4	Deep Learning based semantic segmentation methods	88
4.4.1	State-of-the-art DL pipelines	88
4.4.1.1	FCN-based architectures	88
4.4.1.2	Cascaded architectures	90
4.4.1.3	Input type (2D, 2.5D, 3D, patches, multiphase)	91

4.4.1.4	Training strategies (pre-processing, pre-trained networks...)	93
4.4.1.5	Inference schemes (post-processing, ensemble learning)	94
4.4.1.6	Conclusion	95
5	Histological grade prediction via automatic liver tumors segmentation	97
5.1	Semantic segmentation applied to the study of HCC	97
5.1.1	Motivations	97
5.1.2	Cascaded architecture	98
5.1.3	U-Net network	98
5.1.4	Multiphase information	100
5.1.5	Datasets	101
5.1.6	Experiments	103
5.1.6.1	Data pre-processing	103
5.1.6.2	Training	104
5.1.6.3	Semantic segmentation of liver tissues	105
5.1.6.4	Semantic segmentation applied to weakly annotated datasets	108
5.1.6.4.1	Registration	109
5.1.6.4.2	Automatic liver segmentation	110
5.1.6.4.3	ANTs registration	110
5.1.6.4.4	Automatic multiphase tumor segmentation	112
5.2	Deep Radiomics for histological grade prediction	114
5.2.1	Introduction	114
5.2.2	Related work	115
5.2.3	Prediction of the histological grade	117
5.2.3.1	Experiments and results	117
6	Conclusion and Perspectives	123
6.1	Contributions	123
6.2	Perspectives	126
6.2.1	Semantic Segmentation	126
6.2.2	Dynamic contrast-enhanced images	128
6.2.3	Deep Radiomics	129
Appendix A	Detailed analysis of the DL semantic segmentation reviewed articles	133
A.1	First FCN papers	133

A.2 LITS papers	137
A.3 New hybrid techniques	144
Appendix B Hyperparameters selection	145
B.1 Learning Rate & Decay	147
B.2 Network's depth	148
B.3 Data augmentation	149
Appendix C Radiomics features	153
C.1 First-order statistics	153
C.2 Gray Level Co-occurrence Matrix GLCM Features	156
C.3 Gray Level Run Length Matrix (GLRLM) Features	160
C.4 Gray Level Size Zone Matrix (GLSZM) Features	163
C.5 Gray Level Dependence Matrix (GLDM) Features	166
C.6 Neighbouring Gray Tone Difference Matrix (NGTDM) Features	168
Appendix D Evaluation metrics	171
D.1 Overlap measures	171
D.2 Surface Distance Metrics	171
Bibliography	173

Abbreviations

AFP Alpha-FetoProtein

AJCC American Joint Committee on Cancer

AR Arterial

ASM Active Shape Model

BCLC Barcelona Clinic Liver Cancer

BMI Body Mass Index

CECT Contrast-Enhanced Computed Tomography

CT Computed Tomography

DFS Disease Free Survival

DL Deep Learning

DLR Deep-Learning Radiomics

DSC Dice Similarity Coefficient

EM Expected-Maximisation

FCN Fully Convolutional Networks

FLC fibrolamellar HCC

FLL Focal Liver Lesion

FP False Positive

GC Graph Cut

GLCM Gray Level Co-occurrence Matrix

GT Ground Truth

GVF Gradient Vector Flow

HBV Hepatitis B Virus

HCC Hepatocellular Carcinoma

HCC-CCA mixed Hepatocellular Cholangiocarcinoma

HCR Hand-Crafted Radiomics

HCV Hepatitis C Virus

iCCA Intrahepatic Cholangiocarcinoma

IGA Image-guided Ablation

kNN k-Nearest Neighbours

MAP Maximum A Posteriori

MRI Magnetic Resonance Imaging

MVI MicroVascular Invasion

NAFLD Non Alcoholic-Fatty Liver Disease

NASH Nonalcoholic Steatohepatitis

NECT Non-Enhanced Computed Tomography

PA Probabilistic Atlas

PCA Principal Components Analysis

PDM Point Distribution Model

PSC Primary Sclerosing Cholangitis

PV Portal Venous

RBF Radial Basis Function

RECIST Response evaluation criteria in solid tumours

RFA Radiofrequency Ablation

ROI Region Of Interest

RQS Radiomics Quality Score

SSC Sparse Shape Composition

SSM Statistical Shape Model

SVM Support Vector Machine

TACE Transarterial Chemoembolization

TNM Tumor-Node-Metastases

VOI Volume Of Interest

WHO World Health Organization

Chapter 1

Introduction

1.1 Context

This research work is part of the RADIAL (RADomics Image Analysis of the Liver) project, carried out by the institute for minimally invasive surgery from Strasbourg (IHU), where one of the objectives is to improve the treatment and the prevention of liver cancer.

Being the second leading cause of cancer-related death worldwide, liver cancer is considered as a major public health challenge.

Current ways to characterize the liver cancer are either invasive, through the extraction of tissues samples or non-invasive, typically through interpretation of medical images.

Radiomics is a new field developed in the early 2010s to extract the encoded information within medical images in order to provide a more objective and image-guided assessment of the disease.

Conventional radiomics however suffers from a lot of limitations such as the hand-crafted design of the extracted features or the computation of characteristics only in a manually defined region or volume of interest. Deep learning changed the way to apprehend several medical imaging related problematics thanks to its ability to detect morphological properties in images only by using the pixel intensities as input. The radiomics field was impacted by this new set of algorithms, and a new branch of radiomics was created, where the definition of the region of interest and/or the extraction of features are performed using deep learning approaches.

1.2 Contributions

In our work we present strategies to automatically provide characterization of the liver tumors. We present a review of current liver-related hand-crafted radiomics studies and propose different ways to improve their reproducibility. Often performed manually, the delineations of the tumors can suffer from various bias, such as the inter and intra-observer variability, or its undefined borders leading

to errors when diagnosing, classifying or staging the liver cancer. Therefore, we present a robust automatic pipeline to perform semantic segmentation of liver tumors using both the power of deep cascaded networks architectures and the incorporation of dynamic contrast-enhanced information. A main prerequisite for the liver tumors characterization is the study of dynamic multiphase images, however, the main challenge that we faced in our work is the difficult to obtain (either internally or externally) databases containing liver tumors multiphase images. Consequently, we prove the value of our cascaded architecture by providing missing annotations in a weakly-annotated multiphase CT datasets. To characterize the tumors, we tackle the automatic histological grade prediction and we propose a novel deep radiomics pipeline which uses relevant semantic imaging features to assess the grade in a slice-wise fashion.

1.3 Outline of the thesis

We present the different types of liver cancers, and particularly the hepatocellular carcinoma (HCC), which will be the main topic of our work. We describe the different steps of hepatocarcinogenesis, which transforms healthy hepatic tissues into cancerous ones, by analyzing the evolution of the cells and the associated pathological changes. In case of suspected cancer, the current ways to establish a diagnosis are either through extraction and inspection of tissue samples or non-invasively by analyzing medical images. We outline the different modalities usually exploited to establish a diagnosis, and focus on the computed tomography (CT) for the rest of this research work. We first present the different imaging modalities available to perform a diagnosis, before detailing the two branches of the radiomics field. We review liver radiomics studies and focus on the deep learning radiomics branch for its ability to extract features directly from the input without necessary requiring any prior knowledge. We present our multiphase cascaded architecture to segment the liver and its tumor, before applying this to weakly annotated datasets. We finally use extracted features from our multiphase architecture to predict the histological grade.

The analysis of medical images is commonly performed by the naked eye, but technological advances allow the creation of computer assisted diagnosis tools to investigate the link between imaging features and biological characteristics. We present the radiomics technique allowing the extraction of quantitative features from images. The conventional pipeline is based on engineered features, thus considering the conventional radiomics pipeline as hand-crafted, hence the name of *HCR* (Hand-Crafted Radiomics). To depict the lack of reproducibility present in the vast majority of the *HCR*-liver related studies, we expose the conclusions of our review where *RQS* (Radiomics Quality Score) were analyzed.

The emergence of deep learning (DL) allows creation of algorithms where features are learned directly

from the raw images without extra information. The radiomics field benefited from the development of DL, and a new branch was created, the *DLR* (Deep-Learning Radiomics). We review the DLR studies performed on the HCC, and raise the issue of segmentation, often manually performed and thus suffering from high inter and intra-variability. To combat this variability a lot of work has been done to perform automatic segmentation of liver and its tumors. The first methods developed to perform this task were based on anatomical prior-knowledge or manual interactions, and often failed due to the limited amount of training samples, or when facing pathological cases. Automatic deep learning segmentation can address most of these issues by looking for relevant features directly from the images only. We review the different DL related liver tumors semantic segmentation studies and extract the common key settings to build our own segmentation pipeline. From these key settings, we focus on the cascaded architecture allowing, in the case of tumors segmentation, the implementation of two specialized networks (one to segment the liver on an entire CT slice and the second to delineate the tumors, but only on the liver region).

To prove the ability of DL networks to segment liver tissues, and to prove that a cascaded architecture combined with multiphase CT images allows the best accuracy, we perform semantic segmentation of liver tissues on an internal dataset.

To complete missing annotations in a weakly annotated or non-annotated dataset, we use the same architecture. And finally, we focus on the prediction of the histological grade, being an indicator of the evolution of the disease. We investigate whether the features extracted from our cascaded multiphase semantic segmentation are sufficiently relevant to perform this task. Finally, we present ways to improve our work in the future.

Chapter 2

Liver cancer

In this section, we will describe the anatomical properties of the liver and the major public health problem caused by liver cancers. This section allows us to see the potential value brought by our work regarding the treatment of liver cancer. After a brief introduction about the different types of liver cancers, we will focus on the most common primary one: the **Hepatocellular Carcinoma**. We will review its risk factors, and the way it progresses in the liver, before exposing the different invasive or non-invasive ways to establish the diagnosis of the disease. Finally, we present the different staging systems and treatments available to provide the better chance of survival to the diseased patients. A precise description of the different stages of the disease and the subsequent pathological changes is necessary in order to understand how our work can further be incorporated in the clinical practice.

2.1 Liver cancers description

The liver is a key organ in the human body, responsible for the synthesis of several proteins, and playing a major role during the digestion, particularly with the production of bile that is further stored in the gallbladder. It is also essential for the breakdown of numerous hormones, and it has a central position in the human blood flow system with its unique dual blood supply, being one of the three only portal systems of the human body and the only venous one.

The liver can suffer from various pathologies, which include liver cancer, that is now considered as a major public health challenge due to its high incidence and mortality rates. In their latest statistical report in 2019, the World Health Organization (WHO) ranked it as the fifth cancer type in terms of incidence with about 841,000 new cases annually, and as the fourth cause of cancer-related deaths worldwide, with about 782,000 annual deaths. Mortality and incidence rates are between 2 and 3 times higher among men than women in most regions of the world, making it the second type of cancer in terms of deaths for males [1]. More details about incidence and mortality rates are illustrated in both the figure 2.1 & 2.2.

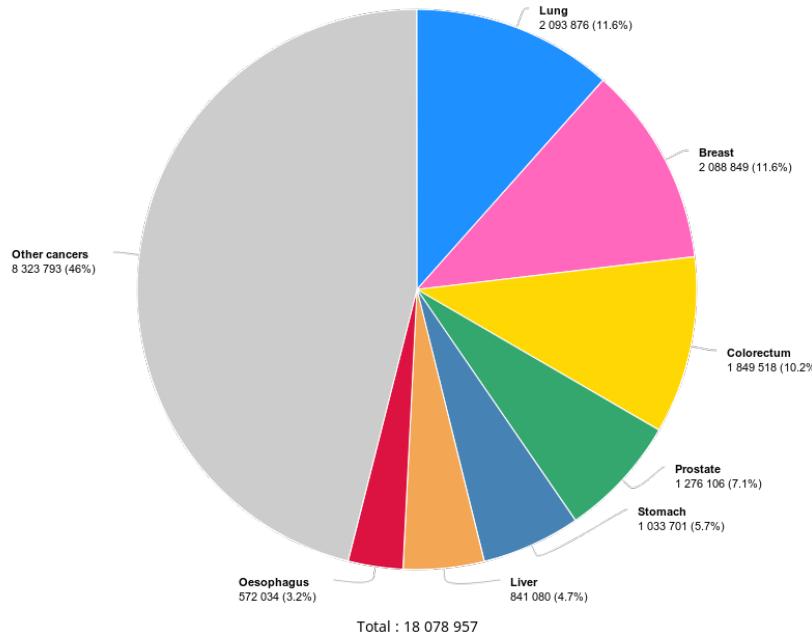


Figure 2.1: Incidence per cancer type in 2018, worldwide, both sexes, all ages, as described by the WHO [2]

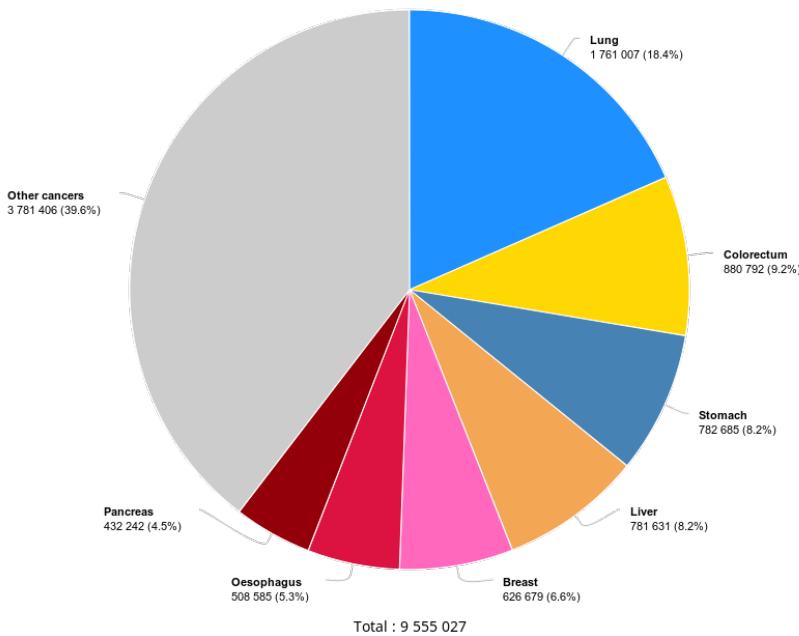


Figure 2.2: Estimated number of cancer-related deaths, per type of cancer in 2018, worldwide, both sexes, all ages, as described by the WHO [2]

The liver cancer can either be referred to as primary, meaning that it is originally growing in the liver itself, or as secondary, in case of extrahepatic cancers that metastasize in the liver. In case of metastases, cancerous cells often originate from the lung, the breast, and some parts of the digestive system, such as the colon, with the colorectal cancer being the main source of extrahepatic

metastases [3,4]. Even though secondary liver cancers are more frequent than primary ones, we have decided to focus on the latter since it is easier to understand the entire process from the appearance of the malignant cells to the physiological changes. Apprehending secondary liver cancers might require to analyze the primary site, something that is beyond the scope of our research work. Primary liver cancers comprise hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (iCCA), mixed hepatocellular cholangiocarcinoma (HCC-CCA) and other rare tumors, notably fibrolamellar HCC (FLC) and pediatric neoplasm hepatoblastoma [5–7]. HCC accounts for nearly 90% of primary liver cancers, with the highest incidence in Asia and in the Sub-Saharan countries, mainly due to the high prevalence of hepatitis B virus (HBV) [5,8]. The second most common one is iCCA, with a noticeable high incidence in Southeast Asian countries, and mainly developed in patients with primary sclerosing cholangitis (PSC), biliary duct cysts, hepatolithiasis, or parasitic biliary infestation with flukes [5,9].

The high incidence of HCC among the primary liver cancers and future collaboration with HCC expert Pr. Baumert leads us to focus our research work on this specific cancer type. We will now detail its risk factors, development, diagnosis methods and treatment techniques.

2.2 Hepatocellular Carcinoma

2.2.1 Risk factors

The main reason leading to HCC is the cirrhotic status of the liver, but its development is often related to the presence of other chronic liver diseases. Its incidence is heterogeneous worldwide because of the variable prevalence of risk factors: most cases occurring in sub-Saharan Africa and eastern Asia are associated with HBV and aflatoxin B1¹ exposure. In the USA, Europe and Japan, Hepatitis C (HCV) and the excessive alcohol consumption are the main risk factors [10]. Other risk factors include the presence of NAFLD (*Non Alcoholic-Fatty Liver Disease*), obesity and diabetes [11]. NAFLD was defined about 20 years ago, and regroups a spectrum of progressive liver diseases that encompasses simple steatosis, nonalcoholic steatohepatitis (NASH) and ultimately cirrhosis [11]. It has been identified as the underlying cause of patients presenting with HCC unrelated to virus and alcohol [12]. Several other studies confirmed that NAFLD is a risk factor for patients with either noncirrhotic liver [13,14] or cirrhotic liver [15–17]. Obesity has been established as a risk factor for several types of cancer, including liver cancer, and a study on 900,000 American adults reported a mortality rate five times higher in patients with a body mass index of 35 kg/m² compared to the group with a normal BMI (Body Mass Index) [18]. Several other studies were conducted to show the relation between HCC and obesity in the UK, Korea, Sweden, Taiwan and also a multicentric European study [19–23].

¹ toxin that can readily on foods and that has been classified as carcinogen by the International Agency for Research on Cancer

Diabetes has also been identified as an independent risk factor for HCC [10], especially types 2 diabetes [24–28].

2.2.2 HCC pathophysiology

2.2.2.1 Liver cancer cells of origin

The basic hepatic cells are divided into parenchymal (hepatocytes, which constitute between 60 and 80% of the total liver mass, and cholangiocytes as depicted in the figure 2.3) and non-parenchymal cells (fibroblasts, stellate cells, Kupffer cells, and endothelial cells) [5].

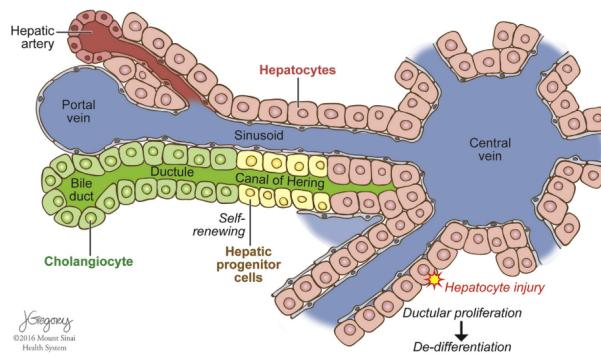


Figure 2.3: Structure of the liver lobule and location of candidate cell of origin of liver cancer. Different types of cancer (ie, HCC, iCCA, mixed HCC-CCA) can originate in the liver, depending on the transformation event and the cell type undergoing neoplastic transformation. ©Sia et al. [5]

The two main primary liver cancers types, HCC and iCCA have been considered to be distinct tumors that originate from specific cell populations. Nonetheless, they have recently been recognized as subtypes of a continuous spectrum of diseases. Therefore, several hypotheses exist about the cells of origin of the different primary liver cancer. One of them suggests that liver tumors can be generated by hepatic progenitor cells, since during liver development, both hepatocytes and cholangiocytes both arise from a common progenitor as depicted in the figure 2.4. However, both tumor subtypes can also arise distinctly from mature parenchymal cells [5].

This latest hypothesis is the one that we will describe in the following, and we will focus specifically on the development of HCCs.

2.2.2.2 Hepatocarcinogenesis

The main process behind the evolution of HCC is called *hepatocarcinogenesis*, which is defined as the progressive transformation of nonmalignant liver cells into HCC [29]. This transformation from nonmalignant liver cells into HCC is not fully understood, but the chronic inflammation present in the liver results in cycles of cell injury-death-regeneration, that stimulate epidemic changes and accumulation of genetic damages [30–33]. The high inter and intra-patient heterogeneity that exists for HCCs is

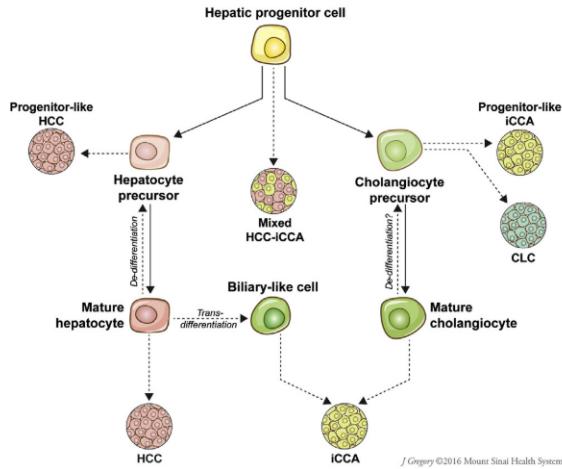


Figure 2.4: Multiple cells of origin of primary liver cancers. HCC and iCCA can develop via transformation of mature hepatocytes and cholangiocytes, respectively. There is evidence that hepatic progenitor cells (HPCs), their intermediate states, or dedifferentiated hepatocytes can form liver tumors with progenitor-like features, including mixed HCC-CCA, such as CLC. Mature hepatocytes can also be reprogrammed into cells that closely resemble biliary epithelial cells and contribute to development of iCCA. ©Sia et al. [5]

explained by the fact that several molecular variants of HCC may be produced, among the patients population, and even within different regions of the same tumor [33, 34]. The *hepatocarcinogenesis* is outlined by the gradual de-differentiation of abnormal nodular lesions, as described in the figure 2.5 [34, 35]. Over time, the more differentiated surrounding tissues are replaced by the growing less differentiated ones.

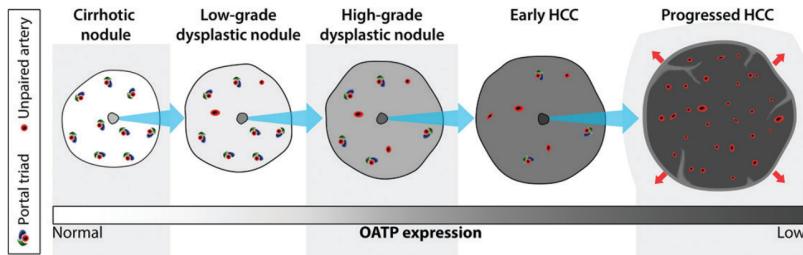


Figure 2.5: Schematic drawing illustrates typical changes in intranodular hemodynamics and OATP (*Organic anionic transporting polypeptides*) expression during multistep hepatocarcinogenesis. As shown, multistep hepatocarcinogenesis is characterized by successive selection and expansion of less-differentiated subnodules within more well differentiated parent nodules. The subnodules grow and eventually replace (blue arrows) the parent nodules. Progressed HCCs show expansile growth (red arrows) and characteristically are encapsulated with fibrous septa. Earlier nodules lack these structures and show replacing growth. During hepatocarcinogenesis, the density of portal triads diminishes while the density of unpaired arteries increases. ©Choi et al. [29]

It is possible that HCCs may arise from malignant cells without following this process, and without transitioning through histologically definable intermediate steps the process is then referred to as “denovo hepatocarcinogenesis” [36].

2.2.2.2.1 The different stages of the hepatocarcinogenesis

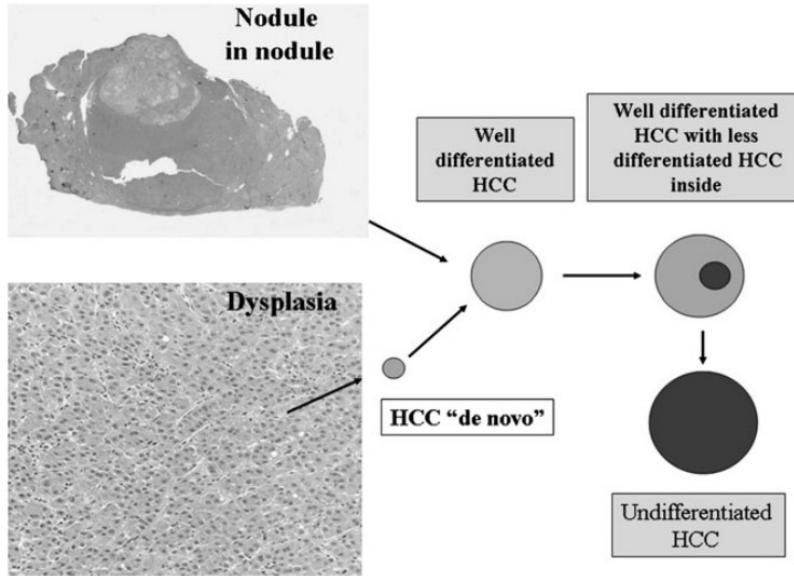


Figure 2.6: HCC can develop inside MRNs (*MacroRegenerative Nodules*). This type of carcinogenesis is called nodule in nodule. However, most HCCs develop outside MRNs, and precancerous lesions can be identified in areas of large and small cell dysplasia or irregular regeneration (de novo carcinogenesis). ©Trevisani et al. [33]

As depicted in the figure 2.6, hepatocarcinogenesis is composed of several steps, where each stage presents specific characteristics:

- *Cirrhotic nodules* are well-defined rounded regions of cirrhotic parenchyma surrounded by scar tissue and measuring less than 15 mm in diameter [37]. Even though they are usually considered as non-malignant, *hepatocytes* that they contain may develop dysplastic features, thus transforming the cirrhotic nodules into a *dysplastic foci* or *nodules*.
- *Dysplastic foci* are not identified via in vivo imaging, but may be recognized histologically. They are not well understood and may develop into *dysplastic nodules*.
- *Dysplastic nodules* correspond to nodular lesions that differ macro and microscopically from the surrounding parenchyma. They are observed in 25% of cirrhotic livers, and can be classified into low-grade (that appear like *cirrhotic nodules*) or high-grade (more similar to well-differentiated HCCs)
- *Early HCCs* grow gradually by replacing the parenchyma, unlike apparent *progressed HCCs* that displace or destroy the liver parenchyma. During their evolution, they tend to surround structures like portal tracts or central veins without destroying them.

They are microscopically indistinguishable from high-grade *dysplastic nodules*, since they tend to be vaguely nodular without having capsules nor distinct margins.

Even if they are considered as ancestors of *progressed HCCs* [35], their progression rate is not clearly defined [38].

- *Progressed HCCs* are generally separated between those measuring less than 2cm and those larger than 2cm.
 - Those smaller than 2cm tend to be nodular with a well-defined margin. They are different from *early HCCs* in the way that they grow by expanding into and compressing the adjacent parenchyma. They tend to be histologically moderately differentiated in the vast majority of the cases, and are associated with vascular invasion and intrahepatic metastasis [35].
 - The large *progressed HCCs* tend to have a more aggressive biological behavior, and are associated with a higher histological grade, with a higher presence of vascular invasion and metastasis. They are histologically composed of poorly differentiated or undifferentiated cancer cells that spread into the surrounding sinusoids, thus often characterized by an ill-defined boundary [33,39–45].
- *Multifocal HCC* can be the result of a synchronous development of several independent liver tumors originating from a primary tumor [33], knowing that patients with HCC are at higher risk of developing new tumors.

2.2.2.2 Physiological changes brought by the hepatocarcinogenesis

These several steps defining hepatocarcinogenesis are most of the time accompanied by some pathophysiological alterations:

- *Angiogenesis*: that is histologically characterized by the development of unpaired (or nontriadal) arteries as depicted in the figure 2.5 and the transformation of hepatic sinusoids into continuous capillaries, a.k.a “sinusoidal capillarization” [46,47].

Besides these changes, the portal tracts, containing both the non-tumoral hepatic arteries and the portal veins, progressively decrease [48].

Whereas the portal inflow to the nodule diminishes, the formation of unpaired arteries causes an increase in arterial flow [46,47]. This difference in blood inflow is such that we observe a total inversion of tendency during the hepatocarcinogenesis, with a decrease of arterial flow accompanied by a preservation of portal venous flow in the early phases, and a decrease of portal blood flow with an increase of arterial flow in the later phases, as depicted in the figure 2.7.

- *Venous drainage*: During hepatocarcinogenesis, the way the blood is drained by the lesions is subject to modifications as depicted in the figure 2.8.

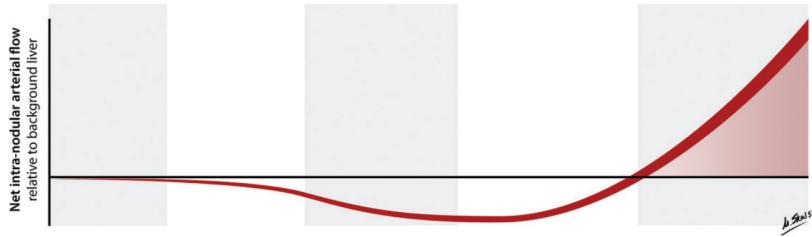


Figure 2.7: The net effect is that intranodular arterial supply diminishes initially and then increases; progressed HCCs typically show arterial hypervascularity compared with background liver ©Choi et al. [29]

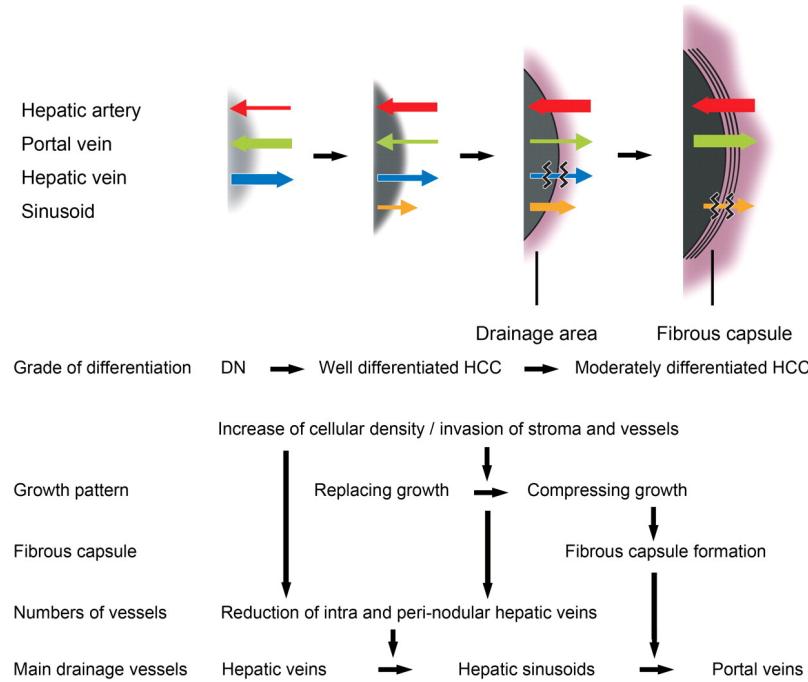


Figure 2.8: Diagram of the mechanism underlying changes in drainage vessels (top) and histologic features (bottom) of HCC during multistep hepatocarcinogenesis. Top: Direction of arrow: direction of flow for that vessel type. Thickness of arrow: volume through that vessel. Wavy lines through arrow: disruption of flow. Gray area: nodule. ©Kitao et al. [49]

First, the blood is evacuated via hepatic veins, then sinusoids are used and replace the hepatic veins that start to be blocked. Later during the process, the sinusoids start to collapse and the blood cannot be delivered through this way, therefore, the only remaining way is through the portal veins [49, 50].

This change along the disease progression may explain the phenomenon of corona enhancement, which appears mainly on progressed hypervascular HCC, and which referred to an enhancement of the peritumoral parenchyma starting a few seconds after the enhancement of the tumor. Worth noting that early HCCs are not concerned by this phenomenon [51].

- **Tumor Capsule and Fibrous Septa:** are observed in about 70% of progressed nodular HCCs [48], with a capsule consisting of two distinct layers.

Some studies have shown that tumors with an intact capsule were associated with a lower recurrence rate than those without (tumors of similar size), suggesting that the capsule may retard the tumor dissemination, especially the tight inner layer that might act as a physical barrier [52].

Even though advanced HCCs with an intact capsule have a more favorable prognosis than HCCs of the same size without any capsule, they have a worse prognosis than early HCCs that are unencapsulated [53].

- *Fat content:* It has been demonstrated that hepatocytes may accumulate fat during the early phases of hepatocarcinogenesis, causing the tumors to be more steatotic.

This steatosis is highly frequent for early HCCs of about 1.5 cm in diameter, but decreases when the size and the grade of the tumor increases. This might probably be because unpaired arteries become more developed once the tumor progresses, resolving ischemic conditions, and provoking the regression of steatosis [48, 54].

- *Iron content* may accumulate early in the process of hepatocarcinogenesis, noticeable essentially in low or high grade dysplastic nodules [55]. In the later stages, hepatocytes become resistant to iron accumulation due mainly to the utilization of iron by neoplastic cells and a higher cellular proliferation, thus early and progressed HCCs are often iron free [56].
- Diminution in the expression of *OATP Transporters*. Some of these transporters belonging to the OATP (*Organic anionic transporting polypeptides*) group are thought to be responsible for the uptake of two hepatobiliary-specific gadolinium-based contrast agents, namely the gadoxetate disodium and gadobenate dimeglumine.

Their expression level tends to be inversely proportional to the advancement of the disease [57, 58].

The different alterations are often the results of early or more advanced stages of hepatocarcinogenesis. They can appear individually or can be combined within the same tumor.

However, the HCC can often later spread either intra or extra-hepatically:

- *Intrahepatic metastasis:* corresponds to one of the most important spread mechanisms of progressed HCCs, and occurs mainly when malignant cells enter portal venules that drain the primary tumor, before spreading into the surrounding parenchyma.

These metastases around the primary tumor in the form of small satellites, and are located within its venous drainage area [59]. Even though malignant cells can easily invade vessels, extrahepatic metastases occurring in organs such as the lungs, the lymph nodes, the bones or the adrenal glands, are late manifestations of the disease [33, 45].

- *Vascular invasion:* is a late manifestation of hepatocarcinogenesis, affecting mainly the progressed HCCs [60], and allows a distinction between primary and secondary liver cancers which uncommonly present an invasion of intrahepatic vessels [61].

These invasions are either classified as micro or macroscopic, and usually are accompanied by a poor prognosis, as they provide a way for cancerous cells to propagate in the liver or systemically. Moreover, patients suffering from cancers with vascular invasion tend to have a higher recurrence rate after ablation, resection and transplantation, therefore, vascular invasion is often considered as a contraindication for these specific treatments [62].

Furthermore, another alteration, called “*Tumor Capsule invasion*” may increase the risk of vascular invasion, as HCC cells may infiltrate through the tumor capsule into the surrounding parenchyma.

All the changes occurring during hepatocarcinogenesis, leading to a more or less aggressive tumor, or even to multiple tumors, need to be assessed as early as possible to provide the best treatment available to the patient in order to increase its chances of survival.

2.2.3 Diagnosis

2.2.3.1 Biopsy

One of the standard ways to determine or confirm the cause and the stage of liver disease, as well as to inform treatment decisions and establish prognosis is to perform a *biopsy*. This surgical procedure consists of the sampling of a specimen measuring typically between 1 and 3 cm in length, with a diameter comprised between 1.2 and 2 mm [63].

The sampled tissues are then reviewed by pathologists, who report the degree of inflammation, steatosis, fibrosis and some other features such as cellular inclusions. The histological assessment is supported by clinicians providing the clinical context to complete the histologic interpretation, and some external blood markers such as the AFP (*Alpha-FetoProtein*) level can be incorporated to attest the presence of a malignant mass [64, 65].

Histological slice grading system was introduced by Edmonson-Steiner and corresponds to the level of cellular differentiation. 4 different groups were initially created and illustrated in the figure 2.9. This grading systems however suffers from several limitations, at such an extent that new grading systems were introduced.

2.2.3.1.1 Histological HCC grading systems

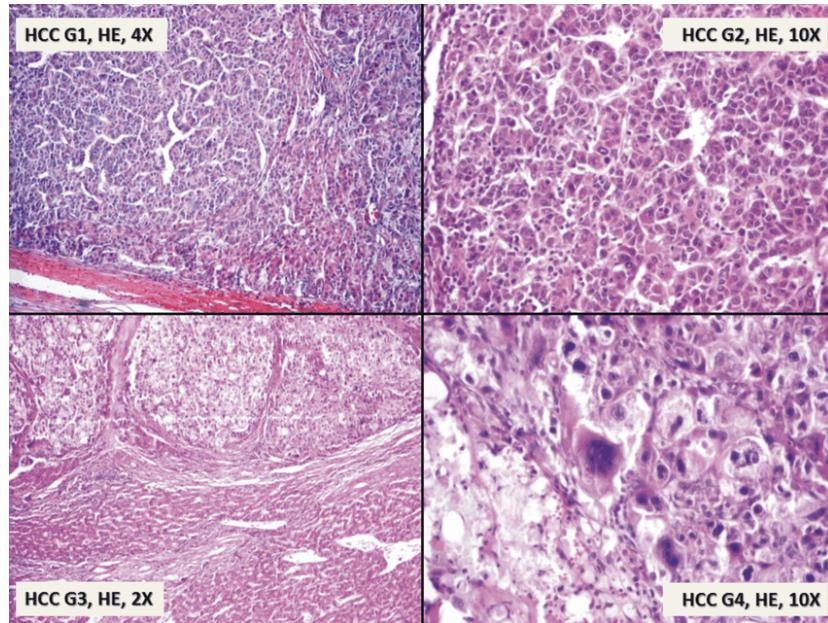


Figure 2.9: Histological grades as illustrated by ©Turdean et al. [66], G1: Well differentiated, G2: Moderately differentiated, G3: Poorly differentiated, G4: Undifferentiated

Several types of histological grading systems exist, with the most frequent ones being ES 1954 (Edmonson-Steiner) [60] and WHO 2010 (World Health Organization) [7]. Differences between these two grading systems are given in the figure 2.10.

Reference	Grades	Architecture	Cytology	Other features
World Health Organization (21)	Well differentiated	Thin trabecular, frequent acinar structures	Minimal atypia	Fatty change is frequent
	Moderately differentiated	Trabecular (3 or more cells in thickness) and acinar	Abundant eosinophilic cytoplasm, round nuclei with distinct nucleoli	Bile or proteinaceous fluid within acini
	Poorly differentiated	Solid	Moderate to marked pleomorphism	Absence of sinusoid-like blood spaces
	Undifferentiated	Solid	Little cytoplasm, spindle, or round-shaped cells	—
Edmonson and Steiner (13)	Grade I	—	—	Areas of carcinoma where distinction from hyperplastic liver is difficult
	Grade II	Trabecular, frequent acini (lumen varying from tiny canaliculi to large thyroid-like spaces)	Resemblance to normal hepatic cells; larger nuclei; abundant acidophilic cytoplasm	Cell borders sharp and clear cut; acini containing bile or protein precipitate
	Grade III	Distortion of trabecular structure, acini less frequent than grade II	Larger, more hyperchromatic nuclei, granular but less acidophilic cytoplasm	Acini are less frequent; tumor giant cells may be numerous
	Grade IV	Medullary, less trabeculae, rare acini	Highly hyperchromatic nuclei, scanty cytoplasm, with fewer granules	Loss of cell cohesiveness; giant, spindle or short-plump cells can be found

Figure 2.10: Features differences between the WHO and the ES1954 histological grading systems as detailed by Martins et al. [67] ©2017 Martins-Filho, Paiva, Azevedo and Alves

Even if ES 1954 is the most widely used grading system, a lot of divergences in the assessment of the grade can be found in the different studies reviewed by Martins et al. [67].

The organization of the patients in different numbers of tiers (3 or 4) G1, G2, G3 and potentially G4, and the possible classification into low or high differentiation level reveals a high variability in the assessment of the histological grade, as seen in the figure 2.11. As an example, some studies suggested that G2 is closer to G1 than to G3 [68,69], therefore, a classification that separates both G1 and G2 in one group and G3 (with potentially G4 if available) in the other can be understood.

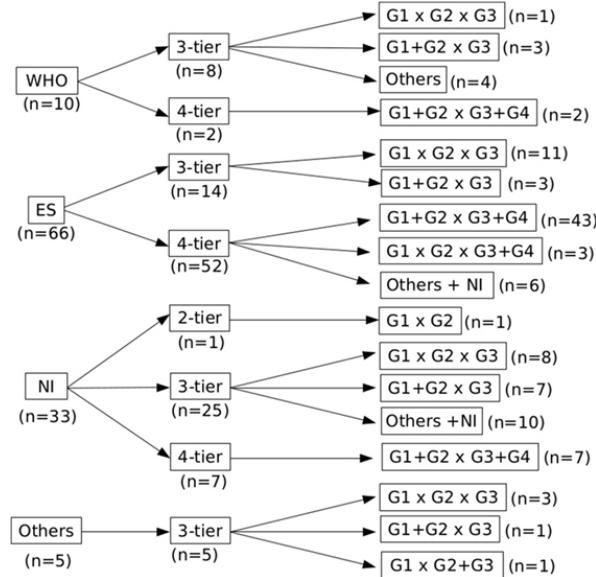


Figure 2.11: Distribution of the studies according to the grading reference, number of tiers, and data analysis, as detailed by Martins et al. [67] ©2017 Martins-Filho, Paiva, Azevedo and Alves

This difference in the classification of the patients was also noticed by Han et al. [68] and detailed in the figure 2.12. They reviewed different articles that used the histological grade to predict the prognosis of the patients, and realized that no clear guidelines were given for tumor presenting regions with heterogeneous grades (two examples of heterogeneous histological HCC slices are given in the figure 2.13).

Table 3 Studies related to histologic differentiation of hepatocellular carcinoma

Author	Definition of heterogeneous histologic grades	Comparison of histologic grades	Clinical significance of histologic grades
Chou	None	G1 vs G2 vs G3 vs G4	Higher histologic grades were more prevalent in microvascular invasion in HCC
Lim	None	G1 vs G2 vs G3	Up-regulation of Notch1 by p53 with Snail up-regulation promotes invasiveness of HCC and increases higher histologic grades
Decaens	None	G1 vs G2 vs G3	Histologic grades of HCC can be used as the new score for patient selection for liver transplantation
Pirisi	None	G1 & G2 vs G3	Histologic grades of HCC are usually underestimated in core biopsy
Zhou	None	G1 & G2 vs G3 & G4	Histologic grading could raise the predictive efficiency of TNM stage for survival of patients with HCC after liver resection
Pawlak	Poorest	G1, G2 & G3 vs G4	Selection of candidates for liver transplantation based on needle core biopsy tumor grade may be misleading
Minagawa	None	G1 vs G2, G3 & G4	On multivariate analysis, well-differentiated HCC is a good prognostic factor for patients with HCC after liver resection
Kim	Predominating	G1 vs G2 vs G3	Higher histologic grades might decrease cancer-free survival rate after radiofrequency ablation for HCC patients
Pawlak	None	G1 & G2 vs G3 & G4	On univariate analysis, higher histologic grade was a significant predictor of survival after hepatectomy for patients with major portal or hepatic vein invasion
Pawlak	Poorest	G1 vs G2 vs G3 & G4	A higher histologic grade can predict occult vascular invasion in HCC larger than 5 cm
Carlisi	None	G1 & G2 vs G3	Histologic grade is one predictor of HCC recurrence after liver transplantation

HCC, hepatocellular carcinoma.

Figure 2.12: Classification groups used by the studies reviewed by ©Han et al. [68]

The two possible solutions were to consider the worst grade (as recommended by the ES grading system), or to consider the most present one (recommended by the WHO grading system), however clear explanations of the decision taken in such cases are often undisclosed in the different studies.

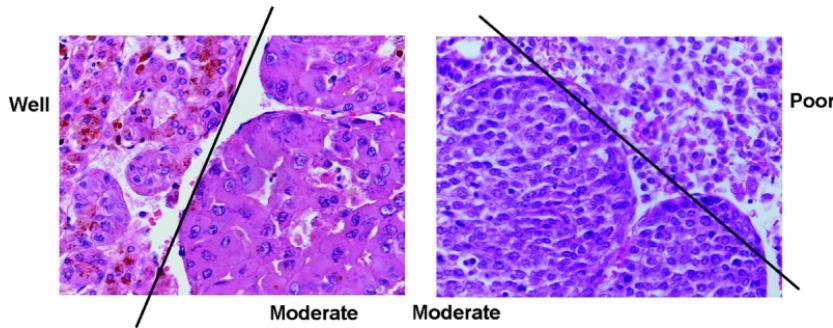


Figure 2.13: Examples of heterogeneous histological slices, as depicted by Pawlik et al. [70] ©2007 Lippincott Williams & Wilkins, Inc

The question regarding which histological HCCs grading system to use remains open. Moreover, various teams proposed to introduce a more complete grading system where the different features are classified separately as suggested by Martins et al. and depicted in the figure 2.14.

It is common to differentiate between the *clinical* and the *pathological* grade, where the *clinical* one corresponds to the assessment performed preoperatively, and the *pathological*² one corresponds to the evaluation of the sample collected during the surgery.

Some studies even suggested that the preoperative histological grade is not accurate enough to be used as a prognostic factor, and that because of the high difference between the NCB (needle core biopsy) grade and the one obtained from the final surgical specimen (resulted from the analysis of samples collected during surgery). Pawlik et al. exposed the concordance in estimating both the clinical and the pathological grades, their results were detailed in the figure 2.15.

As a conclusion, the problem regarding the histological classification of HCC remains open. Two standards are widely used, namely the ES one from 1954 and the WHO from 2010. They both initially contain 4 groups, but differ in the features to consider and the way heterogeneous lesions are supposed to be classified. When different grades are encountered in the lesion specimen, ES 1954 recommends to consider the worst one, whereas WHO 2010 advices to consider the most frequent one. New grading systems were introduced lately in order to overcome these specific problems.

2.2.3.1.2 Limitations

The biopsy suffers from several limitations. Sampling errors are common, as reported by a study where laparoscopic biopsy samples, taken simultaneously from the right and the left lobes of several patients suffering from HCV, were interpreted as cirrhotic in one lobe and fibrotic (F3) in the other in almost 15% of the cases [71]. It has also been proven that the accuracy of both the diagnosis and the staging of the disease highly depend on the size of the specimen, with misinterpretations when too small samples are collected [72].

² Important to distinguish from the pathological stage which mainly corresponds to the TNM.

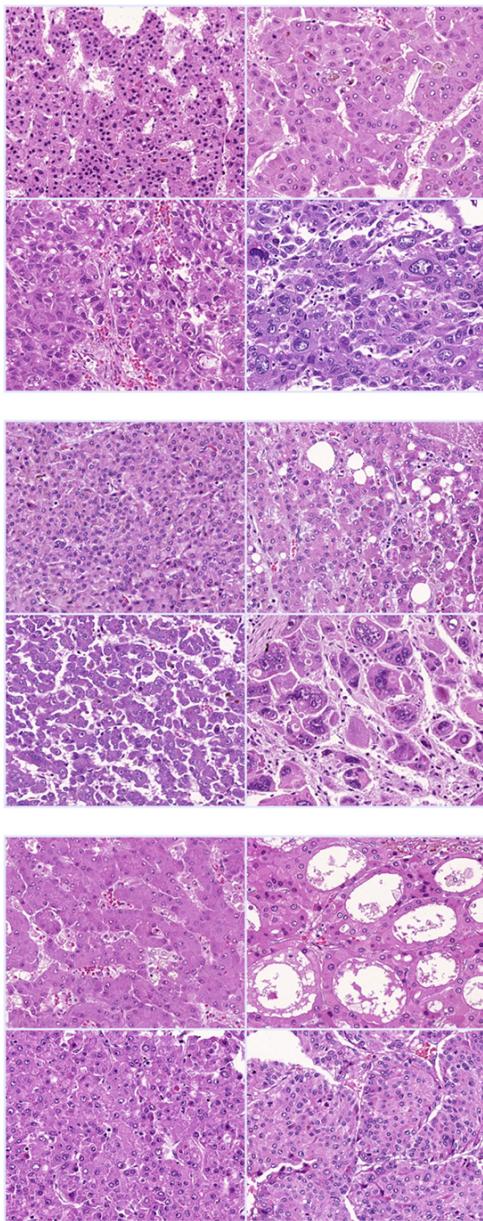


Figure 2.14: New histological HCC grading system proposed by ©Martins et al. [67]

TABLE 3. Concordance of Tumor Grade on Preoperative Needle Core Biopsy Versus Final Surgical Pathology Using 3-Tier Grading System (κ statistic = 0.18)

Preop. Needle Biopsy	Final Surgical Pathology			Total
	Well	Moderate	Poor	
Well	17	14	4	35
Moderate	15	16	13	44
Poor	2	3	9	14
Total	34	33	26	93

Figure 2.15: Concordance of pathological and clinical tumor grade estimation, as detailed by Pawlik et al. © 2007 Lippincott Williams & Wilkins, Inc. [70]

The interpretation of the samples is very subjective as it has been proven in some studies, with the estimation of both inflammatory activity and fat burden that suffered from a high inter and intraobserver variability [73]. Moreover, being a surgical gesture, the biopsy can suffer from adverse events going from simple pain to death in some rare cases [74–77]. Finally, they happen to be costly since they involve an expert gastroenterologist or interventional radiologist, a pathologist, and need to be performed in facilities with adequate equipment.

Even though biopsy is still used to evaluate the degree of fibrosis or cirrhosis, primary liver cancers are now most often diagnosed on the basis of imaging studies alone in clinical practice. Recent advancements in medical imaging enable the radiologist to discern the advancement of focal liver lesions on the basis of vascularity and physiological features, especially through multiphasic, contrast-enhanced, cross-sectional imaging (either computed tomography *CT*, or magnetic resonance *MR*).

2.2.3.2 CT and MR imaging

As explained previously, hepatocarcinogenesis is a long process that can render the tumor very aggressive, and the more the disease evolves, the worse the prognosis will be for the patient, thus an early detection of HCC is critical to improve the survival of affected patients.

The majority of the current guidelines recommend ultrasonography (US) as the primary imaging test for surveillance [29]. CT and MRI are generally not chosen for surveillance but recommended by some guidelines when the US is limited due to obesity or when the risk factors are very high [78, 79].

If the surveillance is positive, the main guidelines advocate in favor of multiphasic CT and MR with extracellular agents for the non-invasive diagnosis and staging [39, 78–80].

Multiphasic CT and MRI can be categorized by the type of contrast agents used:

- Imaging with extracellular contrast agents permits a diagnosis of HCC based on the physiological changes in the blood flow produced by hepatocarcinogenesis. Images are typically acquired before (precontrast) and dynamically after the injection of a contrast agent.

The contrast agent is generally administered at a rate of 4-6 mL/sec for CT and around 2 mL/sec for MRI [81, 82]. The dose is usually adjusted to the body weight, where 1.5-2 mL/kg to achieve an iodine dose of 525 mg/kg of iodine are administered in the case of CT.

Concerning the dynamic contrast enhanced images, three phases are typically acquired, namely the late hepatic arterial, the portal venous, and the delayed phase as illustrated in the figure 2.16:

- The *late hepatic arterial* phase is characterized by the full enhancement of the hepatic artery with all its branches and an enhancement of the portal vein [29]. The portal vein is however not supposed to be enhanced by the antegrade flow [83].

This phase is critical for the characterization of hypervascular HCCs, since it catches the peak arterial enhancement of tumors [84]. The *early hepatic arterial* phase is often omitted by the centers as only hepatic arteries are enhanced but portal veins are not, which render the detection of hypervascular tumors very difficult. One of the main problems associated with this acquisition is the difficulty to detect the peak arterial perfusion, since fixed delay is often not reliable. Some techniques such as contrast agent bolus tracking, or the selection of the best image after multiple continuous acquisitions may be recommended [85].

- The *portal venous* phase is characterized by an enhancement of both hepatic artery (arterial) and portal veins. It is however worth noting that the contrast agent still circulates in the body and will still be present in hepatic arteries as well (but with a lower concentration than for the arterial phases).

The portal venous phase is generally acquired at around 60-80 seconds after the start of the injection, but no clear guidelines are given concerning a precise way to determine the best moment for the acquisition of this given phase.

- The *delayed phase* is acquired 3-5 minutes after the injection and helps to understand how the liver restores the contrast agent back to the rest of the body, especially in case of inhomogeneous washout.

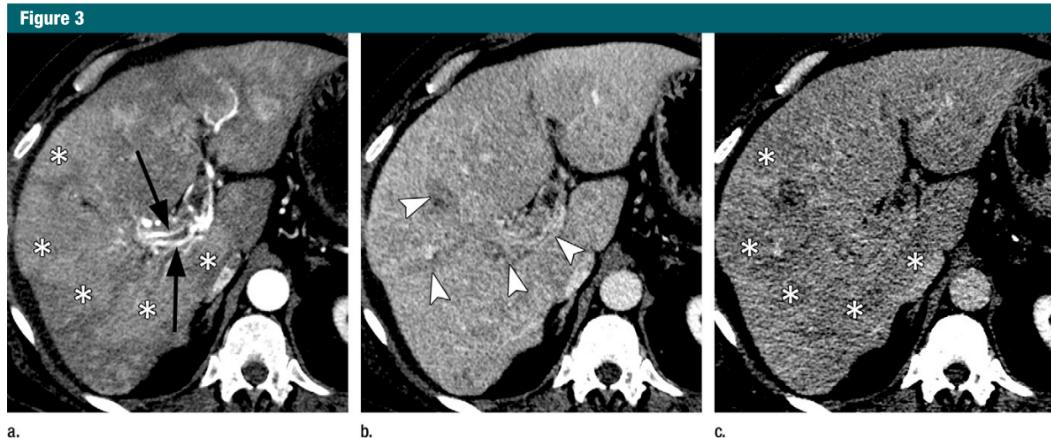


Figure 3: Images in a 64-year-old man with infiltrative HCC and macrovascular invasion. Axial CT images obtained in the (a) late arterial, (b) portal venous, and (c) 3-minute delayed phases after administration of an iodinated contrast agent reveal heterogeneously enhancing soft tissue expanding the lumen of the right portal vein and its branches consistent (arrowheads) with macrovascular invasion by HCC. Note arterial phase hyperenhancing tumoral arteries (arrows), sometimes described as "threads and streaks," within the intraluminal tissue. Note patchy areas (*) of arterial phase hyperenhancement and delayed phase partial washout appearance in the liver parenchyma, consistent with infiltrative HCC.

Figure 2.16: Example of computed tomography images of patients suffering from infiltrative HCC with macrovascular invasion. ©Choi et al. [86]

Both the *portal venous* and the *delayed* phases are critical for the characterization of washout and/or capsule appearance, and are crucial for the differentiation of small HCCs from small ICCs [87,88]. Some centers decide to skip the delayed phase, but it can lead to a loss of information.

Precontrast images are useful in case of detection of enhancement and evaluation of its degree by subtracting the post-contrast image with the precontrast one. They are also useful for patients with iron-rich nodules to detect hyper attenuation before injection of contrast agent, thus avoiding misinterpretation of arterial phase enhancement.

In addition to these classical vascular phases, examinations performed with MR imaging usually include other sequences such as the T1-weighted in-phase, the T2-weighted fast-spin-echo and some diffusion-weighted sequences.

- In order to fully understand the behavior of HCC cells, one can use extracellular agents as explained above, or use intracellular agents. In the latter case, hepatobiliary agents allow a better diagnosis through understanding the hepatocellular function in addition to the vascularity information. These agents first enhance the extracellular space before entering the hepatocytes through the OATP specific receptors. To acquire this hepatobiliary phase, which can exclusively be obtained on MRI, different agents exist, and they mainly differ in their hepatocellular uptake, thus reaching an enhancement at different moments. In the exception to the delayed phase for some specific agents, the normal phases can be acquired, along with some new phases such as the “transitional phase”, which represent a transition from extracellular-dominant to intracellular-dominant enhancement [89]. Examples of images obtained with after a multiphasic MR examination are illustrated in the figure 2.17

Unlike extracellular agents, the hepatobiliary contrast agents show promise for differentiating early HCCs and premalignant nodules from lower-risk nodules [90–96].

Main advantages of choosing CT over MRI examination is that it is widely available, rapid, robust and requires less expertise than MRI when interpreting images. It however exposes patients to radiation doses, and provides relatively lower soft-tissue contrast than when MRI is used. MRI permits the assessment of a great number of tissue characteristics, but is time-consuming, less robust, and prone to artifacts. Even though MR imaging examination may be preferred at some academic centers, there is no consensus for recommending this modality over CT in community or less-specialized centers [65, 86]. Therefore, it has been decided to mainly focus our research work on CT images.

These recent innovations in the medical imaging fields were accompanied by the creation of several clinical criteria that helped the radiologists to take decisions concerning the treatment choice, depending mainly on the characteristics of the tumor.

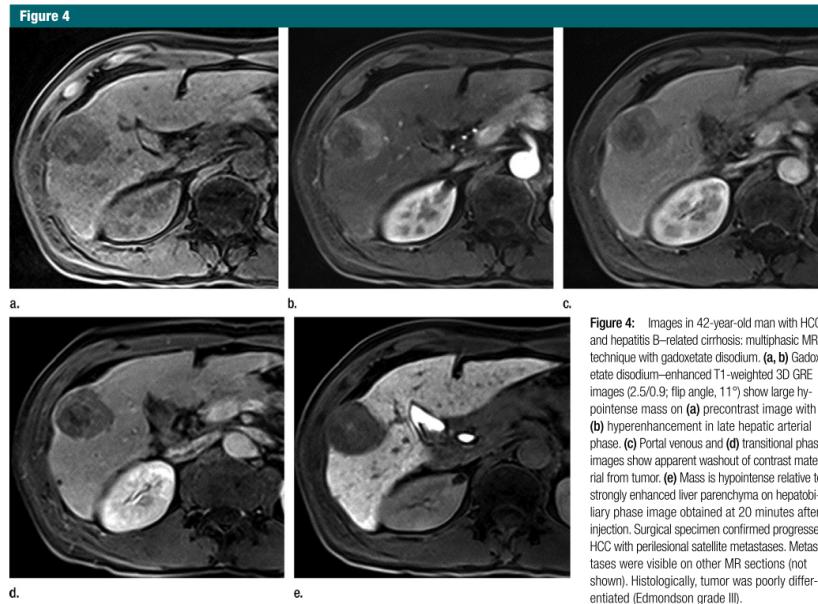


Figure 4: Images in 42-year-old man with HCC and hepatitis B-related cirrhosis: multiphasic MR technique with gadopentetate disodium. (a, b) Gadopentetate disodium-enhanced T1-weighted 3D GRE images (2.5/0.9; flip angle, 11°) show large hypointense mass on (a) precontrast image with (b) hyperenhancement in late hepatic arterial phase. (c) Portal venous and (d) transitional phase images show apparent washout of contrast material from tumor. (e) Mass is hypointense relative to strongly enhanced liver parenchyma on hepatobilary phase image obtained at 20 minutes after injection. Surgical specimen confirmed progressed HCC with perilesional satellite metastases. Metastases were visible on other MR sections (not shown). Histologically, tumor was poorly differentiated (Edmondson grade III).

Figure 2.17: Example of multiphasic MR images of a patient suffering from HCC, ©Choi et al. [86]

2.2.4 Staging & Treatment

One of the most widely used criteria for the assessment of HCCs is the BCLC (*Barcelona Clinic Liver Cancer*) staging system, recently refined by Llovet et al. in 2008, and illustrated in the figure 2.18 [97].

As described in the figure below, the main factors used to build this staging system rely on the *tumor status*, namely, the number and the size of the nodules, the presence or the absence of vascular invasion and extrahepatic spread. Other characteristics such as the *liver function* (as defined by the Child-Pugh classification [98]), the potential presence of portal hypertension and both serum bilirubin and albumin levels), and an indicator called the *general performance status* describing the overall level of functioning of the patient [99].

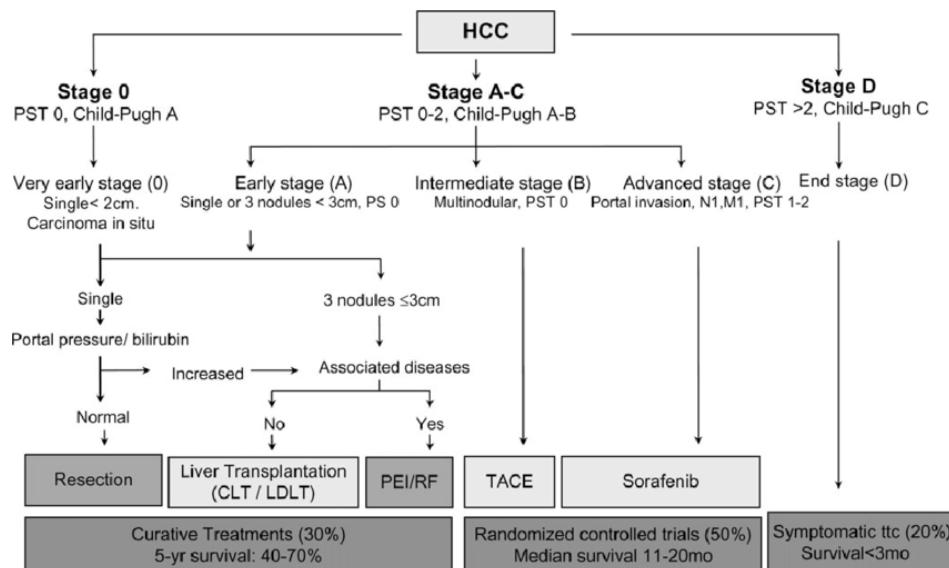


Figure 2.18: BCLC staging classification, as illustrated by Llovet et al. ©Llovet et al. [97]

These criteria were developed to support previous staging systems such as the TNM (*Tumor-Node-Metastases*) as depicted in the figure 2.19, that was introduced and refined by the AJCC (American Joint Committee on Cancer), and it incorporated some independent work such as the *Milan criteria* which concerns only a specific subtype of the TNM staging system [100].

Primary Tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Solitary tumor without vascular invasion
T2	Solitary tumor with vascular invasion or multiple tumors none more than 5 cm
T3a	Multiple tumors more than 5 cm
T3b	Single tumor or multiple tumors of any size involving a major branch of the portal vein or hepatic vein
T4	Tumor(s) with direct invasion of adjacent organs other than the gallbladder or with perforation of visceral peritoneum

Regional Lymph Nodes (N)	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis

Distant Metastasis (M)	
M0	No distant metastasis
M1	Distant metastasis

Figure 2.19: TNM classification as described by the ©AJCC Cancer Staging [101]

One of the main advantages brought by the BCLC staging system is to provide strong guidelines concerning the treatments to follow depending on the disease progression. Hereafter, we detail the different treatment techniques available for patients suffering from HCCs when seeking for curation:

- *Surgical resection*, where the best candidates are patients with solitary tumor and preserved liver functions [10].

A resection is defined as curative if all macroscopic evidence of primary and second tumor are detected in the remnant liver or in any other organs by dynamic CT or MR within 4 weeks after surgery [102].

As mentioned by previous studies, resection remains the treatment of choice in patients without cirrhosis suffering from HCC. Patients with cirrhosis have to be carefully selected to reduce the risk of postoperative liver failure [79, 80, 103].

- *Liver transplantation*, which mainly benefits patients who are not good candidates for surgical resection, and fits better those within the *Milan criteria* (solitary tumor up to 5cm or less than 3 nodules measuring each one less than 3cm) [79, 104]. In theory the transplantation may cure the tumor and the underlying cirrhosis at the same time, however, this procedure suffers from a scarcity in terms of donors, with an increase in the waiting time that led up to 20% of the potential receivers to drop out of the list before finding a donor, even with the existence of bridge therapies that can preserve the patient health during the waiting time [105].

Some modifications have been done to the initial inclusion criteria for the transplant candidates after poor recurrence results were obtained [79].

- Image-guided ablation (IGA) is the most frequently used therapeutic strategy but its efficacy is limited by the size of the tumor, its location and the respiratory motion of the patient.

This technique belongs, just like resections, to primary line treatments for patients with early stage HCC (BCLC stage 0-A) for whom surgical management is contraindicated [65, 80]. It has also been noticed that patients with very small HCCs (less than 2cm in diameter) can benefit from IGA to avoid any surgical procedure [106].

Several methods have been developed to perform the destruction of the tumor, and radiofrequency ablation (RFA) remains the most popular technique. Other techniques (thermal or non-thermal) have been adopted to overcome limitations of RFA.

The risk of complications is higher when the procedure is performed on tumors located along the surface of the liver. Indeed, puncture can cause bleeding, and heating can cause complications by provoking injury on adjacent organs. The intervention should be performed by an oncologist with sufficient experience, in order to assess the risk of causing damages to the gastrointestinal tract [107].

However, RFA techniques have recently been improved by the assistance of 3D navigation systems, allowing better planning with multiple overlapping ablation zones, a more accurate placement of the probes and an assessment of the results intraoperatively, thanks to image fusion. This new generation of techniques is called “*stereotactic RFA*” [108–110].

Therefore, the question whether to choose resection over RFA is still open [65, 109].

- Transarterial Chemoembolization (TACE), which has survival benefits in asymptomatic patients with multifocal disease without vascular invasion or extrahepatic spread can also be considered [10]. TACE can also be applied to patients in earlier stages, who are not suitable for the other therapeutic available treatments. This clinical situation is known as “treatment stage migration” and reflects a certain flexibility in the BCLC clinical interpretation [111].

The same standards are used for the patient follow-up, and can be supported by other metrics such as the RECIST (standing for “*Response evaluation criteria in solid tumours*”) which assess the tumor response, but presents several limitations and has already been modified in the past to incorporate changes [112].

As explained here, general staging systems rely mainly on the macro changes caused by the disease, ignoring many histological measurements, such as the histological grade [60].

Current methods to evaluate the tumor and to propose the best available treatment still suffer from a lot of limitations, especially because they rely on subjective interpretation of images, or clinical data, and also because of the discrete staging systems that are used in a clinical routine.

The main hypothesis of our research work is that medical images may be analyzed and interpreted to provide an objective feedback to the clinicians, and help them with non-invasive tools to give to patients suffering from HCCs the best chances of survival.

The radiomics concept has been introduced to address these issues, by providing an objective evaluation of the medical images at a finer scale than the one used by the medical experts during their visual inspection. It was designed to non-invasively obtain predictive or prognostic models, in order for example to assess the tumor response to a treatment, the recurrence after surgery or predict the type of tumors. Radiomics studies comes with several prerequisites such as the need to respect ethical considerations regarding patient data privacy, or the standardization of the databases [113]. Since its creation, the radiomics field underwent several modifications in its conception to the extent that two different branches are now distinguishable: the hand-crafted radiomics (HCR) and the deep learning radiomics (DLR). The key concepts of the radiomics pipeline and differences between the two existing branches will be discussed in the next chapter.

Chapter 3

Radiomics

3.1 Introduction

As explained previously, there are globally two ways for establishing a diagnosis of diseases such as liver cancers: the biopsy, or the diagnostic imaging. The biopsy, as detailed earlier, suffers from a lot of drawbacks. It remains an invasive procedure, with a high cost in terms of resources, and does not consider the tumor heterogeneity. The diagnostic imaging on the other hand, is not invasive, provides information about the tumor shape, the growth over the time, and is less prone to bias due to tissue heterogeneity. The biopsy remains the only available ground truth since diagnosis imaging still suffer from the medical expert subjectivity. In our work, we are trying to prove that medical images can be used to improve the visual inspection of the medical expert and at long-term replace invasive procedure such as the biopsy.

The recent improvements in the medical imaging field allow acquisition of data enabling a better characterization of the patients phenotype. In the case of brain imaging, the augmentation of contrast, on MR images, thanks to the injection of contrast agents such as gadolinium-based agents (method mentioned before) is an important technique for the evaluation of brain and liver tumors [114–116]. This tool allows a delineation of large tumors and an early detection of small metastatic lesions. The different MRI sequences (e.g. the T1 weighted sequences) also allow an internal separation of the tissue within the same tumor (active vs necrotic part of the tumor) [114].

Support brought by the innovations in the medical imaging field have been demonstrated on other organs such as the liver [117], the breast [118] or the colon [4] with a consequent benefit in terms of diagnosis. However, even though the advancements in the medical imaging fields have allowed these performances, the interpretation of the medical images remains subjective and not quantitative. In order to correctly provide a diagnosis it is necessary to transform images into mineable high-dimensional data.

Introduced in the 80s, CAD (*Computer Assisted Diagnosis*) tools were the first to establish a link between the imaging features and the biological characteristics of the patients [119]. In order to go along with those new systems, standards were introduced such as the one created by the WHO (*World Health Organization*) or the RECIST (*Response Evaluation Criteria in Solid Tumors*) [120], where the objective was to assess the evolution of the disease following the progression of the tumor size, but here again, those criteria suffer from a too high dependence with the observers.

The term *radiomics* was introduced in the early 2010s, allowing the computation of more features than the traditional CADs (more than a thousand vs only a dozen previously) and bringing a more complete diagnosis, since CADs were often limited to distinguish benign vs malignant lesions [121]. This new technique allows some breakthroughs in various applications such as the cancer diagnosis, the detection of the tumors (with the identification of malignant lesions), their classification, the estimation of the patient survival, the prediction of the aggressivity of the tumors, their recurrence, or the advancement of the disease. In the clinical practice, this new method also allows an improvement in the way biopsies are performed, with the identification of the areas where the extraction should be performed [122] or even by prediction when a biopsy is helpful or not [123].

Compared to the above-mentioned criteria based on a naked-eye examination, we are now able to rely on a computer to analyse the gray-levels at a finer scale. Therefore, two approaches exist, the HCR (*Hand-Crafted Radiomics*) based on mathematical engineered features, relying on the textural, intensity and shape based properties of the volume of interest, and the DLR (*Deep-Learning radiomics*), where the extracted features are estimated from the input data without any prior knowledge.

3.2 Handcrafted Radiomics

In this section we will describe the HCR pipeline, by first exposing the different steps of the classical workflow, before analyzing the different studies that used HCR on patients suffering from HCCs. We will conclude with the different improvements that should be brought to enhance the predictive value of radiomics.

A conventional radiomics workflow (based on HCR features) starts with the acquisition and the reconstruction of medical images, followed by the segmentation of those images, which is a critical step since HCR features are extracted from organs or lesions having indistinct borders [122]. Once the different areas were segmented, the features are extracted and quantified, and a statistical analysis is performed to select only the most relevant one. The final step consists in building a model that use the selected features to perform the wanted task, which is often either the tumor characterization or its prognosis. The pipeline is illustrated in the figure below 3.1.

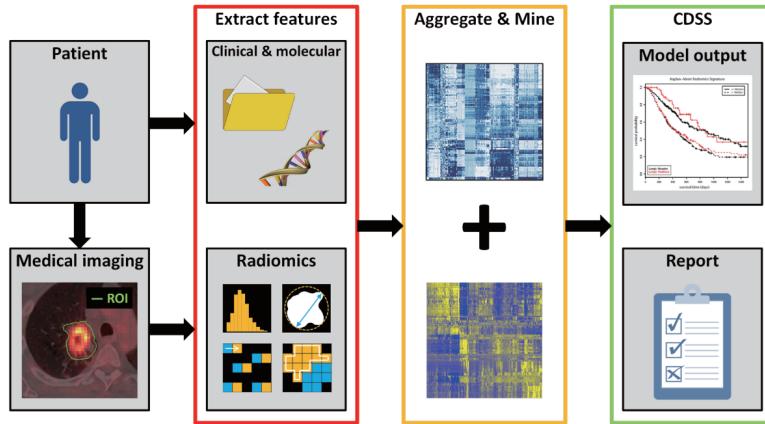


Figure 1 Overview of methodological processes in radiomics: data discovery, collection and preparation, model(s) development/validation and implementation, assessment of clinical utility.

Figure 3.1: Conventional radiomics workflow as depicted by ©Scrivener et al. [124]

We will now describe these different steps, before analyzing recent state-of-the-art HCR studies, before establishing a list of measures needed to be taken in order to improve the quality and the reproducibility of future radiomics works.

3.2.1 HCR workflow

3.2.1.1 Image acquisition and reconstruction

As explained previously (see chapter 2), ultrasonography (US) is the recommended modality as primary imaging test for surveillance. If the surveillance is positive, CT or MR examinations are performed for the diagnosis and the staging of the disease. For the reasons exposed previously, namely the availability, and its robustness when compared to MRI, we will focus on HCR studies based on CT imaging data. Without entering into the details of how a CT scan works, we can assume that performances of the CT imaging depend mainly on some settings such as the slice thickness, the capability for projecting the density variations into image intensities and the reconstruction algorithm which aims at converting tomographic measurements into cross-sectional images. It has been demonstrated that radiomics features can differ between different scanners with the same settings [125]. It is also common to differentiate CT images into two categories, the screening where low dose images are used and the diagnosis with higher quality of contrast obtained with higher doses [126]. It is worth mentioning that the majority of the studies that are being investigated belong to the second category where higher doses are used during the imaging acquisition.

Images are typically combined with other clinical sources when computing the radiomics features. Among them, gene expressions, clinical data such as the age, the gender or the past medical history, blood biomarkers or other prognostic markers such as the tumor size, the stage or the recurrence are the main non-imaging sources of data that are used in the radiomics workflow. However, they can

be difficult to acquire, normalize and integrate in a radiomics pipeline, therefore, features are most commonly extracted from images only.

3.2.1.2 Segmentation

Historically, the segmentation was performed manually, hence, prone to the inter- and intra-observer variability, as depicted below in the figure 3.2.

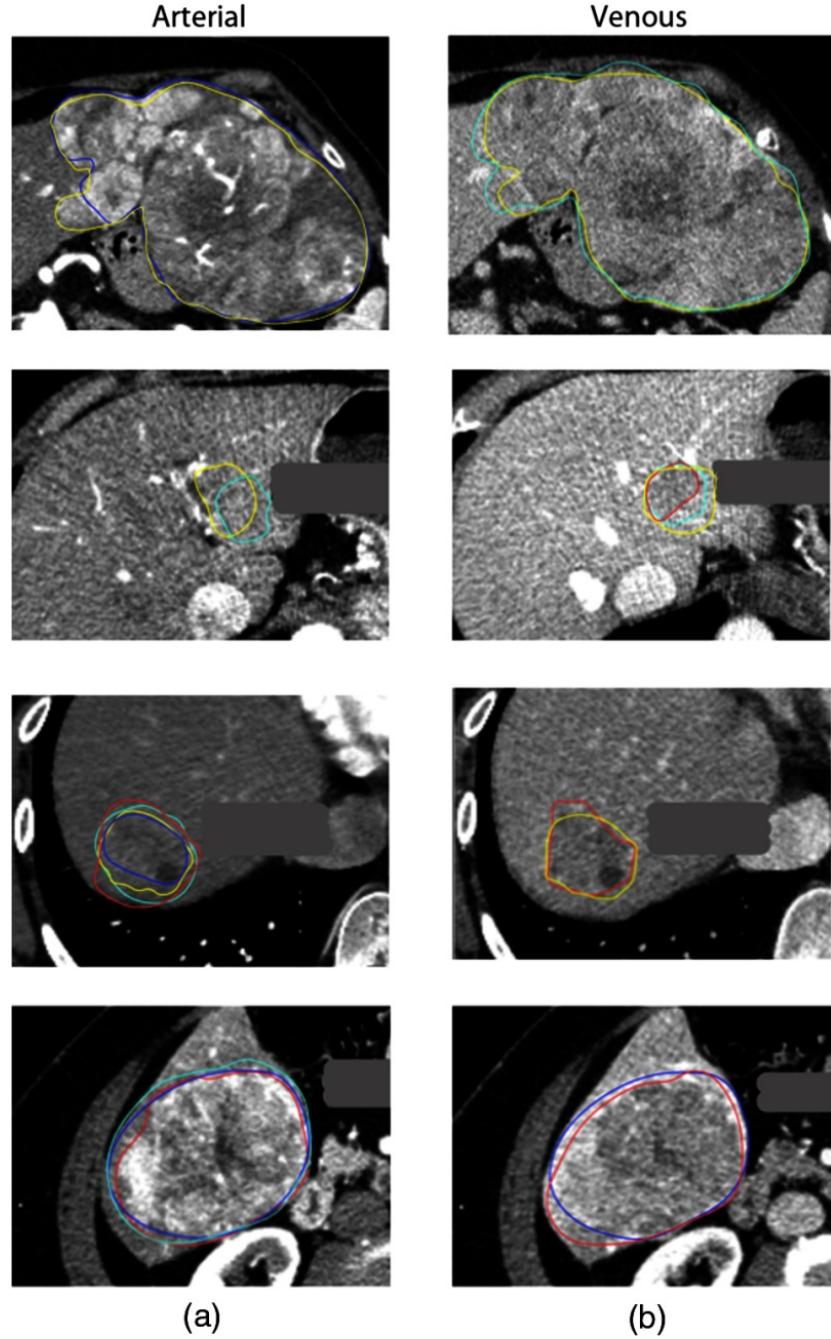


Figure 3.2: Inter-observer variability between 4 different observers on multiphasic images ((a) corresponding to arterial phase images and (b) to portal venous phase images), with a dice similarity coefficient from 0.19 to 0.93, ©Bakr et al. [127]

To reduce this bias, automatic segmentation techniques were developed. These techniques based on

the intensity of the pixels suffer from the fact that the intensity of the pixels, in the case of abdominal organs, is often close from one organ to another. On the other hand, models based on statistical representations often require the computation of an energy function, that can involve a large number of parameters, thus being difficult to compute and optimize (more details concerning the liver semantic segmentation techniques are given in the chapter 4).

3.2.1.3 Features

Once the volume of interest is delineated, different features can be extracted. On one hand, features can be chosen a priori for their ability to translate the physiological behavior expected by the experts. For example, studies performed on the lungs showed a correlation between the textural homogeneity and the survival of the patients [128, 129], or the grade [130]. This knowledge can also be used in the case of brain tumors to assess the response to a treatment, by observing for example the vascular or cellular density [115, 131]. However, a prior knowledge is not always available for the wanted task, therefore, the alternative is to extract a huge quantity of features and to determine the most relevant one by using for example some machine learning algorithms.

Features can be regrouped into different categories depending on their statistical order (first, second or higher order). For the features belonging to the first statistical order, the intensities of the volume of interest are summarized by an histogram, and different values are computed such as the uniformity or the entropy. Even though these features are often sensitive to the acquisition settings and by the way the histogram is computed, they already permitted the prediction of the malignancy of breast lesions [132].

Second order statistical features are meant to extract the textural properties of the volume, by considering the neighboring relationship between pixels. This will play an important role especially for the characterization of the heterogeneity of the tissues. This relation is captured by several descriptive matrices (GLCM, GLRLM, ...) [132].

Shape features are often extracted directly from the VOI in order to analyze its geometrical properties (such as the overall volume occupied by the tumor, its sphericity, its roughness or its fractal dimension).

Those features also permitted the prediction of the response to treatment in previous studies [126].

Finally, higher order features allow the extraction of imaging features in various frequency domains (the Wavelet features are the most commonly used higher order features [126]).

3.2.1.4 Features Selection

As listed here, a large quantity of features can be extracted and they tend sometimes to be highly correlated. The important number of extracted features can often cause overfitting during the creation of a predictive model. A features reduction phase is often required, either in a supervised manner (features are selected for their discriminant power in regards with the wanted task) or in non-supervised manner (the main objective is to suppress the redundant features without considering the different labels) [132].

Among the supervised methods, one can distinguish the univariate ones, where the features are tested one by one depending on their contribution to the final decision (Wilcoxon or Fisher test), from the multivariate where the features are regrouped into subsets before being tested against the output class (such as the Wrapper or the Embedded methods).

Unsupervised methods (such as the PCA: *principal component analysis*) do not consider the label of the data and try to reduce the dimensionality of the features space, however the new principal component are difficult to interpret.

Once the number of features is reduced, the next step is to construct a predictive model, by using either clustering methods (patients are regrouped based on a metric depending on the retained features, but their label can not be inferred from the clustering) or classification ones (where models such as RF: *random forests* or SVM: *support vector machines* are trained from the selected features in order to predict the wanted clinical criteria). Concerning the prediction of the survival, it is common to implement slightly different models such as the *Kaplan-Meier* or the *Cox Proportional Hazard* [133].

In the radiomics studies, one of the main goals is the stability of the features against the preprocessing steps described above. In order to reach this objective, it is possible for the patients to undergo the medical imaging examinations several times (*test-retest*), and the segmentations can be performed by several experts or even by the same expert several times [134].

In summary, when dealing with classical radiomics pipelines, reaching the best results will often depend on the best combinations between the extraction of the features, the technique used to reduce the number of features and the method implemented to create the model. Every modification on the cited steps can have a huge impact on the predictable performances of the created model.

In the next section, we will analyze the different HCR studies performed on patients suffering from HCC and who underwent CT examination. We will first describe the different choices made in each step of the classical pipeline, before presenting ways to improve the quality of future HCR work.

3.2.2 HCR applied to the HCC

In order to analyze the different methods implemented in the HCR field, we reviewed 15 studies performed on patients suffering from HCC and who underwent CT scan examination. Initially, 23 primary liver cancer-related studies have been scanned in our review [135], we then selected the 15 HCC-related ones.

We will first describe the different targets of the studies and the details of the cohorts through the number of patients and the clinical criteria that preceded their selection. We then compare the different imaging acquisition protocols, and the way the regions/volumes of interest are delineated. Finally, we will analyze the different features that appeared to be relevant in the studies, before proposing some tracks to improve the reproducibility and the performances of future radiomics work.

Details concerning the experimental settings of the studies, the endpoints and the different retained features can be found in the table 3.1.

Author	Modality & slice thickness	Mean tumor size	Treatment	#Patients & Inclusion Criteria	Segmentation	Computed features	Retained features	Study endpoints category	Results	%RQS (total points)
Cozzi et al. [136]	NECT 3mm	-	Radiotherapy (volumetric modulated arc therapy)	138 Patients BCCLC stages from A to C, Child-Pugh stages A-B	Segmentation done using the CTV (clinical target volume) which is manually contoured (whole tumor analysis)	35 extracted features 6 geometry and histogram	Compactancy Energy GLNU	OS & local control of the tumor	AUC of the model is 0.80	14 (5)
Zhou et al. [137]	Contrast CT (30 and 60s)	-	Hepatectomy	215 Patients who underwent partial hepatectomy	Largest cross-sectional area of the tumor, manual delineation Exclusion of necrosis	300 features (Mean, SD, Kurtosis, Skewness, GLM) ...)	Histogram features (skewness, energy, means, ...)	Recurrence	First-order statistical features combined with clinical factors can predict early recurrence	25 (9)
Akai et al. [138]	Contrast CT (27-28, 40 and 90s)	3.7cm (2.4-7.0cm)	Hepatectomy	127 patients	Manually setting the ROI to include the tumor within the slice at its max diameter. Single radiologist	96 features (mean, sd, positive value pixels, entropy, kurtosis, skewness)	Entropy, skewness and kurtosis	OS & DFS	First-order statistical features were sufficient to predict postoperative survival	25 (9)
Chen et al. [139]	Contrast CT (25 and 70s)	-	Hepatectomy	61 patients with only one lesion and survival dimension	ROI was delineated around the tumor outline at the longest dimension above 3 months	84 features 12 Gabor 9 Wavelet 7 GLCM	Textural features, Gabor and Wavelet as key features	OS & DFS	Tumor prognosis could be predicted using Gabor and Wavelet responses	17 (6)
Li et al. [140]	Contrast CT (70s) and 1.25mm	8.0cm (5.1-18.7cm)	Hepatectomy or TACE	130 patients treated by LR and 22 by TACE	Irregular ROI manually drawn around the largest-cross sectional tumor outline	27 features (Wavelet)	2 Wavelet features correlated with survival	OS and Treatment sensitivity	Wavelet features correlated with survival suggesting a suitable treatment choice	19 (7)

3.2. HANDCRAFTED RADIOMICS

Author	Modality & slice thickness	Mean tumor size	Treatment	#Patients & Inclusion Criteria	Segmentation	Computed features	Retained features	Retained features category	Study endpoints	Results	%RQS (total points)
Raman et al. [141]	Contrast CT (25s) 3mm	Adenoma 7 ± 3cm FNH	-	80 patients multiple axial slices (from 5 to 10 slices) 17 FNH 19 Adenomas 25 HCCs	ROIs were selected from multiple axial slices (mean, SD, entropy, skewness, kurtosis)	32 features (mean, SD, entropy, skewness, kurtosis)	SD and Mean of Quantitative features	Diagnosis	First-order statistical features able to differentiate 3 types of hypervasculat lesions with a 15% of error rate)	3 (1)	3 (1)
Kuo et al. [142]	Contrast CT (30-35 and 60-70s) 2.5mm	- -	-	30 Patients no patients received chemo before resection	No segmentation by 2 experts.	6 imaging traits (internal arteries, textural heterogeneity, wash-in-wash-out, necrosis, tumor margin)	Tumor margin	MVI status	The tumor margin showed strong correlation with MVI, TNM. Internal arteries showed correlation with MVI	19 (7)	19 (7)
Banerjee et al. [143]	Contrast CT (35, 60-70, 180-300s) 2.5-3mm	2.8cm (1.8-4.5cm)	Hepatectomy or LT ¹	157 patients 72 resection 85 LT ¹ MVI diagnosed in 45 patients	Only imaging features were evaluated by 5 radiologists	3 imaging traits (internal arteries, hypodense halo, tumor-liver difference)	The 3 imaging traits were retained	Semantic	OS and RFS	Combination of the three different imaging traits was correlated with MVI	53 (19)
Renzulli et al. [144]	Contrast CT (25-30, 45-60, 180-300s) 2.5mm	3.3cm (1.8-5.2cm)	Hepatectomy	125 patients where hepatic resection was indicated	Only imaging features were evaluated by 2 radiologists	5 imaging traits Dimensions Lesions number Non-smooth margins TTPVI ²	All except the lesion dimensions	MVI status	The 4 retained traits were correlated with the presence of MVI in HCC	8 (3)	8 (3)

¹ Liver transplantation

² Two-Trait Predictor of Venous Invasion: Internal arteries and Hypoattenuating halo

Author	Modality & slice thickness	Mean tumor size	Treatment	#Patients & Inclusion Criteria	Segmentation	Computed features	Retained features	Study endpoints	Results	%RQS (total points)
Segal et al. [145]	Contrast CT (three-phasic)	-	Hepatectomy	Train: 30 Test: 32 2 radiologists.	No segmentation, visual examination	32 imaging traits (Capsule, Wash-in-Wash-out, Tumor-Liver difference, ...)	Internal arteries and hypodense halo	OS & MVI	Internal arteries combined with hypodense halo can predict OS, MVI	42 (15)
Zheng et al. [146]	Contrast CT (22 and 60s) 5mm	-	Hepatectomy	Train: 212 Test: 107 patients without anticancer therapy	ROI delineated around the tumor outline of the largest cross-sectional area. 2 radiologists	110 GLM features	Quantitative	Recurrence & OS	Textural features sufficient to predict postoperative recurrence and survival	47 (17)
Peng et al. [147]	Contrast CT (30, 60 and 120s) 5mm	4.9-6.4cm	-	Train: 184 Test: 120 Partial hepatectomy with pathologically confirmed HCC	ROI semi-automatically segmented in the largest cross-sectional area by 2 experts	5 imaging traits (tumor margin, peritumoral enhancement, hypoattenuating halo, internal arteries + 8 internal arterioles, tumor-liver difference) & 980 quantitative features	Nonsmooth tumor margins, hypoattenuating halo, and internal arteries + 8 radiomics features	MVI status	Radiological features and a radiomics signature computed with first-order statistical features showed correlation with MVI	47 (17)
Bakr et al. [127]	Contrast CT (AR with bolus tracking, PV, delay) Thickness \leq 3mm	7.4cm	-	28 patients with surgical resection of the tumors (one central, a previously untreated HCC	3 ROIs were placed on different cross sections of the tumors (one central, one superior and one inferior)	464 features (intensity, texture, shape)	Textural features	MVI status	Textural features computed using single- or combined-phased images were correlated with MVI	3 (1)

3.2. HANDCRAFTED RADIOMICS

Author	Modality & slice thickness	Mean tumor size	Treatment	#Patients & Inclusion Criteria	Segmentation	Computed features	Retained features	Study endpoints category	Results	%RQS (total points)
Taonli et al. [148]	Contrast CT (AR with bolus tracking, PV at 70s, Delay at 180s)	5.7±3.2cm	-	38 patients 26 CT/ MRI	Global inspection of the imaging traits and “slice-wise” evaluation for the enhancement ratio and the wash-out ratio	11 imaging traits (wash-in/washout, hypovascular-hypervascular ratio)	Infiltrative pattern, mosaic appearance, presence of MVI, large size	Semantic & Quantitative of MVI & aggressive phenotype	Correlation found between some imaging traits and the aggressive profile of the tumors	19 (7)
Xia et al. [149]	Contrast CT (30, 55, 70, 300s) 2.5-5mm	12 tumors, smaller than 5cm, 26 larger	Hepatectomy or LT	38 patients	Tumor was firstly delineated then divided into 3 spatially distinct sub-regions (using a multi-parametric clustering)	37 features (1st order, geometry, textural) And 4 features for the whole tumor	Volume of transition region & cluster prominence	Quantitative OS	The volume of transition between tumor and liver, and the heterogeneity of the lesion were correlated with survival.	22 (8)

Table 3.1: HCR reviewed studies details

3.2.2.1 Experimental setup

The vast majority of the studies were designed to predict the survival of the patients after surgery, or any other type of treatment [136, 138–140, 143, 145, 146, 149]. In clinical trials, the traditional way to evaluate the survival is through the OS (Overall Survival), which corresponds to the duration from either the date of diagnosis of the disease or the start of its treatment, to either the end of the trial or the death of the patient. Being often assimilated to the survival rate, new finer metrics tend to be preferred such as the DFS (Disease Free Survival), which corresponds to the duration from the beginning of the treatment to the date of the recurrence of the disease.

Other ways to evaluate the response to a given treatment were also evaluated in some of the reviewed studies, such as the presence or absence of recurrence [137, 146], the local control which assess the end of the growth of a tumor [136], and other ways to compute the sensitivity to a treatment [140, 142].

Another important aspect that is often assessed by the reviewed studies is the physiological changes brought by the disease, such as the aggressive profile of the tumor, usually translated by the presence of MVI (*Micro Vascular Invasion*), and its association to genes expression [127, 142–145, 147, 148].

The entire 15 studies had a retrospective design, and the number of selected patients varied from 28 [127] to 319 [146], with a median of 125 patients per study, as illustrated in the figure 3.3.

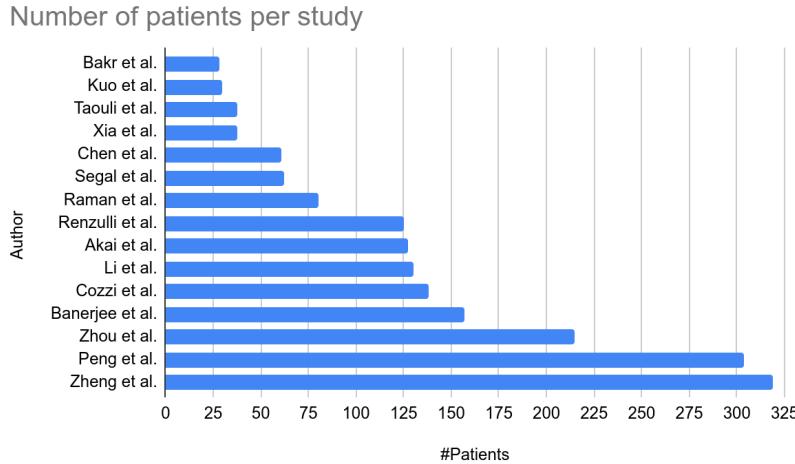


Figure 3.3: Number of patients included in the reviewed HCR studies

The relatively low number of patients can be explained by the stringent inclusion criteria set by the studies (size and number of lesions, baseline imaging examination within a given period of time before initial treatment, ...)

They were all built with data from patients who underwent CT examination, except one, that decided to mix data obtained from both CT and MR examinations [148].

Regarding the CT scan protocol, only one study decided to use images acquired before the injection

of contrast agent [136], whereas two other studies used images acquired at only one phase (early arterial phase for [141], and portal venous phase for [140]), and the remaining studies analyzed multiphase images. Among them, four studies utilized images acquired at both early arterial and portal venous phases [137, 139, 142, 146], while the rest were based on traditional triphasic images [127, 138, 143–145, 147–149]. Concerning the acquisition protocols, early arterial phase images are most of the time acquired around 30s after the injection of contrast agent (between 22 and 35s), and some studies used bolus tracking method to estimate the best acquisition moment instead of using the same timing for all the patients. Portal venous phase images are acquired between 45 and 70s after the injection, and the delayed images obtained during an even larger interval (between 90 and 300s after the injection).

Knowing that images are the key elements in the computation process of the radiomics features, this high variation within the acquisition protocols is the first reason why the standard HCR pipeline should be standardized.

Once the images acquired, the following step consists in selecting the region of interest to compute the features, or evaluate the physiological properties of the lesions to the naked eye.

3.2.2.2 ROI selection and features extraction

The selection of the region/volume of interest and/or the assessment of the physiological characteristics of the tumor is often performed by one or multiple experienced radiologists. This step of the pipeline was performed by only a single expert in rare cases in the reviewed studies [138, 149], whereas it was usually performed by two experts [137, 139–142, 144–148]. When more than one expert is involved, the authors decided to implement ways for quantifying the inter and intra-observer variability, such as the ICC (*Intraclass Correlations Coefficient*) [140, 146], or the Cohen-k statistics [143, 144]. Regarding now the selection of the ROI, we can separate the reviewed studies by the type of features being extracted. On one hand, when using imaging traits corresponding to visual inspection of the tumor characteristics, it is common to use the whole tumor to perform the evaluation. For example when evaluating the vascular invasion, some features like the peritumoral enhancement need to be estimated globally [142–145]. On the other hand, when using computational features, the analysis can be performed on the whole tumor or on a single slice. Some studies decided to compute the features on the entire 3D ROI, as an example Cozzi et al. [136] used the entire tumor independently on the tissues present within it , Xia et al. [149] decided to separate areas based on their textural properties as depicted in the figure 3.4, but the analysis remains global. Other studies placed several ROIs all throughout the tumor (3 ROIs for Bakr et al. and from 5 to 10 for Raman et al.). However, the majority of the studies decided to place the ROI at the largest cross-sectional area [137–140, 146].

Note that some studies decided to compute the features using both the entire volume and the

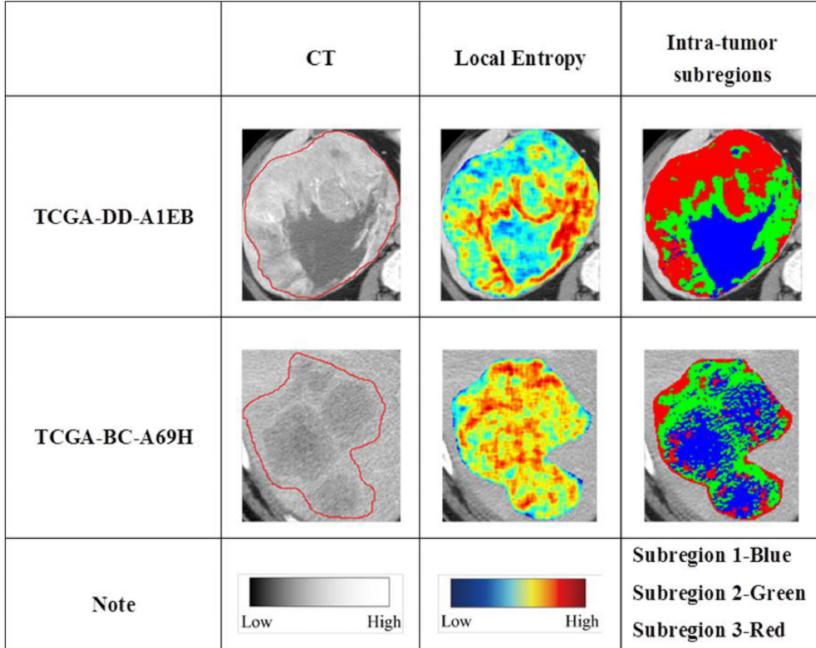


Figure 3.4: Illustration of intratumor partition for two representative patients (TCGA-DD-A1EB and TCGA-BC-A69H). The first column shows the tumor outlines on the original CT image. The second column shows the heatmaps of calculated local entropy on tumor images. The third column shows the three subregions marked with different colors after intratumor partitioning. ©Xia et al. [149]

slice-wise approach (e.g. Taouli et al. [148] evaluated imaging traits globally and computed the ratio using a slice-wise fashion, Peng et al. [147] did the same in their study by computing features using an ROI placed at the largest-cross sectional area, and evaluating imaging traits globally).

Worth mentioning that before the computation of features, it is common to filter the images using different kernel sizes, in order to enhance different elements of the volumes such as the blood vessels for example. All reviewed studies that computed quantitative features filtered their images with a Laplacian of Gaussian algorithm to extract features of specific sizes, as depicted in the figure 3.5.

3.2.2.3 Features selection

Once the ROI delineated and pre-processed, the following step consists in extracting the features, and as explained previously, the choice on which features to extract depends on the type of features the study is going to rely on, quantitative or semantic.

In the case of quantitative features, the reviewed studies often decided to focus on a single category of features (first-order, textural features, higher-order...), thus obtaining a relative small number of features (27 for Li et al., 32 for Raman et al.). However, despite choosing a specific category of features, this number can increase, when combining the native features with the spatial filter, and the different contrast enhanced phases, as in Akai et al. where a total of 96 features are extracted from the initial 6 histogram-based features [138].

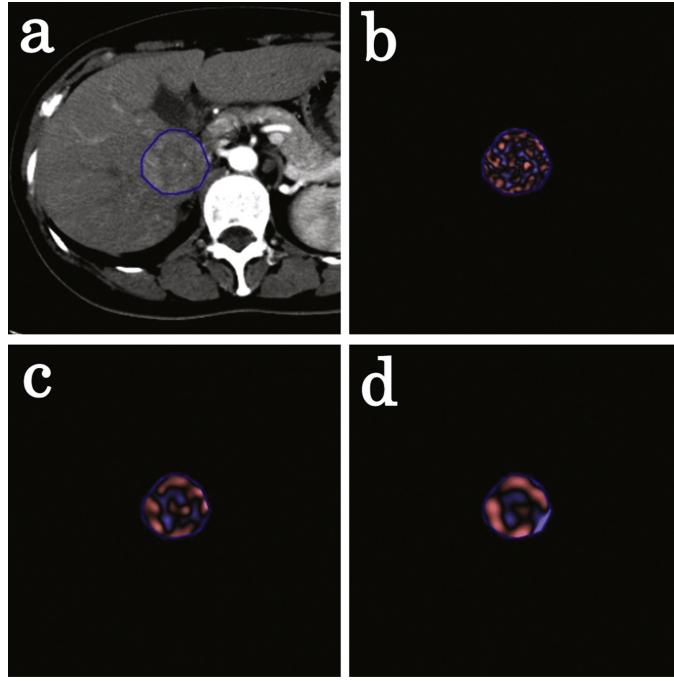


Figure 3.5: Screenshot of the CT texture analysis software. A polygonal ROI was drawn on the tumor (a). Processed images using Laplacian of Gaussian filters with SSF of 2 mm (b), 4 mm (c), and 6 mm (d) were automatically generated. The images were displayed using a red or blue scale showing negative or positive pixels, respectively. ©Akai et al. [138]

Some other studies decided to extract the maximum number of possible features by combining the previously mentioned group of features, and thus obtained 300+ features [127, 137, 147]. The problem in this case, often called “*curse of dimensionality*”, corresponds to a high number of features relative to the number of individuals, and that can cause some troubles when further training the predictable model.

When imaging traits are preferred to classical radiomics features, the number of extracted characteristics is generally below 10, with a high predominance of changes brought by the hepatocarcinogenesis, such as the presence of internal arteries, or the wash-in wash-out effect (example of imaging traits can be observed in the figure 3.6 & 3.7). Even though the number of extracted features is often small, they often correspond to the absence or presence of physiological properties that are sometimes difficult to quantify and that will highly depend on the observer experience.

After the different features are obtained, the next step in the pipeline consists in the selection of the features and the building of the predictive model.

The vast majority of the reviewed studies decided to implement a logistical regression model in order to assess the correlation between features and the study endpoint. Among the existing methods, time-related approaches such as the Cox regression model [136, 140, 143, 146, 149] and the Kaplan-Meier survival analysis [138, 139, 145, 149] are the most commonly used.

Different statistical approaches were used by the studies that tried to determine the link between

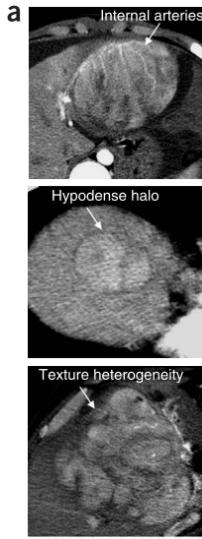


Figure 3.6: From top to bottom: Internal arteries, Hypodense halo, Texture heterogeneity, as illustrated by ©Segal et al. [145]

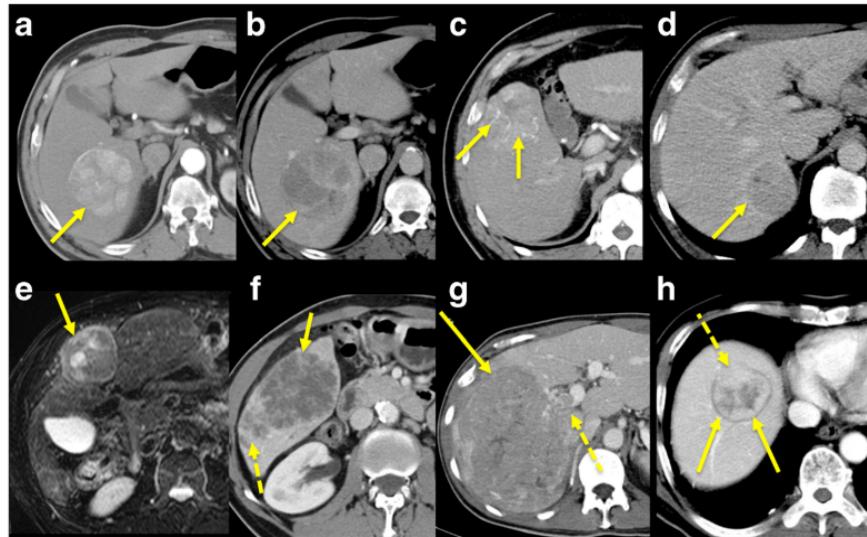


Figure 3.7: (a,b) wash-in/wash-out pattern and mosaic appearance, without capsule/ pseudocapsule. (c) internal arteries (arrows) (d) pseudo-capsule (arrow) (e) [MR image] hyperintense encapsulated with mosaic appearance (arrow) (f) internal necrosis and satellite lesions posteriorly (dashed arrow) (g) right portal vein invasion (dashed arrow) (h) extra-nodular growth anteriorly (dashed arrow) ©Taouli et al. [148]

selected features and study endpoints such as the LASSO (*Least Absolute Shrinkage and Selection Operator*) algorithm [127, 137, 147]. In the other studies, other approaches were implemented, for example, Raman et al. performed a PCA followed by a MANOVA (*Multivariate Analysis Of VAriances*) to create clusters among patients for the classification of hypervasculär lesions, whereas Renzulli et al. evaluated the positive and negative predictive values of the selected features against the microvascular invasion status of the patients.

The list of discriminant features per study is given in the tables 3.2 & 3.3 , with a separation between quantitative and semantic features.

Table 3.2: List of quantitative features used in the reviewed studies

<i>First Order Statistics</i>
<ul style="list-style-type: none"> • Shape [136] • Skewness [137, 138, 140] • Kurtosis [138] • Mean [136, 137, 141] • Energy [136, 137] • Entropy [138, 147] • Peak [127] • Standard deviation [149] • Enhancement ratio [148] • Tumor-Liver difference [148]
<i>Second Order Statistics</i>
<ul style="list-style-type: none"> • Gray Level matrices [136, 146, 147] • Cluster prominence [149]
<i>Higher Order Statistics</i>
<ul style="list-style-type: none"> • Wavelets [127, 139, 140] • Gabor [127, 139]
<i>Morphological features</i>
<ul style="list-style-type: none"> • Tumor margin volume [149] • Tumor size¹ [144, 148]

Table 3.3: List of semantic features used in the reviewed studies

<i>Two Traits Predictor of Venous Invasion</i>
<ul style="list-style-type: none"> • Internal arteries [142–145, 147, 148] • Hypoattenuating halos [143–145, 147]
<i>Intensity-related features</i>
<ul style="list-style-type: none"> • Peritumoral enhancement [144] • Presence of a Tumor-Liver difference [143]
<i>Textural-related features</i>
<ul style="list-style-type: none"> • Non-smooth Tumor Margins [142, 144, 147] • Infiltrative patterns [148] • Mosaic appearance [148]

¹ Not as an exact value, but classified (for example the categories were smaller than 2 cm, between 2 and 5 cm, and larger than 5 cm in [144])

Concerning the studies based on quantitative features we can notice that first-order statistical

features is the most common discriminant type, which is normal because this group of histogram-based characteristics is often implemented in the existing radiomics tools, whereas higher order statistical features require more advanced knowledge to be implemented, and often lack of interpretability.

Even though Li et al. [140] decided to extract only Wavelet features because they consider that the current way of computing textural features is too dependent on the imaging acquisition settings, first and second-order statistical features remain a good indicator of the textural heterogeneity which is often correlated with the physiological advances of the disease.

Worth also noting that imaging traits, often analyzed either alone, or in combination with quantitative features, remain discriminant enough in a lot of studies, because they are directly linked to the physiological changes produced by the disease. Despite needing expertise to be extracted, such as in Segal et al. where 32 different imaging traits were analyzed, their predictable power drives us to consider them in future radiomics studies and focus on a way to quantify them [145].

3.2.2.4 Study reproducibility

Although the majority of the reviewed studies obtained good predictable results in regards with the wanted prediction task, their stability to the experimental settings and their reproducibility remain questionable.

In 2017, Lambin et al. [150] who remains one of the founders of the radiomics fields [151] published a study proposing a way to rethink the HCR pipeline and assess the reproducibility of future radiomics studies . This assessment is performed thanks to the RQS (*Radiomics Quality Score*), which evaluates a total of 16 components with various weights. The evaluation of the different components allows the computation of a score ranging from 0 to 36 points where highest weights are given to criteria allowing a better reproducibility of the study, such as the prospective aspect of the study (7 points over 36) or the presence of a validation step in the proposed workflow (5 points over 36, with a penalty of 5 points when no validation at all is present).

Our reviewed studies were evaluated in regards with the RQS, in consensus with a medical research fellow (Wakabayashi Taïga) [135]. The details of the different scores per study are reported in the table 3.4.

Table 3.4: RQS score details per criteria for the reviewed studies

Criteria	Bakr et al. [127]	Kuo et al. [142]	Taouli et al. [148]	Xia et al. [149]	Chen et al. [139]	Segal et al. [145]	Raman et al. [141]	Renzulli et al. [144]	Akai et al. [138]	Li et al. [140]	Cozzi et al. [136]	Banerjee et al. [143]	Zhou et al. [137]	Peng et al. [147]	Zheng et al. [146]
Image Protocol quality	1	2	2	2	1	2	2	2	2	2	1	2	2	2	2
Multiple segmentations	1	0	0	0	1	1	0	0	0	0	1	1	1	1	1
Phantom Studies	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Multiple time points	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Features reduction	-3	3	3	3	3	3	-3	3	3	3	3	3	3	3	3
Multivariate analysis	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1
Biological correlates	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0
Cut-off analysis	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1
Discrimination statistics	2	1	0	0	1	1	1	1	1	2	1	1	1	1	1
Calibration Statistics	0	0	1	2	0	1	0	0	0	0	1	1	1	0	2
Prospective study	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Validation	-5	-5	2	2	2	0	0	0	3	-5	0	0	5	-5	2
Gold standard comparison	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Clinical utility	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Cost-effectiveness analysis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Open science Data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Total	1	7	7	8	6	15	1	3	9	7	5	19	9	17	17

The mean obtained RQS by the reviewed studies was 8.73 ± 5.57 points, which corresponds to less than 23% of the maximal possible score, and only one study obtained a more than 50% of the maximum points, which translates the lack of robustness of the current HCR state-of-the-art studies applied to the HCC.

The figure 3.8 allows us to see the most respected criteria in regards with the RQS guidelines.

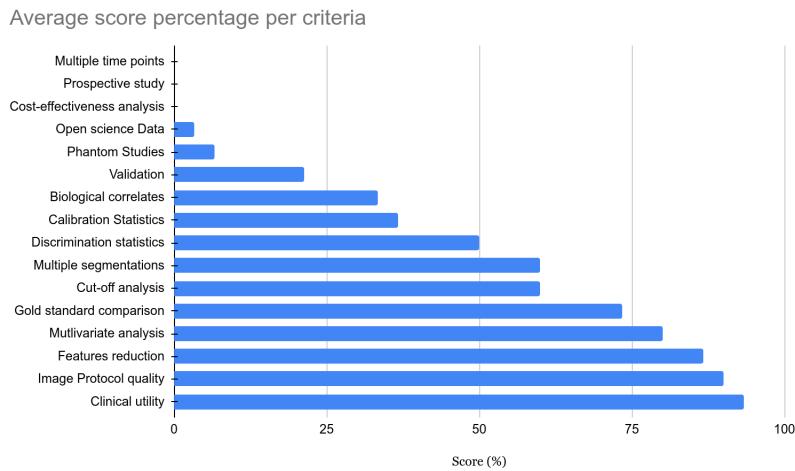


Figure 3.8: Average score percentage per criteria in the RQS for the reviewed studies

The criteria the less respected by the different reviewed studies were all those related with the reproducibility and the robustness. Among them, the analysis of images acquired at multiple times, the implementation of phantom studies to detect inter-scanner differences and features sensitive to those settings, or even the prospective design of the studies that will ensure inclusion of patients undergoing the same protocols were almost always neglected. Even though the different reviewed studies obtained successful results in regards with the wanted task, their robustness relative to the experimental settings and their reproducibility can be questioned, especially when analyzing their results obtained on the radiomics quality scoring system [150].

One way to allow a better reproducibility for the future radiomics studies is to fulfill the maximum possible criteria introduced by the RQS standard, and to use standard ways to compute the quantitative features, thanks to open-source libraries such as pyradiomics [152]. Another way is to reduce as much as possible the impact of human-based interpretation in the HCR workflow. This can be done by replacing the hand-crafted annotations, and the engineered process of features computation and selection by a standardized, automated and data-driven pipeline, similar to what is performed in more recent DLR studies.

3.3 Deep Learning Radiomics

In this section we will first present the value brought by the deep learning in the medical imaging field, before focusing on the differences between the HCR and the DLR strategies. We will then describe in detail the DLR concept before presenting the different reviewed studies tackling liver-related problems using a DLR approach. We will outline the different steps of their pipelines such as the use of multiphasic images, the way they incorporated experts' annotations and their choice regarding the deep network architectures.

3.3.1 DL applied to the medical imaging

The main breakthrough brought by DL was its ability to detect morphological properties in images only by using the pixel intensities as input, whereas traditional machine learning methods often required sophisticated hand-crafted features to achieve descent results [153, 154]. Deep learning networks achieved state-of-the-art results in many medical-related applications among which classification, localization, detection, registration and segmentation [155]. Even with a small number of training cases, those performances were realized thanks to architecture choice, data augmentation or transfer learning [146, 156]. As exposed previously, the automatic segmentation of the tumors brings a volumetric information, that is more powerful than the diameter only, and that allows to compute the tumor burden, which has an importance when estimating the efficiency of a given treatment [157–162].

3.3.2 Difference between HCR and DLR

Being a recent field, radiomics starts to mature when applied to several organs such as the lungs or the breast, but it still struggles when applied to the liver, especially because of the scarcity of available data and the complexity of its anatomy. As exposed previously, the first studies targeting liver cancer almost always rely on hand crafted features. The main limitation observed is the lack of reproducibility originating from engineered features, that sometimes result from complex processes which can be difficult to imitate, and often fail to work on different databases. HCR rely on manual or semi-automatic expert annotations, which often require a complex and time-consuming process, that also provide several constraints, such as the poor inter-expert reproducibility. The extracted features are not necessarily relevant to encode the observed structure, it is needed to increase their number, hence requiring a complex dimensionality reduction step. Another limitation is the difficulty to find the perfect association between features extraction, selection and the statistical analysis used to reach the wanted target.

Those limitations, and the rise of new computational resources associated with the emergence of deep learning techniques allowed the development of a new branch in the radiomics field, called DLR (Deep-Learning Radiomics) [121].

3.3.3 DLR workflow

In this branch of radiomics, the features are extracted through a deep-learning process, avoiding a manual extraction. A neural network can be trained to generate the most relevant features. These features can either be kept in the network for the final pathological target prediction, or used as input in a different model (such as a SVM or a RF). Compared to the HCR, no prior knowledge is required, and the features can be extracted in an end-to-end manner, using only the raw images as input, and without necessarily providing any segmentation. It has also been demonstrated that performances of those networks increase with the size of the training dataset [163]. Eliminating the segmentation phase when evaluating the diagnosis, allows to reduce the workload of experts, and provides a solution to the observer-dependency. When training DLR networks, the original image can be combined with the segmentation or any other pre-processed step result such as the gradient image for example [164] to improve the relevance of the extracted features.

Generally, DLR studies can be classified depending on the type of input used, the training strategy or the type of architecture chosen to extract the features.

As input, deep radiomics networks can consider 2D slices independently, however this technique does not bring sufficient information since the decision mainly depends on the entire volume of interest. The different outputs obtained in a slice-wise fashion can be fused to get a volume-wise decision. The volume by itself can also directly be used as input, however it can raise several issues such as the size of the voxels or the slice thickness. Finally, the classification can be performed by considering a series of volumes corresponding to the entire examination of the patient [165], but here again the question regarding the normalization of the input can be raised.

Once the type of input is chosen, the studies differ depending on the type of network and the training strategy utilized. The networks can be trained *from scratch* using only the available data or a pre-trained architecture can be fine-tuned. In the first case, the obtained network will be specific to the wanted task, but this specialization can also lead to troubles such as overfitting or the sensitivity to class imbalance. The impact of those problematics can be limited with the help of data augmentation (use of existing data to generate new artificial samples) [166], multi-task training (where the number of parameters is limited by the training of several task using the same network) or the incorporation of the proportion of each class present in the data when building the cost function [167]. The other strategy consists in using an architecture pre-trained most of the time on natural images, and then re-train

only a specific part using the desired dataset [168–170]. It is worth noting that this type of training constrains the pipeline to be slice-wise since existing architectures are often pre-trained on 2D images. Finally, the features can be extracted using either a supervised or an unsupervised approach. In the supervised case, the most commonly used networks are based on convolutional layers (CNN), followed by one or multiple dense layers to predict the output class. While the network is trained to perform the classification, the features are extracted either after a fully connected layer [170], or after one of the convolutional layers [171].

Other variants that are also built with convolutional layers as key components can also be implemented (RNN: *Recurrent Neural Networks*, LSTM: *Long Short Term Memory* or *Capsule Net*). Their goal is to get rid of the limitations caused by the input format that need to be fixed, and by the difficulty to consider an entire 3D volume during the training [172]. In the unsupervised case, the objective is to learn the data distribution, so that new data from the same distribution can be generated. The most commonly used architecture in this case is the *auto-encoder*, made up with a part that contracts the information (encoder), in order that the most useful one is conserved to regenerate the original data (decoder). Auto-encoder can be built on top of convolutional layers [168], or trained with the aim of being insensitive to the noise added to the input data (*denoising auto-encoder*) [164, 173]. Following the same principles which are to reconstruct the original input data using only the most relevant features, some studies implemented other type of generative models such as the DBN (*Deep Belief Networks*) [164] or Deep Boltzmann Machines [174].

Some studies are referred to as hybrid, when features are combined with other sources of data (combination of different modalities [133] or association with clinical data such as genomic data [175]), or when only a part of the pipeline implements deep learning methods, either for the extraction of the features [170] or when the decision is taken with a fusion between HCR and DLR features [169].

3.3.4 DLR applied to the liver

Even though the number of studies targeting the liver is increasing, the vast majority of them can be categorized as HCR, and only a few are currently based on deep learning. The main reason behind that is the late emergence of deep learning and the recent outbreak of new architectures and concepts that are often first developed and evaluated in other fields than the medical imaging one (e.g. Residual Network, DenseNet, Capsule Net).

However, we have decided to review the first DLR-liver related studies, in order to understand how DL architectures can successfully be incorporated in a radiomics pipeline. We describe the study endpoints, the data preprocessing, the implementation and training strategies, before analyzing their performances.

A more detailed analysis of the reviewed studies is available in the following link ⁴.

3.3.4.1 Study Endpoint

Within the reviewed DLR studies, the majority of them are targeting a characterization of the tumor, either the classification of FLLs (*Focal Liver Lesions*) [176–179] or the estimation of the fibrosis stage [180], when two of them focused on the response to treatments, either for recurrence after resection [181] or the response after TACE (*TransArterial ChemoEmbolization*) [182].

We selected the studies using CT images to perform their analysis, and realized that the vast majority of them used multiphasic images to perform their research, knowing that the evolution of contrast medium is often correlated with the pathological features of the liver as mentioned in the chapter 2. The only one that used single phase images, was the one targeting an estimation of the fibrosis stage, and their method was based on portal phase images only [180]. Regarding the multiphasic studies, there is no consensus about the delay between the injection of the contrast agent and the acquisition of the different phases. They tend to prefer triphasic acquisition, with images acquired before the injection of contrast agent, at early arterial phase and a third phase, either portal venous [177, 179, 181] or a delayed one [176, 178]. Peng et al. decided to get rid of the NECT (*Non-Enhanced CT*) phase, but still chose a triphasic acquisition (AR, PV, DELAY).

3.3.4.2 Image processing pipeline

Concerning the image processing pipeline, the data used to train the deep networks are most commonly selected via placement of a bounding box by the experts around the hepatic lesion, before a registration between the different phases in case of multiphasic acquisition to compensate the effects of body motion and/or breathing.

The manual delineation of the *ROIs* is usually done by one or more experts on the raw image [176–178, 180, 181], and only one study decided to perform an automatic segmentation of both the parenchyma and the lesion with the application of a random-walker algorithm, before being checked by experts [179]. When the method is based on a 2D approach, selected images are often those presenting the maximal cross-sectional proportion of the lesions, except one study targeting the estimation of the fibrosis stage, that centered the ROI so it displayed the ventral aspect of the liver [180]. Only one study incorporated 3D information in their pipeline, but they also started the placement of the VOI with the slice presenting the maximal proportion of tumor, and extended it to adjacent slices [183].

After placing a bounding box, the images are registered either manually [176, 177, 181] or via the application of a non rigid registration with anatomical constraints [179]. No real registration was

⁴ <https://docs.google.com/spreadsheets/d/1VktRkGiX0J7RzQEVxRq1LxKPfV6YsrzHhctr0xlkEBY/edit#gid=0>

mentioned for two studies [180, 182] but Yamada et al. evaluated the effects of the registration in the prediction performances of their networks (as depicted in the figure 3.9), and after training several networks with misregistered data (shifted, rotated, skewed) they concluded that in the vast majority of the cases, no statistical significance can be found between the performances of the networks trained with manually registered data and those of the networks trained with misregistered data [176].

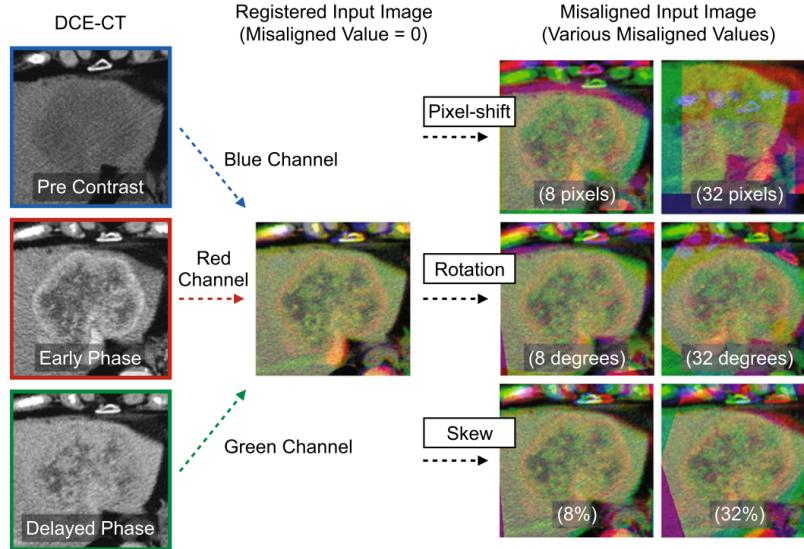


Figure 3.9: Illustration of the manually registered images and the effect of transformations (shift, rotation, skew) in the three phases, by ©Yamada et al. [176]

It is worth noting that some studies performed their analysis on cropped or resized JPEG images [178, 180, 181]. This choice might cause loss of data whereas the raw images are used elsewhere with the native HU intensities.

Finally, in the reviewed studies, only two of them combined images with clinical data [180, 181], whereas the other only used image data, which is comprehensible because clinical data are often difficult to retrieve, and can also be difficult to integrate in a deep network, as illustrated in the figure 3.10.

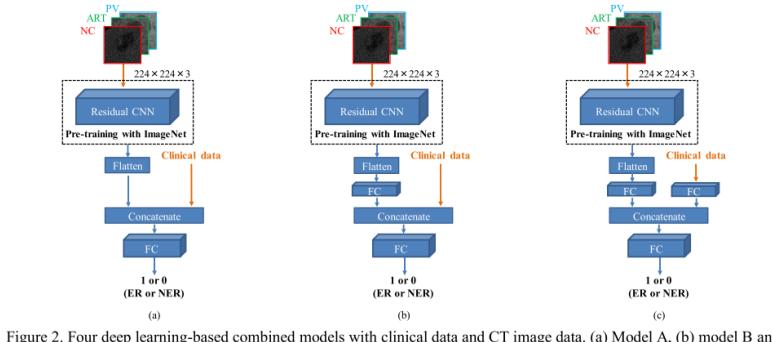


Figure 2. Four deep learning-based combined models with clinical data and CT image data. (a) Model A, (b) model B and (c) model C

Figure 3.10: Illustration given by ©Wang et al. of the different tested deep networks where raw images are combined with the clinical data [181]

3.3.4.3 Architecture choice

The reviewed studies differ mainly in the way they built their deep architecture.

In most of the cases, convolutional layers are used for the extraction of the most discriminant features. The main question being whether to use a pre-trained network or to train a network from scratch. In the case of fine-tuning, the most commonly used architectures are the AlexNet and the ResNet [176, 177, 181, 182], but it is also usual to compare the results obtain by different pre-trained architectures [176, 181]. The general method is to recycle an architecture trained on a huge dataset such as ImageNet, freeze the weights of the early layers (responsible for the high levels features), adjust and train the last layers on the current database to be more specific. An illustration of this process can be found in the figure 3.11.

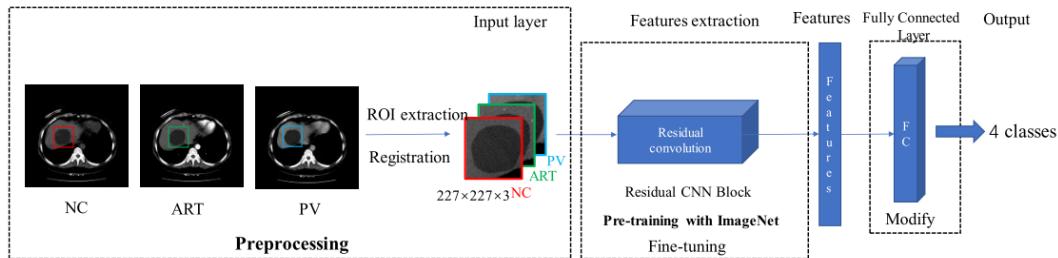


Figure 3. Overview of our method

Figure 3.11: Illustration of the pre-training strategy adopted by ©Wang et al. [177]

The rest of the reviewed studies created a custom architecture and trained it from scratch [178–180]. Two of them used classical convolutional layers followed by max pooling layers, early in the network to extract relevant features, as depicted in the figure 3.12.

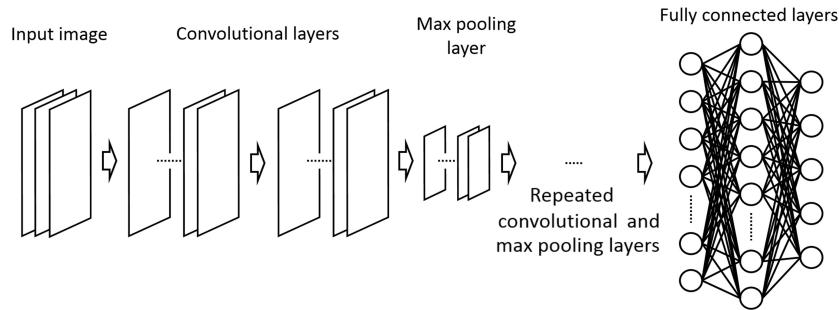


Figure 3.12: Architecture used by ©Yasaka et al. where the features are first extracted through convolutional layers before the prediction is performed using fully connected layers. [178]

Since multiphasic studies often stacked the different phases as channels to feed the deep network, one study decided to extract the temporal information through a different paradigm by using LSTM layers [179]. As depicted in the figure 3.13, their architecture first extracted the features using what they called a “ResGLBlock” per phase, with two scaled data as input (a large one with the entire lesion, and a smaller one with finer details), and conserved the temporal information via bidirectional LSTM

layers.

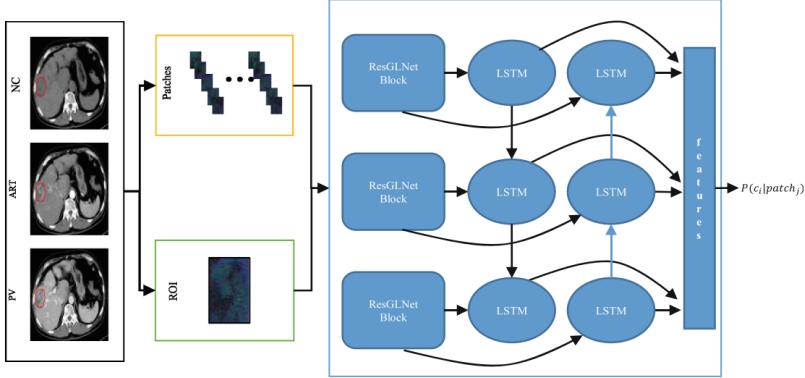


Fig. 1. The flowchart of our framework

Figure 3.13: Architecture used by ©Liang et al. where patches of different scales are extracted from the original images, before being used to train residual networks connected with bidirectional LSTM layers [179]

When using a pre-trained architecture, the input size is often dictated by the native architecture (224x224x3 for example for the ResNet architecture [181,182]), and the different studies need to resample their inputs to fulfill those requirements, which can sometimes affect the performances of the network, whereas custom architectures allow a usage of custom sizes [178–180]. However, the size of the lesions or other extracted ROI is often different from one patient to the other, therefore, this problem is still open.

One way to render the deep networks robust to those changes is through multiscale training. Yasaka et al. for example trained the network with patches cropped at different resolutions from the initial lesion ROI after application of standard geometrical transformations such as rotation or shift [178]. The same process of extracting different patches from an initial ROI is performed by two other studies, where the goal is also to balance the different classes [180, 182].

The networks are then usually trained in a cross validation fashion [176, 178, 180, 181], or validated on external datasets [182] to be less affected by the effect of randomness, and to be less prone to overfitting.

3.3.4.4 Performances

Regarding their performances on their testing sets, the different studies concluded first that fine tuning allows an improvement in the accuracy of the DLR network, when compared to training from scratch (e.g. Wang et al. reported an improvement from 83.7 to 91.2% regarding the FLL type classification accuracy of their model when using a pre-trained network) [176,177]. Several studies also demonstrated that multiphase images increase the performances of the DLR networks, when compared to single phase input only [178]. Instead of training the DLR networks only with images, it is possible to combine them with clinical data which can be difficult to collect, and challenging to integrate in a deep architecture,

but are proven to improve the accuracy of the networks in some cases [181].

Reported results showed moderate to good accuracy for the wanted tasks. For example, the reported mean accuracy is above 0.90 for the studies targeting a classification between Focal Nodular Hyperplasias, Cysts, HCC and Hemangioma (0.91 in both [177,179]), and it slightly drops to 0.84 when more complex categories are integrated (iCC, combined HCC and difference between HCC and early HCCs) [180]. Another study performed the classification between HCCs and non HCCs, by additionally incorporating the differentiation stages for the HCC group, but still had comparable performances than experienced radiologists in the diagnostic performances [176].

The study targeting the estimation of the fibrosis stage reported results less accurate than those obtained using elastography data (MRE: *Magnetic resonance elastography* or TE: *Transient elastography*), but they were the first to perform this analysis on CT images, and their results could be improved with the inclusion of volumetric information, and other sources of data [180].

Finally, the studies predicting a response to a treatment reported a high accuracy with an AUC of 0.82 when predicting the recurrence after TACE [181], and accuracy above 0.83 in the two external validation sets when estimating the response of TACE in HCC [182].

Those results still can be improved, especially by increasing the size of the cohort, or by replacing the manual placement of the bounding boxes with an automatic segmentation method in order to reduce the dependency to single or multiple experts [180,182].

As a conclusion, the different reviewed studies tend to agree on the fact that a multiphase analysis is necessary to precisely describe and encode the pathological features of the disease. For the rest of the pipeline, no real consensus exists but several strategies are implemented especially to compensate for the small size of the databases. Worth also noting that the reviewed studies correspond to the first DLR liver-related applications, therefore, they tend to tackle the less complex problems such as the FLLs classification. With future improvements regarding DL applied to the medical imaging field, and with more publicly available data, the next DLR liver-related studies will be ready to tackle more complex challenges.

In our research work, we conducted a study aiming for the histological grade prediction, consequently, we designed a pipeline following the aforementioned conclusions. The key element in our work was an automatic cascaded liver and tumor segmentation performed on registered multiphase CT images. It is worth noting that we are the first to perform prediction of the histological grade from multiphase CT images using a fully automatic pipeline.

Chapter 4

Segmentation of Liver Tissue in computed tomography

4.1 Introduction

The automatic segmentation of the tumors allows the computation of the entire tumor volume, which corresponds to a better predictor when compared to the diameter only, according to the *Response Evaluation Criteria in Solid Tumor* (RECIST) standard [112,184]. A precise volumetric segmentation of both the liver and its tumors is a prerequisite in case of treatment planning, (*TARE*, *TACE*, *percutaneous thermal ablation*, *radiotherapy surgical resection*,...) especially to adjust the treatment procedures to the volume of the lesions [185–188]. Associating liver and tumors segmentation allows the computation of the *tumor burden*¹, which has an importance in case of surgical resection, or when estimating the efficiency of a given treatment [157–162,189,190].

Relying on human expert delineations currently corresponds to the most common way to assess segmentation methods [191], although it does not correspond to a true gold standard [192] and often requires several experts to segment the different structures for reducing the operator-dependent bias [193,194].

In order to compare digitally obtained segmentations with the expert outputs, different metrics have been implemented over time (see appendix D). However, as explained above, manually segmenting the liver and its different structures is a tedious, time-consuming (up to 90 minutes per patient [195]) and operator-dependent task, as seen in the figure 4.1. Furthermore, the constitution of an open database needs to respect several ethical aspects, and the patient data privacy. This is the main reason why only a small number of samples have been publicly released over time.

As explained in the chapter 2, the liver is an organ with a key function in the human body. Being a

¹ proportion of the liver affected by cancer

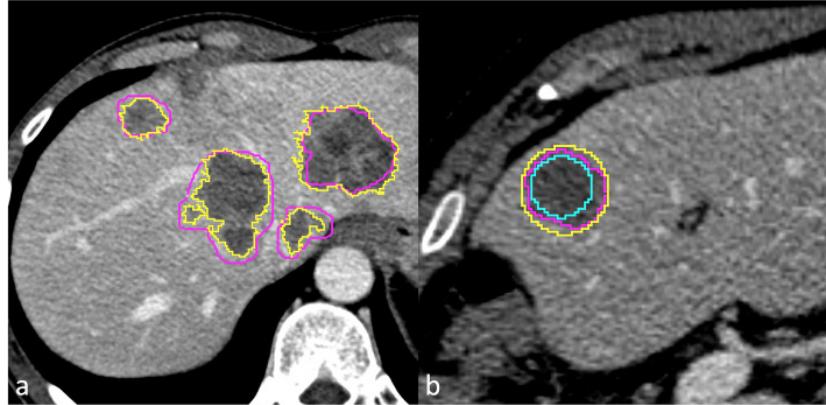


Figure 4.1: Interobserver variability: For each metastasis, the whole lesion volume and the largest axial cross-section were segmented by two readers. a Purple line (reader 1) versus yellow line (reader 2) contouring. The largest two-dimensional (2D) region of interest (ROI) of the main lesion was confronted with two circular ROIs, one inside the metastasis and one outside it. b Purple line (reader 1) 2D versus yellow line (smallest circular ROI inclusive of the whole lesion) versus azure line (largest circular ROI completely inside the lesion). ©Rizzetto et al. [196]

soft organ, it presents a large variation in shape among the population (inter-patient variation), and can also be subject to deformations due to respiratory and heart motions in case of multiphasic acquisition after injection of contrast agents (intra-patient variation).

On computed tomography images, the intensities of the liver itself and those of nearby organs (heart, spleen, stomach) are very similar. The images are often acquired after the injection of contrast agents to offer critical information (e.g. *wash-in/wash-out*). Nonetheless, the acquisition protocol may differ among the institutions (type of contrast agent, moments of injection, injection rate, ...) and the results often differ with the patient characteristics (e.g. blood flow). Moreover, the liver can easily be affected by diseases such as cirrhosis, and is often the site of either primary or secondary cancers with varying contrast levels (hyper-/hypo-intense tumors, peritumoral enhancement) and shapes (non-smooth margins, hypoattenuating halos, ...). All these reasons cause the automatic or semi-automatic segmentation of the liver and its structures to be challenging.

In this chapter, we are first covering the different publicly available datasets as long as the different challenges launched to increase interest in liver semantic segmentation. We are then investigating the first methods developed to perform the semantic segmentation of the liver and/or liver tumors, before presenting most recent advances brought by deep learning (DL).

4.2 Publicly available datasets & open challenges

Up to today, and compared to other organs (brain, breast, lung) [197], only a small number of computed tomography images datasets have been publicly available for liver tissue segmentation purposes. A table presenting details of the publicly available datasets can be found in the table 4.1.

Name	#Cases	Liver GT	Tumor GT	Tumor Type	Other GT	#Experts	Volume size (pixels)	Voxel size (mm)	Challenge	Hidden Cases	Phases
3Dircad-01	20	true	true	Undefined	Blood vessels	1	axial: 512x512 z : 74 - 260	axial:0.56 - 0.87 z: 1 - 4	-	-	Undefined
3Dircad-02	2	true	true	FNH	-	1	axial: 512x512 z: 167 - 219	axial: 0.96 z: 1.8 - 2.4	-	1 AR 1 PV	
Sliver07	30	true	true	Undefined	-	1	-	axial: 0.55 - 0.8 z: 1 - 3	MICCAI07	10	PV
MIDAS	4	false	true	Undefined	-	5	-	-	-	-	Undefined
LITS	131	true	true	Undefined	No	3	axial: 512x512 z: 42 - 1026	axial:0.56 - 1 z: 0.45 - 6	MICCAI17 ISBI 2017	70	Undefined
TCIA	97 ²	false	false	HCC	No	-	axial:512x512 z: 26 - 192	axial: 0.58 - 0.98 z: 0.8 - 5	-	-	mixed
CHAOS	40	true	false	Healthy Livers	No	3	axial: 512x512 z: 77 - 105	axial: 0.7 - 0.8 z: 3 - 3.2	ISBI 2019	-	PV
ImageClef Liver	50	true	true	-	No (Bounding Box)	-	x: 190- 308 y: 213-238 z: 41 - 588	x: 190- 308 y: 213-238 z: 41 - 588	ImageClef 2015	10	Undefined

Table 4.1: Publicly available datasets

Some of the available datasets have been launched by medical-related institutions. IRCCAD (*Research Institute Against Digestive Cancer*) for example, provided both 3DIrcadb sets [198], with the first containing 20 abdominal CT scans, where 15 of the patients suffered from liver tumor, and the second containing only 2 CT scans with patients suffering from FNH (*Focal Nodular Hyperplasia*). The *National library of Medicine* (NLM) offered the MIDAS dataset containing annotated pathological CT scans from 4 patients [199] whereas the cancer imaging archive (TCIA) proposed a set of 97 unlabeled volumes from patients suffering from hepatocellular carcinoma (HCC) [200].

Thanks to the different segmentation challenges launched throughout the time, some other sets have been provided and offered real benchmarks. The *Segmentation of the Liver* (Sliver'07) was organized in conjunction with MICCAI 2007 and offered 40 CT volumes [201], where 30 volumes were available during the challenge, and 10 hidden cases were used to evaluate the proposed methods. In 2015 the Image Clef challenge offered 50 CT scans with expert annotations [202], whereas 201 volumes have been provided by the organizers of both ISBI and MICCAI in 2017 for the *Liver Tumor Segmentation* (LITS) benchmark [191]. In 2019 the organizer of ISBI further granted access to 40 cases for the *Combined (CT-MR) Healthy Abdominal Organ Segmentation* (CHAOS) [203].

In these different datasets, the volumes were annotated by one or multiple experts, with predominantly pixel-wise delineations of the areas of interest (the ImageClef dataset is the only one with bounding box annotations for the lesions). Only one annotation volume per case is given which is designed to translate the consensus among the experts, however, the inter-expert variations can not be correctly caught by this format. Moreover, most of the available databases present heterogeneous data in terms of geometry (number of slices per patient, volume size, resolution), pathology (the number of lesions per volume in the LITS dataset goes from 1 to more than 70) and regarding the acquisition protocols (e.g. the acquisition phase is undefined for the LITS data where arterial and portal venous volumes are mixed up).

4.3 Hand-crafted feature based semantic segmentation methods

A total of 61 studies performing either the segmentation of the liver and/or its tumors were reviewed to introduce the different hand-crafted feature based segmentation methods. Even though the majority of the studies combined several concepts to perform the segmentation, they were classified based on the algorithm using the most domain-specific knowledge [204]. The reviewed studies included in this section were initially dedicated to the liver segmentation³ since this task received initially more interest from the researchers than the liver tumors segmentation task. However, the majority of the detailed methods

³ 48 over the 61 reviewed studies were only dedicated to the segmentation of the liver

can also address the semantic segmentation of liver tumors or the segmentation of other abdominal organs⁴.

The first methods that have been used for the semantic segmentation of liver tissue relied mainly on computer vision and machine learning methods. In most of the cases, prior anatomical knowledge about the intensity, the shape or the position of the liver is incorporated in the process. The different methods are supported by several pre- and post-processing steps, and may require one or multiple interactions with experts.

Among the most utilized methods, we will first present those based on the intensity of the voxels, before mentioning the ones requiring manual interactions such as region growing or graph-theory based methods. We will then introduce the geometric deformable models that can be exploited to add local constraints, before introducing methods requiring prior shape knowledge such as probabilistic atlases (PA) or statistical shape models (SSM). Finally, we will mention the different studies that combined the aforementioned methods with the addition of machine learning based steps.

4.3.1 Image Intensity based methods

Some studies decided to first rely on the gray level distribution of the liver in the training images to get an initial segmentation during the inference, before applying a refinement step consisting on active contours, region growing, or morphological techniques with anatomical constraints.

Lim et al. [205–207], combined prior-knowledge such as the location of the liver and its intensity distribution with active contours to segment the liver. The rough estimation is determined via thresholding and morphological operations in a multiscale fashion. K-means clustering is applied to remove unwanted tissues from the candidate area. Active contour, as described later, is then implemented to refine the liver boundary by first detecting the boundary between liver and adjacent organs. The segmentation pipeline is depicted in the figure 4.2.

Lee et al. [208], proposed a multimodal contextual neural network using pixel intensity and neighborhood dynamic to get a rough liver segmentation. The refinement step is performed using fuzzy rules based on prior knowledge such as location or textural properties. Liu et al. [209], developed a method to segment the liver by first detecting the potential contour implementing a Canny Edge detector algorithm, modified by the template liver position obtained via thresholding. A Gradient Vector Flow (GVF) field computation is used by the snake algorithm to obtain an accurate liver boundary detection. The final segmentation is obtained semi-automatically by asking the user to select the slices where the segmentation gave accurate results, and propagate this segmentation iteratively to adjacent slices. An example is depicted in the figure 4.3.

⁴ 6 studies among the 61 retained ones performed both the segmentation of the liver and its tumors, whereas 7 others segmented both the liver and other abdominal organs

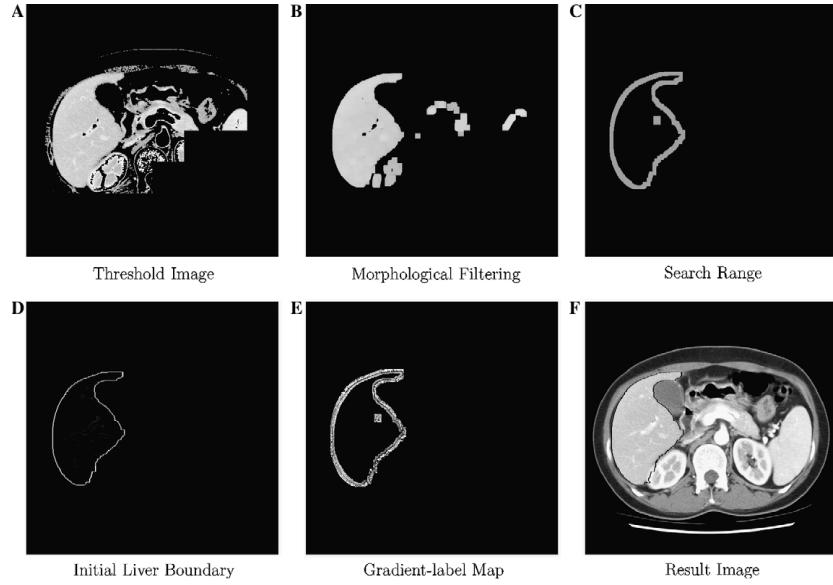


Figure 4.2: Intensity based liver segmentation workflow used by ©Lim et al. [207]

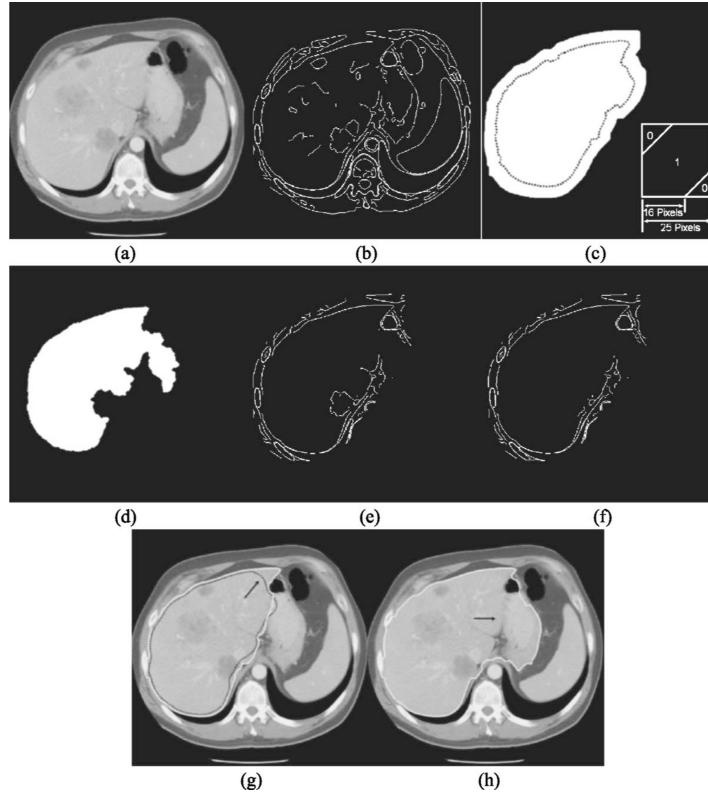


Figure 4.3: Liver segmentation as described by ©Liu et al. [209] : (a) Original image to be segmented, (b) Initial edge map (c) Dilated version of the neighboring slice mask (SSE in right lower corner) (d) The liver template (e) Edge map after modification by the mask and the liver template (f) Final edge map after modification by the concavity removal algorithm (g) The original image overlapped with the final contour (white solid) (h) Leakage if GVF is not applied.

Seo and Park [210] combined thresholding techniques and morphological operations to obtain a final liver segmentation. Specific filters are firstly used to smooth the distribution and reduce the noise. Adaptive multi-modal threshold is then applied to find the intensity range of the contrast enhanced

liver. The left partial histogram threshold (LPTH) algorithm is then implemented to detect the liver and remove neighboring organs, before morphological operations are performed to refine the candidate ROI. Kim et al. [211], obtained a rough segmentation of the liver by comparing the gray level dynamic of the test patient with gray-level distributions obtained from liver ROIs in the training dataset. The ROI is refined using watershed transform followed by anatomical constraints such as the presence of smooth boundaries. Campadelli et al. [212], proposed a new fully automatic liver segmentation technique. They first constrained the research by considering only the area below the heart. Different organs are afterwards classified, in particular the liver which corresponds to the biggest connected component found between the different organs. The liver region is further refined using a region growing method to ensure 3D consistency. Foruzan et al. [213], computed the liver intensity range using expected-maximisation (EM) algorithm, and applied thresholding operation in a bounding box constrained by the position of the ribs and the heart as illustrated in figure 4.4. The final candidate region of interest is then refined by anatomical constraints such as liver anatomy and assumed intensity range.

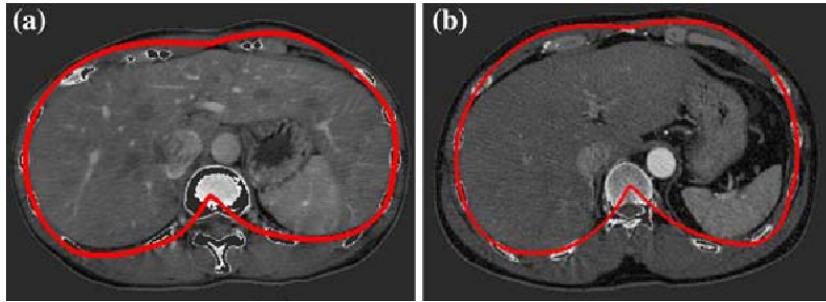


Figure 4.4: Estimation of the liver ROI based on the anatomical rib position as described by ©Foruzan et al. [213]

4.3.2 Region Growing

This set of methods is based on the placement of an initial seed point/region and evolves based on the image properties. As an example, Beck and Aurich [214], proposed to manually place seed points iteratively until the entire liver is segmented by 3-D filling using non-linear interpolation rules. Qi et al. [215], tackled the lesion segmentation problem by using a region growing strategy with manually placed seeds. The lesion is modeled as a mixture of Gaussians. A new voxel is added to the region by comparing its intensity and neighborhood to each of the lesions Gaussian through the Bhattacharyya distance. Since manual placement of seeds still suffers from problems such as inter and intra-operator variability, some studies suggested to automatically define seed points to start the algorithm with.

Rusko et al. [216, 217] proposed to define the original seed region by considering a set of rules such as the intensity range of the liver or its anatomical position relative to the heart. Susomboon et al. [218]

decided to partition the abdominal region via intensity-based EM algorithm to find the distribution of soft tissues in the whole volume. The liver region used as seed was further isolated using quad-tree decomposition and textural features. Kumar et al. [219], proposed a region growing method to segment the liver where the initial seed point corresponds to the centroid of a rough liver region obtained based on several anatomical and intensity constraints. To ensure that the centroid does not lie on lesions or dark objects, its intensity must be in the liver likelihood range. A post-processing step was proposed to delineate potential lesions within the obtained liver area. The process is depicted in the figure 4.5.

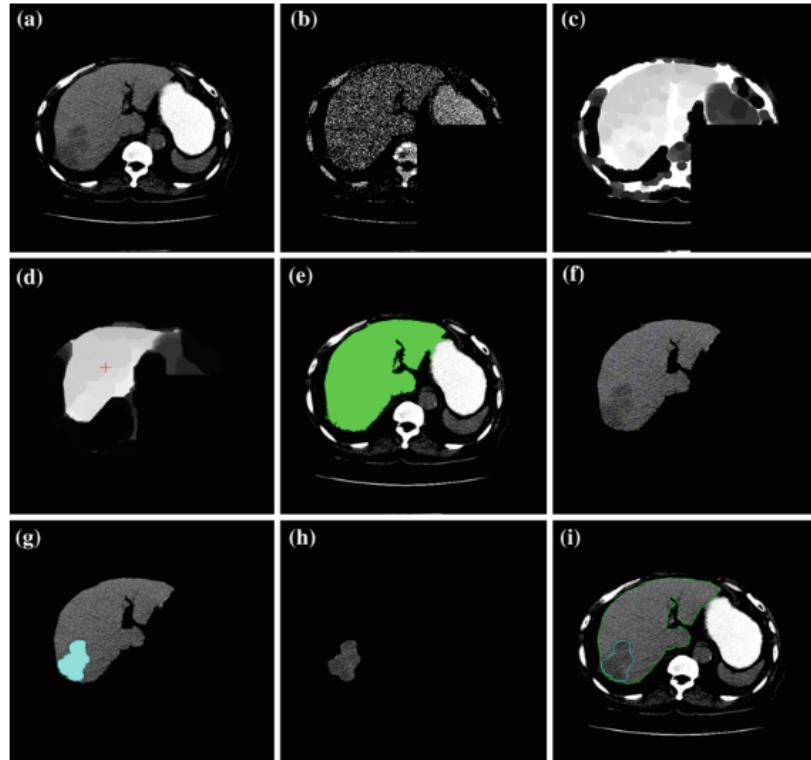


Figure 4.5: Liver segmentation pipeline after the placement of a seed at the estimated liver centroid position as detailed by ©Kumar et al. [219]: (a) Original image, (b) simplified image, (c) eroded image, (d) cross indicates the centroid (seed point) on the eroded image, (e) region grown liver, (f) segmented liver after postprocessing, (g) AFCM clustered lesion, (h) segmented lesion, (i) segmented liver and lesion outlined in the original image

Geometrical rules or local intensity properties often guide the growth of the region. In Rusko et al. Susomboon et al., Kumar et al. the intensity in a given n-voxels-sized neighborhood is analyzed to decide whether or not a point is added to the region [216, 218, 219]. Phole and Toennies decided to iteratively test the homogeneity of the region after adding candidate points [220]. In Beck and Aurich, the filling step is controlled by a nonlinear coupling criterion, which computes the weighted intensity difference between a seed point and its neighborhood. Weights are controlled by a non-normalized Gaussian function applied on both the distance to the seed point and on intensity difference [214]. However region growing methods have the disadvantage to poorly handle pathological livers, and often fail due to leakage in presence of weak boundaries.

4.3.3 Geometric deformable models

Both geometric deformable models (aka active contours) and level set based methods rely on boundary tracking to perform the segmentation.

The basic idea of *active contours* (introduced in 1988) is to come from an initial contour and deform it until it reaches the contour of the object. The deformation function reaches its minimum when the boundary of the object is found. The function used to deform the contour is an energy function that tries to control the smoothness of the curve and that is attracted by the object boundaries.

One of the main drawbacks of this method is that the final curve will always have the same topology as the initial one, meaning for example that multiple objects can not be detected. To overcome those limitations, new models, based on the theory of curve evolution and geometric flows, were designed. Caselles et al. [221] proposed a formulation where finding the contour will be equivalent to finding a geodesic curve of minimal weighted length, thus the formulation of *geodesic active contours*.

In the active contours formulation, a parametric characterization of the boundary is utilized, whereas in the level set paradigm, they are embedded as a time-dependent partial differential equation (PDE). Osher and Fedkiw [222] proposed a definition, where the boundary of an object can be regarded as a zero-level set of a time-dependent function which is evolving based on the speed at which the contour evolves along its normal direction. The negative values correspond to voxels belonging to the liver regions, and positive values to the outside. This paradigm was first implemented by Pan and Dawant [223], who used the level set technique for the automatic segmentation of liver in CT scans. They realized that the classical formulation can not accurately handle cases where weak boundaries are present in noisy or non-uniform images. Therefore, they decided to use an accumulative curvature-dependent speed function that would depend on the front past history. In order to be compliant with abdominal CT scans segmentation, a priori anatomical knowledge was added to the process by considering the distance from the liver to the skin and the ribs. Initial detection of the skin and the ribs is illustrated in the figure 4.6.

Wang et al. [224] combined probabilistic atlas (PA) and level set to perform the segmentation of the liver. Training volumes were pre-processed before the liver region was segmented, and stored with the denoised CT volumes. During the inference phase, the different training denoised CT volumes are used to compute the weighted PA for the current test patient. The liver ROI is then extracted using the weighted PA, and improved by estimating 5 different regions (heart, liver, right kidney, spleen and bone). Finally, a combined shape-intensity model is generated from the training samples, and used to refine the mask by level-set based segmentation.

Garamendi et al [225], proposed to segment the liver using a geometric level set method. After manually placing a circle in the liver region to segment, different candidates ROI were iteratively obtained by

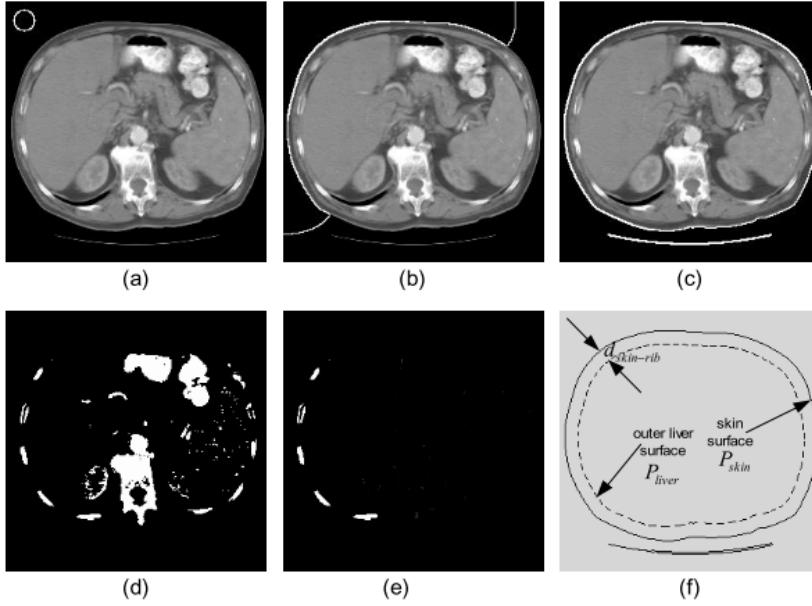


Figure 4.6: Skin surface localization (a-c) followed by the extraction of the ribs (d-f) as described by ©Pan et al. [223]

computed gray-level properties inside and outside the given region, and shifting the boundary by looking at local information around the border pixels. This procedure is repeated until the border stabilizes. Chi et al. [226] proposed a segmentation technique based on active contour. They placed the initial boundary by analyzing anatomical properties of the patient such as the position of the rib cage, the heart and the lungs. K-means clustering is added to reduce the variation present in the candidate surface by detecting blood vessels, remaining heart tissues, the liver, the kidney and the tumors. Finally, active contour is used based on GVF (*gradient vector flow*) combined with a distance transform formulation which detects the liver boundary based on the mask obtained for the previous slice . Furukawa et al. [227] combined rough segmentation based on MAP estimation and level set method to achieve liver segmentation. To perform the rough segmentation, the lungs are initially segmented and used to normalize the different images. The liver as long as 3 other components (right-kidney, heart and other tissues/organs or muscles) are segmented using MAP with a PA as reference. Active contours are used to refine the segmentation by re-using the definition provided by Catelles et al. and adding a term based on the distance to the human body contours.

Lee et al. [228] computed a speed image through a fast marching algorithm, by looking at gradient magnitude in the original image. A rough segmentation is retrieved by performing a 2.5D propagation after placing seed points on the top and the bottom of the liver. A level set method is implemented to refine the obtained area. The evolution of the level-set based contour is controlled by the initial speed image values and by the curvature. Vessels near the liver boundary can be added to the final map using a rolling ball algorithm. Wimmer et al. [229] semi-automatically segmented the liver by extending 2D

cross-sectional annotations, in 3D using RBF and applied a level set technique to refine the obtained volume. The initial area is obtained by manually performing 2 cross-sectional segmentations per axis, and extending them in 3D via RBF after sampling some points along the given contours, as depicted in the figure 4.7. The level set reconstruction surface was then used to refine the obtained shape. The function used takes into account the points initially present after manual segmentation. This method however suffers from errors when an initial surface includes high-contrast boundaries not belonging to the liver, because they will tend to attract the surface during the deformation process.

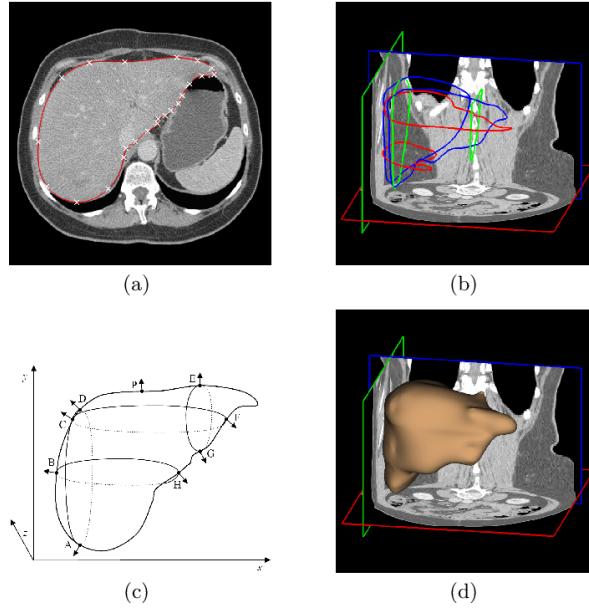


Figure 4.7: Semi automatic segmentation pipeline as proposed by ©Wimmer et al. [229]: (a) Manual delineation of a liver cross-section; spline control points are depicted in white (b) Six cross-sectional contours are defined, two for each orthogonal view orientation. (c) Visualization of calculated 3-D normals (d) Rendering of the interpolated surface.

Later, Wimmer et al. [230] used an implicit active shape model to perform the segmentation of the liver. To build the active shape model, one segmentation is used as reference, and the rest of the training segmentations are aligned to it. To convert the obtained masks into an implicit representation, a signed distance map is computed for each mask where grid points are assigned positive or negative Euclidean distance to the boundary. The shape model is expressed based on the principal component analysis of the signed distance maps. A level set function is formulated based on the given shape model. To build the boundary, all grid points in a narrow-band around the zero-level set are considered, and their boundary profile is computed. A probability is assigned to each point by considering the learned appearance model. Candidates points are projected back to the zero level set, and the points with the highest probability are kept. The obtained probability map is used in GVF (*Gradient Vector Flow*) to get the best contour.

Massoptier and Casciaro used a patch-based intensity approach to distinguish hepatic tissues from other

abdominal organs, and refine the obtained boundary by implementing active contour with GVF. The liver segmentation was followed by a liver lesions segmentation via k-means clustering [231]. Wimmer et al. [232] built later a shape model based on nonparametric density estimates. The appearance model is obtained by computing profiles for points in a narrow band along the zero-level set boundary. To evaluate the boundary probability for the points present in the narrow-band, they used a kNN classifier. A region model built upon a cascade of classifiers is implemented to complete the boundary model, and prevent it from stopping at local extremum. Finally, a shape model based on Parzen density estimation constrains the level set evolution. Song et al. [233] pre-segmented the liver using an edge detector, and refined the segmentation with a curvature-based level set algorithm . Three B-splines surface models are built knowing the position of the lung, and assuming that the bottom of the liver corresponds to the bottom of the left lobe. Those three surface models are implemented in a graph-based optimal surface fitting scheme to remove possible false positives and the remaining part is used as initialization for the level set evolution.

Suzuki et al. [234] preprocessed the image via anisotropic diffusion filtering to reduce the noise on the CT images, and combined it with a scale-specific gradient magnitude filter to enhance the liver boundaries. The obtained image is handled by experts to put seed points for the fast marching level-set algorithm. Next to the boundary estimation, the geodesic active contour level set segmentation refined the initial contour. Their technique uses the gradient-enhanced image as a speed function, so that the front expansion speed slows down in regions having high gradient and accelerates where the gradient is low. The different steps of their method is depicted in the figure 4.8.

To segment the liver, Platiero et al. [235] first aligned the different training shapes, and captured their variability via principal components analysis when building the level set function. The narrow-band technique is used to select boundary candidates whose profiles are computed. Their method, incorporating shape based priors as long as edge and region-based knowledge is then implemented. Jimenez et al. [236] proposed an optimized level set where the parameters are defined at each stage by means of multi-curvature, and where the liver segmentation is iteratively corrected in a pyramidal process. A fine details strategy tries to prevent problems that can come from the use of a narrow-band technique when searching for the boundary, by removing outlier pixels. An additional step is semi-automatically used to impose local curvature constraints.

Algorithms based on active contours were historically considered as one of the most utilized segmentation methods [204]. Later after the introduction of active-contours and level set, a new algorithm called Graph Cut segmentation was introduced by Boykov et al. to propose an alternative to boundary-based approaches [237].

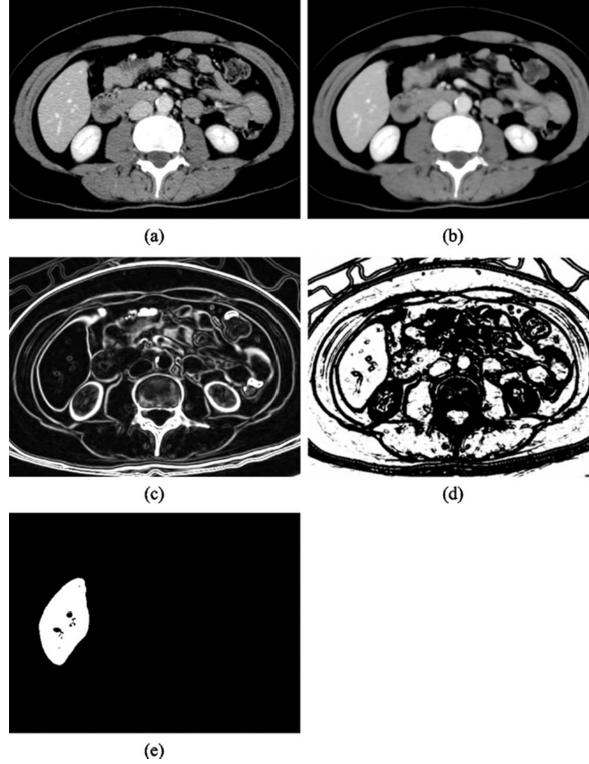


Figure 4.8: The different steps of the segmentation process developed by Suzuki et al. [234]: (a) Original CT image. (b) Anisotropic diffusion noise reduction. (c) Scale-specific gradient magnitude calculation. (d) Nonlinear grayscale conversion. (e) Geodesic active contour segmentation.

4.3.4 Graph-theory based methods

The random walker algorithm, which was introduced by Grady et al. [238] and the Graph Cut (GC) method correspond to semi-automatic⁵ methods where the user provides seeds for both the background and the region to segment. They both interpret pixels of the image as nodes on a graph where edges represent adjacency between pixels. The weights on the edges correspond to the similarity between adjacent pixels⁶.

The GC will then try to find the minimum cost function between all possible cuts on the graph to separate the object from its background.

GC often suffers from the “small cut” problem where only the seeds are separated from the rest of the image, whereas the random walker method does not suffer from this problem since it is not aiming for the smallest boundary, as represented in the figure 4.9.

The random walker algorithm has been implemented by Maier et al. [239] to proceed the segmentation of the liver. Seed points were automatically detected in several regions in a slice-wise fashion. Based on the skin surface and some anatomical assumptions, the rib cage and liver seed points are generated. Background seeds are obtained by detecting the air, the fat and the bones. Dong et al. [240] proposed a

⁵ can be fully automatic in some cases ⁶ The random walker will compute, for each unlabeled pixel, the highest probability that a random walker reaches the first labeled pixel whereas the GC will try to find the optimal cut for the entire image considering all the labeled seed, thus being more likely to fail in case of weak boundaries [238]

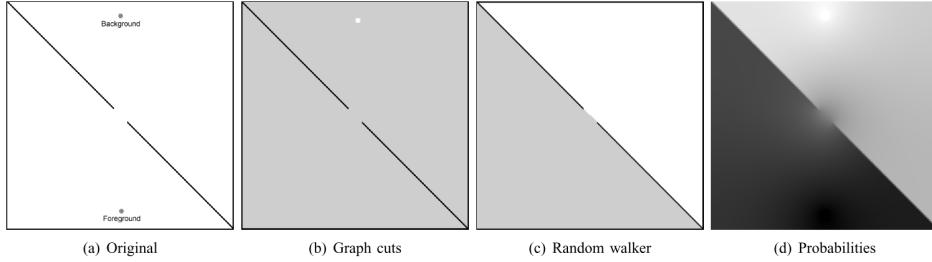


Figure 4.9: Comparison of random walker algorithm to graph cuts for a weak boundary with small seeds. Note that a 4-connected graph was used in these experiments. (a) Original (synthetic) image created with a diagonal black line with a section completely erased. (b) Graph cuts solution Since surface area of seeds is smaller than the weak boundary, the smallest cut minimally surrounds the seeds. (c) Random walker solution. (d) Probabilities associated with the random walker algorithm offer a notion of segmentation confidence at each pixel. ©Grady et al. [238]

random walker implementation where adjacent slices are considered as prior knowledge when segmenting the liver and the spleen in the current slice. Two slices are chosen so that they present the largest cross sectional area of both the liver and the spleen. Seeds are manually placed on these slices and updated on adjacent slices via GC (*Gaussian Mixture Model*) and intensity constraints through *Narrow Band Thresholding* (NBT), before the modified random walker segment the organs in a slice-wise fashion. Details of the execution can be found in the figure 4.10.

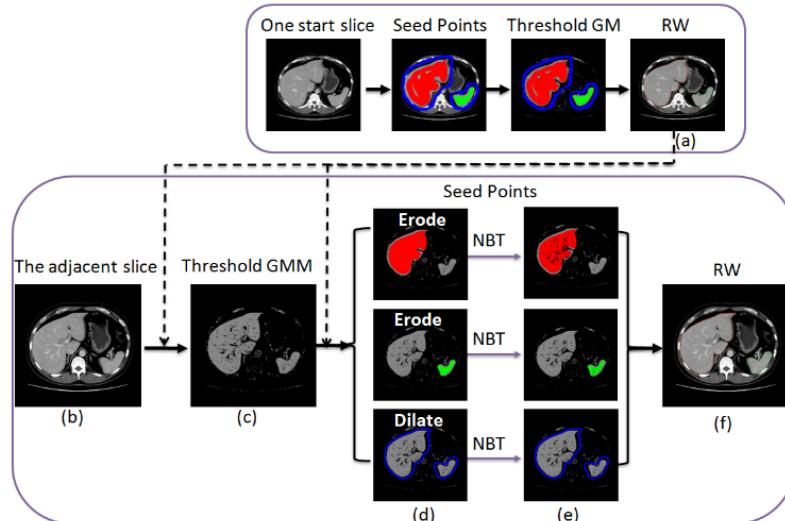


Figure 4.10: Segmentation Pipeline detailed by ©Dong et al. [240]: (a) The segmented liver (red) and spleen (green) of the previous slice; (b) The current slice; (c) Candidate pixel by applying a threshold to the GC; (d) The rough objects (red and green) and background (blue) seed points; (e) The fine seed points using the NBT method; (f) The segmentation results

Beichel et al. [241] used a graph-cut method to propose an initial liver region which can interactively be segmented by the users. The cost of the graph cut segmentation is specified by taking region and boundary properties into account. The region term evaluates the gray-level similarity in the neighborhood of a given point and compares it to the intensity of the seed points. The boundary term is computed from the local gradient information. Massoptier and Casciaro [242], introduced a GC

method initialized by an adaptive threshold to perform the liver segmentation. The thresholding was implemented by detecting the most likely liver intensity based on patch partition of the abdominal slices. The energy function took both a region and a boundary-based term into account. Shimizu et al. [243] combined a statistical atlas-based approach and a graph cuts algorithm to segment the liver. The graph supports an energy function including a local term, a boundary term and a shape-based term computed from the gradient of the initial region.

The key part of graph cut methods is to define a relevant energy function, which can sometimes be difficult.

Probabilistic atlases (PA) and Statistical Shape Models (SSM) both utilize prior information on the liver. The main goal of those techniques is to improve the segmentation by not relying only on the gray-level distribution since it was proven not to be sufficient, especially knowing the intensity similarity that the liver presents with close organs [244–247].

4.3.5 Probabilistic Atlases

To construct a probabilistic atlas, images from the training set are registered to a reference image, whereas manual delineations are warped onto the template image and averaged to get a probability of belonging to the liver for each voxel of the space. The generated atlas is then incorporated in the segmentation process. The studies differ one another on the way they construct the atlas, and on how it is incorporated in the segmentation process. Park et al. [245] used the MIAMI method (*Mutual Information for Automatic Multimodality Image Fusion*) which combines a TPS-based (*Thin Plate Splines*) registration method with MI (*Mutual Information*) as a similarity metric to segment multiple abdominal organs. The TPS is a registration method that uses control points as constraints to the interpolating function. They used 36 control points that were placed in different organs by experts, and they performed the registration separately for each organ to not be biased by the presence of more control points on the liver surface. The different points can be seen in the figure 4.11.

Zhou et al. [244] extracted the so-called “anatomical structure” of each patient, which is composed of the bone structure (determined using gray-level thresholding) and the diaphragm (based on the shape of the air within the lungs) to obtain the liver segmentation. A matrix transformation of the TPS is calculated between each one of the training cases and a predefined standard anatomical structure [248]. The atlas is then computed by combining the obtained positions using a voting strategy. When performing the segmentation, the same strategy is performed: the current patient’s anatomical structure is estimated, and deformed to the standard one. The gray-level distribution is also used to compute the final probability as well as the spatial location as depicted in the figure 4.12.

Slagmolen et al. [246] resampled all training images to a selected one using affine registration with

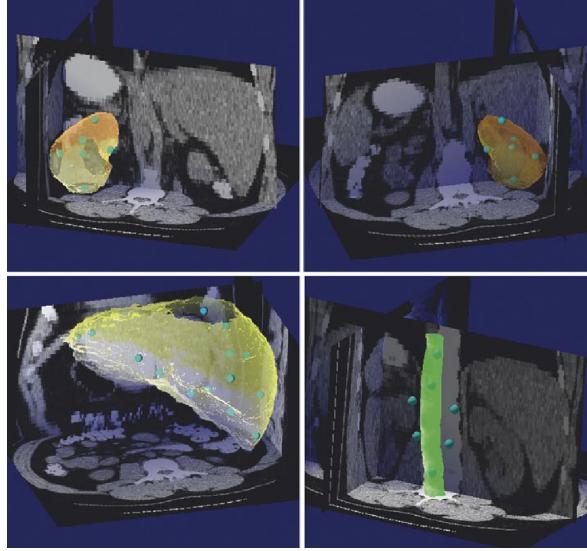


Figure 4.11: Example of control points distribution used in the MIAMI method as described by ©Park et al. [245]

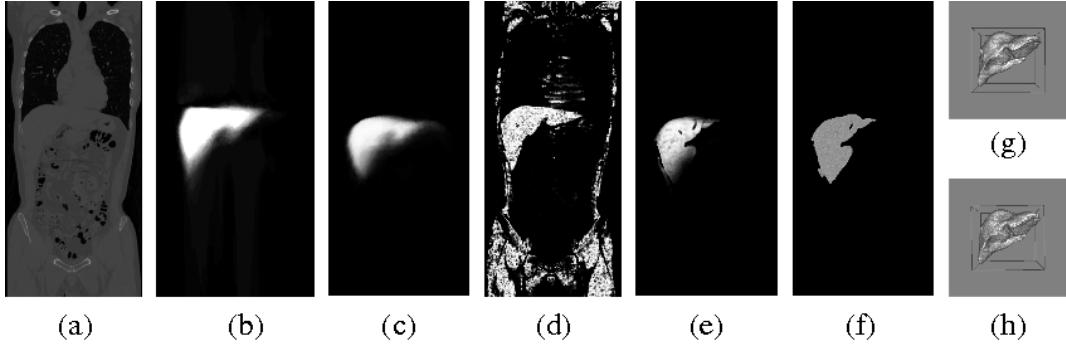


Figure 4.12: Example of liver segmentation with the different slices results as described by Zhou et al. [244]: (a) original CT image (1 coronal slice), (b) liver atlas, (c) $P(B)$: liver probability on spatial location, (d) $P(C|B)$: liver probability on density, (e) $P(A)$: liver probability, (f) liver segmentation result (1 coronal slice), (g) ground truth of liver (3-D view), (h) liver segmentation result (3-D view)

MI as the similarity metric, followed by a non-rigid registration guided by a surface distance metric when segmenting the liver. A mean morphology is defined by averaging, for each patient in the training set, its deformation field to every other image. Each patient is then deformed using its corresponding mean deformation field. And finally the obtained images are averaged on their intensities to obtain the atlas. Van Rikxoort et al. [247] decided to first find the vertical range that most likely contains the liver by looking at HU values covering the lungs and extract slices above them. They trained a kNN classifier with both gray level features (obtained from both the gradient computation and the Gaussian derivatives at various scales), positional features through the coordinates in the liver space, and an additional feature obtained by computing the proportion of the liver above, behind and to the left of a given voxel. No concrete atlas was created here, but the current test volume is registered to each training sample in a *non-rigid multi-atlas segmentation*, before applying the classification to refine the obtained region. Li et al. [249] constructed a liver and a rib-cage probabilistic atlas . Both atlases

were built iteratively to reduce the dependence to the reference patient, before a liver segmentation is obtained. Linguraru et al. [250] created the atlas by normalizing the liver coordinates of each patient relative to the xiphoid, and used those coordinates in the size-preserving affine registration process, by randomly selecting a patient as reference.

In those different studies, the segmentation method applied is often related to the way the atlas is created. Therefore Park et al. decided to complete the segmentation step by incorporating the atlas information in a bayesian framework which combines MAP and EM, with gray level distribution as basis [245]. The segmentation is then refined using a MRF (*Markov Random Field*) regularization step. In the other studies, non-rigid or elastic registration is performed from the atlas to the target case [246, 249, 250], after constraining the input, using either manually defined ROI around the liver [246], or analyzing the gaussian intensity distribution to remove irrelevant parts [249].

4.3.6 Statistical Shape Models

Statistical shape models on the other hand try to extract features of the model instead of building an average map.

They were first described by Cootes et al. [251], who introduced the concept of Active Shape Models (ASM) which are built from a set of segmented training images and consist of two parts, a geometrical model, and a local appearance model. The geometrical model describes the shape and is represented by a PDM (*Point Distribution Model*), a dense collection of landmark points on the surface of the object. Given the location of the landmarks for each training case, a principal component model can be built in order to approximate all valid shapes. The local appearance model describing the boundary, is used as an additional one to detect the modeled shape in the new image. It is based on the local gray value appearance around the boundary. A PCA or a kNN model can be used to determine the possible profiles at each landmark. The model is then guided by internal forces that should keep the shape of the deformable model similar to the one of the underlying SSM, and external forces that drive the deformable surface towards the best fit to the data.

To perform the segmentation of the liver, Montagnat and Delingette [252] followed this scheme by deforming the model with global constraints (shape constraints and transformation constraints) to reduce the degrees of freedom, and registered the obtained model on the volumetric data including an additional external force based on local gradient computation. Lamecker et al. [253] proposed a model that combined global constraints that minimized the geometric distortion for shape correspondence, and local constraints by considering the intensity profiles for the boundary search. The evolution of the boundary search for one example can be seen in the figure 4.13.

Heimann et al. [254] used an evolutionary algorithm to create a population of possible liver shape

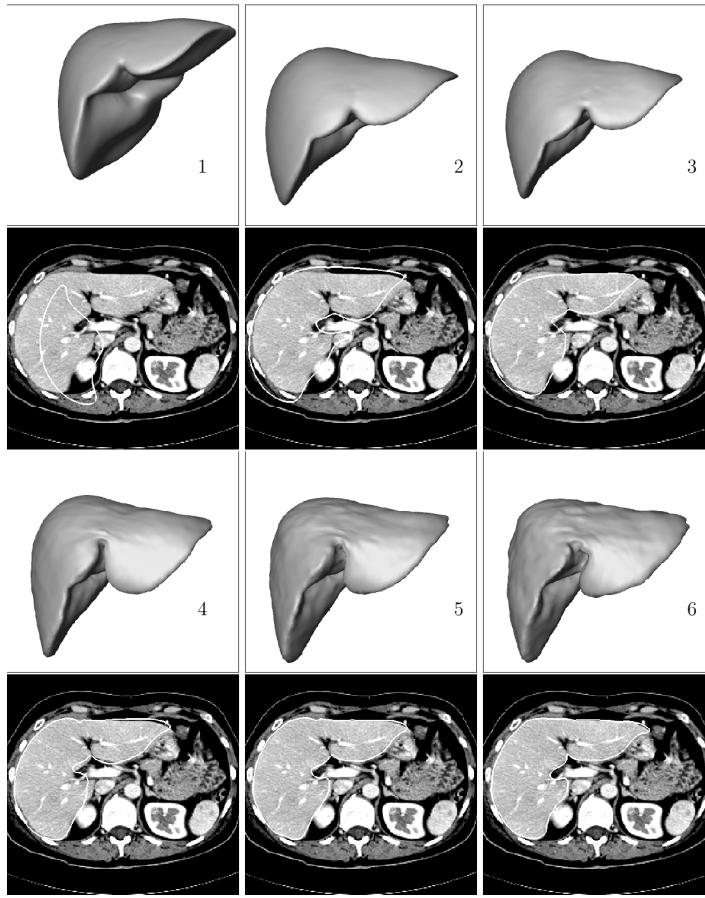


Figure 4.13: Evolution of boundary search in one example as described in ©Lamecker et al. [253]: At each of the six steps the model at its current position and shape is shown as a 3D surface. Below each surface one axial slice with the intersection line of the surface is displayed. Step 1: initial positioning of the mean model in the CT data such that the upper borders of the bounding boxes match. Step 2: result of the initial optimization of the position parameters. Step 3-5: results of the combined position and shape optimization with increasing number of shape modes (3, 9, 21). Step 6: final result (42 shape modes)

configurations and evaluated all the population members using a local appearance model based on a kNN-classifier utilizing the profiles. Saddi et al. [255] constrained the boundary of the liver to fit the global-shape learned from the training samples, and refined it locally using a template-matching algorithm. For both problems, they used the intensity distribution inside and outside the liver to find the best region to segment. Zhang et al. [256] performed the detection of the liver using a 3-D generalized Hough transform so that each vertex on the surface of the average shape model is stored in a table, by considering its coordinates relative to the liver centroid. The most probable liver centroid is computed for a given test image. A search is then performed to determine the points belonging to the boundary knowing the intensity distribution inside the liver and the properties of the gradient along the boundary. An optimal surface detection algorithm is finally applied to refine the obtained boundary by using a graph-search strategy. The different steps of the algorithm can be visualized in the figure 4.14.

Ling et al. [257] went further by proposing a hierarchical segmentation of the liver in a coarse-to-fine

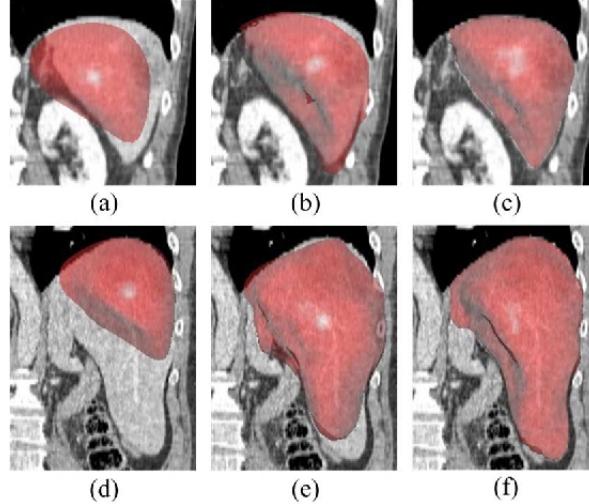


Figure 4.14: State of the surface mesh in a coronal slice from back after each segmentation step. The top row shows a relatively easy case while the bottom row shows a relatively difficult case. (a) and (d) After localization by 3-D GHT. (b) and (e) After model subspace initialization. (c) and (f) Final result after optimal surface detection. ©Zhang et al. [256]

fashion. For the training process, a hierarchical shape model is obtained by downsampling the resolution of an initially obtained dense mesh. Each layer of the pyramid then contained the mean shape as long as the different modes that capture shape variations for the given resolution. During the inference, the pose estimation is performed using a MSL-based approach (*Marginal Space Learning*) to reduce the dimensionality of the research [258]. The model is then upsampled to the finest resolution using local boundary refinement. At the finest resolution, a patch-based approach is implemented to refine the obtained boundary. Seghers et al. [259] used multiple local shape models instead of a global one to segment the liver. A grid search is used for the landmarks candidates detection, then, an intensity model selects the best landmarks in a given test image by analyzing the intensity profiles. This search is performed in a multi-resolution manner by changing the size of the grid. A shape model was built at each edge of the geometrical model, where an edge is connecting two landmarks. Erdt et al. [260] tried to determine the regions where high deformations are expected by looking at the curvature. They first built the model based on the data present in the training set, and then constrained it locally using the curvature and the gradient at each landmark point. Examples of how the incorporation of local curvature can improve the liver segmentation results are depicted in the figure 4.15.

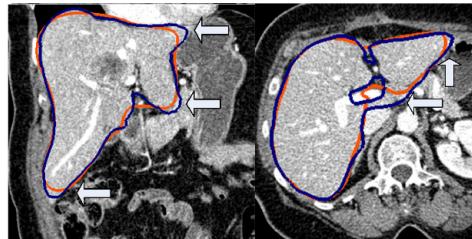


Figure 4.15: Example of how the incorporation of the curvature improved the quality of the segmentation as reported by ©Erdt et al. [260]

A common drawback of SSMs is their lack of flexibility. To combat this limitation, some studies combined the SSM with a free form deformation step to achieve better performances. For example, Kainmüller et al. [261] built a SSM based on a PCA analysis of the given shapes present in the training dataset, and to add flexibility to the model, a free-form deformation step based on the computation of 3-vector fields was conducted. During the deformation, they considered the general intensity distribution of the liver, the local displacements and they also tried to preserve the shape features of the initial model.

Okada et al. [262] combined both PA and SSM in their study to segment the liver. The PA was constructed from the different training cases by first registering the volumes to a standardized patient space using non-rigid registration of the abdominal cavity. The registration target was defined by experts describing the standard position and shape of the liver. The atlas was then computed by averaging the binary registered liver masks. The multilevel shape model (ML-SSM) is obtained by dividing the liver shape recursively into patches, and performing a PCA for each one of them. An example of construction of ML-SSM can be seen in the figure 4.16. For each test case, the atlas was combined with the gray-level distribution of the patient to obtain an initial liver region. A shape model was then built from this area and the ML-SSM is employed to segment the different patches, by adding an adhesive constraint to eliminate inconsistencies between adjacent patches. They later improved the efficiency of their method by constructing the PA hierarchically, where the different structures are constrained by the shape of the organ in the next highest hierarchy level. Three levels were handled, the abdominal cavity, the liver, and finally both the vena cava & the gallbladder. They extended the ML-SSM to an organ-based relationship and defined the MO-SSM (Multi-Organ SSM) [262]. The MO-SSM is constructed by combining ML-SSM for the different organs of interest. As a result they improved the accuracy in the regions concerned by the MO-SSM.

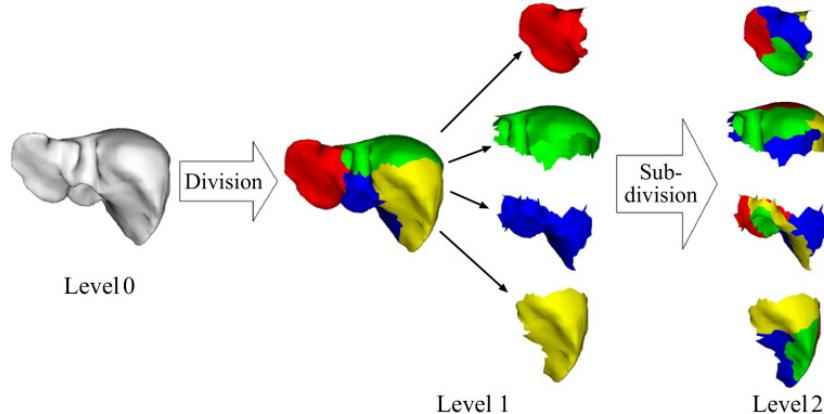


Figure 4.16: Hierarchical division of the liver shape for multi-level statistical shape model as described by ©Okada et al. [262]

Apart from the lack of flexibility, SSMs suffer from a low number of training samples, and also

hardly deal with low contrast between the liver and its surrounding or within the liver directly [253, 255].

Examples of those limitations are depicted in the figure 4.17.

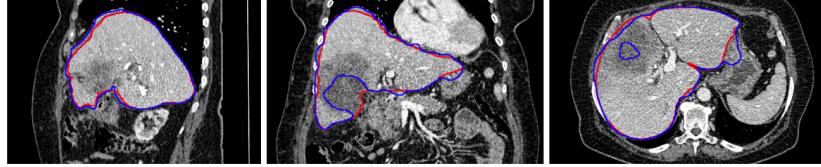


Figure 4.17: Example of errors obtained during the segmentation process as exposed by ©Saddi et al. [255]

4.3.7 Methods integrating machine learning based steps

Some studies decided to combine the aforementioned techniques with the introduction of machine learning methods such as shallow neural networks or SVM (*Support Vector Machine*) to segment the liver and potentially its tumor.

Tsai and Tanashi [263] used a neural network based on regional histogram to get a rough liver segmentation. Several post-processing steps were added, such as Laplacian filtering to obtain the boundary followed by a smoothing using B-spline functions. Gao et al. [264] first located the liver by extracting peaks on gray-level histograms, and applied domain-based knowledge to remove irrelevant tissues. The liver region is refined based on adjacent slices properties, before a boundary refinement is performed via a deformable contour technique. Schimdt et al. [265] incorporated a set of intensity distributions, neighboring relationship between organs and geometrical constraints to segment different parts of the abdomen. Their study was based on several anatomical assumptions, but fails when the liver presents a slightly non-standardized aspect, especially when large tumors are present. Freiman et al. [266] applied a multi-resolution, multi-class smoothed Bayesian classification followed by morphological adjustments and an active contour refinement for the segmentation of the liver. The classification was performed by iteratively looking at liver and potential tumor gray-level distributions. MAP rule is applied for the two classes. The liver region is adjusted using morphological operations, before an active contour refinement step based on gray-level intensity gives the final segmentation. Freiman et al. [267] further proposed a method for the semi-automatic segmentation of liver tumors, where a SVM classification is employed to separate healthy and tumorous tissues. A 3D energy function is then applied using affinity constraints to get the final VOI. Florin et al. [268] proposed a shape model to segment the liver, that will describe the shape variations only using a small number of key slices. This model, called Sparse Information Model, consists of key slices that are selected to be sufficient so when combined with an interpolation function, the 3D volume can be reconstructed. Those key slices are further used for the segmentation but the result is highly sensitive to initializations.

To segment the liver, Schenk et al. [269] combined the “livewire” algorithm, which is a semi-automatic 2D segmentation method, with a shape-base model to approximate the contour between the manually defined slices. The interpolation used to combine the user-defined segmentations is an object-based interpolation considering the distance transforms. Maklad et al. [270] proposed to segment the liver by analyzing the blood vessels structure. Abdominal blood vessels are first enhanced via bias field correction as depicted in the figure 4.18, and then extracted through thresholding and region growing techniques. Those vessels are further classified into hepatic and non-hepatic ones by applying distance transforms. The boundary between the liver and its surrounding organs is then constructed equidistantly from the hepatic vessels and the non-hepatic ones. A post-processing step is finally applied to refine the obtained area, by filling holes and classifying boundary tumors as belonging or not to the liver based on their relation to hepatic blood vessels.

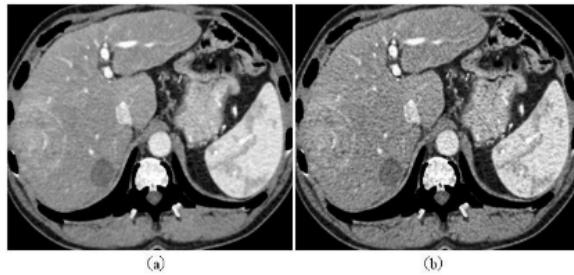


Figure 4.18: Results of the blood vessels enhancement: (a) before enhancement (b) after enhancement, ©Maklad et al. [270]

Chartrand et al. [271] started from a set of user-defined contours, and obtained an initial shape using variational interpolation. Boundary points were then extracted using a template matching method. Intensity profiles were computed in a narrow band around the previously established boundary, before a non-rigid registration scheme based on Laplacian mesh optimization deformed it to the real liver boundaries. Goryawala et al. [272] first extracted 5 different regions on the volumes using a k-means algorithm (the liver, surrounding tissues, peripheral muscles, rib/spinal cord and the air) where clusters are initialized by user-defined seed points on the central liver slice. An intensity based region growing algorithm is then applied using two more seed points on the top and bottom of the liver, and refined using a 3-axis growing strategy. The slice presenting the largest cross-sectional liver part is automatically identified before being proposed to the user to get a precise boundary which will then be copied in all slices containing liver voxels. This boundary is finally refined using a localized region growing method. One advantage of the presented method is that it seems to be weakly influenced by the user-defined initializations. Li et al. [273] computed the total variation and the L1 norm to initialize the liver shape, before applying a level set method guided by local and global energies, and refining the obtained region using gray-level co-occurrence matrices.

Mostafa et al. [274] proposed an artificial bee colony clustering technique to segment the liver, which is then refined using morphological operations . The final step consists of a region growing method to enhance the segmentation obtained previously. Shi et al. started with a blood vessel shape initialization, and deformed the liver shape using a region-specific deformable framework [275]. The blood-vessel shape allows the algorithm to get a more accurate initialization based on the patient specificities. Al-Shaikhli et al. [276] proposed a level set formulation guided by a combined region-based and voxel-wise cost function. For the global image term, textural features (GLCM) as long as intensity-based and volume properties were computed. Initialization is performed based on prior knowledge about liver geometry and intensity distribution. The local information is represented by the shape prior obtained from a hierarchical patch-based division of the liver. The newly constructed level-set formulation is then iteratively calculated to get the liver boundary. Xu et al. [277] performed a multi-organ segmentation via context learning followed by SIMPLE atlas selection (*Selective and Iterative Method for Performance Level Estimation*). Context-Learning was done via intensity distribution analysis with a GC. The different organs were modelised by several atlases using the SIMPLE method. During inference the different organs segmentations were fused using joint level fusion before a final abdominal segmentation can be obtained.

Wang et al. [278] learned the liver shape model from a set of training samples, by implementing a Sparse Shape Composition (SSC). When segmenting a new case, they first initialized the liver boundary that is then used to construct a rough polygonal mesh representation of the liver. The mesh is finally refined via homotopy-based optimization using the SSC as reference. Huang et al. [279] first localized the liver using a trained AdaBoost classifier where image features (intensity based and contextual information) were treated as input. The liver was then registered using the SSM incorporating all training shapes in combination with a kNN classification. A Free-form deformation is finally applied to refine the segmentation using the SSM to compute forces applied to the mesh. The entire pipeline is depicted in the figure 4.19.

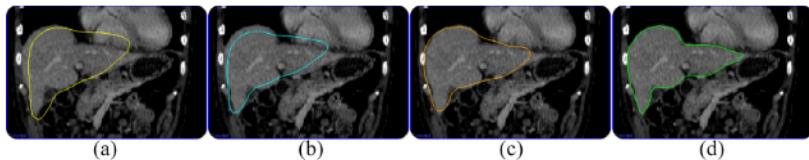


Figure 4.19: The four steps of the liver segmentation framework: (a) liver model location; (b) registration with liver distance map; (c) shape fitting under appearance guidance; (d) free-form deformation. ©Huang et al. [279]

Anter et al. [280] first detected the liver boundary using fuzzy c-means clustering based on the gray level local distribution, then refined the liver segmentation map via connected component analysis. Conze et al. [281] semi-interactively performed the segmentation of the parenchyma, and the separation

between the active and the necrotic part of the lesions on multiphase CT images. Images were registered between the different phases using PV (Portal Venous) phase images as reference, and supervoxels⁷ were computed at different scales using a SLIC (simple linear iterative clustering) algorithm on PV phase images. A random forest was trained on a set of retained supervoxels, and used to get a label prediction. They proved the value of using dynamic CECT and implementing hierarchical multi-scale supervoxels to better segment the liver tissues.

It has been proven that methods relying only on intensity are not robust enough to produce acceptable results especially when dealing with some primary tumors that present high textural heterogeneity. Prior knowledge and engineered features were added over time to increase the performances, but remained sensitive to the size of the dataset used in the training process, and the amount of interaction to incorporate in the workflow. Historically, semi-automatic methods were used for liver tissue segmentation, while they now tend to be replaced by automatic techniques, offering reproducibility and fewer interactions with experts. However, newer methods still suffered when dealing with pathological livers, often presenting irregular shape or intensity patterns. Recently, deep learning has changed the way of comprehending different computer vision related problems, especially in the medical imaging field, where ,for example, state-of-the-art semantic segmentation methods rely now on these techniques.

4.4 Deep Learning based semantic segmentation methods

4.4.1 State-of-the-art DL pipelines

The different DL studies introduced for the liver tissue semantic segmentation purpose share some common properties that will be exposed hereafter. We will fist detail the different types of DL architectures that have been developed to perform this task, before introducing the sequential segmentation concept shared by almost all of the reviewed studies. The details regarding the networks input type, the training strategies, and the inference schemes were then discussed. Details regarding the reviewed studies can be found in the appendix A.

4.4.1.1 FCN-based architectures

The use of fully convolutional networks (FCN) was quickly democratized, and popular imaging-related architectures such as VGG-16 or AlexNet were transformed to FCNs by replacing dense layers with convolution layers [282,283]. An illustration of the resulting architecture is given in the figure 4.20.

The FCN allows an architecture to predict as output an object having the same size as the input. Those architectures were perfectly suitable for the required segmentation tasks.

⁷ group of voxels sharing the same textural properties

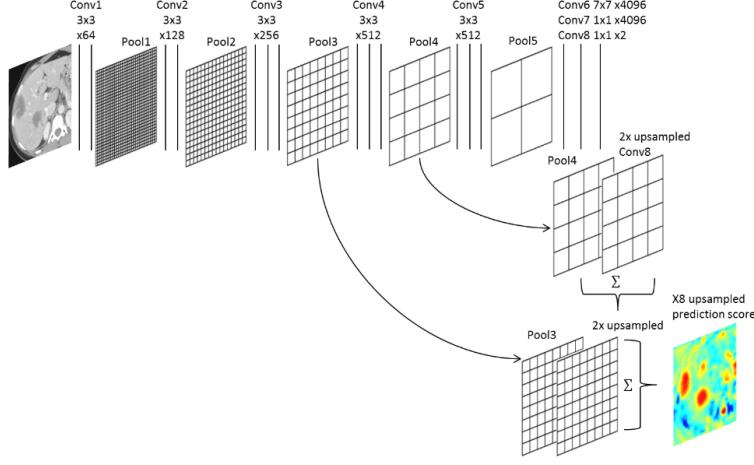


Figure 4.20: Example of a FCN architecture implemented to perform the liver tumor segmentation task, as described by ©Bellver et al. [283]

The most recent studies showed a predominance for U-Net like networks [284, 285]. The U-Net was introduced by Ronneberger et al. [286] initially for the segmentation of cells in microscopic images. An overview of its architecture is given in the figure 4.21. It consists of two parts, the first one where the information of the image is compressed, going from the extraction of low-level features, to the extraction of more semantic-related features, and the second part where the information is resampled back to the original image resolution. The main advantages of this architecture reside in the fact that it allows the network having both the input and the output sharing the same resolution, and also the so-called *skip-connections* that were introduced to reinject the features learned from the contraction part, in the decoding part.

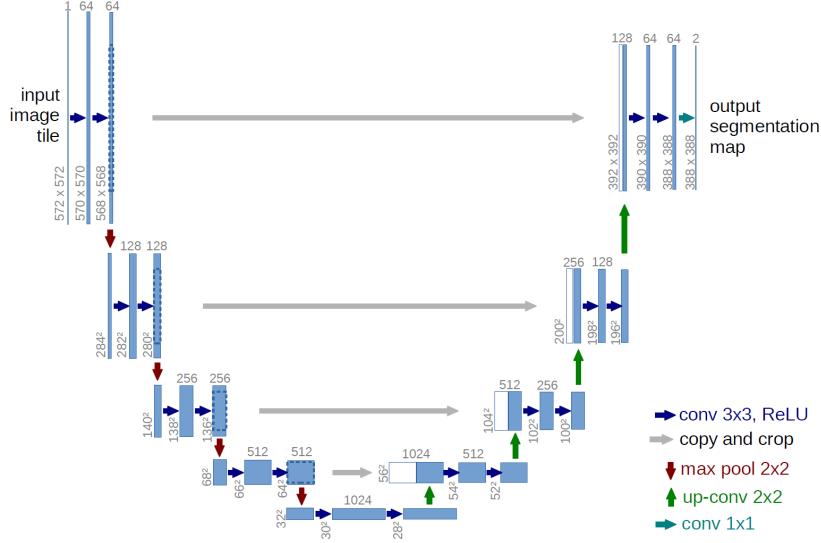


Figure 4.21: U-Net architecture as originally implemented and described by ©Ronneberger et al. [286]

Two other well-known architectures named ResNet and DenseNet were implemented in some other

studies, either alone or in combination with the U-Net [287–291].

The residual network (*ResNet*) developed by He et al. [292] in 2015 introduced the concept of residual connections, that help deep networks overcome the gradient vanishing problem. At each stage of the network, identity connections, as illustrated in the figure 4.22, allow the information to be passed from one block to the following, so that early features are not lost.

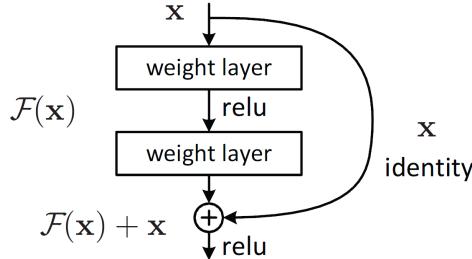


Figure 4.22: Residual connection as described by ©He et al. [292]

Densely connected networks (*DenseNet*) introduced by Huang et al. [293] in 2017 share the same motivation of allowing early layers features being kept all throughout the network. Contrary to the ResNet, the DenseNet requires less parameters and enables faster computation since it uses concatenation whereas ResNet added features map that needed to be kept [293]. Example of DenseNet architecture built for a classification task is depicted in the figure 4.23.

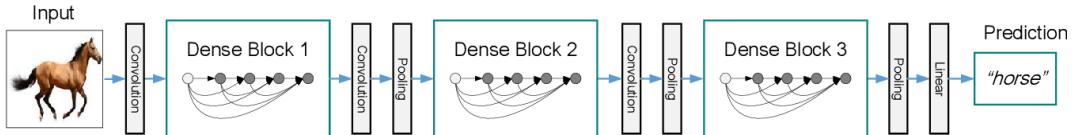


Figure 4.23: A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature-map sizes via convolution and pooling. As described by ©Huang et al. [293]

4.4.1.2 Cascaded architectures

To segment the liver and the lesions it might contain, the vast majority of the studies used a cascaded architecture where a coarse liver segmentation is performed, before either being refined [285] or directly used to segment the lesions [282, 287, 290, 291, 294]. An example of cascaded architectures is given in the figure below.

Interestingly, some studies decided to extract features before combining them to get a final segmentation map. For example, Bi et al. proceed first to an extraction of both liver and lesion features, before combining them with the original image to perform a pixel-wise categorical classification, as illustrated in the figure 4.25 [289].

Whereas Vorontsov et al. [284] separated the extraction of liver features and lesion features, as detailed in the figure 4.26. They added a classifier at each step which used features concatenated from

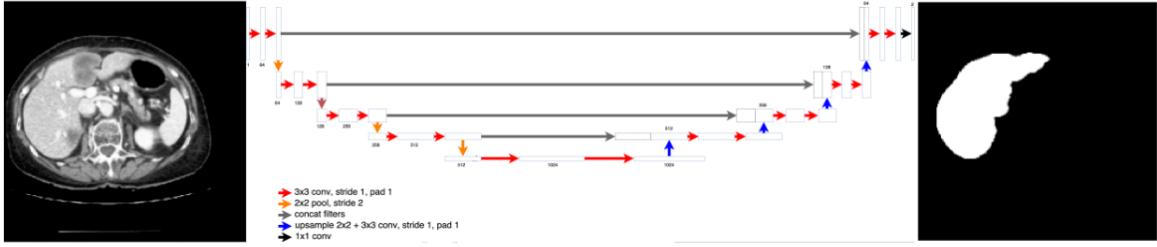


Figure 5: Step 1 of Cascaded FCN: The first U-Net learns to segment livers from a CT slice.

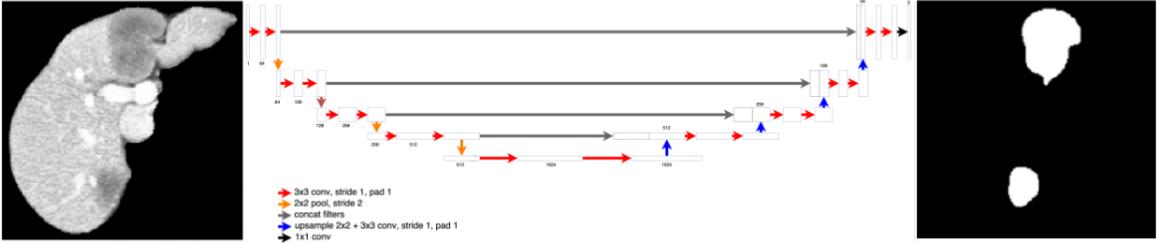


Figure 6: Step 2 of Cascaded FCN: The second U-Net learns to segment lesions from a liver segmentation mask segmented in step 1 of the cascade

Figure 4.24: Cascaded liver and lesion segmentation architecture as initially implemented by ©Christ et al. [294]

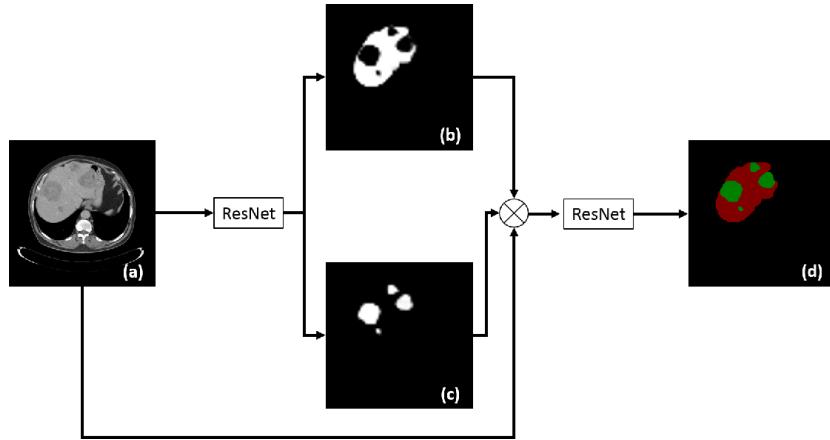


Figure 4.25: Custom cascaded architecture implemented by ©Bi et al. [289] where the liver and lesions features are first extracted before being combined to produce the final segmentation map

3 adjacent slices to get a segmentation of the middle slice.

4.4.1.3 Input type (2D, 2.5D, 3D, patches, multiphase)

The first deep-learning related study that performed the semantic segmentation of the liver or its lesions, used patches as input. Li et al. [295] trained a classification model where patches were considered as positive if at least 50% of its pixels are tumoral. Details of their network are given in the figure 4.27.

Later, Frid-Adar et al. [296] also implemented a patch-based segmentation, but by separating *non-lesion* patches on whether they are located in the *liver interior* or in the *liver boundary*. The classification was performed through a multi-scale approach, followed by a CNN-based FP reduction

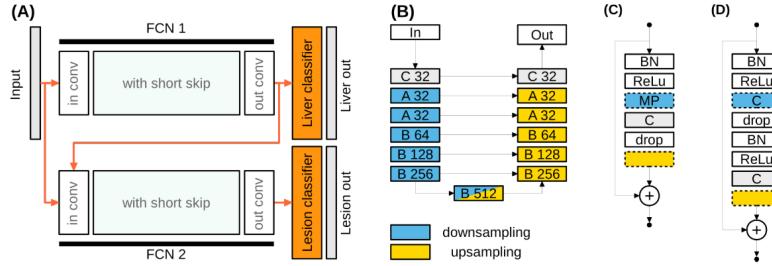


Fig. 1. (A) Two FCNs, FCN 1 and 2, each take a 2D axial slice as input. FCN 1 produces a segmentation mask for the liver; FCN 2 for lesions. The latent representation produced by FCN 1 is passed as an additional input to FCN 2. (B) FCN structure with the number of convolution filters noted in each block. Blocks coloured blue perform downsampling while those coloured yellow perform upsampling. “C” denotes a 3×3 pixel convolution layer; “A” and “B” denote blocks A and B, shown in (C) and (D), respectively. “BN”, “ReLU”, and “MP”, denote batch normalization, rectified linear units, and max pooling, respectively. Blocks with dashed lines are used in only the upsampling or the downsampling path, as denoted by colour.

Figure 4.26: Architecture implemented by ©Vorontsov et al. [284] where the liver and lesions features are computed separately before being combined using a classifier.

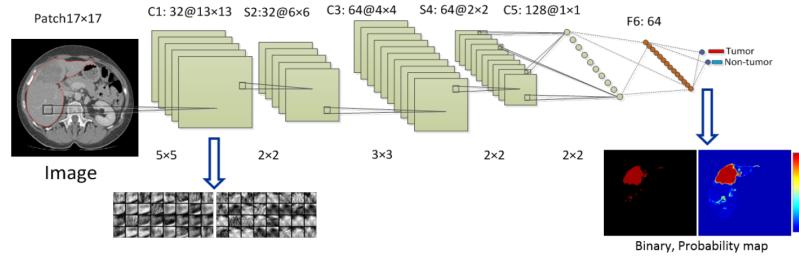


Figure 4.27: Patch-wise architecture as detailed and described by ©Li et al. [295]

step.

However, the majority of the studies implemented a slice-by-slice segmentation [288–290], whereas some others chose to introduce volumetric consistency by using 3 or sometimes 5 adjacent slices as input in a 2.5D-manner [283, 285, 287].

Worth noting that only a few studies incorporated 3D layers in their architecture, such as Li et al. who computed 3D inter-slice features in 12-slices blocks all along the different CT scans [291]. Rafiee et al. [297] decided to incorporate 3D layers only in the encoding part of their *3D-2D-FCN* model. The connection between the encoding part and the decoding part was done by custom *skip connections* to concatenate the middle slice of the 3D volume with its corresponding 2D features map in the decoding part. Dou et al. were the first to build an entire 3D network with large kernels, and had intermediate supervised layers to combat the gradient vanishing problem [298]. Details of their network are given in the figure 4.28.

Concerning the incorporation of temporal information, Sun et al. were the first to utilize multiphase information during the DL-driven automatic segmentation of liver lesions [299]. One FCN network per channel was trained to extract features before they were merged through a features fusion layer.

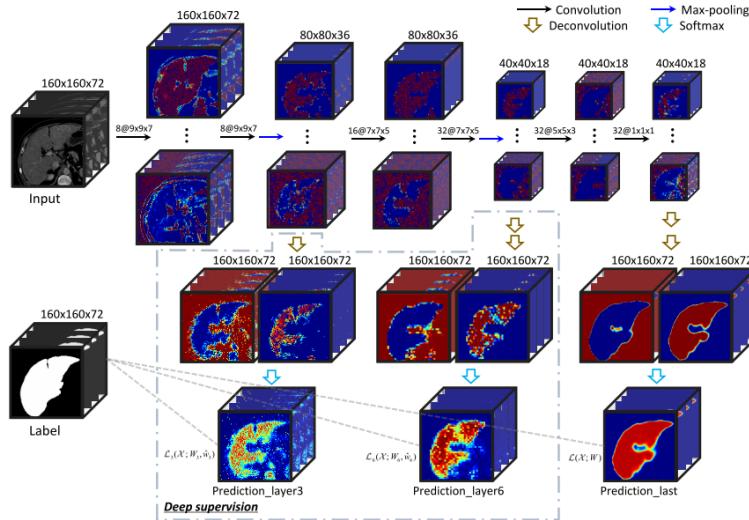


Figure 4.28: Full 3D architecture as detailed and implemented by ©Dou et al. [298]

4.4.1.4 Training strategies (pre-processing, pre-trained networks...)

Regarding the training strategies, no real consensus can be established.

Concerning the pre-processing steps, the majority of the studies decided to perform the liver coarse segmentation on a down-sampled representation of the images to limit the computational cost, before using the original resolution to perform the lesion segmentation in order to not lose any details [284, 285, 287, 290, 291].

Data normalization was also commonly applied after clipping the HU intensities to a given range, but no consensus can be deduced concerning the clipping range. Interestingly, Kaluva et al. decided to apply three different windowing ranges to the original image when carrying out the lesion segmentation which helped the network to better segment the tumor and its boundary than when only one range was applied. Almost all the reviewed studies performed data augmentation to artificially increase the size of the dataset, by most of the time employing classical geometrical transformations such as rotations, flips, ships, scalings or elastic deformations [282–285, 287, 289, 291, 294, 296, 297].

It is also worth noting that no consensus can also be deduced regarding the training from scratch or the application of fine-tuning after pre-training the architecture on another dataset such as ImageNet [283, 289, 294]. The same differences are remarkable for the choice of slices to keep during the training. Yuan et al. and Kaluva et al. decided to keep only liver slices plus a certain margin to train the liver segmentation network, whereas only slices presenting one or multiple lesions were used to train the lesion segmentation network [285, 290]. Bi et al. selected half of the slices to present both the liver and the lesions, whereas none of them were present in the other half [289]. The type of slices chosen is directly linked to the loss function used to train the network, where some studies implemented a weighted version of the cross entropy (categorical or binary) [282, 283, 287, 294], and some others

decided to directly associate the target metric with the loss function through the Dice or Jaccard distance [284, 285, 288]

4.4.1.5 Inference schemes (post-processing, ensemble learning)

Finally, the inference schemes present also some differences among the different proposed methods.

The vast majority of the studies applied a post-processing step to refine the obtained liver segmentation by extracting the largest connected component [283, 285, 287, 290, 291]. They also often took advantage of the cascade paradigm where lesions activations outside of the predicted liver can be removed [284, 285, 291]. In other studies, the reduction of the FP (*False Positives*) was done through an additional classifier filtering the different lesions candidates. Chelbus et al. [288] for example implemented an object based random forest classifier, Bellver et al. proposed a patch-based lesion detector as depicted in the figure 4.29 [283], whereas Frid-Adar et al. integrated an additional CNN designed to detect FP [296].

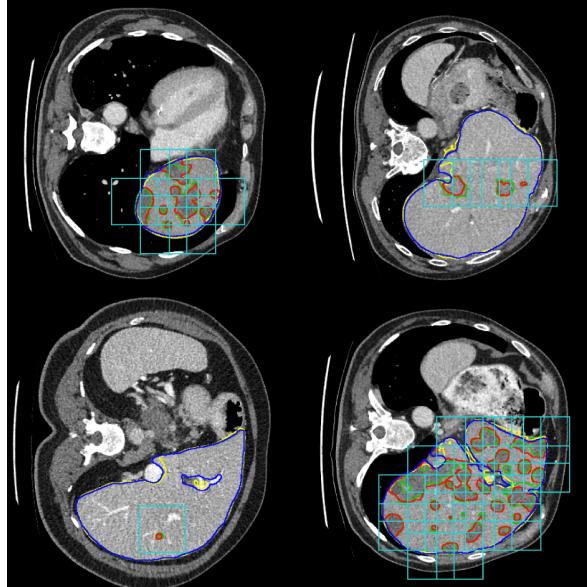


Figure 4.29: Results of the lesion detection network. Blue and red lines indicate the liver and lesion ground truth, respectively. Yellow and green lines are the segmentation results for liver and lesion. The light blue bounding boxes are the windows detected as having a lesion. All positive pixels at the output of the segmentation network will be removed if they disagree with the results of lesion detection, as reported by ©Bellver et al. [283]

Some other studies decided to add a more sophisticated step by implementing Conditional Random Fields (CRF) [294, 297, 298], however, these models tend to be difficult to train, and may also increase the inference time.

The combination of different networks in a so-called *ensemble-learning* way was also commonly realized by the different studies. Yuan et al. combined the different networks obtained through a cross-validation process to give the final segmentation [285]. Whereas Chlebus et al. trained one network per axis [288]. Bi et al. decided to train different networks following a multiscale-strategy [289],

and Vorontsov et al. trained networks on differently oriented images [284].

4.4.1.6 Conclusion

As exposed previously, no real consensus can be found to design a generic DL-related liver tissue semantic segmentation method. However, the state-of-the-art studies share some common properties such as their cascaded architecture, the use of both pre- and post-processing or the use of FCN-based networks as basis. The different teams that participated in either the MICCAI17 or ISBI17 LITS challenge were reviewed by the organizers, who came to the same conclusions [191]. They also realized that the best tumor-segmentation results were obtained for large lesions, and within a specific lesion-liver HU difference interval (a difference between -10 and -60). They also noticed that top-ranked methods used some 3D approaches in their architecture, showing perspectives to capture the whole volume context in the future. They are planning to organize again the same type of competition in the future, by particularly providing multiple ground truths and potentially splitting the lesion segmentation task into large and small lesions, since current methods still struggle in segmenting the small lesions.

More recent DL-related studies have been using publicly available databases such as LITS as a way to compare themselves to state-of-the-art methods. As a matter of facts, the key common elements that have been proposed previously served as a basis for new studies.

In the following section, we will present our experiments that were conducted to automatically perform the semantic segmentation of the liver and its potential tumors. Inspired by the above mentioned work, a sequential architecture consisting of several specialized networks was trained by considering the dynamic of the liver thanks to contrast-enhanced CT images. The relevant features computed by our deep cascaded multiphase architecture will subsequently be extracted to predict the histological grade.

Chapter 5

Histological grade prediction via automatic liver tumors segmentation

In this chapter we present the experimental results of our work. We first describe more precisely the different key concepts to be implemented such as the cascaded architecture or the incorporation of the multiphase images. We then expose our semantic segmentation experiments where the best performances were obtained by stacking several specialized networks with dynamic contrast-enhanced images as input. After validating the value brought by our semantic segmentation architecture to provide additional annotations in the case of weakly annotated datasets, we describe our DLR pipeline which incorporates the relevant semantic imaging features for the prediction of the histological grade.

5.1 Semantic segmentation applied to the study of HCC

As explained previously, state-of-the-art deep learning techniques allow modern hardware to perform automatic segmentation of both the liver and its structures with a precision close to the one obtained by the experts themselves. In this section, we will first describe the implemented networks and the different used datasets, before presenting the experiments. After selecting an optimal set of hyperparameters, we evaluated the advantages brought by a cascaded architecture, and the way to implement it, before describing the way to incorporate the multiphase information in our architecture. The results and the conclusions of this work were presented in the literature [300].

5.1.1 Motivations

We implemented different architectures to validate the hypothesis that current deep networks can perform automatic delineation of both the liver and its tumors. The wash-in wash-out being an important characteristic of the tumor (see chapter 2), we decided to design an architecture that will

try to extract features encoding this specific information. We wanted to prove that incorporating the dynamic information through the use of contrast-enhanced images can improve the performances of the network, therefore, we compared results obtained using single and multi-phase networks.

We built several architectures to confirm that the cascaded approach corresponds to the best paradigm when dealing with small multiphase datasets. Finally, we proved the ability of our cascaded architecture to provide additional annotations in weakly annotated datasets.

It is worth noting that we are the first to automatically delineate HCC on multiphase CT images using a DL architecture, and that features computed by the our architecture will further be used in a radiomics purpose to predict the histological grade.

5.1.2 Cascaded architecture

Even though no consensus was made regarding existing methods, we noticed that several studies implemented a sequential pipeline that can be modeled as cascaded architecture. The cascade consists in several networks trained to perform a specific simpler task, and that are sequentially connected, to provide a final annotation map. In our case, the goal is the segmentation of the liver and its internal structures with the differentiation between parenchyma, and both the active and the necrotic parts of the hepatic tumor. Our cascade will consequently be composed of 3 networks, as depicted in the figure 5.1. The first one will be specialized to segment the liver in abdominal axial slices, the second will delineate the contours of the tumor in the predicted liver ROI, whereas the final one will differentiate between active and necrotic parts within the obtained tumor ROI [300]. Each one of the given networks will share the same *U-Net* architecture.

5.1.3 U-Net network

As a basis architecture, we have decided to implement *U-Net* like networks because it has been previously used for the semantic segmentation task, and has proven to give good results even with a small number of training samples.

The original *U-Net* architecture was developed by Ronneberger et al. [286] and initially designed for the delineation of cells in microscopic images. As detailed previously, the network can be divided into two subparts, a contraction one where the information contained within the images is compressed, through the extraction of high to low level features, and a decoding part where the compressed information is used to reconstruct a high resolution segmentation. The architecture was initially composed of 19 convolutional layers, with a rectified linear unit function as activation. The input image had an initial size of 512×512 pixels, and at each stage of the encoding part, 2 convolutional layers are stacked with an increased number of filters. The initial pair of convolution layers used 64 filters each, whereas the

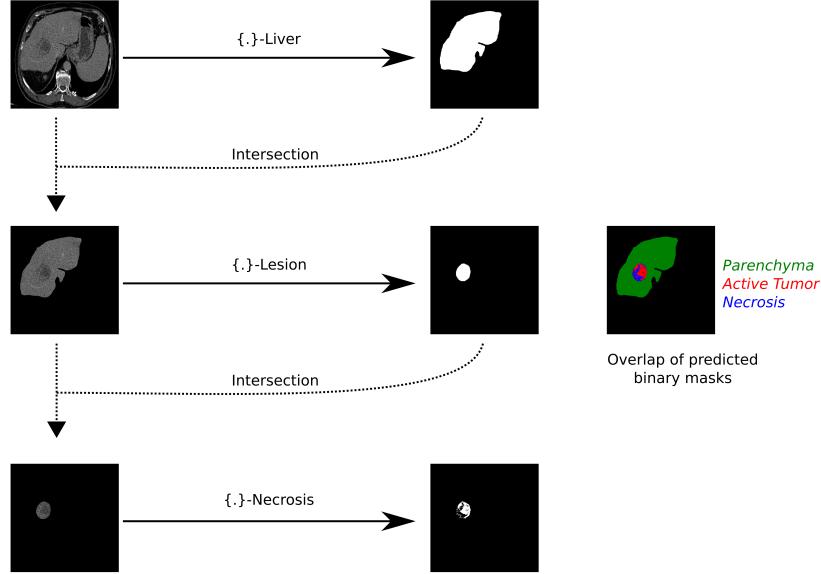


Figure 5.1: Cascaded network: The first network takes as input a CT image and segments the liver. The resulting segmentation map is used to remove non-liver pixels in the input data of the second network which performs the segmentation of lesions. The last network segments the necrosis within the lesions. The three binary masks are combined in the final segmentation map.

final stage of the encoding stage used 1024 filters. After each pair of convolutional layers, a max pooling layer is applied to halve the spatial dimension of the features maps. As a result, a $30 \times 30 \times 1024$ features map is produced at the bottleneck of the network. This representation is then reformatted in the decoding part to obtain a segmentation map with a size equivalent to the one of the input image. In order to increase the spatial dimension of the features maps, Up-Convolutional layers are implemented. It is worth noting that the number of filters used in each of the pair of convolutional layers is decreased from the bottleneck to the final layer of the network. The last convolutional layers will map the obtained features to the final number of dimensions of the segmentation maps, which will correspond to the number of classes to predict. The final layer implements a softmax function, to simulate a prediction of appartenance to each one of the output class.

In comparison to the original architecture, we have implemented zero-padding convolutions to preserve the image size and obtain a segmentation map with the same spatial dimension as the input image. We conserved the same settings as in the original architecture concerning the number of filters to use at each stage, starting with a pair of convolutions of 64 filters each, and reaching a $32 \times 32 \times 1024$ features map in the bottleneck part of the network.

The same naming-system as in our study will be used here [300]. Single-phase elementary networks will be referred to by both the input phase and the segmentation target, as an example, PV-Lesion will refer to the network responsible for the segmentation of the lesion, with PV (Portal Venous) phase images as input. The complete *U-Net* architecture for this specific elementary network is depicted in the figure 5.2.

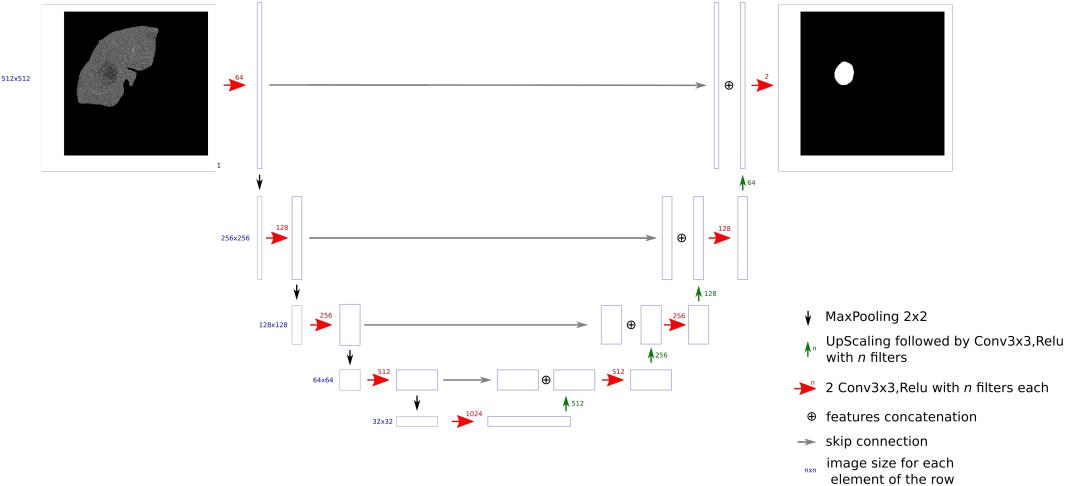


Figure 5.2: PV-Lesion network used to segment lesions within the liver with a PV image as input

In order to evaluate the improvements brought by the cascaded architecture, we also trained the original *U-Net* architecture to perform simultaneously the whole internal tissues segmentation task. The same naming system as previously will be used where “Full” corresponds to the simultaneous segmentation task, thus, AR-Full will refer to the network dedicated to the segmentation of both the parenchyma, the active and the necrotic part of the lesions simultaneously, with AR images as input. An illustration of the network is given in the figure 5.3.

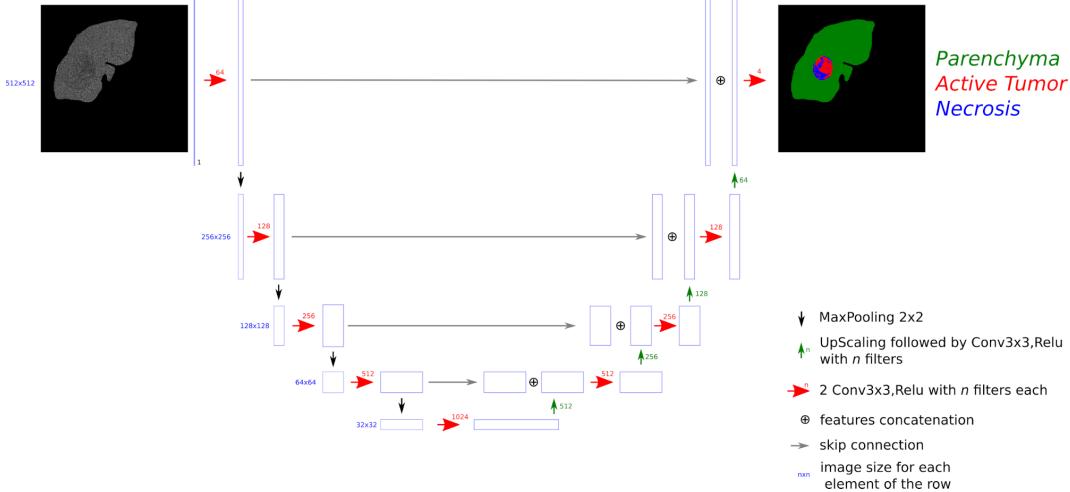


Figure 5.3: AR-Full refers to the network trained with AR images as input (values outside the liver are masked), and that outputs a label map, with parenchyma, active and necrotic parts annotated

5.1.4 Multiphase information

Only the AR and PV phases were considered in the multiphase networks because NECT phase images do not provide enough inter-tissue contrast.

In order to incorporate the multiphase information in our pipeline, we investigated 2 different strategies. The first one, referred to as DMP (Dimensional MultiPhase), consists in concatenating both the AR and

PV images as input to the network (see figure 5.4). The second one referred to as MPF (MultiPhase Fusion), consists in performing both the encoding and the decoding separately for each phase, before merging the output maps (simple addition on the obtained features maps), as depicted in the figure 5.5.

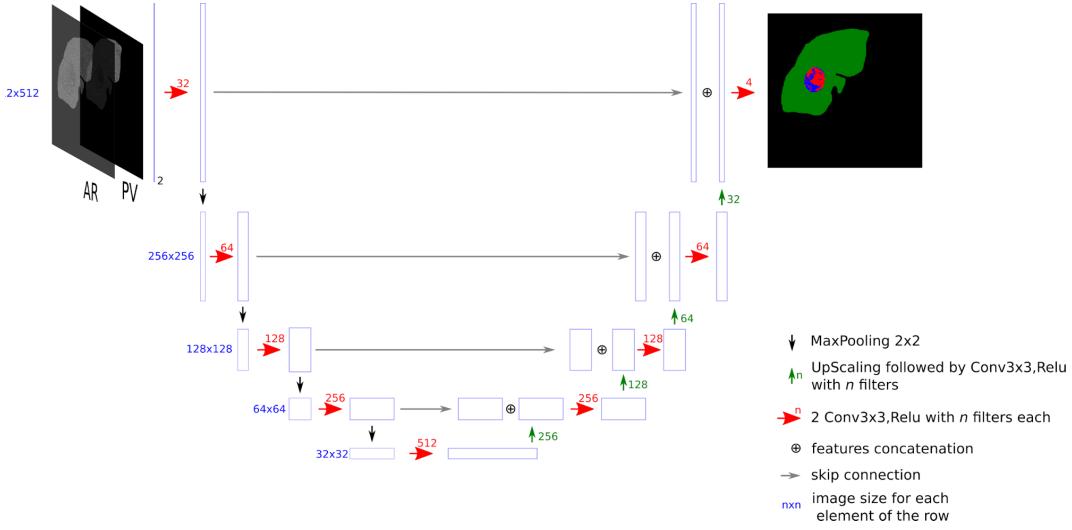


Figure 5.4: DMP-Full network that combines the AR and the PV images as an input to segment the parenchyma and both the active and the necrotic parts of the lesions. Here, the two channels are considered as features for the first layer

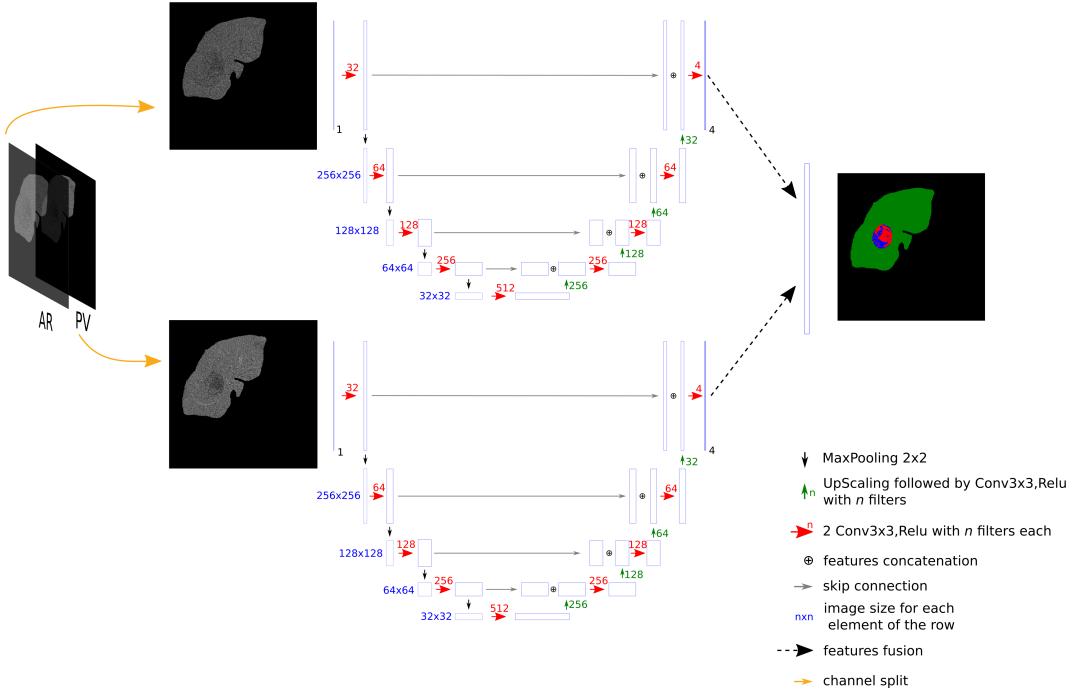


Figure 5.5: MPF-Full network: initially, AR and PV images are processed separately. The resulting maps are merged (by simple addition) at the end

5.1.5 Datasets

The different datasets used in our research work are detailed in the table 5.1.

Db Name	Db number of cases	retained	Axial voxel size (mm)	Slice thickness (mm)	Available contrast-enhanced phases	Contains Tumor	Tumor type	Liver truth	Tumor Ground truth	Necrosis Ground truth	#Experts
TheraHCC-dB	104 2D slices	0.66-0.97	0.7-1.25	NECT, AR & PV	Yes	HCC	true	true	true	true	4
LITS-dB	131 3D vol.	0.55-1	0.45-6	Single phase images but mixed (AR & PV)	Yes	Mixed	true	true	false	false	3
3Dlrcad-dB ¹	15 3D vol.	0.57-0.87	1.25-4	Single phase images but mixed (AR & PV)	Yes	Mixed	true	true	false	false	-
TCIA-dB ²	18 3D vol.	0.62-0.90	2.5-7	AR&PV	Yes	HCC	false	true	true	true	1
G-dB	79 3D vol.	0.58-0.98	0.8-5	AR&PV	Yes	HCC	false	true	false	false	1

Table 5.1: Datasets used in our experiments (NECT: Non-Enhanced Contrast CT, AR: Arterial, PV: Portal Venous)

¹ Subset of the 3Dlrcad-01 presented in the table 4.1 where only the 15 volumes with liver tumors were retained
² Subset of the one presented in the table 4.1 where only the 18 exploitable volumes with both AR and PV phases were retained

As we can see in the table, and except for **LITS-dB** and **3DIrcad-dB** (**3DIrcad-dB** being a subset of the **LITS-dB** [191]), they are all different in their construction. Differences can be found on the annotated areas (chosen sparse slices or entire 3D volumes) and on the annotated tissues that can be found in the dataset (some of them only contain ground truth annotations for the liver and the tumors it might contain such as **LITS-dB**, whereas some others contain only ground truth annotation for the tumor such as **TCIA-dB**). Another crucial difference concerns the contrast enhanced phases available in each dataset (**LITS-dB** contains only single phase images whereas **TCIA-dB**, **TheraHCC-dB** and **G-dB** present multiphasic images).

To prove the ability of the deep learning to perform the automatic segmentation of both the liver and its internal tissues, such as the parenchyma and both the active and the necrotic part of the tumor, we first performed our experiments on **TheraHCC-dB** since this database was previously used for the same task [281], because it presents multiphase images and finally because this is the only available dataset with complete ground truth for both liver parenchyma, active and necrotic part of the tumors.

TheraHCC-dB is composed of images from seven patients, all suffering from *HCC* and who underwent CECT (Contrast-Enhanced Computed Tomography) examinations, resulting in a total of 13 CT sequences.

More details about the standard CECT examination protocol can be found in the chapter 2.2.3.2. In our case, images were acquired at 4 different moments: one before the injection of the contrast medium (NECT: Non Enhanced CT), and the 2 others after the injection to reflect both the arterial (AR) (~20-25s after injection) and the portal venous (PV) phases (~60-70s after injection).

Eight regularly sampled slices across each one the 13 sequences were segmented by 4 experts, resulting in 104 labeled slices.

The segmentation maps obtained from each one the 4 experts were fused using the STAPLE algorithm to reach a consensus map [301].

5.1.6 Experiments

5.1.6.1 Data pre-processing

The first task to implement in this study was the inter-phase registration so that environmental effects, such as respiratory motions, will not affect the performances of the networks, and to ensure that a given voxel is at the exact same position for the different CECT volumes of a patient. The registration was performed using a diffeomorphic deformable registration algorithm, where the PV images were used as reference, since they contained the original expert annotations [281, 282, 294, 302].

Another bias that can affect the deep semantic segmentation networks training is the heterogeneous

image sizes and voxel resolutions present within the training images. To avoid this bias, it has been decided to scale them so that they all have a 512×512 axial size³ and an isotropic voxel resolution of 0.97mm^2 .

The data normalization is another aspect that needs to be considered before feeding the images in the deep network. In order to reduce the effect of extreme values from regions present in the tomographic images (such as the bones or the air), and to enhance the intensity of the liver voxels, we first clipped the *HU* values to be in the range $[-100, 400]$, corresponding to the most commonly observed liver intensities range. The retained intensities were finally mapped to the interval $[0, 1]$ ⁴.

5.1.6.2 Training

In order to validate our hypotheses, we have decided to first run experiments on the **3Dircad-dB** to set the most crucial hyperparameters such as the learning rate, the decay, the depth of the network or the type and the amount of data augmentation. The selection process is detailed in the appendix B, and here is the list of the chosen hyperparameters:

- Lr: 1e-4
- Decay: 1e-4
- Number of epochs: 20
- Optimizer: Adam
- Number of filters at bottleneck: 1024
- Input image size: 512
- Data augmentation
 - Rotations in the interval $[0, 40]$
 - Translations with shift in the interval $[-0.1, 0.1]$
 - Horizontal and vertical flips
- Augmentation factor: 20

When transferring those settings to the **TheraHCC-dB**, we implemented both translation and the addition of gaussian noise in the training since it slightly improved the performances. In order to remove any bias, the same set of hyperparameters has been used for both the **{.}-Liver**, **{.}-Lesion**, **{.}-Necrosis**

³ Either cropping or padding was performed so that the images have the same size ⁴ DL studies usually tend to perform data standardization but here *HU* values are already normalized. The used range helps us to avoid potential bias caused by images acquired at a delayed phase with a different contrast

and $\{\cdot\}$ -Full networks when training the **TheraHCC-dB**, regardless of the type of input (single phase or multiphase).

The same set of hyperparameters was used to train both liver and lesions segmentation networks on respectively the **LITS-dB** and the **G-dB**.

We first present the results obtain when performing the semantic segmentation of liver tissues on the **TheraHCC-dB**, before applied the resulting architecture to provide additional annotations to both **TCIA-dB** and **G-dB**.

5.1.6.3 Semantic segmentation of liver tissues

In this section, mean DSCs were computed in a slice-wise fashion for each target class and the different methods were statistically compared using the Wilcoxon signed paired rank tests, since slice-wise DSCs did not follow a normal distribution. We first compared $\{\text{NECT, AR, PV, DMP, MPF}\}$ -Liver networks to evaluate which phase allows better liver segmentation. We then trained $\{\text{NECT, AR, PV, DMP, MCF}\}$ -Lesion and $\{\text{NECT, AR, PV, DMP, MCF}\}$ -Necrosis networks separately on the **TheraHCC-dB** by masking all values outside the liver using ground truth annotations in order to assess whether multiphase information is really useful for the segmentation. The results are given in table 5.2.

Multiphase performed significantly better than single-phase for segmenting the liver (DMP vs PV, $P = 0.001$; DMP vs AR, $P = 0.005$; DMP vs NECT, $P < 0.001$) and the active part of the lesions (DMP vs PV, $P < 0.001$; DMP vs AR, $P = 0.003$; DMP vs NECT, $P < 0.001$). When comparing single-phase alone, PV achieved significantly better DSCs than AR or NECT for all the segmentation tasks except for the liver segmentation. When comparing multiphase methods, DMP carries out significantly better than MPF for the segmentation of the liver (DMP vs MPF, $P = 0.004$), the parenchyma (DMP vs MPF, $P < 0.001$) and the active part of the lesions (DMP vs MPF, $P = 0.005$).

Since both DMP-Lesion and DMP-Necrosis led to the best results, we combined them in a cascade as explained before, and compared it to both $\{\text{NECT, AR, PV, DMP, MPF}\}$ -Full networks. We evaluated them in terms of liver tissue classification performance on images where the values outside the ground truth liver area were masked. The mean DSCs are reported in table 5.3. Examples of segmentation results are given in figure 5.6.

The results highlighted that the cascaded version performed significantly better than $\{\cdot\}$ -Full networks for segmenting the active part of the lesion (Cascaded DMP vs PV-Full, $P = 0.001$). The resulting segmentation maps allow us to estimate the necrosis rate (which will be the ratio between the necrotic and the tumor areas). This metric is commonly used for diagnosis and

Table 5.2: Segmentation results using single-phase vs multiphase methods on TheraHCC-dB.

Input	Target	Network				
		NECT	AR	PV	DMP	MPF
Raw CT	Liver	81.1 \pm 27.7	89.5 \pm 13.2	88.7 \pm 11.4	89.9 \pm 15.6	88.2 \pm 16.0
True liver mask	Parenchyma	86.5 \pm 13.7	82.5 \pm 18.6	88.7 \pm 15.4	90.5 \pm 13.2	86.9 \pm 17.8
True liver mask	Lesion	77.4 \pm 24.1	77.4 \pm 20.2	87.8 \pm 9.7	88.5 \pm 11.7	86.6 \pm 10.3
True lesion mask	Necrosis	67.5 \pm 15.5	69.7 \pm 16.3	77.8 \pm 12.4	78.5 \pm 13.3	78.8 \pm 11.7
True lesion mask	Active Tumor	65.6 \pm 20.4	63.9 \pm 22.6	71.6 \pm 20.7	75.5 \pm 17.4	73.2 \pm 18.6

Table 5.3: Segmentation results using $\{\cdot\}$ -Full vs cascaded architectures on TheraHCC-dB.

Target	Network					
	NECT-Full	AR-Full	PV-Full	DMP-Full	MPF-Full	Cascaded DMP
Parenchyma	85.3 \pm 14.9	84.1 \pm 17.9	87.0 \pm 19.0	82.9 \pm 19.7	87.9 \pm 15.9	90.5 \pm 13.2
Necrosis	62.6 \pm 18.6	63.5 \pm 21.5	75.7 \pm 14.4	73.7 \pm 14.1	75.6 \pm 13.4	75.8 \pm 15.1
Active Tumor	42.2 \pm 24.0	43.2 \pm 26.1	53.5 \pm 24.2	51.3 \pm 25.6	52.0 \pm 23.3	59.6 \pm 22.5

prognosis of the treatment outcome. In this configuration⁵, our workflow provided estimates of this valuable biomarker with a mean error rate of 13.0 %, which is accurate enough for clinical application.

When compared with another study that have been using a different dataset composed of MR images, we achieved slightly better segmentation results [303]. We evaluated our method on the same database used in [281], where a manual expert interaction was required for the segmentation phase, which is not the case in the present deep learning approach. To allow fair comparison, the evaluation was conducted on the areas where all the experts reached an agreement as in [281]. The mean patient-wise segmentation DSCs are depicted in table 5.4. Our method enabled a better segmentation of the lesions and both necrotic and active parts. Therefore, we were able to predict the patient-wise necrosis rate with a slightly better precision.

Table 5.4: Average patient-wise segmentation DSCs (on full agreement expert area) with the semi-interactive approach of [281] and the cascaded DMP method

	Semi-interactive method [281]	ours
Parenchyma	93.7 \pm 3.4	92.2 \pm 4.7
Lesion	90.7 \pm 6	91.8 \pm 4
Necrosis	83.0 \pm 12.9	83.6 \pm 11.7
Active Tumor	75.2 \pm 10.9	82.0 \pm 6.4
Necrosis rate error	7.84 \pm 4.4	7.10 \pm 1.4

We finally combined the DMP-Liver, the DMP-Lesion and DMP-Necrosis networks in a cascaded fashion, as described in the figure 5.1. This allowed us to perform a fully automatic segmentation of the raw (unmasked) CT images. We reached average slice-wise DSCs of 78.3 ± 22.1 for the segmentation of the parenchyma, 50.6 ± 24.6 for the segmentation of the active part, and 68.1 ± 23.2 for the necrotic part of the lesions. We also provided a necrosis rate per patient with a mean error of 15.9% when

⁵ after using the liver expert GT as mask for the first step

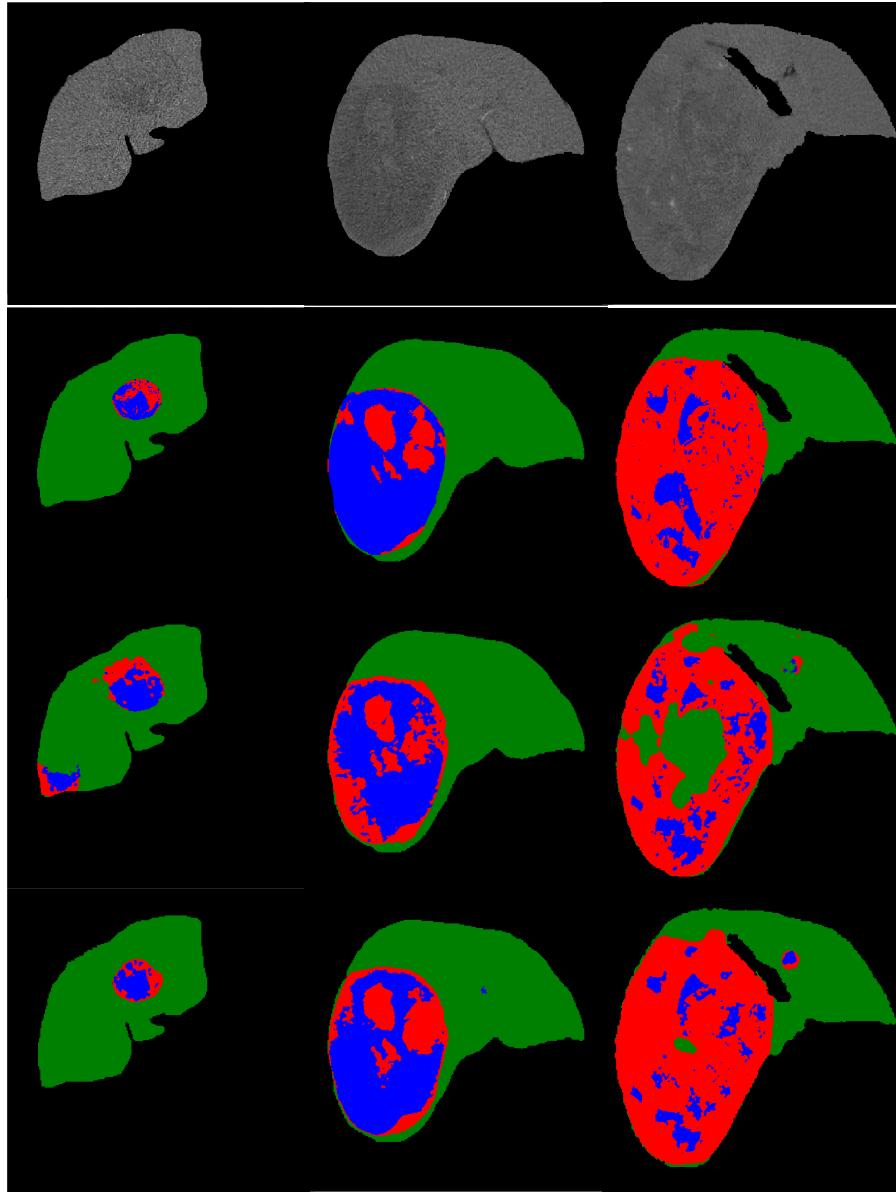


Figure 5.6: From top to bottom : Raw images with HU values inside the liver, Ground truth, DMP-Full segmentation, Cascaded DMP segmentation

compared to expert necrosis rate estimation⁶. Examples of fully automatic segmentation results are given in figure 5.7.

When comparing results obtained by the specialized networks, we proved that the addition of the multiphase information provided better segmentation results than when only single phase images are used. Statistically significant improvement was obtained for the segmentation of both the liver and the active part of the tumors. We also investigated the performances of the single phase networks, and noticed that the PV phase was the one providing the best results, where significant improvement was obtained for all the segmentation tasks, except for the liver segmentation. Regarding the internal liver tissues segmentation, the elementary networks providing the best results were the DMP-Lesion and DMP-Necrosis ones. When combined in a cascaded way, they performed better than each one of the

⁶ Here the estimation is performed on the raw images in an automatic fashion

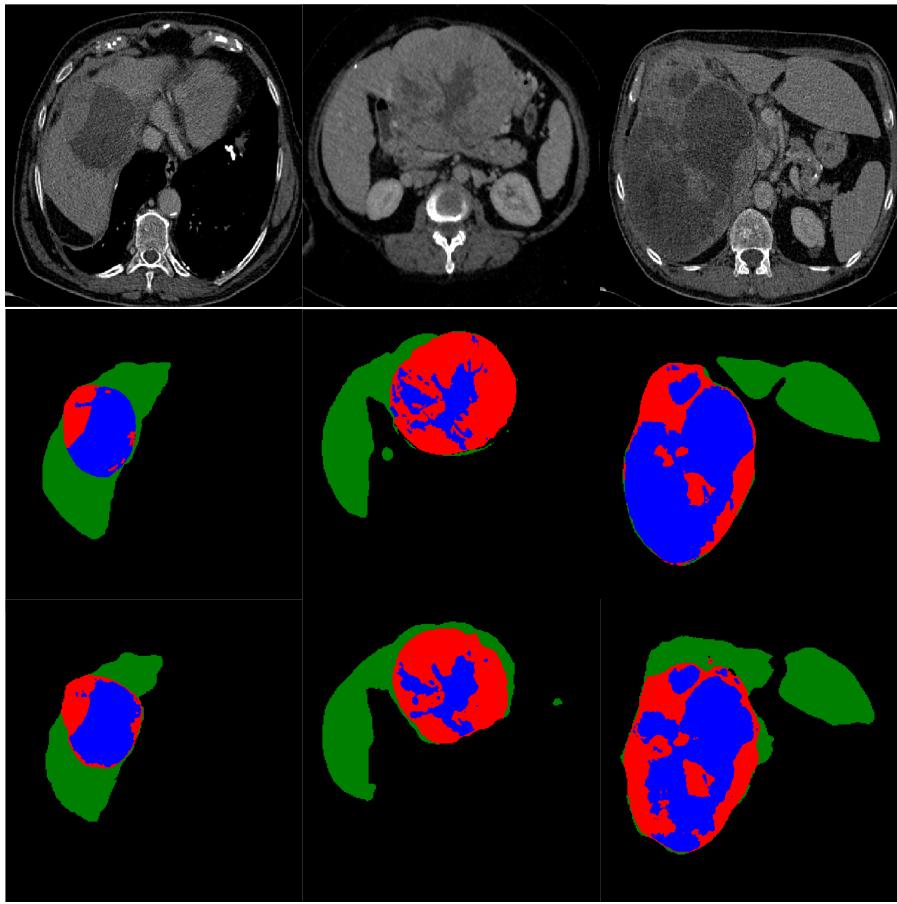


Figure 5.7: From top to bottom : Raw images, ground truth and results of the fully automatic segmentation of liver tissue

{.}-Full architecture, with a statistically significant improvement for the segmentation of the active part of the lesions, meaning that combining several specialized networks together is more efficient than training a single network addressing all tasks simultaneously.

With the same experimental conditions, our solution provided a better accuracy for the segmentation of the lesions, and for both their active and the necrotic parts, than the one obtained using a semi-automatic technique with expert interaction [281, 300]. We also obtained equivalent results than those from a similar study considering MR images as input [304].

We concluded that the combination of multiphase registered images used in a cascaded architecture allows us to automatically perform the semantic segmentation of liver tissues.

5.1.6.4 Semantic segmentation applied to weakly annotated datasets

We proved that semantic segmentation of liver tissues can be performed using DL through robust CV training on both **3Dircad-dB** and **TheraHCC-dB**. We also demonstrated that both a cascaded architecture and the use of multiphase information allows a better segmentation accuracy than using single phase images only. We will now extend our work to prove the ability of semantic segmentation

architecture to provide missing annotations into weakly annotated datasets such as **TCIA-dB**.

TCIA-dB originally only contained raw images without annotations, so an expert performed the delineations for both the tumor and the necrosis areas on PV images only.

To achieve a multiphasic semantic segmentation of the liver tissues we have to ensure that both the liver and its internal structures such as the potential tumors are located at the same spatial position between the different CECT volumes.

5.1.6.4.1 Registration

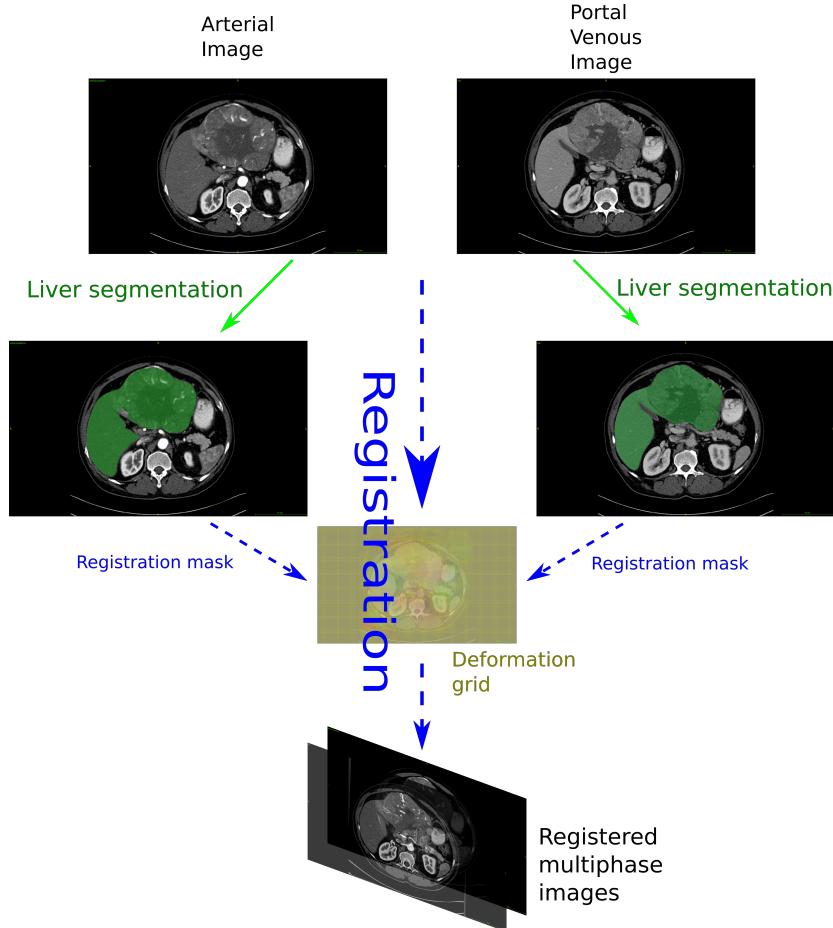


Figure 5.8: Illustration of the registration pipeline applied to the images of **TCIA-dB**. The first green arrows correspond to the liver segmentation using a network trained on the **LITS-dB**. Dashed blue arrows correspond to the ANTs registration pipeline, where a dilated version of the predicted liver annotations maps are used as registration masks. The ANTs algorithm implements 3 transformations: a rigid, an affine, and a diffeomorphic Syn transformation that computes a deformation grid [302]. The 3 steps of the ANTs algorithm allows us to obtain registered multiphase images.

One way to perform a registration is to implement a series of transformations (rigid or non-rigid) that will match a moving volume to a target one. Each step of the registration is controlled by a similarity loss. When dealing with CT volumes the available losses controlling the different steps of the registration pipeline can be affected by areas of the body with a high gradient such as the bones for example. When registering two CT volumes, one can either directly use both the entire volumes or

constraints the registration to a specific area (aka mask). The liver being a soft organ, it easily moves with the respiratory motions, in order to reduce the effect of the neighboring parts of the abdomen, we have decided to restrict the computation of the similarity metrics to the dilated liver mask area so that the registration algorithm mainly focuses on the gradient present along its borders.

Consequently, the first step for the **TCIA-dB** registration is to perform a liver semantic segmentation (green arrows in the figure 5.8).

5.1.6.4.2 Automatic liver segmentation

In the available datasets (as presented in the table 5.1), only **TheraHCC-dB** and **LITS-dB** contained expert liver delineation.

TheraHCC-dB contains annotations only on sparse slices across the liver, whereas **LITS-dB** contains full 3D pixel-wise liver annotation but it only contains monophase images, without any information regarding the acquisition phase (AR, PV and potentially **DELAY** volumes are mixed in the dataset). Our experiments however showed that a liver segmentation network trained on sufficiently enough cases is able to perform the semantic segmentation of both AR and PV raw images independently. We trained our network called **CECT-Liver** on the 131 volumes of the **LITS-dB** using the same hyperparameters as the ones detailed previously. When testing the **CECT-Liver** network on **TheraHCC-dB** we obtained a mean slice-wise DSC of 90.4 ± 17.5 on the PV images and 86.9 ± 19.1 on the AR images. These results, close to those obtained in our previous work using a CV approach [300], proved that **CECT-Liver** can perform liver segmentation on both AR and PV unseen images.

The **CECT-Liver** network was also able to segment unseen volumes of the **TCIA-dB** in both AR and PV phases even when the liver presents a big lesion, as we can see in the figure 5.9.

Once a robust liver segmentation was obtained, we implemented our registration pipeline.

We have decided to implement the registration pipeline using ANTs [305], since it has already been used for liver CT scans registration [306, 307].

5.1.6.4.3 ANTs registration

The classical ANTs registration pipeline is made of 3 steps. The first two steps consist of linear transformations, where a rigid transformation is first applied, followed by an affine transformation. The last one is generally non-linear. In our pipeline, we used a **Syn** (*Standard Symmetric normalisation*) transformation, which processes a gradient field, determining how each point of the space will shift [302].

The MI (Mutual Information) was used as loss function for the first two steps because it has the advantage of computing the similarity at a large scale. Linear rigid and affine transformations tend to roughly bring both volumes in the same space so we decided not to use a finer similarity metric. For the **Syn** transformation, we used the CC (Cross Correlation) as loss function. The CC has the

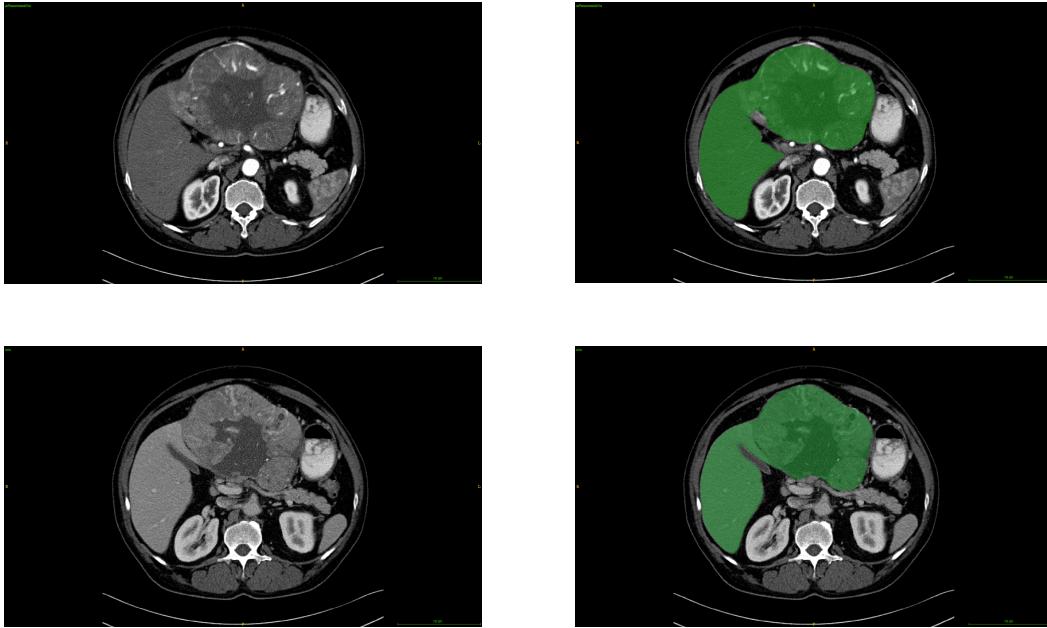


Figure 5.9: Example of liver segmentation using the CECT-Liver network on TCIA-dB patients (Top row AR images, bottom row: PV images, left: Raw images, right : liver segmentation as overlay).

advantage of looking precisely in a region around each voxel when computing the similarity between the two volumes.

During the registration process, the liver segmentation was used as registration mask, and we decided to set the PV volume as target (fixed) volume since it usually presents the finer voxel resolution when compared to AR (or DELAY) volume, and since it contains the original expert annotations.

1. We initially resampled the AR volume so it has the same resolution as the corresponding PV volume.
2. We performed the liver segmentation using CECT-Liver on both the PV and the resampled AR volumes in a slice-wise manner with a classical post-processing consisting in applying a binary opening operation to the mask and conserving the big connected component (see green arrows in the figure 5.8). We obtained a liver mask for both the PV and the resampled AR volumes.
3. We applied the registration using both dilated version of the two masks obtained at the end of the second step as registration masks (as depicted by the dashed blue arrows in the figure 5.8) (we dilated the liver mask with a SSE of 5cm when setting the registration mask in order to counter any error in the segmentation process, and to always have both the liver and its border included in the registration mask)

Table 5.5: Registration pipeline

The registration allows us to obtain both a new registered arterial volume, a transformation matrix and a deformation field volume. In the axial plane, the obtained deformation fields tend to present high deformation at both the top and the bottom of the liver (which can be explained anatomically since they are surrounded by the air and will be more subject to deformation than central areas of the liver) whereas central areas of the liver present high deformation close to the border. Examples of

obtained deformation fields at the end of the registration pipeline are depicted in the figure 5.10.

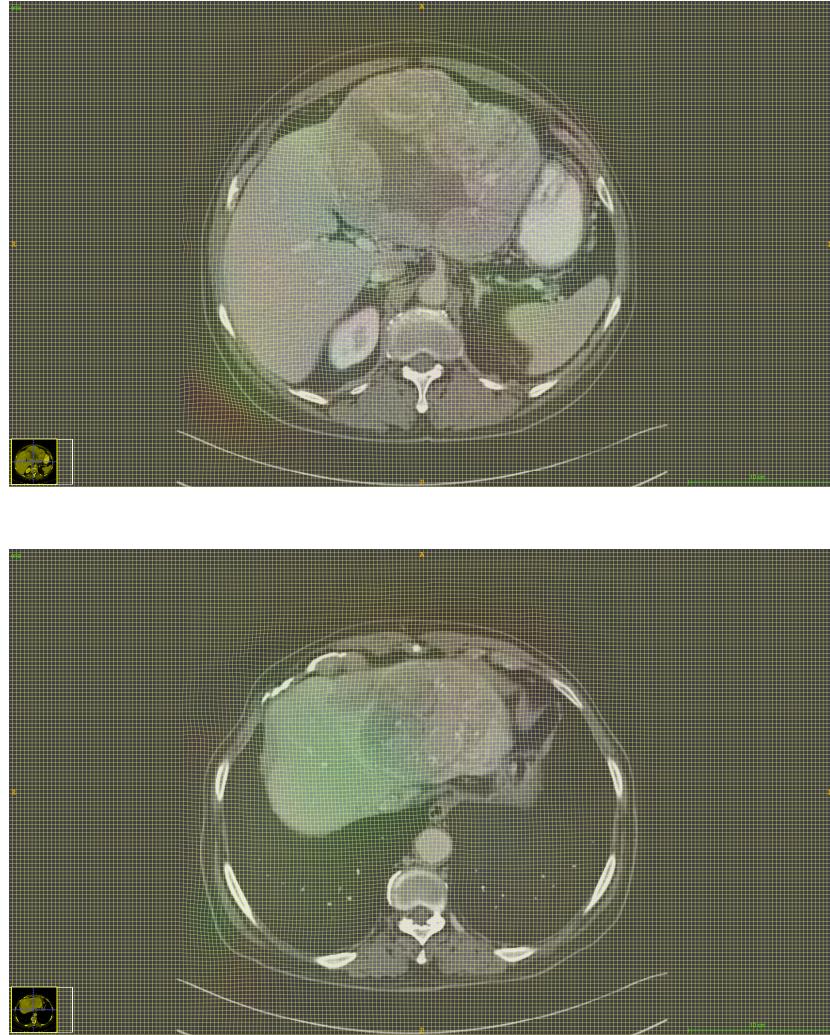


Figure 5.10: Example of deformation grids obtained after applying the registration pipeline on **TCIA-dB** patients

One way to assess the precision of the registration is to apply the transformation matrix to the initial resampled AR liver mask and to compute the DSC with the target PV liver mask. When applying this evaluation, we obtained a mean patient-wise DSC of 92.8 ± 3.8 between the target PV and the registered AR liver masks at the end of the registration step on the **TCIA-dB**, sufficient to consider the registration as successful when compared with those obtained by state-of-the-art registration methods applied to the liver (where mean liver masks dices obtained after a subject-to-subject registration varied between 0.96 and 0.91 when applied to respectively the SLIVER and the LITS datasets) [306].

The complete registration pipeline applied to the **TCIA-dB** is detailed in the table 5.5.

5.1.6.4.4 Automatic multiphase tumor segmentation

Once both the AR and PV volumes of the **TCIA-dB** were registered, and the liver masks obtained, we built the second step of our cascaded architecture, dedicated to the multiphasic tumor segmentation.

In order to train such a network, we created a multiphase (arterial and portal venous) database where the temporal volumes of a given patient are registered and where segmentations of both the liver and tumor regions are available. **TheraHCC-dB** and **G-dB** were the only two datasets containing multiphasic images (as we can see in table 5.1), however **G-dB** initially contained ground truth annotation only for the tumors, with segmentations performed only on the AR phase images. Since **G-dB** contains more cases than **TheraHCC-dB**, we have decided to build our multiphasic tumor segmentation network on this dataset. However in order to train a tumor segmentation network, our cascaded architecture requires to have the liver segmentation mask as input for the second step (as depicted in the figure 5.4), therefore we had to obtain the liver segmentation mask for the patients of the **G-dB**.

AR and PV volumes of the **G-dB** dataset were initially not registered so we applied the same registration pipeline as the one used for **TCIA-dB** (see table 5.5). After application of the procedure to each patient of **G-dB**, we used the resulting transformation matrix to transform the AR tumor mask to the PV space as depicted by the red and green dashed arrows in the figure 5.11.

We then fused the liver and the tumor masks to obtain a multiclass segmentation mask that can fit both the PV and the registered AR volumes, as depicted in the figure 5.12.

We were able thanks to our cascaded architecture and a robust liver segmentation network to provide additional annotations to the volumes present in **G-dB** that originally contained only experts annotations for the tumor area on AR volumes.

Once a complete database where both the liver and the tumor segmentation masks were available, and where AR and PV volumes were registered, we trained a robust multiphase tumor segmentation network.

We trained both a DMP-Tumor and a MPF-Tumor segmentation network on the registered **G-dB** dataset, and evaluated them on **TCIA-dB** which contained expert tumor annotations. We evaluated both DMP and MPF architectures since no statistical differences were available when comparing results obtained for the tumor segmentation in our previous work on **TheraHCC-dB** [300].

After training both architectures with the same parameters, we obtained a mean patient-wise DSC of 73.2 ± 20.6 with MPF architecture versus 64.9 ± 27.2 when using the DMP when evaluating the models on the **TCIA-dB** patients. An example of prediction on the **TCIA-dB** is depicted in figure 5.13.

Those results obtained on an external dataset tend to demonstrate a satisfactory precision of the tumor segmentation when training a cascaded architecture with a sufficient number of cases. We retained the MPF-Tumor segmentation network since it performed significantly better than the DMP one ($p = 0.02$ using a Wilcoxon signed paired rank test on the patient-wise DSC). We confirmed the benefit of the cascaded architecture since these results were obtained using an architecture where the first network was trained on **LITS-dB** and the second on **G-dB**. We were also able to use a monophase

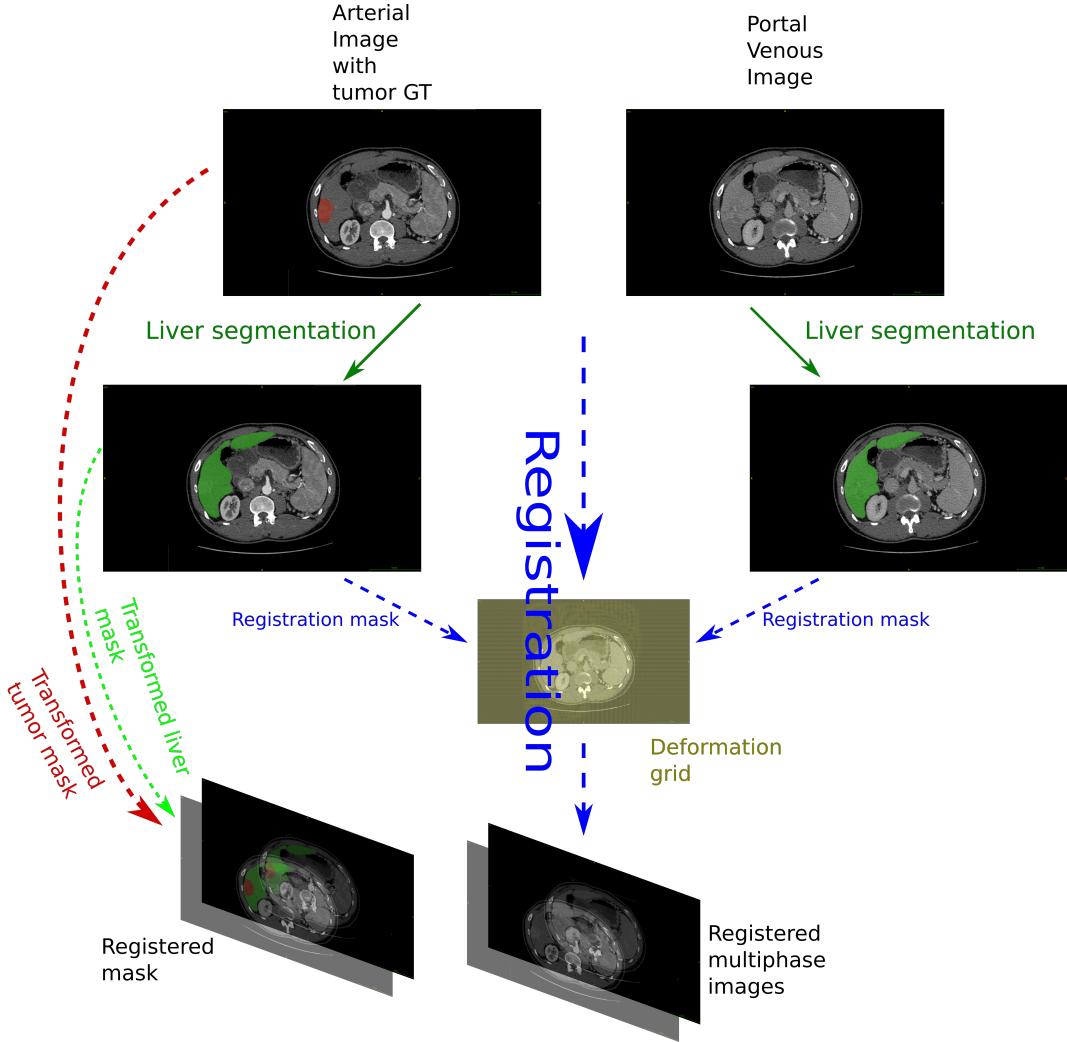


Figure 5.11: Illustration of the registration pipeline applied to **G-dB**. A similar approach as the one applied to **TCIA-dB** is performed to obtain the registered multiphase images. A final step is added here to transform both the tumor and the liver masks using the registration transformation matrix (red and green dashed arrows)

network for the first step and a multiphasic network for the second. This work proved the ability of deep learning (semantic segmentation network) combined with image processing (registration) to enhance and complete weakly annotated databases (both **TCIA-dB** and **G-dB** were enhanced in the same way).

5.2 Deep Radiomics for histological grade prediction

5.2.1 Introduction

The main goal of our work is to use imaging features to better characterize the liver cancers. We decided to extract relevant imaging features from our multiphasic cascaded semantic segmentation to perform the prediction of histological grade. The prediction of the histological grade of the tumor was rarely studied, since only one study used deep learning tools to perform this task but with MR images as input [183] (see section 2.2.3.2 to get more details on both modalities and our choice to focus on

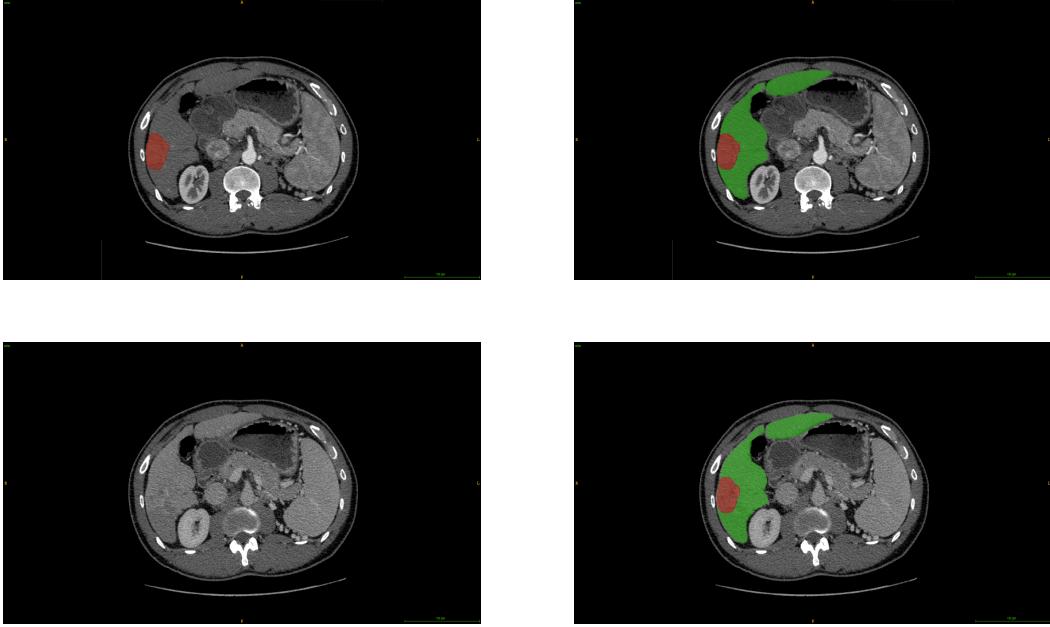


Figure 5.12: Example of a patient from **G-dB**, obtained after enhancing the dataset with our semantic segmentation network and our registration pipeline. Top row: AR_registered volume with original tumor expert segmentation, bottom row: PV_volume, left: original raw volumes, right: segmentation mask overlay where the parenchyma is obtained through our segmentation pipeline and the tumor was initially delineated by an expert then transformed to fit the target volume space.



Figure 5.13: Example of an image from the **TCIA-dB**, with the obtained predicted tumor segmentation using the MPF-Tumor segmentation network (left: raw, middle: expert annotation, right: obtained segmentation)

CT imaging). Indeed, this task appears to be more challenging than previous DLR liver-related work where either the type of FLL or the treatment response (such as the recurrence) were predicted (see section 3.3), however, being able to perform such a prediction could highly improve the automatic characterization of liver tumors, as seen in the section 2.2.3.1.1. We will now review the only study in the literature that addressed the problem of predicting the histological grade of HCCs through a DL architecture with medical images, before presenting our own automatic DL pipeline.

We now present our contribution.

5.2.2 Related work

To our knowledge, only one study tackled the problem of estimating the histological grade of HCCs using a DL architecture, but with MR images as input [183].

Yang et al. incorporated 42 patients suffering from HCC in their study, resulting in a total of 51

HCCs. Each lesion was analyzed by 2 experienced pathologists who estimated their histological grade after microscopic examination (the lesions were classified as well, moderately and poorly differentiated, following the WHO classification system [7]). The extracted tissues were obtained through either biopsy (12 patients) or after surgical removal (2 liver transplants and 28 liver resection). All the 42 patients underwent pre-operative multiphasic MR imaging examinations and images were available at 5 different phases (precontrast, late arterial, portal venous, equilibrium and delayed phases). They obtained a dataset composed of 9 well, 7 poorly and 35 moderately differentiated HCCs.

For each patient, a ROI was placed by one expert at the maximal axial cross-sectional area to entirely cover the tumor. The ROI was copied in the 2 slices above and below the chosen one, to obtain a 3D volume. Intensities of each volume were normalized and 4D tensors were created for each patient so that each tensor had a $32 \times 32 \times 5 \times 5$ shape (32×32 corresponding to the resampled axial ROI dimension, the third dimension being the number of retained slices, and the last dimension being the dynamic temporal evolution of the ROI with the 5 phases).

The used architecture is depicted in the figure 5.14. It first splits the 4D tensors into 5 3D objects so that each slice is treated separately. Each 3D volume was then processed by 2 convolutional, 2 max pooling and 1 fully connected layer. The features of each slice were then concatenated, before a second fully connected layer followed by a dense layer with a softmax activation function outputs the probability of belonging to each one of the three classes (well, moderately or poorly differentiated).

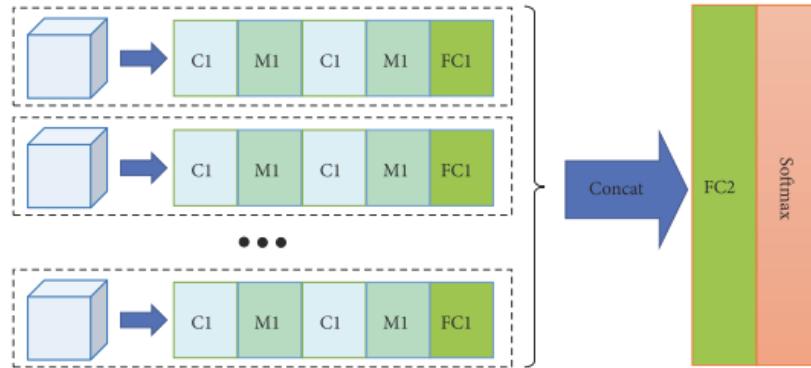


Figure 5.14: MCF-3DCNN architecture as detailed by ©Yang et al. [183]

During the training process, they implemented a label-shuffling method to overcome the problem of imbalanced data. Furthermore, to avoid the effect of overfitting, they trained their network with augmented data (original images were transposed, rotated, and flipped), a learning rate reduction and the addition of dropout (rate = 0.5). Using their architecture they were able to correctly classify the HCCs into the 3 differentiation groups with a mean accuracy of **74%**.

Their study however suffers from a lot of limitations such as the reduced size of the cohort, the imbalanced data and the fact that the analysis was only performed in a manually drawn VOI.

We have decided to tackle the same issue, but we implemented a fully automatic pipeline where both the segmentation and the grade prediction steps were performed by DL networks.

5.2.3 Prediction of the histological grade

In order to perform the prediction of the histological grade, our idea is to use the imaging features retained for the liver tissue segmentation, especially those used to segment tumoral structures. We believe that one easy way to extract relevant imaging features is to use those retained by the semantic segmentation networks, thus the better the accuracy regarding the semantic segmentation of the liver tissues, the more accurate the histological grade prediction will be. As explained previously, we proved that a cascaded architecture combined with the use of temporal contrast enhanced images allows a better delineation of liver tissues. We therefore proved that the designed cascaded architecture can be applied to provide missing annotations into weakly annotated datasets such as **TCIA-dB**. It has been proven that using temporal information can improve the accuracy of the grade prediction, by exploiting the wash-in wash-out specific features [308]. To conduct our DLR study, we first performed a multiphasic semantic segmentation of the **TCIA-dB**, before predicting the histological grade. After obtaining the final cascaded architecture, we built our network dedicated to the histological grade prediction.

5.2.3.1 Experiments and results

TCIA-dB contains images from 18 patients, where 9 were diagnosed with a grade 3 (G3), 7 with a grade 2 (G2) and 2 with a grade 1 (G1). In order to obtain a balanced training dataset, it has been decided to split them into two groups, the first containing G1 and G2 patients, and the second containing G3 patients, as it has been done previously in the literature since G2 was considered as being closer to G1 than to G3 [68,69]. Patients from the first group (G1 and G2) were considered as having a low grade (LG), whereas those from the second group have a high grade (HG). As explained previously, to train a network dedicated to predict the histological grade, we have decided to focus on what we called the relevant imaging semantic features. We therefore extracted the features from the second network of our cascaded architecture, to focus on the temporal behavior of the tumor.

The retained network (MPF-Tumor architecture as illustrated in the figure 5.5) is made of 2 classical U-Net networks, where each one takes either the AR or the PV image as input. We believe that the compressed information present in the bottleneck part of the network can be sufficient to encode the useful information present in the image (the U-Net will work as an auto-encoder for the semantic information). Therefore, we extracted for each patient of the **TCIA-dB**, this encoded information in a

slice-wise manner, represented by two $32 \times 32 \times 512$ features cubes (one per phase in the MPF-Tumor architecture) per slice as depicted in the figure 5.15.

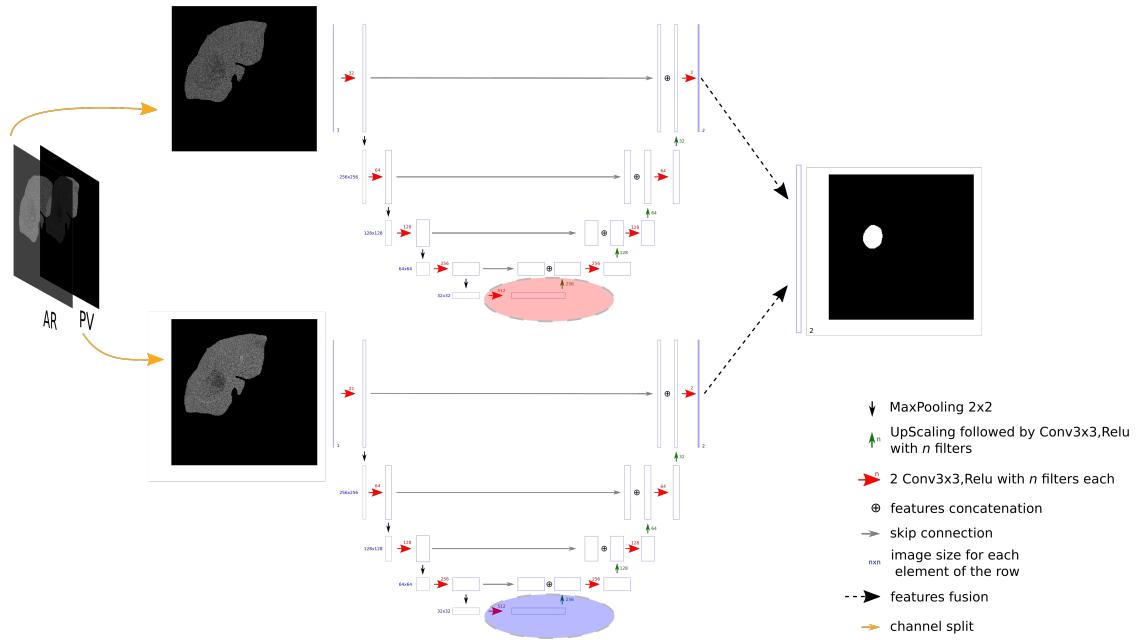


Figure 5.15: Red and blue areas correspond to the bottleneck part of the U-Net network where the features extraction is performed. Each image is then represented as a $32 \times 32 \times 512$ features cube.

Before applying the extraction of the features, we normalized the dimension of the different volumes of the dataset, so that each voxel measures $0.68 \times 0.68\text{mm}$ in the axial plane (because it corresponds to the resolution of the images used to train our semantic segmentation network), and that the volumes have a 2.5mm z-spacing (corresponding to the spacing of the majority of the PV volumes in **TCIA-dB**).

We finally built an architecture responsible for the grade prediction. We focused only on the centrally located tumor slices, since the histological grade corresponds to a measurement of the evolution of the disease, which tends to have more physiological effects at the center of the tumor. Centrally located slices will therefore exhibit the highest grade for a given patient.

Therefore, a slice-wise architecture was built, first because the features were computed in a slice-wise manner and second because the histological grade tends to be heterogeneous in the lesion, meaning that a slice-wise approach allows us to give a finer prediction to find potential areas with a more advanced disease.

The architecture depicted in the figure 5.16 works as a dimensionality reduction algorithm, where the first 1D convolutional layers are dedicated to reduce the number of features initially present ($32 \times 32 \times 512$). The dimensionality reduction step is performed for each phase separately, before the remaining features are combined (simple addition in the features space). A final dense layer takes the remaining features as input and computes the probability of belonging to each class thanks to a softmax activation function (LG vs HG).

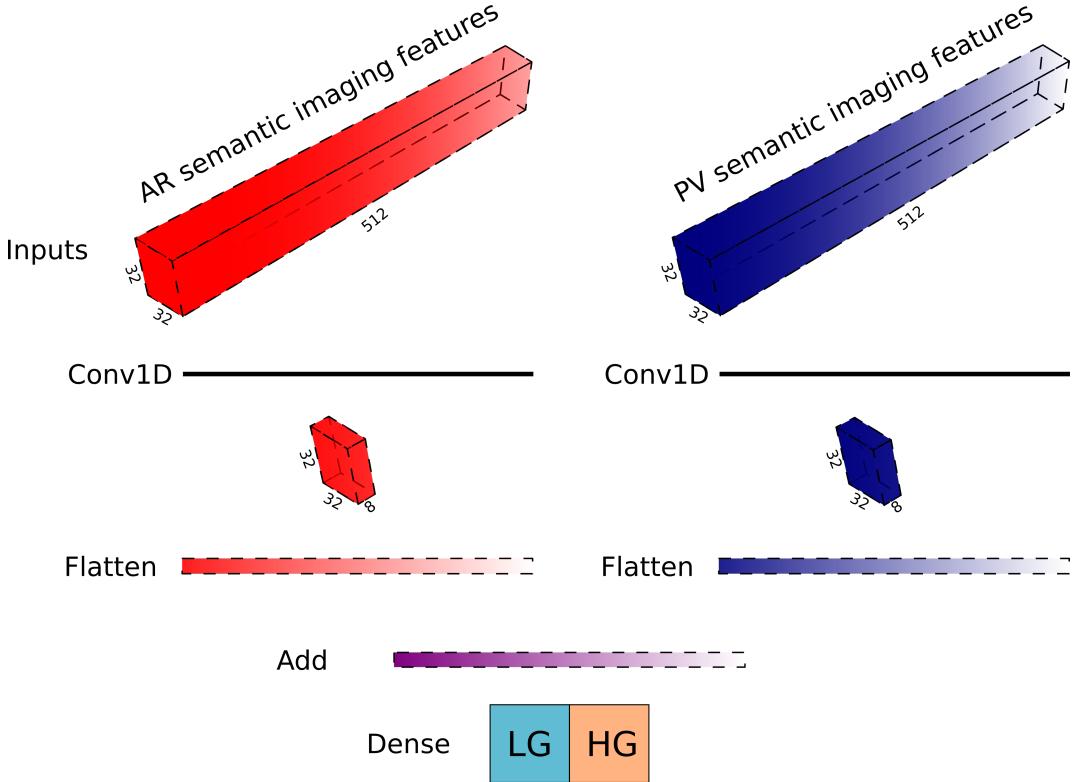


Figure 5.16: Slice-wise histological grade prediction using both AR and PV retained semantic imaging features

When training the network, we decided to consider that each centrally-located tumor slice of a given patient will have a higher probability of exhibiting the highest grade, which in this case corresponds to the observed patient-wise grade (ES1954 histological grading system [60]).

Knowing the composition of **TCIA-dB** (9 high grade patients vs 9 low grade ones), we performed a 9-fold CV training, so that each patient is at least present once in the testing set, and so that the training and the test sets contains both the same number of patients per class (7 patients from each class in the training set and 1 patient from each class in the testing set).

After testing several combinations for the hyperparameters, we fixed the number of retained features to 8 as depicted in the figure 5.16 (meaning that after the features dimensionality reduction, we obtained a $32 \times 32 \times 8$ cube per phase), and we considered a 2cm volume (corresponding to 8 centrally located slices with a 2.5mm spacing) when training/testing our architecture.

With our CV-training, we were able to correctly predict the patient-wise histological grade of 15 patients among 18, as detailed in the table 5.6 (a patient is considered as being correctly predicted when at least half of the retained slices were annotated with the correct GT class). When considering a slice-wise prediction, we were able to correctly predict ~74% of the slices.

It has been observed that the classification confidence was often correlated with the distance to the center of the tumor, where central tumor slices were correctly classified with a higher probability than

distant slices, which were either classified with a lower confidence, or even misclassified, as we can see in the figure 5.17. The quality of the semantic segmentation also often guided the accuracy of the prediction, as depicted in the figure 5.18.

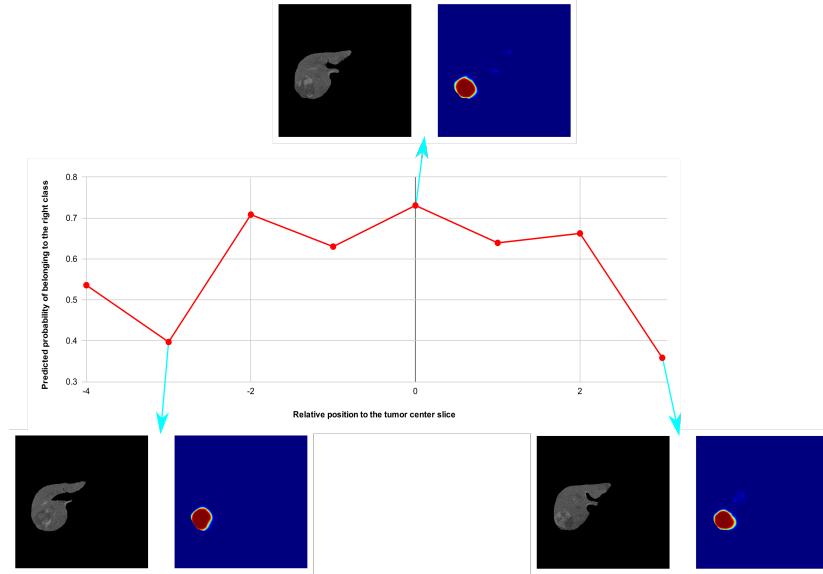


Figure 5.17: Example of slice-wise histological grade prediction of one patient from the TCIA-dB. Here we observed that the confidence was correlated with the relative position to the center. Even though the semantic segmentation of the tumor was correctly performed, distant slices tend to be either misclassified or classified with a lower confidence than central slices. Details of tumor semantic segmentation are given for 3 slices (raw image where pixels outside the predicted liver are masked in the left side, and tumor segmentation heatmap in the right side)

Table 5.6: Confusion matrix regarding the patient-wise histological grade prediction

		True grade		<i>Total</i>
		LG	HG	
Predicted grade	LG	7	1	8
	HG	2	8	10
		<i>Total</i>	9	18

Those results provided a more detailed prediction than the one consisting of a single patient-wise classification. Being able to compute the histological grade locally (here in a slice-wise fashion) allows us to visually focus on the heterogeneous regions that are crucial when needing to establish a diagnosis. Our pipeline can also provide a map of the best biopsy sites that will further be necessary in the clinical practice to either evaluate the progression of the disease or its prognosis.

We successfully proved that multiphase images incorporated in a cascaded architecture corresponds to the best combination when performing semantic segmentation of liver and its tumors. Moreover, our preliminary results proved that imaging features extracted from our cascaded multiphase segmentation

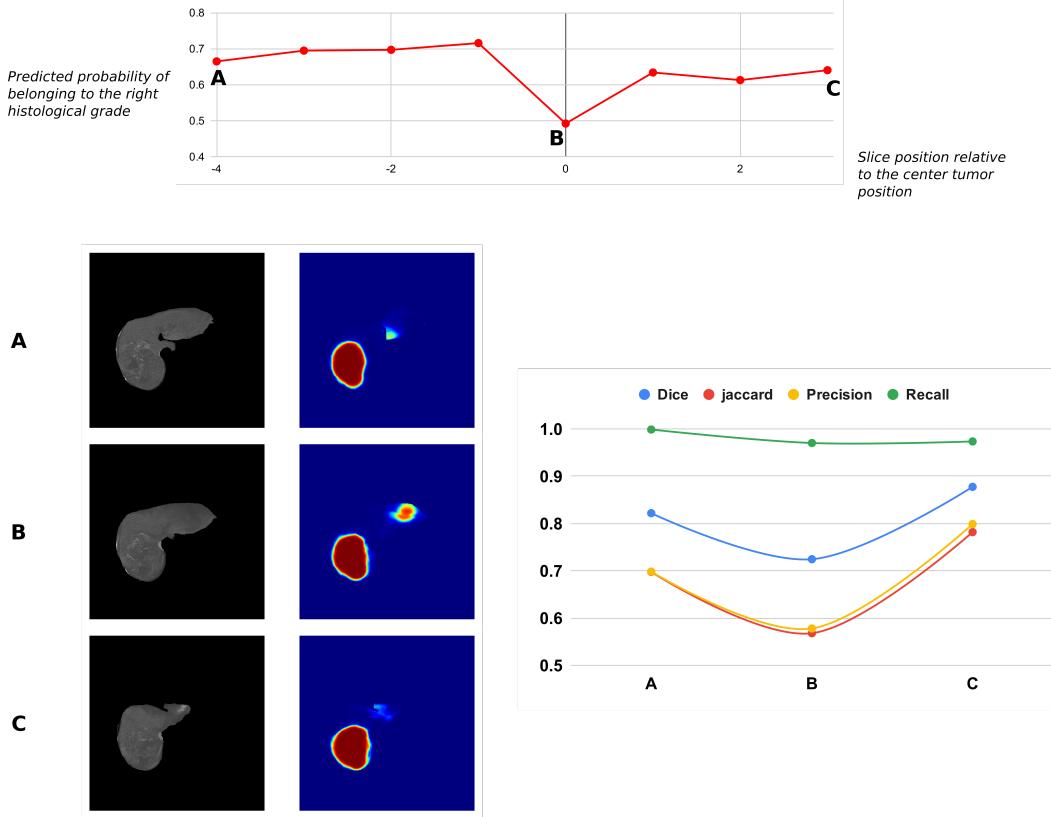


Figure 5.18: Example of slice-wise histological grade prediction of one patient from the TCIA-dB. Here we observed that the confidence regarding the grade prediction was correlated with the quality of the semantic segmentation. An accurate semantic segmentation is therefore a prerequisite to extract relevant features. Details of tumor semantic segmentation are given for 3 slices (raw image where pixels outside the predicted liver are masked in the left side, and tumor segmentation heatmap in the right side), two are correctly segmented and their histological grade was correctly predicted whereas one of them contained mis-segmented areas leading to a mis-classification of the histological grade.

architecture were relevant enough to be incorporated in a deep radiomics pipeline for the histological grade prediction. Our semantic segmentation results outperformed those reported on the same database using an ensemble classifier and requiring expert interactions during the segmentation process. Our histological grade prediction results were on par with the ones reported by the only study performing histological grade prediction but using a different dataset composed of MR images and using manually drawn regions of interest.

In the following section we present the different axes of improvement to improve the present work.

Chapter 6

Conclusion and Perspectives

6.1 Contributions

The purpose of the thesis was to propose solutions to better use medical images in order to provide a better characterization of the liver tumors. We successfully presented results proving the ability of deep learning to accurately delineate the liver and its tumors on dynamic CECT images. Moreover, we presented preliminary results proving the ability of automatically extracted deep multiphase imaging features to predict the histological grade of HCCs.

After describing the potential causes that can lead to the different types of liver cancer, we motivated our choice to mainly focus on hepatocellular carcinoma (HCC). This primary cancer type, which is the most frequent one, was described through its origin, the way it is usually diagnosed, classified and treated. The different steps transforming healthy hepatic cells to cancerous ones have been described, along with the analysis of the physiological changes often caused by this process. To better characterize the liver tumors, we have decided to rely on medical images solely, through a quantitative and objective analysis. Regarding the modality choice, we retained the computed tomography because it currently corresponds, with MRI, to the most widely used modality for a non-invasive assessment of the disease. CT was chosen over MRI mainly since it is usually the first choice for less-specialized centers, and because it can be more easily interpreted than MRI. With our medical studies review, we showed that the use of dynamic temporal images is a prerequisite to conduct an imaging based liver cancer study.

In the present work, we presented the two paradigms allowing the computer assisted analysis of medical images, either with engineered features, or thanks to deep-learning. One of the key aspects of medical images analysis being the segment, we compared the different semantic segmentation architectures, and evaluated the benefit brought by multiphase images which was most of the time neglected by previously existing studies. Then we extracted relevant features from our semantic segmentation architecture to tackle the prediction of the histological grade. Each step of our research

work required imaging databases precisely annotated by experts in association with researchers to improve their applicability.

In the clinical practice, images are usually analyzed by the experts with the naked eye, but the technological advancements allowed the creation of computer assisted diagnosis tools (CADs), where a few number of imaging features were initially used to differentiate benign and malignant lesions. A new technology called radiomics has been developed to compute a higher number of features but it took a long time before this technology was applied to the liver, especially because of the scarcity of publicly available datasets. We provided a detailed description of the radiomics pipeline, based on a manual segmentation of the ROI, followed by the extraction of a high number of engineered quantitative features, thus being called *HCR* (Hand-Crafted Radiomics).

We presented our review, where a total of 15 *HCR* studies performed on HCC patients have been analyzed. We evaluated them against the radiomics quality score (*RQS*) which has been developed to assess the robustness and the reproducibility of radiomics studies. We pointed out the lack of reproducibility of the studies, with a mean *RQS* of $8.73 +/- 5.57$ points out of a possible maximum value of 36 points.

Several important criteria were found as being ignored by the majority of the studies, such as a prospective design, the use of open-sourced data, the evaluation of the prediction on a validation dataset, or the extraction of features at multiple timepoints.

The emergence of deep learning has changed the way a lot of imaging related problems are comprehended. The radiomics field has been also impacted by this novel set of algorithms, and a new paradigm called *DLR* (Deep-Learning Radiomics) has been initiated, where one or several steps of the radiomics pipeline are performed by deep learning algorithms.

We believe that the key part of the radiomics pipeline lies in the segmentation of the ROI. This step has been found to suffer from a high inter- and intra-observer variability, thus having consequences for the accuracy of the final prediction.

We reviewed the different studies using deep learning architectures to perform automatic segmentation of the liver and its tumors. We extracted the common key settings shared by the majority of them, such as the cascaded architecture or the use of fully convolutional networks. However, we realized the lack of studies presenting results obtained from multiphase images, which has been a key element in our work.

We performed the automatic segmentation of an internal dataset composed of 104 sparse biphasic liver slices obtained from patients suffering from HCC (images available before the injection of contrast medium: Non-Enhanced CT, and at both arterial and portal venous phases). Considering how challenging the segmentation of liver tissues is, the amount of data, and the success of such an architecture,

we decided to train several specialized networks in a cascaded way.

When evaluating the performances of each specialized network, we validated the hypothesis that the use of multiphase information allows a better accuracy than single phased based networks for each task, with significant difference obtained for the segmentation of the liver (mean DSC of 89.9 ± 15.6 for the multiphase network *vs* 89.5 ± 13.2 for the best single phase network, $p = 0.005$) and the active part of the lesions (mean DSC of 75.5 ± 17.4 *vs* 71.6 ± 20.7 , $p = 0.001$).

Regarding single phase networks, the *PV* phase was the one allowing the most accurate segmentation, with significant difference vs *AR* and *NECT* for the segmentation of the parenchyma (mean DSC of 88.7 ± 15.4), the lesion (mean DSC of 87.8 ± 9.7), and both the necrotic (mean DSC 77.8 ± 12.4) and the active (mean DSC of 71.6 ± 20.7) part of the lesion.

We validated the hypothesis that several specialized networks combined in a cascaded architecture perform better than a single network addressing all the tasks simultaneously (obtained mean slice-wise DSC of 90.5 ± 13.2 for the parenchyma, 75.8 ± 15.1 for the necrosis and 59.6 ± 22.5 for the active part of the tumor when using the cascaded architecture with the liver GT mask as input).

In a fully automatic manner (without using the liver GT mask), we were able to reach promising results regarding the size of the dataset, with a mean slice-wise DSC of 78.3 ± 22.1 for the parenchyma, 50.6 ± 24.6 for the active tumor and 68.1 ± 23.2 for the necrotic part of the tumor.

We were also able to automatically compute the necrosis rate of the tumors, with a mean error of 15.9% when compared with the experts obtained rates. This prediction is accurate enough to consider the predicted necrosis rate when evaluating the treatment outcomes.

To overcome the limited size of multiphase datasets, we defined a strategy to “augment” the weakly annotated or non-annotated available datasets. With a robust registration algorithm and our cascaded architecture where specialized networks are trained on a sufficient amount of data, we were able to perform the semantic segmentation of liver and its tumors on unseen cases. We validated this assumption by performing the segmentation of TCIA tumors, and obtained a mean patient-wise DSC of 73.2 ± 20.6 .

To predict the histological grade, we proved that features learned by our multiphase semantic segmentation network are relevant to perform this task. We predicted the grade for central tumor slices, and after a CV training, we correctly predicted 74% of them. When considering the patient histological grade as being the most frequent one in the central tumor slices, we were able to correctly classify 15 patients over the 18 of the dataset. Regarding the small number of patients involved in the study, the results that we presented are only preliminary and further studies are required to confirm them.

As a conclusion, our research work is the first to propose a fully automatic *DLR* pipeline with multiphase images as input. Our strategy is dedicated to small databases where the entire annotations are often

not available but the same workflow can be applied for larger datasets. Our *DLR* architecture was dedicated, in this research work, to the prediction of the histological grade of HCC patients, but our strategy can be extended to other characteristics of HCC or other types of liver cancers.

Regarding the lack of publicly available datasets, we encourage the creation of an open 3D CECT dataset containing liver volumes of both healthy and diseased patients with precise and complete pathological information, and expert delineations of liver, hepatic vessels and lesions for each phase.

6.2 Perspectives

Several axes of improvement can be considered regarding our research work. We will discuss potential improvements to bring to the semantic segmentation of liver tumors, before raising issues regarding the current standards for dynamic CECT acquisition and their impact to future multiphasic CT-based radiomics studies. Finally we present ways to incorporate our work in other DLR-liver related studies, by encouraging the creation of open and precisely annotated CECT datasets.

6.2.1 Semantic Segmentation

As discussed before, the HCR paradigm often requires manual experts segmentations, and several studies already discussed the inter- and intra-observer variability present among the obtained delineations. This variability is translated in the quality of the retained features, which tend not to be robust enough when evaluating the predictive strength of the model.

The DLR paradigm allows one or multiple steps of the radiomics pipeline to be implemented with deep learning approaches. We decided to focus on the segmentation step and performed it automatically thanks to deep neural networks, thus reducing the number of potential biases. The automatic segmentation, as performed in our research work, however requires a high number of training cases to offer a realistic delineation of the targeted tissues.

In the case of liver tumors, a cascaded architecture stacking a first network responsible for the segmentation of the liver and a second dedicated only to detect the tumor within the obtained liver mask is currently the method allowing to obtain the most accurate results. State-of-the-art liver segmentation methods currently allow us to reach annotations similar to those obtained by the experts with mean DSC often above 0.95-0.96. We believe that there is no real need to improve the liver segmentation in the future, and that more interest needs to be shown in tumor and other liver tissues segmentation.

Regarding the segmentation of liver tumors, current state-of-the-art results were obtained thanks to post-processing steps such as false positives (FP) filtering method to reduce the number of objects misclassified as lesions. It might be that current semantic segmentation architectures still lack the

ability to discriminate between real lesions and other areas of the liver sharing the same textural properties such as regions close to the vessels as illustrated in the figure 6.1.



Figure 6.1: Tumor candidates marked with a dashed white line were classified as false positives, as described by **Chelbus et al.** [288].

This low recall might be explained by the poor quality of the currently available segmentation datasets. Liver annotations often contain non-hepatic areas (such as air most of the time) or imperfections close to the organ borders. When performing the automatic liver segmentation, the deep neural networks will most of the time be able to avoid these regions. However, in case they are integrated in the liver mask, they will often be misclassified as tumors by the second network in the cascade (responsible for the tumor segmentation).

To overcome this issue, publicly available datasets such as the LIST-dB need to be properly resegmented. They usually only contain annotations for the parenchyma and/or the tumors, but a third class (or more) could be incorporated, in order to consider tissues belonging to none of these groups such as the cysts or the blood vessels.

As explained in the section 4.4.1, current state-of-the-art tumor segmentation results are often obtained with 2D or 2.5D networks, but the post-processings steps usually help to improve the obtained accuracy (via 3D CRF or even simple morphological operations). Several studies implemented a U-Net like architecture, where the information is compressed before being decoding to obtain the final segmentation map. One of the key concepts in these architectures is the propagation of features learned in the earlier layers later in the network, either thanks to skip-connections, residual units or even densely connected layers. Another way to perform the segmentation could be by implementing a multi-scale pyramidal architecture where tumor-related features can be learned directly from the raw images at multiple scales instead of being learned from previously computed features, which is often the case in the aforementioned architectures.

It might also be interesting to perform the segmentation with a full 3D architecture. Some studies already investigated this solution [298], but they could not yet outperform state-of-the-art results. Some other concepts could be incorporated in future architectures. Jin et al. [309] for example proposed a network that integrated both U-Net and attention residual mechanism to proceed the segmentation of both the liver and the lesions. The residual attention mechanism has been introduced in 2017 by Wang et al. [310] to perform image classification, with the idea that the attention mechanism can help the network focusing on specific parts of the image. The study from Jin et al. was the first to use the attention mechanism for semantic segmentation purposes. On the hidden test set of LITS, they outperformed a lot of 2D-based methods, but were still far from the top-ranked teams. However, new paradigms such as the self-attention mechanism, in combination with state-of-the-art 2D and 3D architectures are certainly an avenue for the improvement of the automatic liver and tumors segmentation tasks [311].

6.2.2 Dynamic contrast-enhanced images

In our research work, the key element is the extraction of features from dynamic contrast-enhanced images. However, it has been very difficult to collect dynamic CECT based images, since only a few publicly available datasets contain this type of volume.

There is a huge room for improvement regarding this specific type of dataset. It has been mentioned in the chapter 2, that the use of images obtained after the injection of contrast medium largely improves the visual and the automatic quality of the diagnosis. Nonetheless, we realized that multiphasic images could be better used in either the semantic segmentation and/or the radiomics field.

The first issue is related to the acquisition protocol required to obtain this type of images. Most of the time, the different retained phases, namely arterial, portal venous or delayed phases are acquired following either visual inspection of the radiologists¹, or after a specific duration following the injection moment that can differ from one study to the other. Another criteria to consider is the physiological characteristics of the patient such as its weight that will determine how the contrast medium is diffused through the liver.

When dealing with multiphase databases, we believe that these key elements need to be considered, at least the moment of injection and/or the duration between the injection and the acquisition. Currently, only the DICOM format allows the addition of such metadata (it is only possible to record the exact acquisition moment but not the injection moment) but they are barely reported and are often left to default values.

Obtaining such data seems really challenging, hence, some techniques need to be implemented to deal

¹ The arterial phase is sometimes acquired with the help of a bolus tracking technique

with current multiphase images. For example, we believe that a “phase correction system”, responsible for the normalization of images belonging to the same phase, could be implemented . The automatic detection of the amount of contrast medium still in the liver could also help to predict the exact duration between the injection of the contrast medium and the acquisition.

In our research work, we have decided to implement an histogram specification algorithm to “normalize” images from a given phase. The first step requires to compute a mean histogram per phase (which requires to have a sufficient amount of CT images correctly labeled), before transforming the volumes of a given patient to match the obtained mean histogram, as depicted in the figure 6.2.

In the case of multiphase images, the expert performing the segmentation will usually try to obtain only one segmentation volume shared by all the registered volumes. This representation might be coherent for the liver mask, but we believe that it needs to be modified for internal liver tissues. The lesions, for example, will have a different behavior from one phase to the other (hypo- or hyper-dense tumors for example), consequently, one segmentation per phase should be performed. This might potentially help the deep neural networks to better understand the dynamic of the lesions, and also address the problem of small mis-registration when only one ground truth segmentation map is shared by all the available phases images.

One long-term objective could be to create, with the help of radiological experts, a dataset containing registered multiphase CT volumes with expert annotations performed following the aforementioned protocol.

6.2.3 Deep Radiomics

Regarding our DLR study, we believe that the prediction of the histological grade can be improved. As an example we advocate for a grade prediction at a finer scale, but it will require to know the exact position of the extracted sample, or to obtain a map of the heterogeneous regions after surgical removal of liver tissues through liver resection or liver transplant for example.

In our research work we investigated a patch-wise prediction of the histological grade but the results were not as promising as those obtained through our slice-wise approach, especially because our patch-wise semantic segmentation network needs to be improved.

We are considering the relevant imaging features as being the ones extracted from the bottleneck part of our U-Net network but other types of auto-encoders could be investigated to better extract the relevant information from the raw images, or from the retained features. We could have also incorporate clinical data (AFP (Alpha-FetoProtein) levels, age, ...) in our pipeline, however such data are often difficult to retrieve, and challenging to combine with imaging features, consequently, we decided to focus only on images in the current work.

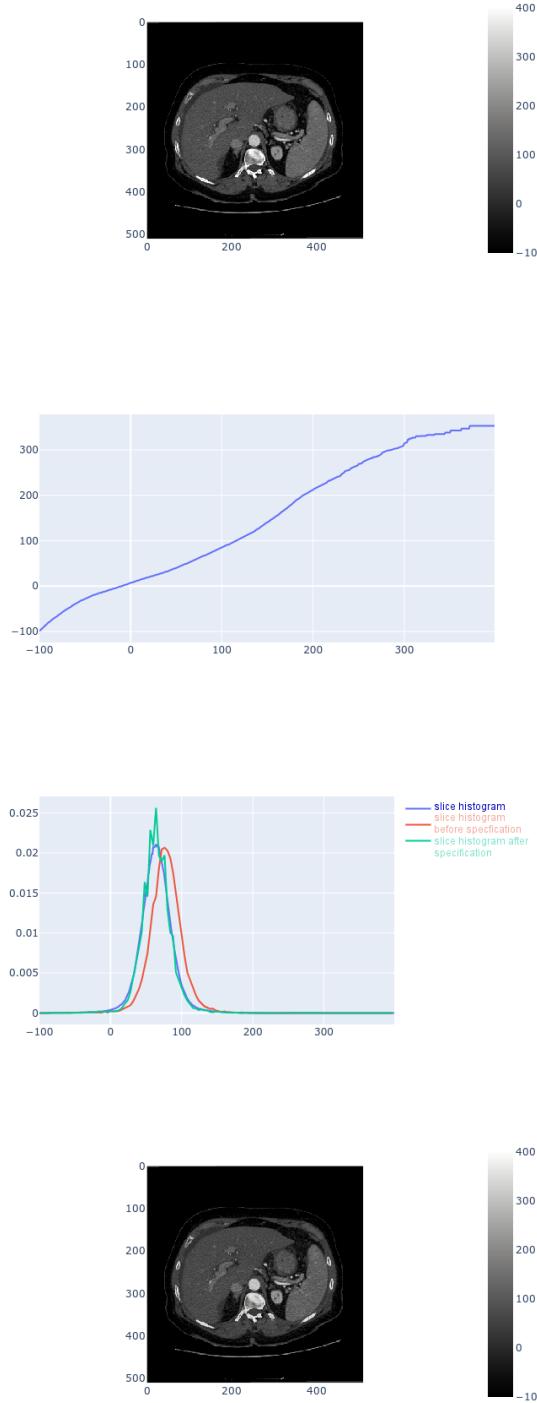


Figure 6.2: Histogram specification experiment conducted to normalize images from the same phase, top row: slice before histogram specification; second row: mapping function for the chosen range $[-100, 400]$; third row: representation of the mean phase histogram, the current slice histogram and the transformed one after application of histogram specification; bottom row: slice after histogram specification.

It is worth also noting that the current registration steps still suffers from limitation essentially because it performed using the predicted liver masks as registrations masks. However, the automatic liver delineation is not perfect, and therefore this can result in slightly misregistered volumes. Therefore, we

advocate an incorporation of an automatic deep learning based registration step in our DLR pipeline, or a workflow requiring no registration at all (for example by considering features learned in each phase separately and combining them for the grade prediction).

If more interest is shown in the future towards the prediction histological grade, we believe that the existing grading systems have to be standardized by considering more criteria than just the worst or the most frequent grade present in the histological slices, which is currently the case.

Our DLR architecture obtained promising preliminary results for the prediction of the histological grade, but we believe that the same architecture can be adapted to focus on other tasks, such as the prediction of recurrence after treatment using longitudinal studies.

Appendix A

Detailed analysis of the DL semantic segmentation reviewed articles

A.1 First FCN papers

Li et al. (2015)

The aim of the study was to perform tumor segmentation using a classification model with patches as input and convolutional layers introduced early in the network to recognize low-level features. A patch was considered as positive if at least 50% of its pixels are tumoral. Five different patch sizes were tested to train a CNN dedicated to perform a binary classification between tumoral and non-tumoral labels. Using cross-validation on a DB containing 26 CECT volumes, it was noticed that 17x17 patches were the one that reached the best accuracy. The CNN achieved slightly better performance than classical ML techniques built on hand-crafted features (AdaBoost, RF and SVM).

Ben-Cohen et al. (2016)

Ben-Cohen et al. (2016), used the famous VGG-16 architecture to construct his own network. The dense layers were converted into convolutional layers, especially the output classification layer that was removed and replaced by a 1x1 convolutional layer with 2 channels for a pixel-wise classification. The remaining layers were distributed in a FCN-8s-like architecture allowing to combine low-layer information with high-layer information. They also decided to provide not only one slice as input, but 3 successive slices (namely *2.5D input*) to help the network retrieve the 3D information. The output of the network corresponding to the segmentation of the middle slice. In order for the 3D information to be consistent from one patient to another, they interpolated the slices in the z-axis to have a fixed 1mm-spacing.

Data augmentation was used to generate natural-like images in order to artificially increase the size of the database. In their case, 4 different synthetic images were generated from one raw image by applying a scaling in the range [0.8, 1.2].

Two different databases were used to train their model, the SLIVER07 dataset was used to train the liver segmentation part, and an internal database of 20 patients, with only 68 liver-segmented slices, and 43 lesions-segmented slices, was used to train the second part of the architecture.

To evaluate their model, it was decided to consider the entire segmentation for the first stage by computing the Dice, whereas an object-based detection metric was implemented for the second stage, by analyzing the false positive rate per case (FPC).

The liver segmentation performance was evaluated on the 68 liver-annotated slices of their internal dataset, and they obtained a dice of 0.89 (note that only a slight improvement was obtained thanks to the 2.5D input: 0.88 vs 0.89).

The accuracy of the lesion detection model was compared to both a patch-based approach where 17x17 patches were sampled within the liver, and a sparsity-based model. The FCN achieved better results than the other two methods reaching a FPC of 0.74.

When using the entire cascaded architecture, they realized that the automatic liver segmentation often doesn't include darker borders, and that those pixels could further be classified as FP, so the fully automatic segmentation leads to an improvement in FPC.

They also conducted some synthetic experiments and realized that the network mostly relies on the gray level differences.

Christ et al. (2016)

Christ et al., provided a cascaded architecture for the segmentation of both the liver and the potential lesions it contains. Their motivations were similar since they believed that the network in the first step will learn filters specific to the detection and the segmentation of the liver in an abdominal axial CT image, while considering that the liver ROI will help reduce the number of FP for the lesion segmentation task. The U-Net architecture introduced by **Ronneberger et al** which was used as a basis, was chosen for its ability to combine low and high-level features through so-called *skip-connections*. In order to prove their hypothesis, they worked with the 3DIRCADb dataset containing 20 patients, where only the 15 patients presenting hepatic tumors were retained.

The images were pre-processed by first clipping the HU values in the range [-100, 400], followed by a histogram equalisation. Data augmentation was also used to increase the size of the database via rotations, flips and addition of gaussian noise, so they were able to train their models with more than 20k 2-D slices. In order not to train their networks from scratch, they used the pre-trained UNet

from Ronneberger et al. trained on cell image data. A weighted loss function was implemented to combat the class imbalance during training, and they applied a 3D conditional random field (CRF) as a post-processing step on the obtained probability maps, to refine the results.

They however noticed that finding the appropriate weights for the CRF remains an open problem. They obtained a Dice of 0.93, increased at 0.94 after applying the 3D CRF, for the liver segmentation, and a Dice of 0.56 for the lesions. They applied the same cross-validation learning on an internal database of 100 CT scans (Train: 60, Val: 20, Test: 20) and they reached a mean dice of **0.91** for the liver and **0.61** for the lesions.

Dou et al. (2016)

In 2016, the first fully 3D neural network dedicated to the segmentation of the liver was developed by **Dou et al.** Their 3D CNN network was built with large kernels, and had intermediate supervised layers to combat the gradient vanishing problem. In order to implement deeply-supervised mechanism, 3D Deconvolutional layers were built, to bridge coarse feature volumes, to dense probability ones. The SLIVER07 database containing 30 patients was evaluated, where 20 patients were used to train the networks, and 10 to test it. The global loss of the network incorporated an intermediate loss at each of those stages, and it was noticed that this mechanism allowed a better convergence than a classical 3D-CNN. A fully connected CRF model was also applied in the transverse plane to refine the obtained volume. The results obtained in a cross-validation manner confirm the gain brought by the deeply-supervised mechanism. On the test set, the results were on par with state-of-the-art “classical” methods (level-set, shape-based model...), but with a faster processing time.

Rafiei et al. (2018)

3D convolutional layers were also used by **Rafiei et al (2018)** to perform the liver segmentation task. Their architecture incorporated an encoding part where 3D layers only were used, whereas the decoding part consisted only of 2D layers. At each iteration, a block of 512x512x38 slices was fed to the network, and the decoding part was mainly intended to capture the 3D shape of the liver. The connection between the encoding part and the decoding part is done by custom skip connections, which first select the center slice in the 3D volume, and then crop it to meet the required size for the concatenation. To combat the classical problem of segmentation errors at the edges of the liver, they used a spatial-weighted-cross-entropy loss function, where pixels close to the liver border will be assigned a higher weight than those in the center. To help the generalization abilities of the network, a dropout layer was added just after the encoding part. Once the segmentation performed, they applied a fully connected CRF model only on the border pixels especially to reduce computation time.

Training and evaluation was completed on the MICCAI 2015 database containing 30 patients. Data augmentation was utilized to combat the lack of data, and 7 images were generated from one original image by rotation in the range [-30, 30]°.

They slightly improved the results obtained by **Heinrich et al** [4], who used a multi-altas patch-based method (92.95 vs 93.52 on the average Dice), but with a high reduction of inference time (~1000s vs ~60s)

Sun et al. (2017)

Sun et al., were the first in 2017 to use the multiphase information in a deep CNN for the lesion segmentation task by developing a so-called multi-channel fully convolutional networks (MC-FCN). To implement the architecture, they transformed the AlexNet to a FCN by converting fully connected layers into convolutional layers. They started by training their architecture with images from the different phases separately, then they realized that the weights were very different from one phase to another. So they built the MC-FCN so it can take 512x512x3 volumes as input, where each channel corresponds to a different phase. The 3 channels were separated early in the network by a so-called *slice layer*, before feature extraction was performed for each phase using the FCN. Finally, a feature fusion layer is employed to merge the information from the different phases, and the output is cropped to 512x512x2.

They tried to perform the segmentation of the lesions within the liver, but the network failed to converge, certainly due to the lack of sufficient training data, even when using a large number of iterations. To solve this issue, they first pre-trained the network to segment the liver in the image before fine-tuning it for the lesion segmentation.

They finally proved by the obtained results, an improvement of the segmentation accuracy when using the multiphase information.

Frid-Adar et al. (2017)

Frid-Adar et al. implemented a patch-based liver lesions detection method. Contrary to **Li et al. (2015)**, their main hypothesis was that non-lesion patches are different if they are located in the liver interior or in the liver boundary. They proposed to classify the patches into lesion and non-lesion categories, where lesion patches include lesion-boundary patches, and non-lesion patches contain both normal-interior patches and normal-boundary patches. An additional step was added to reduce the number of FP, with a binary classification CNN trained only on lesion patches and normal-boundary patches.

They also decided to use two different scales to capture fine (20x20) and global spatial information

(50x50), before concatenating the obtained features from both sides at the end of the network. Patches were randomly sampled to respect class-balancing, and data augmentation was applied via random flips and rotations [5°, 130°, 300°]. Intensity was shifted so they all share the same mean intensity. 140 000 patches per class were used for the training, and 2 initialisation schemes were tested: a training from scratch, and a fine-tuning after training the network on the Cifar-10 dataset. The network was tested on a database containing 132 CT scans, with a total of 498 metastases. Images were first resampled to get a 0.71-fixed spacing, which corresponded to the finest available spacing in the dataset. They confirmed the improvement brought by the separation of the non-lesion patches into two different categories (TPR of 85.9 vs 80 when using a strict binary classification). They also found that fine-tuning from Cifar-10 was not as successful as training from scratch (82.8 vs 85.9 on the TPR) .

A.2 LITS papers

In 2017, the Liver Tumor Segmentation (LITS) benchmark was created, where the goal was to have a standard database with a sufficient amount of humanly annotated data to assess the quality of a given model to perform one or multiple tasks, among which, the liver segmentation, the tumor segmentation and detection, and the tumor burden estimation.

The dataset contains a total of 201 CT scans, acquired in 7 different hospitals, and reviewed by 3 independent radiologists. The visible (131 volumes) and the hidden part (70 volumes) were separated so that each institution has the same proportion on both sides.

A high heterogeneity is present in the database, where from 0 up to 75 tumors are present per volume, and an average tumor-liver HU difference of 31.94 HU.

Two challenges have been launched by ISBI and MICCAI, aiming to segment this database. Some of the top-ranked studies that participated have been reviewed since they shared some common specifications.

Li et al. (2017)

Li et al. (2017) combined both 2D and 3D information to get the final classification. They first trained a Residual Network to get a coarse liver segmentation. Then, a 2.5D DenseUNet (densely connected path with UNet-like connections) architecture was developed to get a 64-channels features map corresponding to the 2D features. A pixel-wise classification step was added to get a pixel-wise 2D classification. The binary maps were stacked to a 12-slices 3D block, and combined as input with the original 3D volume to obtain a 64-channels features map encoding the 3D inter-slice information.

The 64-channels feature maps encoding both the 2D and the 3D information were merged so that the last layer of their architecture could extract the so-called *hybrid features*, before a pixel-wise probability

is generated.

For the coarse segmentation step, they decided to resample the images so they have a fixed spacing of $0.69 \times 0.69 \times 1.0 \text{ mm}^3$, but for lesion segmentation, the network was trained with images at their original resolution.

From their experiments, they realized that the DenseUNet performed better than simple DenseNet, confirming the importance of long skip-connections. Interestingly, the 2D version of this DenseUNet allowed a better learning behaviour, when compared to the 3D version, and this hypothesis was confirmed by the prediction results (improvement of 0.89 points for the Dice per case, and 0.3 points for the Dice global). They also proved that the hybrid features help the network improve the obtained results.

Using the final architecture, they got a mean dice per case of **0.72** and a dice global of **0.82**.

Han et al. (2017)

Han et al. (2017) decided to combine the residual network and the UNet, by building a resnet-like network with UNet skip connections between encoding and decoding parts.

They also provided 5 stacked slices as input in a 2.5D fashion, so the network can rely on the information provided by adjacent slices.

The first step of their model consisted of getting a coarse liver segmentation, before being refined in the second step which also segmented the lesions it may contain.

A weighted cross-entropy loss function with empirically chosen weights was implemented to train the network. The entire CT slices resampled to a fixed resolution of $1 \times 1 \times 2.5 \text{ mm}^3$ were used to train the liver segmentation, whereas only liver slices with their original resolution were utilized for the lesions segmentation step.

During inference, the input volume is resampled to $1 \times 1 \times 2.5 \text{ mm}^3$, then the liver segmentation is performed before getting the largest 3D connected component. The original resolution is then used to perform the lesion segmentation in order to get finer details from the liver and its lesions. A post-processing step was employed so that connected components considered as lesions will be removed if the maximal probability value is less than a given threshold.

The entire process allowed them to reach a mean dice per case of **0.67**.

Yuan et al. (2017)

Yuan et al. (2017) proposed a hierarchical convolution-deconvolution neural network to carry out the segmentation of both the liver and the lesions.

The first step of their architecture consisted of getting a coarse liver segmentation on the entire 3D

volume, before the second step refined the probability map, and the final one executed the lesions segmentation.

The same architecture was used for the three different steps, where 2.5D input was chosen, and training was performed with the Jaccard distance as loss function. This function takes advantage from the fact that it maximizes directly the performance that is computed from the output of the network. The architecture is composed of a so-called *convolutional path* where the information is compressed, and a *deconvolutional path*, where the full image resolution is recovered. Dropout layers were also used at the end of the convolution path, and the other just before the last deconvolution layer.

During the first step of their cascade, images were down-scaled to 128x128 and the volumes were resampled with a slice thickness of 3mm. The training process of this specific step was made only from liver slices plus 5 slices above and below it. For the inference, the segmentation was performed on each slice, and pixels were considered as being part of the liver after applying a thresholding of 0.5 on the probability map, and after extracting the largest connected component.

To train the second step, which consisted of the refinement of the previously obtained liver area, the original volume was firstly resampled to get a fixed slice-thickness of 2 mm. The liver 3D bounding box was extracted from the annotated liver, with 10 more voxels in the 3 directions, and the axial dimension was adjusted to 256x256 pixels.

The inference was achieved by extracting the slices from the predicted liver area, and applying the prediction with the second network. The same post-processing step was implemented to conserve the 3D consistency.

In order to train the last network, the liver VOI was extracted as previously, but by keeping the original resolution to avoid losing small details. The training was completed using only lesion slices, and an additional image was added beside the original image, where intensity difference between lesion and liver voxels was enhanced via 3D regional histogram equalization.

During the inference, all activations outside the predicted liver were removed.

Their method achieved average DSCs of **0.96** for liver segmentation (Dice global of 0.967), **0.66** for tumor segmentation (Dice global of 0.82), and a root mean square error (RMSE) of **0.02** for tumor burden estimation.

Bellver et al. (2017)

The cascade paradigm was also the key element in the study of **Bellver et al. (2017)**.

The main architecture they used was the DRIU (deep retinal image understanding), consisting of VGG16 backbone with last Dense Layers replaced by a convolutional layer. The network was pre-trained on the ImageNet dataset, before fine-tuned with CT images having intensity clipped to the range

[-150, 250] and further normalized. To train the network, the weighted binary cross entropy (BCE) was implemented in order to deal with class balancing. When training the network responsible for the segmentation of the lesion, they decided to suppress the loss outside the liver area, and considered only pixels inside the liver region when computing the weights of their loss function. Each one of the networks were trained with 2.5D input to exploit the 3D information. After executing both the liver and lesions segmentation, they developed a lesion detector to reduce the number of FP. A patch-based approach was used for the training of the detector, with patches of 80x80 pixels sampled through the liver (having at least 25% overlap with the liver), and classified as positive if at least 50 pixels of the region belong to a lesion. The detector was trained using a ResNet-50-like architecture, pre-trained again on ImageNet, and that was customized by replacing the last classification layer by a binary-classification layer (healthy/non-healthy).

A post-processing invoking a fully-connected 3D CRF model was further called to refine the obtained regions.

Their results proved that the detector helps the network to reduce the number of FP, and also confirmed that learning to segment the lesions only from the liver area helps the network.

With their method, they obtained a Dice per case of **0.59** on the test set of LITS.

Chlebus et al. (2018)

Chlebus et al. (2018) submitted their results to both ISBI and MICCAI challenges while targeting only the lesion segmentation. In their study they not only described the methods used to generate their segmentation, but they challenged the LITS dataset by comparing it with manual annotations from another expert.

3 different U-Net-like architectures (axial, sagittal, coronal) on 4 resolution levels with additional short skip connections dedicated to the segmentation of tumors only were trained on the LITS database. The loss function considered only the voxels belonging to the liver. In order to reduce the number of FP, an additional step was added where hand-crafted features were computed on the 3D objects obtained after the segmentation step. In order to assess the performance of the neural network, a human (MTRA: medical-technical radiology assistant) was asked to segment the LITS dataset. The dice per case between MTRA and LITS was 0.70. The neural network reached a DSC of 0.51 on the LITS (4.6 FP/case). After using the classifier, the Dice was improved to 0.58, and the number of FP was reduced to 0.7/case. They finally reached a dice per case of **0.68** for the lesion and **0.96** for the liver.

Among the several hand-crafted features, the most relevant one for detecting the FP were shape-based features as long the distance from the object to the boundary.

Kaluva et al. (2018)

Kaluva et al. (2018) proposed to build their model with the DenseNet architecture as a basis. The DenseNet is constructed such that a given layer l is connected to each of its predecessors $l-1, l-2, \dots, 0$ in a concatenated manner. The number of feature maps at a given stage is known as *growth rate*. The number of feature maps in the classical DenseNet grow linearly with the depth. After each dense block, a Transition Down (TD) layer is introduced to reduce the spatial dimensionality of the feature maps. Their cascaded model consisted of 2 steps, one specified on the liver segmentation, and the second one on the lesions segmentation. To train the liver segmentation network, the original CT images were downsampled to 256x256 pixels before normalizing HU intensities clamped to the range [-100, 300]. No downsampling was performed when training the lesions segmentation network, and comparing to the first network, 3 differents HU windows ([0, 100], [-100, 200] and [-100, 400]) were applied on the original image and combined to get a 3-channels input. To deal with class imbalancing, the first network was trained only on liver slices (with 10 slices above and below it), whereas the second one was trained only on lesion slices (plus 5 slices above and below). They also decided to connect the input and the output of each Dense block in the downsampling part, in a ResNet-like fashion.

Their first network was trained thanks to a spatial-weighted-cross-entropy as loss function with a higher loss at the edges of the liver, relative to the interior. The same loss function was used for the second network, with a higher weight given to the tumor region, and was combined with the dice loss.

As post-processing, they simply decided to mask all tumor activations that were located outside the predicted liver.

On the 70 volumes of the hidden set (LIST) the average Dice for the liver was **0.91** and **0.72** for the lesions.

Bi et al. (2017)

Bi et al. (2017) also developed a cascaded architecture where a first ResNet segment both the liver and the lesion, and the second one combining the original image with the two obtained probability maps to produce the final segmentation map.

Images were preprocessed by first clipping the intensities to the range [-160, 240], before applying a min-max normalization. To train their networks, they used a total of more than 8,000 randomized slices, where half of them presented both liver and lesions, and the other half where the liver was not present at all.

The original network was pre-trained on the ImageNet dataset for 100 epochs, before being fine-tuned with augmented data from LITS (random scaling, crops and flips).

During inference, a multiscale-strategy was used by resizing the original image to different scales from 512x512 to 640x640 with a step of 32 at each stage, and averaging the obtained outputs.

The use of the multiscale-strategy slightly improves the results obtained using only their cascaded ResNet. A slight improvement was also notable between their cascaded version and the classical ResNet. They finally obtained a Dice of **0.64** for the segmentation of the lesions.

Vorontsov et al. (2018)

Voronstov et al. (2018) decided also to use a cascaded architecture where two U-Net-like networks were combined, but they decided to build a model allowing an end-to-end training of the entire cascade. The first network took as input an axial slice and predicted the probability of each pixel to belong to the liver, and the second one took as input both the original slice and the output of the first network. The latter implemented short skip connections in a resnet-fashion to propagate the liver localisation information to all the layers.

Images were pre-processed by dividing the input intensities to 255, and clipping the remaining values to the range [-2, 2], and only liver slices were used to train the network with the dice loss as loss function. Data augmentation was applied via flips, rotations, zoom and elastic deformations. The network was pre-trained on the down-sampled images (256x256) then fine-tuned on the full resolution slices.

To improve the segmentation, a classifier was added at the end of each network, aiming to introduce a so-called *cross-slice context*. The segmentation was performed on 3 adjacent slices, and the output features were combined by a classifier to predict the middle-slice segmentation. During inference, 3 different networks were associated via a voting strategy, where each one provided a prediction corresponding to the average of 4 oriented images after flipping the original one. The final liver segmentation corresponded to the largest connected component, and lesions were kept when belonging to a dilated version of the predicted liver. This prevents missing lesions when the liver is slightly under-segmented.

They finally obtained a Dice per case of **0.66** for the lesions and of **0.95** for the liver.

Bilic et al. (2019) LITS Review

The majority of the papers that were submitted for these 2 challenges share some common properties, and they were reviewed by the organizers themselves in **Bilic et al. (2019)**.

They started with a reminder about the composition of the database, where the 201 total CT scans were acquired in 7 different hospitals, and reviewed by 3 independent radiologists. The visible (131 volumes) and the hidden part (70 volumes) were separated so that each institution has the same proportion on both sides.

A high heterogeneity is present in the database, where from 0 up to 75 tumors are present per volume, and an average tumor-liver HU difference of 31.94 HU.

Regarding the submitted methods, a U-Net-like network was often chosen as a basis, with modifications by incorporating residual connections, or adjusting the input resolution. No direct 3D methods on the original images, but some methods used 3D convolutions for the lesions segmentation task, with smaller input resolution images. Some other techniques used 2.5 input.

Images were pre-processed by clipping the HU values, but without a clear consensus regarding the range to apply. The images are then often normalized and resampled to a specific resolution. The size of the dataset is often increased after the application of data augmentation, which corresponds most of the time to standard geometrical transformations.

The most commonly used loss functions were the weighted cross entropy loss function, having the advantage of solving the class imbalance problem. The Dice loss and the Jaccard loss were also used and are useful because they allow the network updates its weights to reach the best scores regarding the final used metric.

After performing the segmentations, the vast majority of the teams used post-processing techniques to refine their results. Among these techniques, they reported the extraction of connected components, the removal of lesions activation outside of the liver, the fill of erroneous holes in the lesions shape via morphological operations, the application of connected CRF, or the use of a classifier (such as Random Forest) to detect the FP.

Ensemble learning was also applied by the top-ranked methods, either via axis-related networks, the combination of different rescaled input data, or the use of different networks trained in a CV-manner. Regarding the results, they realized that the best tumor-segmentation results were obtained for large lesions, and within a specific lesion-liver HU difference interval (a difference between -10 and -60).

The top-ranked methods used some 3D approaches in their architecture (**Chlebus et al, Li et al**), showing perspectives to capture the whole volume context in the future.

The introduction of such a standard database led to a rise in the interest towards automatic liver and lesions segmentation. The organizers are planning to relaunch the same type of competition in the future, by particularly providing multiple ground truth (instead of only 1 for LITS) and potentially splitting the lesion segmentation task into large and small lesions, since current methods still struggle in segmenting the small lesions.

A.3 New hybrid techniques

Jin et al. (2018)

Jin et al. (2018) proposed a network that integrated both U-Net and attention residual mechanism to proceed the segmentation of both the liver and the lesions.

The residual attention mechanism has been introduced in 2017 by Wang et al. [310] to perform image classification, with the idea that attention mechanism can help the network focusing on specific parts of the image.

The study from Jin et al. was the first to use the attention mechanism for semantic segmentation purpose.

Their model contained 3 stages, the first one to get a coarse liver segmentation from 2D images, the second one to refine the obtained volume of interest using a 3D architecture, and the last one to segment tumors from the precise liver VOI.

Each network was built according to the U-Net shape but using residual blocks as key components. The combination of both residual blocks and attention mechanism allows the network to learn on deep architectures, but with a focus on localizations that are relevant for the requested task. The images were pre-processed by first clipping the HU intensities to the range [-100, 200], before applying a zero-mean and a min-max normalization. The network was then trained using the dice loss as loss function. In the first stage, a set of 2/3 liver images and 1/3 randomly picked slices were down-sampled to 256x256 pixels to train the network. The coarse liver segmentation was obtained through the extraction of the largest connected component.

Large 3D patches of 224x224x32 slices were extracted from both the lesions and non-lesions area and were provided to train the second network.

Finally, for the lesions segmentation task, they tested 3 different patch-sizes (32x32x32, 64x64x32 and 128x128x32) before realizing that the larger the patch was, the richer the extracted information will be. The patches were randomly sampled within the liver.

Since the model was trained on 5 CV-folds, they implemented a voting strategy from the five resulting networks to get the final segmentation.

They obtained a Dice per case of **0.59** for the lesion segmentation task on the hidden set of LITS, outperforming a lot of 2D-based methods, but still far from the top-ranked teams.

Appendix B

Hyperparameters selection

In this chapter, we reported the experiments that we run on the 3DIrcad-dB to find the best sets of hyperparameters for the tumor segmentation networks. Results were reported here to allow better readability in of our complete semantic segmentation work (see chapter 5.1.6.2)

Here is the list of hyperparameters to set in our case:

- Learning rate
- Decay
- Number of epochs
- Depth of the network, modeled in our case by the number of filters used in the bottleneck of our U-Net architecture
- Initial image size
- Type and amount of data augmentation
- Dropout
- Input type, either single slice or 2.5D

Several techniques exist for the determination of the best set of hyperparameters, such as the grid search or the bayesian hyperparameters optimization. However given the number of hyperparameters and the time required to perform one training¹, it would have been too costly to implement those techniques that better suits small architectures and less complex problems, so we have decided to follow an heuristic approach to find the best set of parameters.

We first tried to set the best pairs of learning rate and decay, by freezing the rest of the parameters.

¹ Evaluating one set of hyperparameters can take up to 48h using a single our networks implemented using Keras with a Tensorflow backend and trained on a single NVIDIA GTX 1070 graphics card with 8 GB VRAM.

For the chosen optimizer, the learning rate and the decay will determine how the gradient descent is performed. A too large initial learning rate will cause the network to continuously overshoot the minimum, whereas a too small one will require a large amount of iterations to reach the minimum, as depicted in the figure B.1 hereafter.

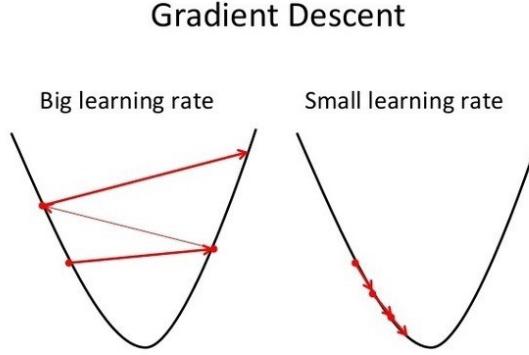


Figure B.1: Illustration of the gradient descent when the learning rate is too large (left) or too small (right)

Once the best candidates were found, we tried to modify the depth of the network. In the case of the U-Net architecture, one easy way to change the depth of the network is to add or remove one stage in both the encoding and the decoding parts. This can be directly modeled by the number of filters used in the bottleneck part of the network (originally 1024). In our case, using 1024 filters in the bottleneck part of the network, will correspond to an architecture with almost 32M parameters. Setting these first 3 parameters (learning rate, the decay and the depth of the network) gave us a baseline segmentation accuracy that we tried to improve with the data augmentation. As exposed previously, the standard way to apply data augmentation is to perform geometrical transformations to the original images so that the network will further be invariant to those modifications.

The classical transformations applied are:

- Translation: Each translation is performed in one randomly chosen direction, where the delta value corresponds to the amplitude of the shift in terms of number of pixels (e.g shift of 0.1 for an initial size of 256 corresponds to a shift of 26 pixels in one direction).
- Flip : Horizontal or Vertical flip of the original image.
- Rotation : Apply a rotation of a given angle to the original image.

We decided in our case to only investigate the data augmentation that will prevent the aspect of the liver and its internal tissues. In the case of elastic deformation, we can not testify that the retained features will still be discriminant in our deep radiomics approach for the grade classification if the aspect of the internal liver tissues is changed. We have performed for each of the following experiments

a 5-Fold Cross-Validation training on the 3DIRcad-dB (see table 5.1 for more details), and the reported results correspond to the mean slice-wise dice for the given region (parenchyma or lesion).

B.1 Learning Rate & Decay

We first evaluated the best pairs of **learning rate** and **decay** when the other parameters such as the number of epochs or the depth of the network are fixed.

Here are the results obtained with the following fixed parameters:

- Optimizer: Adam
- Number of epochs: 10
- Number of filters at bottleneck: 512
- Input image size: 256
- No data augmentation

Table B.1: Parenchyma segmentation results (mean slice-wise % Dsc)

		Decay				
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Lr	10^{-3}	82.9	91.9	90	80.3	9.6
	10^{-4}	81.3	90.8	90.8	91.9	93.7
	10^{-5}	32.4	76.7	88	89.1	89.7

Table B.2: Tumor segmentation results (mean slice-wise % Dsc)

		Decay				
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Lr	10^{-3}	7.9	23.5	24.5	16.1	1.5
	10^{-4}	6	15.6	15.6	23.6	21.7
	10^{-5}	5.4	5.4	14	17	15.2

Since those two metrics are highly related to the number of voxels of each class, we decided in order to combine them to incorporate the number of voxels of each one of the two classes to create a new weighted metric that will help us select the best networks.

Table B.3: Segmentation results obtained with the combined metric

		Decay				
		10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Lr	10^{-3}	11.6	26.9	27.7	19.3	1.9
	10^{-4}	9.7	19.3	19.3	27.0	25.2
	10^{-5}	6.7	8.9	17.6	20.5	18.9

We can see that using a too small initial learning rate (1e-5) will not be enough, for the given number of iterations, to approach the minimum during the gradient descent. On the contrary, when

the learning rate is set at a very high value such as 1e-3, with a small decay (1e-5 or lower in our case), the network continuously overshoots the minima, ending with a very large loss.

We selected the pairs providing the best accuracy and investigated the behavior of the gradient descent when slightly changing the depth of the network.

B.2 Network’s depth

Once the pairs of learning rate and decay selected, we have changed the depth of the network. This setting is controlled in our pipeline by a parameter called “bottleneck filters”.

In the above mentioned experiments, this number was set to 512, and we tried to change it with the following values: **{128, 256, 512, 1024}**.

Here are the reported results for both the segmentation of the parenchyma and the tumor when changing the depth of the network:

Table B.4: Parenchyma segmentation results (mean slice-wise % Dsc)

Lr	Decay	# Bottleneck filters			
		128	256	512	1024
10^{-3}	10^{-2}	76.1	92.1	91.9	7.2
10^{-3}	10^{-3}	76.5	88.7	87	0
10^{-4}	10^{-4}	87.1	92.8	93.9	94.2
10^{-4}	10^{-5}	89.7	92.6	89.9	93.6

Table B.5: Tumor segmentation results (mean slice-wise % Dsc)

Lr	Decay	# Bottleneck filters			
		128	256	512	1024
10^{-3}	10^{-2}	18.5	17.3	23.5	2.8
10^{-3}	10^{-3}	23.5	22.1	24.5	0
10^{-4}	10^{-4}	14.6	22	25.9	27.6
10^{-4}	10^{-5}	12.2	19	15.2	26

Table B.6: Segmentation results obtained with the combined metric

Lr	Decay	# Bottleneck filters			
		128	256	512	1024
10^{-3}	10^{-2}	21.3	21.0	26.9	3.0
10^{-3}	10^{-3}	26.1	25.4	27.6	0.0
10^{-4}	10^{-4}	18.2	25.5	29.2	30.9
10^{-4}	10^{-5}	16.0	22.6	18.9	29.3

We can see that a too large network depth can cause problems, probably because the size of the training dataset is too small (potential overfitting).

For the next experiments, we have decided to fix the learning rate to **1e-4** and the decay to **1e-4**, with a number of filters at the bottleneck stage of the network of **1024**, similar to the original architecture [286].

Once those hyperparameters were set, we investigated the amount and the type of data augmentation to apply during the training.

B.3 Data augmentation

We investigate the impact of geometrical data augmentation by first fixing the amount of data augmentation to 4, meaning that each image in the training set will be used to create 4 new images.

The other hyperparameters are fixed as follow:

- Lr: 1e-4
- Decay: 1e-4
- Optimizer: Adam
- Number of epochs: 10
- Number of filters at bottleneck: 1024
- Input image size: 256

Translation

We first investigated the value brought by translations during the data augmentation process, by changing the shift from large to small deltas.

Here are the results for each of the chosen sets (mean slice-wise % Dsc, expect for the weighted combined metric):

Shift delta	Parenchyma	Tumor	Combined
No Augm.	94.2	27.6	30.9
[-0.1, -0.05, ..., 0.1]	94.8	32.6	35.7
[-0.1, 0, 0.1]	95	32.5	35.6
[-0.2, 0.1, ..., 0.2]	95	34.6	37.6
[-0.3, 0.2, ..., 0.3]	94.6	32	35.1

We can see that all the experiments provided a better accuracy than the baseline, with a better accuracy obtained for images augmented with a shift in the range [-0.2, 0.2].

Flip

We also evaluated the value brought by flipped images during the data augmentation process. Flip can only be performed horizontally or vertically, thus a maximum of 2 images can be generated for a given initial image.

Here are the results obtained when generating only flipped images during the data augmentation process (mean slice-wise % Dsc, expect for the weighted combined metric)::

Flip	Parenchyma	Tumor	Combined
No Augm.	94.2	27.6	30.9
With Flip	94.7	31.6	34.7

We can see here that flip as data augmentation brings a clear value to the segmentation performances of the network.

Rotation

When applying rotation, we have to carefully choose the angles used during the rotation since too small angles can cause overfitting and too large angles can be useless because representing unrealistic data.

Here are the results obtained for the chosen rotation sets (mean slice-wise % Dsc, expect for the weighted combined metric)::

Rotation angle	Parenchyma	Tumor	Combined
No Augm.	94.2	27.6	30.9
[0, 40]	94.9	35.9	38.8
[0, 80]	87.1	36	38.5
[0, 90]	87.3	36.2	38.7
[0, 270]	94.6	34.7	37.6

Reported results showed that experiments involving training images obtained after rotations always improved the baseline accuracy of the network. It is worth noting that rotations performed with angles in the range [0,40] gave results slightly better than when larger angles are used. Given the results of the above mentioned experiments, we have decided for the data augmentation to generate new images using **rotations with an angle included in the interval [0, 40]°, both horizontal and vertical flips, and translation in the range [-0.2, 0.2]**.

These 3 geometrical transformations will randomly be combined for the next experiments.

We first investigated what happened when both augmentation functions are combined, by incrementally increasing the augmentation factor, and reported the results hereafter (mean slice-wise % Dsc, expect for the weighted combined metric)::

Epochs	Augm. Factor	Parenchyma	Tumor	Combined
10	4	94	26.6	29.9
10	8	94.6	32.3	35.4
10	12	94.8	34.2	37.2
10	16	94.9	34.9	37.9
10	20	94.8	40.5	43.2

We can see that increasing the number of generated images improves the segmentation performances of the network.

For this network, we haven't seen any value when adding a dropout with a rate of 0.2, however this might be discussed because with a higher number of epochs, the addition of dropout might be beneficial.

Epochs	Augm. Factor	Dr	Parenchyma	Tumor	Combined
10	20	0	94.8	40.5	43.2
10	20	0.2	94.7	38.4	41.2

Using the original image size (512x512) seems also to improve the accuracy of the network, but in these experiments, we trained our networks with a reduced resolution (256x256) to reduce the training time.

Epochs	Augm. Factor	ImgSize	Parenchyma	Tumor	Combined
10	20	256	94.8	40.5	43.2
10	20	512	95.1	42.7	45.3

We finally increased the number of training epochs, and changed the augmentation factor, to realize that the best accuracy was obtained with an augmentation factor of 20 and a network trained with 30 epochs.

Augmenting the data with a factor higher than 20 (e.g 24 in the table below) tends not to improve the performances of the network. The same phenomenon is observed when the number of training epochs is increased at more than 30, probably because the network will start to overfit, due to a too

large number of training iterations, or due to the presence of images that are too similar.

Epochs	Augm. Factor	ImgSize	Parenchyma	Tumor	Combined
10	20	256	94.8	40.5	43.2
20	16	256	95	43.4	45.9
30	16	256	95.1	43.2	45.8
20	20	256	95	41.1	43.8
30	20	256	95.1	44.9	47.4
36	20	256	95	42.8	45.4
30	24	256	95.2	44.9	47.4

The final set of hyperparameters for the segmentation of both the parenchyma and the tumor is the following:

- Lr: 1e-4
- Decay: 1e-4
- Number of epochs: 30
- Optimizer: Adam
- Number of filters at bottleneck: 1024
- Input image size: 256
- Data augmentation
 - Rotations in the interval [0, 40]
 - Horizontal and vertical flip
- Augmentation factor: 20

Appendix C

Radiomics features

Here we detail the different formula to compute the first-order and textural radiomics features, as described by *PyRadiomics* (<https://pyradiomics.readthedocs.io/en/latest/>).

C.1 First-order statistics

Let:

- X be a set of N_P voxels included in the ROI,
- $P(i)$ be the first order histogram with N_g discrete intensity levels, where N_g is the number of non-zero bins,
- $p(i)$ be the normalized first order histogram and equal to $\frac{P(i)}{N_P}$.

We can define the following features:

1. Energy:

$$energy = \sum_{i=1}^{N_P} (X(i) + c)^2$$

2. Total Energy

$$total\ energy = V_{voxel} \sum_{i=1}^{N_P} (X(i) + c)^2$$

3. Entropy:

$$entropy = - \sum_{i=1}^{N_g} p(i) \log_2 (p(i) + \epsilon)$$

Here, ϵ is an arbitrarily small positive number ($\approx 2.2 \times 10^{-16}$)

4. Minimum:

$$\text{minimum} = \min(X)$$

5. 10th Percentile

The 10th percentile of X , a robust alternative to the minimum gray-value.

6. 90th Percentile

The 90th percentile of X , a robust alternative to the maximum gray-value.

7. Maximum:

The maximum gray level intensity within the ROI.

$$\text{maximum} = \max(X)$$

8. Mean:

The average gray level intensity within the ROI.

$$\text{mean} = \frac{1}{N_P} \sum_{i=1}^{N_P} X(i)$$

9. Median:

The median gray level intensity within the ROI.

10. Interquartile Range:

$$\text{interquartile range} = P_{75} - P_{25}$$

Here P_{25} and P_{75} are the 25th and 75th percentile of the image array, respectively.

11. Range:

The range of gray values in the ROI.

$$\text{range} = \max(X) - \min(X)$$

12. Mean Absolute Deviation (MAD):

MAD is the mean distance of all intensity values from the Mean Value of the image array.

$$MAD = \frac{1}{N_P} \sum_{i=1}^{N_P} |X(i) - \bar{X}|$$

Where \bar{X} is the mean of X

13. Robust Mean Absolute Deviation (rMAD):

$$rMAD = \frac{1}{N_{10-90}} \sum_{i=1}^{N_{10-90}} |X_{10-90}(i) - \bar{X}_{10-90}|$$

14. Root Mean Square (RMS):

$$RMS = \sqrt{\frac{\sum_{i=1}^{N_P} (X(i) + c)^2}{N_P}}$$

Here, c is an optional value which shifts the intensities to prevent negative values in X . This ensures that voxels with the lowest gray values contribute the least to RMS, instead of voxels with gray level intensity closest to 0.

15. Standard Deviation

$$\text{standard deviation} = \sqrt{\frac{1}{N_P} \sum_{i=1}^{N_P} (X(i) - \bar{X})^2}$$

Standard Deviation measures the amount of variation or dispersion from the Mean Value. By definition, standard deviation $= \sqrt{\text{variance}}$.

16. Skewness:

$$\text{skewness} = \frac{\frac{1}{N_P} \sum_{i=1}^{N_P} (X(i) - \bar{X})^3}{\left(\sqrt{\frac{1}{N_P} \sum_{i=1}^{N_P} (X(i) - \bar{X})^2} \right)^3}$$

where \bar{X} is the mean of X .

Skewness measures the asymmetry of the distribution of values about their mean value.

17. Kurtosis:

$$\text{kurtosis} = \frac{\frac{1}{N_P} \sum_{i=1}^{N_P} (X(i) - \bar{X})^4}{\left(\sqrt{\frac{1}{N_P} \sum_{i=1}^{N_P} (X(i) - \bar{X})^2} \right)^2}$$

Where \bar{X} is the mean of X .

Kurtosis is a measure of the “peakedness” of the distribution of values in the image ROI. A higher kurtosis implies that the mass of the distribution is concentrated towards the tail(s) rather than towards the mean. A lower kurtosis implies the reverse: that the mass of the distribution is concentrated towards a spike near the Mean value.

18. Variance:

$$\text{variance} = \frac{1}{N_P} \sum_{i=1}^{N_P} (X(i) - \bar{X})^2$$

19. Uniformity:

$$\text{uniformity} = \sum_{i=1}^{N_g} P(i)^2$$

C.2 Gray Level Co-occurrence Matrix GLCM Features

A normalized GLCM is defined as $P(i, j; \delta, \alpha)$, a matrix with size $N_g \times N_g$ describing the second-order joint probability function of an image, where the $(i, j)^{\text{th}}$ element represents the number of times the combination of intensity levels i and j occur in two pixels in the image, that are separated by a distance of δ pixels in direction α . The distance δ from the center voxel is defined as the distance according to the infinity norm. For $\delta=1$, this results in 2 neighbors for each of 13 angles in 3D (26-connectivity).

Let:

- ϵ be an arbitrarily small positive number ($\approx 2.2 \times 10^{-16}$)
- $P(i, j)$ be the co-occurrence matrix for an arbitrary δ and α ,
- $p(i, j)$ be the normalized co-occurrence matrix and equal to $\frac{P(i, j)}{\sum P(i, j)}$,
- N_g be the number of discrete intensity levels in the image,
- $p_x(i) = \sum_{j=1}^{N_g} P(i, j)$ be the marginal row probabilities,
- $p_y(i) = \sum_{i=1}^{N_g} P(i, j)$ be the marginal column probabilities,
- μ_x be the mean gray level intensity of p_x ,
- μ_y be the mean gray level intensity of p_y ,
- σ_x be the standard deviation of p_x ,
- σ_y be the standard deviation of p_y ,

- $p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j)$, where $i + j = k$, and $k = 2, 3, \dots, 2N_g$,
- $p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j)$, where $|i - j| = k$, and $k = 0, 1, \dots, N_g - 1$,
- $HX = -\sum_{i=1}^{N_g} p_x(i) \log_2 [p_x(i) + \epsilon]$ be the entropy of p_x ,
- $HY = -\sum_{i=1}^{N_g} p_y(i) \log_2 [p_y(i) + \epsilon]$ be the entropy of p_y ,
- $HXY = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log_2 [P(i, j) + \epsilon]$ be the entropy of $P(i, j)$,
- $HXY1 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log_2 (p_x(i)p_y(j) + \epsilon)$,
- $HXY2 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log_2 (p_x(i)p_y(j) + \epsilon)$

We can define the following features:

1. Autocorrelation

$$\text{auto correlation} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijp(i, j)$$

2. Joint Average

$$\text{joint average} = \mu x = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) i$$

3. Cluster Prominence:

$$\text{cluster prominence} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 p(i, j)$$

4. Cluster Shade:

$$\text{cluster shade} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^3 p(i, j)$$

5. Cluster Tendency:

$$\text{cluster tendency} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^2 p(i, j)$$

6. Contrast:

$$\text{contrast} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j)$$

Contrast is a measure of the local intensity variation, favoring values away from the diagonal ($i = j$).

A larger value correlates with a greater disparity in intensity values among neighboring voxels.

7. Correlation:

$$\text{correlation} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i j p(i, j) - \mu_x \mu_y}{\sigma_x(i) \sigma_y(j)}$$

8. Difference Average:

$$\text{difference average} = \sum_{k=0}^{N_g-1} k p_{x-y}(k)$$

9. Difference Entropy:

$$\text{difference entropy} = \sum_{k=0}^{N_g-1} p_{x-y}(k) \log_2(p_{x-y}(k) + \epsilon)$$

10. Difference Variance:

$$\text{difference variance} = \sum_{k=0}^{N_g-1} (k - DA)^2 p_{x-y}(k)$$

11. Joint Energy:

$$\text{joint energy} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i, j))^2$$

12. Joint Entropy:

$$\text{joint entropy} = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log_2(p(i, j) + \epsilon)$$

13. Informal Measure of Correlation (IMC) 1:

$$IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$$

14. Informal Measure of Correlation (IMC) 2:

$$IMC2 = \sqrt{1 - e^{-2(HXY^2 - HXY)}}$$

15. Inverse Difference Moment (IDM)

$$IDM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + |i-j|^2}$$

16. Maximal Correlation Coefficient (MCC)

$$MCC = \sqrt{\text{second largest eigen value of } Q}$$

$$Q(i,j) = \sum_{k=0}^{N_g} \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$$

17. Inverse Difference Moment Normalized (IDMN):

$$IDMN = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + \left(\frac{|i-j|^2}{N_g^2}\right)}$$

18. Inverse Difference (ID):

$$ID = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + |i-j|}$$

19. Inverse Difference Normalized (IDN):

$$IDN = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + \left(\frac{|i-j|}{N_g}\right)}$$

20. Inverse Variance:

$$\text{inverse variance} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{|i-j|^2}, i \neq j$$

21. Maximum Probability:

$$\text{maximum probability} = \max(p(i, j))$$

22. Sum Average:

$$\sum \text{average} = \sum_{k=2}^{2N_g} (kp_{x+y}(k))$$

23. Sum Entropy:

$$\sum \text{entropy} = - \sum_{k=2}^{2N_g} p_{x+y}(k) \log_2(p_{x+y}(k) + \epsilon)$$

24. Sum of Squares:

$$\sum \text{squares} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_x)^2 p(i, j)$$

C.3 Gray Level Run Length Matrix (GLRLM) Features

A GLRLM quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value. In a gray level run length matrix $P(i, j \vee \theta)$, the $(i, j)^{\text{th}}$ element describes the number of runs with gray level i and length j occur in the image along angle θ .

Let:

- N_g be the number of discrete intensity values in the image,
- N_r be the number of discrete run lengths in the image,
- N_p be the number of voxels in the image,
- $N_z(\theta)$ be the number of runs in the image along angle θ , which is equal to $\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j \vee \theta)$, and $1 \leq N_z(\theta) \leq N_p$,
- $P(i, j \vee \theta)$ be the run length matrix for an arbitrary direction θ ,
- $p(i, j \vee \theta)$ be the normalized run length matrix, defined as $p(i, j | \theta) = \frac{P(i, j \vee \theta)}{N_z(\theta)}$,

We can define the following features:

1. Short Run Emphasis (SRE):

$$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i, j \vee \theta)}{j^2}}{N_z(\theta)}$$

2. Long Run Emphasis (LRE):

$$LRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 P(i, j \vee \theta)}{N_z(\theta)}$$

3. Gray Level Non-Uniformity (GLN):

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} p(i, j \vee \theta) \right)^2}{N_z(\theta)}$$

4. Gray Level Non-Uniformity Normalized (GLNN):

$$GLNN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} p(i, j \vee \theta) \right)^2}{N_z(\theta)^2}$$

5. Run Length Non-Uniformity (RLN):

$$RLN = \frac{\sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} P(i, j \vee \theta) \right)^2}{N_z(\theta)}$$

6. Run Length Non-Uniformity Normalized (RLNN):

$$RLNN = \frac{\sum_{j=1}^{N_r} \left(\sum_{i=1}^{N_g} P(i, j \vee \theta) \right)^2}{N_z(\theta)^2}$$

RLNN measures the similarity of run lengths throughout the image, with a lower value indicating more homogeneity among run lengths in the image. This is the normalized version of the RLN formula.

7. Run Percentage (RP):

$$RP = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i, j \vee \theta)}{N_p}$$

8. Gray Level Variance (GLV):

$$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j \vee \theta) (i - \mu)^2$$

Here, $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j \vee \theta) i$

9. Run Variance (RV):

$$RV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j \vee \theta) (j - \mu)^2$$

Here, $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j \vee \theta) j$

10. Run Entropy (RE):

$$RE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j \vee \theta) \log_2(p(i, j \vee \theta) + \epsilon)$$

Here, ϵ is an arbitrarily small positive number ($\approx 2.2 \times 10^{-16}$)

11. Low Gray Level Run Emphasis (LGLRE):

$$LGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i, j \vee \theta)}{i^2}}{N_z(\theta)}$$

12. High Gray Level Run Emphasis (HGLRE):

$$HGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 P(i, j \vee \theta)}{N_z(\theta)}$$

13. Short Run Low Gray Level Emphasis (SRLGLE):

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i, j \vee \theta)}{i^2 j^2}}{N_z(\theta)}$$

14. Short Run High Gray Level Emphasis (SRHGLE):

$$SRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i, j \vee \theta) i^2}{j^2}}{N_z(\theta)}$$

15. Long Run Low Gray Level Emphasis (LRLGLE):

$$LRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \vee \theta) j^2}{i^2}}{N_z(\theta)}$$

16. Long Run High Gray Level Emphasis (LRHGLE):

$$LRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j \vee \theta) i^2 j^2}{N_z(\theta)}$$

C.4 Gray Level Size Zone Matrix (GLSZM) Features

A GLSZM describes the amount of homogeneous connected areas within the volume, of a certain size and intensity, thereby describing tumor heterogeneity at a regional scale. A voxel is considered connected if the distance is 1 according to the infinity norm (26-connected region in 3D). In a GLSZM $P(i, j)$, the $(i, j)^{th}$ element equals the number of zones with gray level i and size j appear in image. Contrary to GLCM and GLRLM, the GLSZM is rotation independent, with only one matrix calculated for all directions in the ROI. The mathematical formulas that define the GLSZM features correspond to the definitions of features extracted from the GLRLM.

Let:

- N_g be the number of discrete intensity values in the image,
- N_s be the number of discrete zone sizes in the image,
- N_p be the number of voxels in the image,
- N_z be the number of zones in the ROI, which is equal to $\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i, j)$, and $1 \leq N_z \leq N_p$
- $P(i, j)$ be the size zone matrix,
- $p(i, j)$ be the normalized size zone matrix, defined as $p(i, j) = \frac{P(i, j)}{N_z}$.

We can define the following GLSZM features:

1. Small Area Emphasis (SAE):

$$SAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i, j)}{j^2}}{N_z}$$

2. Large Area Emphasis (LAE):

$$LAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i, j) j^2}{N_z}$$

3. Gray Level Non-Uniformity (GLN):

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_s} P(i, j) \right)^2}{N_z}$$

GLN measures the variability of gray-level intensity values in the image, with a lower value indicating more homogeneity in intensity values.

4. Gray Level Non-Uniformity Normalized (GLNN):

$$GLNN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_s} P(i, j) \right)^2}{N_z^2}$$

GLNN measures the variability of gray-level intensity values in the image, with a lower value indicating a greater similarity in intensity values. This is the normalized version of the GLN formula.

5. Size-Zone Non-Uniformity (SZN):

$$SZN = \frac{\sum_{j=1}^{N_s} \left(\sum_{i=1}^{N_g} P(i, j) \right)^2}{N_z}$$

6. Size-Zone Non-Uniformity Normalized (SZNN):

$$SZNN = \frac{\sum_{j=1}^{N_s} \left(\sum_{i=1}^{N_g} P(i, j) \right)^2}{N_z^2}$$

7. Zone Percentage (ZP):

$$ZP = \frac{N_z}{N_p}$$

8. Gray Level Variance (GLV):

$$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i, j) (i - \mu)^2$$

Here, $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i, j) i$

9. Zone Variance (ZV):

$$ZV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j) (j - \mu)^2$$

Here, $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j) j$

10. Zone Entropy (ZE):

$$ZE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j) \log_2 [p(i,j) + \epsilon]$$

Here, ϵ is an arbitrarily small positive number ($\approx 2.2 \times 10^{-16}$)

11. Low Gray Level Zone Emphasis (LGLZE):

$$LGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2}}{N_z}$$

12. High Gray Level Zone Emphasis (HGLZE):

$$HGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} i^2 P(i,j)}{N_z}$$

13. Small Area Low Gray Level Emphasis (SALGLE):

$$SALGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2 j^2}}{N_z}$$

14. Small Area High Gray Level Emphasis (SAHGLE):

$$SAHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)i^2}{j^2}}{N_z}$$

15. Large Area Low Gray Level Emphasis (LALGLE):

$$LALGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)j^2}{i^2}}{N_z}$$

16. Large Area High Gray Level Emphasis (LAHGLE):

$$HGLZE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} i^2 j^2 P(i,j)}{N_z}$$

C.5 Gray Level Dependence Matrix (GLDM) Features

A Gray Level Dependence Matrix (GLDM) quantifies gray level dependencies in an image. A gray level dependency is defined as the number of connected voxels within distance δ that are dependent on the center voxel. A neighboring voxel with gray level j is considered dependent on center voxel with gray level i if $|i - j| \leq \alpha$. In a gray level dependence matrix $P(i, j)$ the $(i, j)^{th}$ element describes the number of times a voxel with gray level i with j dependent voxels in its neighborhood appears in image.

Let:

- N_g be the number of discrete intensity values in the image,
- N_d be the number of discrete dependency sizes in the image,
- N_z be the number of dependency zones in the image, which is equal to $\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j)$
- $P(i, j)$ be the dependence matrix,
- $p(i, j)$ be the normalized dependence matrix, defined as $p(i, j) = \frac{P(i, j)}{N_z}$.

We can define the following GLDM features:

1. Small Dependence Emphasis (SDE):

$$SDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j)}{i^2}}{N_z}$$

A measure of the distribution of small dependencies, with a greater value indicative of smaller dependence and less homogeneous textures.

2. Large Dependence Emphasis (LDE):

$$LDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j) j^2}{N_z}$$

3. Gray Level Non-Uniformity (GLN):

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_d} P(i, j) \right)^2}{N_z}$$

4. Dependence Non-Uniformity (DN):

$$DN = \frac{\sum_{j=1}^{N_d} \left(\sum_{i=1}^{N_g} P(i, j) \right)^2}{N_z}$$

5. Dependence Non-Uniformity Normalized (DNN):

$$DN = \frac{\sum_{j=1}^{N_d} \left(\sum_{i=1}^{N_g} P(i, j) \right)^2}{N_z^2}$$

6. Gray Level Variance (GLV):

$$GLV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) (i - \mu)^2, \text{ where } \mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} i p(i, j)$$

7. Dependence Variance (DV):

$$DV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) (j - \mu)^2, \text{ where } \mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} j p(i, j)$$

8. Dependence Entropy (DE):

$$\text{Dependence Entropy} = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) \log_2 [p(i, j) + \epsilon]$$

9. Low Gray Level Emphasis (LGLE):

$$LGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j)}{i^2}}{N_z}$$

10. High Gray Level Emphasis (HGLE):

$$HGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j) i^2}{N_z}$$

11. Small Dependence Low Gray Level Emphasis (SDLGLE):

$$SDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j)}{i^2 j^2}}{N_z}$$

12. Small Dependence High Gray Level Emphasis (SDHGLE):

Measures the joint distribution of small dependence with higher gray-level values.

13. Large Dependence Low Gray Level Emphasis (LDLGLE):

$$LDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i,j)i^2}{j^2}}{N_z}$$

14. Large Dependence High Gray Level Emphasis (LDHGLE):

$$LDHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i,j) i^2 j^2}{N_z}$$

C.6 Neighbouring Gray Tone Difference Matrix (NGTDM) Features

A NGTDM quantifies the difference between a gray value and the average gray value of its neighbors within distance δ . The sum of absolute differences for gray level i is stored in the matrix. Let X_{gl} be a set of segmented voxels and $x_{gl}(j_x, j_y, j_z) \in X_{gl}$ be the gray level of a voxel at position (j_x, j_y, j_z) , then the average gray level of the neighborhood is:

$$\bar{A}_i = \bar{A}(j_x, j_y, j_z) \quad (C.1)$$

$$= \frac{1}{W} \sum_{k_x=-\delta}^{\delta} \sum_{k_y=-\delta}^{\delta} \sum_{k_z=-\delta}^{\delta} x_{gl}(j_x + k_x, j_y + k_y, j_z + k_z) \quad (C.2)$$

$$(C.3)$$

where $(k_x, k_y, k_z) \neq (0, 0, 0)$ and $x_{gl}(j_x + k_x, j_y + k_y, j_z + k_z) \in X_{gl}$

Here, W is the number of voxels in the neighborhood that are also in X_{gl} .

We can define the following NGTDM features:

1. Coarseness:

$$Coarseness = \frac{1}{\sum_{i=1}^{N_g} p_i s_i}$$

2. Contrast:

$$Contrast = \left(\frac{1}{N_{g,p}(N_{g,p}-1)} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_i p_j (i-j)^2 \right) \left(\frac{1}{N_{v,p}} \sum_{i=1}^{N_g} s_i \right), \text{ where } p_i \neq 0, p_j \neq 0$$

3. Busyness:

$$\text{Busyness} = \frac{\sum_{i=1}^{N_g} p_i s_i}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |ip_i - jp_j|}, \text{ where } p_i \neq 0, p_j \neq 0$$

4. Complexity:

$$\text{Complexity} = \frac{1}{N_{v,p}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j| \frac{p_i s_i + p_j s_j}{p_i + p_j}, \text{ where } p_i \neq 0, p_j \neq 0$$

5. Strength:

$$\text{Strength} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p_i + p_j) (i - j)^2}{\sum_{i=1}^{N_g} s_i}, \text{ where } p_i \neq 0, p_j \neq 0$$

Appendix D

Evaluation metrics

Here we detail the most commonly used evaluation metrics.

D.1 Overlap measures

Dice and Jaccard are defined as follow:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \quad Jaccard(A, B) = \frac{|A| \cap |B|}{|A \cup B|}$$

Volume overlap error (VOE) :

$$VOE(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

Relative volume difference (RVD), corresponding to an asymmetric measure and defined as follow:

$$RVD(A, B) = \frac{|B| - |A|}{|A|}$$

D.2 Surface Distance Metrics

This type of metric evaluate the quality of the segmentation by measuring the distance between the surfaces of reference and the predicted areas.

Let $S(A)$ denote the set of surface voxels of A , the shortest distance between an arbitrary voxel v and $S(A)$ will be defined as :

$$d(v, S(A)) = \min_{s_A \in S(A)} \|v - s_A\|, \text{ where } \|\cdot\| \text{ corresponds to the Euclidean distance.}$$

We can compute the following metrics :

The Average Symmetric Surface Distance (ASD):

$$\text{ASD} = \frac{1}{|S(A)| + |S(B)|} \left(\sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A)) \right)$$

The Maximum symmetric surface distance(MSD):

$$MSD(A, B) = \max \{ \max_{s_A \in S(A)} d(s_A, S(B)), \max_{s_B \in S(B)} d(s_B, S(A)) \}$$

Bibliography

- [1] F. Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel;, Lindsey A. Torre;, and Ahmedin Jemal. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 2018.
- [2] World health organization. <http://www.who.int/news-room/fact-sheets/detail/cancer>. Accessed: 2020-10-20.
- [3] Morten Høyer, Anand Swaminath, Sean Bydder, Michael Lock, Alejandra Méndez Romero, Brian Kavanagh, Karyn A. Goodman, Paul Okunieff, and Laura A. Dawson. Radiotherapy for liver metastases: A review of evidence. *International Journal of Radiation Oncology Biology Physics*, 82(3):1047–1057, 2012.
- [4] Dushyant V. Sahani, Mohammad Ali Bajwa, Yasir Andrabi, Surabhi Bajpai, and James C. Cusack. Current status of imaging and emerging techniques to evaluate liver metastases from colorectal carcinoma. *Annals of Surgery*, 259(5):861–872, 2014.
- [5] Daniela Sia, Augusto Villanueva, Scott L. Friedman, and Josep M. Llovet. Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. *Gastroenterology*, 152(4):745–761, 2017.
- [6] Rafael Lozano, Mohsen Naghavi, Kyle Foreman, Stephen Lim, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859):2095–2128, 2012.
- [7] F. T. Bosman, F. Carneiro, R. H. Hruban, and N. D. Theise. *WHO classification of tumours of the digestive system*. Number Ed. 4. World Health Organization, Geneva, Switzerland, 2010.
- [8] Lindsey A. Torre, Freddie Bray, Rebecca L. Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65(2):87–108, 2015.

- [9] John Bridgewater, Peter R. Galle, Shahid A. Khan, Josep M. Llovet, Joong Won Park, Tushar Patel, Timothy M. Pawlik, and Gregory J. Gores. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *Journal of Hepatology*, 60(6):1268–1289, 2014.
- [10] Alejandro Forner, María Reig, and Jordi Bruix. Hepatocellular carcinoma. *The Lancet*, 391(10127):1301–1314, apr 2018.
- [11] Andrea Marengo, Chiara Rosso, and Elisabetta Bugianesi. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. *Annual Review of Medicine*, 67(1):103–117, 2016.
- [12] Jorge A. Marrero, Robert J. Fontana, Grace L. Su, Hari S. Conjeevaram, Dawn M. Emick, and Anna S. Lok. NAFLD may be a common underlying liver disease in patients with hepatocellular carcinoma in the United States. *Hepatology*, 36(6):1349–1354, 2002.
- [13] Valérie Paradis, Stéphane Zalinski, Emna Chelbi, Nathalie Guedj, Françoise Degos, Valérie Vilgrain, Pierre Bedossa, and Jacques Belghiti. Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liverfibrosis: A pathological analysis. *Hepatology*, 49(3):851–859, 2009.
- [14] Jessica Dyson, Bryan Jaques, Dipankar Chattopadyhay, Rajiv Lochan, Janine Graham, Debasish Das, Tahira Aslam, et al. Hepatocellular cancer: The impact of obesity, type 2 diabetes and a multidisciplinary team. *Journal of Hepatology*, 60(1):110–117, 2014.
- [15] Robert J. Wong, Ramsey Cheung, and Aijaz Ahmed. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology*, 59(6):2188–2195, 2014.
- [16] Mustafa S. Ascha, Ibrahim A. Hanouneh, Rocio Lopez, Tarek Abu Rajab Tamimi, Ariel F. Feldstein, and Nizar N. Zein. The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. *Hepatology*, 51(6):1972–1978, 2010.
- [17] Sahil Mittal, Yvonne H. Sada, Hashem B. El-Serag, Fasiha Kanwal, Zhigang Duan, Sarah Temple, Sarah B. May, Jennifer R. Kramer, Peter A. Richardson, and Jessica A. Davila. Temporal trends of nonalcoholic fatty liver disease-related hepatocellular carcinoma in the veteran affairs population. *Clinical Gastroenterology and Hepatology*, 13(3):594–601.e1, 2015.
- [18] Eugenia E. Calle, Carmen Rodriguez, Kimberly Walker-Thurmond, and Michael J. Thun. Overweight, Obesity, and Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults. *New England Journal of Medicine*, 348(17):1625–1638, apr 2003.

- [19] Chi Ling Chen, Hwai I. Yang, Wei Shiung Yang, Chun Jen Liu, Pei Jer Chen, San Lin You, Li Yu Wang, Chien An Sun, Sheng Nan Lu, Ding Shin Chen, and Chien Jen Chen. Metabolic Factors and Risk of Hepatocellular Carcinoma by Chronic Hepatitis B/C Infection: A Follow-up Study in Taiwan. *Gastroenterology*, 135(1):111–121, 2008.
- [20] Sabrina Schlesinger, Krasimira Aleksandrova, Tobias Pischon, Veronika Fedirko, Mazda Jenab, Elisabeth Trepo, Paolo Boffetta, Christina C Dahm, Kim Overvad, Anne Tjønneland, Jytte Halkjær, and Guy Fagherazzi. Abdominal obesity , weight gain during adulthood and risk of liver and biliary tract cancer in a European Cohort. 000(May), 2012.
- [21] Claudine Samanic, Wong Ho Chow, Gloria Gridley, Bengt Jarvholm, and Joseph F. Fraumeni. Relation of body mass index to cancer risk in 362,552 Swedish men. *Cancer Causes and Control*, 17(7):901–909, 2006.
- [22] Sang Woo Oh, Yeong Sook Yoon, and Soon Ae Shin. Effects of excess weight on cancer incidences depending on cancer sites and histologic findings among men: Korea National Health Insurance Corporation study. *Journal of Clinical Oncology*, 23(21):4742–4754, 2005.
- [23] G. D. Batty, M. J. Shipley, R. J. Jarrett, E. Breeze, M. G. Marmot, and G. Davey Smith. Obesity and overweight in relation to organ-specific cancer mortality in London (UK): Findings from the original Whitehall study. *International Journal of Obesity*, 29(10):1267–1274, 2005.
- [24] Hiroshi Noto, Keiichiro Osame, Takehiko Sasazuki, and Mitsuhiko Noda. Substantially increased risk of cancer in patients with diabetes mellitus: A systematic review and meta-analysis of epidemiologic evidence in Japan. *Journal of Diabetes and its Complications*, 24(5):345–353, 2010.
- [25] Louise Wideroff, Gloria Gridley, Lene Mellemkjaer, W. H. Chow, Martha Linet, Shannon Keehn, Knut Borch-Johnsen, and Jørgen H. Olsen. Cancer incidence in a population-based cohort of patients hospitalized with diabetes mellitus in denmark. *Journal of the National Cancer Institute*, 89(18):1360–1365, 1997.
- [26] Hashem B. El-Serag, Thomas Tran, and James E. Everhart. Diabetes Increases the Risk of Chronic Liver Disease and Hepatocellular Carcinoma. *Gastroenterology*, 126(2):460–468, 2004.
- [27] J. A. Davila, R. O. Morgan, Y. Shaib, K. A. McGlynn, and H. B. El-Serag. Diabetes increases the risk of hepatocellular carcinoma in the United States: A population based case control study. *Gut*, 54(4):533–539, 2005.
- [28] Manami Inoue. Diabetes Mellitus and the Risk of Cancer. *Archives of Internal Medicine*, 166(17):1871, sep 2006.

- [29] Jin Young Choi, Jeong Min Lee, and Claude B. Sirlin. CT and MR imaging diagnosis and staging of hepatocellular carcinoma: Part I. Development, growth, and spread: Key pathologic and imaging aspects. *Radiology*, 272(3):635–654, 2014.
- [30] Snorri S. Thorgeirsson and Joe W. Grisham. Molecular pathogenesis of human hepatocellular carcinoma. *Nature Genetics*, 31(4):339–346, 2002.
- [31] Rajagopal N. Aravalli, Erik N.K. Cressman, and Clifford J. Steer. Cellular and molecular mechanisms of hepatocellular carcinoma: An update. *Archives of Toxicology*, 87(2):227–247, 2013.
- [32] Rachel I. Brody and Neil D. Theise. An inflammatory proposal for hepatocarcinogenesis. *Hepatology*, 56(1):382–384, jul 2012.
- [33] Franco Trevisani, M. C. Cantarini, J. R. Wands, and M. Bernardi. Recent advances in the natural history of hepatocellular carcinoma. *Carcinogenesis*, 29(7):1299–1305, 2008.
- [34] Catherine Frenette and Robert G. Gish. Hepatocellular Carcinoma: Molecular and Genomic Guideline for the Clinician. *Clinics in Liver Disease*, 15(2):307–321, 2011.
- [35] Young Nyun Park. Update on precursor and early lesions of hepatocellular carcinomas. *Archives of Pathology and Laboratory Medicine*, 135(6):704–715, 2011.
- [36] Ken Ichi Taguchi, Yoshiki Asayama, Shin Ichi Aishima, Hidehiro Nishi, et al. Morphologic approach to hepatocellular carcinoma development in man: De novo or the so-called 'dysplastic nodule-carcinoma' sequence? *Oncology Reports*, 9(4):737–743, July 2002.
- [37] A. Zimmermann. Terminology of nodular hepatocellular lesions. *Hepatology*, 22(3):983–993, sep 1995.
- [38] Korosh Khalili, Tae Kyoung Kim, Hyun Jung Jang, Leyla Kochak Yazdi, Maha Guindi, and Morris Sherman. Indeterminate 1-2-cm nodules found on hepatocellular carcinoma surveillance: Biopsy for all, some, or none? *Hepatology*, 54(6):2048–2054, 2011.
- [39] Masatoshi Kudo. Real practice of hepatocellular carcinoma in Japan: Conclusions of the Japan society of hepatology 2009 kobe congress. *Oncology*, 78(SUPPL. 1):180–188, 2010.
- [40] R.Palmer Beasley, Chia-Chin Lin, Lu-Yu Hwang, and Chia-Siang Chien. HEPATOCELLULAR CARCINOMA AND HEPATITIS B VIRUS. *The Lancet*, 318(8256):1129–1133, nov 1981.
- [41] Hashem B. El-Serag. Hepatocellular Carcinoma. *New England Journal of Medicine*, 365(12):1118–1127, sep 2011.

- [42] György Baffy, Elizabeth M. Brunt, and Stephen H. Caldwell. Hepatocellular carcinoma in non-alcoholic fatty liver disease: An emerging menace. *Journal of Hepatology*, 56(6):1384–1391, 2012.
- [43] Katherine A. McGlynn and W. Thomas London. The Global Epidemiology of Hepatocellular Carcinoma: Present and Future. *Clinics in Liver Disease*, 15(2):223–243, 2011.
- [44] Gia L. Tyson and Hashem B. El-Serag. Risk factors for cholangiocarcinoma. *Hepatology*, 54(1):173–184, 2011.
- [45] Neil Theise, M.P. Curado, and S. Franceschi. Hepatocellular carcinoma. *WHO Classification of Tumours of the Digestive System*, pages 205–216, 01 2010.
- [46] William C. Palmer and Tushar Patel. Are common factors involved in the pathogenesis of primary liver cancers? A meta-analysis of risk factors for intrahepatic cholangiocarcinoma. *Journal of Hepatology*, 57(1):69–76, 2012.
- [47] William B. Coleman. Mechanisms of Human Hepatocarcinogenesis. *Current Molecular Medicine*, 3(6):573–588, 2005.
- [48] Masamichi Kojiro and Tania Roskams. Early hepatocellular carcinoma and dysplastic nodules. *Seminars in Liver Disease*, 25(2):133–142, 2005.
- [49] Azusa Kitao, Yoh Zen, Osamu Matsui, Toshifumi Gabata, and Yasuni Nakanuma. Hepatocarcinogenesis : Multistep Changes of Drainage Vessels at CT during Arterial Portography and Purpose : Methods : Results : Conclusion :. *Ra*, 252(2):605–614, 2009.
- [50] Kazuhiko Ueda, Osamu Matsui, Yasuhiro Kawamori, Yasuni Nakanuma, Masumi Kadoya, Jun Yoshikawa, Toshifumi Gabata, Akitaka Nonomura, and Tsutomu Takashima. Hypervasculär hepatocellular carcinoma: Evaluation of hemodynamics with dynamic CT during hepatic arteriography. *Radiology*, 206(1):161–166, 1998.
- [51] Osamu Matsui, Satoshi Kobayashi, Junichiro Sanada, Wataru Kouda, Yasuji Ryu, Kazuto Kozaka, Azusa Kitao, Koichi Nakamura, and Toshifumi Gabata. Hepatocelluar nodules in liver cirrhosis: Hemodynamic evaluation (angiography-assisted CT) with special reference to multi-step hepatocarcinogenesis. *Abdominal Imaging*, 36(3):264–272, 2011.
- [52] Irene O.L. Ng, Edward C.S. Lai, Sheung T. Fan, and Matthew M.T. Ng. Tumor encapsulation in hepatocellular carcinoma. A pathologic study of 189 cases. *Cancer*, 70(1):45–49, 1992.

- [53] Tomohiro Iguchi, Shinichi Aishima, Kensaku Sanefuji, Nobuhiro Fujita, Keishi Sugimachi, Tomonobu Gion, Akinobu Taketomi, Ken Shirabe, Yoshihiko Maehara, and Masazumi Tsuneyoshi. Both fibrous capsule formation and extracapsular penetration are powerful predictors of poor survival in human hepatocellular carcinoma: A histological assessment of 365 patients in Japan. *Annals of Surgical Oncology*, 16(9):2539–2546, 2009.
- [54] Rumiko Kutami, Yutaka Nakashima, Osamu Nakashima, Koji Shiota, and Masamichi Kojiro. Pathomorphologic study on the mechanism of fatty change in small hepatocellular carcinoma of humans. *Journal of Hepatology*, 33(2):282–289, 2000.
- [55] Masamichi Kojiro, Ian R. Wanless, Venancio Alves, et al. Pathologic diagnosis of early hepatocellular carcinoma: A report of the international consensus group for hepatocellular neoplasia. *Hepatology*, 49(2):658–664, feb 2009.
- [56] Kurinchi Gurusamy. Trace element concentration in primary liver cancers - A systematic review. *Biological Trace Element Research*, 118(3):191–206, 2007.
- [57] Azusa Kitao, Osamu Matsui, Norihide Yoneda, Kazuto Kozaka, Rieko Shinmura, Wataru Koda, Satoshi Kobayashi, Toshifumi Gabata, Yoh Zen, Tatsuya Yamashita, Shuichi Kaneko, and Yasuni Nakanuma. The uptake transporter OATP8 expression decreases during multistep hepatocarcinogenesis: Correlation with gadoxetic acid enhanced MR imaging. *European Radiology*, 21(10):2056–2066, 2011.
- [58] Takahiro Tsuboyama, Hiromitsu Onishi, Tonsok Kim, Hirofumi Akita, Masatoshi Hori, Mitsuaki Tatsumi, Atsushi Nakamoto, Hiroaki Nagano, Nariaki Matsuura, Kenichi Wakasa, and Kaname Tomoda. Hepatocellular carcinoma: Hepatocyte-selective enhancement at gadoxetic acid-enhanced MR imaging - Correlation with expression of sinusoidal and canalicular transporters and bile accumulation. *Radiology*, 255(3):824–833, 2010.
- [59] Yutaka Nakashima, Osamu Nakashima, Masatoshi Tanaka, Koji Okuda, Miwako Nakashima, and Masamichi Kojiro. Portal vein invasion and intrahepatic micrometastasis in small hepatocellular carcinoma by gross type. *Hepatology Research*, 26(2):142–147, 2003.
- [60] Steiner Edmondson HA and PE. Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies. *Cancer*, 7(3):462–503, 1954.
- [61] Kunio Okuda. Hepatocellular carcinoma: Clinicopathological aspects. *Journal of Gastroenterology and Hepatology (Australia)*, 12(9-10):314–318, 1997.

- [62] Josep M. Llovet, Josep Fuster, and Jordi Bruix. The Barcelona approach: Diagnosis, staging, and treatment of hepatocellular carcinoma. *Liver Transplantation*, 10(2 SUPPL. 1):115–120, 2004.
- [63] Arturo A. Bravo, Sunil G. Sheth, and Sanjiv Chopra. Liver Biopsy. *New England Journal of Medicine*, 344(7):495–500, feb 2001.
- [64] Dou Sheng Bai, Chi Zhang, Ping Chen, Sheng Jie Jin, and Guo Qing Jiang. The prognostic correlation of AFP level at diagnosis with pathological grade, progression, and survival of patients with hepatocellular carcinoma. *Scientific Reports*, 7(1):1–9, 2017.
- [65] Julie K. Heimbach, Laura M. Kulik, Richard S. Finn, Claude B. Sirlin, Michael M. Abecassis, Lewis R Roberts, Andrew X. Zhu, M Hassan Murad, and Jorge A. Marrero. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*, 67(1):358–380, jan 2018.
- [66] S. Turdean, Simona Gurzu, M. Turcu, S. Voidazan, and Anca Sin. Current data in clinicopathological characteristics of primary hepatic tumors. *Romanian Journal of Morphology and Embryology*, 53(3 SUPPL.):719–724, 2012.
- [67] Sebastiao N. Martins-Filho, Caterina Paiva, Raymundo Soares Azevedo, and Venancio Avancini Ferreira Alves. Histological grading of hepatocellular carcinoma-a systematic review of literature. *Frontiers in Medicine*, 4(NOV):1–9, 2017.
- [68] Dai Hoon Han, Gi Hong Choi, Kyung Sik Kim, Jin Sub Choi, Young Nyun Park, Seung Up Kim, Jun Yong Park, Sang Hoon Ahn, and Kwang Hyub Han. Prognostic significance of the worst grade in hepatocellular carcinoma with heterogeneous histologic grades of differentiation. *Journal of Gastroenterology and Hepatology (Australia)*, 28(8):1384–1390, 2013.
- [69] Jessica Zucman-Rossi, Augusto Villanueva, Jean Charles Nault, and Josep M. Llovet. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology*, 149(5):1226–1239.e4, 2015.
- [70] Timothy M. Pawlik, Ana L. Gleisner, Robert A. Anders, Lia Assumpcao, Warren Maley, and Michael A. Choti. Preoperative Assessment of Hepatocellular Carcinoma Tumor Grade Using Needle Biopsy. *Annals of Surgery*, 245(3):435–442, mar 2007.
- [71] Arie Regev, Mariana Berho, Lennox J. Jeffers, Clara Milikowski, Enrique G. Molina, Nikolaos T. Pyrsopoulos, Zheng Zhou Feng, K. Rajender Reddy, and Eugene R. Schiff. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *American Journal of Gastroenterology*, 97(10):2614–2618, 2002.

- [72] Guido Colloredo, Maria Guido, Aurelio Sonzogni, and Gioacchino Leandro. Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: The smaller the sample, the milder the disease. *Journal of Hepatology*, 39(2):239–244, 2003.
- [73] P. Bedossa. Intraobserver and Interobserver Variations in Liver Biopsy Interpretation in Patients with Chronic Hepatitis C. *Hepatology*, 20(1):15–20, jul 1994.
- [74] Don C. Rockey, Stephen H. Caldwell, Zachary D. Goodman, Rendon C. Nelson, and Alastair D. Smith. Liver biopsy. *Hepatology*, 49(3):1017–1044, 2009.
- [75] Laurent Castera, Isabelle Negre, Kamran Samii, and Catherine Buffet. Patient-administered nitrous oxide/oxygen inhalation provides safe and effective analgesia for percutaneous liver biopsy: A randomized placebo-controlled trial. *The American Journal of Gastroenterology*, 96(5):1553–1557, 2001.
- [76] Leonard B. Seeff, Gregory T. Everson, Timothy R. Morgan, Teresa M. Curto, William M. Lee, Marc G. Ghany, Mitchell L. Shiffman, Robert J. Fontana, Adrian M. Di Bisceglie, Herbert L. Bonkovsky, and Jules L. Dienstag. Complication Rate of Percutaneous Liver Biopsies Among Persons With Advanced Chronic Liver Disease in the HALT-C Trial. *Clinical Gastroenterology and Hepatology*, 8(10):877–883, 2010.
- [77] F. Piccinino, E. Sagnelli, G. Pasquale, G. Giusti, A. Battocchia, et al. Complications following percutaneous liver biopsy. *Journal of Hepatology*, 2(2):165–173, 1986.
- [78] Masao Omata, Laurentius A. Lesmana, Ryosuke Tateishi, Pei Jer Chen, et al. Asian pacific association for the study of the liver consensus recommendations on hepatocellular carcinoma. *Hepatology International*, 4(2):439–474, 2010.
- [79] Josep M. Llovet, Michel Ducreux, Riccardo Lencioni, Adrian M. Di Bisceglie, Peter R. Galle, Jean Francois Dufour, Tim F. Greten, Eric Raymond, Tania Roskams, Thierry De Baere, Michel Ducreux, Vincenzo Mazzaferro, Mauro Bernardi, Jordi Bruix, Massimo Colombo, and Andrew Zhu. EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. *Journal of Hepatology*, 56(4):908–943, 2012.
- [80] Jordi Bruix and Morris Sherman. Management of hepatocellular carcinoma: An update. *Hepatology*, 53(3):1020–1022, 2011.
- [81] W. Schima, C. Kulinna, A. Ba-Ssalamah, and T. Grünberger. Multidetektor-CT (MDCT) der leber. *Radiologe*, 45(1):15–23, 2005.

- [82] Alex Frydrychowicz, Meghan G. Lubner, Jeffrey J. Brown, Elmar M. Merkle, Scott K. Nagle, Neil M. Rofsky, and Scott B. Reeder. Hepatobiliary MR imaging with gadolinium-based contrast agents. *Journal of Magnetic Resonance Imaging*, 35(3):492–511, 2012.
- [83] Roberto Miraglia, Giada Pietrosi, Luigi Maruzzelli, Ioannis Petridis, Settimo Caruso, Gianluca Marrone, Giuseppe Mamone, Giovanni Vizzini, Angelo Luca, and Bruno Gridelli. Predictive factors of tumor response to trans-catheter treatment in cirrhotic patients with hepatocellular carcinoma:A multivariate analysis of pre-treatment findings. *World Journal of Gastroenterology*, 13(45):6022–6026, 2007.
- [84] Myeong Jin Kim, Young Choi Jin, Seok Lim Joon, Yong Kim Jin, Hee Kim Joo, Taik Oh Young, Hye Yoo Eun, Joon Chung Jae, and Whang Kim Ki. Optimal scan window for detection of hypervascular hepatocellular carcinomas during MDCT examination. *American Journal of Roentgenology*, 187(1):198–206, 2006.
- [85] James P. Earls, Neil M. Rofsky, Douglas R. DeCorato, Glenn A. Krinsky, and Jeffrey C. Weinreb. Hepatic arterial-phase dynamic gadolinium-enhanced MR imaging: Optimization with a test examination and a power injector. *Radiology*, 202(1):268–273, 1997.
- [86] Jin Young Choi, Jeong Min Lee, and Claude B. Sirlin. CT and MR imaging diagnosis and staging of hepatocellular carcinoma. Part II. Extracellular agents, hepatobiliary agents, and ancillary imaging features. *Radiology*, 273(1):30–50, 2014.
- [87] Riccardo Iannaccone, Andrea Laghi, Carlo Catalane, Plinio Rossi, Filippo Mangiapane, Takamichi Murakami, Masatoshi Hori, Francesca Piacentini, Italo Nofroni, and Roberto Passariello. Hepatocellular carcinoma: Mole of unenhanced and delayed phase multi-detector mew helical CT in patients with cirrhosis. *Radiology*, 234(2):460–467, 2005.
- [88] Jordi Rimola, Alejandro Forner, Maria Reig, Ramon Vilana, Carlos Rodríguez De Lope, Carmen Ayuso, and Jordi Bruix. Cholangiocarcinoma in cirrhosis: Absence of contrast washout in delayed phases by magnetic resonance imaging avoids misdiagnosis of hepatocellular carcinoma. *Hepatology*, 50(3):791–798, 2009.
- [89] Yuko Nakamura, Naoyuki Toyota, Shuji Date, Seitaro Oda, Tomohiro Namimoto, Yasuyuki Yamashita, Toru Beppu, and Kazuo Awai. Clinical significance of the transitional phase at gadoxetate disodium-enhanced hepatic MRI for the diagnosis of hepatocellular carcinoma: Preliminary results. *Journal of Computer Assisted Tomography*, 35(6):723–727, 2011.
- [90] Yukihisa Takayama, Akihiro Nishie, Tomohiro Nakayama, Yoshiki Asayama, Kousei Ishigami, Daisuke Kakihara, Yasuhiro Ushijima, Nobuhiro Fujita, Masakazu Hirakawa, and Hiroshi Honda.

- Hypovascular hepatic nodule showing hypointensity in the hepatobiliary phase of gadoxetic acid-enhanced MRI in patients with chronic liver disease: Prediction of malignant transformation. *European Journal of Radiology*, 81(11):3072–3078, 2012.
- [91] Takashi Kumada, Hidenori Toyoda, Toshifumi Tada, Yasuhiro Sone, Masashi Fujimori, Sadanobu Ogawa, and Teruyoshi Ishikawa. Evolution of hypointense hepatocellular nodules observed only in the hepatobiliary phase of gadoxetate disodium-enhanced MRI. *American Journal of Roentgenology*, 197(1):58–63, 2011.
- [92] Satoshi Kobayashi, Osamu Matsui, Toshifumi Gabata, Wataru Koda, Tetsuya Minami, Yasuji Ryu, Kazuto Kozaka, and Azusa Kitao. Relationship between signal intensity on hepatobiliary phase of gadolinium ethoxybenzyl diethylenetriaminepentaacetic acid (Gd-EOB-DTPA)-enhanced MR imaging and prognosis of borderline lesions of hepatocellular carcinoma. *European Journal of Radiology*, 81(11):3002–3009, 2012.
- [93] Young Kon Kim, Won Jae Lee, Min Jung Park, Seong Hyun Kim, Hyunchul Rhim, and Dongil Choi. Hypovascular hypointense nodules on hepatobiliary phase gadoxetic acid-enhanced MR images in patients with cirrhosis: Potential of DW imaging in predicting progression to hypervasculat HCC. *Radiology*, 265(1):104–114, 2012.
- [94] Tomoko Hyodo, Takamichi Murakami, Yasuharu Imai, Masahiro Okada, Masatoshi Hori, Yuki Kagawa, Sachiko Kogita, Seishi Kumano, Masatoshi Kudo, and Teruhito Mochizuki. Hypovascular nodules in patients with chronic liver disease: Risk factors for development of hypervasculat hepatocellular carcinoma. *Radiology*, 266(2):480–490, 2013.
- [95] Min Jung Park, Young Kon Kim, Min Woo Lee, Won Jae Lee, Young Sun Kim, Seong Hyun Kim, Dongil Choi, and Hyunchul Rhim. Small hepatocellular carcinomas: Improved sensitivity by combining gadoxetic acid-enhanced and diffusion-weighted MR imaging patterns. *Radiology*, 264(3):761–770, 2012.
- [96] Seong Hyun Kim, Seung Hoon Kim, Jongmee Lee, Min Ju Kim, Yong Hwan Jeon, Yulri Park, Dongil Choi, Won Jae Lee, and Hyo K. Lim. Gadoxetic Acid–Enhanced MRI Versus Triple-Phase MDCT for the Preoperative Detection of Hepatocellular Carcinoma. *American Journal of Roentgenology*, 192(6):1675–1681, jun 2009.
- [97] Josep M. Llovet, Adrian M. Di Bisceglie, Jordi Bruix, Barnett S. Kramer, Riccardo Lencioni, Andrew X. Zhu, Morris Sherman, Myron Schwartz, Michael Lotze, Jayant Talwalkar, and Gregory J. Gores. Design and endpoints of clinical trials in hepatocellular carcinoma. *Journal of the National Cancer Institute*, 100(10):698–711, 2008.

- [98] RN Pugh, IM Murray-Lyon, JL Dawson, MC Pietroni, and R Williams. Transection of the oesophagus for bleeding oesophageal varices. *The British journal of surgery*, 60(8):646–649, August 1973.
- [99] M. M. Oken, R. H. Creech, and T. E. Davis. Toxicology and response criteria of the Eastern Cooperative Oncology Group, 1982.
- [100] Vincenzo Mazzaferro, Enrico Regalia, Roberto Doci, Salvatore Andreola, Andrea Pulvirenti, Federico Bozzetti, Fabrizio Montalto, Mario Ammatuna, Alberto Morabito, and Leandro Gennari. Carcinomas in Patients With Cirrhosis. *The New England Journal of Medicine*, 334(11):693–699, 1996.
- [101] Stephen B. Edge and Carolyn C. Compton. The american joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of Surgical Oncology*, 17(6):1471–1474, 2010.
- [102] Tian Yang, Chuan Lin, Jian Zhai, Song Shi, Min Zhu, Nan Zhu, Jun Hua Lu, Guang Shun Yang, and Meng Chao Wu. Surgical resection for advanced hepatocellular carcinoma according to Barcelona Clinic Liver Cancer (BCLC) staging. *Journal of Cancer Research and Clinical Oncology*, 138(7):1121–1129, 2012.
- [103] C. Verslype, O. Rosmorduc, and P. Rougier. Hepatocellular carcinoma: ESMO-ESDO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 23(SUPPL. 7), 2012.
- [104] Vincenzo Mazzaferro, Sherrie Bhoori, Carlo Sposito, Marco Bongini, Martin Langer, Rosalba Miceli, and Luigi Mariani. Milan criteria in liver transplantation for hepatocellular carcinoma: An evidence-based analysis of 15 years of experience. *Liver Transplantation*, 17(S2):S44–S57, oct 2011.
- [105] Josep M. Llovet, Josep Fuster, and Jordi Bruix. Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: Resection versus transplantation. *Hepatology*, 30(6):1434–1440, 1999.
- [106] Alejandro Forner, Josep M. Llovet, and Jordi Bruix. Hepatocellular carcinoma. *The Lancet*, 379(9822):1245–1255, 2012.
- [107] David J. Breen and Riccardo Lencioni. Image-guided ablation of primary liver and renal tumours. *Nature Reviews Clinical Oncology*, 12(3):175–186, 2015.

- [108] Stéphanie Perrodin, Anja Lachenmayer, Martin Maurer, Corina Kim-Fuchs, Daniel Candinas, and Vanessa Banz. Percutaneous stereotactic image-guided microwave ablation for malignant liver lesions. *Scientific Reports*, 9(1):1–8, 2019.
- [109] Reto Bale, Peter Schullian, Gernot Eberle, Daniel Putzer, Heinz Zoller, Stefan Schneeberger, Claudia Manzl, Patrizia Moser, and Georg Oberhuber. Stereotactic Radiofrequency Ablation of Hepatocellular Carcinoma: a Histopathological Study in Explanted Livers. *Hepatology*, 70(3):840–850, 2019.
- [110] Gregor Laimer, Peter Schullian, Nikolai Jaschke, Daniel Putzer, Gernot Eberle, Amilcar Alzaga, Bruno Odisio, and Reto Bale. Minimal ablative margin (MAM) assessment with image fusion: an independent predictor for local tumor progression in hepatocellular carcinoma after stereotactic radiofrequency ablation. *European Radiology*, 30(5):2463–2472, 2020.
- [111] Marta Burrel, María Reig, Alejandro Forner, Marta Barrufet, Carlos Rodríguez De Lope, Silvia Tremosini, Carmen Ayuso, Josep M. Llovet, María Isabel Real, and Jordi Bruix. Survival of patients with hepatocellular carcinoma treated by transarterial chemoembolisation (TACE) using Drug Eluting Beads. Implications for clinical practice and trial design. *Journal of Hepatology*, 56(6):1330–1335, 2012.
- [112] E A Eisenhauer, P Therasse, J Bogaerts, L H Schwartz, D Sargent, R Ford, J Dancey, S Arbuck, S Gwyther, M Mooney, L Rubinstein, L Shankar, and L Dodd. New response evaluation criteria in solid tumours : Revised RECIST guideline (version 1 . 1). *European Journal of Cancer*, 45(2):228–247, 2008.
- [113] An Tang, Roger Tam, Alexandre Cadrin-Chênevert, Will Guest, Jaron Chong, Joseph Barfett, Leonid Chepelev, Robyn Cairns, J. Ross Mitchell, Mark D. Cicero, Manuel Gaudreau Poudrette, Jacob L. Jaremko, Caroline Reinhold, Benoit Gallix, Bruce Gray, Raym Geis, Timothy O’Connell, Paul Babyn, David Koff, Darren Ferguson, Sheldon Derkatch, Alexander Bilbily, and Wael Shabana. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Canadian Association of Radiologists Journal*, 69(2):120–135, 2018.
- [114] Antonios Drevelegas. Imaging of brain tumors with histological correlations. *Imaging of Brain Tumors with Histological Correlations*, pages 1–432, 2011.
- [115] Mu Zhou, Lawrence Hall, Dmitry Goldgof, Robin Russo, Yoganand Balagurunathan, Robert Gillies, and Robert Gatenby. Radiologically Defined Ecological Dynamics and Clinical Outcomes in Glioblastoma Multiforme: Preliminary Results. *Translational Oncology*, 7(1):5–13, 2014.

- [116] Yee Liang Thian, Angela M. Riddell, and Dow Mu Koh. Liver-specific agents for contrast-enhanced MRI: Role in oncological imaging. *Cancer Imaging*, 13(4):567–579, 2013.
- [117] Fergus Davnall, Connie S P Yip, Gunnar Ljungqvist, Mariyah Selmi, Francesca Ng, Bal Sanghera, Balaji Ganeshan, Kenneth A. Miles, Gary J. Cook, and Vicky Goh. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? *Insights into Imaging*, 3(6):573–589, 2012.
- [118] Bas B Koolen, Marie-jeanne T F D Vrancken Peeters, and Tjeerd S Aukema. 18F-FDG PET / CT as a staging procedure in primary stage II and III breast cancer : comparison with conventional imaging techniques. pages 117–126, 2012.
- [119] Kunio Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4-5):198–211, jun 2007.
- [120] C. Carl Jaffe. Measures of response: RECIST, WHO, and new alternatives. *Journal of Clinical Oncology*, 24(20):3245–3251, 2006.
- [121] Parnian Afshar, Arash Mohammadi, Konstantinos N. Plataniotis, Anastasia Oikonomou, and Habib Benali. From Hand-Crafted to Deep Learning-based Cancer Radiomics: Challenges and Opportunities. 2018.
- [122] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*, 278(2):563–577, 2016.
- [123] Ying Liu, Jongphil Kim, Yoganand Balagurunathan, Qian Li, Alberto L. Garcia, Olya Stringfield, Zhaoxiang Ye, and Robert J. Gillies. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clinical Lung Cancer*, 17(5):441–448.e6, sep 2016.
- [124] Madeleine Scrivener, Evelyn E. C. de Jong, Janna E. van Timmeren, Thierry Pieters, Benoît Ghaye, and Xavier Geets. Radiomics applied to lung cancer: a review. *Translational Cancer Research*, 5(4):398–409, 2016.
- [125] Roberto Berenguer, María Del Rosario Pastor-Juan, Jesús Canales-Vázquez, Miguel Castro-García, María Victoria Villas, Francisco Mansilla Legorburo, and Sebastià Sabater. Radiomics of CT features may be nonreproducible and redundant: Influence of CT acquisition parameters. *Radiology*, 288(2):407–415, 2018.
- [126] Rajat Thawani, Michael McLane, Niha Beig, Soumya Ghose, Prateek Prasanna, Vamsidhar Velcheti, and Anant Madabhushi. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer*, 115(June 2017):34–41, 2018.

- [127] Shaimaa Bakr, Sebastian Echegaray, Rajesh Shah, Aya Kamaya, and John Louie. Noninvasive radiomics signature based on quantitative analysis of computed tomography images as a surrogate for microvascular invasion in hepatocellular carcinoma: a pilot study. *Journal of Medical Imaging*, 4(04):1, 2017.
- [128] Balaji Ganeshan, Elleny Panayiotou, Kate Burnand, Sabina Dizdarevic, and Ken Miles. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival. *European Radiology*, 22(4):796–802, 2012.
- [129] Yoganand Balagurunathan, Yuhua Gu, Hua Wang, Virendra Kumar, Olya Grove, Sam Hawkins, Jongphil Kim, Dmitry B. Goldgof, Lawrence O. Hall, Robert A. Gatenby, and Robert J. Gillies. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Translational Oncology*, 7(1):72–87, 2014.
- [130] Omar S. Al-Kadi and D. Watson. Texture analysis of aggressive and nonaggressive lung tumor CE CT images. *IEEE Transactions on Biomedical Engineering*, 55(7):1822–1830, 2008.
- [131] Yue Cao, Christina I. Tsien, Vijaya Nagesh, Larry Junck, Randall Ten Haken, Brian D. Ross, Thomas L. Chenevert, and Theodore S. Lawrence. Clinical investigation survival prediction in high-grade gliomas by MRI perfusion before and during early stage of RT. *International Journal of Radiation Oncology Biology Physics*, 64(3):876–885, 2006.
- [132] Vishwa Parekh and Michael A. Jacobs. Radiomics: a new application from established techniques. *Expert Review of Precision Medicine and Drug Development*, 1(2):207–226, 2016.
- [133] Anastasia Oikonomou, Farzad Khalvati, Pascal N. Tyrrell, Masoom A. Haider, Usman Tarique, Laura Jimenez-Juan, Michael C. Tjong, Ian Poon, Armin Eilaghi, Lisa Ehrlich, and Patrick Cheung. Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. *Scientific Reports*, 8(1):1–11, 2018.
- [134] Joost J.M. van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G.H. Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo J.W.L. Aerts. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21):e104–e107, nov 2017.
- [135] Taiga Wakabayashi, Farid Ouhmich, Cristians Gonzalez-Cabrera, Emanuele Felli, Antonio Saviano, Vincent Agnus, Peter Savadjiev, Thomas F. Baumert, Patrick Pessaux, Jacques Marescaux, and Benoit Gallix. Radiomics in hepatocellular carcinoma: a quantitative review, sep 2019.

- [136] Luca Cozzi, Nicola Dinapoli, Antonella Fogliata, Wei-Chung Hsu, Giacomo Reggiori, Francesca Lobefalo, Margarita Kirienko, Martina Sollini, Davide Franceschini, Tiziana Comito, Ciro Franzese, Marta Scorsetti, and Po-Ming Wang. Radiomics based analysis to predict local control and survival in hepatocellular carcinoma patients treated with volumetric modulated arc therapy. *BMC Cancer*, 17(1):829, 2017.
- [137] Ying Zhou, Lan He, Yanqi Huang, Shuting Chen, Penqi Wu, Weitao Ye, Zaiyi Liu, and Changhong Liang. CT-based radiomics signature: a potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma. *Abdominal Radiology*, 42(6):1695–1704, jun 2017.
- [138] H. Akai, K. Yasaka, A. Kunimatsu, M. Nojima, T. Kokudo, N. Kokudo, K. Hasegawa, O. Abe, K. Ohtomo, and S. Kiryu. Predicting prognosis of resected hepatocellular carcinoma by radiomics analysis with random survival forest. *Diagnostic and Interventional Imaging*, 99(10):643–651, 2018.
- [139] Shuting Chen, Yanjie Zhu, Zaiyi Liu, and Changhong Liang. Texture analysis of baseline multiphasic hepatic computed tomography images for the prognosis of single hepatocellular carcinoma after hepatectomy: A retrospective pilot study. *European Journal of Radiology*, 90:198–204, 2017.
- [140] Meng Li, Sirui Fu, Yanjie Zhu, Zaiyi Liu, Shuting Chen, Ligong Lu, and Changhong Liang. Computed tomography texture analysis to facilitate therapeutic decision making in hepatocellular carcinoma. *Oncotarget*, 7(11):13248–13259, 2016.
- [141] Siva P. Raman, James L. Schroeder, Peng Huang, Yifei Chen, Stephanie F. Coquia, Satomi Kawamoto, and Elliot K. Fishman. Preliminary Data Using Computed Tomography Texture Analysis for the Classification of Hypervascular Liver Lesions. *Journal of Computer Assisted Tomography*, 39(3):1, 2015.
- [142] Michael D. Kuo, Jeremy Gollub, Claude B. Sirlin, Clara Ooi, and Xin Chen. Radiogenomic Analysis to Identify Imaging Phenotypes Associated with Drug Response Gene Expression Programs in Hepatocellular Carcinoma. *Journal of Vascular and Interventional Radiology*, 18(7):821–830, 2007.
- [143] Sudeep Banerjee, David S. Wang, Hyun J. Kim, Claude B. Sirlin, Michael G. Chan, Ronald L. Korn, Aaron M. Rutman, Surachate Siripongsakun, David Lu, Galym Imanbayev, and Michael D. Kuo. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology*, 62(3):792–800, 2015.

- [144] Matteo Renzulli, Stefano Brocchi, Alessandro Cucchetti, Federico Mazzotti, Cristina Mosconi, Camilla Sportoletti, Giovanni Brandi, Antonio Daniele Pinna, and Rita Golfieri. Can Current Preoperative Imaging Be Used to Detect Microvascular Invasion of Hepatocellular Carcinoma? *Radiology*, 279(2):432–442, may 2016.
- [145] Eran Segal, Claude B. Sirlin, Clara Ooi, Adam S. Adler, Jeremy Gollub, Xin Chen, Bryan K. Chan, George R. Matcuk, Christopher T. Barry, Howard Y. Chang, and Michael D. Kuo. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nature Biotechnology*, 25(6):675–680, 2007.
- [146] Bo Hao Zheng, Long Zi Liu, Zhi Zhi Zhang, Jie Yi Shi, Liang Qing Dong, Ling Yu Tian, Zhen Bin Ding, Yuan Ji, Sheng Xiang Rao, Jian Zhou, Jia Fan, Xiao Ying Wang, and Qiang Gao. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients. *BMC cancer*, 18(1):1148, 2018.
- [147] Jie Peng, Jing Zhang, Qifan Zhang, Yikai Xu, Jie Zhou, and Li Liu. A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis b virus-related hepatocellular carcinoma. *Diagnostic and Interventional Radiology*, 24(3):121–127, 2018.
- [148] Bachir Taouli, Yujin Hoshida, Suguru Kakite, Xintong Chen, Poh Seng Tan, Xiaochen Sun, Shingo Kihira, Kensuke Kojima, Sara Toffanin, M. Isabel Fiel, Hadassa Hirschfield, Mathilde Wagner, and Josep M. Llovet. Imaging-based surrogate markers of transcriptome subclasses and signatures in hepatocellular carcinoma: preliminary results. *European Radiology*, 27(11):4472–4481, 2017.
- [149] Wei Xia, Ying Chen, Rui Zhang, Zhuangzhi Yan, Xiaobo Zhou, Bo Zhang, and Xin Gao. Radiogenomics of hepatocellular carcinoma: multiregion analysis-based identification of prognostic imaging biomarkers by integrating gene data—a preliminary study. *Physics in Medicine & Biology*, 63(3):035044, feb 2018.
- [150] Philippe Lambin, Ralph T.H. Leijenaar, Timo M. Deist, Jurgen Peerlings, Evelyn E.C. De Jong, Janita Van Timmeren, Sebastian Sanduleanu, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14(12):749–762, 2017.
- [151] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G P M Van Stiphout, Patrick Granton, Catharina M L Zegers, Robert Gillies, Ronald Boellard, André Dekker, and Hugo J W L Aerts. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, 2012.
- [152] Joost J.M. Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G.H. Beets-Tan, Jean Christophe Fillion-Robin, Steve Pieper, and

- Hugo J.W.L. Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107, 2017.
- [153] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(1995):60–88, 2017.
- [154] Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological Physics and Technology*, 10(3):257–273, 2017.
- [155] Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep Learning Applications in Medical Image Analysis. *IEEE Access*, 6:9375–9379, 2017.
- [156] Zilong Hu, Jinshan Tang, Ziming Wang, Kai Zhang, Lin Zhang, and Qingling Sun. Deep learning for image-based cancer detection and diagnosis - A survey. *Pattern Recognition*, 83:134–149, 2018.
- [157] Paolo G. Gobbi, Chiara Broglia, Giuseppe Di Giulio, Monica Mantelli, Paola Anselmo, Francesco Merli, Pier L. Zinzani, Gabriele Rossi, Vincenzo Callea, Emilio Iannitto, Marco Paulli, Lorena Garioni, and Edoardo Ascari. The clinical value of tumor burden at diagnosis in Hodgkin lymphoma. *Cancer*, 101(8):1824–1834, 2004.
- [158] Lars Bornemann, Volker Dicken, Jan Martin Kuhnigk, Dag Wormanns, Hoen Oh Shin, Hans Christian Bauknecht, Volker Diehl, Michael Fabel, Stefan Meier, Oliver Kress, Stefan Krass, and Heinz Otto Peitgen. OncoTREAT: A software assistant for cancer therapy monitoring. *International Journal of Computer Assisted Radiology and Surgery*, 1(5):231–242, 2007.
- [159] Claus Peter Heußel, S. Meier, S. Wittelsberger, H. Götte, P. Mildenberger, and H. U. Kauczor. Quantitative CT-verlaufskontrolle von lebermalignomen nach RECIST und WHO im vergleich zur volumetrie. *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgebenden Verfahren*, 179(9):958–964, 2007.
- [160] Jan Martin Kuhnigk, Volker Dicken, Lars Bornemann, Annemarie Bakai, Dag Wormanns, Stefan Krass, and Heinz Otto Peitgen. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Transactions on Medical Imaging*, 25(4):417–434, 2006.
- [161] Michael Puesken, Boris Buerke, Joachim Gerss, Barbara Frisch, Florian Beyer, Matthias Weckesser, Harald Seifarth, Walter Heindel, and Johannes Wessling. Prediction of lymph node manifestations in malignant lymphoma: Significant role of volumetric compared with established metric lymph

- node analysis in multislice computed tomography. *Journal of Computer Assisted Tomography*, 34(4):564–569, 2010.
- [162] Hans Christian Bauknecht, Valentina C. Romano, Patrik Rogalla, Randolph Klingebiel, Claudia Wolf, Lars Bornemann, Bernd Hamm, and Patrick A. Hein. Intra-and interobserver variability of linear and volumetric measurements of brain metastases using contrast-enhanced magnetic resonance imaging. *Investigative Radiology*, 45(1):49–56, 2010.
- [163] Jie Zhi Cheng, Dong Ni, Yi Hong Chou, Jing Qin, Chui Mei Tiu, Yeun Chung Chang, Chiun Sheng Huang, Dinggang Shen, and Chung Ming Chen. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Scientific Reports*, 6(March):1–13, 2016.
- [164] Wenqing Sun, Bin Zheng, and Wei Qian. Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Computers in Biology and Medicine*, 89:530–539, 2017.
- [165] Wei Shen, Mu Zhou, Feng Yang, Di Dong, Caiyun Yang, Yali Zang, and Jie Tian. Learning from Experts: Developing Transferable Deep Features for Patient-Level Lung Cancer Prediction. volume 8150, pages 124–131. 2016.
- [166] Devinder Kumar, Audrey G. Chung, Mohammad J. Shaifee, Farzad Khalvati, Masoom A. Haider, and Alexander Wong. Discovery Radiomics for Pathologically-Proven Computed Tomography Lung Cancer Prediction. volume 1, pages 54–62. 2017.
- [167] Amir Jamaludin, Timor Kadir, and Andrew Zisserman. SpineNet: Automatically pinpointing classification evidence in spinal MRIs. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9901 LNCS:166–175, 2016.
- [168] Oier Echaniz and Manuel Graña. Ongoing Work on Deep Learning for Lung Cancer Prediction. volume 10338 of *Lecture Notes in Computer Science*, pages 42–48. Springer International Publishing, Cham, 2017.
- [169] Benjamin Q. Huynh, Hui Li, and Maryellen L. Giger. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3):034501, 2016.
- [170] Rahul Paul, Samuel H Hawkins, Yoganand Balagurunathan, Matthew B Schabath, Robert J Gillies, Lawrence O Hall, and Dmitry B Goldgof. Deep Feature Transfer Learning in Combination with

- Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma. *Tomography*, 2(4), 2016.
- [171] Zeju Li, Yuanyuan Wang, Jinhua Yu, Yi Guo, and Wei Cao. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Scientific Reports*, 7(1):1–11, 2017.
- [172] Shekoofeh Azizi, Sharareh Bayat, Pingkun Yan, Amir Tahmasebi, Jin Tae Kwak, Sheng Xu, Baris Turkbey, Peter Choyke, Peter Pinto, Bradford Wood, Parvin Mousavi, and Purang Abolmaesumi. Deep recurrent neural networks for prostate cancer detection: Analysis of temporal enhanced ultrasound. *IEEE Transactions on Medical Imaging*, 37(12):2695–2703, 2018.
- [173] Bum Chae Kim, Yu Sub Sung, and Heung Il Suk. Deep feature learning for pulmonary nodule classification in a lung CT. *4th International Winter Conference on Brain-Computer Interface, BCI 2016*, pages 2–4, 2016.
- [174] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, 101(1):569–582, nov 2014.
- [175] Nastaran Emaminejad, Wei Qian, Yubao Guan, Maxine Tan, Yuchen Qiu, Hong Liu, and Bin Zheng. Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients. *IEEE Transactions on Biomedical Engineering*, 63(5):1034–1043, 2016.
- [176] Akira Yamada, Kazuki Oyama, Sachie Fujita, Eriko Yoshizawa, Fumihiro Ichinohe, Daisuke Komatsu, and Yasunari Fujinaga. Dynamic contrast-enhanced computed tomography diagnosis of primary liver cancers using transfer learning of pretrained convolutional neural networks: Is registration of multiphasic images necessary? *International Journal of Computer Assisted Radiology and Surgery*, 14(8):1295–1301, 2019.
- [177] Weibin Wang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, Qingqing Chen, Dong Liang, Lanfen Lin, Hongjie Hu, and Qiaowei Zhang. Classification of Focal Liver Lesions Using Deep Learning with Fine-Tuning. In *Proceedings of the 2018 International Conference on Digital Medicine and Image Processing - DMIP '18*, pages 56–60, New York, New York, USA, 2018. ACM Press.
- [178] Koichiro Yasaka, Hiroyuki Akai, Osamu Abe, and Shigeru Kiryu. Deep learning with CNN showed high diagnostic performance in differentiation of liver masses at dynamic CT. *Radiology*, 286(3—March):887–896, 2018.

- [179] Dong Liang, Lanfen Lin, Hongjie Hu, Qiaowei Zhang, Qingqing Chen, Yutaro Iwamoto, Xianhua Han, and Yen-Wei Chen. Combining Convolutional and Recurrent Neural Networks for Classification of Focal Liver Lesions in Multi-phase CT Images. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11071 LNCS, pages 666–675. 2018.
- [180] Koichiro Yasaka, Hiroyuki Akai, Akira Kunimatsu, Osamu Abe, and Shigeru Kiryu. Deep learning for staging liver fibrosis on CT: a pilot study. *European Radiology*, 28(11):4578–4585, 2018.
- [181] Weibin Wang, Qingqing Chen, Yutaro Iwamoto, Xianhua Han, Qiaowei Zhang, Hongjie HU, Lanfen LIN, and Yen-Wei Chen. Deep Learning-Based Radiomics Models for Early Recurrence Prediction of Hepatocellular Carcinoma with Multi-phase CT Images and Clinical Data. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4881–4884. IEEE, jul 2019.
- [182] Jie Peng, Shuai Kang, Zhengyuan Ning, Hangxia Deng, Jingxian Shen, Yikai Xu, Jing Zhang, Wei Zhao, Xinling Li, Wuxing Gong, Jinhua Huang, and Li Liu. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *European Radiology*, 30(1):413–424, 2020.
- [183] Da-wei Yang, Xi-bin Jia, Yu-jie Xiao, Xiao-pei Wang, Zhen-chang Wang, and Zheng-han Yang. Noninvasive Evaluation of the Pathologic Grade of Hepatocellular Carcinoma Using MCF-3DCNN: A Pilot Study. *BioMed Research International*, 2019:1–12, apr 2019.
- [184] Xiao Dan Ye, Zuguo Yuen, Jian Zhang, and Zheng Yuan. Radiological biomarkers for assessing response to locoregional therapies in hepatocellular carcinoma: From morphological to functional imaging (Review). *Oncology Reports*, 37(3):1337–1346, 2017.
- [185] A. Al-Nahhas, T. Szyszko, P. Tait, O. Damrah, and R. Canelo. Selective Internal Radiation therapy. In *Liver and Biliary Tract Surgery*, volume 10, pages 409–418. Springer Vienna, Vienna, nov 2014.
- [186] R. Yamada, M. Sato, M. Kawabata, H. Nakatsuka, K. Nakamura, and S. Takashima. Hepatic artery embolization in 120 patients with unresectable hepatoma. *Radiology*, 148(2):397–401, 1983.
- [187] Kathy S. Albain, R. Suzanne Swann, Valerie W. Rusch, Andrew T. Turrisi, Frances A. Shepherd, Colum Smith, Yuhchyau Chen, Robert B. Livingston, Richard H. Feins, David R. Gandara, Willard A. Fry, Gail Darling, David H. Johnson, Mark R. Green, Robert C. Miller, Joanne Ley, William T. Sause, and James D. Cox. Radiotherapy plus chemotherapy with or without surgical

- resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial. *The Lancet*, 374(9687):379–386, 2009.
- [188] S Rossi, M Di Stasi, E. Buscarini, P Quaretti, F Garbagnati, L Squassante, C T Paties, D E Silverman, and L Buscarini. Percutaneous RF interstitial thermal ablation in the treatment of hepatic cancer. *American Journal of Roentgenology*, 167(3):673–759, sep 1996.
- [189] Bernard Nordlinger, Marguerite Guiguet, Jean Christophe Vaillant, Pierre Balladur, Karim Boudjema, Philippe Bachellier, and Daniel Jaeck. Surgical resection of colorectal carcinoma metastases to the liver: A prognostic scoring system to improve case selection, based on 1568 patients. *Cancer*, 77(7):1254–1262, 1996.
- [190] S. Jagannath, W. S. Velasquez, S. L. Tucker, L. M. Fuller, P. W. McLaughlin, J. T. Manning, L. B. North, and F. C. Cabanillas. Tumor burden assessment and its implication for a prognostic model in advanced diffuse large-cell lymphoma. *Journal of Clinical Oncology*, 4(6):859–865, 1986.
- [191] Patrick Bilic, Patrick Ferdinand Christ, Eugene Vorontsov, Grzegorz Chlebus, Hao Chen, Qi Dou, Chi-Wing Fu, Xiao Han, Pheng-Ann Heng, Jürgen Hesser, Samuel Kadoury, Tomasz Konopczynski, Miao Le, Chunming Li, Xiaomeng Li, Jana Lipkovà, John Lowengrub, et al. The Liver Tumor Segmentation Benchmark (LiTS). pages 1–43, 2019.
- [192] Tobias Heimann, B. van Ginneken, M.A. Styner, Yulia Arzhaeva, Volker Aurich, Christian Bauer, Andreas Beck, Christoph Becker, Reinhard Beichel, et al. Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, aug 2009.
- [193] Sebastian Echegaray, Olivier Gevaert, Rajesh Shah, Aya Kamaya, John Louie, Nishita Kothary, and Sandy Napel. Core samples for radiomics features that are insensitive to tumor segmentation: method and pilot study using CT images of hepatocellular carcinoma. *Journal of Medical Imaging*, 2(4):041011, 2015.
- [194] Jan Hendrik Moltz, Lars Bornemann, Jan Martin Kuhnigk, Volker Dicken, Elena Peitgen, Stephan Meier, Hendrik Bolte, Michael Fabel, Hans Christian Bauknecht, Markus Hittinger, Andreas Kießling, Michael Pusken, and Heinz Otto Peitgen. Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans. *IEEE Journal on Selected Topics in Signal Processing*, 3(1):122–134, 2009.
- [195] Akshat Gotra, Lojan Sivakumaran, Gabriel Chartrand, Kim Nhien Vu, Franck Vandenbroucke-Menu, Claude Kauffmann, Samuel Kadoury, Benoît Gallix, Jacques A. de Guise, and An Tang.

- Liver segmentation: indications, techniques and future directions. *Insights into Imaging*, 8(4):377–392, 2017.
- [196] Francesco Rizzetto, Francesca Calderoni, Cristina De Mattia, Arianna Defeudis, Valentina Giannini, Simone Mazzetti, Lorenzo Vassallo, Silvia Ghezzi, Andrea Sartore-Bianchi, Silvia Marsoni, Salvatore Siena, Daniele Regge, Alberto Torresin, and Angelo Vanzulli. Impact of inter-reader contouring variability on textural radiomics of colorectal liver metastases. *European Radiology Experimental*, 4(1), 2020.
- [197] Grand challenge. <https://grand-challenge.org>. Accessed: 2020-10-20.
- [198] 3dircadb. <https://www.ircad.fr/research/3d-ircadb-01/>. Accessed: 2020-10-20.
- [199] Midas. <http://insight-journal.org/midas/collection/view/38>. Accessed: 2020-10-20.
- [200] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, Dec 2013.
- [201] Bram van Ginneken, Tobias Heimann, and Martin Styner. 3D segmentation in the clinic: A grand challenge. *International Conference on Medical Image Computing and Computer Assisted Intervention*, 10:7–15, 2007.
- [202] Imageclef. <https://www.imageclef.org/2015/liver#access>. Accessed: 2020-10-20.
- [203] Chaos. https://chaos.grand-challenge.org/Combined_Healthy_Abdominal_Organ_Segmentation/. Accessed: 2020-10-20.
- [204] Mehrdad Moghbel, Syamsiah Mashohor, Rozi Mahmud, and M. Iqbal Bin Saripan. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artificial Intelligence Review*, 50(4):497–537, 2018.
- [205] Seong-Jae Lim, Yong-Yeon Jeong, Chil-Woo Lee, and Yo-Sung Ho. Automatic segmentation of the liver in CT images using the watershed algorithm based on morphological filtering. *Medical Imaging 2004: Image Processing*, 5370(December 2015):1658, 2004.
- [206] Seong-Jae Lim, Yong-Yeon Jeong, and Yo-Sung Ho. Segmentation of the Liver Using the Deformable Contour Method on CT Images. In Yo-Sung Ho and Hyoung Joong Kim, editors, *Advances in Multimedia Information Processing - PCM 2005*, volume 3767 of *Lecture Notes in Computer Science*, pages 570–581. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

- [207] Seong Jae Lim, Yong Yeon Jeong, and Yo Sung Ho. Automatic liver segmentation for volume measurement in CT Images. *Journal of Visual Communication and Image Representation*, 17(4):860–875, 2006.
- [208] Chien Cheng Lee, Pau Choo Chung, and Hong Ming Tsai. Identifying Multiple Abdominal Organs from CT Image Series Using a Multimodule Contextual Neural Network and Spatial Fuzzy Rules. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):208–217, 2003.
- [209] Fan Liu, Binsheng Zhao, Peter K. Kijewski, Liang Wang, and Lawrence H. Schwartz. Liver segmentation for CT images using GVF snake. *Medical Physics*, 32(12):3699–3706, 2005.
- [210] Kyung Sik Seo and Jong An Park. Improved automatic liver segmentation of a contrast enhanced CT image. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3767 LNCS:899–909, 2005.
- [211] Pil Un Kim, Yun-Jung Lee, Youngjin Jung, Jin Ho Cho, and Myoung Nam Kim. Liver extraction in the abdominal CT image by watershed segmentation algorithm. *IFMBE Proceedings*, 14(1):2563–2566, 2007.
- [212] Paola Campadelli, Elena Casiraghi, and Andrea Esposito. Liver segmentation from computed tomography scans: A survey and a new algorithm. *Artificial Intelligence in Medicine*, 45(2-3):185–196, 2009.
- [213] Amir H. Foruzan, Reza Aghaeizadeh Zoroofi, Masatoshi Hori, and Yoshinobu Sato. Liver segmentation by intensity analysis and anatomical information in multi-slice CT images. *International Journal of Computer Assisted Radiology and Surgery*, 4(3):287–297, 2009.
- [214] Andreas Beck and Volker Aurich. Hepatux - a semiautomatic liver segmentation system. *3D Segmentation in The Clinic: A Grand Challenge*, pages 225–233, 2007.
- [215] Yingyi Qi, Wei Xiong, Wee-Kheng Leow, Qi Tian, Jiayin Zhou, Jiang Liu, Thazin Han, Sudhakar K Venkatesh, and Shih-chang Wang. Semi-automatic segmentation of liver tumors from CT scans using Bayesian rule-based 3D region growing. *MICCAI Workshop*, 41(43):1–10, 2008.
- [216] L Rusko and G Bekes. Fully automatic liver segmentation for contrast-enhanced CT images. *International Conference on Medical Image Computing and Computer-Assisted Intervention. Segmentation in the Clinic: a grand challenge, 2007*, pages 143–150, 2007.
- [217] László Ruskó, György Bekes, and Márta Fidrich. Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images. *Medical Image Analysis*, 13(6):871–882, 2009.

- [218] Ruchaneewan Susomboon. A hybrid approach for liver segmentation. *Proc. MICCAI Workshop 3-D Segmentat. Clinic: A Grand, Challenge,,* i:151–160, 2007.
- [219] S. S. Kumar, R. S. Moni, and J. Rajeesh. Automatic liver and lesion segmentation: A primary step in diagnosis of liver diseases. *Signal, Image and Video Processing*, 7(1):163–172, 2013.
- [220] Regina Pohle and Klaus D. Toennies. Segmentation of medical images using adaptive region growing. In Milan Sonka and Kenneth M. Hanson, editors, *Medical Imaging 2001: Image Processing*, volume 4322, pages 1337–1346, jul 2001.
- [221] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic Active Contours. In *International Journal of Computer Vision*, volume 22, pages 61–79. 1997.
- [222] Stanley Osher and Ronald Fedkiw. *Signed Distance Functions*, volume 153. 2003.
- [223] Shiyan Pan and Benoit M. Dawant. Automatic 3d segmentation of the liver from abdominal ct images: a level-set approach. *Medical Imaging 2001: Image Processing*, 4322:128–138, 2001.
- [224] Jinke Wang, Yuanzhi Cheng, Changyong Guo, Yadong Wang, and Shinichi Tamura. Shape–intensity prior level set combining probabilistic atlas and probability map constrains for automatic liver segmentation from abdominal CT images. *International Journal of Computer Assisted Radiology and Surgery*, 11(5):817–826, 2016.
- [225] J. F. Garamendi, N. Malpica, J. Martel, and E. Schiavi. Automatic segmentation of the liver in CT using level sets without edges. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4477 LNCS(PART 1):161–168, 2007.
- [226] Ying Chi, Peter M M Cashman, Fernando Bello, and Richard I Kitney. A Discussion on the Evaluation of A New Automatic Liver Volume Segmentation Method for Specified CT Image Datasets. *Heart*, pages 167–175, 2007.
- [227] D Furukawa, A Shimizu, and H Kobatake. Automatic Liver Segmentation Method based on Maximum A Posterior Probability Estimation and Level Set Method, T. *3D Segmentation in The Clinic: A Grand Challenge*, pages 117–124, 2007.
- [228] Jeongjin Lee, Namkug Kim, Ho Lee, Joon Beom Seo, Hyung Jin Won, Yong Moon Shin, and Yeong Gil Shin. Efficient Liver Segmentation exploiting Level-Set Speed Images with 2 .5D Shape Propagation. *3D Segmentation in the Clinic: A Grand Challenge*, pages 189–196, 2007.

- [229] Andreas Wimmer, Grzegorz Soza, and Joachim Hornegger. Two-stage semi-automatic organ segmentation framework using radial basis functions and level sets. *3D segmentation in the clinic: a grand challenge, LNCS, Springer Berlin, Heidelberg*, 1(c):207–214, 2007.
- [230] Andreas Wimmer, Joachim Hornegger, and Grzegorz Soza. Implicit active shape model employing boundary classifier. *Proceedings - International Conference on Pattern Recognition*, (June 2014), 2008.
- [231] Laurent Massoptier and Sergio Casciaro. A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans. *European Radiology*, 18(8):1658–1665, 2008.
- [232] Andreas Wimmer, Grzegorz Soza, and Joachim Hornegger. Organ Segmentation. *Image (Rochester, N.Y.)*, pages 26–33, 2009.
- [233] Yi Song, Andy J. Bulpitt, and Ken W. Brodlie. Liver Segmentation Using Automatically Defined Patient Specific B-Spline Surface Models. In *Professional differences in interprofessional working*, volume 22, pages 43–50. 2009.
- [234] Kenji Suzuki, Ryan Kohlbrenner, Mark L. Epstein, Ademola M. Obajuluwa, Jianwu Xu, and Masatoshi Hori. Computer-aided measurement of liver volumes in CT by means of geodesic active contour segmentation coupled with level-set algorithms. *Medical Physics*, 37(5):2159–2166, 2010.
- [235] Carlos Platero, María Carmen Tobar, Javier Sanguino, José Manuel Poncela, and Olga Velasco. Level set segmentation with shape and appearance models using affine moment descriptors. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6669 LNCS:109–116, 2011.
- [236] Daniel Jimenez-Carretero, Laura Fernandez-De-Manuel, Javier Pascau, Jose M. Tellado, Enrique Ramon, Manuel Desco, Andres Santos, and Maria J. Ledesma-Carbayo. Optimal multiresolution 3D level-set method for liver segmentation incorporating local curvature constraints. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 3419–3422, 2011.
- [237] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts 1 Energy minimization in early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [238] Leo Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006.

- [239] F Maier, A Wimmer, G Soza, J N Kaftan, D Fritz, and R Dillmann. Automatic Liver Segmentation using the Random Walker Algorithm. *Proceedings des Workshops vom 6. bis 8. April 2008 in Berlin Bildverarbeitung für die Medizin 2008*, pages 56–61, 2008.
- [240] Chunhua Dong, Yen Wei Chen, Lanfen Lin, Hongjie Hu, Chongwu Jin, Huajun Yu, Xian Hua Han, and Tomoko Tateyama. Simultaneous segmentation of multiple organs using random walks. *Journal of Information Processing*, 24(2):320–329, 2016.
- [241] Reinhard Beichel, Thomas Pock, Christian Janko, Roman B. Zotter, Bernhard Reitinger, Alexander Bornik, Kalman Palagyi, Erich Sorantin, Georg Werkgartner, Horst Bischof, and Milan Sonka. Liver segment approximation in CT data for surgical resection planning. 5370(316):1435, 2004.
- [242] Laurent Massoptier and Sergio Casciaro. Fully automatic liver segmentation through graph-cut technique. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pages 5243–5246, 2007.
- [243] Akinobu Shimizu, Keita Nakagomi, Takuya Narihira, Hidefumi Kobatake, Shigeru Nawano, Kenji Shinozaki, Koich Ishizu, and Kaori Togashi. Automated Segmentation of 3D CT Images Based on Statistical Atlas and Graph Cuts. In Bjoern Menze, Georg Langs, Zhuowen Tu, and Antonio Criminisi, editors, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 6533 of *Lecture Notes in Computer Science*, pages 214–223. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [244] Xiangrong Zhou, Teruhiko Kitagawa, Takeshi Hara, Hiroshi Fujita, Xuejun Zhang, Ryujiro Yokoyama, Hiroshi Kondo, Masayuki Kanematsu, and Hiroaki Hoshi. Constructing a probabilistic model for automated liver region segmentation using non-contrast X-ray torso CT images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4191 LNCS:856–863, 2006.
- [245] Hyunjin Park, Peyton H Bland, and Charles R Meyer. Construction of an abdominal probabilistic atlas and its application in segmentation. 22(4):483–492, 2003.
- [246] Pieter Slagmolen, An Elen, Dieter Seghers, Dirk Loeckx, Frederik Maes, and Karin Haustermans. Atlas based liver segmentation using nonrigid registration with a B-spline transformation model. *Proceedings of MICCAI workshop on 3D segmentation in the clinic: a grand challenge*, pages 197–206, 2007.
- [247] E. M. Rikxoort, Yulia Arzhaeva, and B. Ginneken. Automatic segmentation of the liver in computed tomography scans with voxel classification and atlas matching. 2007.

- [248] X. Zhou, T. Kitagawa, K. Okuo, T. Hara, H. Fujita, R. Yokoyama, M. Kanematsu, and H. Hoshi. Construction of a probabilistic atlas for automated liver segmentation in non-contrast torso CT images. *International Congress Series*, 1281:1169–1174, 2005.
- [249] Chang Yang Li, Xiuying Wang, Stefan Eberl, Michael Fulham, Yong Yin, and Dagan Feng. Fully automated liver segmentation for low- and high-contrast ct volumes based on probabilistic atlases. *Proceedings - International Conference on Image Processing, ICIP*, pages 1733–1736, 2010.
- [250] Marius George Linguraru, Zhixi Li, Furhawn Shah, See Chin, and Ronald M. Summers. Automated liver segmentation using a normalized probabilistic atlas. *Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 7262(February 2009):72622R, 2009.
- [251] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application, 1995.
- [252] J. Montagnat and H. Delingette. Volumetric medical images segmentation using shape constrained deformable models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 1205, pages 13–22. 1997.
- [253] H Lamecker, T Lange, and M Seebass. Segmentation of the liver using a 3D statistical shape model. *Konrad-Zuse-Zentrum für Informationstechnik Berlin*, 09(April):27, 2004.
- [254] Tobias Heimann, Sascha Münzing, Hans Peter Meinzer, and Ivo Wolf. A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4584 LNCS:1–12, 2007.
- [255] Kinda Anna Saddi and Christophe Chefd. Global-to-Local Shape Matching for Liver Segmentation in CT Imaging. *MICCAI (October 2007)*, pages 207–214, 2007.
- [256] Xing Zhang, Jie Tian, Kexin Deng, Yongfang Wu, and Xiuli Li. Automatic liver segmentation using a statistical shape model with optimal surface detection. *IEEE Transactions on Biomedical Engineering*, 57(10 PART 2):2622–2626, 2010.
- [257] Haibin Ling, S. Kevin Zhou, Yefeng Zheng, Bogdan Georgescu, Michael Suehling, and Dorin Comaniciu. Hierarchical, learning-based automatic liver segmentation. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.
- [258] Yefeng Zheng, Adrian Barbu, Bogdan Georgescu, Michael Scheuering, and Dorin Comaniciu. Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and

- steerable features. *Proceedings of the IEEE International Conference on Computer Vision*, pages 1–8, 2007.
- [259] Dieter Seghers, Pieter Slagmolen, Yves Lambelin, Jeroen Hermans, Dirk Loeckx, Frederik Maes, and Paul Suetens. Landmark based liver segmentation using local shape and local intensity models. *Proc. Workshop of the 10th Int. Conf. on MICCAI, Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, (May 2014):135–142, 2007.
- [260] Marius Erdt, Sebastian Steger, Matthias Kirschner, and Stefan Wesarg. Fast automatic liver segmentation combining learned shape priors with observed shape deviation. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, pages 249–254, 2010.
- [261] Dagmar Kainmueller, Thomas Lange, and Hans Lamecker. Shape constrained automatic segmentation of the liver based on a heuristic intensity model. *Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge*, pages 109–116, 2007.
- [262] Toshiyuki Okada, Ryuji Shimada, Masatoshi Hori, Masahiko Nakamoto, Yen Wei Chen, Hironobu Nakamura, and Yoshinobu Sato. Automated Segmentation of the Liver from 3D CT Images Using Probabilistic Atlas and Multilevel Statistical Shape Model. *Academic Radiology*, 15(11):1390–1403, 2008.
- [263] Du Yih Tsai and Nobutaka Tanahashi. Neural-network-based boundary detection of liver structure in CT images for 3-D visualization. *IEEE International Conference on Neural Networks - Conference Proceedings*, 6:3484–3489, 1994.
- [264] Luomin Gao, David G. Heath, Brian S. Kuszyk, and Elliot K. Fishman. Automatic liver segmentation technique for three-dimensional visualization of CT data. *Radiology*, 201(2):359–364, 1996.
- [265] Günter Schmidt, Maria Athelogou, Ralf Schönmyer, Rene Korn, and Gerd Binnig. Cognition network technology for a fully automated 3d segmentation of liver. *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, pages 125–133, 2007.
- [266] Moti Freiman, Ofer Eliassaf, Yoav Taieb, Leo Joskowicz, and Jacob Sosna. A Bayesian approach for liver analysis: Algorithm and validation study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5241 LNCS(PART 1):85–92, 2008.

- [267] Moti Freiman, Ofir Cooper, Dani Lischinski, and Leo Joskowicz. Liver tumors segmentation from CTA images using voxels classification and affinity constraint propagation. *International Journal of Computer Assisted Radiology and Surgery*, 6(2):247–255, 2011.
- [268] Charles Florin, Nikos Paragios, Gareth Funka-Lea, and James Williams. Liver segmentation using sparse 3D prior models with optimal data support. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4584 LNCS(June 2014):38–49, 2007.
- [269] Andrea Schenk, Guido Prause, and Heinz Otto Peitgen. Efficient semiautomatic segmentation of 3D objects in medical images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1935:186–195, 2000.
- [270] Ahmed S. Maklad, Mikio Matsuhiro, Hidenobu Suzuki, Yoshiki Kawata, Noboru Niki, Noriyuki Moriyama, Toru Utsunomiya, and Mitsuo Shimada. Blood vessel-based liver segmentation through the portal phase of a CT dataset. *Medical Imaging 2013: Computer-Aided Diagnosis*, 8670(December 2013):86700X, 2013.
- [271] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. DeGuise. SEMI-automated liver CT segmentation using Laplacian meshes. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 641–644. IEEE, apr 2014.
- [272] Mohammed Goryawala, Seza Gulec, Ruchir Bhatt, Anthony J. McGoron, and Malek Adjouadi. A low-interaction automatic 3D liver segmentation method using computed tomography for selective internal radiation therapy. *BioMed Research International*, 2014(1), 2014.
- [273] Dengwang Li, Li Liu, Jinhu Chen, Hongsheng Li, and Yong Yin. A multistep liver segmentation strategy by combining level set based method with texture analysis for CT images. *IEEE International Conference on Orange Technologies, ICOT 2014*, pages 109–112, 2014.
- [274] Abdalla Mostafa, Ahmed Fouad, Mohamed Abd Elfattahc, Aboul Ella Hassanien, Hesham Hefny, Shao Ying Zhu, and Gerald Schaefer. CT liver segmentation using artificial bee colony optimisation. *Procedia Computer Science*, 60(1):1622–1630, 2015.
- [275] Changfa Shi, Yuanzhi Cheng, Fei Liu, Yadong Wang, Jing Bai, and Shinichi Tamura. A hierarchical local region-based sparse shape composition for liver segmentation in CT scans. *Pattern Recognition*, 50:88–106, 2016.

- [276] Saif Dawood Salman Al-Shaikhli, Michael Ying Yang, and Bodo Rosenhahn. Automatic 3D Liver Segmentation Using Sparse Representation of Global and Local Image Information via Level Set Formulation. pages 1–10, 2015.
- [277] Zhoubing Xu, Ryan P. Burke, Christopher P. Lee, Rebeccah B. Baucom, Benjamin K. Poulose, Richard G. Abramson, and Bennett A. Landman. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Medical Image Analysis*, 24(1):18–27, aug 2015.
- [278] Guotai Wang, Shaoting Zhang, Hongzhi Xie, Dimitris N. Metaxas, and Lixu Gu. A homotopy-based sparse representation for fast and accurate shape prior modeling in liver surgical planning. *Medical Image Analysis*, 19(1):176–186, 2015.
- [279] Cheng Huang, Xuhui Li, and Fucang Jia. Automatic liver segmentation using multiple prior knowledge models and free-form deformation. *CEUR Workshop Proceedings*, 1194:22–24, 2014.
- [280] Ahmed M. Anter, Aboul Ella Hassanien, Mohamed A. Abu ElSoud, and Mohamed F. Tolba. Neutrosophic Sets and Fuzzy C-Means Clustering for Improving CT Liver Image Segmentation. In Pavel Kömer, Ajith Abraham, and Václav Snášel, editors, *Advances in Intelligent Systems and Computing*, volume 303 of *Advances in Intelligent Systems and Computing*, pages 193–203. Springer International Publishing, Cham, 2014.
- [281] Pierre-Henri Conze, Vincent Noblet, François Rousseau, Fabrice Heitz, Vito de Blasi, Riccardo Memmo, and Patrick Pessaux. Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic contrast-enhanced CT scans. *International Journal of Computer Assisted Radiology and Surgery*, 12(2):223–233, 2017.
- [282] Avi Ben-Cohen, Idit Diamant, Eyal Klang, and Michal Amitai. Deep Learning and Data Labeling for Medical Applications. 10008:77–85, 2016.
- [283] Miriam Bellver, Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Xavier Giro-i Nieto, Jordi Torres, and Luc Van Gool. Detection-aided liver lesion segmentation using deep learning. (Nips), 2017.
- [284] Eugene Vorontsov, An Tang, Chris Pal, and Samuel Kadoury. Liver lesion segmentation informed by joint liver segmentation. *Proceedings - International Symposium on Biomedical Imaging*, 2018-April:1332–1335, 2018.
- [285] Yading Yuan. Hierarchical Convolutional-Deconvolutional Neural Networks for Automatic Liver and Tumor Segmentation. i:3–6, 2017.

- [286] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. pages 1–8, 2015.
- [287] Xiao Han. Automatic Liver Lesion Segmentation Using A Deep Convolutional Neural Network Method. 2017.
- [288] Grzegorz Chlebus, Andrea Schenk, Jan Hendrik Moltz, Bram van Ginneken, Horst Karl Hahn, and Hans Meine. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Scientific Reports*, 8(1):1–7, 2018.
- [289] Lei Bi, Jinman Kim, Ashnil Kumar, and Dagan Feng. Automatic Liver Lesion Detection using Cascaded Deep Residual Networks. 2017.
- [290] Krishna Chaitanya Kaluva, Mahendra Khened, Avinash Kori, and Ganapathy Krishnamurthi. 2D-Densely Connected Convolution Neural Networks for automatic Liver and Tumor Segmentation. X(X):1–4, 2018.
- [291] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng Ann Heng. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Transactions on Medical Imaging*, (1):1–13, sep 2017.
- [292] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *Enzyme and Microbial Technology*, 19(2):107–117, dec 2015.
- [293] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:2261–2269, 2017.
- [294] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A. Elshaera, Jana Lipkova, Sebastian Schlecht, Freba Ahmaddy, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Felix Hofmann, Melvin D Anastasi, Seyed-Ahmad Ahmadi, Georgios Kaassis, Julian Holch, Wieland Sommer, Rickmer Braren, Volker Heinemann, and Bjoern Menze. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. pages 1–20, 2017.
- [295] Wen Li, Fucang Jia, and Qingmao Hu. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *Journal of Computer and Communications*, 03(11):146–151, 2015.

- [296] Maayan Frid-adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. *Patch-Based Techniques in Medical Imaging*, volume 10530 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham, 2017.
- [297] Shima Rafiei, Ebrahim Nasr-Esfahani, Kayvan Najarian, Nader Karimi, Shadrokh Samavi, and S. M.Reza Soroushmehr. Liver Segmentation in CT Images Using Three Dimensional to Two Dimensional Fully Convolutional Network. *Proceedings - International Conference on Image Processing, ICIP*, pages 2067–2071, 2018.
- [298] Qi Dou, Hao Chen, Yueming Jin, Lequan Yu, Jing Qin, and Pheng-Ann Heng. 3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9901 LNCS, pages 149–157. jul 2016.
- [299] Changjian Sun, Shuxu Guo, Huimao Zhang, Jing Li, Meimei Chen, Shuzhi Ma, Lanyi Jin, Xiaoming Liu, Xueyan Li, and Xiaohua Qian. Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs. *Artificial Intelligence in Medicine*, 83:58–66, 2017.
- [300] Farid Ouhmich, Vincent Agnus, Vincent Noblet, Fabrice Heitz, and Patrick Pessaux. Liver tissue segmentation in multiphase CT scans using cascaded convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, 2019.
- [301] Simon K. Warfield, Kelly H. Zou, and William M. Wells. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.
- [302] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, feb 2008.
- [303] Fan Zhang, Junlin Yang, Nariman Nezami, Fabian Laage-gaupp, Julius Chapiro, Ming De Lin, and James Duncan. Liver Tissue Classification Using an Auto-context-based Deep Neural Network with a Multi-phase Training Framework. In Wenjia Bai, Gerard Sanroma, Guorong Wu, Brent C Munsell, Yiqiang Zhan, and Pierrick Coupé, editors, *Patch-Based Techniques in Medical Imaging*, pages 59–66, Cham, 2018. Springer International Publishing.
- [304] Fan Zhang, Junlin Yang, Nariman Nezami, Fabian Laage-Gaupp, Julius Chapiro, Ming De Lin, and James Duncan. Liver tissue classification using an auto-context-based deep neural network

- with a multi-phase training framework. In Wenjia Bai, Gerard Sanroma, Guorong Wu, Brent C Munsell, Yiqiang Zhan, and Pierrick Coupé, editors, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11075 LNCS, pages 59–66, Cham, 2018. Springer International Publishing.
- [305] Brian B Avants and Nick Tustison. Advanced normalization tools (ants).
- [306] Shengyu Zhao, Yue Dong, Eric I.Chao Chang, and Yan Xu. Recursive Cascaded Networks for Unsupervised Medical Image Registration. *arXiv*, pages 10600–10610, 2019.
- [307] Shengyu Zhao, Tingfung Lau, Ji Luo, Eric I.Chao Chang, and Yan Xu. Unsupervised 3D End-to-End Medical Image Registration with Volume Tweening Network. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1394–1404, 2020.
- [308] Daisuke Okamoto, Kengo Yoshimitsu, Akihiro Nishie, Tsuyoshi Tajima, Yoshiki Asayama, Kousei Ishigami, Masakazu Hirakawa, Yasuhiro Ushijima, Daisuke Kakihara, Tomohiro Nakayama, Yunosuke Nishihara, Shinichi Aishima, Akinobu Taketomi, Junji Kishimoto, and Hiroshi Honda. Enhancement pattern analysis of hypervascular hepatocellular carcinoma on dynamic MR imaging with histopathological correlation: Validity of portal phase imaging for predicting tumor grade. *European Journal of Radiology*, 81(6):1116–1121, 2012.
- [309] Qiangguo Jin, Zhaopeng Meng, Changming Sun, Leyi Wei, and Ran Su. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. (October):1–13, 2018.
- [310] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January(1):6450–6458, 2017.
- [311] Xiao Chen, Lanfen Lin, Dong Liang, Hongjie Hu, Qiaowei Zhang, and Yutaro Iwamoto. A DUAL-ATTENTION DILATED RESIDUAL NETWORK FOR LIVER LESION CLASSIFICATION AND LOCALIZATION ON CT IMAGES College of Computer Science and Technology , Zhejiang University , China Department of Radiology , Sir Run Run Shaw Hospital , China College of Informa. *2019 IEEE International Conference on Image Processing (ICIP)*, pages 235–239, 2019.