

Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN
INGEGNERIA BIOMEDICA, ELETTRICA E DEI SISTEMI

Ciclo 33

Settore Concorsuale: 09/G2 - BIOINGEGNERIA

Settore Scientifico Disciplinare: ING-INF/06 - BIOINGEGNERIA ELETTRONICA E
INFORMATICA

NEW METHODOLOGIES IN CT PERFUSION AND MRI ANALYSIS TO DEVELOP
CANCER IMAGING BIOMARKERS

Presentata da: Margherita Mottola

Coordinatore Dottorato

Michele Monaci

Supervisore

Alessandro Bevilacqua

Co-supervisore

Mauro Ursino

Esame finale anno 2021

Keywords

Imaging biomarker

Signal processing

Reproducibility

Machine learning

Cancer imaging

External reviewers

Prof. **Filippo Molinari**,
Politecnico di Torino, Italy
email: filippo.molinari@polito.it

Prof. **Francesca Odone**,
University of Genova, Italy
email: francesca.odone@unige.it

Abstract

Medical imaging innovations have paved the way to quantitative imaging in oncology for developing cancer imaging biomarkers (IBs). At present, diagnostic imaging still mostly relies on morphometric analysis, based on visual assessments, or semi-automatic tools. However, different cancer features, including angiogenesis and heterogeneity, are known to be present far before that morphological changes become visible. Therefore, proper tools are required to perform functional assessments for early diagnosing the onset of either primary cancer or metastasis, predicting cancer aggressiveness, analysing cancer features at microscopic level and developing and monitoring therapies targeted against specific cancer subtypes.

This Thesis focuses on two distinct imaging technologies, Computed Tomography perfusion (CTp) and multiparametric Magnetic Resonance Imaging (mpMRI) exploited for investigating diverse cancer features, occurring on different primary sites. I developed a methodology in CTp which adopts a voxel-based image analysis approach and extended its use to mpMRI, for performing in both cases quantitative local analyses and achieving precision and accurateness at single-voxel level. This is expected to improve reproducibility of measurements and characterize biological behaviours in the highest detail, thus facilitating the clinical interpretation of the outcomes.

In quantitative applications, both CTp and mpMRI encounter technical limitations. While CTp enables the analysis of tissue hemodynamics for monitoring angiogenesis and cancer response to anti-angiogenic drugs, it has not entered the clinical routine yet due to different perfusion computing methods yielding unreproducible results. Instead, machine learning applied to mpMRI analysis may be exploited to extract imaging features representative of cancer heterogeneity and employ these features to build predictive models. However, the recent employment of software generating with a semi-automatic approach hundreds of features, regardless of their clinical meaning, have led to a distrustfulness of clinicians towards machine learning applications and weakened their clinical appealing accordingly.

In CTp, my contributions focus on investigating whether, and under what conditions, two of the widely adopted perfusion methods, Maximum Slope (MS) and

Deconvolution (DV), could yield reproducible perfusion parameters. To this end, I developed signal processing methods to model the first pass kinetics in hepatic perfusion study and remove any numerical cause hampering MS and DV reproducibility. In mpMRI analysis, I propose a new approach to extract imaging features based on the local computation of first order features, which preserves spatial reference, thus enabling their representation through colorimetric maps and making them easy to be interpreted, accordingly. The new imaging features are exploited to develop predictive and prognostic models.

On the methodological side, as regards CTP, my research unveiled the cause of MS and DV non-reproducibility: MS and DV represent two different states of the system. Transport delays invalidate MS assumptions which require correction. By formalizing the right MS formulation, I have achieved the voxel-based equivalence of the two methods. On the clinical side, my radiomics approach disclosed that, in rectal cancer, patients responding to neoadjuvant chemoradiation show, at pre-therapy mpMRI, only few sparse coarse subregions with altered tissue density. Instead, in prostate cancer, asymmetry gradients generated by the disproportion between high- and low- diffusivity gland component volumes predict cancer clinical significance which prescribes curative treatments.

Contents

Abstract	v
Contents	vii
1 Introduction and Thesis overview	1
2 Imaging modalities and techniques	9
2.1 Computed Tomography	9
2.2 Magnetic Resonance Imaging	16
2.2.1 T1- and T2- weighted MRI	19
2.2.2 Diffusion Weighted Imaging	20
2.3 Dynamic contrast-enhanced imaging techniques	24
3 Hallmarks of cancer and tumour biomarkers	29
3.1 Cell proliferation and tumour growth	29
3.2 Tumour angiogenesis	31
3.3 Tumour heterogeneity	35
3.4 From clinical to imaging biomarkers	37
3.5 The role of biomarkers in clinical practice	38
3.5.1 Diagnostic biomarkers	39
3.5.2 Prognostic biomarkers	40
3.5.3 Predictive biomarkers	42
4 Quantitative CTp image analysis	45
4.1 From image sequences to TCCs	45
4.2 Physiology of the organ: single or dual input model	48
4.3 Kinetics of the contrast agent	50
4.3.1 First pass or recirculation	51
4.3.2 Mono and multicompartamental model	53
4.4 TCC fitting models	55
4.4.1 The lagged normal	56
4.4.2 The sigmoid	58
4.4.3 The gamma variate	60

4.5	Computing methods for perfusion parameters	63
4.5.1	Maximum Slope	64
4.5.2	Compartmental model	67
4.5.3	Deconvolution method	72
4.6	Reproducibility of perfusion parameters: an open issue	78
5	Quantitative imaging for a radiomic approach	81
5.1	Radiomics in oncology: is there a novelty?	82
5.2	Fundamentals of Machine Learning	83
5.3	Binary discrimination	89
5.3.1	Statistical hypothesis testing	90
5.3.2	The ROC curve	92
5.4	A radiomic workflow	95
5.5	Image processing	95
5.6	Feature generation	98
5.7	Data preparation	106
5.7.1	Outlier removal	106
5.7.2	Data standardization and normalization	107
5.8	Shrinkage methods for subset selection	108
5.9	The SVM classifier	112
5.10	k-fold Cross Validation	117
5.11	The challenge: standardization and reproducibility	119
6	A methodological approach for MS and DV reproducibility	121
6.1	Case of study: PIXEL multicentre study	122
6.2	CTp study of liver through MS and DV	125
6.2.1	Model selection	125
6.2.2	Data preparation	126
6.3	Extracting the first pass signal	127
6.4	Modelling the first pass through signal fitting	128
6.5	Computation of perfusion parameters	130
6.6	The <i>generalized</i> MS	133
6.7	Reproducibility assessment	134
7	Experimental results: voxel-based reproducibility of BF	137
7.1	Analysis of the effects of CTp fitting errors on BF values	137
7.1.1	Analysis of residuals	139
7.1.2	Implications on BF values	142
7.2	Reproducibility of BF in the multicentre PIXEL study	143
7.2.1	Analysis of fitting parameters	144
7.2.2	Voxel-based reproducibility of MS and DV	146

7.3	The <i>equivalence times</i> of MS and DV	152
7.4	A theoretical proof by a single-input model case study	159
7.5	Assessing perfusion after Sorafenib and SIRT treatments	163
8	Radiomics: methodological and clinical studies	169
8.1	A reproducibility study of first order and texture features	169
8.1.1	Data preparation	170
8.1.2	Image processing	171
8.1.3	Feature extraction	174
8.1.4	Robustness analysis	175
8.1.5	Experimental results	176
8.2	A novel approach for radiomic feature computation	182
8.2.1	Why new imaging features	183
8.2.2	Local computation of parametric feature maps	184
8.2.3	A new set of imaging features	189
8.3	Discriminate nCRT responder from non-responder in Rectal Cancer	191
8.3.1	Study population	192
8.3.2	Image protocol acquisition	193
8.3.3	Data preparation and feature generation	194
8.3.4	Statistical analysis and feature selection	194
8.3.5	Experimental results	195
8.4	Exploiting radiomics in Prostate Cancer (PCa)	198
8.4.1	Study population	201
8.4.2	Objectives of the studies	203
8.4.3	Data preparation and feature generation	205
8.4.4	Preliminary staging of four progressive PCa risk levels	206
8.4.5	Prediction of clinically significant PCa	208
9	Conclusions	219
	Appendix	223
	CTp software development and management	223
	List of Abbreviations	229
	List of figures	233
	List of tables	237
	Bibliography	239
	Acknowledgements	261

Chapter 1

Introduction and Thesis overview

Bioengineering provides a large part of the medical innovations, regarding advanced technologies and analytic tools and techniques, which address different medical fields including oncology, for cancer diagnosis, assessment of patient's prognosis, monitoring new therapeutic protocols and therapy response. Accordingly, bioengineers constitute the bridge between innovation and real clinical application. In fact, the numerical analysis of medical data provides clinical evaluations with objectivity and robustness, whilst mathematical modelling advances the understanding of cancer mechanisms, through enabling the investigation of cause-effect relationships at progressive degrees of complexity. Working in multidisciplinary teams, bioengineers serve as figures to support the clinical interpretation of mathematical and modelling aspects of cancer mechanisms and to associate the information stemming from data analysis to the underlying biological processes, thus identifying the factors promoting cancer development and its sustainment into the human body [1].

In this Thesis, I focus on signal and image processing techniques to contribute at improving oncologic practice by providing reliable and non-invasive tools for predicting cancer aggressiveness and treatment effects. In particular, I address medical imaging to investigate those cancer biological properties, such as angiogenesis and heterogeneity, which are known to play a key role in determining the degree of cancer aggressiveness and treatment response, and whose quantitative evaluation is objectively challenging.

In this regard, angiogenesis is the growth, promoted by vascular endothelial growth factors, of new blood vessels from pre-existing ones and their proliferation into cancer to increase the nutrient and oxygen supply. Yet more, the degree of disorganization and disruption of the new vessels is known to correlate with cancer aggressiveness [2]. Instead, heterogeneity refers to the (spatial) variations occurring in cancer proliferation, responsible for the differentiation of cancers in multiple subtypes. Cancer heterogeneity occurs at a macroscopic level, with structural alterations also visible in the radiological images, as well as at the deeper underlying

layers, with modifications of biochemical functions and gene expression. Hence, the dynamic and microscopic nature of angiogenesis requires dynamic imaging modalities, such as the Computed Tomography perfusion (CTp), through which performing a parametric analysis of dynamic signals, arising from a tracer injected into the vascular network. Instead, in order to study cancer heterogeneity, medical imaging technologies, such as multiparametric Magnetic Resonance Imaging (mpMRI), are fit for objectively measuring both salient and latent image properties that can often capture different tissue heterogeneity levels with a great accuracy. Prospectively, numerical analysis of mpMRI contributes to advance precision medicine and develop patient oriented treatments targeted against specific cancer subtypes.

At present, several technical and analytical open issues still hinder quantitative imaging referred to both CTp and mpMRI in the clinical practice. CTp suffers from the side-effect of radiation dose exposure, which has to be minimized as much as possible. Then, the self-limited image quality makes achieving reliable quantitative measurements harder. In addition, for most of anatomic districts, CTp lacks of any ground-truth and different computing methods for perfusion parameters yield unreproducible results. Thus increases the difficulty of validating them and procrastinates the entrance of CTp in the clinical routine. As regards the numerical analysis of mpMRI, descriptive and more often inferential statistics techniques used for the extraction of imaging features yield a huge amount of data, this contributing to make their clinical interpretation very difficult indeed.

In this Thesis, my contribution, inspired to the principles of quantitative imaging, is based on a dedicated analysis of each stage of a processing pipeline to minimize uncertainty and variance of measures, thus identifying four main progressive analytic levels: (i) single-voxel, (ii) a region of neighbouring voxels, (iii) the entire region of interest, oriented to (iv) a single-patient analysis. The methodology has been structured throughout CTp analysis, by deepening signal processing, theoretical and mathematical modelling, and data analysis. Then, the same principles have been adopted, in clinical classification problems, to setup a machine learning pipeline, including the computation of new imaging features, to be reliably used in cancers occurring in different anatomic districts, even with small datasets.

In CTp imaging, I have dealt with the reproducibility of two widely adopted perfusion methods, the Maximum Slope (MS) and the Deconvolution (DV) assuming the *common agreement* principle, to mutually validate them. As the reference organ, I have addressed the liver, where perfusion studies are still harder due to the dual vascular contribution coming from the arterial and portal vein circulation. To cope with the reproducibility issue, I have developed three main signal processing methods, voxel oriented, which involve (i) the extraction of the end of the first passage of the tracer, data-driven (ii) modelling of the first passage and (iii) iterative DV computation. These methods have allowed separating the tissue signals referred

to each vascular input contribution and computing MS and DV independently on them. Yet more, these methods have permitted to exclude numerical reasons responsible for such a non-reproducibility of MS and DV and to ascribe the cause to theoretical reasons. Hence, I have investigated whether and under what conditions the two methods could be equivalent and I have proposed an alternative formulation of MS, the generalized MS (gMS), accounting for transport delays between input and tissue.

Instead, in mpMRI analysis, I have adopted a local approach to characterize each voxel based on its neighbourhood, extracting imaging features by considering a patch of neighbouring voxels as a structural unit, this recalling the concept of *the smallest informative unit*, already known in medicine. This approach aims at detecting image units large enough to be considered homogeneous and sufficiently numerous to allow investigating their heterogeneity. Hence, I have developed some algorithms, employing machine learning techniques, to identify potential predictive and prognostic imaging biomarkers of rectal and prostate cancer (PCa).

In CTP analysis, the main finding is to discover the cause of MS and DV non-reproducibility: MS and DV are namely representing two different states of the system, due to vascular transport delays. Moreover, the gMS can account for them and yield MS and DV equivalence, accordingly. In this regard, I have achieved a proof of concept on a dataset of seventy-five patients belonging to the multi-centre PIXEL study [3], where voxel-based equivalence of gMS and DV has been attained. Moreover, I have verified the theoretical hypothesis of the gMS under simpler modelling conditions, by exploiting a case study made of two CTP examinations of colorectal cancer, which is described through a single input model. In this case, where no delays were expected between input and tissue, the equivalence of MS and DV has been achieved. Then, in a clinical application, the liver perfusion study has been applied, on data belonging to the SARAH study [4], to compare the effects on perfusion parameters of Sorafenib and Selective Internal Radiotherapy (SIRT) in advanced hepatocarcinoma (HCC) at one-month follow-up.

In mpMRI analysis, my research has led to detect very promising IBs, with either a predictive or prognostic role. I have identified a single feature on a dataset of forty locally advanced rectal cancer (LARC) able to discriminate patients showing moderate and complete response to neo-adjuvant chemoradiotherapy (nCRT) from those having poor and minimal response. As regards PCa study, I have achieved a preliminary stratification of patients, through a single feature, into four progressive PCa risk levels, and I have developed a predictive model using a couple of imaging features to detect clinically significant PCa (csPCa).

Besides the present introductory Chapter, the Thesis content is organised as follows, also described by the flowchart in Figure 1.1:

- Chapter 2 provides insights into CT and mpMRI technologies, routinely used

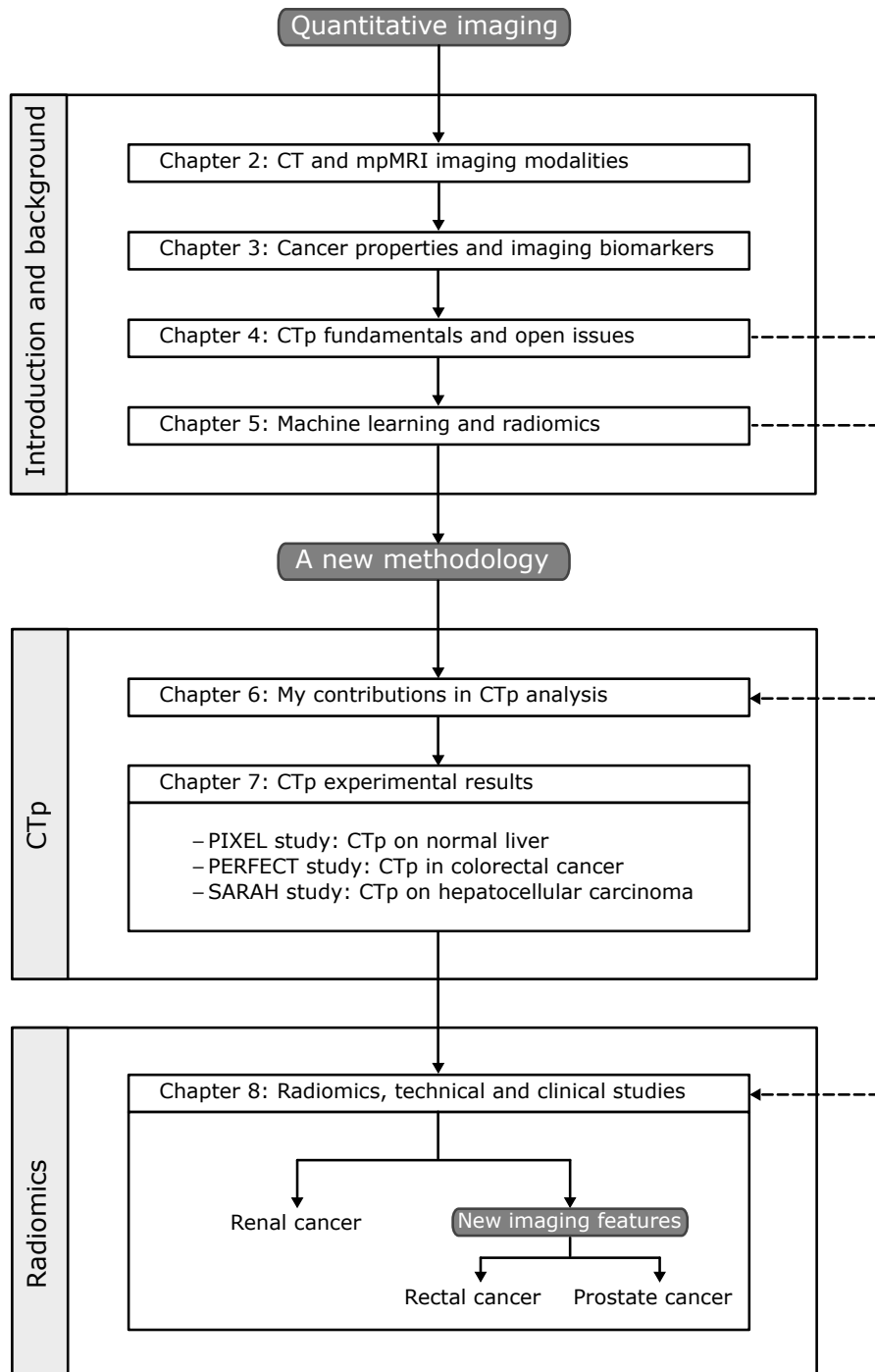


Fig. 1.1: Thesis content flowchart. Chapters from 2 to 5 introduce the background and state of the art regarding CTp, mpMRI, and machine learning in medical imaging. Chapter 6 presents the new voxel-based methodology developed for CTp analysis, whilst Chapter 7 reports CTp experimental results. Chapter 8 describes my contributions in radiomics: a radiomic study on renal cancer carried out exploiting features from the state of the art and the translation of the methodology developed in CTp to mpMRI analysis, with the setup of new imaging features, used in the clinical studies on rectal and prostate cancer.

in the clinical practice for cancer diagnosis, therapy assessment, and patient monitoring. In particular, the physics principles of CT and MRI are discussed, also reporting how CT and mpMRI allows acquiring either dynamic or functional sequences, besides static morphological ones.

- Chapter 3 deepens the description of relevant cancer features, including cell proliferation and tumour growth, angiogenesis, and heterogeneity, which prompt the main advances in imaging acquisition protocols, image analysis, and cancer care. Hence, the Chapter explains how evolutions in cancer research, promoting tailored treatments, have highlighted the need for reliable biomarkers to orient the clinical decision making. At the end, the evolution from clinical to imaging biomarkers is presented, even discussing their current utilization into clinical practice.
- Chapter 4 presents the fundamentals of CTP image analysis through the description of an operating workflow split into five main stages: (i) the extraction of dynamic signals from image sequences, (ii) the choice of either a single or dual input model depending on the physiology of the organ under investigation, (iii) the selection of either a mono or multicompartmental model depending on whether the perfusion study focuses on only the first passage of the tracer or even on the recirculating phase, (iv) the fitting of vascular and tissue CTP signals, and (v) the computation of perfusion parameters through MS, DV, and compartmental methods. Lastly, open issues of CTP analysis are discussed.
- Chapter 5 presents the fundamentals of machine learning and its application to medical imaging, recently referred to as radiomics. The Chapter also provides some elements of statistics necessary to understand the principles of binary discrimination and classification. Throughout this Chapter an example of radiomic workflow is presented. Then, for each operating stage, some of the methods known in the literature are explained, since they have been used in radiomic applications presented subsequently. In particular, the shrinkage methods for feature selection, Support Vector Machine (SVM) classifier, and k-fold Cross Validation (CV) for model selection are deepened. At the end, open issues of radiomics are discussed.
- Chapter 6 explains my methodological contributions to pursue the reproducibility of MS and DV. First, the Chapter presents PIXEL, a French multi-centre study on liver perfusion used as the reference for developing the methods for CTP analysis. Second, my preprocessing choices are presented, including model selection and data preparation. Then, the Chapter explains the methods developed to (i) extract the end of the first passage of the tracer in the liver, (ii) model the first pass with data-driven parametric functions, (iii) compute perfusion parameters with MS and DV, after setting up an iterative procedure to solve DV and separate the contributions of tissue signals referred to each of

the two hepatic vascular inputs, (iv), investigate the theoretical and modelling equivalence of MS and DV, thus formulating the gMS method. Moreover, the reproducibility assessment criteria are discussed. Yet more, the software pipeline developed for CTP analysis is presented in Appendix.

- Chapter 7 presents the experimental results of the voxel-based reproducibility of MS and DV. First, a preliminary study investigates the effects of fitting errors on perfusion parameters. Then, by including seventy-five liver CTP examinations of PIXEL study, the Chapter starts showing the efficacy of computational methods in reducing variance and improve precision of measurements, thus yielding (i) an excellent matching of colorimetric perfusion maps and (ii) voxel-based agreement of BF measurements. Then, I prove that the gMS can catch the time when MS and DV are representing the same status of the system, thus yielding equal voxel-based measurements. Then, equivalence of MS and DV is shown also on a case study of two colorectal cancer CTP examinations. At the end, a clinical application is presented, where CTP is exploited to compare two different treatments in advanced HCC.
- Chapter 8 presents my contributions in radiomics. In particular, a methodological study assessing the reproducibility of the most frequently adopted radiomic features is reported, carried out during my PhD research period abroad. Then, I propose a new approach to compute imaging features giving relevance to local variability of image values and facilitating the clinical interpretation of feature values. The Chapter discusses the application of this approach in clinical radiomic studies on LARC to predict the nCRT response, and on PCa to perform a preliminary staging of PCa risk levels and classify csPCa.
- Chapter 9 provides concluding remarks and hints for future works.

Research presented in this Thesis have been carried out within the:

- Computer Vision Group (CVG), Advanced Research Center on Electronic System (ARCES), University of Bologna, Italy
Head: Prof. Alessandro Bevilacqua

and in collaboration with the following institutions:

- Department of Radiology, Beaujon Hospital, University Paris VII “D. Diderot”, Clichy, France.
Chair: Prof. Valérie Vilgrain
- Department of Radiology, Addenbrooke’s Hospital, Cambridge, United Kingdom.
Chair: Prof. Fiona Gilbert
- Diagnostic Imaging Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori “Dino Amadori” – IRST S.r.l., Meldola (FC), Italy.

Head: Dr. Domenico Barone

- Department of Radiology, S. Orsola-Malpighi Hospital, Bologna, Italy.

Chair: Prof. Rita Golfieri

Chapter 2

Imaging modalities and techniques

In oncology, different imaging modalities are conventionally adopted for tumour diagnosis, therapy assessment, and patient monitoring. CT and MRI are largely employed as anatomic imaging modalities and provide high resolution information regarding tissue morphology, such as tumour location and size, and also structural changes involving the adjacent organs [5]. Moreover, both CT and MRI can also be exploited with dynamic contrast-enhanced (DCE) imaging techniques, namely DCE-CT and DCE-MRI, in order to derive functional information of tumour microvasculature. In DCE-CT and DCE-MRI, repeated image acquisitions are performed with the conventional CT and MR scanners in order to characterize the vascular tumour pathways by following the dynamics of a contrast agent (CA) intravenously injected [6]. In compliance with the multidisciplinary nature of this Thesis, in the following, I address the physics principles of CT and MRI, together with their clinical impact, though the most of the technical details could already be known by bioengineers reading this Thesis.

2.1 Computed Tomography

CT is the widest spread modality for morphological and structural imaging, that since its introduction, in the early seventies, has revolutionized the field of radiology. Yet more, it is still progressing with continuous advances in terms of technologies and image reconstruction techniques. These developments have lead to an effective synergy between the fields of radiology and oncology, that, in its turn, exceptionally improves oncologic patient care.

In CT imaging modality, X-rays are directed to thin sections of the body from multiple orientations. X-rays intensity decrease along the path, depending on the X-ray energy, path length, and density of the encountered structures. Hence, attenuated X-rays, detected after they have passed through the volumetric sections

of human body, are processed in order to reconstruct images of the distribution of X-ray attenuation coefficients in the analysed volume. Therefore, two basic components constitute the CT scanners, an X-ray source, that is in the most of cases an X-ray tube producing a polychromatic X-ray beam, or a synchrotron producing a monochromatic X-ray beam, and X-rays detectors, which receive the attenuated X-rays and convert them in a digitally processable signal. The X-ray tube is located on a structure rotating around a fixed support called gantry, in which the patient bed is moved during acquisitions. CT scanners currently adopted in clinic refer to the third and fourth generation of scanners. While in CT scanners of the third generation, X-rays tube and detectors are opposite located on the rotating structure and move synchronously, in the fourth generation CT scanners, a ring of fixed detectors is used, and only the X-ray tube moves around the entire rotating line. In planar fan beam CT scanner configuration, X-rays are collimated and detected by linear detector arrays. In this case, collimation reduces the scattering of X-rays, so that a lower number of spurious X-rays is received by detectors. The employment of linear arrays limits the CT acquisition at one slice at a time. Instead, in cone-beam CT scanner configuration, X-rays detectors are planar, and the collimator is removed. Thousands of slices can be acquired at a time, thus decreasing image noise during reconstruction. Finally, the parallel beam configuration adopts a synchrotron as beam line X-ray source, with the advantage of volumetric acquisition in the absence of distortion, although the size of the acquirable volume is limited by the beam line configuration. X-ray tube parameters, together with the intrinsic property of the detectors, strongly affect the quality of the reconstructed CT images. In particular, the peak X-ray energy determine the X-ray spectrum, X-ray tube current affects X-ray intensity, while the focal spot size influences the spatial resolution by varying the Field of View (FOV) of the acquisition. X-rays detectors are made of scintillators generally realized with cesium iodide, gadolinium oxysulfide, and sodium metatungstate. Material, size and geometry determine the number of scintillation events detected and counted. Currently adopted CT detectors show a compact design made of three essential layers. The first one is the scintillator to convert X-rays to light, the second one is the photodiode, to convert light to current, and at last there is the mechanical and electrical supporting layer. Scintillator is a particular luminescent material converting high-energy photons into visible light, suitable to be processed by a photodiode. Important requirements for scintillators are optical transparency in order to ensure high efficacy of X-ray conversion, high X-ray stopping power, good spectral match with the underlying photo-detector, so that crystals and polycrystalline are the most suitable types of scintillators [7].

X-ray attenuation mainly depends upon two types of interactions between photons and matter, the Compton scattering and the photoelectric effect. Both of them are X-ray absorption processes. Based on the Compton effect, the interaction of the

photon (X-ray) with free electrons produces a photon scattering (change of direction) and release energy to the electron. The scattered rays have different wavelength and energy with respect to the incident X-ray, and it is possible to reconstruct the intensity of the scattered X-rays starting from the intensity and the wavelength of the incident X-ray, which are known, and from the X-ray scattering angle, which is retrieved from the corresponding direction of the detector that detects the scattered X-ray. In particular, the difference in wavelength between the incident and scattered X-rays is called Compton shift. This is a kind of partial absorption and it dominates at high X-ray energy, where absorption mostly depends on material density. The photoelectric effect occurs when a X-ray interacts with matter, and in particular with an electron of an atom, thus producing the ejection of the electron from its K-shell, the innermost shell of the atom. In order to produce the electron ejection, the energy of the incident photon must be greater or equal to the one of the electron in its shell. Then, the removed electron is completely absorbed in the process, and, in order to stabilize the atom that has lost one of its electrons, an outer electrons fills that vacancy. The energy lost by the outer electrons to enter the unstable atom is emitted as X-ray radiation with attenuated intensity with respect to the incident X-ray. Therefore, the photoelectric effect, predominantly occurring at low energy, strongly relies upon the atomic number of the attenuating medium, the energy of the incident photon, and also the density of the medium [8].

Signals received are processed with image reconstruction algorithms in order to reconstruct the local X-rays attenuation coefficient, μ , of each point within the body section, thus producing cross-sectional radiodensity images - slices - of the body. Once a sufficient number of body slices are collected, they are stacked together, thus mimicking a volumetric imaging in which the smaller unit is represented by the voxel. For CT scanners, a transverse plane here referred to as x - y plane is scanned and the z -axis is oriented perpendicular to the scan or image plane, therefore aligned longitudinally with the body. Sagittal and coronal sections are approximated by y - z and x - z planes, respectively. Local attenuation coefficients are computed by means of the Inverse Radon transform based on the following relationship (Equation 2.1):

$$I = I_0 \cdot e^{-\mu \cdot d} \quad (2.1)$$

where I , is the signal intensity being attenuated by the object, I_0 is the primary intensity, and d is the absorber thickness. This relationship is valid for the simplest case of homogeneous object (constant density) invested by a monochromatic radiation, such as the one produced by a well-calibrated synchrotron. According to Equation 2.1, μ can be easily computed as follows in Equations 2.2, 2.3:

$$\ln \left(\frac{I_0}{I} \right) = \mu \cdot d \quad (2.2)$$

$$\mu = \frac{1}{d} \cdot \ln \left(\frac{I_0}{I} \right) \quad (2.3)$$

Indeed, body tissues are inhomogeneous structures (with at least small variations of local densities), invested from polychromatic X-ray beams, so that Equation 2.1 should be adjusted into Equation 2.4:

$$I = \int_0^{E_{max}} I_0(E) \cdot e^{-\int_0^d \mu(E) ds} dE \quad (2.4)$$

where, the contribution to the total attenuation resulting from each ray path interval depends on the local value of μ (a distribution of values), and is expressed by the integral over μ along the ray path. In addition, Equation 2.4 introduces the dependency of attenuation values on energy, E . Accordingly, the integration over all energy intervals is considered. In practice, Equation 2.4 allows estimating a distribution of attenuation coefficient, $\mu = \mu(x, y)$, at a fixed z -position. Dependency of attenuation from energy is known to cause so called beam hardening artefacts. Low-energy photons are absorbed by the tissue more than high-energy photons, thus increasing the mean energy value of the X-ray beam leaving the body section. This phenomenon is referred to as “hardening” of the beam. Beam hardening artefacts creates streaks and dark bands in CT images between two dense objects [9].

The most widely used reconstruction method for CT images is the filtered back-projection (FBP). According to FBP, each signal received by the X-ray detector, describing the attenuation of the beam for a given path, is backprojected along the viewing direction from which it is detected. Hence, the contributions arising from multiple backprojected directions are added to each voxel intersecting that ray direction. Resulting attenuation coefficient of each voxel is given by the sum of non-negative values corresponding to each projection. The major artefact of FBP reconstruction is that positive attenuation coefficients are also assigned to voxels that do not contains any object, thus resulting in blurring edges in the reconstructed CT images. Iterative reconstruction algorithms and combinations of multiple CT scans can reduce image blurring as well as the beam hardening CT artefacts above mentioned. The basic idea of iterative reconstruction is a trial and error approach: for each X-ray beam direction, a forward projection is computed by summing the intensities of all voxels along that path. This projection is compared based on different statistical metrics with the actual recorded projection, and the process is iterated until convergence is reached to the desired similarity value. If on the one hand, iterative reconstruction algorithms result in a better image quality, with lower noise and artefacts, on the other hand they require high computation time, which limits their large employment [10].

Actually, reconstructed images are displayed as CT numbers, CT_N , expressed in Hounsfield Units (HU), thus both removing any dependence of the X-ray spectrum

and expressing values in a more convenient size. Conversion from local μ values to CT_N is defined as follows in Equation 2.5:

$$CT_N = \frac{\mu_x - \mu_{water}}{\mu_{water}} 1000 \quad (2.5)$$

where, μ_x is the attenuation coefficient of each point of the image - pixel - whilst μ_{water} is the attenuation coefficient of water employed for normalization. By definition, -1000 HU refers to vacuum and 0 HU to water. The CT values of air and water are independent of the energy of the X-rays, thus constituting two calibration points. Indeed, a single voxel unit may contain tissues with different attenuation values, so that the corresponding CT_N approximates the weighted sum of the different attenuation values. The X-ray detector's limit to differentiate tissues under the voxel unit size is referred to as partial volume effect. Accordingly, the CT_N of a voxel affected by the partial volume effect is described by Equations 2.6, 2.7:

$$CT_N = \nu_1 CT_{N1} + \nu_2 CT_{N1} + \dots \nu_i CT_{Ni} \dots + \nu_n CT_{Nn} \quad (2.6)$$

$$\sum_i \nu_i = 1 \quad (2.7)$$

where, ν_i are the partial volume elements, resulting as the weights of the linear combination of each relative CT_N . Lung tissue and fat show negative CT values, whilst most other body areas such as connective tissues and most soft tissue organs show high CT positive values. CT values of bone and contrast media strongly depend on energy and increase with reduced voltage settings. From a theoretical point of view, the Hounsfield scale is no upper bounded, although in practice, CT medical scanners usually provide a range of values between -1024 HU and +3071 HU corresponding to 4096 (2^{12}) levels, with 12 bits per pixel required. However, all the levels cannot be evaluated in a single view by the human eye. Therefore, CT images are often displayed by selecting a restricted window of grey levels, so that CT values exceeding window limits are saturated to the minimum or maximum value of the chosen range. The window is set up by its width (WW) which affects image contrast and by its level (WL, i.e. the centre of the window) which changes image brightness. For instance, Figure 2.1: shows two examples of WW/WL values commonly employed for visualizing lung (Figure 2.1 (a), WW=1500 HU and WL=-400 HU) and abdomen (Figure 2.1 (b), WW=400 HU and WL=60 HU) CT images. In addition, Figure 2.2 shows the same image as Figure 2.1 (b) after adopting the HU window of Figure 2.1 (a). One can see that an erroneous set up of WW/WL values hampers the correct interpretation of the anatomic image. In fact, the most appropriate HU window often depends on the anatomic region under investigation. Image contrast is defined by the difference in intensity of two neighbouring image areas. That is, CT images show high contrast in those areas where structures with high attenuation

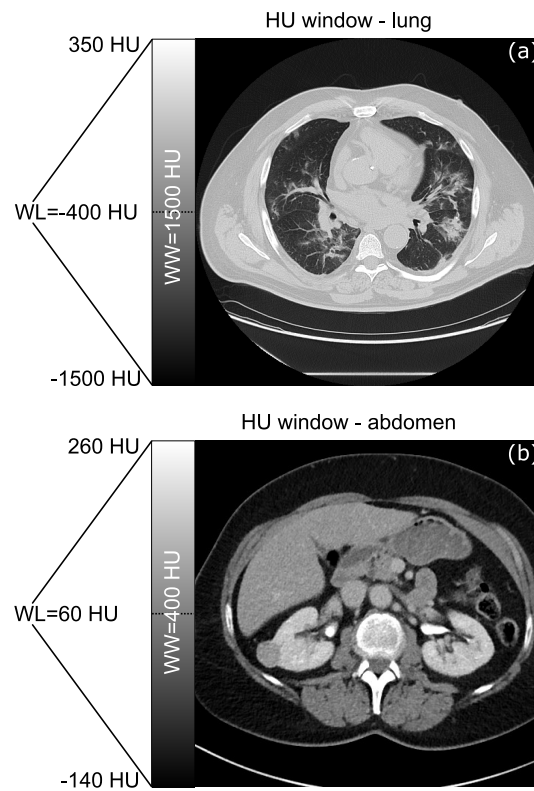


Fig. 2.1: Examples of HU windows for the visualization of lung (a) and abdomen (b) CT images, by properly setting the WW and WL values.

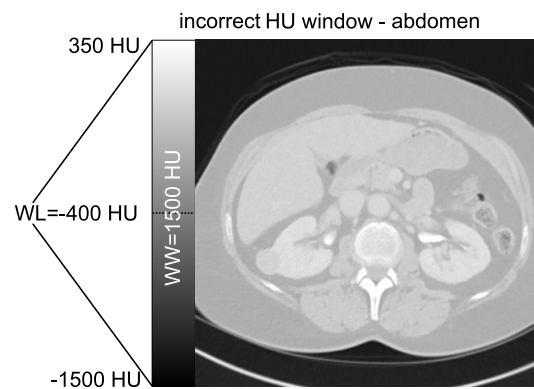


Fig. 2.2: Abdomen CT image of Figure 2.1 (b) displayed using the HU window in Figure 2.1 (a).

values are adjacent to ones with low attenuation values. Accordingly, high density structures can be easily characterized with CT, while soft and low density tissues are completely hidden in most practical cases, for which other morphological imaging

modalities result more proper [9].

CT image resolution is measured through the voxel size, that refers both to the spatial resolution as pixel spacing and temporal resolution as frequency of sampling along the z -axis. Image voxel size mainly depends upon three factor, that is size of the FOV, acquisition matrix and slice thickness. While FOV is a parameter dependent on the focal spot size, which can be adjusted according to the tissue being imaged and patient size, the acquisition matrix depends on physical factors, that is the number of detector rows of a CT scanner and detector sizes which state the minimum value of pixel dimension. Currently, multi-slice CT scanners are employed, with detector rows ranging from 256 to 320, z -coverage up to 160 mm, and pixel size of 0.6 mm. Multi-slice CT scanners allow acquiring multiple slices (equals to the number of detector rows) per gantry rotation (with rotation time equals to 0.28 s), thus strongly reducing overall acquisition length. Once FOV has been fixed, the digital pixel spacing in the image is stated by Equation 2.8:

$$\text{pixel spacing} = \frac{\text{FOV}}{\text{matrix size}} \quad (2.8)$$

Actually, image resolution relies on a compromise between image quality and patient dose exposure [8]. In particular, CT dose refers to the energy absorbed from radiation per mass of tissue and is measured in Gray (mGy). For practical reasons, it is commonly expressed as CT Dose Index (CTDI) that measures the dose absorption per each slice, by accounting for the different distribution of radiation between the centre and the periphery, and also considering intrinsic patient's characteristics, such as body weight or height [11]. It is also employed the Dose Length Product (DLP, measured in mGy/cm), where the CTDI is multiplied for the scan length (measured in cm). Moreover, the DLP can be used to compute the effective dose exposure (measured in mSv) of the patients, that is given by Equation 2.9:

$$E = (k \times DLP) \quad (2.9)$$

where k coefficients are tabular values referred to adult or pediatric patients over different ages and body regions, measured in $\text{mSv}/\text{mGy}\cdot\text{cm}$ [12].

CT image quality is substantially determined by X-ray beam width that varies the number of photons per voxels, so that the increase of photon numbers can reduce image noise. Differently, dose absorption relies on the concentration of photons per voxels. For instance, if keeping fixed the dose exposure, a better image quality can be attained by enlarging voxel sizes, thus increasing the number of photons per voxels accordingly. Indeed, X-ray beam width depends on design characteristics, and acquisition factors of CT protocols, that is voltage (kVp) and tube current (mA or mAs). Standard values of voltage and tube current range within kVp:[80÷200] and mA=[90÷130], respectively. Image noise (mainly quantum noise) inversely relate

with the square root of the number of photons in a voxel, while tube current, rotating time, and slice thickness all directly correlate with number of photons in the voxel. For instance, by increasing mA by four fold, will reduce by one half image noise.

While for many years the only method adopted for slice acquisitions was the so called “scan and step”, that is acquisition of slices by progressive z -values without overlapping, a recent common alternative is the spiral or helical scanning. Accordingly, the patient body is moved continuously along z -axis during the rotation of the gantry around it. This leads another acquisition parameter to be properly set before acquisition, the Pitch factor, which expresses the ratio between the distance at which the patient is moved and the X-ray beam width. The main advantages refer to the shortness of acquisition time and dose reduction, albeit with the risk of worsening image quality. In fact, for helical CT scans the CTDI should be divided by the Pitch value to account for the spreading of radiation in a wider area, that reduces photons concentration per slice. Moreover, the spreading of X-rays can originate blurring artefacts in CT images. Therefore, on the one hand helical scans offer more flexibility in terms of image reconstruction, that is images at several slice thickness and orientations can be reconstructed in post-processing. On the other hand, a CT image with high detail can be reconstructed only if that detail is present in the original scan. For these reasons, axial CT acquisitions can still be the proper solution in many practical cases [9].

2.2 Magnetic Resonance Imaging

MRI measures different characteristics of tissues, by using different energies and physical concepts. Incidentally, one of the advantages is that ionizing radiations are no longer needed. In particular, MRI exploits the phenomenon of Nuclear Magnetic Resonance (NMR) to generate morphological images along multiple axes of human body. To recall the basic principle of NMR, nuclei contain protons, positive charged particles which cause nuclei to exhibit a magnetic field. In addition, protons within a nucleus, bear a spin angular momentum, I , that can be either $I=0$ or $I = 1/2$. Only nuclei whose protons have non null angular moment will be affected by magnetic fields. Human body is predominantly composed of hydrogen (one proton and one electron) as a part of water molecules, and its nucleus has $I = 1/2$. Therefore, MRI exploits the signals arising from the magnetization of hydrogen nuclei to generate images of body sections. Of course, all nuclei within human body with an odd number of protons will have a half integer angular moment I , however, MRI is based on hydrogen basically due to its elevate concentration in human body, and its strong magnetic moment.

From quantum mechanics theory, nuclei with $I = 1/2$ under an external magnetic

field, B_0 , measured in Tesla (T), have two possible energy states, parallel (α , low energy) and antiparallel (β , high energy) to B_0 . The energy difference between the two states and so, the disparity between parallel and antiparallel spins, increases as the field B_0 increases. In addition, each charged particle has a magnetic moment $\vec{\mu}$, defined according to Equation 2.10:

$$\vec{\mu} = \gamma \cdot \vec{J} \quad (2.10)$$

where, \vec{J} is the total angular momentum and γ is called gyromagnetic ratio, a constant specific for each nucleus. Moreover, the disparity of parallel and antiparallel spins generates a residual magnetization, M , also causing a degree of polarization of the matter. Because of the presence of the magnetic moment $\vec{\mu}$, each charged nucleus precesses about the axis of B_0 with an angular frequency, ω_0 , called Larmor Frequency and expressed by Equation 2.11:

$$\omega_0 = \gamma B_0 \quad (2.11)$$

Generally, these nuclei within the static field B_0 , are described as in the ground state energy level of an harmonic oscillator, that is an unperturbed state. When perturbed, nuclei are promoted to higher or lower energy levels, and seek to retrieve their equilibrium ground state when perturbation ends. Hence, unperturbed nuclei are described by the energy level as in Equation 2.12

$$E_n = \left(n + \frac{1}{2} \right) \cdot h\omega_0 \quad (2.12)$$

where h is the constant Plank and n is the quantized energy level, that is $n=1$ if referring to α and $n=2$ if referring to β . This leads the energy difference between α and β expressed as in Equation 2.13:

$$\Delta E = h\omega_0 \quad (2.13)$$

Hence, the phenomenon of NMR occurs when some nuclei change their energy state (α or β) due to a perturbing magnetic source. In MR, the NMR is induced by applying a second magnetic field B_1 , perpendicularly to the static field B_0 (typical B_0 values are 1.5-3 T). The resonance condition causing the transition between α and β energy levels is described by Equation 2.14:

$$h\omega_1 = |E_\alpha - E_\beta| \quad (2.14)$$

where $\omega_1 = \gamma B_1$, that is the resonance frequency due to B_1 , that must be in the order of the energy distance between the two spin states to perturb their equilibrium state. In MRI, B_1 is applied as short radiofrequency (RF) pulses that tip off $\vec{\mu}$ from the axis

of B_0 of a so called flip angle proportional to the magnitude of the magnetic field [13]. During relaxation, that is returning to equilibrium state, the absorbed energy (and subsequently emitted) produces a typical Free Induction Decay (FID) signal. This signal is detected and amplified and constitutes the primary source for MR image reconstruction. In particular, in order to spatially localize the nuclei emitting FID signals, RF pulses of magnetic gradients are applied along three orthogonal distances. Also, RF pulses are repeated so that multiple FID signals are averaged to improve the signal-to-noise (SNR) ratio. At last, averaged FID signals are Fourier transformed, to reconstruct the MR image [14].

Spatial resolutions on reconstructed images is defined according to Equation 2.15, that states the relationship between acquired MR images and the physical object:

$$I(x, y, z) = O(x, y, z) \otimes h(x, y, z) \quad (2.15)$$

Here, the acquired image, $I(x, y, z)$ is given by the convolution of the physical object, $O(x, y, z)$, with a 3D Point Spread Function (PSF). In practice, two different objects can be detected if their distance is not less than the full width at half maximum (FWHM) of the PSF. PSF in MRI are generally anisotropic, since the acquisition process is different for each dimension. In fact, slice resolution is also strictly linked to the frequency response of the RF gradient pulses for slice selection. As for CT imaging, the digital pixel spacing of the image is related to FOV and acquisition matrix (Equation 2.8). However, by considering the width of the PSF inversely related to the number of samples acquired in the corresponding dimension, data truncation losses can occur even at small pixel sizes [15].

A MRI scanner is composed by several hardware components to generate, encode, and receive the signals that are converted into images. First, a magnet that generates a strong static field in the space where images must be acquired. Second, a system of gradient coils and gradient amplifiers to create the magnetic field gradient system. Then, an RF amplifier and RF transmit coil to activate resonating nuclei, and RF receive hardware components to detect the signals back emitted by resonating nuclei. RF signals are received and collected by a spectrometer, until they are enough to reconstruct the image. In current MR scanners, cryogenic superconducting magnets of 1.5-3 T are employed to generate the static field. In this regard, the major issues refer to the combination of large field strength, high field homogeneity at the centre of the magnet's volume, and high field stability. Superconducting magnets, and in particular coil windings, are cooled through immersion in liquid helium at a temperature of 4.3 K, where their electrical resistance becomes zero. At this state, if an external power supply is applied, the inductive current generated can be maintained even when the power supply is removed, thus also producing and maintaining the static magnetic field. The amount of inhomogeneities in the magnetic field causes, in a proportional manner, spatial distortion

effects. Actually, MRI signals are weak, and small external RF interferences can degrade signal and image quality. Therefore, the MR scanner room is fully shielded from external sources, thus also preventing interference from the MR system to RF-sensitive external instrumentals. The spectrometer of MR systems serves for RF generation and transmission, and reception of RF signals, so that it includes a variety of RF components. Typically, the spectrometer is integrated in a high-speed digital computer to allow a real time precise RF pulse and waveform generation and to enable precise data acquisition timing. RF transmission is realized by a cascade of RF amplifiers and RF coils, that can serve as both transmitting and receiving antennas. While uniform RF excitation fields have a great relevance for achieving a good image quality, uniform receive sensitivity is not a requirement, thus being even advantageous when many receive coils are used together. In current MR scanners, generally distinct RF coils are used for transmission and reception. In particular, volume coils are used for RF transmission and surface of array coils for reception, since they yield a higher SNR than volume ones. Actually, phased array coils are used for reception, consisting of multiple coils allowing receiving with a higher SNR multiple signals, that are combined to reconstruct one image. The improvement of SNR is allowed by separate preamplifiers and receiver channels. Recent MRI scanners have up to 64 receiver channels and around 200 coil elements [16].

2.2.1 T₁- and T₂- weighted MRI

After NMR perturbation, the signal observed varies over time depending on the nature of the tissue itself and the interactions within the tissue and external forces. There are two main types of magnetization relaxation, that is longitudinal (T_1) and transverse (T_2) relaxation. The first refers to the spin-lattice energy transfer occurring when nuclei lose the energy absorbed through RF, the second refers to the spin-spin energy transfer, due to interactions among nuclei. Both T_1 and T_2 times, needed for longitudinal and transverse relaxation, respectively, are property of the analysed tissue. In particular, T_1 corresponds to the time taken by the longitudinal magnetization to recover of the 63% (T_1 recovery). Accordingly, T_1 limits the time TR, that is the repetition time for RF pulses. T_2 is the time took for the 63% of the transverse magnetization to be lost due to dephasing (T_2 decay). In practice, it is the time between the RF pulse and the MR signal, or Time of Emission (TE). Body tissues can be differentiated according to their own T_1 and T_2 times, thus obtaining the corresponding T_1 - and T_2 - weighted MR sequences. In T_1 -weighted images, in order to emphasize the differences between T_1 times of different tissues, TR is required to be very short. And also, TE must be short to limit the effects of T_2 time. A low signal is measured in T_1 -weighted images for tumours, necrotic areas, sclerosis, calcification, whilst high signals are generated by fat and contrast

media. Instead, in order to emphasize tissue differences according to the T_2 time, TE must be long. In this case, tissues such as fat, with short T2 time, result dark, whilst water, tumour, and inflammation are bright. T_2 relaxation also impacts on MR images spatial resolution: the higher degree of relaxation, the narrower the PSF. Nowadays, 3T-MRI systems can reach in-plane resolutions lower than 1 mm² with 3 mm of slice thickness. T_1 - and T_2 - weighted images are the reference morphological MR sequences in tumour studies [17].

2.2.2 Diffusion Weighted Imaging

The application of RF pulses according to different protocols allows MRI images to show several properties of the tissue of interest. The variety of tissue properties that can be pointed out through images is one of the major strength of MRI if compared to CT modality, and shows its relevance in oncologic imaging, where with a wider amount of information related to tumour, diagnosis and treatments can be more precise and accurate. Diffusion Weighted Imaging (DWI) is based on the attenuation of T_2 signals according to the capability of water molecules to diffuse (due to Brownian motion) in that region. Diffusion is measured through the diffusion coefficient, D , that describes the rate of diffusion and is dependent upon temperature, the size of particles, and viscosity of the tissue. The free diffusion of water molecules at normal body temperature is $D=3 \times 10^{-3}$ mm²/s. However, water molecules in human tissues are not free to move, since they encounter macromolecules, cell membranes, and other microstructures that obstacle their motion. High signals in DWI emphasize areas of restricted diffusion, that suggests tissue diseases.

In practice, a T_2 -weighted image must be obtained with no diffusion attenuation, before applying gradient RF pulses along multiple directions. Generally, DWI acquires isotropic images by employing three gradients of same amplitude along the three x , y , and z axes [17]. In particular, when applying RF pulses for DWI, the spin's magnetic moment acquires a phase shift so that the signal received in MRI is due to the bulk magnetization, which is the sum of the magnetization vectors of all particles with unknown location. In particular, the bulk signal attenuation, $S(TE)$, with respect to the original signal, S_0 , is given by Equation 2.16 [13]:

$$\frac{S(TE)}{S_0} = e^{-bD} \quad (2.16)$$

where, b is a diffusion attenuation factor chosen before image acquisitions. Generally, the choice of the appropriate b -value mostly depends on the degree of anisotropy in the tissue of interest. Low b -values ($b=50/100$ s/mm²) sequences have a contrast similar to T_2 -weighted images but also added suppression of intravascular signal

intensity, thus increasing lesion conspicuity. Moreover, DWI images has generally a lower in-plane resolution than the corresponding slice in T_2 -weighted sequences, with pixel spacing between $[1\div 1.5] \text{ mm}^2$ [18]. In this regard, Figure 2.3 shows

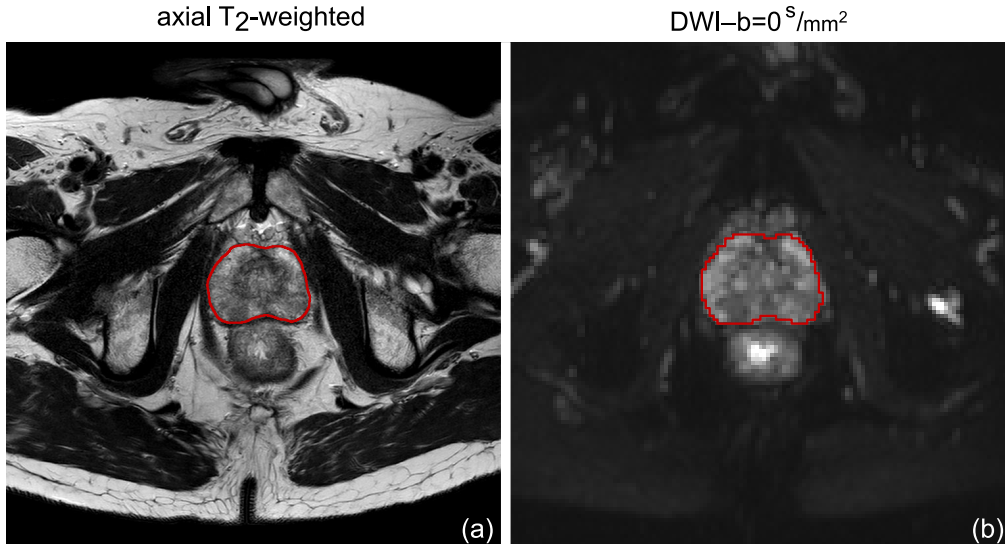


Fig. 2.3: T_2 -weighted image (a) and DWI at $b=0 \text{ s/mm}^2$ image (b) of prostate, outlined in red. T_2 -weighted image has a higher in-plane resolution, thus representing anatomy with more detail.

an axial T_2 -weighted image (Figure 2.3 (a)) and a DWI image (Figure 2.3 (b)) acquired with $b=0 \text{ s/mm}^2$, where, in both images, prostate Region of Interest (ROI) is outlined in red. One can appreciate the higher morphological detail in T_2 -weighted image rather than DWI one having a lower in-plane resolution. Because of the different resolution, higher in T_2 -weighted than in DWI, the prostate contouring has a smoother appearance in Figure 2.3 (a), whilst it seems rawer in Figure 2.3 (b). High b -values will emphasize the contrast between restricted areas and those ones with normal diffusion, which can be very useful in tumour localization. Figure 2.4 compares DWI images acquired at four different b -values, that is $b=0, 150 \text{ s/mm}^2$ as low b -values, and $b=800, 2000 \text{ s/mm}^2$ as high b -values. In all images, prostate (red) and tumour (green) ROIs are outlined. From $b=0 \text{ s/mm}^2$ to $b=2000 \text{ s/mm}^2$ there is a progressive loss of anatomical references with enhancement of only the tumour area. Actually, high b -value DWI sequences can suffer from the so called susceptibility artefact arising from the air-tissues boundaries, that generates high signal not to be diagnosed as abnormal [19].

Assuming isotropic diffusion in two different areas with diffusion coefficients D_A and D_B , respectively, and signal intensities S_{0A} and S_{0B} when no diffusion is applied (i.e. $b=0 \text{ s/mm}^2$), the contrast (ΔS_{max}) between these two signals is maximized if

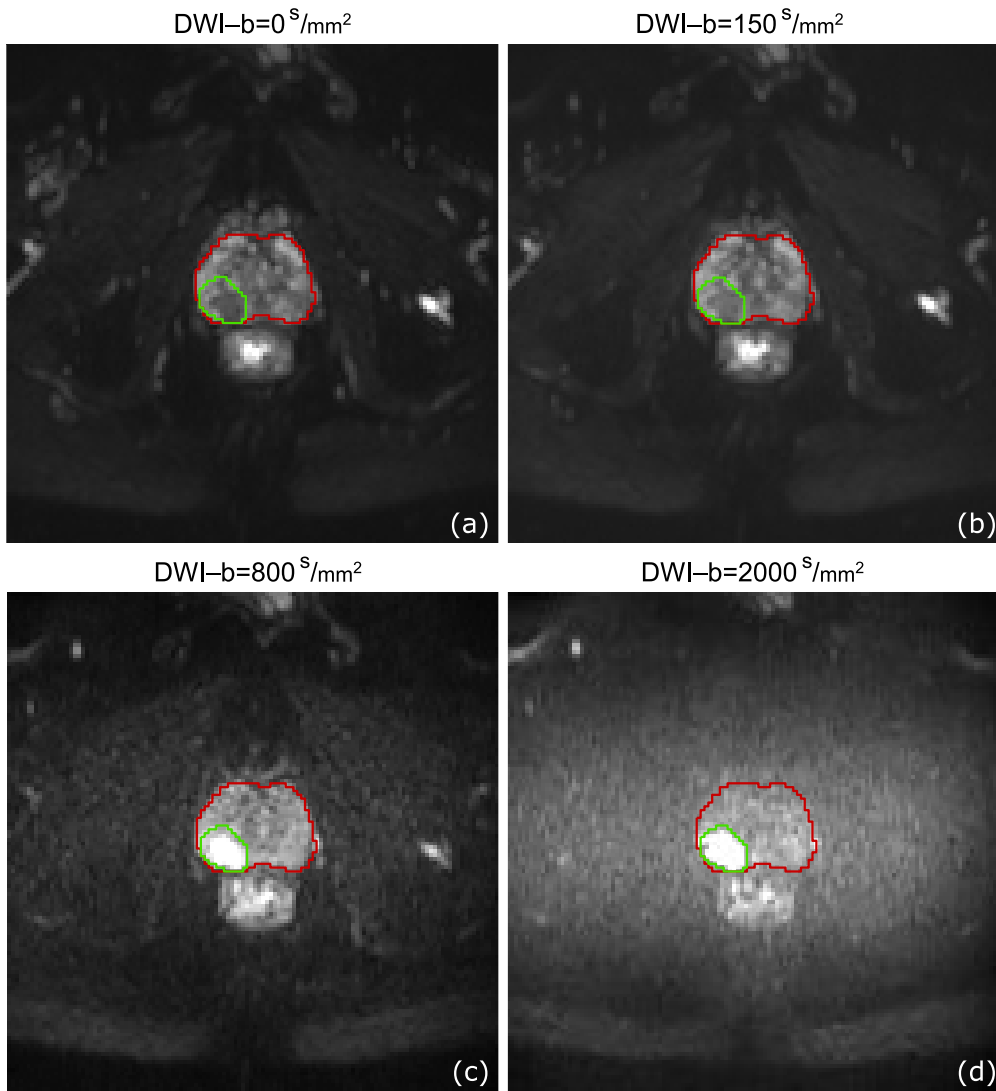


Fig. 2.4: DWI images acquired at four different b-values (low b-values in a,b and high b-values in c,d) with prostate and tumour ROIs outlined in red and green, respectively. From $b=0 \text{ s/mm}^2$ (a) to $b=2000 \text{ s/mm}^2$ one can appreciate the enhancement of DWI high signals in tumour and loss of morphological detail of the surrounding prostate gland.

b -value corresponds to Equation 2.17 [13]:

$$b_{\Delta S_{max}} = \frac{\ln\left(\frac{D_A}{D_B}\right) - \ln\left(\frac{S_{0A}}{S_{0B}}\right)}{D_A - D_B} \quad (2.17)$$

In biological tissues, DWI signals correlate to tissue cellularity. Different features, commonly associated with malignancy and hypercellular metastatic tissues, including water diffusion from extra-cellular to intra-cellular spaces, restriction of

cellular membrane permeability, can be detected by DWI imaging, this explaining the increased role of this modality in oncology. Moreover, due to short acquisition times, that is not more than 5 minutes if more than three different b -values-weighted sequences are acquired, DWI is easily incorporated in routine MRI protocols [19].

Apparent Diffusion Coefficient

Actually, for a quantitative evaluation of DWI sequences, the coefficient D should be computed from Equation 2.16. To this end, the Apparent Diffusion Coefficient (ADC) parametric maps have been introduced. By recalling Equation 2.16, the D can be estimated by exploiting at least a couple of b -values-weighted DWI images, according to Equation 2.18:

$$D = \frac{\ln\left(\frac{S_1}{S_2}\right)}{b_2 - b_1} \quad (2.18)$$

In practice, in human tissues where water molecules cannot freely diffuse, the real diffusion D corresponds to the ADC, measured in one direction. Accordingly, the mean ADC value (ADC_m) is calculated from the gradients applied along the three directions as in Equation 2.19 [13]:

$$ADC_m = \frac{ADC_x + ADC_y + ADC_z}{3} \quad (2.19)$$

ADC parametric maps can represent at voxel level the hidden physical processes occurring at a micrometric scale. A higher number of b -values sequences allows a more accurate computation of ADC maps, where Equation 2.16 is solved through linear regression. The downside is significant motion artefacts that can jeopardize the reliability of ADC maps, especially when more than two b -values are considered. On the contrary of original DWI signals, regions of restricted diffusion are represented by low ADC values (dark signals), that are lower for malignant lesions than benign or healthy tissues [19]. For instance, Figure 2.5 compares ADC (a) image and DWI (b) at high b -value ($b=2000 \text{ s/mm}^2$) where prostate and tumour ROIs are outlined in red and green, respectively. The malignant area is represented by hypo-intensity in ADC maps and high DWI signal. Recently, most of clinical interest on ADC maps relies on therapy assessment based on changing of ADC values at follow up examinations after treatments. With effective treatments, in the short period one can observe a persistent high signal in DWI, but increased (toward normalization) ADC values, while in the long term successful therapies will normalize the DWI signal too [20]. Actually, modern mpMRI protocols are based on the acquisition and joint interpretation of multiple sequences, that is anatomic T_1 - and T_2 - weighted, DWI sequences, and ADC maps, together with functional MRI sequences through dynamic acquisition techniques for a comprehensive view of tumour characteristics [21].

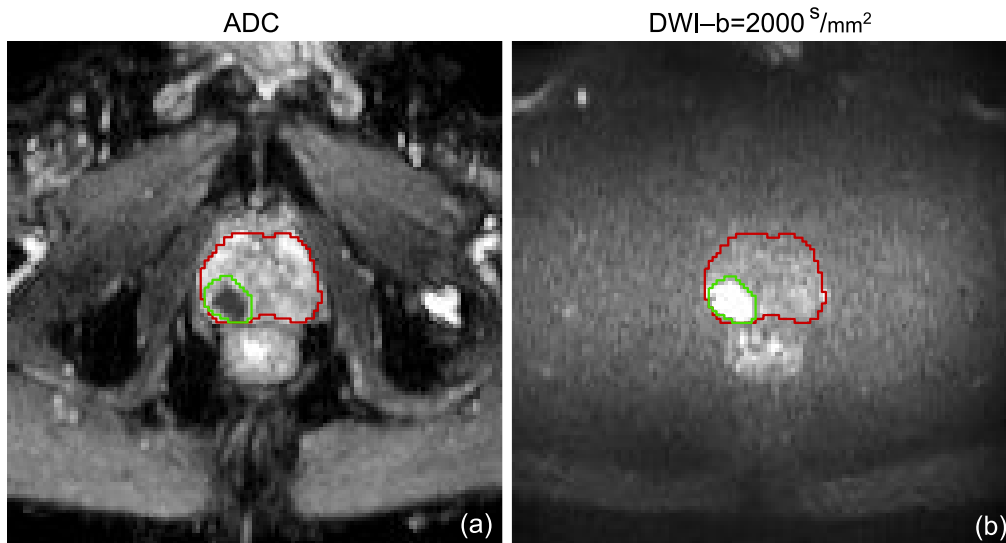


Fig. 2.5: Comparison of ADC (a) and DWI at $b=2000 \text{ s/mm}^2$ (b) images, where prostate (red) and tumour (green) ROIs are outlined. As one can notice, the malignant area is characterized by hypo-intensity in the ADC map and hyper-intensity in DWI.

2.3 Dynamic contrast-enhanced imaging techniques

DCE-CT and DCE-MRI are well-established imaging techniques used to describe the microvasculature of tissue and organs, and any potential alteration hinting at malignancy. Moreover, functional imaging of the tissue hemodynamics allows monitoring tailored treatments based on anti-angiogenic drugs, that act as inhibitors of vascular tumour growth [22].

DCE protocols require the acquisition of a baseline image, in the absence of a CA, followed by a series of images acquired over time after the injection of the CA [23]. The temporal changes in the enhancement of the CA allow characterizing the vascular functionality of the tissue of interest, through the analysis of the so-called Time Attenuation Curves (TACs) or Time Intensity Curves (TICs) retrieved from DCE-CT and DCE-MRI sequences, respectively. Figure 2.6: shows an example of DCE-CT sequence acquired over time after the injection of the CA and, in green, the enhancement of the CA within the tissue. Each image of the whole sequence contributes to reconstruct one sample of the green curve.

Curves of CA over time represents three main phases of tissue perfusion that is, entry of the CA into the capillary network from the arterial systems, circulation in the vascular network, and exit into the venous network. Depending on the CA and tissue under investigations, leakage phenomena can be observed too, from the intravascular space to the interstitium [24].

The presence of CA in blood vessels modifies X-ray attenuation on CT images

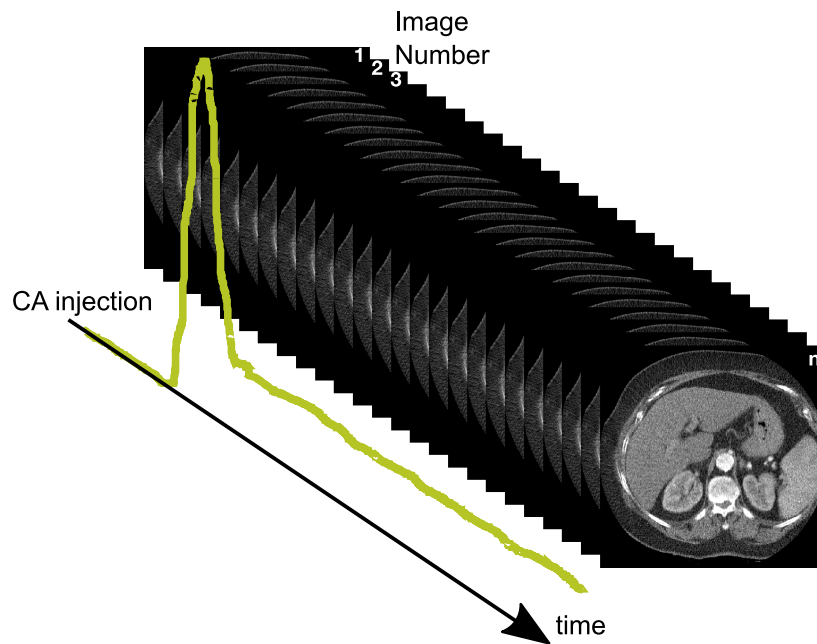


Fig. 2.6: An example of DCE-CT protocol, which requires the acquisition of repeated CT scans over time in order to reconstruct the enhancement of the CA within the tissue, represented by the green curve.

in a linear way and signal intensity on MRI in a non-linear manner. CA employed in DCE-CT are iodinated media, containing iodine atoms, that are administered mainly through intravenous injection. As explained before (Sect. 2.1), interactions between photons also depends on the atomic number of the matter. Since iodine has $Z=53$, very high if compared to most tissues in the body, the administration of iodinated CA produced image contrast due to differential photoelectric absorption. Moreover, the main advantage of using iodine is that the energy level of its electron shell is 33.2 keV, similar to the energy of X-rays commonly employed in CT imaging. The radiodensity of iodinated CA ranges between [25-30] HU per milligram of iodine per milliliter at a tube voltage of 100-140 kVp. Collateral effects and allergic reactions still remain the main issue related to the use of CA. In this regard, widespread iodinated CA are water soluble, non-ionic (organic) compound that can reduce side effects because not dissociate into single components.

As regards DCE-MRI, images can be acquired using both T_2 - or T_1 - weighted sequences, thus allowing exploiting two different contrast mechanisms to reduce T_2 - or T_1 - relaxation times, respectively. One approach is based on the susceptibility effect, that is local magnetic inhomogeneities arising between structures with different susceptibility, such as the CA and the surrounding tissues, that lead to a negative enhanced T_2 -weighted images. The susceptibility effect is properly exploited in those tissues, such as brain, having specific vascular barriers that can prevent the passage

of the CA into the interstitial space, thus allowing catching with high temporal resolution (1 Hz frequency sampling) the first passage of the CA into the tissue. In tumours, where substantial CA leakages occur from blood vessels, the measurements of reduction of T_2 -relaxation time due to the susceptibility effect might be very inaccurate, so that the contrast mechanism most commonly adopted is based on negative enhanced T_1 -weighted sequences (0.2 Hz frequency sampling) produced by the reduction of T_1 relaxation time, mostly derived from the CA which is exited into the interstice [25]. CA for DCE-MRI reduces T_1 or T_2 time in tissue by affecting the amplitude and time scale of the magnetic field variation. The majority of CA employed in DCE-MRI are paramagnetic agents, such as transition or rare earth metals, containing unpaired electrons generating, in their turn, large residual magnetic dipoles. Accordingly, strong magnetic field originate and protons in the proximity experience a reduction of the relaxation times of water molecules. Today, most of MRI CA are based on gadolinium (Gd), a silvery earth metal. In particular, the most common oxidation state of Gd is Gd^{3+} , holding seven unpaired electrons, so promising to originate the largest dipole based on the number of unpaired electrons. Gd^{3+} improves image contrast by increasing T_1 relaxation time, thus appearing bright in T_1 -weighted DCE-MRI sequences. However, Gd^{3+} is a toxic material, and the limited sensitivity of MRI (on the contrary to the high sensitivity of PET imaging) requires high concentration dose, which further emphasizes collateral effects for patient. Significant quantities of Gd^{3+} ($Gd^{3+} > 0.1$ g) must be administered to detect contrast enhancement. Generally, to constrain toxicity, Gd^{3+} is stabilized with an organic chelator, a molecule with high solubility, able to involve Gd^{3+} and prevent pH precipitation. The downside is that the longer distance between water molecules and Gd^{3+} due to chelator, the weaker the contrast in DCE imaging, thus originating some issues related to the needed dose of paramagnetic agent, that however still remains lower than the one required of CA for DCE-CT [13].

The main challenge of the DCE imaging techniques still remains the quantification of the so-called perfusion parameters, that is measures of vascular functionality of the tissue of interest computed by processing the extracted TACs and TICs. The lack of consensus on methods and software for quantitative analysis of DCE images still hampers DCE-CT and DCE-MRI techniques to enter the clinical routine [26], except for some anatomic districts such as brain [27] and heart [28]. If compared to DCE-MRI, DCE-CT show higher simplicity in terms of both acquisition and processing, and also benefits from large availability, low cost, and patient acceptability [29]. On the contrary, the major limit refers to the radiation dose exposure. In fact, the need to reduce X-ray exposure limits the total number of slices acquired and requires to use low tube currents and voltages, thus potentially leading to an increase of image noise [30]. DCE-MRI saves patients from radiation exposure and allows imaging on multiple and arbitrary planes on the internal body with high ei-

ther spatial or temporal resolution, set according to the characteristics of the tumour under investigation. That is, for semi-quantitative analysis where tumour heterogeneity has to be probed at the finest possible level, a high spatial resolution is recommended, whilst high temporal resolutions are needed for quantitative analysis for which CA enhancement requires to be high-frequency sampled to allow accurate model fittings [31]. Although both refers to DCE imaging, data acquisition and processing can be very heterogeneous between DCE-CT and DCE-MRI. Protocols and analytic methods mainly depends on various factors, including the acquisition strategy, referred to selection of the temporal resolution, also as a consequence of the injection dose, mode and type of CA, data preprocessing (motion correction and conversion of signals into TACs/TICs), and, not least, the analysis mode (based on Region of Interests, ROIs, or individual voxels). A detailed workflow for DCE-CT imaging analysis will be discussed in Chapter 4.

Chapter 3

Hallmarks of cancer and tumour biomarkers

Tumour exhibits a set of hallmarks functions and characteristics, varying over spatio and time, and most of them currently being studied [2]. In the era of personalized medicine, understanding with the finest detail the mechanisms underlying tumour biology and inter-relationships occurring within tumour habitat, aims at improving the efficacy of target and tailored therapies, through defining new administration protocols, with the highest possible specificity for the innumerable tumour subtypes [32]. In the following, three main hallmarks of cancer are highlighted, from cell proliferation and tumour growth (Sect. 3.1) to tumour angiogenesis (Sect. 3.2) and heterogeneity (Sect. 3.3). On the ground of these biological mechanisms, cancer research has focused on studying and monitoring cancer evolution over spatio and time, thus highlighting the need for reliable biomarkers for improving the clinical decision making. Hence, the origin of biomarkers, from clinical to imaging ones, and their current employment into clinical practice is deepened in Sects. 3.4 and 3.5, respectively.

3.1 Cell proliferation and tumour growth

Tumour originates from proliferating cells, carrying oncogenic and mutation factors, which acquire through a multistep process tumorigenic features, turning over time into malignant ones. Besides masses of proliferating cells, tumours are referred to as complex tissues, where different cellular subtypes interact with each other and actively participate to the developments of a set of hallmark functions. Also normal cells constituting the tumour stroma, play a role in tumorigenesis since realizing the tumour microenvironment which facilitates tumour proliferation and growth [2]. Genomic instability is the fundamental underlying condition of tumour proliferation

(and also heterogeneity), which induces a series of correlated hallmarks functions of cancer cells, including the sustainment of proliferative signalling, tumour growth, resistance to cell death [33]. In particular, one of the distinctive traits of cancer cells is the capability to promote chronic proliferation. Normal tissues are highly regulated in a controlled system alternating production of growth-promoting signals, cell growth and cell division, thus preserving tissue architecture in terms of number of cells and their organization. Instead, cancerous cells enable alternative signalling pathways, largely regulated by growth factors, and modify the normal cell-cycle, also influencing other biological properties, survival, and metabolism. Promoted by mutated growth factors, hyper-cellularization occurring in tumours dramatically changes tissue structure in terms of the spatial relationships between cancer cells and the surrounding microenvironment. Moreover, regulation of cancer cell proliferation occurs between cells and their immediate neighbours, thus making the comprehension of the proliferation and growth signalling systems very challenging and not yet enough exhaustive [2].

Oncogenic factors in proliferating cancer cells are known to modify, in the so called metabolic reprogramming, the pathways of nutrient acquisition and assimilation of carbon into macromolecules such as lipids, proteins, and nucleic acids, thus sustaining tumour growth. In particular, glucose and glutamine are the two metabolic pathways considerably reprogrammed, for instance by the Ras-related oncogenes. These factors regulate both the nutrient uptake and utilization, thus supporting biosynthesis, redox homeostasis and cell survival. Cancer cells develop the ability to withstand the hard environmental conditions optimising the nutrient resources at the disposal, and also compensating for the loss of glucose or glutamine with alternative metabolic pathways [34].

One of the major issues related to metabolic reprogramming is therapy resistance. For instance, mitochondrial metabolism is one of the alternative pathway of the resistant cancer cells [35], so that mitochondria are recently being considered as promising targets for the development of novel anticancer drugs. In this regard, PET functional imaging which relies in most of clinical applications on glucose metabolism, can contribute in detecting the reprogrammed metabolic processes [36]. Recent studies have proved the efficacy of different imaging modalities in correlating with overproliferation factors suggesting tumour growth. To this purpose, beside PET imaging [37], which is nominally the most specific modality, an increasing role has been gained by functional MRI sequences, including DWI and ADC parametric maps, according to well-established relationships between cell proliferation and water diffusivity [38].

3.2 Tumour angiogenesis

Tumour growth and invasion of the surrounding tissues are strongly promoted by a process referred to as angiogenesis, that is the growth of new blood vessels from pre-existing ones and their proliferation into cancer in order to ensure the supply of nutrients and oxygen. Hence, angiogenesis is another relevant hallmark of tumour progression, which also acts as promoting factor for the development of metastasis and their metabolic sustainment over time [39]. Indeed, tumour are made of cancer cells and stromal cells, which create, as above mentioned, the proper microenvironment for tumour development. In particular, tumour growth and increase in aggressiveness is mostly due to cell-to-cell communications occurring between tumour and stromal cells, also involving signalling pathways for the energetic refill and vascular supply. Under the promotion of endothelial cells, which are a particular kind of stromal cells, angiogenesis occurs [40]. It starts from the degradation of the basement membrane of existing vessels, followed by the migration of endothelial cells in the interstice, where they proliferate and create new vessel connections, thus regenerating the basement membrane and promoting anastomosis. Figure 3.1 shows

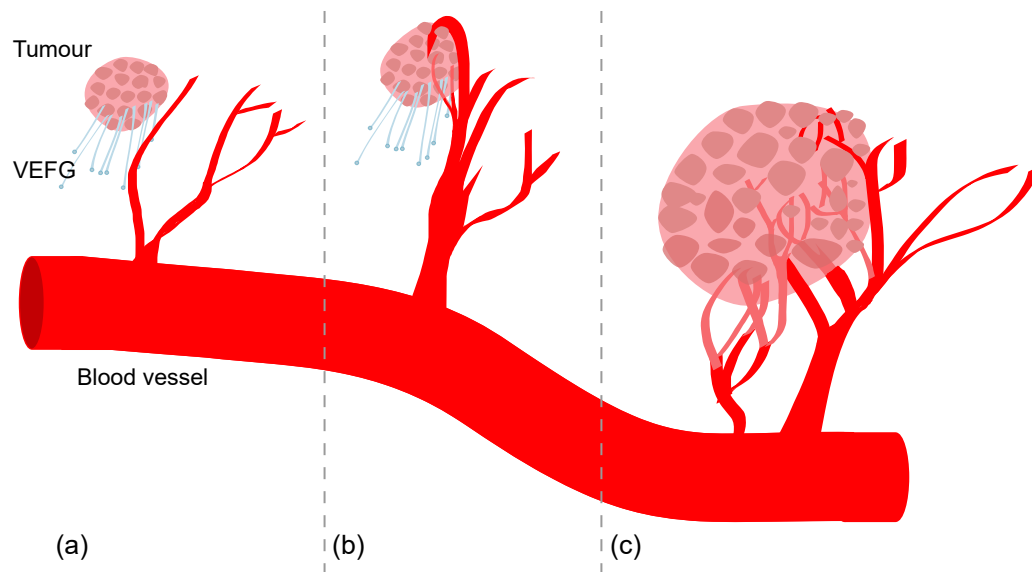


Fig. 3.1: Three stages of tumour angiogenesis: (a) tumour secretes VEGF; (b) VEGF increases blood vessels growth and their proliferation into tumour; (c) tumour with increased blood supply

tumour angiogenesis split into three subsequent phases. First, tumour promotes the production of cellular growth factors. Second, growth factors favour blood vessels growth and their proliferation into tumour. Third, tumour receives an increased

blood supply. Actually, not all tumours induce angiogenesis as soon as they start to grow, but they turn into an angiogenic phenotype when the so called angiogenic switch occurs. Basically, the switch is due to an unbalanced increased production of angiogenic factors with respect to angiogenesis inhibitors [41]. For instance, an over-expression of angiogenic factors can be promoted by Ras-related oncogenes – which also regulate the metabolic reprogramming in proliferating tumour cells (as explained above) – thus showing that the distinct hallmarks of cancers can also be regulated by the same operating factors [2]. Angiogenesis does not only serve for the metabolic sustainment of tumours, but also offers tumour metastatic cells the ways to enter the circulatory system, thus explaining the increase of tumour aggressiveness.

The primary factor promoting angiogenesis is the vascular endothelial growth factor (VEGF) that can be produced by hypoxic tumour cells, endothelial cells, and the so called tumour-associated macrophages. VEGFs include the originally identified glycoproteins called VEGFA, and the family of VEGF-related polypeptides. In particular, VEGFA is reported as both the main endothelial growth factor and a regulator of vascular permeability, largely modified by angiogenesis [42]. Blood vessels produced by angiogenesis are distort and enlarged, with an excessive vessel branching. Microhaemorrhages and leakage phenomena are also present, together with abnormal levels of endothelial cell proliferation and apoptosis [2]. In this regard, Figure 3.2 shows the comparison between the normal and tumour vasculature. While normal tissue are organized in hierarchical pathways transporting blood from arteries to venous vessels, tumour vasculature results locally interrupted, irregular, and disorganized.

Understanding the mechanism underlying angiogenesis is at the basis of cancer research towards the development of effective anti-angiogenic drugs. The US Food and Administration (FDA) has approved for the first time, in 2004, bevacizumab, an anti-angiogenic drug, for metastatic colorectal cancer, also applied in kidney and lung tumour care. Entering of bevacizumab in clinical practice has strongly promoted research in molecular target therapies, and after that, a conspicuous number of anti-angiogenic drugs have been licensed [43]. Like bevacizumab, anti-angiogenic drugs have been initially proposed as VEGFs inhibitors, under the hypothesis that blocking angiogenesis would have reduced tumour growth and improved overall survival of oncologic patients. Actually, this has resulted as a transient effect of anti-angiogenic drugs, which instead act to normalize tissue vasculature, with a cytostatic rather than a cytotoxic effect, thus consequently improving the efficacy of conventional therapies in reaching the target site. Moreover, a normalized vasculature may reduce the shedding of metastatic cells from the primary tumour. In addition, anti-angiogenic drugs decrease tumour hypoxia and improve the absorption of macromolecules from vessels to tumours. Hence, for instance, radiation therapy

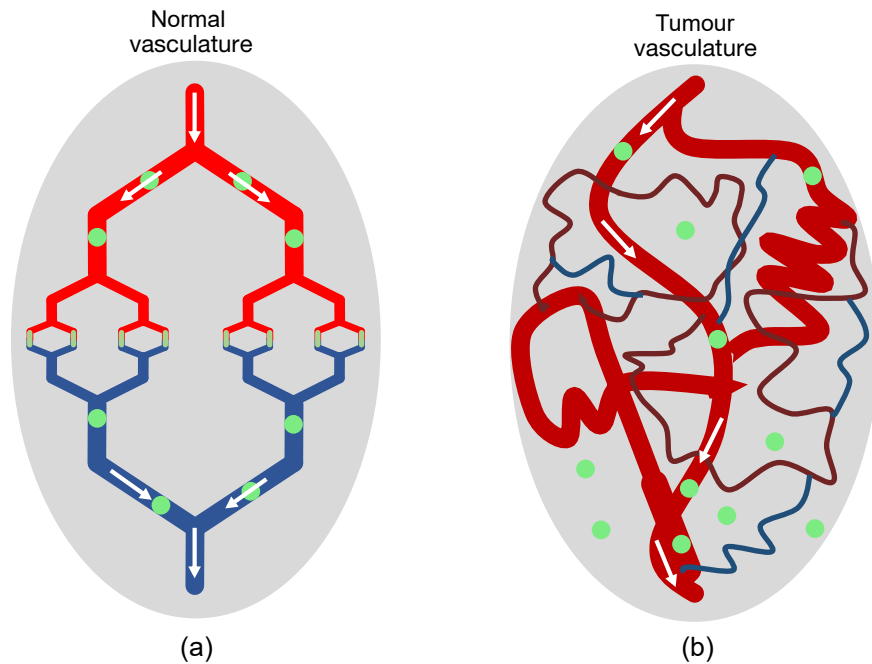


Fig. 3.2: Normal (a) and tumour (b) vasculature

is reported as having a better outcome when delivered during the normalization window of tumours under the effects of anti-angiogenic factors [44].

Assessment of angiogenesis by perfusion imaging

Quantifying tumour angiogenesis has significant implications in clinical management of patients, primarily for tumour staging and monitoring therapy response [45]. The assessment of therapy efficacy is conventionally based on tumour size changes, which may result inadequate for evaluating response to anti-angiogenic drugs. From a clinical point of view, a more reliable indicator is time to progression. However, time to progression requires very large studies with numerous patients in order to be validated as clinical endpoint, with the risk of treating ineffectively patients for a long time [46]. Alternatively, clinical standard adopt biological markers derived from the histopathological study, which is exploited for characterizing tumour features, associated to several biological functions, most of which referring to the surrounding tumour habitat. Histopathological markers for angiogenesis include the microvascular density (MVD), microvascular cell proliferation (MVCPC), and total vascular area (TVA), albeit measurements are limited by local tumour heterogeneity and the reduced number of biopsy samples [45]. In addition, there is a variety of tumour subtypes showing different responses to therapy even if being assigned to

the same grade of aggressiveness, therefore biological markers result in most of cases insufficient to provide a comprehensive view of the disease [47].

The amount of blood normally exchanged into the vascular network is modified by angiogenesis, therefore evaluating tissue perfusion may be very useful to measure hemodynamic changes and identify which perfusion parameters correlate better to patient prognosis [48]. From a theoretical point of view, the study of tissue perfusion dates from the nineties, when in physiological modelling studies, perfusion has been defined as the transport of blood to a unit volume of tissue per unit of time [49]. Advances in cancer care through novel anti-angiogenic drugs have renewed the interest in the study of tissue perfusion, also strongly promoted by improvements in dynamic imaging technologies. In fact, dynamic imaging offers an incredible opportunity to monitor tissue hemodynamics through the study of the kinetics of a tracer injected into the vascular compartment. DCE-CT and DCE-MRI techniques, also referred to as perfusion imaging, allow deriving from the concentration of CA over time perfusion parameters, including Blood Flow (BF), Blood Volume (BV), Mean Transit Time (MTT), Permeability Surface (PS), Transfer constant (K_{trans}), Extraction Fraction (EF), and Perfusion Index (PI), which represent the status of tissue vasculature. Ultimately, perfusion parameters allow detecting any abnormality indicative of angiogenesis, and also observing improvement or worsening depending on the efficacy of a treatment [50]. It is worth noting that as morphological changes in tumour size occurs later than variations in microcirculation, the perfusion study, although with a limited translation into clinical routine so far, can be a very useful tool for assisting the clinical decision making [48]. Perfusion imaging can measure the blood flow within the capillary network of tumour tissues and towards the extravascular space. In fact, microvasculature involving mainly capillaries, is characterized by a slow flow, which permits exchanges with the supplied tissue. Capillary permeability allows a compound of nutrients, hormones, and mediators to diffuse into the interstitial space.

Referring to perfusion parameters, BF measures the quantity of blood reaching one unit of tissue volume per unit of time, and is expressed in $[\text{ml}/\text{min}\cdot 100\text{g}]$ or $[\text{ml}/\text{min}\cdot 100\text{ml}]$, if considering the unit of volume at the denominator with unit density. Accordingly, BF is equivalently expressed in $[\text{min}^{-1}]$.

BV represents the fraction of blood volume contained in capillaries with respect to a unit tissue of volume. This ratio is expressed in $[\text{ml}/100\text{ml}]$ or in %.

MTT is the mean time, measured in $[s]$ in which blood flow passes through the capillary network, that is the time between the inflow through a large artery and its outflow through a vein.

PS indicates the flow of molecules through the capillary membrane in a unit tissue of volume, therefore expressed like BF in $[\text{ml}/\text{min}\cdot 100\text{g}]$. In many cases, PS depends not only from the status of vasculature, but also from the intrinsic properties

of the CA, which determine its tendency to exit the intravascular compartment.

K^{trans} is a combination of BF and PS derived under proper modelling conditions, also depending from the mathematical model employed to describe the kinetics of the CA. Accordingly, K^{trans} is measured in [ml/min·100g]. Physiologically, K^{trans} measures the component of the flow that reaches the extravascular space, this means that, in practice, it describes the delivery of nutrients to tissue cells.

EF is the volume of CA within the interstitial compartment, that is the extravascular component of CA in unit tissue of volume. Like BV, EF is measured in [ml/100ml] or in %.

Finally, in particular organs, such as lung and liver, where blood is supplied to the capillary network from two main vessels, PI measures the proportional contribution at the inflow of a tissue unit of the two inputs, that is a ratio of the BF coming from one of the two inputs, over the total BF [24].

Several research studies have demonstrated a correlation between perfusion parameters and biological markers of angiogenesis [51], and diagnosis [52] and staging [53] of different tumour subtypes, albeit the exact relationships between them are not yet well-known. Nevertheless, some frequently observed phenomena can be mentioned: BV and BF both correlate with MVD, however, in most studies BV has resulted to perform better than BF in depicting MVD status. In fact, BF measurements can be affected by other factors, such as cardiac output, above all in tumour vessels lacking of auto-regulatory mechanisms, where if the cardiac output falls down, one can observe a misleading decrease of BF [54], accordingly. In addition, MTT is expected to decrease in presence of arteriovenous shunts, typical occurring in tumours, while BV and permeability-related perfusion parameters have shown to increase in response to tumour adaptation to hypoxia, that is definitely associated to modifications towards more aggressive tumour phenotypes and therapy resistance [55]. Depending on the tracer kinetic model adopted, different perfusion parameters can be computed. In this regard, the major cause hampering perfusion imaging to enter clinical practice relies on the difficulty to achieve reproducible results with different mathematical computing methods. A methodological contribute in quantitative CTp imaging is given in Chapter 6, after explaining the theoretical background in Chapter 4.

3.3 Tumour heterogeneity

At the smallest clinically detectable tumour size, of approximately 0.5 cm³ [56], cancer cells are reasonably homogeneous, but they show a strong tendency to differentiate in tumour subtypes when proliferating [2]. In fact, tumours exhibit substantial spatial variations at macroscopic level, and also in gene expression, biochem-

istry, and histopathology, referred to as tumour heterogeneity. Its origin arises from both genomic and phenotypic causes, which may influence prognosis and response to therapies [57]. By conceiving the tumour as a molecularly dynamic disease, also consequently to angiogenesis, mapping tumour heterogeneity over spatio and time is a further essential requirement for developing effective and durable therapeutic treatments [58].

According to the Nowell's theory of clonal evolution, cancers arise from single cells and subsequently develop genomic instability, that is known as one of the major causes of tumour heterogeneity, and occurs consequently to high rate errors during DNA replication. Then, evolutionary selection pressures on expanding tumours affected by genomic instability, lead to the emergence of genetic mutations and heterogeneous subpopulations of tumours [59]. Nowadays, the different techniques for genoma sequencing have allowed identifying the so called genetic signature of many mutagenic processes, and even chemotherapy itself is known to create and enforce genomic instability. Actually, it is such a combination of genomic instability with other factors which cooperates for tumour malignancy, metastasis development, and tumour progression [58]. Moreover, the tumour microenvironment has an effect on the different cellular phenotypes that may proliferate into different tumour regions. For instance, any tumour contains areas of hypoxia and normoxia. In the hypoxic region, cells with anaerobic metabolism will be advantaged and will repopulate that tumour region, whilst the opposite will occur in a normoxia area of the same tumour, thus increasing the phenotypic heterogeneity within tumour [60].

Indeed, tumour heterogeneity comprises both intra-tumour and inter-tumour heterogeneity. While the first refers to the different lesions present in an individual tumour, the second indicates differences between tumours within individual patients. In particular, some relevant features can be highlighted for the intra-tumour heterogeneity. It can be dynamic and tends to increase together with the tumour growth. Accordingly, high levels of heterogeneity are often associated to poor prognosis, due to the cell resistance to therapies. Moreover, in practical cases, heterogeneity may show dependencies not only from the underlying tumour biology, but also based on the imaging test used [32].

In the era of personalized cancer medicine, the investigation of intra-tumour and inter-tumour heterogeneity results very challenging for both clinicians and researchers, especially for the development of reliable clinical indicators of tumour progression and aggressiveness. Heterogeneity is a clear limit to pursue personalized treatments if only biopsy-based approaches are considered for tumour characterization, since the discrete sampling of a needle biopsy, although being the standard of care for clinical assessment of prognosis and aggressiveness of suspicious malignant tissue areas, cannot represent the entire tumour phenotypes, and can miss some lesions with a poor prognosis signature, thus leading to underestimate tumour

aggressiveness [61]. Moreover, quantitative measurements commonly employed in clinics may be misleading in depicting tumour characteristics, since are often based on average values computed over the whole heterogeneous tumour, thus resulting dramatically unspecific [59]. Lastly, another clinical implication of heterogeneous tumours is the biochemical recurrence. Depending on the grade of intra-tumour heterogeneity, some cells can contain a pre-existing resistant clone that is free to repopulate the tumour even after a chemotherapeutic treatment.

Each tumour results as a uniquely molecularly characterized microenvironment, this explaining the meaning of inter-tumour heterogeneity. In practice, this limits the applicability of the cancer biomarkers over distinct types of tumours, and also among the same tumours that have evolved along different carcinogenic pathways [60]. Altogether, insights into tumour heterogeneity strongly support an approach aimed at a local regional quantitative assessment, which may increase the overall accuracy in detecting even the smallest morphological and phenotypic subtypes of tumour populations, consequently improving efficacy of target therapeutic strategies [32].

3.4 From clinical to imaging biomarkers

From a clinical point of view, effectiveness of cancer care shows a notable dependency from the availability of markers of tumour features and biological functions. The ways through which these tumour markers can be derived determine how often they can be measured, and which degree of specificity they hold. A biomarker has been defined by the FDA as a “*characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention*” [62]. In a non-invasive approach body fluids such as plasma, serum or urine can be analysed to study the presence and concentration of specific substances. For instance, the Prostate Specific Antigen (PSA) level in blood plasma represents a clinical biomarker of PCa expression. Alternatively, clinical biomarkers of tumour stage and therapy response are derived from immunohistochemistry and histopathology, both requiring a tissue sample to be analysed [63]. Medical imaging in cancer care has a fundamental relevance since allows a real time monitoring of the disease, without tissue destruction, at different spatio and temporal resolutions, also with a minimal invasiveness [64]. Its current clinical employment in cancer management is substantially based on visual assessments or morphometric semi-quantitative analyses, thus lacking of any approved standardized quantitative approach [43]. In agreement with the concept of precision medicine, the role of medical imaging has evolved within oncological practice, from being primarily a diagnostic tool to be integrated in clinical decision making for tumour staging, prognosis, gauging the effectiveness of therapies [65]. The spatial and temporal tumour heterogeneity arising

from local variations in tumour growth, cell metabolism, vasculature, oxygenation, can be quantitatively measured by computing, from morphological and functional imaging (e.g. CT, MRI, and PET) innumerable imaging features representing various aspects of tissue properties [66].

Indeed, the employment in computer vision of imaging features with predictive purposes dates from the sixties, whilst its application in medical imaging started in the eighties [67]. Actually, most of early applications refer to Computer Aided Detection (CADe) and Diagnosis (CADx) systems, the former with the aim of detecting, localizing and segmenting tumour lesions, the latter employed to distinguish benign and malignant lesions. Since the last two decades, quantitative approaches in medical imaging started to be applied for developing imaging biomarkers (IBs) of tumour prognosis and therapeutic endpoints [68]. In 2007, the Radiologic Society of North America (RSNA) organized the Quantitative Imaging Biomarkers Alliance (QIBA) involving researchers, healthcare professionals, industry stakeholders, and scientists to regulate the use of IBs in clinical trials and promote standardized approaches of quantitative imaging, from imaging feature extraction to data analysis and display and reporting methods, with the aim of accelerating IBs translation into clinical practice [69]. With the rapid increase of technologies and computing resources allowing reconstructing higher quality images with incredibly low processing times, the interest in IBs is strongly grew, from both clinical and research side. If compared to biological and chemical biomarkers, currently employed in cancer care, IBs benefit from non-invasiveness, immediateness of measurements, high level and broad extent of information intrinsically detected. In fact, quantitative imaging techniques for feature extraction (e.g. histogram and texture analysis, Fourier and wavelet transforms) allow analysing image properties in a transformed domain (e.g. probability, frequency, or time-frequency), where measurements represent tumour biology even beyond what images offer at a visual-based inspection [70].

3.5 The role of biomarkers in clinical practice

In clinical practice, biomarkers are organized into three main categories, that is *diagnostic*, *prognostic* and *predictive* biomarkers.

For the sake of clarity, biomarkers need to be distinguish from *surrogate endpoints*, that are intended to substitute a clinical endpoint, that is a characteristic reflecting patient feel and symptomatology. Biomarkers can refer to *surrogate endpoints* before being definitely translated into clinics, when larger prospective trials are performed to confirm the association between the measurement and the clinical outcome [43]. The application of *surrogate endpoints* takes an additional complexity and is limited to those cases in which the pathophysiology of the disease is fully

understood, otherwise they might generate pitfalls in the disease assessment [71].

Diagnostic biomarkers are used for detecting or confirming the presence of a disease or any its subtype, or a characteristic condition of clinical interest. As mentioned above, PSA is the *diagnostic* biomarker employed for PCa detection, or for instance, Carcinoembryonic Antigen (CEA) levels in blood is substantially raised by the colorectal cancer. Indeed, measurements of biomarkers have not perfect clinical and analytical sensitivity and specificity, so that the clinical applicability of a biomarker relies on the best trade-off between these two features, also depending on the role of the biomarker itself. *Diagnostic* biomarkers are likely employed in screening tests, thus requiring high sensitivity rather than high specificity. This means that a number of diseases higher than the real one can be incorrectly detected, with the advantage of reducing at the minimum level the miss of undetected tumours [72]. *Prognostic* and *predictive* biomarkers both provide information on the status of a tumour or disease, but while *prognostic* biomarker quantify the severity condition and tumour aggressiveness irrespective to treatment, *predictive* biomarkers substantially quantify the benefit received from a treatment compared to the baseline condition [73]. Actually, all biomarkers can have a certain degree of both *prognostic* and *predictive* role, and recognising their difference can have a strong impact on the management of disease. On the one hand, a *prognostic* biomarker that is incorrectly interpreted as *predictive* one may lead to an overestimation of the benefits of the treatment, for instance a pharmaceutical drug, that consequently will be considered as a tailored treatment for a subtype of disease, thus limiting its availability to a specific portion of the population, and its cost will increase accordingly. On the other hand, if a *predictive* biomarker is intended as *prognostic*, its power will be underestimated, since its effect on a specific subtype of disease will be hidden in the more general context of patient prognosis [74].

3.5.1 Diagnostic biomarkers

All altered cellular processes and modified cell biologic properties, constituting the hallmarks of tumour development and growth, can potentially serve as *diagnostic* biomarkers for cancer. Since the most of cancer cell mutations occur at molecular level, involving DNA, RNA, microRNAs and proteins, molecular biomarkers are currently the most widely adopted for cancer detection. Moreover, molecular *diagnostic* biomarkers have gained a particular interest due to the worldwide diffusion of the personalized medicine approach for cancer treatment. Accordingly, the employment of target molecules for cancer diagnosis advantages the development of specific drugs against these target molecules [75]. Conventionally, cancer is diagnosed through morphological assessments based on imaging, and antibody staining in biopsy or tissue samples. However, recent developments have made possible the

detection of cancer from a variety of molecular indicators present in body fluids such as saliva, urine, broncho alveolar and ductal lavage, thus improving the specificity of the diagnosis itself and increasing the precision of the detected tumour features [76]. Actually, due to the high level of heterogeneity of tumour subtypes and molecular processes involved in malignancy, there is a plethora of *diagnostic* biomarkers currently employed, each referring to a specific clinical application, and most of them rely on molecular indicators [77]. In this context, diagnostic imaging still remains the reference tool for assessing those macroscopic morphological features of tissues that may hint at tumour proliferation. The employment of IBs in most of cases provides a companion diagnostic status, together with molecular indicators, in order to reveal information of prognosis and therapy response [78].

3.5.2 Prognostic biomarkers

The underlying idea of clinical *prognostic* biomarkers is cancer staging, at diagnosis as well as in follow-up examinations, that is, measuring the extent of severity of cancer disease. Cancer staging plays a fundamental role in defining the most proper treatment and assessing the changes of a successful clinical outcome, thus definitely quantifying patient prognosis [79]. Cancer staging can be performed through morphological assessments of tumour extensions, dynamic analysis of tissue perfusion, and biological evidences of tumour patterns at hystopathology. In particular, the stage of a disease does not only reflect the rate of growth and extension, but also the type of the relationships with the surrounding tissue microenvironment.

The most widely adopted staging system is the TNM, currently at the eighth edition, that code the extent of primary tumour (T), regional lymph nodes (N), and distant metastases (M), thus defining stage groups according to T, N, and M status. At the basis of the TNM classification there is the clinical observation that survival rates are higher for localized diseases rather than tumours that extend beyond the original tumour site. Each of the T, N, and M stages is further split in different subclasses of progressive tumour extension. Moreover, in clinical practice, TNM classification is specified for each tumour type or small groups of them sharing some similar behaviour in tumour growth. However, some general rules can be highlighted. The class T comprises the stage Tx, referring to primary tumour that cannot be assessed, T0, with no evidence of primary tumour, the stage Tis, meaning carcinoma in situ, and four progressive stages from T1 to T4 describing the increase in size and local extent of the primary tumour. Similarly, the stage N comprises the groups Nx, N0, and three progressive classes from N1 to N3 describing the increasing involvement of regional lymph nodes. Finally, M0 and M1 stages describe the absence and presence, respectively, of distant metastases. Actually, there are two different types of clinical and pathological TNM classification. The first of them is

a pretreatment classification, essential to select the most adequate therapy. This clinical TNM staging arise from information collected through imaging, endoscopy, biopsy, and surgical exploration. Instead, the pathological classification is performed after surgery, and it is used to provide additional information to estimate prognosis and end outcome, as well as to guide adjuvant therapy planning [80].

Clinical pre-treatment cancer staging can be only partially derived from imaging, since the visual-based interpretation of CT, MRI, or PET images results unspecific to differentiate benign from malignant lesions, as well as the degree of lesion aggressiveness. For instance, the American College of Radiology (ACR) has proposed a standardized Reporting and Data Systems (RADS) for different tumours, such as prostate (PI-RADS), liver (LI-RADS), colon (C-RADS) and others, with the aim to reduce variability in reporting image findings [81]. Different RADS also attempt to provide a risk assessment of tumour malignancy from images, thus scoring tumour lesions with numbers corresponding to qualitative interpretations in terms of “normal”, “negative”, “benign”, “probably benign”, “high risk”, or “definitely malignant”, that however have shown in clinical practice a limited and unspecific *prognostic* value [82].

In clinical practice, cancer stage-related pathological information are derived from biopsy, that is currently the reference procedure for obtaining pathological proof of tumour metastases or to determine the spread of a tumour. In patients with more than one primary tumour and diffuse metastasis, biopsy is useful to determine which primary tumour originate metastases. Moreover, a biological inspection of biopsy samples is exploited to obtain a phenotypic representation of the tumour [83]. Accordingly, different staging system based on biopsy outcomes have been developed for several types of tumours, such as the Gleason Score (GS) for PCa [84]. Biopsy has been proposed as a random-sampling-based approach for studying some portions of tissue, guided generally by ultrasonography to reach the site of interest. Nowadays more image-guided procedures are also available, such as MRI-guided biopsy or MRI-fusion biopsy, which improve the safety and the efficacy of the procedure by assisting the needle exploration of tumour site with more detailed morphological images, thus leading to sample more accurately the most supposedly aggressive areas. However, underestimation of tumour aggressiveness still occurs, and despite developments in execution techniques, biopsy still remains an invasive examination, that can originate considerable side effects, including bleeding and infection [85]. Current guidelines suggest reducing the biopsy rate especially for low-grade diseases, thus endorsing IBs for revealing tumour prognosis [86].

3.5.3 Predictive biomarkers

Imaging is currently exploited for objectively characterizing tumour response and defining clinical requirements for novel anticancer agents. The availability of *predictive* biomarkers allowing an early assessment of tumour response plays an important role in defining and adjusting target cancer therapies. Among the several criteria for evaluating tumour response, the RECIST system, introduced in 2000, is nowadays, with the RECIST version 1.1 (RECIST1.1) the most widely employed standardized method for solid tumours, based on a morphological evaluation of tumour burden size changes over time, after treatments. RECIST1.1 criteria establish four main categories of tumour response according to the size of tumour burden, defined through the sum of measurements of unidimensional diameters of target lesions. In particular, the longest diameter, greater or equals to 10 mm, is considered for nonnodal target lesions and the shortest axis, greater or equals to 15 mm, for target lymph nodes, with a maximum number of assessed target lesions equals to five, up to two lesions in any organ. Hence, tumour response is considered complete if all nonnodal target lesions disappear and the shortest axis of nodal lesions becomes lower than 10 mm, while, response is considered partial if more than the 30% of decrease of all target lesion measurements is observed between baseline and follow-up assessment. Moreover, tumour is considered stable if no size changes are reported, or in progression if an increase of more than the 20% and 5 mm in the sum of all measurements is registered, or one or more lesions appear, or progression is observed in previous non target lesions [87].

Despite its large employment, RECIST1.1 system has well-known limitations, such as in case of lesions that do not meet criteria to be defined as target lesions, or in case of lesions with particular growth patterns and localization. For instance, lesions with complex shapes and irregular margins do not allow accomplishing the standardize procedure for measuring lesion diameters. Moreover, RECIST1.1 criteria have been initially conceived for assessing the effects of cytotoxic drugs, such as chemotherapy, that can be considered an effective treatment if leads to tumour shrinkage and limits the growth of new lesions, since mostly acting to inhibit cell duplication. However, as explained above (Sect. 3.2), new target therapies act on different tumour mechanisms, including hormonal interferences, signals transduction, inhibition of angiogenesis, therefore, a tumour regression in the short time after treatment is just minimally observed, and such tumour response criteria based on size and morphological tumour changes can reasonably underestimate the effects of treatments. Modified RECIST criteria have also been proposed to overcome the limitations of RECIST1.1 in monitoring new therapies, although some limitations still remain [88]. In the end, it is worth noting that RECIST criteria report up to 10% and 30% of error in intra-observer and inter-observer comparisons, respec-

tively [89]. In all, this strongly motivates research in IBs for improving the accuracy of therapy response evaluation.

Chapter 4

Quantitative CTP image analysis

This Chapter presents the fundamental of CTP analysis, from the theoretical comprehension of tissue modelling to the operative pipeline adopted to compute perfusion parameters, thus deriving insights into tumour vascularity, angiogenesis, hemodynamic evolution of malignancy.

First, the extraction of Time Concentration Curves (TCCs) from CTP images is described in Sect. 4.1. Second, the Chapter focuses on modelling aspects. In particular, the choice of the model depending on the physiology of the organ (Sect. 4.2) and kinetics of the CA (Sect. 4.3). Then, the most used fitting models for vascular and tissue TCCs are addressed in Sect. 4.4. Hence, the computation of perfusion parameters through different methods is presented in Sect. 4.5. In the end, an overview of the major factors contributing to hamper the translation of CTP into clinics for most organs, is provided in Sect. 4.6.

4.1 From image sequences to TCCs

As introduced in Chapter 3, Sect. 3.2, CTP image analysis allows measuring tissue perfusion, modified by angiogenesis or anti-angiogenic therapies.

Dynamic image protocols require repeated scans of a tissue of interest acquired before, during, and after the intravenous injection of a CA, in order to follow the vascular dynamics of the CA into tissue [90]. Basically, information contained in the stack of image slices acquired repeatedly over time is represented through time signal, the TCCs. These measure the enhancement of the CA in terms of its concentration, compared to the baseline value, which represents the CA unenhanced stage [23]. Before computing perfusion parameters, some preprocessing steps are needed, including segmentation of ROIs, image denoising, computation and subtraction of the baseline value from the original TCCs.

As regards segmentation, a tissue ROI is generally outlined on a reference slice

and co-aligned over all the slices of the entire sequence. Accordingly, a time signal arises from each pixel within the ROI, where each sample refers to the value of that pixel in an image slice. Figure 4.1 shows a stack of CTP images, and the corre-

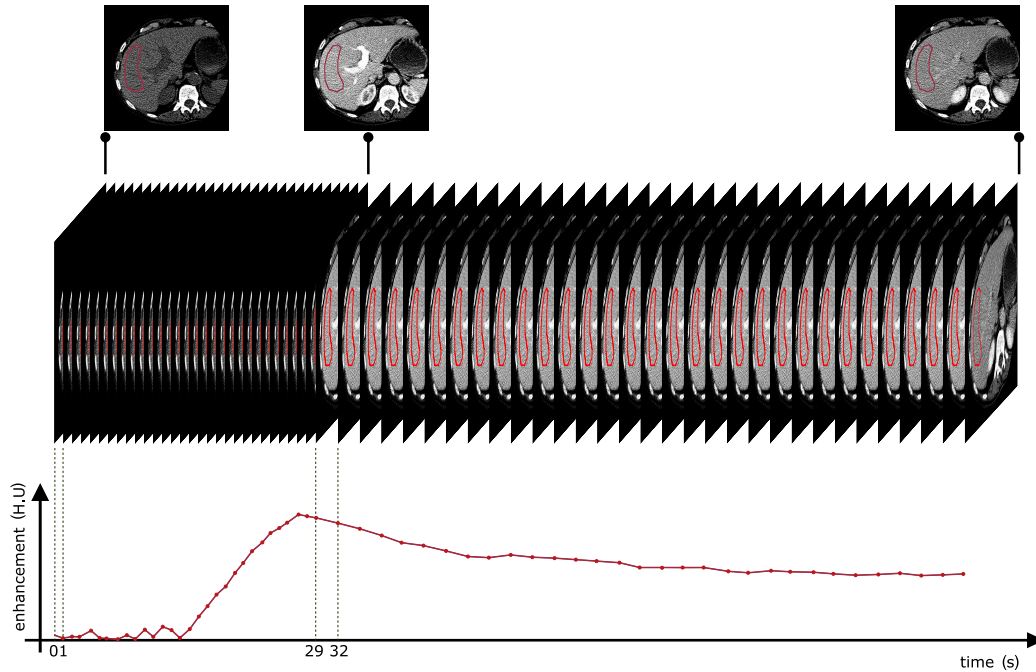


Fig. 4.1: A stack of sixty DCE-CT examinations acquired each second for the first thirty seconds and each three seconds for the remaining time. In red, a ROI is drawn on the hepatic tissue and the signal derived from the median value of the ROI is represented, where each sample of the signal corresponds to one image of the entire stack.

sponding time signal derived from the median pixel within the segmented ROIs in the liver. Generally, the temporal resolution of the signals retrieved varies depending on the imaging modality considered. Moreover, depending on whether the CTP study focuses on the CA first passage only or even on its recirculating phase, the temporal resolutions can be varied during acquisitions (Sect. 4.3). The example in Figure 4.1 represents a hepatic perfusion study mainly focused on the first passage of the CA into tissue, therefore images are acquired each second for the first 30 seconds, and each three seconds for the remaining samples. Based on the tissue under investigation, a further ROI is drawn on at least one vessel, representing the tissue's main vascular supply, and a time signal is derived accordingly [91]. For instance, Figure 4.2 shows for the same example in Figure 4.1, the signals extracted from the aortic and portal vein ROIs, which represent the two main hepatic vascular inputs. The segmentation of the input vessels shall prefer a feeding vessel big enough to avoid partial volume effects [92]. In addition, the vascular input has to be chosen at the minimum distance from the tissue being analysed, in order to reduce the delays

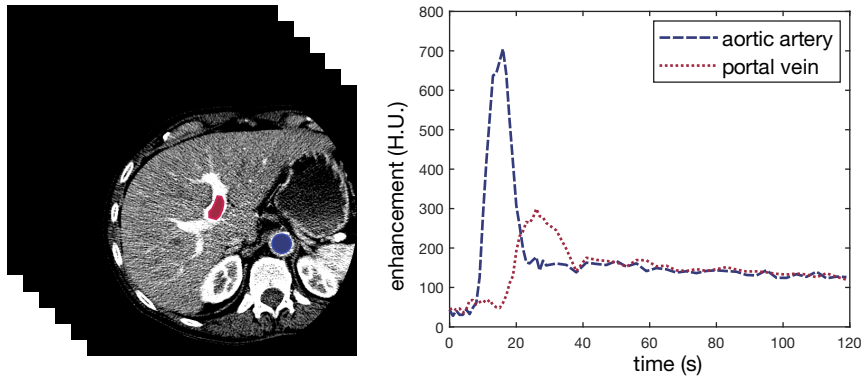


Fig. 4.2: Hepatic vascular inputs in DCE-CT examinations of liver: in blue, the signal extracted from the aortic artery; in green, the signal from the portal vein.

effects on the computed perfusion parameters.

The need for longer acquisition times for dynamic analysis rises the issue of the excessive patient radiation exposure, which imposes to diminish CT tube current and voltage at the expense of image quality. Accordingly, in the clinical practice, to reduce image noise, slice thickness is enlarged and images are reconstructed with a lower in-plane resolution [91]. Even considering to decrease the number of slices acquired, that is the coverage of the volume of interest or the sampling frequency, CTP images still have a poorer quality if compared to standard static morphological CT images. Image noise can dramatically degrades the quality of signals and the accurateness of computing methods applied on TCCs, as well as the reproducibility of results, thus jeopardizing the reliability and clinical usefulness of the CTP itself [50]. Moreover, both respiratory and peristaltic motion of patients represents one additional source of noise in dynamic signals. Acquisition are often carried out in breath-hold conditions, although this limits their temporal duration [91]. Hence, in most of cases, unreliable pixels need to be discarded before any perfusion quantitative measurement [93]. Exploiting spatio-temporal filters for CT images is the most commonly solution adopted to correct for the presence of acquisition noise and motion artefacts. In particular, edge-preserving median filters are largely applied to CTP images to remove spurious pixels while preserving information content of the structures and their boundaries. Commonly, a kernel size of 3 [94] or 5 [28] pixels is adopted, or even 7 pixels as reported in Figure 4.3. In particular, it shows the effects of an anisotropic median filter with kernel size of 7 pixels on some TCCs extracted from original (a) and filtered (b) images, where the mean of coefficient of variation (CV) of all signals, largely employed to measure signal reliability in perfusion analysis [95], diminishes by 32% referring to original images and 4% considering the filtered ones.

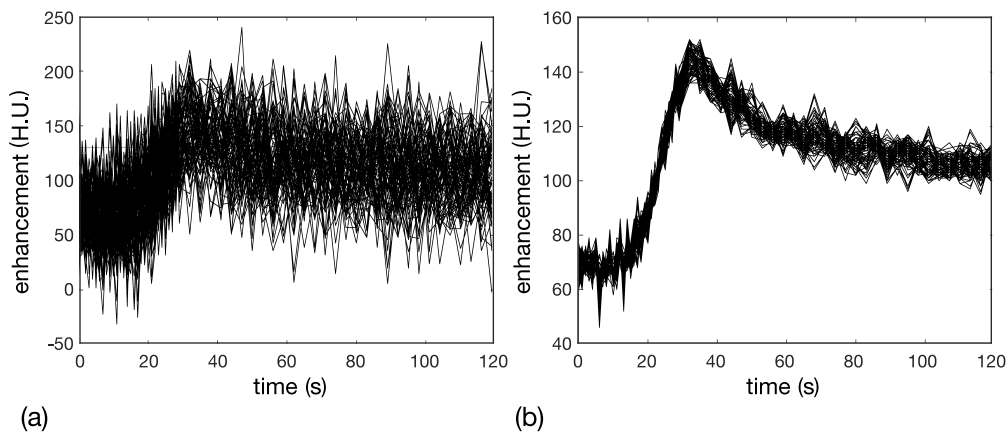


Fig. 4.3: Comparison between signals extracted from CTP before (a) and after (b) filtering images with a median filter with kernel of size 7 pixels.

The last preprocessing stage refers to the baseline computation and its removal from the original TCCs to achieve as many TACs. The literature proposes different approaches for the computation of baseline values, based on voxel-based measurements or referred to the entire ROI. For instance, one of the most widely used approach is to adopt for all the TCCs within the ROI, one global baseline value only, computed as the mean of all the TCCs at the first time instant [96]. Instead, two different voxel-based approaches consider one baseline value for each ROI's TCC, derived from the first [51] or the mean of the first four or five image slices [97]. Actually, image noise influences the computation of perfusion parameters as well as the extraction of CA baseline values, above all if only one or a few samples are considered for the computation. Therefore, in order to improve the robustness of baseline values, a very recent study [98] proposes a voxel-based approach based on an adaptive computation of the maximum number of baseline samples, thus exploiting as much as possible signal's information from which deriving a more reliable baseline value. Finally, Figure 4.4 shows the TACs corresponding to the TCCs reported in Figure 4.3(b).

4.2 Physiology of the organ: single or dual input model

TACs measure the average concentration of CA within image voxel which varies consequently to its flow within the tissue. A basic assumption of CTP imaging is that each voxel is supplied by an impermeable artery which transports the CA to a permeable capillary, from which, in its turn, it comes out into the interstitial space. Then, CA is driven out of the voxel through an impermeable vein [99]. Accordingly, tissue voxel is mathematically modelled using a single input model,

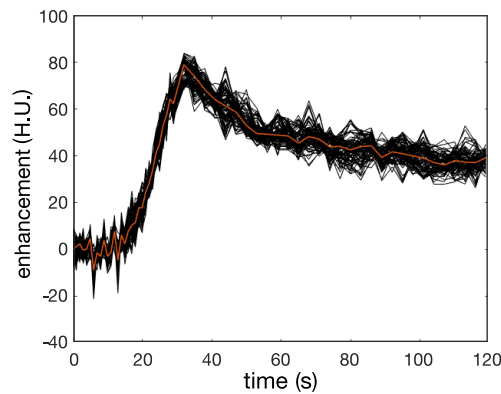


Fig. 4.4: TACs achieved from the TCCs in Figure 4.3 after baseline removal. In red, the median TAC.

where the tissue box, independently from its internal structures defined based on the CA kinetic model adopted (Sect. 4.3), is linked to the vascular network through one input and output pathways, as shown in Figure 4.5 (a). Actually, some human

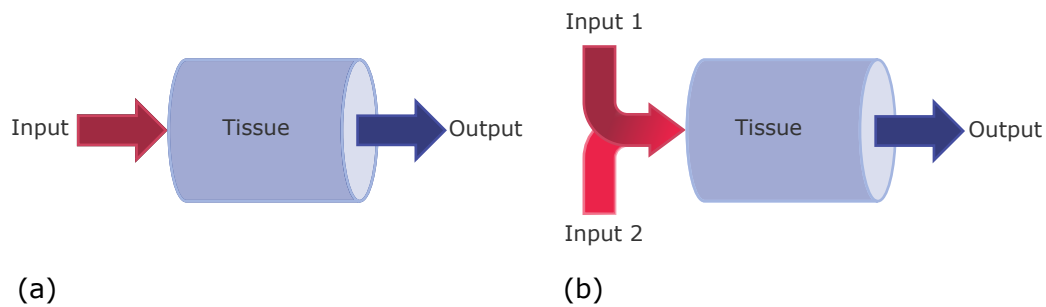


Fig. 4.5: Selection of the model to be used for studying tissue perfusion based on the physiology of the organ: a single input model with one only inflow and outflow (a) and a dual input (b) model accounting for the two vascular inputs supplying lung and liver.

organs, including lung and liver, receive a substantial blood supply from two main vessels, therefore the employment of such a single input model may underestimate any quantitative measurement of tissue perfusion [100]. In those cases, a more accurate representation of tissue physiology is achieved by considering a dual input model, as the one shown in Figure 4.5 (b), where the two inputs mix with each other at the inflow of the tissue capillary bed and the tracer concentration observed in the tissue and reaching the unique outflow pathway results from the proportional contributions of the inputs themselves.

For instance, in lung, blood supply is split between the pulmonary and bronchial circulation, therefore the two main input vessels considered in perfusion analysis are the pulmonary and aortic arteries [101], whilst in liver, the blood reaches the

tissue from the aorta and the portal vein [102]. Actually, because of difficulties in imaging the second vascular supply, single input models have been largely adopted for lung and liver too. In particular, in lung, before the widespread of 320-row CT systems, the z-coverage provided by the previous CT scanners did not allow imaging the pulmonary artery within the maximum temporal limit imposed by dynamic acquisitions of thorax [103], and the same happened in liver, where in order to image the portal vein or its major branches, the region of the hepatic hilum need to be included, with consequent technical difficulties due to the narrow z-coverage [104]. In addition, in liver, the use of single input model has been largely justified by considering that the major blood supply in hepatic tumours arises from the arterial inflow [105]. Moreover, in some cases, the portal vein is not visible in images and this hampers the correct ROI drawing, therefore the single input model results to be the only practicable solution for performing perfusion analysis [106].

Although the adoption of such a dual input model may increase both the clinical accuracy of the results and allow a wider comprehension of more biological mechanisms underlying different clinical diseases, dual input models complicate the mathematical modelling due to the presence of time delays, that need to be taken into account [107]. In fact, on the one hand, the delay between the two inputs makes very challenging to estimate the mixing proportion of the vascular contributions which determines the tissue tracer enhancement. On the other hand, there is a further delay between the input vessels and the delivery of CA at tissue capillaries which complicates the tissue modelling. As we will see, even small delays in the arrival of the CA at the tissue may strongly influence the computation of perfusion parameters.

4.3 Kinetics of the contrast agent

CTp imaging allows studying different tissue hemodynamic features depending on the CA adopted and injection modality, besides the choice of the tracer kinetic model. Generally, to fulfil the assumptions of different mathematical models used to compute perfusion parameters, a short sharp injection of CA is recommended [55], thus ensuring that the CA concentration in the input vessel reaches its peak before the maximum tissue enhancement. In particular, this requirement is necessary when compartmental analysis (Sect. 4.5.2) is performed. Accordingly, CA is generally administered through a bolus of 40-50 ml at 5-10 ml/s. Some computing methods, like Deconvolution, are less affected by slower injection rates, albeit faster injections are useful since they improve the SNR of the TCCs.

Also intrinsic properties of the CA have an effect on the tissue hemodynamics assessed through a perfusion study. One can distinguish mainly two types of CA,

intravascular and extravascular. Early applications of CTP imaging primarily considered extravascular CAs, which pass across the capillary membrane towards the extravascular space. Later on, intravascular iodinated CAs have been developed to obtain more vascular-specific perfusion parameters and a prolonged tissue enhancement. In fact, intravascular CAs have shown to improve the computation of perfusion parameters, since a prolonged and emphasized tissue enhancement facilitate the mathematical modelling of the TCCs. For this reason, they are, nowadays, the most frequently adopted in perfusion studies [96].

4.3.1 First pass or recirculation

The study of perfusion may be limited to the first pass of CA into the vascular compartment of the tissue of interest, or include the CA recirculating phases, when CA is progressively washed out by the vascular circulatory system over subsequent cardiac cycles. Accordingly, CTP acquisitions last from 40 to 60 s if perfusion analysis focus on the first pass, otherwise they cover a longer time interval, up to 2 or 3 min [54]. During first pass, CA has mainly an intravascular distribution, whereas in the recirculating phase CA moves towards the extravascular space. A short and sharp CA bolus injection is important to obtain the characteristic rapid increase in up-slope of the CA enhancement, which reaches a peak and then decreases to the baseline value [108]. Accordingly, a high temporal resolution is needed in first pass acquisitions in order to catch the rapid variations of the signals, while, to reduce radiation exposure, a lower resolution is accepted in the CA delayed phase [54]. From a clinical point of view, it could be interesting to assess variations of perfusion parameters between first pass and recirculation, also considering that perfusion parameters in each of the two phases can have a different prognostic or predictive role. In this regard, a different selection of the end of the first pass can strongly vary the values of the computed perfusion parameters [109]. Figure 4.6 shows a representative vascular TAC of the aortic artery where the two kinetic phases are highlighted. Actually, distinguishing between the two kinetic phases of the CA on the TACs extracted from image sequences can be very challenging. Some simulation studies show that the CA kinetics limited to the first pass of CA is represented by a TAC decaying to the baseline, after its peak [110], and it is well-known that when adopting sharp and short bolus injections, the recirculation contributes to TACs mainly after that its peak occurs, in vessels as well as in tissue curves [111]. An additional complexity arises when considering a dual input physiological model, where the tissue enhancement results from the overlapping of two different kinetic phases of CA within the respective input vessels. Some studies define the end of the first pass as the lowest point reached after the maximal peak and before recirculation peak, and accordingly, some software packages have implemented an automatic

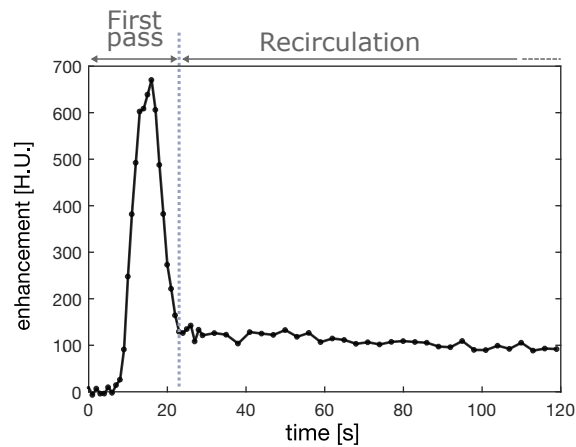


Fig. 4.6: Kinetics of the CA comprising two main phases, the first pass, when CA mainly flows into the intravascular compartment, and the recirculation, when CA moves towards the extravascular compartment.

computation of the first pass time, also to be evaluated as an additional perfusion parameter [109]. However, two major considerations are needed. First, computing the end of the first pass as the first local minimum after the signal's peak leads to non-robust and unreliable measurements, which mostly suffer from noise, variations in sampling frequency, motion artefacts, and ROI positioning. Second, except for arterial input TACs, tissue curves do not show a clear local minimum after the peak, but the descending phase of the signal after the peak is progressively integrated in the subsequent re-enhancement of the signal itself due to recirculation. Therefore, this method cannot be applied and leads to a strongly approximate solution.

From a methodological point of view, the correct separation of first pass and recirculation still remains an open issue and, at the best of my knowledge, none alternative solutions have been proposed so far. Actually, the problem has had a limited interest in research studies since one of the early changes in tissue perfusion due to angiogenesis and tumour growth is observed in alterations of vascular permeability, thus requiring the study of the entire vascular dynamics, including the recirculation phase (i.e. when CA extravasates), and generally performed through multi-compartmental computing methods [108]. However, it is worth notice that this lack limits the applicability of some perfusion methods and does not allow the comparison of results achieved with different perfusion methods, thus delaying the standardization and reproducibility of results of the CTP, and ultimately contributing to postpone its translation into the clinical practice.

4.3.2 Mono and multicompartmental model

To accomplish the clinical purpose, perfusion imaging should refer to the first pass only or include the recirculating kinetic phase of the CA. Accordingly, tissue modelling may be limited to represent the intravascular compartment, thus adopting a monocompartmental model, or can be extended to describe also the extravascular space with a multicompartmental model. A monocompartmental model is suitable when perfusion imaging protocol adopts an intravascular CA (i.e. with a negligible dispersion in the interstice during the first pass) and is focused on the first pass phase of the CA kinetics. By assuming a single compartment, one can measure a reduced number of perfusion parameters, namely BF, BV, MTT, and PI if a dual input organ is considered. Instead, a multicompartmental model is needed to describe the recirculation of CA into the vascular network, which include its extravasation through highly permeable capillaries [108].

Perfusion compartmental models aim at describing the exchanges which occur between tissue and blood within the capillary bed, using simplified boxes made of one or multiple interacting chambers. From a biological point of view, as shown in Figure 4.7 (a), capillaries are constituted of erythrocytes within the blood plasma (PSL), and separated from the extracellular space (EES) by a thin layer of endothelial cells, semi-permeable to small molecules, whereas tissue includes cells, which are separated by frost protein fibers called interstice, generally referred to as the entire EES. Accordingly, whilst BF refers to the flow within PSL, BV is generally split into two components BV_e and BV_p , referred to EES and PLS, respectively. Similarly, the blood flow from PLS to EES is assessed through the permeability-related perfusion parameters. From a mathematical point of view, this biological system is represented by a compartmental model as the one shown in Figure 4.7 (b). The internal cylinder indicates the PLS, whilst the external one refers to EES.

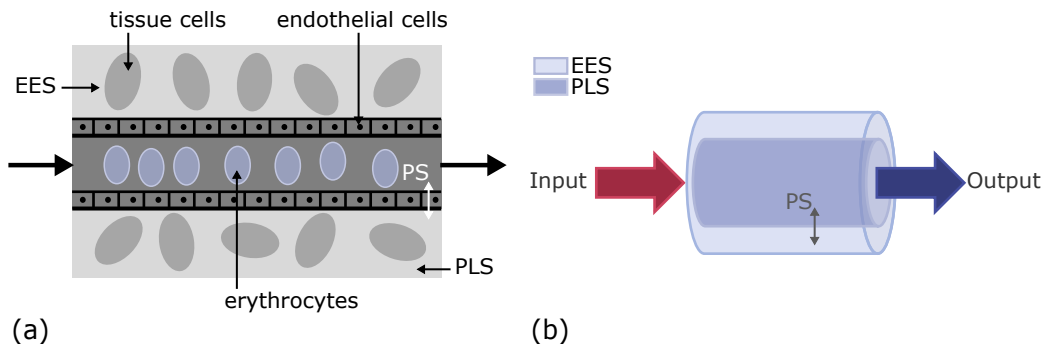


Fig. 4.7: Representation of a tissue capillary through a biological (a) and mathematical (b) model. The CA entering the compartment is driven into the PLS and exits towards the EES by passing across the permeable endothelial cells of the membrane.

Capillary permeability ensures the exchanges between the two regions. Figure 4.8 reports a simplified version of the capillary modelling through compartments. In

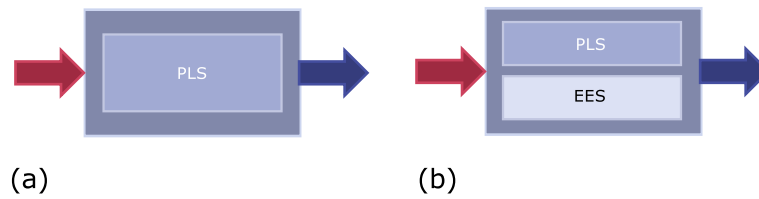


Fig. 4.8: Representation of simplified version of mono (a) and multicompartmental (b) models corresponding to the system reported in Figure 4.7(b).

particular, Figure 4.8 (a) describes a monocompartmental model representing the PLS only, whilst Figure 4.8 (b) a multicompartmental model, including both PLS and EES [112].

Models employed to describe tissue perfusion exchanges are assumed to be well-mixed or plug-flow compartments, as shown in Figure 4.9. In the first case, a

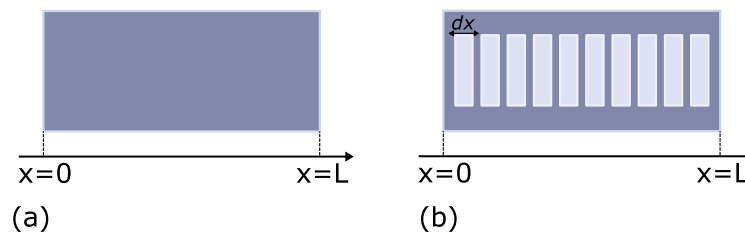


Fig. 4.9: Compartmental models distinguished in well-mixed (a) and plug-flow (b) systems. In the first case, there is an instantaneous mixing of CA within the PLS and its concentration is spatially homogeneous, whilst in the second case, the concentration of CA within the compartment is spatio dependent. Accordingly, to account for each position, the entire length L of the capillary is split into infinitesimal elements dx .

well-mixed compartment (Figure 4.9 (a)), the concentration of CA is considered to be uniform in the unit volume, along the whole length L , for each sampling time. Accordingly, the CA flowing towards the outlet is directly proportional to its concentration in the entire compartment. Well-mixed systems neglect spatial heterogeneities in CA concentration within the capillary bed, which can be caused, for instance, by the multiple contributions arriving at tissue inflow from small arterioles. In the second case, a plug-flow compartment (Figure 4.9 (b)), the assumption of spatial homogeneity in CA concentration is relaxed. Moreover, a plug-flow system requires as additional assumption that the diameter of a single erythrocyte is larger than that one of a single capillary, thus forcing erythrocytes to travel in single file through the capillary, therefore with a uniform velocity. Hence, in a capillary of a given length L , the concentration of CA is spatial dependent. Accordingly, L is

split into small disks with infinitesimal thickness dx and the concentration of CA is usually referred to each of them [113].

4.4 TCC fitting models

In most practical cases, the pipeline to compute the perfusion parameters from a CTP study requires to model also the TACs through a fitting procedure. The main aim of such a fitting stage is to remove residual noise from signals [114], thus benefiting the computing methods in achieving more robust measurements than those ones computed on raw data [115]. For this purpose, numerous mathematical fitting methods are at the disposal and a detailed explanation of all of them is beyond the scope of this manuscript. However, it is worth noting that a major classification of fitting models distinguishes between non-parametric and parametric ones. The former extract information exclusively from data, thus suffering from noise and presence of outliers and hampering the clinical interpretation of the model itself. The latter are defined through a finite number of known parameters with well-established mathematical relationships. Moreover, parametric fitting models are based on physiological models, and have been already used by the literature, since they explain physiological and pathological behaviours, thus definitely contributing to enrich the clinical interpretation of the quantitative perfusion analysis. For these reasons, parametric models are by far the most frequently adopted class in perfusion studies [95],[116]. I also adopted parametric models for fitting the vascular and tissue signals because the sparse and non-uniform sampling of extracted TACs did not make the employment of non-parametric models physiologically acceptable. Therefore, I preferred having reasonably a fitting error higher than that one which I could have obtained using non-parametric models, but preserving the correspondence between the fitting model and the physiological meaning of the signals. Nevertheless, the employment of non-parametric models can be addressed in a future stage of this research.

Rather than on the whole TAC, including CA recirculation, a fitting procedure can be performed exclusively in some portions of the TACs, for instance in order to extract the contribution referred to the first pass only, thus allowing assessing vascular hemodynamics into the two CA phases separately.

Different parametric models have been largely employed in TAC modelling. In the following, the most popular methods will be detailed, one of which primarily employed for vascular TACs, the Lagged Normal (LN) model (Sect. 4.4.1), whilst some others largely used for fitting the tissue TACs, the sigmoid (Sect. 4.4.2) and Gamma Variate (Sect. 4.4.3). In this regard, in the absence of specific values, fitting parameters are unitless.

4.4.1 The lagged normal

The LN model has been parametrized with the specific aim of describing the dispersion of an indicator in arterial or, in general, large vessels [117]. Initially, the distribution of an indicator within the vascular network has been described through a normal density curve, as reported in Equation 4.1:

$$f_1(t) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-t_c}{\sigma}\right)^2}, & \text{for } t \geq 0 \\ 0, & \text{for } t < 0 \end{cases} \quad (4.1)$$

where, $f_1(t)$ is a Gaussian curve of unit area, standard deviation σ , and central time t_c , around which transit times are symmetrically distributed. However, later on, experimental observations have highlighted that indicator curves obtained in mixing chambers tend to be always skewed, so that a first order exponential term has been introduced, as the one formalized in Equation 4.2:

$$f_2(t) = \begin{cases} \frac{1}{\tau} \cdot e^{-\frac{t}{\tau}}, & \text{for } t \geq 0 \\ 0, & \text{for } t < 0 \end{cases} \quad (4.2)$$

where, $f_2(t)$ is the exponential function with decay time, τ . Hence, from the combination of the two contributions arising from Eqs. 4.1 and 4.2, the formulation of the LN model is derived and expressed through a differential equation, as reported in Equation 4.3:

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-t_c}{\sigma}\right)^2} - \tau \frac{df(t)}{dt} \quad (4.3)$$

Equation 4.3 can also be represented by means of three more practical parameters enabling data-driven criteria of selection, the skewness (s), the mean time (t_m), and the Relative Dispersion (RD), which allow expressing the three LN parameters, as described by Eqs. 4.4, 4.5, and 4.6:

$$\tau = RD \cdot t_m \cdot \left(\frac{s}{2}\right)^{\frac{1}{3}} \quad (4.4)$$

$$\sigma = \sqrt{(RD \cdot t_m)^2 - \tau^2} \quad (4.5)$$

$$t_c = t_m - \tau \quad (4.6)$$

In particular, s , as it is defined for a Gaussian process, determines the degree of asymmetry, thus assuming positive values for the LN model, whereas t_m indicates the mean transit time of the LN, which combines the effects of Gaussian and exponential decay times. For instance, Figure 4.10 (a) shows an example of LN function in order to appreciate the right-skewed decay, characteristic of the model, where t_m is

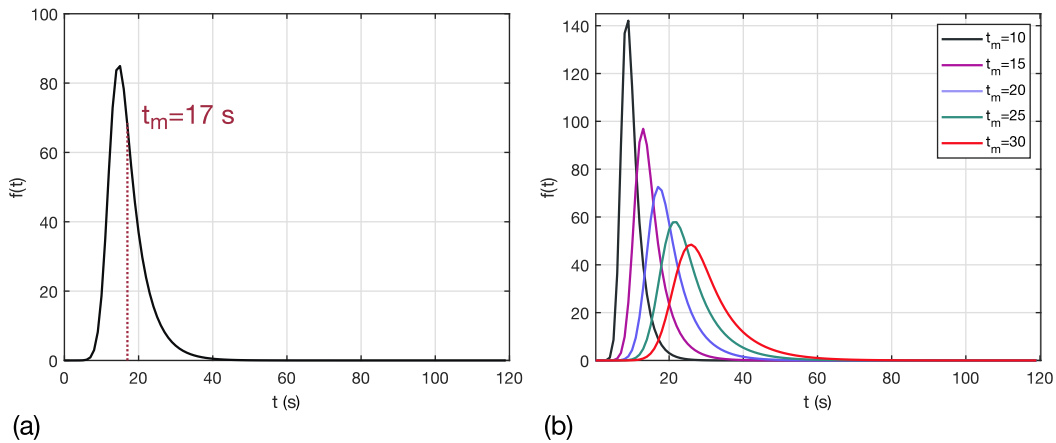


Fig. 4.10: (a) Representative behaviour of a LN fitting model, with such a right-skewed decay, where the parameter t_m is highlighted. In (b), it is reported the variation of the LN model when t_m ranges between 10 s and 30 s with step equals to 5 s

highlighted. Figure 4.10 (b) shows the behaviour of the LN as t_m varies between $t_m = [10 \div 30]$ s, a reasonable interval for vascular TCCs, with step equals to 5 s. Accordingly, the shifting of t_m towards higher values induces larger and lower functions. Finally, RD describes the variance of the LN normalized with respect to t_m and is expressed by Equation 4.7:

$$RD = \frac{\int_0^{\infty} f(\tau)(\tau - t_m)^2 d\tau}{t_m} \quad (4.7)$$

$$RD \simeq \frac{\int_{t_{min}}^{t_{max}} f(\tau)(\tau - t_m)^2 d\tau}{t_m}$$

Figure 4.11 shows how LN shape varies as RD and s range between $RD = [0.3 \div 1.1]$ [a.u.] and $s = [0.5 \div 1, 7]$ [a.u.], respectively. Actually, since Equation 4.3 describes a probability density function, it is scaled, for normalization purposes, by a global factor which represents the area under the observed curve (AUC).

Figure 4.12 shows an example of LN fitting performed on the TAC extracted from the aortic artery in a hepatic CTp examination, where $AUC=6078$, $s=1.09$, $RD=0.3$, and $t_m=22$. In this case, the LN function fits the vascular TAC during the first pass phase of the CA, then decays to the baseline value.

So far, the LN is still the widely adopted model for representing TACs extracted from large vessels [116], and strongly suggested for applications in liver studies [118].

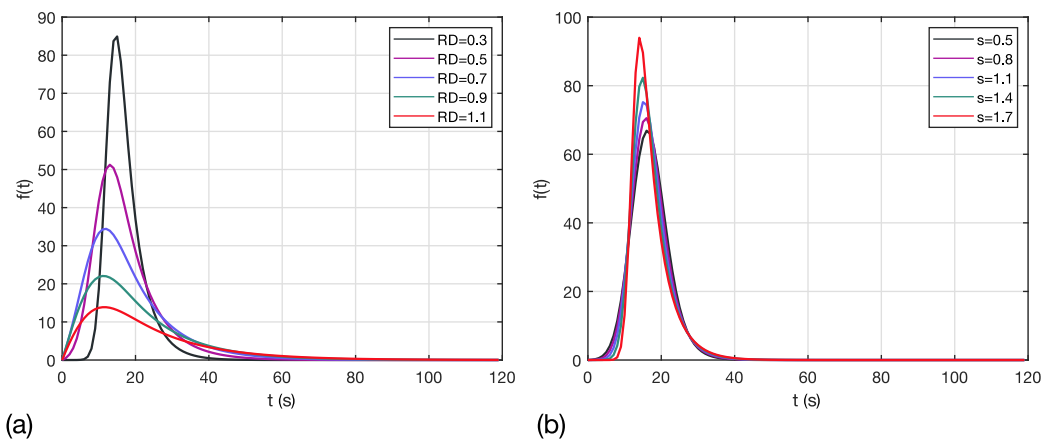


Fig. 4.11: Example of LN model with $Area = 800$, $t_m = 17$ s, and varying RD (a) and s (b), respectively. In particular, when RD varies between $RD = [0.3 \div 1.1]$, $s = 1.5$ (a), and when $s = [0.5 \div 1, 7]$, RD is kept fixed at 0.3. The two parameters show an opposite effect on the shape of the LN. Higher RD values induces larger and lower functions, whilst higher s values leads to more asymmetric functions, spread in a narrower range, therefore reaching a higher peak value.

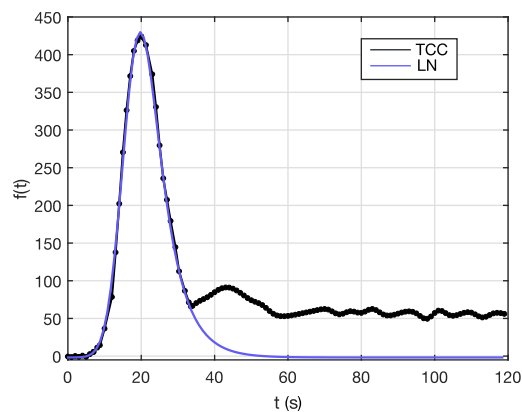


Fig. 4.12: Example of LN fitting of a vascular TCC extracted from the aortic artery in liver CTp examination. In this case, $AUC = 6078$, $s = 1.09$, $RD = 0.3$, and $t_m = 22$

4.4.2 The sigmoid

The employment of the sigmoid model mostly relies on applications focused on the study of only the first pass CA kinetic phase [119]. In fact, this model is particularly suitable to represent the up-slope signal curves due to the arrival of CA in tissue. Varying model parameters allows achieving a very accurate fitting of the ascending phase of the TACs, thanks to its flexibility with respect to experimental data, albeit limited up to the TAC peak value. From Hill's equation, the sigmoid

model is described by Equation 4.8:

$$f(t) = E_0 + (E_{max} - E_0) \frac{t^\alpha}{(EC_{50} + t)^\alpha} \quad (4.8)$$

where, E_{max} is the maximal concentration reached (i.e. the saturation value), therefore corresponding to the TAC's peak value, whilst E_0 represents the baseline concentration (i.e. before the enhancement). EC_{50} represents the time instant when $E(t)$ reaches the half-value of E_{max} , whilst α is the parameter that determines the slope of the curve's ascent, mostly affecting the shape of the model. A graphical explanation of E_0 , E_{max} , and EC_{50} is provided in Figure 4.13 (a), showing a

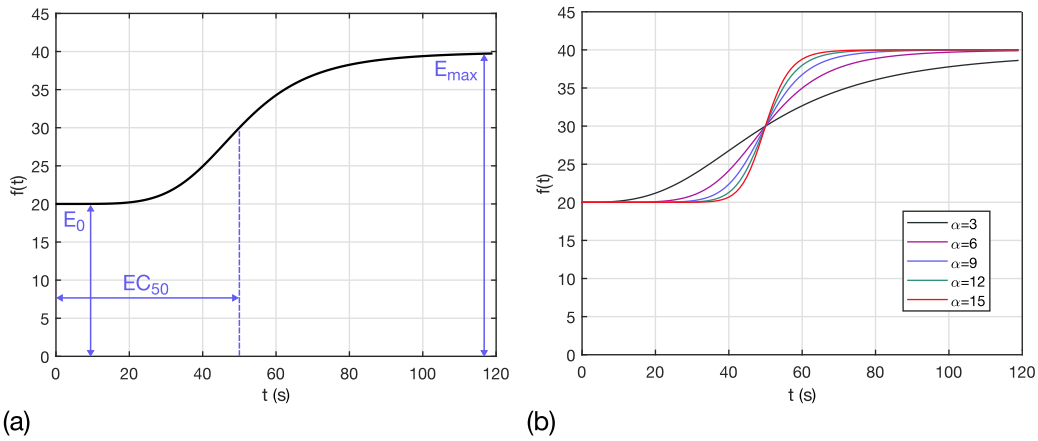


Fig. 4.13: In (a), the representative behaviour of a sigmoid model with the three parameters E_0 , E_{max} , and EC_{50} highlighted in order to clarify their meaning; in (b), variation of the sigmoid as $\alpha = [3 \div 12]$, which modifies the up-slope of the curve.

representative behaviour of a sigmoid model. In addition, Figure 4.13 (b) shows the variation of the sigmoid's up-slope depending on the α parameter, in this case ranging within $\alpha = [3 \div 15]$ with step 3.

Figure 4.14 (a) shows an example of a tissue TCC extracted from a liver CTp examination, fitted by a sigmoid model, with $E_0 = 65$, $E_{max} = 147$, $EC_{50} = 22.9s$, and $\alpha = 8.4$. As one can see, the sigmoid fits the TCC samples up to the peak value, then reaching saturation. In particular, Figure 4.14 (b) highlights the fitting model achieved during the ascending phase of the signal. Accordingly, a large use of this model is diffused in perfusion studies exploiting the MS (Sect. 4.5.1) to compute perfusion parameters, where BF is derived from only the portion of signal comprises between the first enhancement and peak samples [93].

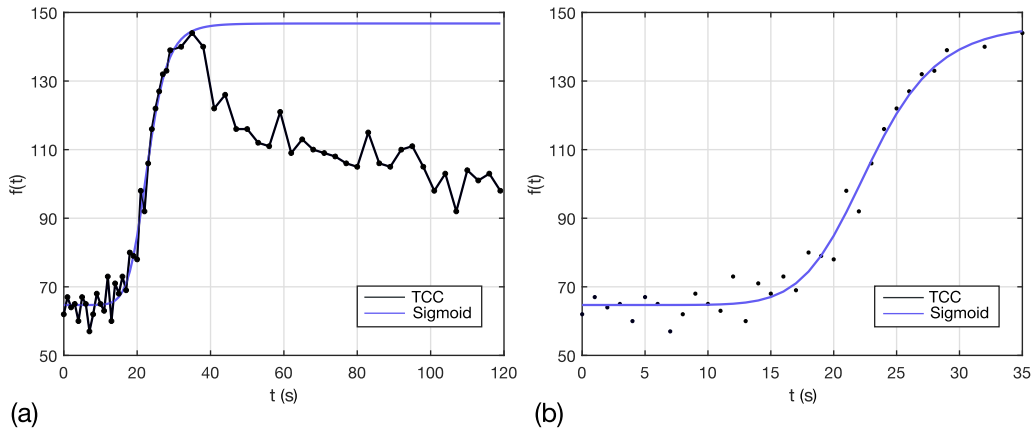


Fig. 4.14: In (a), the fitted signal; in (b), the fitted signal focused on the ascent phase of the signal.

4.4.3 The gamma variate

More than the sigmoidal model, the GV [120] is nowadays the reference model for representing tissue TACs, both in phantom [121],[122] and human studies[123], also exploited for representing the enhancement in large vessels, like the portal vein in the hepatic perfusion [124].

According to its original formalization [125], the GV model is described by Equation 4.9:

$$f(t) = K(t - t_0)^\alpha \cdot e^{-\left(\frac{t-t_0}{\beta}\right)} \quad (4.9)$$

which is valid for $t > t_0$, where t_0 is the location time, the first time instant when $f(t)$ is non-null. Moreover, K is a global scale factor, and α and β are shape and scale factors, respectively. Figure 4.15 shows the effect of varying t_0 in Equation 4.9, which substantially causes a right-time-shift of the curve, thus representing a delayed arrival of the CA within tissue.

Since the early applications, GV function has proved to be very suitable in representing tracer dilution curves. In addition, this model, whose mathematical derivation can be found in detail in [126], allows a physiological interpretation of the parameters, thus being suggested for its application in perfusion studies. In brief, by following what explained for compartmental models in Sect. 4.3.2, a blood vessel can be modelled as a series of mixing chambers with equal volume V , which receives from the vascular input a blood flow at a rate, Q . Accordingly, the parameter α in Equation 4.9 is related to the number of mixing chambers within the vessel's model, whilst β is given by the ratio of Q and V , as expressed in Eqs. 4.10 and 4.11:

$$\alpha = n - 1 \quad (4.10)$$

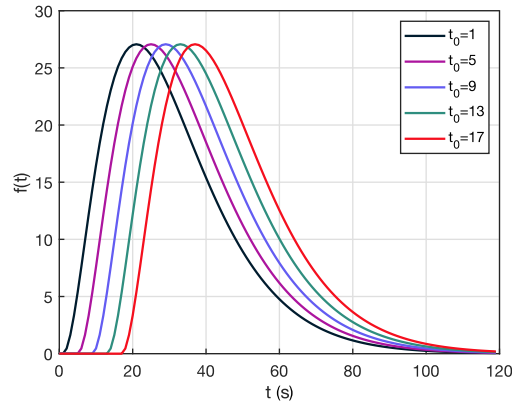


Fig. 4.15: Example of GV model with $K = 0.5$, $\alpha = 2$, and $\beta = 10$, with $t_0 = [1 : 4 : 20]$. Variation of t_0 causes a right-shift of the curve over time.

$$\beta = \frac{V}{Q} \quad (4.11)$$

Hence, α and β have an opposite meaning. In practice, by assuming a constant Q , the higher α (i.e., the higher the number of mixing chambers) the lower V , that means that each chamber is expected to have a reduced V in order to preserve the rate Q . Accordingly, an opposite behaviour is expected in the shape of the GV function by varying, in Equation 4.9, α or β , respectively. In this regard, Figure 4.16 shows some

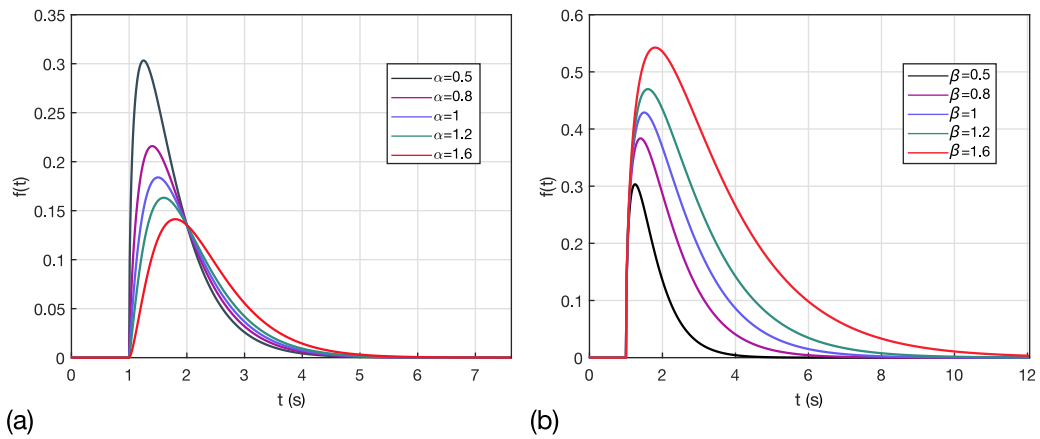


Fig. 4.16: Example of a GV function with varying α (a) and β (b), respectively, and others parameters kept fixed at $t_0 = 1$, and $K = 1$. In particular, while α varies, $\beta = 0.5$ and viceversa. If only α is varying, the GV curves reproduce slower mean transit times as the value of α increases (a), whilst if only β is varying, the curves describes a higher rate Q for higher β values.

GV curves achieved when one between α or β is varying within $[0.5, 0.8, 1, 1.2, 1.6]$ and the other parameters are kept fixed at $t_0 = 1$ and $K = 1$. In particular, in

Figure 4.16 (a), α is varying and $\beta = 0.5$, and the opposite is shown in Figure 4.16 (b). By assuming a constant β , lower α values describe a slower mean transit time of CA within the tissue compartment, whilst, if α is constant, lower values of β can reproduce higher flow rate of CA entering the compartment.

A simplified version of Equation 4.9 is obtained by considering $t_0 = 0$, thus achieving Equation 4.12:

$$f(t) = Kt^\alpha \cdot e^{-\left(\frac{t}{\beta}\right)} \quad (4.12)$$

Hence, Equation 4.12 can be exploited to derive the relationship between α and β and the time t_{max} , that is the time instant when the GV peak value occurs, by setting its first derivative to zero. Therefore, from Equation 4.13:

$$\begin{aligned} f'(t_{max}) &= 0 \\ &= K \left[\alpha t_{max}^{\alpha-1} \cdot e^{-\frac{t_{max}}{\beta}} - \frac{e^{-\frac{t_{max}}{\beta}}}{\beta t_{max}^\alpha} \right] \\ &= K t_{max}^{\alpha-1} e^{-\frac{t_{max}}{\beta}} \left[\alpha - \frac{t_{max}}{\beta} \right] \end{aligned} \quad (4.13)$$

Equation 4.14 can be derived, also expressed in the form of Equation 4.15:

$$t_{max} = \alpha \cdot \beta \quad (4.14)$$

$$\beta = \frac{t_{max}}{\alpha} \quad (4.15)$$

Equation 4.14 is also graphically described by Figure 4.17 where the GV model is represented according to the variation of the ratio between α and β . One can see that t_{max} of each curve moves towards lower or higher values according to Equation 4.14. Finally, K can be expressed in terms of $f(t_{max})$, thus achieving Equation 4.16:

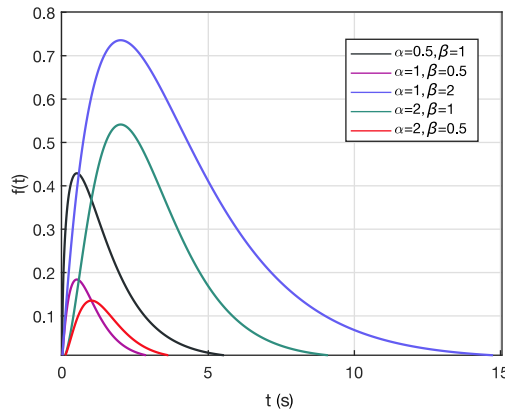


Fig. 4.17: A graphical representation of the relationship between α and β as reported in Equation 4.14. t_{max} of each curve varies depending on the value of α and β .

$$K = f(t_{max}) \cdot t_{max}^{-\alpha} \cdot e^{\alpha} \quad (4.16)$$

Figure 4.18 shows the fitting of the same TCC reported in Figure 4.14 (a) through

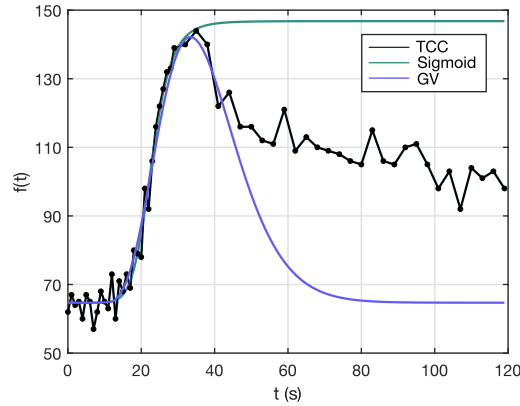


Fig. 4.18: Comparison between the sigmoid fitting reported in Figure 4.14(a) and the GV fitting, which allows fitting the first pass kinetics of the CA rather than the only ascending phase of the TCC as the sigmoidal model does.

a GV function, with $K \sim 10^{-6}$, $\alpha = 6.96$, $\beta = 3.99$, and $t_0 = 6$ s (estimated through a minimization procedure employing the Interior Point, IP, algorithm). It is worth noting, as highlighted in Sect. 4.4.2, that while the fitting with a sigmoid model is limited to the ascending phase of the TCC, the GV model allows fitting the entire first pass kinetics of the CA. Moreover, a mixture of GV functions can be used to fit the whole signal, thus including the recirculation.

4.5 Computing methods for perfusion parameters

Several computing methods can be applied on the TACs, and in most of cases, on their fitted versions, in order to derive perfusion parameters, which allow quantifying tissue perfusion. In this regard, in Chapter 3, Sect. 3.2, the most widely adopted perfusion parameters have been introduced.

A major classification of perfusion computing methods relies on the theoretical principle exploited for carrying out the mathematical modelling. A method, on the ground of the Indicator Dilution Theory (IDT), adopts a deconvolution approach on a monocompartmental model. Accordingly, it allows computing BF, BV, and MTT and neglects the computation of permeability-related perfusion parameters. Instead, numerous other methods are based on the Fick's principle, and by adopting mono or multicompartamental models, provide analytical solutions to a variable number of differential equations which describe each of the compartment involved in the model. The higher number of compartments, the more detailed the descrip-

tion of diffusive exchanges, and the greater number of the perfusion parameters computed, accordingly. Among compartmental models, MS exploits Fick's principle and the mass conservation's law under the assumption of no venous outflow from the compartment, and provides an approximate solution of the basic differential equation representing the monocompartmental model. It allows extracting only the BF, referred to the first pass CA kinetic phase. Although allowing a limited representation of tissue perfusion, MS has been largely employed due to its low complexity and robustness, and as such it is detailed in Sect. 4.5.1. In addition, the most diffused compartmental methods are presented in Sect. 4.5.2, limiting the description to those ones employing bicompartmental models. At the end, the DV method is explained and discussed in Sect. 4.5.3.

4.5.1 Maximum Slope

MS is based on the assumption of a monocompartmental model as the one represented in Figure 4.19, describing a unit of tissue volume where a blood flow carrying

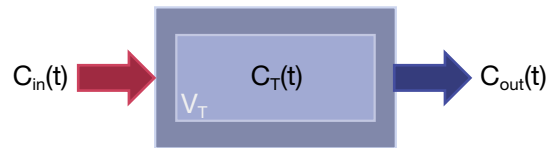


Fig. 4.19: The diffusion of a tracer is described through the Fick's principle using a single input monocompartmental model. The BF within the unit of tissue volume V_T , is measured by applying the mass conservation's law. $C_{in}(t)$, $C_T(t)$, and $C_{out}(t)$ represent the CA concentration at the arterial input, within the compartment, and at the venous outflow, respectively.

on a certain quantity of CA comes from an input artery and is driven out through a venous vessel. In particular, in order to fulfil the monocompartment hypothesis, MS assumes the conservation of the mass under the assumption of no venous outflow [127]. Accordingly, on the ground of the mass conservation's law, the Fick principle describes the diffusion of a tracer through a single input compartmental model, as expressed by Equation 4.17:

$$\frac{dC_T(t)}{dt} = \frac{\text{BF}}{V_T} [C_{in}(t) - C_{out}(t)] \quad (4.17)$$

where, $C_T(t)$ is the CA concentration within the tissue volume, V_T , where a certain amount of BF is exchanged. $C_{in}(t)$ represents the CA input concentration usually referred to as the arterial blood plasma concentration $C_A(t)$, whilst $C_{out}(t)$ refers to the venous outflow. Actually, a minimum transit time exists before the injected CA reaches the venous circulation, when it is assumed to be still inside tissue.

Therefore, under the assumption of no venous outflow, $C_{out} \approx 0$, and Equation 4.17 can be simplified as follows (Equation 4.18):

$$\frac{dC_T(t)}{dt} \approx \frac{BF}{V_T} C_{in}(t) \quad (4.18)$$

This also implies that $C_T(t)$ reaches its maximum slope in the correspondence of the maximum value of $C_{in}(t)$. Assuming V_T as a normalization factor represented by a constant volume unit, BF is given by Equation 4.19 [128], that is the MS of the tissue TAC normalized by the maximum value of the input TAC.

$$BF \approx \frac{\left. \frac{dC_T(t)}{dt} \right|_{max}}{\left. C_{in}(t) \right|_{max}} \quad (4.19)$$

In addition, Figure 4.20 provides a graphical representation of the computation of

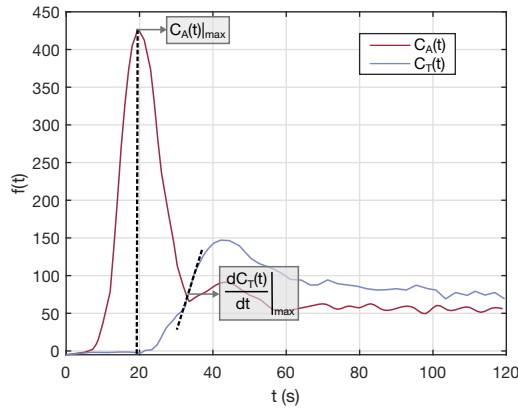


Fig. 4.20: Graphical representations of the terms of Equation 4.19 which define the formulation of the MS method for computing the BF by assuming a single input model.

BF with MS as in Equation 4.19.

The original MS formulation, thought for single input monocompartmental models, needs to be extended to be applied also in dual input models (Sect. 4.2). For instance, in case of a liver CTP study, the quantitative assessment of tissue perfusion has to account for the two contributions arising from the arterial and portal vascular circulation. In practice, the computation of BF in Equation 4.19, is split into two additional terms, according to Equation 4.20:

$$BF = aBF + pBF \quad (4.20)$$

where, aBF refers to the BF arriving from the arterial vessel, while pBF is the BF received from the portal vein. Therefore, by substituting Equation 4.19 into each term of Equation 4.20, the dual input expression of the MS method is achieved, as

reported in Equation 4.21:

$$\text{BF} \approx \frac{dC_{T_A}(t)|_{max}}{C_A(t)|_{max}} + \frac{dC_{T_P}(t)|_{max}}{C_P(t)|_{max}} \quad (4.21)$$

where, $C_{T_A}(t)$ and $C_{T_P}(t)$ are the components of tissue TAC, each referred to a different input, that is the arterial ($C_A(t)$) and portal ($C_P(t)$) vessels, respectively.

However, under the MS approach $C_{T_A}(t)$ and $C_{T_P}(t)$ cannot be analytically separated. Commonly, they are approximated by exploiting the TAC extracted from the spleen. In particular, the peak's time of the splenic TAC is assumed as the end of the arterial phase and the begin of the portal phase on the tissue TAC. Accordingly, the two components $C_{T_A}(t)$ and $C_{T_P}(t)$ are given by Eqs. 4.22 and 4.23 [129]:

$$C_{T_A}(t) \approx C_T(t)|_{t \in [0, C_S(t_{max})]} \quad (4.22)$$

$$C_{T_P}(t) \approx C_T(t)|_{t \geq C_S(t_{max})} \quad (4.23)$$

where $C_S(t)$ is the mean TAC extracted from the spleen and t_{max} is the time instant when its peak occurs. In this regard, Figure 4.21 extends the graphical representa-

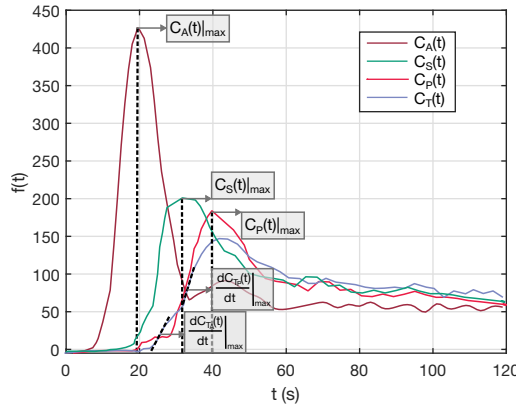


Fig. 4.21: Graphical representations of the terms of Equation 4.21 which define the extension of the MS method for computing the BF in a dual input model.

tion of Figure 4.20 with the representation of the terms involved in the computation of the dual input MS (Equation 4.21) for the representative case of hepatic perfusion study. Actually, in case of dual input MS, besides BF, a further perfusion parameter can be computed, that is the PI. In case of liver, PI is commonly referred to as Hepatic PI (HPI), and is expressed by Equation 4.24.

$$\text{HPI} = \frac{\text{aBF}}{\text{aBF} + \text{pBF}} \quad (4.24)$$

The main advantage of employing MS for the computation of perfusion parameters is

that it strongly allows reducing acquisition times, and, in case of CTP, the radiation dose administered to patient, accordingly. Moreover, MS is a simple method, with a low computational complexity, therefore implemented in many software packages commercialized for perfusion analyses [130]. Yet more, according to its mathematical formulation, MS yields robust measurements, albeit sometimes poorly accurate. In this regard, it is worth noting that, as shown in Figure 4.20, the peak's time of the input function may not correspond to the time of the MS on the tissue TAC, and this happens in most practical cases. Such a delay between the two TACs is known to cause underestimated BF measurements. In addition, the MS method is based on the assumption of no venous outflow, which is valid for very few seconds, therefore often in the MS's time instant a venous outflow already exists, thus leading to inaccuracies in the BF estimates [131].

The biggest open issue relies on the application of MS when studying perfusion of dual input organs. MS is commonly used in most recent works assessing, for instance, liver perfusion with the approximate solution above presented (Equation 4.21) and none alternative approach have been proposed so far [132]. Such an approximative solution makes substantially not comparable BF values computed with MS to the measurements achieved with other methods, thus ultimately jeopardizing the reliability of the measures themselves. It is also worth noting that MS has been conceived to compute BF only. Accordingly, in many applications, MS is used in association with the Patlak model – one of the compartmental methods which is explained in the next Sect. 4.5.2 – to estimate more perfusion parameters.

4.5.2 Compartmental model

Dual input single compartmental models, like the one presented in Figure 4.22, can be described, alternatively to the MS method, by the Materne one [133]. Ac-

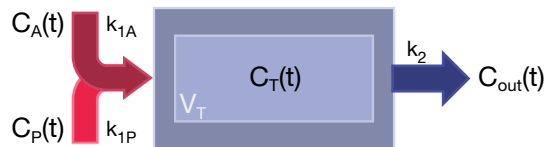


Fig. 4.22: Representation of a dual input single compartmental model, referring for instance to hepatic perfusion, described by the Materne method exploiting the Fick's principle.

cordingly, the differential equation in Equation 4.17 valid for the single input model in Figure 4.19, can be rewritten as in Equation 4.25:

$$\frac{dC_T(t)}{dt} = k_{1A}C_A(t) + k_{1P}C_P(t) - k_2C_T(t) \quad (4.25)$$

where k_{1A} and k_{1P} are the arterial and portal venous inflow rate constant, while k_2 is the outflow rate constant. By solving for $C_T(t)$, also including two time delay parameters, τ_A and τ_P , accounting for the transit time from the artery (generally the aorta) and portal vein to the tissue, one achieves Equation 4.26:

$$C_T(t) = \int_0^t k_{1A}C_A(t' - \tau_A) + k_{1P}C_P(t' - \tau_P)e^{-k_2(t-t')} dt' \quad (4.26)$$

where $C_T(0)=0$ is assumed. Hence, an unweighted least squares fit of the experimental $C_T(t)$ through the model reported in Equation 4.26 allows achieving the parameters k_{1A} , k_{1P} , and k_2 . Accordingly, total tissue perfusion is calculated as follows (Equation 4.27):

$$BF = \frac{k_{1A} + k_{1P}}{EF} \quad (4.27)$$

where EF is assumed to be unit in the normal liver, that means that CA is free to access the EES, thus leading to Eqs. 4.28,4.29:

$$aBF = k_{1A} \quad (4.28)$$

$$pBF = k_{1P} \quad (4.29)$$

Finally, MTT is derived from k_2 , according to Equation 4.30:

$$MTT = k_2^{-1} \quad (4.30)$$

The methods presented so far limits the quantification of tissue perfusion at the PLS. Extending quantitative perfusion analysis to the EES can be performed through exchange or uptake models. The major difference between them relies on whether the assumption that CA cannot accumulate into the EES is fulfilled or not. In case of exchange models, not allowing the CA accumulation, a backward flux of CA exists from EES to the PLS. In this regard, the simplest method to describe the tracer's flow between PLS and EES exploits the two compartments exchange model (2CX) [113], reported in Figure 4.23. In order to fulfil the mass conservation's law

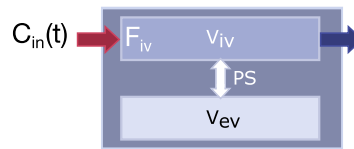


Fig. 4.23: Representation of the 2CX model: the CA enters in the PLS, with volume v_{iv} and then reaches the EES, with volume v_{ev} . Exchanges between PLS are bidirectional and measured through the parameter PS .

imposing no CA accumulation within the system, the exchanges between the PLS

having volume v_{iv} , and EES with volume v_{ev} are assumed as being bidirectional and quantified by the parameter PS. The CA enters the PLS, where its flow is F_{iv} and moves towards the EES. The total tissue concentration of CA is a weighted sum of that one in PLS ($C_{iv}(t)$) and ESS ($C_{ev}(t)$), respectively, as expressed by Equation 4.31:

$$C_T(t) = v_{iv}C_{iv}(t) + v_{ev}C_{ev}(t) \quad (4.31)$$

Accordingly, the EES's inflow is $PS C_{iv}(t)$ and it is equal to the outflow out of ESS coming back to the PLS. Then, the model in Figure 4.23 is represented by Eqs. 4.32,4.33:

$$V_{iv} \frac{dC_{iv}(t)}{dt} = F_{iv}(C_{in}(t) - C_{iv}(t)) + PS(C_{ev}(t) - C_{iv}(t)) \quad (4.32)$$

$$V_{ev} \frac{dC_{ev}(t)}{dt} = PS(C_{iv}(t) - C_{ev}(t)) \quad (4.33)$$

The solution of these equations provide F_{iv} , v_{iv} , v_{ev} , and PS.

While the 2CX model describes two well-mixed compartments, an alternative modelling is represented by the tissue homogeneity (TH) model, where the PLS is represented as a plug-flow system (Figure 4.24). Since the EES is represented

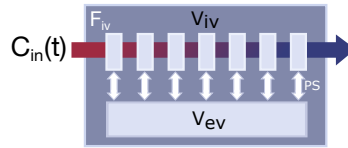


Fig. 4.24: Representation of the THM where the well-mixed EES of the 2CX model is substituted by a plug-flow system

by a well-mixed system as well as the 2CX model, Equation 4.33 is still valid for representing the ESS in the TH model. On the contrary, Equation 4.32 needs to be converted into Equation 4.34, in order to fulfil the assumptions of a plug-flow system:

$$\frac{V_{iv}}{L} \frac{\partial C_{iv}(x, t)}{\partial t} = -F_{iv} \frac{\partial C_{iv}(x, t)}{\partial x} + \frac{PS}{L} (C_{ev}(x, t) - C_{iv}(x, t)) \quad (4.34)$$

where the boundary condition is $C_{iv}(0, t) = C_{in}(t)$. Actually, Equation 4.34 does not have a solution in the time domain, therefore the TH model has had a limited diffusion. The adiabatic approximation (AATH) has been proposed in order to circumvent this lack. The AATH model relies on considering the capillary wall impermeable to CA along its length, whilst exchanges are admitted only at the venous end of the capillary, where $x = L$ (Figure 4.25). Therefore, Equation 4.34

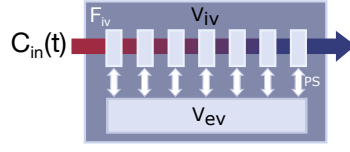


Fig. 4.25: Representation of the AATH model, where the CA exchanges are assumed to happen at the venous outlet only, where $x = L$, thus allowing solving the Equation 4.34 in the time domain, considering $PS = 0$.

can be simplified into Equation 4.35, where PS is assumed to be zero.

$$\frac{V_{iv}}{L} \frac{\partial C_{iv}(x, t)}{\partial t} = -F_{iv} \frac{\partial C_{iv}(x, t)}{\partial t} \quad (4.35)$$

Accordingly, the EES is described by assuming v_{iv} tending to zero and $C_{in}(t) = C_{iv}(L, t)$ at the venous outlet, as given by Equation 4.36 [134]:

$$V_{ev} \frac{dC_{ev}(t)}{dt} = PS (C_{in}(t) - C_{ev}(t)) \quad (4.36)$$

By assuming that also the PLS is represented through a plug-flow system, the distributed parameter (DP) model is derived, as shown in Figure 4.26: The EES is

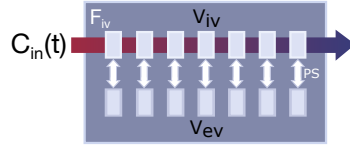


Fig. 4.26: Representation of the DP model: both the PLS and EES compartments are described through plug-flow systems.

modelled through a series of infinitesimal compartments where, in contrast to the TH model, in the EES, exchanges are allowed only between adjacent capillary walls. Consequently, both $C_{iv}(t)$ and $C_{ev}(t)$ depend on the position, x , as expressed by Equation 4.34 and Equation 4.37:

$$V_{ev} \frac{\partial C_{ev}(x, t)}{\partial t} = PS (C_{iv}(x, t) - C_{ev}(x, t)) \quad (4.37)$$

Actually, the assumption of no parallel transport in the EES is not justified by any biological or physical evidence. However, it allows achieving an analytical solution for the equations of the DP model [135].

From all the previous models, one can derive the Tofts model by assuming, as a further assumption, an infinite CA infusion, that means infinite F_{iv} . Accordingly,

the $C_{ev}(t)$ can be expressed as in Equation 4.38:

$$C_{iv}(t) = K^{trans} e^{-t\left(\frac{K^{trans}}{v_{ev}}\right)} \int_0^t C_{iv}(\tau) d\tau \quad (4.38)$$

Hence, Equation 4.31 can be rewritten according to the extended Tofts model, which include the PLS space too, as in Equation 4.39 [24]:

$$C_T(t) = v_{iv} C_{in}(t) + K^{trans} e^{-t\frac{K^{trans}}{v_{ev}}} \int_0^t C_{iv}(\tau) d\tau \quad (4.39)$$

Currently, the Tofts model is the most widely used method in perfusion studies, since it has proved to yield reproducible results, albeit with a limited physiological interpretation. Moreover, comparison between different centres is constrained by the adoption of exactly the same perfusion protocol. In fact, K^{trans} substantially combines the effects of physiological conditions referred to capillary permeability and acquisition conditions. In fact, the model requires a moderate temporal resolution, because the use of two only parameters might not allow achieving a good fit of the up-slope phase of the TACs. In this regard, the extended model, with one additional parameters has proved to increase both the goodness of fit and the physiological interpretation of results, preserving the robustness of measurements [24].

Under the conditions of small PS, large v_{ev} , and short acquisitions times, the outflow from the EES is negligible and CA tends to accumulate. Accordingly, the uptake model describes the CA concentration into the EES as in Equation 4.40:

$$v_{ev} \frac{dC_{ev}(t)}{dt} = \text{PSC}_{iv}(t) \quad (4.40)$$

Actually, in case of uptake models, the EES is not accessible, therefore the number of perfusion parameters computed is limited to F_{iv} , PS, and v_{iv} [113]. A further simplification of the uptake model is achieved with by Patlak method, which assumes an infinite F_{iv} too, thus reducing the computable parameters at PS and v_{iv} [136]. Moreover, this also leads PS to be equal to K^{trans} . Hence, Equation 4.39 can be rewritten as in Equation 4.41:

$$C_T(t) = v_{iv} C_{in}(t) + K^{trans} \int_0^t C_{iv}(\tau) d\tau \quad (4.41)$$

The two parameters, K^{trans} and v_{iv} can be estimated by linear regression. In fact, by assuming $y(t) = \frac{C_T(t)}{C_{in}(t)}$ and $x(t) = \frac{\int_0^t C_{in}(\tau) d\tau}{C_{in}(t)}$, Equation 4.41 can be rewritten as in Equation 4.42:

$$y(t) = v_{iv} + K^{trans} x(t) \quad (4.42)$$

Generally, a graphical solutions of this equation is provided through the Patlak-

Rutland plot [137].

Compartmental methods for perfusion analysis have had a wide diffusion so far, thus being implemented in many software packages [138]. However, a complete reproducibility of results in multicentre studies has not been reached yet. From a mathematical point of view, they can yield robust measurements, while, from a clinical point of view, they might lead to difficulties in interpreting results. Differential equations are solved in most of cases through fitting procedures, by which perfusion parameters are derived by minimization. First, this does not allow separating the different kinetic phases of the CA and, above all, compartmental methods may force the representation of tissue perfusion with a specific model, thus hiding or missing some particular phenomena.

4.5.3 Deconvolution method

The DV method is grounded on the IDT, under the assumption of system linearity and time-invariance [139]. As well as the MS method, DV adopts a monocompartmental model like the one reported in Figure 4.19, without no additive hypothesis on the internal chambers. Hence, the output function $C_{out}(t)$ is conceived as the convolution of the input function $C_{in}(t)$ with the system's impulse response function, $h(t)$, representing the probability density function of the transit times of CA molecules which flow from the inlet towards the outlet (Equation 4.43).

$$\begin{aligned} C_{out}(t) &= C_{in}(t) \otimes h(t) \\ C_{out}(t) &= \int_0^t C_{in}(\tau)h(t - \tau)d\tau \end{aligned} \quad (4.43)$$

In particular, $h(t)$ is a characteristics of both the tracer and system and it is measured as $1/t$. However, perfusion imaging does not allow extracting the TACs from the venous outflow because it is not possible to identify a principal outflow vessel, and venous blood is a “mixed” blood arising from different regional areas. Therefore, the analysis of tracer diffusion is performed by considering the tracer which remains into the tissue rather than exiting the compartment, over time. To this purpose, we shall hypothesize to collect in a single container all the outflow from the system. In this case, the total amount of CA at the outlet, is given by the cumulative frequency function expressed by Equation 4.44

$$H(t) = \int_0^t h(\tau)d\tau \quad (4.44)$$

where, $H(t)$ tends to one as t tends to infinite. Accordingly, $h(t)$ and $H(t)$ have the same time delay, and when $h(t)$ is at its maximum, $H(t)$ has a flex point, then $H(t)$ approaches unity as $h(t)$ reaches zero. Hence, by looking inside the compartment,

the function $1 - H(t)$ represents the fraction of the input within the system, at each time instant t , that is the system's impulse residue function, $R(t)$ (Equation 4.45).

$$\begin{aligned} R(t) &= 1 - H(t) \\ R(t) &= 1 - \int_0^t h(\tau) d\tau \end{aligned} \quad (4.45)$$

For $t = 0$, when CA has been injected within the system, $R(0)$ is equal to 1, while it tends to zero when all tracer molecules have left the tissue. Alternatively, Equation 4.45 can be written as in Equation 4.46:

$$h(t) = -\frac{dR(t)}{dt} \quad (4.46)$$

Figures 4.27 (a,b) show $h(t)$ and $H(t)$, respectively, whilst Figure 4.27 (c) reports the corresponding $R(t)$. In particular, the time delay of $H(t)$ for leaving the zero line equals the plateau's length of $R(t)$ at unit value, before decaying. For an ideal instantaneous injection, $R(t)$ is a first order exponential decay function [140]. However, since the input function, differently from ideality, is not instantaneous, $R(t)$ may undergo dispersion, so that in practical cases, $R(t)$ assumes the bell-shape reported in Figure 4.27 (d), with $R(t_{max}) < 1$ and occurring for $t > 0$.

Hence, rewriting Equation 4.17 into its integral form, and considering a unit volume V_T , Equation 4.47 is obtained accordingly, which represents the CA concentration within the compartment through the IDT.

$$C_T(t) = BF \int_0^t (C_{in}(\tau) - C_{out}(\tau)) d\tau \quad (4.47)$$

By substituting Equation 4.43 into Equation 4.47, we obtain Equation 4.48:

$$C_T(t) = BF \int_0^t C_{in}(\tau) d\tau - \int_0^t C_{in}(\tau) h(t - \tau) d\tau \quad (4.48)$$

where, in the first integral, $C_{in}(t)$ can be equivalently expressed by the convolution with the delta function, $\delta(t)$, thus achieving Equation 4.49:

$$C_T(t) = BF \int_0^t \left(\int_0^t C_{in}(\xi) \delta(\tau - \xi) d\xi - \int_0^t C_{in}(\xi) h(\tau - \xi) d\xi \right) d\tau \quad (4.49)$$

that can be rearranged into Equation 4.50:

$$C_T(t) = BF \int_0^t C_{in}(\xi) \left(\int_0^t \delta(\tau - \xi) - h(\tau - \xi) d\tau \right) d\xi \quad (4.50)$$

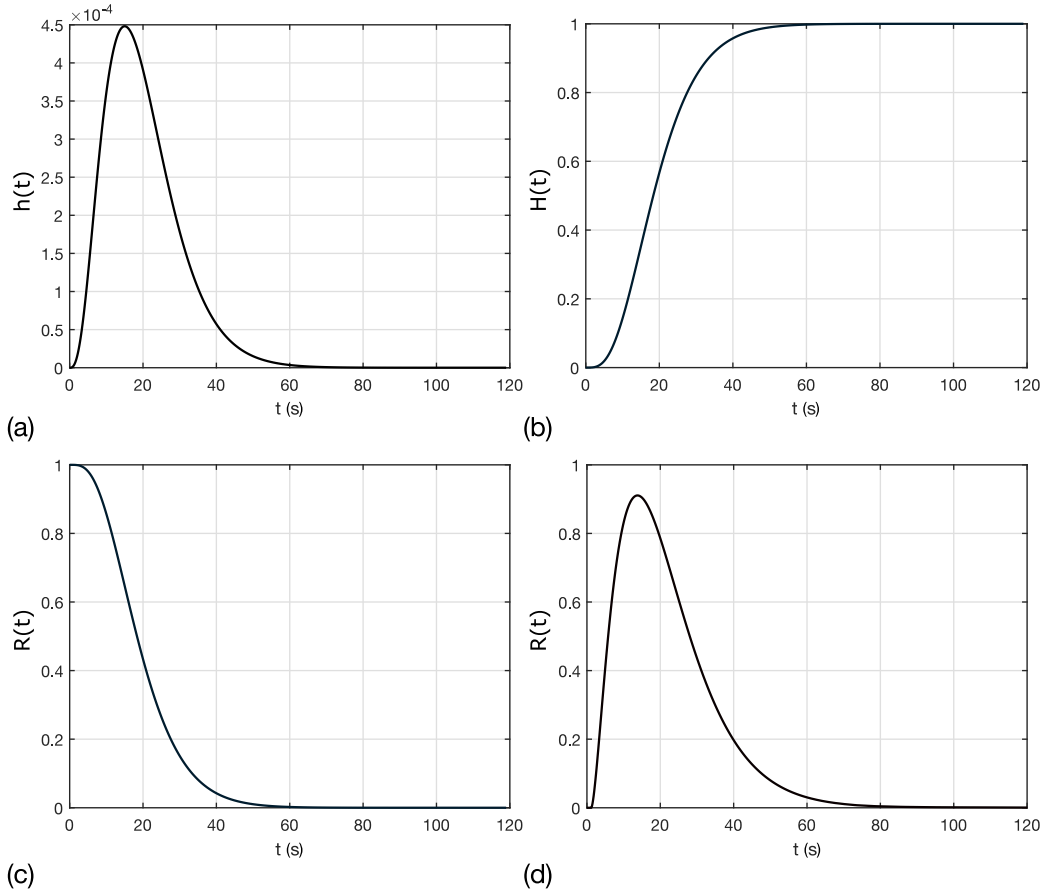


Fig. 4.27: The functions $h(t)$, $H(t)$, and $R(t)$ are reported in (a), (b), and (c), respectively, according to their relationship presented in Eqs. 4.44, 4.45. $h(t)$ (a) is a probability density function representing transit times of CA molecules within the system, and its integration over time yields $H(t)$ (b), which tends to unit, when all molecules have exited the compartment. Accordingly, $R(t)$, describing the tracer molecules which remains within the compartment, decays from unit value to zero. In real cases, when tracer injections are not perfectly instantaneous, $R(t)$ assumes the bell-shape reported in (d), it is spread over time, with $R(t_{max}) < 1$ occurring for $t > 0$.

By considering that for $t > 0$ we have (Equation 4.51):

$$R(t) = 1 - \int_0^t h(\tau) d\tau = \int_0^t (\delta(\tau) - h(\tau)) d\tau \quad (4.51)$$

and by substituting $\tau' = \tau - \xi$, we achieve (Equation 4.52):

$$\int_0^t (\delta(\tau - \xi) - h(\tau - \xi)) d\tau = \int_{-\xi}^{t-\xi} (\delta(\tau') - h(\tau')) d\tau' = R(t - \xi) \quad (4.52)$$

Hence, Equation 4.50 can be rewritten into Equation 4.53:

$$C_T(t) = BF \int_0^t C_{in}(\xi)R(t - \xi)d\xi \quad (4.53)$$

Then, a tissue TAC is described by the convolution of $C_{in}(t)$ with $R(t)$ (Equation 4.54) [114]:

$$C_T(t) = C_{in} \otimes BF \cdot R(t) \quad (4.54)$$

Since perfusion imaging allows measuring $C_T(t)$ and $C_{in}(t)$, a DV procedure allows recovering $R^*(t)$, which is a BF-scaled version of $R(t)$. Accordingly, BF can be estimated as the initial value of $R^*(t)$ (Equation 4.55):

$$BF = R^*(t)|_{t=t_0} \quad (4.55)$$

or, in practical cases, BF is derived from the maximum value of $R^*(t)$ (Equation 4.56):

$$BF = R^*(t)|_{t=t_{max}} \quad (4.56)$$

Therefore, once $R(t)$ is known, it is possible to compute $h(t)$ too from Equation 4.46, in order to measure the MTT, according to Equation 4.57:

$$MTT = \int_0^\infty th(t)dt \quad (4.57)$$

Then, since BF and MTT are computed from $R^*(t)$ and $h(t)$, respectively, BV is derived by applying the Central Volume Theorem (CVT), as expressed by Equation 4.58:

$$BV = MTT \cdot BF \quad (4.58)$$

Alternatively, BV is computed according to Equation 4.59, and MTT from the CVT as in Equation 4.60:

$$BV = \int_0^t R^*(t)dt \quad (4.59)$$

$$MTT = \frac{BV}{BF} \quad (4.60)$$

Actually, DV means solving an ill-posed inverse problem, for which different approaches have been proposed in the literature, in order to achieve accurate estimates of $R^*(t)$. Model dependent techniques, with a limited utilization, perform a parametric deconvolution by assuming a predefined model for the $R^*(t)$, thus implicitly imposing a priori knowledge of tissue microvasculature. If on the one hand this reduces the freedom's degrees and consequently the computational complexity, on the other hand small uncertainties in modelling assumptions may lead to totally incorrect solutions. For this reason, in most practical cases, non parametric DV

techniques are adopted, which mainly exploit an algebraic approach. This implies a reformulation of the convolution problem (Equation 4.54) in its matrix form, by assuming that for discrete small time instants, $R^*(t)$ and the TACs are constant.

To this purpose, Equation 4.54 is represented according to Equation 4.61:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (4.61)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$ represent $C_{in}(t)$ and $C_T(t)$, respectively, with n the number of TCC samples. In particular, \mathbf{A} expresses $C_{in}(t)$ in Equation 4.62:

$$\mathbf{A} = \begin{bmatrix} C_{in}(t_1) & 0 & 0 & \dots & 0 \\ C_{in}(t_2) & C_{in}(t_1) & 0 & \dots & 0 \\ C_{in}(t_3) & C_{in}(t_2) & C_{in}(t_1) & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ C_{in}(t_N) & C_{in}(t_{N-1}) & \dots & \dots & C_{in}(t_1) \end{bmatrix} \quad (4.62)$$

whereas \mathbf{b} is the voxel-based $C_T(t)$ (Equation 4.63):

$$\mathbf{b} = \begin{bmatrix} C_T(t_1) \\ C_T(t_2) \\ \dots \\ \dots \\ C_T(t_N) \end{bmatrix} \quad (4.63)$$

Performing DV means estimating \mathbf{x} from Equation 4.61, as expressed by Equation 4.64:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \quad (4.64)$$

whose best solution is given by Equation 4.65, through the least-square minimization:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \quad (4.65)$$

However, \mathbf{A} is known to be ill-conditioned and several regularization methods are proposed in the literature for improving DV accurateness, including Tikhonov method [141], truncated Singular Value Decomposition (tSVD), circular tSVD (cTSVD) [142]. In particular, cTSVD is a well-established technique for CTP, allowing for time delays between the vascular input and the tissue curves [143]. In practice, in order to prevent aliasing in circular deconvolution, $C_{in}(t)$ and $C_T(t)$ are first zero-padded for $L = 2n$ samples. Then, the circular square matrix $\mathbf{A}^c \in \mathbb{R}^{L \times L}$ is implemented

according to Equation 4.66 [144]:

$$\mathbf{A}^c_{i,j} = \begin{cases} C_{in}(t_{i-j+1}), & \text{for } j \leq i \\ C_{in}(t_{L+i-j+1}), & \text{for } j > i \end{cases} \quad (4.66)$$

The cTSVD solution is achieved by SVD decomposition of \mathbf{A}^c , so that its inverse matrix $\mathbf{A}^{c-1} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T$, where $\mathbf{\Sigma}$ is the diagonal matrix of the singular values sorted in descending order, and \mathbf{V} and \mathbf{U} contain the left- and the right-singular vectors, respectively. In order to reduce the oscillation of the solution, the less representative singular values in $\mathbf{\Sigma}$ are removed [145].

By solving the algebraic decomposition, DV allows estimating the three perfusion parameters which mainly represent the intravascular compartment, that is BF, BV, and MTT. From a theoretical point of view, the DV approach can be exploited to compute the three perfusion parameters both including and excluding the recirculating phase of the CA. However, in practice, the last case depends on the availability of proper methods to extract the TCC referred to the first pass only from the longer signals, as explained in Sect. 4.3.1.

Moreover, like for the MS method, an additional complexity arises when performing DV in dual input models. By recalling what explained in Chapter 3, Sect. 3.2, dual input perfusion introduces an additional perfusion parameters that in case of liver perfusion, is the HPI, whose formulation has been previously reported for the MS method. In practice, a dual input model means that Equation 4.54 has to be expressed as in Equation 4.67:

$$C_T(t) = [\text{HPI} \cdot C_A(t) + (1 - \text{HPI}) \cdot C_P(t)] \otimes R^*(t) \quad (4.67)$$

where, $C_{in}(t)$ is split into the linear combination of the two inputs, weighted by the HPI. Similarly, Equation 4.67 can be rewritten and split into the two Eqs.4.68.4.69:

$$C_{T_A}(t) = [\text{HPI} \cdot C_A(t)] \otimes R^*(t) \quad (4.68)$$

$$C_{T_P}(t) = [1-\text{HPI} \cdot C_P(t)] \otimes R^*(t) \quad (4.69)$$

where, $C_{T_A}(t)$ and $C_{T_P}(t)$ are the TACs referred to $C_A(t)$ and $C_P(t)$, respectively. It is worth noting that in Equation 4.67, both $R^*(t)$ and HPI are unknown, this strongly increasing the computational complexity of the DV procedure. Some techniques of blind deconvolution through minimization procedures have been proposed to solve Equation 4.67, albeit with a limited diffusion. In this regard, a novel methodological approach is proposed in the next Chapter 6.

4.6 Reproducibility of perfusion parameters: an open issue

CTp data analysis requires a strict methodological pipeline, where several aspects are considered, including acquisition protocol, image and signal processing, mathematical modelling of tissues and tracers' dynamic and kinetic properties, implementation of computational procedures for deriving perfusion parameters. Ultimately, perfusion parameters may be used to derive IBs for tumour diagnosis, prognosis, and therapy monitoring [146].

All the methodological stages discussed so far have an impact on the numerical perfusion values obtained with the different computing methods, this rising the issue of proving the reproducibility and clinical reliability of results. In fact, the lack of standardized experimental conditions for achieving reproducible results still represents the major cause hampering CTp, as well as the other dynamic imaging techniques, to become a consolidated approach used in the clinical routine. This has led CTp to undergo alternating phases of higher and lower interest from the medical community, which, despite the well-known potentiality and usefulness of the tool, has constrained CTp to a secondary role with respect to other imaging modalities, because of some technical issues. In fact, the lack of reproducibility of perfusion parameters does not justify the additional patient radiation exposure, so that CTp acquisitions are performed in most of cases only within clinical trials.

Actually, CTp is rising a renewed interest in the clinical research field, with more than 120 scientific works in the last years (according to PubMed database) addressing CTp applications in liver [147], head and neck [148], lungs [149], abdomen [150], and kidneys [151]. Three wide European multicentre liver CTp studies also exist (SARAH [4], PIXEL [3], and PROSPeCT [152]), enrolling more than 300 patients each, to evaluate promising IBs in predicting tumour development and patient prognosis. Many efforts are being addressed towards the development of motion correction artefacts [153], improve image quality with new reconstruction algorithms [154], and reducing dose radiation exposure [147], however, some difficulties still remain to have different computing methods and software yielding comparable results. In addition, no consensus exists on which the best method is for perfusion analysis referred to different anatomic districts and acquisition protocols. In several occasions, the employment of software packages have further emphasized the variability of results, especially among different institutions and scanners [155]. Moreover, it is not well known how they work, from image preprocessing to parameter computation, and this limits clinicians to be confident with the tool, and researchers to address the causes of such a variability.

Proving the reproducibility of CTp perfusion parameters has an additional complication derived by the lack of any ground truth. Hence, reproducibility of measurements would lead, implicitly, to accuracy of results. Accordingly, CTp reproducibil-

ity can be pursued based on the *common agreement* approach [156]. This principle is largely adopted in other quantitative imaging applications, and it results the only feasible also in CTP. The mutual concordance of many computing methods becomes the objective reference for bounding solutions as close as possible around the truth value.

Many studies report variations of up to 30% between perfusion values, depending on the computing methods chosen [157]. In addition, very few methodological studies deal with how to improve CTP reproducibility and even less [112] focus on the modelling aspects rather than on the computational ones. Non reproducible results may arise, for instances, from different modelling of time delays or compartmental inner structures. However, if the reproducibility of two different methods cannot be achieved based on their different theoretical underlying assumptions, one should at least define the exact modelling causes of such a diversity and, accordingly, which contribution is needed (either computable or not) in order to obtain the equivalence of measurements.

Actually, very few works perform a quantitative comparative analysis of different computing methods and in all cases the comparison refer to aggregate data, without reporting patient-wise voxel-based correlations of perfusion parameters. In this regard, a voxel-based approach has a fundamental importance. Reproducibility of CTP means equivalence of single perfusion values, not just of global averages arising from entire ROIs. Moreover, a voxel-based approach allows detecting the unreliable voxels, due to noise or artefacts, which may lead to misleading perfusion parameters [93]. One should also consider that global perfusion estimates cannot depict small local changes in tissue vasculature, thus resulting in poorly accurate clinical assessments and definitely reducing the prognostic and predictive role of CTP itself.

Reproducibility of perfusion parameters also depends on the availability of multicentre studies in order to confirm results achieved in a single institution. In fact, numerous factors can further jeopardize repeatability and reproducibility of parameters among different centres, including the employment of multiple CT scanners, acquisition protocols, tracer infusion protocols, and different training and expertise of technicians performing the examination [54]. Recently CTP, as well as other dynamic imaging techniques, have been integrated in the so called “*omics*” imaging, which combines information derived from clinics with structural and functional imaging, in order to enrich the processes of tumour diagnosis, management, and clinical decision making. Then, the reproducibility of perfusion parameters has become an urgent need because its lack can dramatically impacts on the reliability of all measurements derived.

The next Chapter 6 shows my methodological contributions to boost CTP reproducibility of two of the widely adopted computing methods, MS and DV, referring to PIXEL, a multicentre hepatic CTP study.

Chapter 5

Quantitative imaging for a radiomic approach

Radiomics is the application of machine learning methods and techniques to medical images to extract features and classify them based on different clinical purposes. The medical interest is mainly focussed on detecting patients with tumour and those without, predicting grades of therapy responses, and stratifying tumour stages. In practice, this field of application results limited by the number of cases at the disposal. Enrolling hundreds of patients is very difficult, and yet more, even if a consistent number of patients is initially included, there might be several technical and clinical reasons for excluding them during data analysis. For instance, exclusion criteria might be the presence of image noise and artefacts, employment of non standardized acquisition protocols, the presence of other comorbidities confusing the clinical assessment. Accordingly, in most situations, the variety of machine learning methods applicable reliably is limited. In this regard, linear classifiers are good candidates for working even with a few samples, to reduce overfitting and improve model generalizability. Moreover, they offer simplicity and computational attractiveness.

First, this Chapter offers a critical discussion on radiomics, proposed as the novelty of radiology (Sect. 5.1). Then, fundamentals of machine learning are provided in Sect. 5.2. Elements of statistics are provided in Sect. 5.3 to better understand binary discrimination. Hence, the radiomic workflow is presented in Sect. 5.4. The subsequent sections explain the theory of some machine learning methods and techniques widely employed in radiomics for image preprocessing (Sect. 5.5), feature generation (Sect. 5.6), preparation (Sect. 5.7), and selection (Sect. 5.8), model development (Sect. 5.9) and selection (Sect. 5.10). Most of these methods have been applied in clinical studies presented in Chapter 9, therefore the theoretic description has been enriched with the discussion of some critical issues encountered during application.

5.1 Radiomics in oncology: is there a novelty?

Over the past decade, the field of radiology has been apparently renewed with the conceptualization, in 2012, of “radiomics” as the new emerging field of extracting from radiological images a huge number of quantitative features, which may reflect the underlying pathophysiology of a tissue, and provide information of tumour stage and prognosis. Radiomics has been defined as belonging “*to the last category of innovations in medical image analysis*” [158].

Computer vision experts may reasonable argue that none innovation has been introduced in medical imaging with radiomics, since automatic machine learning techniques to derive quantitative features and large-scale data analysis methods are at least fifty years old [159]. Then, which is the real novelty? Perhaps, one might answer that radiomics, the machine learning with a new guise, has been disclosed to radiologists.

Actually, the event has brought a real revolution in radiology. The number of published radiomic studies incredibly exploded, as well as the the number of free and commercial software packages for automatic radiomic analysis. Radiomics has become the focus of international conferences and congress events, where clinicians and radiologists consider themselves experienced enough to divulge methods and techniques that normally belong to competence of computer scientists and engineers. Most of recent radiomic studies are written and reviewed by radiologists, thus being published in prestigious clinical journals, albeit in some cases far beyond from being scientifically grounded.

From a theoretical point of view, such a wide clinical interest can promote research in improving and innovating techniques for automatic medical image analysis, however, an inappropriate use of the tools at the disposal can, at the same time, delay the translation of radiomic outcomes into clinics because of the hard difficulty to achieve standardized protocols and reproducible results.

The principle of “radiologists teaching to radiologists” successfully applied in medical practice, cannot perform well when the matter of teaching is radiomics and machine learning. The process of translating machine learning methods within a new field of applications, such as radiology, is really delicate because the correctness of procedures must be preserved while their explanation to radiologists needs to be simplified to let them become familiar and confident in the methodology. First of all, radiomics is an ideological revolution of radiology, which has to be led from radiologists to convince their colleagues to open their mind to a renewed medical practice. Then, radiomics is a methodological revolution which needs to be carried out by expert scientists. Only this synergy will allow achieving reliable results, which can really improve efficacy of medical practice.

5.2 Fundamentals of Machine Learning

The purpose of this Section is to provide the fundamentals of machine learning mostly dealing with topics or methods used in the next Chapters.

Among main goals of machine learning are pattern recognition and classification, that is processing of input data collected from different kind of sensors, such as images and signals, in order to assign them to a finite number of classes or categories [160]. In particular, a pattern is any measurement extracted from the input data which needs to be classified [161]. Figure 5.1 reports a schematic workflow

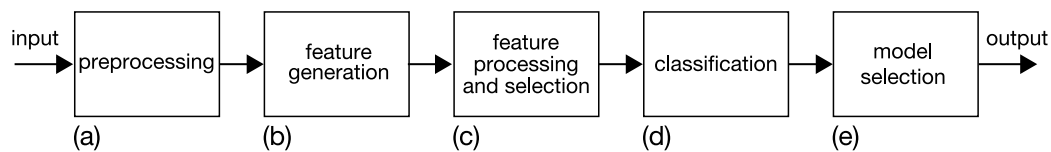


Fig. 5.1: A workflow representing the machine learning methodology made of five main steps: input preprocessing (a), feature generation (b), feature processing and selection (c), development of the classifier (d), and model selection (e).

of a machine learning process, made of five stages. As first, the input data are preprocessed in order to standardize the subsequent feature extraction procedure and improve the accurateness of measurements. Second, features or patterns are generated according to different techniques chosen based on the specific pattern recognition task. Then, data are prepared before developing the classifier. This means that data features are processed with the aim of making them homogeneous, thus allowing feature comparisons in a multivariate approach. Hence, features are selected in order to train a classifier employing only those features having the highest discriminative capability. The fourth stage is supervised classification, where the selected patterns are assigned to a target label. Finally, the performance of the classifier is assessed and the best model is selected.

Input preprocessing

Input preprocessing is the first procedure required from the machine learning pipeline and has a crucial role for the validity of subsequent measurements, and, definitely, for training correctly the classifier. A machine learning pipeline can receive several kinds of input data, such as images, texts, audio, signals, and all of them need to be converted in something understandable by machines. Accordingly, input preprocessing is the first stage of data elaboration which aims at identifying the component of interest (i.e. the signal) in the input data, by removing the effects of noise and artefacts, and standardizing the input data to ensure the comparability

of results.

As the first thing, a preprocessing workflow requires a comprehensive analysis of the inputs received, for detecting any inconsistency within data and verifying the correct acquisition of inputs and the availability of all information which will be exploited during the whole machine learning process. The workflow adopted for input preprocessing is not general purpose, but strongly depends on specific machine learning applications. Generally, it can include the use of filtering methods for denoising and removing artefacts, thus also enhancing the object of interest in the input data.

Input preprocessing plays a crucial role for data standardization. In many practical cases, inputs are heterogeneous. For instance, data can be collected from sensors using different sampling frequencies or data might be acquired from multiple signal sources. Therefore, in case of heterogeneous datasets, resampling procedures can be applied to uniform the sampling frequency, or proper normalization factors need to be detected and applied to data in order to compensate for the different acquisition sources.

Ultimately, the input preprocessing relies on a series of procedures, exploited to reduce the variability of data due to the external sources, generically referred to as noise, and preserving the real variability of data, which constitutes the information to be learnt by the classifier.

Feature generation

Features are measurable properties or characteristics derived from a phenomenon being studied, including both visible and latent aspects. The process of feature generation is based on a multiplicity of methods and techniques which allows studying the input data in a multi-layer approach.

Generally, the features extracted can be categorical or numerical. Categorical features assume a fixed and finite number of values. In particular, categorical features are nominal, if they not imply any dependence of order or rank, like days, months, or the Boolean values (true and false), whilst, they are ordinal, if they naturally imply a kind of sorting, such as the three categories small, medium, and large, or temporal ordering. Instead, numerical features are continuous or discrete variables which do not refer to a predefined set of possible values. Among numerical features, some of them refer to interval measurements, for which the differences between values are relevant, for instance dates, increments, whilst other features are employed as absolute measurements, like length, counts, age.

The main issue in feature generation is the selection of the proper class of features with respect to the specific classification task. In fact, obviously, an infinite number of features can be generated from input data and innumerable techniques

can be adopted to further increase the number of extracted features, but this may lead to an exponential rise of the computational complexity, loss of efficiency and meaningfulness of procedures. Hence, feature generation requires a deep analysis of input data for detecting the most proper way of measuring and extracting the relevant information.

Once features have been extracted from data, they constitute a vector of multiple observations organised in a tabular form. Depending on the specific classification task, the feature vector, that is a N -by- p matrix, where N are the samples or observations and p are the extracted features, can be also associated to a vector of target labels, eventually exploited by the classifier as learning examples. As such, the feature matrix is processed in the subsequent steps.

Feature processing and selection

Feature processing operations aim at improving the quality and homogeneity of the dataset, which will be employed for training the classifier. Procedures include detecting and removing records with invalid values, outlier clipping, standardizing, scaling, and normalizing numerical values, adjusting anomalous distributions, and treating missing values. Then, this intermediate step prior to training the classifier, accomplishes the crucial role of dimensionality reduction and feature selection.

In fact, the major problem occurring when dealing with high dimensional feature vectors (i.e., a very large number of features) is the consequent high number of parameters to be learnt by the classifier, which, especially for small datasets (i.e., a few number of observations) may lead to overfitting, large variance of estimates, and low generalizability. Moreover, in many cases several features are highly correlated, therefore carry on a similar information content, thus resulting redundant. Accordingly, most features can be removed from the original feature vector in order to achieve a subset of the most informative features which hold the greater discriminative power.

In practice, dimensionality reduction may be performed through two different approaches. First, a subset of the original feature vector can be achieved through combination or selection of the original features, where combination means applying transformation techniques which convert the original feature space into one with a lower dimension. Some examples of these methods are the PCA and the Independent Component Analysis (ICA). Accordingly, the classifier is trained on the transformed feature vector [161].

In most applications, working with transformed variables may yield an additional complexity in interpreting feature meaning and results. Therefore, the preferred choice is the identification within the original dataset of the most discriminative subset referring to the classification task. The two main different selection approaches

are filtering and wrapper methods, respectively. In particular, filtering methods are based on correlation measures between the features and the output variable, whilst wrapper methods performs preliminary training steps for each feature subset to detect the most effective ones. Accordingly, wrapper methods are computationally more expensive than filtering methods, which employ in most of cases fast and easy statistical techniques. However, wrapper methods yield a more accurate selection of the best subset of features. In fact, whilst filtering methods simply discard the features with the lowest discriminative power, often using univariate analysis techniques, wrapper methods select a feature subset performing a multivariate analysis and detecting the features which may yield the highest accuracy of results.

Wrapper methods generally exploit classification and regression techniques. In fact, in a combined workflow, many classification algorithms such as Recursive Feature Selection Support Vector Machine (RFE-SVM) or decision trees (DTs) are employed also for feature selection. Finally, a very popular class of regression-based methods largely employed for feature selection are shrinkage methods, which will be presented later in Sect. 5.8 [162].

Classification

The core of machine learning is partitioning the entire feature space into two or more regions where distinct patterns are grouped. Accordingly, machine learning algorithms design the decision boundary for partitioning learning data in a m -dimensional feature space, m representing the length of the vector of the selected features. Then, the decision boundary is used to make predictions on new data [161].

Machine learning problems are split into unsupervised and supervised learning [163]. In unsupervised learning, the examples are not labelled, therefore the dataset for training is constituted uniquely by the feature vector. Accordingly, the main goal of unsupervised learning is clustering, that is the extraction, automatically from the data, of groups of observations sharing similar characteristics measured by proper metrics or distance functions. Yet more, unsupervised learning is used for density estimation of distribution of data within the feature space or for visualization purposes, requiring the down-projection from a high-dimensional space to a bi- or three-dimensional one. Instead, supervised learning requires that the training set contains, besides the feature vector, a target vector where each observation is assigned to a specific class or category. Supervised machine learning algorithms are exploited in the classification tasks. Actually, a further distinction is needed for classification problems. While classification namely refers to the assignment of each input to one among a finite number of discrete classes or categories, the problem is called regression if the output is represented by continuous variables [164].

In this Thesis, I employed supervised learning algorithms. Data employed in ma-

chine learning studies, are commonly partitioned into two or three subsets, namely training and test sets, and when possible, as in our case, a validation set. The initial model is trained on the training set, made of learning examples. Then, the trained model is validated on the validation set in order to adjust the hyperparameters of the model and estimate the prediction error, accordingly. Finally, to provide an unbiased evaluation of the predictive performance of the validated model, it is tested on the holdout test set, where none of the examples has been used in the previous stages. Hence, the test of the final predictive model allows assessing the generalization error of the model itself.

Splitting training and test sets from the original dataset can be performed randomly, whilst the splitting of training and validation subsets is realized through proper resampling methods. In fact, while the hold-out test set does not enter directly in the model development phase, training and validation sets are used jointly during model training and validation. Most common resampling methods are k-fold Cross Validation (CV), Leave-one-out (LOO)-CV, multiple random splits, bootstrap. These methods require multiple runs of training, where original samples are included in the training set, other times constitute the out-of-bag samples, where the model is assessed and validated. Hence, the classification block in the machine learning pipeline provides multiple competing models, from which the optimal one needs to be selected and definitely tested.

The most diffused machine learning methods for training classifiers include DTs, neural networks (NNs), Bayesian networks (BNs), support vector machines (SVM). DTs are logic-based algorithms. A DT is realized through learning simple decision rules inferred by features. The root node of the tree is composed by the feature that best divides the training set based on a specific metric. Accordingly, the methods based on decision trees differ on the decision metric adopted to split the tree at each node. Two common approaches avoid overfitting in DTs. One is to stop the splitting before a perfect separation of the training data is achieved. Another one is to prune the tree, by imposing a limited number of leaves [165]. DTs are non parametric models, therefore easily interpretable and slightly affected by outliers. However, the computational complexity may become inefficient when working with very large datasets, and at the same time, overfitting may occur when working with small datasets. An other frequent limit is represented by the so called data fragmentation problem, which means that in some nodes there are not enough samples to make stastically significant decisions [166].

NNs are based on the perceptron concept, where decision is derived from a threshold value applied on a weighted combination of the inputs (i.e. features) of the model. However, perceptron algorithm just works for linearly separable problems. Hence, NNs realize a multi-layer perceptron-based model, where innumerable units, called neurons, make connections with each other. These units generally lie in

the intermediate hidden layer, between the input and output layers. NNs have great potentiality, although being time consuming and computationally demanding. NNs do not fit for small dataset because of overfitting risk. Moreover, NNs do not allow controlling the internal branching of the network, thus making sometimes very challenging the task of interpreting results [165].

BNs belong to statistical learning algorithms and are probability models, providing an estimate of the probability that a given observation falls within a certain class, rather than simply performing a classification. In particular, BNs are graphical models, based on Directed Acyclic Graphs (DAGs), representing probability relationships among features. The nodes in DAGs are in a one-to-one correspondence with the features [165]. In particular, each node, corresponding to each feature, is associated to a conditional probability established between the specific node and the set of its parents in the graph. DAG does not allow cycles, and features are organized in a topological sorting, where each feature comes before its descendent ones. The design of a DAG requires the knowledge of the marginal probability of the root nodes and the conditional probabilities of each non-root node, given all the combinations of possible parents [161]. Accordingly, the process of realizing a BN can be divided into two different tasks, the learning of the DAG and the determination of its parameters. Designing the DAG is the crucial point for the construction of a BN, above all if its structure is unknown. Different methods have been proposed for this purpose, albeit not reported in this manuscript. BNs have great potentiality for working with small datasets, and differently from DTs or NNs, they include the possibility to account for prior information. However, wrong hypothesis on the DAG structure due to non-statistically representative small datasets can lead to strong inaccuracies on the results.

Finally, SVM is a very useful tool for solving linear and non-linear problems, well working with small datasets. Since it is adopted in machine learning applications of this Thesis, a detailed description of the methodology is provided in the present Chapter [165].

Model selection

The last stage involved in a machine learning pipeline accomplishes the task of selecting the optimal model from a set of candidate deriving from the classification block. In fact, classification may involve the training on the same dataset of different classifiers, or of the same classifier using different hyperparameters.

A class of metrics for model selection are information criteria based on likelihood functions and applicable for parametric model-based problems. These metrics give a penalty to the candidate models based on the expected optimistic bias of the trained model. In fact, the trained models are expected to overestimate generalization per-

performances, therefore, the candidate trained models are ranked after being weighted by a penalty factor which consider both the performance achieved on the training set and the complexity of the model. Generally linear models, or models with few parameters are preferred because they show better generalization performances on the test set, with a lower probability of overfitting training data. Popular probabilistic measures for model selection are the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan and Quinn (HQ), and Bridge Criterion (BC).

The CV technique has been mentioned above for model validation, introduced into different forms, such as k-fold CV or LOO-CV based on the criterion adopted for splitting the dataset into training and validation sets. Then, CV can also be exploited for model selection. By computing the error loss function for each trained model, that one with the smallest average validation loss is selected and it is retrained on the complete training set. CV for model selection do not require parametric models, thus allowing a wide range of applications [167].

Finally, model selection can be performed in a very easy computational approach by directly comparing the predictive performance of models. For instance, one can refer to the Receiver Operating Characteristic (ROC) curve, which will be explained in detail in the following Section. Several metrics can be derived after selecting the optimal working point of the model, generally where the Youden cut-off is located, on its corresponding ROC curve. Besides sensitivity and specificity, comparison may refer to Positive Predictive Value (PPV), accuracy (ACC), informedness (I), AUC, F1-score. The main advantage of these metrics is the easiness of interpretation of both model and results. Accordingly, most of them are detailed later.

5.3 Binary discrimination

Thinking at machine learning applications into medical field, we can imagine that machine learning benefit clinical practice realizing a system from which obtaining objective decisions deriving from the experience of a huge team of clinicians and the outcomes of billions of patients worldwide [159].

Actually, at the basis of each clinical decision there is a simple process of discrimination, which means separating patients in classes, and in the simplest case, into two classes, speaking of binary discrimination. From a statistical point of view, binary discrimination relies on testing the statistical hypothesis of difference of a feature value between two classes. Hence, the aim of binary discrimination is to find out a decision threshold, either a single value, a line, or a surface to separate two classes. For instance, patients are separated between healthy or with disease, therapy responders or non-responders, high stage or low stage tumour, and so on. Substantially, patients are categorized as positive or negative with respect to a clin-

ical status or end-point. In the following, some elements of statistics are provided, including the fundamentals of statistical hypothesis testing and a very common method employed to assess the performance of a binary discrimination.

5.3.1 Statistical hypothesis testing

Statistical hypothesis testing is the reformulation of the discrimination problem into two alternative hypothesis, H_0 and H_1 , expressed as follows:

H_0 : the value of the feature differs significantly

H_1 : the value of the feature does not differ significantly

where, H_0 is called null hypothesis and H_1 is the alternative hypothesis. The choice for accepting the null or alternative hypothesis depends on experimental evidence derived from statistical information. For instance, given two classes, one can test the null hypothesis referring to the difference between the mean values, therefore the null hypothesis tests whether this difference is significantly different from zero. For the sake of simplicity, we can consider a general case of a random variable $x = x_1, x_2, \dots, x_N$, with N observations, which is defined by a probability function, $p_Z(Z; \theta)$, as the one shown in Figure 5.2, with an unknown parameter, θ . In particular, Z is referred to as *test statistic*, that is a computed quantity for

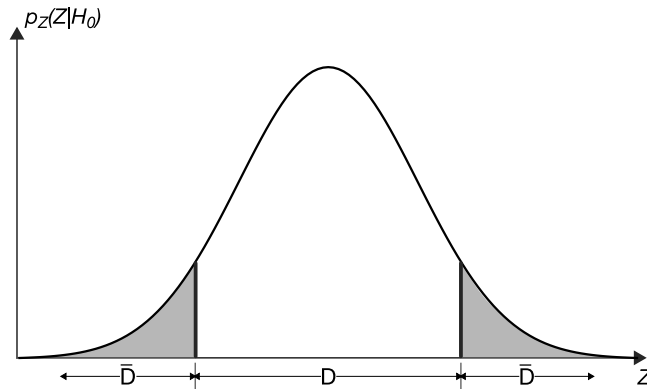


Fig. 5.2: Acceptance and rejection regions for hypothesis testing. If the value of the statistic test falls within D , H_0 is accepted, otherwise rejected.

which a probability is determined [168]. Hence, D is the area where H_0 is accepted, whilst \bar{D} is the complementary area, where H_0 is rejected. In practice, for each value of Z , if its probability falls within D , H_0 is considered true. Otherwise, if the probability of Z falls within \bar{D} , H_1 is considered true. The limit between D and \bar{D} is defined by the significance level α of the test, that is the probability chosen as the criterion for rejecting H_0 . In practice, α is the probability of rejecting the null

hypothesis when it is true. By assuming $\alpha = 0.05$ the area of \bar{D} should be equal to 0.05, this implying that the two tails in Figure 5.2 have area equals to 0.025 each. Hence, with $\alpha = 0.05$ we expect to reject H_0 if it is true, 5% of times. Based on the different statistical tests adopted, chosen according to prior information about the distribution of a variable, the value of the statistic test probability is tabulated with the statistic test value. This means that in practical cases, when performing a statistical test the probability of the computed statistic test, namely the p -value, is compared to the significance level of the test itself, as follows:

- if $p\text{-value} > \alpha$ H_0 is accepted
- if $p\text{-value} \leq \alpha$ H_0 is rejected

Being based on probability distributions, some errors are expected to occur when taking a certain decision. In this regard, Table 5.1 summarizes the two types of errors occurring in hypothesis testing. On the one hand, the error committed when

Table 5.1: Errors occurring in hypothesis testing

	If H_0 is true	If H_0 is false
If H_0 is rejected	Type I error	No error
If H_0 is not rejected	No error	Type II error

rejecting the null hypothesis if it is true is called Type I error and is expected to occur with a frequency of α . On the other hand, the error of not rejecting H_0 if it is false is referred to as Type II error and is represented by β . Accordingly, the power of a statistical test is $1 - \beta$, that is the probability of rejecting H_0 if it is false [168].

Let us consider the example case of testing the null hypothesis of null mean of a population, $\mu=0$. Hence, the statistic test Z is expressed as in Equation 5.1:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \quad (5.1)$$

where \bar{X} and $\sigma_{\bar{X}}$ are the sample mean and standard deviation. Hence, if H_0 is not rejected (Equation 5.2):

$$P \left[-Z_{0.05} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq Z_{0.05} \right] = 0.95 \quad (5.2)$$

where the significance level α is assumed to be 0.05. Equation 5.2 can also rewritten as Equation 5.3, in terms of the so called confidence limits.

$$P [\bar{X} - Z_{0.05}\sigma_{\bar{X}} \leq \mu \leq \bar{X} + Z_{0.05}\sigma_{\bar{X}}] = 0.95 \quad (5.3)$$

In particular it follows (Equations 5.4, 5.5):

$$L_1 = \bar{X} - Z_{0.05}\sigma_{\bar{X}} \quad (5.4)$$

$$L_2 = \bar{X} + Z_{0.05}\sigma_{\bar{X}} \quad (5.5)$$

where L_1 and L_2 are respectively the lower and upper confidence limits, whilst the distance between them defines the confidence interval (Equation 5.6):

$$\bar{X} \pm Z_{0.05}\sigma_{\bar{X}} \quad (5.6)$$

Hence, the significance level of the test, α , implicitly defines also the confidence interval (CI), which is $1 - \alpha$. Generally, confidence limits are an estimate of the precision of measures [168].

To the aim of a binary discrimination, the statistical hypothesis testing is a useful and easy tool for detecting and discarding irrelevant features, having a poor discriminative capability. However, it does not provide any information regarding the goodness of discrimination, in other words the overlapping between the two classes. For instance, the mean values of two classes may be significantly different but their large spread may lead to a poor separation between classes [161].

5.3.2 The ROC curve

A binary discrimination is graphically assessed through the ROC curve, widely employed for visualizing, organizing and selecting classifiers based on their performance. As such, ROC curves are also extensively used in the medical decision making community for evaluating the performance of supervised classifiers.

The output of a classifier can be a continuous variable which is converted into predicted class labels according to a threshold value or a binary discrete variable representing the class membership. By comparing the predicted class labels to the actual ones, four different cases may substantially occur, which are outlined in Table 5.2, known as confusion matrix. In particular, if a positive (\mathbf{p}') instance is correctly classified as positive (\mathbf{p}), it is referred to as True Positive (TP), whilst if it is misclassified as negative (\mathbf{n}) it is counted as False Negative (FN). Similarly, if a negative (\mathbf{n}') instance is correctly classified as negative, it is referred to as True Negative (TN), whilst if it is misclassified as positive it is counted as False Positive (FP). Given a two-by-two confusion matrix, several metrics can be calculated accordingly and reported in Table 5.3. In particular, sensitivity and specificity corresponding to the TP rate (TPR) and TN rate (TNR), respectively, refer in medical practice to the probability of detection and false alarm rate. Moreover, the count of FP and FN represent the Type I and Type II errors, respectively, mentioned in Sect. 5.3.1.

Table 5.2: Confusion matrix of a supervised binary classifier, where **p** and **n** are the positive and negative classes, respectively.

		Predicted class		total
		p	n	
Actual class	p'	True Positive	False Negative	P'
	n'	False Positive	True Negative	N'
total		P	N	

Table 5.3: Metrics derived from a two-by-two confusion matrix

Metric	Formula
sensitivity (TPR)	$\frac{TP}{TP+FN}$
specificity (TNR)	$\frac{TN}{TN+FP}$
accuracy (ACC)	$\frac{TP+TN}{TP+TN+FP+FN}$
positive predictive value (PPV)	$\frac{TP}{TP+FP}$
negative predictive value (NPV)	$\frac{TN}{TN+FN}$
false negative rate (FNR)	$\frac{FN}{TP+FN}$
false positive rate (FPR)	$\frac{FP}{TN+FP}$
false discovery rate (FDR)	$\frac{FP}{TP+FP}$
false omission rate (FOR)	$\frac{FN}{TN+FN}$
F ₁ score	$\frac{2TP}{2TP+FP+FN}$
F ₂ score	$\frac{5TP}{5TP+FP+4FN}$

The ROC curve is defined in the space given by the FPR (x -axis) and TPR (y -axis), with $FPR=1 - TNR$. Accordingly, each confusion matrix is represented through a single point in the ROC space. For instance, Figure 5.3 (a) shows the representative points of five different discrete classifiers in the ROC space. The ideal best classifier is located on the left corner of the space, where both sensitivity and specificity are equal to the 100%, whilst the point located on the diagonal $y = x$,

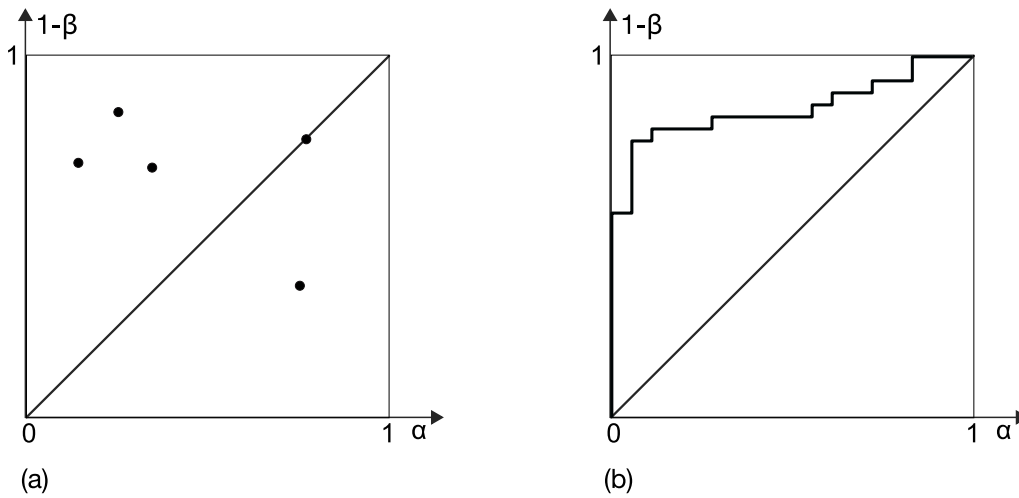


Fig. 5.3: ROC space represented by the FPR (x -axis) and TPR (y -axis) where each point represents the performance of a discrete binary classifier (a). Instead, a ROC curve is derived from a continuous classifier as the threshold, T , applied on the output variable is varied (b).

has same performance of a random assignment.

If a classifier yields a continuous variable as output on which a threshold is applied to achieve a binary outcome (i.e. to predict the positive or negative class), the classifier itself is represented through a curve in the ROC space, representing its performance as the threshold varies. Hence, Figure 5.3 (b) shows an example of ROC curve. If the ROC curve is in the lower right triangle of the ROC space, the classifier performs worse than the random guessing, whilst, a classifier performs as much better as its ROC curve departs from the diagonal towards the left upper triangle of the ROC space. A very common metric employed to compare ROC curves is the AUC. Since the AUC is a portion of the unit square given by TPR and FPR ranging within the interval $[0, 1]$, its value is between 0 and 1 as well. Hence, the random classifier has $AUC=0.5$. The output of a continuous classifiers can also be seen through plotting two probability density functions, $\omega_0(T)$ and $\omega_1(T)$ corresponding to the probability of assigning an instance to the negative and positive class, respectively, with T the threshold parameter, varied in order to have a binary outcome from a continuous variable. Hence, the $TPR(T)=\int_T^\infty \omega_1(x)dx$, whilst $FPR(T)=\int_T^\infty \omega_0(x)dx$. Accordingly, Figure 5.4, shows the corresponding portions of TP, TN, FP, and FN under the probability density function curves.

Given a ROC curve, the threshold T to achieve the best possible classifier is selected from the Youden Index (YI) which identifies the point on the ROC curve most distant from the diagonal. That point maximizes the informedness (I) of the classifier, where $I=\text{specificity}+\text{sensitivity}-1$, used in practical cases to chose the best trade-off between sensitivity and specificity metrics [169].

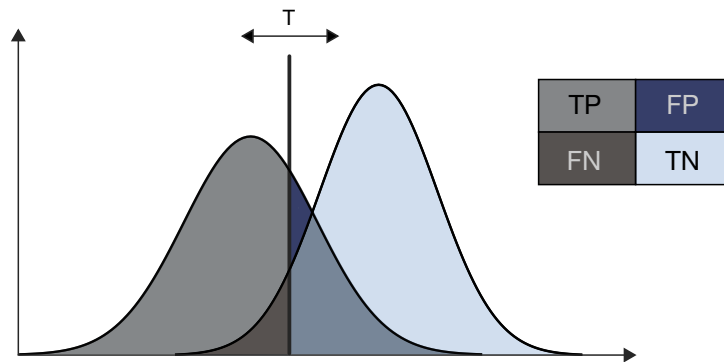


Fig. 5.4: Portions of TP, TN, FP, and FN under the two probability density functions $\omega_1(T)$ and $\omega_0(T)$, referring to the positive and negative class, respectively.

5.4 A radiomic workflow

On the ground of what stated in Sect. 5.1 and explained in Sect. 5.2, a radiomic approach relies on the application of machine learning methodology to radiological image analysis. Accordingly, a radiomic workflow consists of three main steps, highlighted in Figure 5.5. The first step is image processing (Figure 5.5 (1)), including image acquisitions, preprocessing and preparation (e.g. image denoising and resampling), and ROIs segmentation. Then, radiomics features are extracted (Figure 5.5 (2)) from the ROIs. In most of applications, the radiomic features include shape features which assess the morphometric characteristics of tumour ROIs, first order statistical features that are histogram-based descriptors, second and higher order texture features calculating the statistical spatial relationships between neighbouring pixels, and features derived applying different filters or transform on images, such as wavelet transforms [70]. Then, radiomic features are processed and analysed (Figure 5.5 (3)) in order to select, either in an univariate or multivariate approach, the features with the highest discriminative capability to be applied for developing the predictive model. Hence, this last third step includes also model development, selection and assessment of the classification performance [68].

5.5 Image processing

A radiomic workflow requires acquisition of images according to a specific standardized protocol, based on the clinical purpose, imaging modality, and patient characteristics, with the aim of achieving reproducible results in all the subsequent steps of quantitative imaging, including feature extraction and their application for clinical purposes [68]. Several studies report numerous sources of noise and variabil-

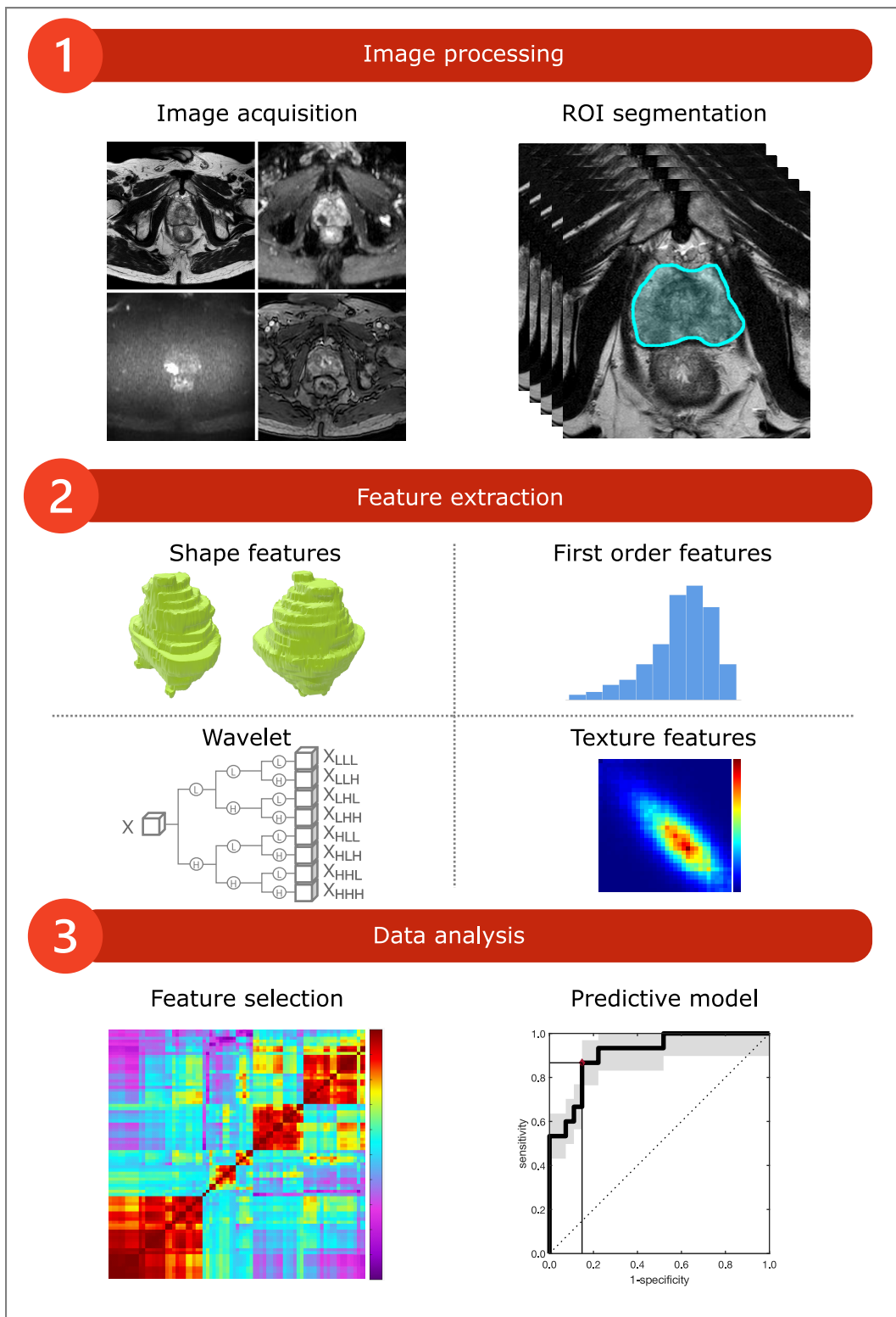


Fig. 5.5: The radiomic workflow consisting of three main steps: (1) image preprocessing, including image acquisition and preparation (e.g. ROI segmentation); (2) feature extractions, such as shape features, first order and texture features, and wavelet transforms; (3) radiomic data analysis including the selection of the most discriminative radiomic features and development of the predictive model.

ity in the early steps of the radiomic workflow which hamper the clinical reliability of the results, besides delaying the translation of radiomics outcomes into clinics and weakening its potential role in clinical decision-making [170]. Since data variability is in most of cases unavoidable, the only way to cope with varying scanners, technologies, acquisition protocols and image reconstruction algorithms is to set up properly preprocessing steps, to reduce as much as possible, the influence of these sources of noise on the results. So far, radiomic studies have been performed mostly in single centre institutions, whereas multicentre analyses are needed to prove the clinical meaningfulness of radiomics as a diagnostic, prognostic or predicting tool. Variability of data collected in multicentre study is incredibly high and complex to handle, and pre-treatment of data plays a crucial role [171].

Since the initial spread of radiomics [158], CT has been the most widely adopted modality in radiomic studies, mainly due to its broad availability. Later on, MRI has gained an increasing role, so that nowadays, among more than one thousand scientific papers published on radiomic studies in 2020, around eight hundred are equally split between CT and MRI (according to PubMed database). Subregions of tumour habitat as well as metastatic lesions and normal tissues are generally imaged and analysed exploiting radiomic approaches. Among several applications, CT is the reference imaging technology for radiomic studies regarding head and neck [172] and lung cancers [173], whilst MRI is largely adopted in the study of breast [174], prostate [175], abdominal tumours [176]. Also PET radiomics is a growing field and most of radiomic studies refer to the use of ^{18}F -FDG. In addition, the availability of hybrid systems has also offered new opportunities to associate incremental radiomic information derived from both morphological and functional sequences with the tumour tissue biology and behaviour. Actually, although PET imaging has a great promising role in radiomics since being highly specific in catching abnormalities in tumour biology, it shows several technical limitations. In fact, if compared to CT and MRI, PET images show lower resolution and higher noise, and often finding the correlation of radiomic features derived from PET and morphological imaging can result very challenging [177].

Data preparation of acquired images requires segmentation of ROIs from which radiomic features are subsequently extracted. The process is normally carried out manually by expert radiologists which identify the most informative tissue area in order to answer the clinical question and delineate for each image slice a contour of the selected ROI [178]. Indeed, a large number of tools for automatic segmentation have been proposed, although with a limited success, because of the difficulties to achieve a large-scale fully validated segmentation tool. More often, semi-automatic methods are used, which also allows manual refinements of segmentations. For instance, region growing and watershed are two widely used methods in the majority of CAD systems [68]. In particular, in the region growing approach, after the

operator-dependent selection of a pixel seed, its neighboring pixels are iteratively added to the foreground region based on a similarity criterion [179], whereas, the watershed method requires a manual selection of a ROI on a single slice in order to generate an initial three-dimensional surface of the lesion, refined with active geometric contours [180]. If semi-automatic and, above all, automatic tools allow strongly reducing the time required for segmentations, they often need manual refinements, since noise in the image, low contrast, smoothness of edges may cause poorly accurate segmentations. In some cases, reproducibility of radiomic features extracted from semi-automatic or automatic regions have shown higher reproducibility than those derived from manual segmented ROIs [181]. However manual segmentation, although being time consuming, still remains the ground truth in order to achieve the highest accuracy of the results. Accordingly, until now, in morphological imaging most of radiomic studies adopt manually segmented ROIs [178].

Depending on the image acquisition protocols and the dataset being analysed, data preparation can involve also filtering, denoising, resampling. In fact, in most of cases radiomic studies are conducted on retrospective datasets. This means that data are originally acquired for different research purposes and raw data are not usually stored, so that often reconstructing images differently is not possible. Images may need to be pre-processed in order to reduce their differences. For instance, heterogeneous resolutions can be interpolated to a common voxel size, or images can be filtered to achieve similar noise characteristics [171]. It is worth noting that preprocessing can dramatically alter information retained by data, it can result very challenging to be performed while preserving the original information and guaranteeing the full validity of the subsequent computational steps. A data-driven selection for the most proper solution is recommended.

5.6 Feature generation

The process of feature generation is a key step in classification. Several computer vision methods can be applied to the field of medical imaging for feature generation and, based on both the clinical objective and either characteristics or dimensionality of data, a researcher is required to choose the most appropriate strategy [182]. In oncological applications, the features extracted from radiological images can allow measuring objectively the properties of normal and abnormal structures and highlighting latent properties, expected to be more powerful in predicting tumour aggressiveness and therapy response than visual findings. By assuming a coarse-grain taxonomy, imaging features can be organized into three main categories, morphological features, grey level (GL)-based, and filter-based features [182]. In the following, I address just some of the imaging features, currently the most widely exploited

in radiomic studies [70]. In this regard, my discussion focuses on shape features as morphological descriptors [183], first order features [70], texture features of second [184] and higher orders [185], as GL-based descriptors, and wavelet transforms in the frequency domain, as filter-based features [186].

Table 5.4 provides an overview of the main characteristics of the shape, first

Table 5.4: Overview of the main characteristics of the most adopted radiomic features, where D stands for diagnosis and P for both prognosis and prediction

Features	Category	Property	Role	Complexity	Interpretability
Shape	Structural	Morphology	D	low	simple
First order	Statistical	Histogram	D, P	low	simple
Texture	Statistical	Texture	D, P	high	difficult
Wavelet	Filter-based	Texture	D, P	medium	difficult

order, texture, and wavelet features included in the discussion. In particular, the comparison among these four classes of features relies on five indicators, category, property, role addressed in clinical applications, computational complexity, and clinical interpretability. Shape features are the only ones which measure structural properties based on geometrical and morphological descriptors of the tissue ROIs, without any relationship with the image GL-distributions. Hence, information content provided by basic morphological descriptors is strictly correlated to the visual interpretation of images performed by radiologists [70]. Accordingly, shape features, also with a very low computational complexity, enable simple clinical interpretability. In clinical applications, shape features based on simple morphological descriptors can have good performance if applied for diagnosis (referred to as D in Table 5.4) [187]. In fact, on several occasions, the increase or decrease of tumour size are late descriptors of tumour progression or good response to anti-cancer therapies, respectively. Thus, basic morphological features can have a poorer prognostic or predictive value than other features referred to GL descriptors. On the contrary, first order, texture, and wavelet-based features are frequently adopted not only for diagnosis, but also for prognosis and prediction of therapy response (both referred to as P in Table 5.4). Several studies adopt for clinical classification purposes, a compound of different feature classes in order to catch various aspects of the tissue under investigation [188]. While first order features measure the statistical properties of GL-based histograms, with a low computational complexity and simple clinical interpretability, both texture and wavelet-based features can provide measure of the texture of the imaged tissue. In particular, the former are based on statistical properties of bi-dimensional GL histograms, the latter rely on the characteristics of the filters employed for the frequency transform. Then, the computational complexity and clinical interpretability rise accordingly.

Shape features

Shape features are descriptors of geometry and morphology of the ROI independent from the GL distributions, calculated on the segmentation mask rather than directly on images [70]. Generally, multi slice ROIs are collected to allow computing 3D shape features, referred to the entire segmented volume. In this regard, ROI mask volume is not conceived as a collection of voxels, each one contributing with a single volume element, but it is described through the set of coordinates of voxel centres, thus avoiding partial volume effects often occurring in ROI segmentation. Accordingly, to represent the outer structure, a surface mesh is computed, thus reducing the overestimation of the surface area when all single voxels are considered as single elements. A triangle mesh is derived from the surface of the ROI using different meshing algorithms, such as the Marching Cubes [189], that is one of the most widely adopted ones. A volumetric mesh is drawn with triangle faces and vertices covering the entire volume ROI: faces are delimited with lines, that are connection between two adjacent vertices, and two lines only can share one vertex. To compute the mesh volume, all faces should have the same orientation of the normal of the face. Consequently, the correct orientation of the mesh faces leads to all face edges being shared between exactly two faces. The most widely used shape features include *mesh surface*, *volume surface*, *surface area to volume ratio*, *asphericity*, *sphericity*, *spherical disproportion*, *maximum diameter*, *major axis length*, *minor axis length*, *least axis length*, *elongation*, *flatness*, *volume density* [183].

Although shape features have a very low computational cost and allow an easy clinical interpretation, they strongly depend on the segmentation, relying on both intra-reader and inter-reader variability. Moreover, shape features perform a morphological assessment of tumour, which may have mainly a diagnostic role, thus not representing tissue heterogeneity [68]. Shape features are computed to provide a quantitative measure of tumour extent and structure, although none of the hidden properties of the tumour is caught. Shape features are also a valid tool for measuring objectively image findings that radiologists are used to assess qualitatively. According to what reported in one of the most recent radiomic multicentre studies [173], also in agreement with the majority of scientific articles in the literature, when compared with statistical descriptors or higher order texture features or filter transforms, shape features suffer from a lower sensitivity. Therefore in the most of cases, other feature classes are adopted.

First order features

First order (FO) features describe the GL distribution of pixels without considering any spatial relationship between them. Basically, FO features are histogram de-

scriptors such as *mean*, *median*, *skewness*, *kurtosis*, *minimum* and *maximum value*, *percentile*, computed referring to the 3D volume distribution of GLs arising from multi-slice segmented ROIs. Most of these FO features represent the average (e.g. *mean*, *median*, etc.) and variation of data (e.g. *interquartile range*, *percentile*, etc.), whilst some of them focus on the shape of the distribution, usually represented by moments (e.g. *skewness*, *kurtosis*, etc.). FO features can also account for the diversity of GL values, by applying concepts derived from information theory to quantify the heterogeneity of image values. Some examples of this kind of FO features are *entropy* and *uniformity*. In particular, *entropy* measures the information content of the image, that is quantifies uncertainty by weighting the information content by its probability: the higher the entropy, the more heterogeneous the image content. In practice, FO metrics measuring diversity of values are not computed directly on the histogram of GLs, but after its rebinning, applied in order to derive the relative frequentist probability of each range of GLs [190]. FO features are robust descriptors, reported as the most repeatable and reproducible features [191], also allowing an easy computation and clinical interpretation, since representing at the coarse level tissue inhomogeneities and abnormalities. The highest number of software packages for radiomic analysis shares the same computational implementation of FO features, thus also explaining their higher reproducibility if compared with other radiomic features [192]. However, since FO features arise from global observation of tissue ROIs, their sensitivity in catching small local abnormalities can be low, thus limiting a wide applicability of FO features for personalized medicine. FO features lack the spatial references in characterizing the GL distributions of tissue structures, this being one of the major reasons why the texture features have gained such a large interest and application in radiomic studies [190].

Second order texture features

Texture is one of the characteristic properties of images, one of the aspects being used by human eye in interpreting pictorial information of images, thus unravelling the structural arrangement of surfaces and their relationship to the surrounding environment. Texture features are calculated in the spatial domain and mostly rely on the assumption that the texture information of an image is contained in the average spatial relationships among GLs. Therefore, the extraction of texture features is based on computing a set of GL spatial-dependence probability distribution matrices [184].

Originally, texture analysis was thought for assessing surface texture in 2D images, and it has been later extended to 3D volume. Second order texture features are based on GL co-occurrence matrices (GLCMs), which express the combination of discretised GLs of neighbouring pixels, or voxels in 3D GLCMs, along specific image

directions [184]. In practice, GLCMs are bi-dimensional histograms quantifying how much frequently similar pixel-values are located next to each other, by considering the neighbourhood as a unique direction at a certain distance (d). Therefore, second order statistics are derived from the spatial relationship of two pixel values and not from the values themselves. The HU values within an image ROI often need to be quantised into a fixed number of GLs (e.g. 32, 64, 128), before computing the GLCM, a matrix in which both rows and columns represent a set of possible image values. Given d , each value (i, j) of the GLCM, $P_d[i, j]$ counts all pairs of pixels having GLs i and j , separated by the so-called displacement vector, established in order to examine the texture based on d and the orientation. By convention, pixels to the east of the reference pixel are at 0° , to the north-east at 45° , the north at 90° , and north-west at 135° . When considering 2D GLCMs, these four symmetric unique directions are considered, thus representing a 8-connected neighbourhood. In particular, the displacement vectors depending on d and corresponding to these four directions are reported in Table 5.5. When considering 3D GLCMs, thirteen

Table 5.5: Four unique directions of 2D GLCM

Displacement vector	Direction ($^\circ, \theta$)
(0, d, 0)	0
(-d, d, 0)	45
(-d, 0, 0)	90
(-d, -d, 0)	135

unique symmetric directions are investigated, thus corresponding to a 26-connected neighbourhood, and reported in Table 5.6. In particular, by considering the orientation of the three x, y, z -axes, the angles θ and ϕ are measured between x and y , and y and z , respectively. The values on the main diagonal of GLCMs represent the joint probabilities of finding identical pixel-values in the original image, therefore by moving away from the main diagonal, GLCMs represent the increasing difference between intensity values in the original image. The texture features most widely extracted from GLCMs are the ones proposed in the early seventies by Haralick [184]. All these metrics weight the entries of the GLCM by some value depending on what properties need to be highlighted. For instance, *angular second moment* squares the elements of the GLCM and then sums them up. Therefore, the lower the difference between values, the higher the value of uniformity. *Contrast* quantifies the degree of different intensity values along given orientation and distance. Since the texture is an intrinsic property of an image, results have to be rotationally invariant, therefore second order features are derived from 2D GLCMs

Table 5.6: Thirteen unique directions of 3D GLCM

Displacement vector	Direction ($^{\circ}$, θ, ϕ)
(d, 0, d)	(0,45)
(d, 0, 0)	(0,90)
(d, 0, -d)	(0,135)
(d, d, d)	(45,45)
(d, d, 0)	(45,90)
(d, d, -d)	(45,135)
(0, d, d)	(90,45)
(0, d, 0)	(90,90)
(0, d, -d)	(90,135)
(-d, d, d)	(135,45)
(-d, d, 0)	(135,90)
(-d, d, -d)	(135,135)
(0, 0, d)	(0,0)

after direction-weighting the GLCMs themselves [190]. GLCMs and corresponding second order texture features have low computational complexity – albeit greater than shape and FO features – and are computed in all radiomic packages [192]. Nevertheless, some open issues are still present, referred to their most reliable and reproducible employment in radiomic studies. For instance, no agreement still exists on how to aggregate GLCM information with the highest reproducibility in clinical applications, that is whether extracting single representative texture features from 3D GLCMs or averaging texture features of multi-slices. Actually, most clinical studies rely on 2D GLCMs-based features due to their lower computational complexity than 3D GLCMs. However, characterizing tumour with multi-slice images (i.e. features derived from single slices) can lead to a high approximation of the entire tumour volume, thus also jeopardizing the accurateness of the description of tumour heterogeneity, especially when a low spatial resolution along the axial direction increases the partial volume effects. In this regard, recent studies report a better clinical correlation achieved with 3D texture features and more reproducibility of 3D rather than 2D features [193]. Moreover, the choice of the distance d at which consider the neighbouring pixels, affects the performance of second order texture features in classification. Finally, preprocessing steps of image denoising, filtering and quantization, can dramatically affect image texture, thus making more difficult the reproducibility of results [183].

Higher order texture features

While second order texture features consider spatial relationships between two neighbouring pixels, higher order statistical features assess the relationship between three or more pixels. The original and easiest way to derive higher order statistical texture features has been proposed in 1975 by Galloway, with the introduction of the GL Run Length matrices (GLRLMs), where a GL run is defined as a line of pixels in a certain direction with the same intensity value. Accordingly, the number of pixel constituting the run line is called run length and the number of its occurrence is the run length value. In a GLRLM, the rows represent the discretized GLs, the columns the run lengths. In practice, a run length describes the neighbouring pixels having the same GL in a specific orientation or direction. The most commonly used texture features derived from the GLRLMs are long run emphasis, short run emphasis, GL uniformity, and run length uniformity. Generally, the combinations of these parameters are used to constitute a feature vector. Similarly to GLCMs, the GLRLMs are computed for different directions and then averaged to obtain rotationally invariant results [185]. Higher order texture features have both a higher computational complexity and a harder clinical interpretation. Despite frequently included in software tools for radiomic analyses, their clinical application in classification purposes is not widespread.

It is worth noting that the behaviour of texture features, referring to both second and higher order statistics, is frequently affected by imaging modalities and their acquisition parameters [186]. For instance, slice thickness is one of the factors mainly affecting values and reproducibility of texture features. It is well-recognised that a thin slice thickness can improve measurements of tumour volume and its changes over time, thus also improving clinical assessments. However, if referring to the computation of texture features, one should also consider some other factors. If on the one hand, thinner slice images are often associated with higher noise levels, which can negatively affect texture features, on the other hand, thicker slices can lead to blur images, making the characteristic texture patterns more difficult to identify, thus resulting in the loses of some texture details due to a poorer spatial resolution along the axial direction. Therefore, the clinical reliability of texture features mostly relies on achieving the best trade-off between feature reproducibility and capability of characterizing tissue texture in detail [193].

Wavelet

Images, acquired and normally interpreted in the spatial domain, can be transformed into the frequency domain in order to emphasize frequency-based features. In fact, instead of assigning an intensity value to each couple of spatial coordinates

of the image, in the frequency domain images are described through the pattern and rate at which intensity values change along different directions [190].

One of the most powerful mathematical transformations, largely exploited in image analysis, are the discrete wavelet transforms, which are non-stationary signals alternating low frequency components such as the background and high frequency components such as edges and structures. In particular, discrete wavelet transforms represent images through a basic function, called “mother wavelet”, properly scaled and shifted in order to fit the original signal. Hence, the wavelet can localize the original signal in both time and frequency domains, and this is referred to as the main advantage of wavelet transform if compared to Fourier transform that represent signals in frequency only. Moreover, wavelet transforms perform a multi-resolution analysis, through which the original image is decomposed into low- and high-frequency components at different scales. The process is realized through a bank of filters, the original image is convolved with the mother wavelet, in a single and repeated downsampling step along each direction (i.e., rows and columns). Therefore, wavelet decomposition yields a set of wavelet coefficients corresponding to different scales and directions [186].

Given an image X , and let L and H be respectively a low-pass and high-pass filter, an example of the wavelet decompositions of X is referred to as X_{LLH} , interpreted as the application of a L filter along x- and y- directions, and a H filter along z-direction, thus being constructed as in Equation 5.7:

$$X_{LLH}(i, j, k) = \sum_{p=1}^{N_L} \sum_{q=1}^{N_L} \sum_{r=1}^{N_H} L(p)L(q)H(r)X(i+p, j+q, k+r) \quad (5.7)$$

where, N_L is the length of the filter L, and N_H is the length of H. Each component such as X_{LLH} of the original image is called sub-band, and first-order and texture features can be derived on each sub-band of the transformed image [61].

The choice of the mother wavelet dramatically impacts on the information content preserved by the decompositions, and consequently, on the accuracy of any post-processing procedure, such as the computation of radiomic features. In this regard, the key characteristics for the selection of the mother wavelet are the number of the filter coefficients, the number of vanishing moments, symmetry, orthogonality, and computational complexity. For instance, a large number of coefficients produces smoother decompositions but with a higher computational complexity [194]. The number of vanishing moments controls the ability of the mother wavelet of enhancing structures that can be approximated with high-degree polynomials. In particular, the number of vanishing moments equals the degree of the polynomial of the wavelet scaling function up to which the derivative of the Fourier transform is zero, and this means that up to that degree of polynomial function, the wavelet scal-

ing function alone can be used to represent those structures. In addition, symmetry affects the equality of decomposition of the image, independently from the forward or reverse direction of convolution. Finally, orthogonality establishes to what extent decompositions are correlated along the different directions. For all these reasons, the identification of the optimal mother wavelet can be very challenging and also strictly dependent from the imaging modality, acquisition parameters, and tumour phenotype being analysed [186]. Some recent studies strongly promote the use of wavelet transforms, which are reported as having a high prognostic value of tumour aggressiveness [195], and also allow achieving more robust and reproducible results of first order and texture features computed on decomposed images if compared to feature values computed on the original images. Of course, working in a transformed image domain yields an additional complexity in interpreting outcomes and correlating them with the underlying biological interpretation [193].

5.7 Data preparation

As explained above, in Sect. 5.2, homogeneity of data is an essential property for training the classifier correctly. In the following, two basic aspects of data preparation will be presented, outlier removal and data normalization and standardization, which have been employed in clinical applications presented in the subsequent Chapters.

5.7.1 Outlier removal

The process of outlier removal, herein presented, mainly refers to data distributed according to a random process. Accordingly, radiomic features are considered as discrete random variables.

Given a random variable, the mean value is the reference for localizing the expected value of the variable itself, while the standard deviation gives a measure of dispersion of the range of the variable. Accordingly, an outlier is defined as a feature which is very distant from the mean, given a certain threshold, usually expressed in terms of standard deviation. A distance of two or three times the standard deviation covers the 95% or the 99% of the distribution, respectively. Therefore, all data points beyond this limit, can be considered outliers, that is not representative of the real behaviour of the random variable. Depending on the original shape of the distribution, outlier removal can also be performed on one site only. The presence of outliers may produce large errors during training, therefore in most practical cases they are removed.

However, if the dataset at the disposal is not big enough to completely drop off

outliers, some other techniques are used to prevent the influence of outliers on the model. For instance, when designing the classifier it may be adopted a cost function that is less sensitive to outliers. Alternatively, after removing outliers, these may be treated as missing values, therefore predicted heuristically, by assuming the expected value or the extreme value of the distribution after outlier removal. In particular, in this last case, prediction of the missing value accounts for the the prior information regarding its original localization within the distribution and attempts to preserve it [161].

It is worth noting that this kind of outlier removal process is employed in univariate feature analysis, when each feature is considered and assessed separately. As such, the process is largely used when working with few samples, where multivariate feature analysis results difficult to perform reliably.

5.7.2 Data standardization and normalization

Commonly, feature values lie within very different dynamic ranges. When conducting a multivariate analysis, features need to be comparable. In fact, having a high variability among feature ranges may lead the features with the wider range and higher absolute values to predominate the classifier even if they have a lower discriminative power than features with narrower ranges and lower values. To prevent this, data are generally rescaled into similar ranges, through linear and non-linear methods. The most popular methods are data standardization and $[0, 1]$ -normalization. Actually, data standardization is normalization through mean and variance. In practice, given a feature variable of N observations, the value of each of them is converted as in Equation 5.10, through the operators described in Equations 5.8, 5.9:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (5.8)$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (5.9)$$

$$\hat{x} = \frac{x - \bar{x}}{\sigma} \quad (5.10)$$

Hence, the standardized features have null mean and unit variance. As a consequence, the process yields features lying into similar ranges, symmetric with respect to zero.

Instead, data normalization, also called min-max feature scaling, provides features within the range $[0,1]$. Hence, data are rescaled according to Equation 5.11:

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5.11)$$

Actually, Equation 5.11 can also be rewritten as Equation 5.12 to represent the more general case of feature scaling within the interval $[a, b]$, with $b > a$.

$$\hat{x} = a + \frac{(x - x_{min})(b - a)}{x_{max} - x_{min}} \quad (5.12)$$

Data normalization modifies feature range without affecting the original variance of data. Therefore, normalized features have identical ranges and different variances. The choice of preserving the original variance of data generally depends on applications, although one should consider that variance means information, therefore features with higher variance hold intrinsically a greater discriminative power. Moreover, selecting the most proper data normalization method may depend on the classifier adopted. Different classifiers are known to perform better when working with standardized or normalized data [161].

5.8 Shrinkage methods for subset selection

If a given feature vector X is constituted by m features, the methods for subset selection allow reducing the number of m features to p , with $p \leq m$. Shrinkage methods are based on the idea of exploiting linear regression for feature selection. Accordingly, being based on linear methods, they are particularly suitable in case of a limited number of observations, low SNR or sparse data.

By considering the feature vector $X^T = (X_1, X_2, \dots, X_m)$, the linear model for predicting through regression the real output Y is expressed as in Equation 5.13:

$$Y = f(X) = \beta_0 + \sum_{j=1}^m X_j \beta_j \quad (5.13)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_m)^T$ is the vector of coefficients, which need to be estimated, generally through least squares minimization. Hence, given a feature vector of N observation, each $x_{ij} = (x_{i1}, x_{i2}, \dots, x_{im})$ is a vector of feature measurements for the i -th observation. Accordingly, the minimization of the residual sum of squares is given by Equation 5.14:

$$\begin{aligned} RSS(\beta) &= \sum_{i=1}^m (y_i - f(x_i))^2 \\ &= \sum_{i=1}^m (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \beta_j) \end{aligned} \quad (5.14)$$

Actually, linear regression has two major disadvantages, that is prediction accuracy and interpretation. The least square estimates have often low bias but large variance.

An useful way to improve their accuracy is by shrinking some coefficients to zero. In addition, a large number of predictors makes the interpretation of the linear combination which will determine the classifier performance more difficult [196].

Shrinkage methods arise from linear regression by adding a regularization term to the error function, also in order to control over-fitting. For instance, *ridge* regression imposes a penalty on the size of coefficients, so that the vector of coefficients is derived from Equation 5.15:

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^m \beta_j^2 \right\} \quad (5.15)$$

In particular, $\lambda \geq 0$, is the regularization parameter, the penalty factor, which determines the amount of shrinkage. The higher the value of λ , the more coefficients are shrunk towards zero. Then, Equation 5.15 can be rewritten into Equation 5.16:

$$\begin{aligned} \hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} & \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \beta_j)^2 \\ \text{subject to} & \sum_{j=1}^m \beta_j^2 \leq t \end{aligned} \quad (5.16)$$

which makes the constrain t on the coefficients explicit. Generally, the ridge solution depends on the variance of features, therefore the feature vector needs to be standardized (i.e. null mean, unit variance) before solving Equation 5.16. In addition, thanks to its penalty term, the ridge regression induces the shrinkage of correlated variables.

An alternative method to ridge regression is represented by the *lasso* shrinkage method. Similarly to the ridge method, Equation 5.16 is expressed for lasso as in Equation 5.17 [196]:

$$\begin{aligned} \hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} & \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \beta_j)^2 \\ \text{subject to} & \sum_{j=1}^m |\beta_j| \leq t \end{aligned} \quad (5.17)$$

Accordingly, Equation 5.17 can be rewritten as Equation 5.18, its equivalent Lagrangian form:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^m |\beta_j| \right\} \quad (5.18)$$

Hence, while the penalty term in Equation 5.15 expresses the L_2 -norm ($\lambda \sum_{j=1}^m \beta_j^2$) that one in Equation 5.17 is the L_1 -norm ($\lambda \sum_{j=1}^m |\beta_j|$). Accordingly, while ridge regression has closed solution, the solution of lasso is non-linear in y_i and, in particular, it results a quadratic programming problem. However, modern algorithms allow providing an estimate for ridge and lasso in approximately the same amount of time. The major advantage for preferring lasso than ridge in practical applications of feature subset selection is that by the nature of its expression, and considering small constraints t , lasso leads some coefficients to be exactly zero. For a graphical comparison of ridge and lasso regression methods, one can refer to Figure 5.6: By

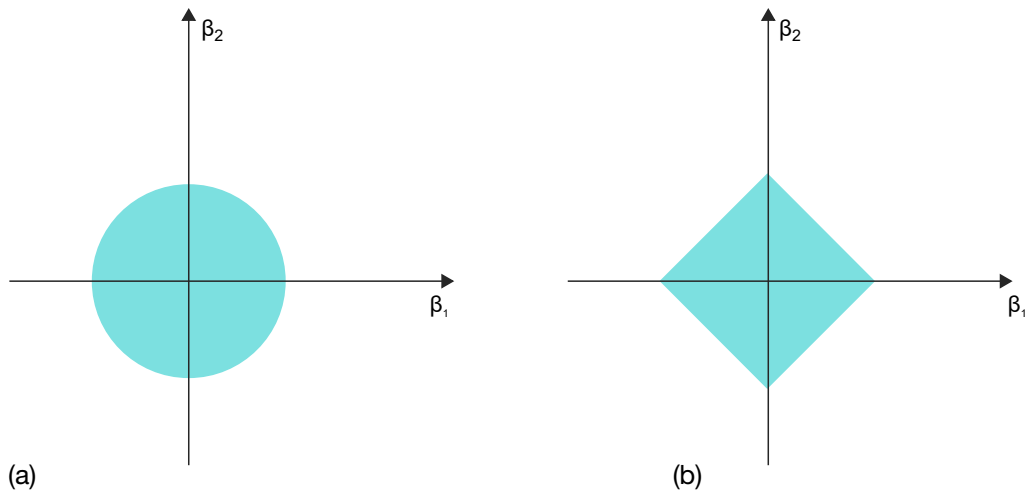


Fig. 5.6: The constraint regions in a bidimensional case of ridge (a) and lasso regression methods (b): a disk and a diamond, respectively. The presence of corners in (b) leads some coefficients to be exactly zero.

assuming the case of two features, in Figure 5.6 (a) it is reported the constraint region for ridge, that is the disk $\beta_1^2 + \beta_2^2 \leq t$, whilst in Figure 5.6 (b), it is shown the constraint region for lasso, that is the diamond $|\beta_1| + |\beta_2| \leq t$. The presence of corners in the constraint region of lasso leads the coefficient β_j to be exactly zero when the solution occurs at that corner. Accordingly, for higher dimensional spaces, the diamond becomes a rhomboid and so on.

Actually, feature selection through shrinkage methods as ridge and lasso can be generalized by employing a generic L_p -norm, like in Equation 5.19 [164]:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^m x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^m |\beta_j|^p \right\} \quad (5.19)$$

for $p \geq 0$. Some representative examples of constraint regions for different values of p are provided in Figure 5.7:

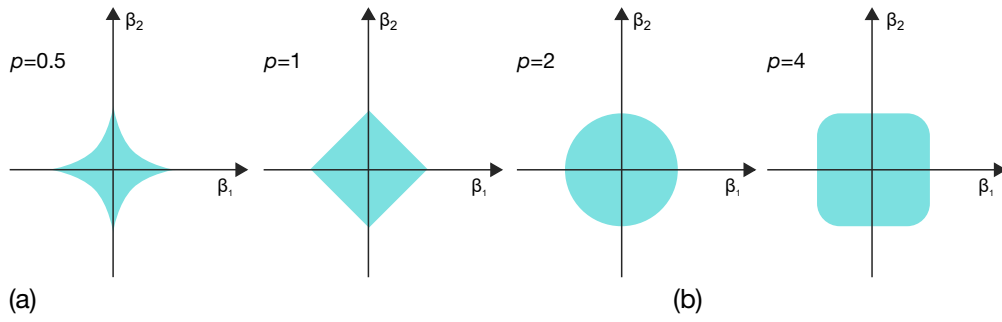


Fig. 5.7: The constraint regions in a bidimensional case for $p=0.5$ (a), $p=1$ (b), $p=2$ (c), and $p=4$ (d), which determines the type of the L_n -norm ($\lambda \sum_{j=1}^m |\beta_j|^p$).

The cases for $p < 1$ have non-convex constraint regions, thus resulting in more difficult solutions. The optimal values should range within $p = [1, 2]$, however, only the case $p = 1$ (i.e. lasso) leads the term $|\beta_j|^p$ to be not differentiable at 0, and so corresponding coefficients to be exactly zero. A compromise between lasso and ridge has been proposed with the *elastic net* method. The penalty term is further modulated through a parameter α which establishes the weight of the L_2 - and L_1 -norm, as in Equation 5.20:

$$\lambda \sum_{j=1}^m (\alpha \beta_j^2 + (1 - \alpha) |\beta_j|) \quad (5.20)$$

The benefit introduced by the elastic net is that it is possible to achieve some coefficients to be exactly zero while removing also highly correlated features [196].

When using the lasso regularization method and partially, also when adopting the modified elastic net version, a certain amount of correlation can still be present in the subset of selected features. This practically means that some of those correlated features can further be removed. This was the case encountered also in an application of the lasso method presented in Chapter 9. Moreover, when working with small dataset, in order to prevent overfitting, it is important to reduce as much as possible the number of features used to train the classifier. To this purpose, in a very simple approach, a correlation analysis can be performed after lasso regularization in order to remove highly correlated features. Then, a further selection may be performed from uncorrelated features, through the multivariate analysis of the discriminative capability of small subsets limited to just two or three features, thus allowing detecting the most powerful features to train the classifier.

5.9 The SVM classifier

The original formulation of the SVM classifier dates to the sixties, and it still now offers a very wide range of applications when working with medical data. Initially, it has been formulated for solving linear classifications problems, but it has also been extended to non linear problems, through the so called kernel trick [197]. In particular, the SVM classifier has been conceived for two linearly separable classes. Hence, let x_i with $i=1,2,\dots,N$ be a feature vector of the entire training set, X , and let the target vector be made of two separable classes, ω_1 and ω_2 . The goal for linearly separating the two classes is finding out the hyperplane, a linear separation surface, expressed by Equation 5.21:

$$g(x) = \mathbf{w}^T \mathbf{x} + w_0 = 0 \quad (5.21)$$

that classify the training data exactly into two different spaces. In particular $w = [w_1, w_2, \dots, w_l]^T$ is called weight vector and w_0 is the threshold. Accordingly, let x_1 and x_2 be two points on the decision hyperplane, they can be substituted into Equation 5.21, thus achieving the relationship in Equation 5.22:

$$\begin{aligned} 0 &= \mathbf{w}^T \mathbf{x}_1 + w_0 = \mathbf{w}^T \mathbf{x}_2 + w_0 \\ \mathbf{w}^T (\mathbf{x}_1 - \mathbf{x}_2) &= 0 \end{aligned} \quad (5.22)$$

Hence, from Equation 5.22, since $(x_1 - x_2)$ lie on the hyperplane, the vector w is necessarily orthogonal to the hyperplane. In particular, Figure 5.8 shows the graphical representation of a decision hyperplane in a bidimensional space, by assuming $w_1 > 0$, $w_2 > 0$, and $w_0 < 0$. In practice, w and w_0 determine direction and intercept of the hyperplane, respectively. Figure 5.8 also highlights d and z , which are respectively expressed by Equations 5.23, 5.24:

$$d = \frac{w_0}{\sqrt{w_1^2 + w_2^2}} \quad (5.23)$$

$$z = \frac{|g(\mathbf{x})|}{\sqrt{w_1^2 + w_2^2}} \quad (5.24)$$

Hence, $|g(\mathbf{x})|$ represents the Euclidean distance of \mathbf{x} from the decision hyperplane, and it takes positive values on one side of the decision boundary, and negative ones on the other side. This is expressed by Equation 5.25:

$$\begin{aligned} \text{If } \mathbf{w}^T \mathbf{x} + w_0 &\geq 0 & \mathbf{x} &\in w_1 \\ \text{If } \mathbf{w}^T \mathbf{x} + w_0 &\leq 0 & \mathbf{x} &\in w_2 \end{aligned} \quad (5.25)$$

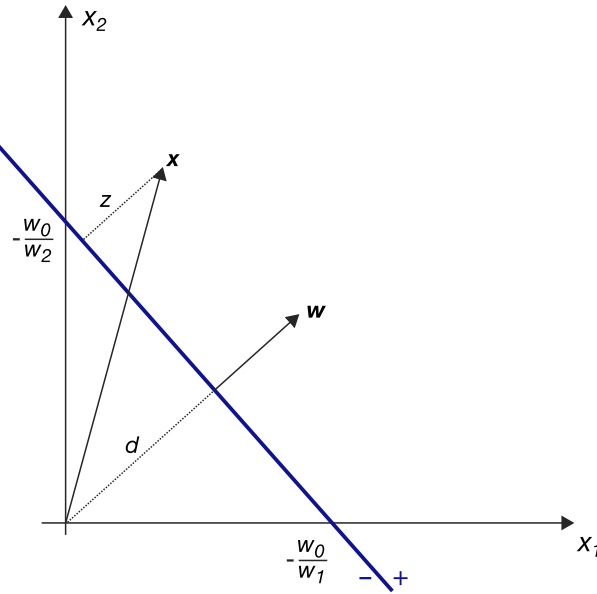


Fig. 5.8: The graphical representation in a bidimensional space of a decision hyperplane, as formulated by Equation 5.21

Actually, the “best” hyperplane refers to the generalization performance of the classifier, that is, not merely a correct classification of training data, but the capability to classify correctly new data. In order to identify the optimal hyperplane through Equation 5.21, the SVM introduces the concept of *margin* of the hyperplane. Accordingly, in order to treat equally the two classes, the searched hyperplane needs to be at the same distance (i.e. margin) from the nearest points belonging to each of the classes. As one can see in Figure 5.9 even by assuming equal margin for the two classes, two different solutions are still possible. One has margin $2z_1$, and the other one has margin $2z_2$. Again, the SVM adopts the criterion of the *maximum margin* to derive the more generalizable decision hyperplane. Mathematically, Equation 5.24 can be scaled in order to have $g(x) = 1$ in the nearest point towards w_1 , and $g(x) = -1$ in the nearest point towards w_2 . This can also be expressed by Equation 5.26

$$\frac{1}{\|\mathbf{w}\|} + \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (5.26)$$

which also requires to fulfil Equation 5.27:

$$\begin{aligned} \mathbf{w}^T \mathbf{x} + w_0 &\geq 1, \quad \forall x \in w_1 \\ \mathbf{w}^T \mathbf{x} + w_0 &\leq -1, \quad \forall x \in w_2 \end{aligned} \quad (5.27)$$

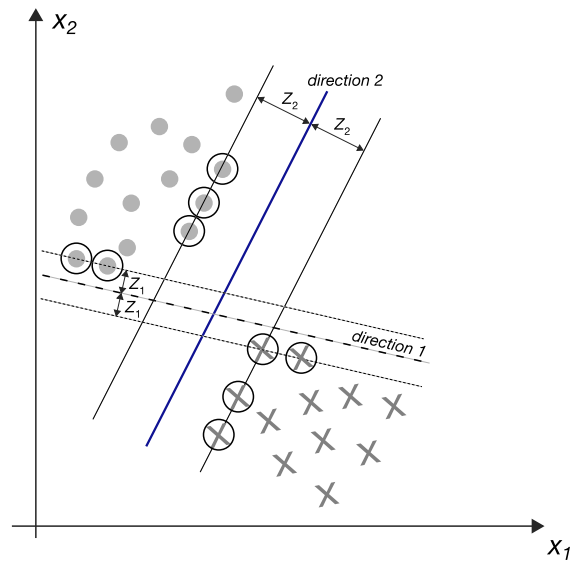


Fig. 5.9: The graph shows two possible decision hyperplane with different margins, $2z_1$ and $2z_2$ corresponding to direction 1 and 2, respectively. The SVM adopts the maximum margin criterion, therefore direction 2 is chosen accordingly as the decision hyperplane.

The problem of finding the maximum margin is solved by minimizing the expression in Equation 5.28:

$$\begin{aligned} \text{minimize } J(\mathbf{w}) &\equiv \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{subject to } &y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1, \quad i = 1, 2, \dots, N \end{aligned} \quad (5.28)$$

By exploiting the Lagrange multipliers, one can achieve Equations 5.29, 5.30:

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i \quad (5.29)$$

$$\sum_{i=1}^N \lambda_i y_i = 0 \quad (5.30)$$

where λ is the vector of Lagrange multipliers. This means that the vector w is given by the linear combination of observations having $\lambda_i \neq 0$, which are called *support vectors* (SVs) and lie on one of the two hyperplanes, defined accordingly by Equation 5.31:

$$\mathbf{w}^T \mathbf{x} + w_0 = \pm 1 \quad (5.31)$$

It is worth noting that SVM classifier only depends on the SVs, that are the most critical points within the training set. This explains the reason why the SVM works very well with few samples. Moreover, the SVM is a convex problem which provides

a global and unique solution.

Of course, the SVM classifier is applicable to non separable classes. In this case, samples of the training set may belong to one of the following three categories, also shown in Figure 5.10

- Vectors out of the decision surface and are correctly classified
- Vectors inside the band and correctly classified (placed in squares in Figure 5.10)
- Vectors inside the band and misclassified (placed in circles in Figure 5.10)

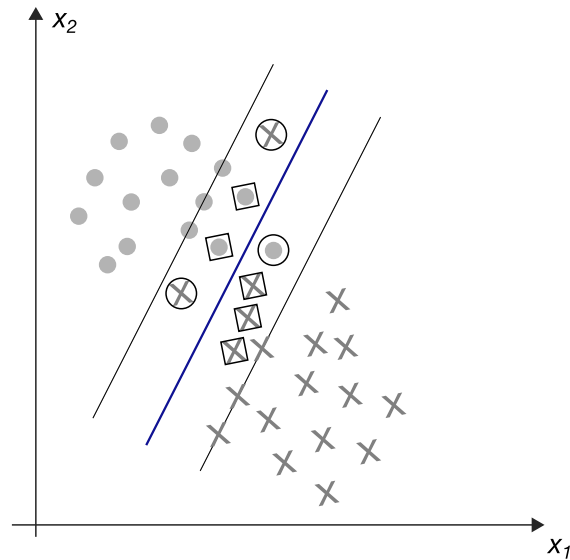


Fig. 5.10: A representative case of non separable classes where some points fall within the decision surface. Among these ones, those points of each class correctly classified are placed within squares, whilst the misclassified samples are placed within circles.

By introducing a so-called “slack variable” ξ_i which allows some points to be within the margin, the three categories correspond respectively to $\xi_i = 0$, $0 < \xi_i \leq 1$, and $\xi_i > 1$. Hence, for non separable classes, the goal of SVM is to have the maximum margin possible while minimizing the points having $\xi_i > 0$. Equivalently, the SVM minimizes the cost function reported in Equation 5.32:

$$J(\mathbf{w}, w_0, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N I(\xi_i) \quad (5.32)$$

where ξ is the vector of ξ_i , and (Equation 5.33)

$$I(\xi_i) = \begin{cases} 1 & \xi_i > 0 \\ 0 & \xi_i = 0 \end{cases} \quad (5.33)$$

The parameter C is referred to as the misclassification cost. Accordingly, the minimization problem becomes (Equation 5.34):

$$\begin{aligned} \text{minimize } & J(\mathbf{w}) \equiv \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{subject to } & y_i(\mathbf{w}^T \mathbf{x} + w_0) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \\ & \xi \geq 0, \quad i = 1, 2, \dots, N \end{aligned} \quad (5.34)$$

By adopting the Lagrangian multipliers, it is possible to demonstrate that the parameter C , in practice, constrains the value of λ_i . This means that in linear separable classes, $C \rightarrow \infty$. The slack variables ξ_i are not directly treated when solving the problem in Equation 5.34, but they are indirectly defined through C .

Non linear case

SVM methodology can also be exploited for non linear cases, by assuming a possible transformation from the input feature space to a k -dimensional space where the non linear classification problem is converted into a linear one (Equation 5.35).

$$\mathbf{x} \in \mathbf{R}^1 \rightarrow \mathbf{y} \in \mathbf{R}^k \quad (5.35)$$

By recalling Equations 5.21 and 5.29, and substituting the latter within the former, one can achieve Equation 5.36 [161]:

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{w}^T \mathbf{x} + w_0 \\ &= \sum_{i=1}^{N_s} \lambda_i y_i \mathbf{x}_i^T \mathbf{x} + w_0 \end{aligned} \quad (5.36)$$

where N_s is the number of SVs. The expression reported in Equation 5.36 allows us to adopt what has been commonly defined as the “kernel trick” for SVM classifiers. Basically, a kernel is defined through the inner product of a mapping function from the feature space to a k -dimensional space, as expressed by Equation 5.37:

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}') \quad (5.37)$$

where, $k(\mathbf{x}, \mathbf{x}')$ is a symmetric function, so that $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$ and $\phi(\mathbf{x})$ is the mapping function of \mathbf{x} , $\mathbf{x} \rightarrow \phi(\mathbf{x})$. Accordingly, the expression in Equation 5.36 can be seen through the adoption of an identity mapping function which defines the linear kernel (Equation 5.38) [164].

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}' \quad (5.38)$$

Very commonly adopted kernels are the polynomials and radial basis functions, which are reported in Equations 5.39,5.40.

Polynomials

$$k(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + 1)^q, \quad q > 0 \quad (5.39)$$

where q is the degree of the polynomial.

Radial basis function

$$k(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{\sigma^2}\right) \quad (5.40)$$

However, they have not been used in our applications.

Finally, when the optimal kernel has been chosen, the minimization problem of SVM becomes the one reported in Equation 5.41, and the decision hyperplane is the one given by Equation 5.42:

$$\begin{aligned} \max_{\lambda} \left(\sum_i \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \right) \\ \text{subject to } 0 \leq \lambda_i \leq C, \quad i = 1, 2, \dots, N \\ \sum_i \lambda_i y_i = 0 \end{aligned} \quad (5.41)$$

$$\text{assign } \mathbf{x} \text{ in } \omega_1(\omega_2) \text{ if } g(\mathbf{x}) = \sum_{i=1}^{N_s} \lambda_i y_i k(\mathbf{x}_i, \mathbf{x}) + w_o > (<) 0 \quad (5.42)$$

5.10 k-fold Cross Validation

CV is a method for validating the performance of a classifier, which allows providing a more accurate estimate of how the model will generalize with unknown samples. Actually, different approaches of CV are used. Herein, explanation is focused on k-fold CV. As previously introduced, in machine learning applications the original dataset is split into training and test sets, as shown in Figure 5.11 (a). Then, the training set is used for model development and the test set for performance assessment. It is worth noting that model development in this case includes both feature selection and tuning hyperparameters of the classifier.

A k-fold partitioning of the original dataset allows estimating the generalization performance of the model on different subsets and accordingly selecting the model with the lowest generalization error. Figure 5.11 (b,c,d) reports three different ways of performing k-fold CV, depending on how the feature selection and model development processes are integrated throughout the whole cycle. In the first case

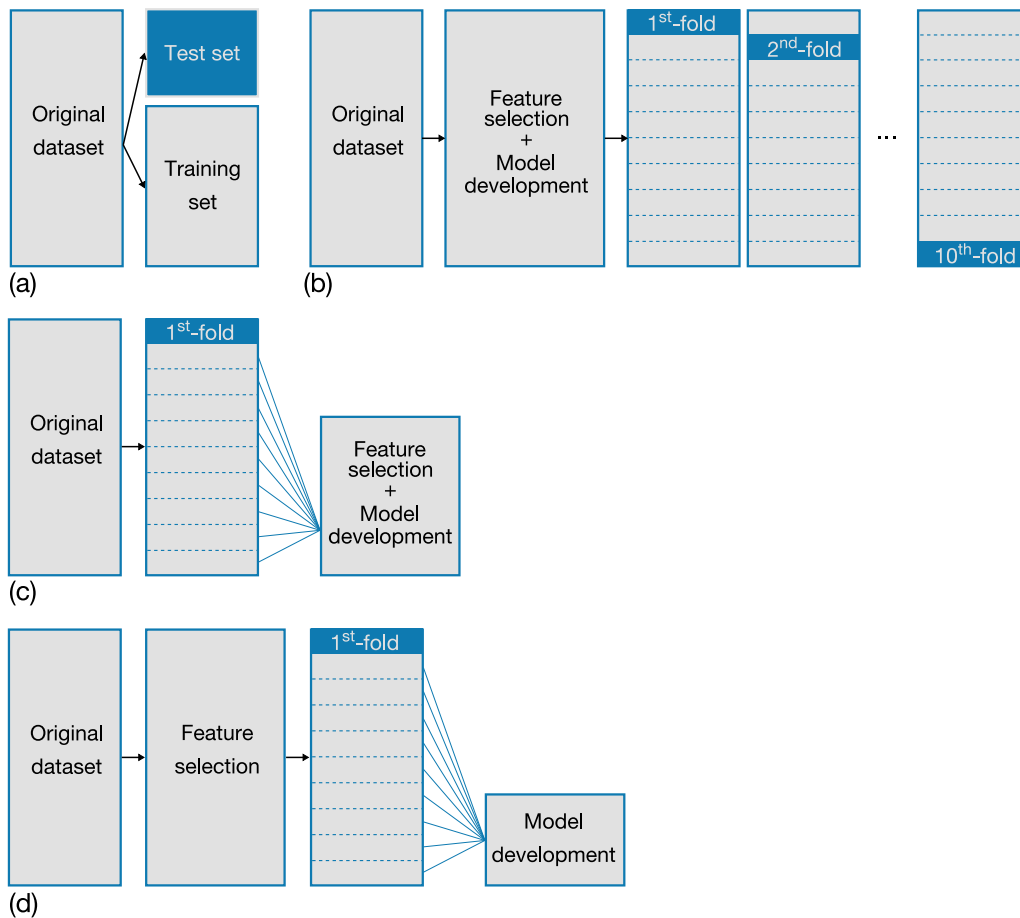


Fig. 5.11: (a) The original dataset is split into training and test set. (b) A k -fold procedure can be performed after training the model on the entire dataset, for estimating the generalization performance of the model itself and selecting the most generalizable model. The dataset is split into k -folds, the model is trained in $(k-1)$ -folds and tested on the remaining one. (c,d) Two alternative approaches of repeated CV are presented: (c) feature selection and model development are performed on $(k-1)$ -folds, (d) feature selection is performed on the entire dataset and model development on $(k-1)$ -folds.

(Figure 5.11 (b)), the model is developed on the original entire dataset, which is later partitioned into k -folds. Each of the k -folds is iteratively employed for validation. Common values of k are three, five, and ten, generally chosen based on the size of the original dataset and the balance between positive and negative samples. In fact, when splitting the dataset into k -folds one should avoid to make unbalance datasets between positive and negative classes, for excluding any selection bias on the model performance. Moreover, with k -fold CV, at the end all samples from the wider dataset have been employed for both training and validation. To overcome this issue, an alternative approach is the repeated CV, graphically represented in Figure 5.11 (c). The original dataset is partitioned into k -folds before any model

development. Accordingly, k -runs of training are performed, and for each run, the model is validated on one fold and trained on the $(k-1)$ -folds. An additional way of performing k -fold CV is also that shown in Figure 5.11 (d). In this case, the phases of feature selection and classifier setup are separated. That is, feature selection is performed as the first step, to speed up the computational time required by each run. Then, for each run the model is trained (with hyperparameter tuning) on the $(k-1)$ -folds and validated on the k^{th} fold.

Once performed k -fold CV, it is very common to estimate the generalization capability of the final model by averaging the performance achieved on each validation set. Alternatively, in an application-specific approach, the final model can be selected by adopting a specific selection criterion based on the metrics computed for assessing the performance of the models. Actually, numerous methods for aggregating data after performing CV also exist, albeit not reported in this Thesis.

In most of practical applications, CV is repeated several times, in order to remove any dependence from the initial selection of training and test sets [198].

5.11 The challenge: standardization and reproducibility

Non-standardized procedures and non-reproducible results are the main issues hindering radiomics from being a validated clinical practice. In this regard, many factors jeopardize the repeatability of radiomic feature values and classification outcomes, thus hampering radiologists' confidence in radiomic usefulness, and delaying its translation into clinics [199]. Nowadays, there is a broad awareness of the need to reduce variability of results, although solution seems to be far yet.

In a whole radiomic workflow, innumerable factors are sources of variability of results. Actually, one should distinguish between reproducibility of radiomic features in terms of acquisition protocols, technical and computational aspects, and reproducibility of features in classification, which can be pursued by preventing overfitting and influence of outliers on training.

Some radiomic studies perform a dedicated analysis on reproducibility of radiomic features, using phantoms to explore the effects of variable acquisition parameters, such as tube current in CT imaging [170], or voxel size [200]. Other radiomic studies assess the effects of varying segmentations on first or second order texture features, for instance, in non-small cell lung cancer [181],[201] and rectal cancer[202]. Recently, the Image Biomarker Standardization Initiative (IBSI) [183] has involved several researchers and radiologists from different countries to define global guidelines for standardization of radiomics, based on the idea that the standardization will increase as well as reproducibility of radiomic features, if a pre-defined methodology will be shared and applied across different centres and stud-

ies. Hence, IBSI provides practical recommendations and a tentative standardized framework. In agreement, Radiomics Ontology, accessible at the NCBO BioPortal (<https://bioportal.bioontology.org/ontologies/R0>) provides a semantic framework for radiomic features sharing the nomenclature proposed by IBSI. Moreover, the NRG Oncology investigators proposed the radiomics quality score (RQS) to evaluate the quality of radiomic studies [203].

The development of fully automated frameworks for radiomic classifiers [204] as well as the availability of many software packages of public domain or commercial provided by different vendors for radiomic studies have been considered a marked step forward towards the standardization of processes, but actually, on several occasions, software packages have shown a great variability of measurements [192]. Moreover, the recent increase of high-performance computing resources in entry-level workstations and the growth of automatic tools for radiomic analyses, has made them popular and accessible to research groups with different expertises, this incredibly increasing the variability of results. In addition, less experienced researchers in quantitative medical images use radiomic software packages as black-boxes, without full awareness of data processing choices. Accordingly, it is hard to believe that this approach could be the effective solution for the standardization issue. Referring to variability of radiomic features, the authors in [205] say: *“to unravel this knot, one needs to start from the end and first standardize the software framework used to extract radiomic features”*. Actually, a framework is made by a series of consequent analytical choices that cannot perform at the same way on different inputs or conditions. On the contrary, applying the same scheduled analytical framework on different input data may lead to the opposite effects of increasing non-reproducibility of radiomic features. Variability of input data, strongly depending on acquisition conditions, parameters, scanners, and patients should be faced with a dedicated specific analysis aiming at applying the most proper solution chosen in a case-by-case approach. The reliable employment of software for radiomic analysis should be postponed, in order to favour primarily the methodological validation of procedures, and only later on to let them be widely accessible thanks to commercialized software. Otherwise, the risk is not only to delay the clinical translation of radiomics, but even to waste this great opportunity to improve and benefit the clinical decision making with the quantitative radiomic contribution.

Chapter 6

A methodological approach for MS and DV reproducibility

This Chapter presents my methodological contributions in CTP analysis, aimed at pursuing the reproducibility of MS and DV in hepatic CTP, by assuming the *common agreement* principle. In particular, this work refers to PIXEL, a multicentre study on liver perfusion, presented in Sect. 6.1. Accordingly, the preliminary methodological choices for analysing hepatic CTP examinations are introduced in Sect. 6.2, including the selection of a monocompartmental model for analysing the first pass kinetics of CA and data preparation procedures. On the one hand, this Chapter proposes novel methods for: (i) extracting the end of the first passage of the CA from tissue TACs (Sect. 6.3), (ii) modelling the first passage through a fitting procedure (Sect. 6.4), (iii) computing perfusion parameters through MS and DV, independently on the same signals, after exploiting an iterative procedure based on a deconvolution process to separate arterial and portal contributions from tissue TACs (Sect. 6.5). On the other hand, this Chapter questions the validity of the MS assumptions which may lead MS to represent a different system's status with respect to DV, and proposes a reformulation of the MS method, the gMS, to allow temporal and modelling equivalence of MS and DV (Sect. 6.6). Then, the criteria for investigating the reproducibility of measurements at voxel level are detailed in Sect. 6.7. In order to accomplish CTP analysis, I devised a software for handling the whole process of the scientific study. The main goal of software design was aimed at keeping data consistency. This has been done by embedding, in the output structures (matrices, figures, etc.), tags referring to the input and the code release that generated it. In this regard, details are presented in Appendix.

6.1 Case of study: PIXEL multicentre study

The difficulties of CTP and, more generally, of DCE imaging modalities to enter the clinical practice are also due to the limited number of multicentre studies at the disposal, through which validating IBs against the variability of scanners, acquisition protocols, expertise of technicians and operators [206].

As mentioned in Sect. 4.6, three wide CTP multicentre studies exist. During my PhD course, I have had the opportunity to work on data belonging at each of them. Mainly, I have worked on data from the PIXEL (Perfusion Index: Evaluation for Liver metastases) project and the methodology for reproducibility of MS and DV has been developed and tested primarily on these data. For this reason, I provide first an overview of aims and characteristics of the PIXEL study. In addition, I have applied the methods developed for pursuing reproducibility of MS and DV also on data belonging to PROSPeCT and SARAH projects. Accordingly, these two multicentre studies will be briefly introduced in Chapter 7, before presenting the experimental results of some specific applications.

The availability of PIXEL data has been possible thanks to the collaboration between the CVG and Prof. V. Vilgrain who has directed the PIXEL project. In particular, PIXEL is a French multicentre study – initially enrolling around four-hundred patients in nineteen national Centres – aiming at assessing the predictive role of the HPI for the onset of hepatic metastases in patients with non-metastatic colorectal cancer (CRC) diagnosis, before the administration of any treatment. Secondary aims of the study are: (i) identifying a cut-off HPI value for predicting the onset of hepatic metastases within three years from CRC diagnosis, and (ii) assessing the role of other perfusion parameters and measuring the inter-reader variability of perfusion parameters.

PIXEL inclusion criteria can be summarized as follows:

- age ≥ 18 years old
- no previous cancer
- non metastatic CRC
- providing written informed consent
- patient followed by one of the involved Centre

In addition, non-inclusion criteria are:

- presence of hepatic metastases at CRC diagnosis
- chronic hepatopathy
- administration of chemotherapy after execution of CTP examination
- surgery of CRC after CTP after execution of CTP examination
- allergic reaction to iodine CA
- renal failure hampering the administration of iodine CA

- pregnant patient

CTp imaging protocol has required:

- patients to breath slowly during the whole acquisition time
- administration of 40 ml of iodinated CA contemporaneously with acquisitions, with a concentration of 350 mgI/ml, followed by 20 ml of saline solution
- acquisition of images every 1 s for the first 30 s, and 3 s after, up to 2 min
- CT tube current and voltage at 100 mA and 80 kV, respectively (1 s rotation time, 100 mAs exposure)

Patients were enrolled since 2008 to 2011 and all were monitored for three years from inclusion. Actually, fourteen Centres and 315 patients have been definitely included. Table 6.1 resumes the included Centres, CT scanner employed in each Centre, and number of enrolled patients.

For each patient, CTp image sequences, collected according to the Digital Imaging and Communication in Medicine (DICOM) standard, have been carefully analysed in order to retrieve all needed information for perfusion analysis. In particular, the availability of acquisition time instant of each CTp is mandatory for extracting TCCs from images. Accordingly, CTp examinations where the acquisition time was not available have been excluded. In addition, the enhancement of the portal vein in several cases is not clearly detectable in CTp images, this preventing the outlining of the ROI of the portal vein and, consequently, the computation of dual input hepatic perfusion. For these reasons, some CTp examinations have been excluded. Table 6.2 resumes the relevant information of patients selected for perfusion analysis from each Centre, also highlighting the number of patients who developed metastases (column “Met.”) within three years from CRC diagnosis and those who have not (column “Not-met.”).

As one can see from Table 6.2, the number of patients who have developed metastases within three years from CRC diagnosis is really small. This has been the major limitation hampering the detection and validation of an IB predictive of the onset of hepatic metastases. Moreover, it is also worth noting that the PIXEL study has been probably the first multicentre CTp study. Accordingly, a series of technical issues occurred when collecting data, which forced in many cases a subsequent selection of patients during data processing, this making the proper comparison of results of the different Centres very challenging if not, in some cases, impossible. For instance, some Centres adopted a different image acquisition protocol or a different protocol for CA injection. In some cases, as it happens in Centre 16, the sampling frequency varies also within the Centre itself.

To the purpose of this Thesis, data from PIXEL study have been analysed to investigate methodological aspects of perfusion computing methods, and the comparison between Centres was just a secondary aim. Therefore, among the whole

Centres at the disposal, the most populous Centres were chosen and considered for the reproducibility study, that is Centres 1 and 16, by selecting from Centre 16 those patients sharing the same CTp protocol as Centre 1.

Table 6.1: PIXEL Centres and enrolled patients

Centre (C)	Centre name	CT scanner	Patients
1	Beaujon	GE Lightspeed VCT	71
2	Hegp Broussais	GE Lightspeed VCT	32
3	Ambroise Paré	Philips MX8000 IDT 16	18
6	Henri Mondor A. Chenevier	GE Lightspeed VCT	1
7	Pitié Salpêtrière	Philips Brilliance 64	7
8	Chu Nantes	GE Lightspeed VCT	20
9	Chru Angers	Philips MX8000 IDT 16	44
10	Haut-Lévêque Bordeaux	Siemens Definition 64	13
12	Claude Huriez	Philips Brilliance 40	11
15	Institut Gustave Roussy Villejuif	GE Lightspeed VCT	23
16	Chu Amiens	GE Lightspeed Pro 32	49
17	Institut mutualiste Montsouris	GE Discovery CT 750 HD	20
18	Chu Caen Hôpital Cte De Nacre	Philips Brilliance 40	2
19	Cabinet d'hépatologie et de gastro entérologie	Philips MX8000 IDT 16	4

Table 6.2: Patients of PIXEL selected for perfusion analysis

Centre (C)	Met.	Not-met.	Total patients
1	6	65	71
2	3	29	32
3	1	17	18
6	0	1	1
7	0	7	7
8	2	18	20
9	4	40	44
10	0	13	13
12	1	10	11
15	2	22	24
16	5	42	47
17	1	19	20
18	1	1	2
19	2	2	4
TOTAL	28	286	314

6.2 CTp study of liver through MS and DV

By facing the CTp reproducibility issue, the simplest operating conditions have been considered, referring to the choice of the compartmental model, the kinetic phase of the CA, and the computational methods for estimating perfusion parameters. Therefore, this Thesis adopts a monocompartmental model and focuses on the first pass of the CA, studied through MS and DV methods, being the former the most simple and robust to compute and the latter the most precise one [207]. Accordingly, this reproducibility study addresses the computation of BF values, the only perfusion parameter computable with the MS method and even the most common parameter considered to early detect tumour changes in diverse anatomic districts [91].

By considering the study of the dual input liver perfusion, in Chapter 4, I have presented some open issues related to MS and DV methods. Here, I briefly recap these issues in Table 6.3 and all of them are addressed throughout this Chapter.

Moreover, before deepening the reproducibility issues, the stages of model selection and data preparation are highlighted.

6.2.1 Model selection

Figure 6.1 reports the dual input monocompartmental model adopted for perfusion analysis. By recalling what previous described, the input of the system is the

Table 6.3: Summary of the open issues of MS and DV methods applied to the hepatic perfusion study

Method	Open issues
MS	<ul style="list-style-type: none"> • Approximate separation of $C_{TA}(t)$ and $C_{TP}(t)$ by the spleen's peak • Presence of delays between input TAC's peak and tissue maximum slope
DV	<ul style="list-style-type: none"> • Extraction of TAC referred to CA first pass only • Solving deconvolution with two unknown parameters

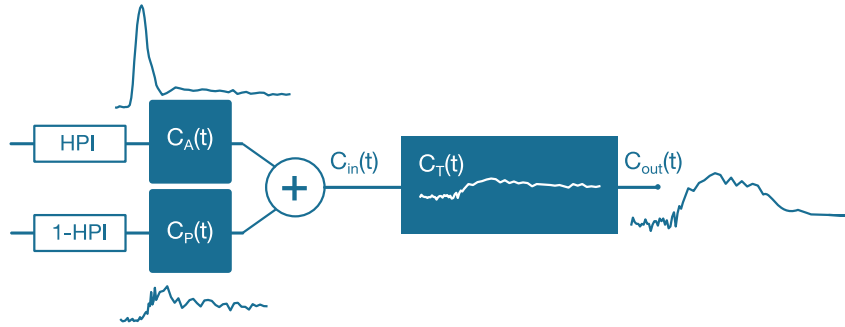


Fig. 6.1: Hepatic dual-input one compartmental model

linear combination of $C_A(t)$ and $C_P(t)$ weighted by the HPI (Equation 6.1).

$$C_{in}(t) = \text{HPI} \cdot C_A(t) + (1 - \text{HPI}) \cdot C_P(t) \quad (6.1)$$

CTp sequences allow extracting voxel-based signals, $C_T(t)$, which are practically given by the sum of the two contributions $C_{TA}(t)$ and $C_{TP}(t)$, due to $C_A(t)$ and $C_P(t)$, respectively, as formalized in Equation 6.2.

$$C_T(t) = C_{TA}(t) + C_{TP}(t) \quad (6.2)$$

Accordingly, the major issues regard the computation of the HPI and the extraction of the two tissue contributions from the whole tissue signals.

6.2.2 Data preparation

First, the analysis of CTp examinations requires ROI segmentation and image preprocessing in order to extract the TCCs from the entire dynamic sequence. Hence, in each CTp examination, a central representative slice is selected and two ROIs are first drawn, on the aorta and the liver, respectively. The procedure of ROI placement on liver is carried out manually, by an expert radiologist, with a great care, excluding large vessels, such as portal vein or hepatic artery. Then, a ROI

outlines the portal vein, and it is aligned over time, on each sampling instant, to compensate for motion in the subsequent CTP scans [208]. Actually, the variability in ROI outlining does not affect the findings of the voxel-based analysis performed, this making a second segmentation, commonly required for studying reproducibility on lesion analysis, useless. In addition, a denoising procedure is performed on CTP images, by referring to the portions of image highlighted with the vascular and tissue ROIs. Hence, a median 2D filter with $[5 \times 5]$ pixel size is adopted for aortic and portal vein ROIs, from which one mean TCC ($C_A(t)$ and $C_P(t)$, respectively) is achieved. Instead, a median 3D filter with $[9 \times 9 \times 3]$ voxel size is chosen for tissue ROIs from which single voxel-based signals are derived ($C_T(t)$), after excluding voxels undergoing dynamic artefacts [93]. Since vascular TCCs, which are averaged ROI signals, present limited noise in the unenhanced portion, the baseline value is computed for them by averaging the first five samples. Instead, by adopting a voxel-based approach for tissue, baseline values are computed for each voxel-based TCC according to the method presented in [209]. After baseline removal from vascular and liver TCCs, the corresponding TACs are achieved. Hereinafter, $C_A(t)$, $C_P(t)$, $C_T(t)$ refer to the TACs.

Finally, since CTP acquisition protocols generally adopt higher sampling frequency during the first pass than in the CA recirculating phase, a preliminary non-parametric fitting of the real TACs is performed to up-sample signals, so as to work on a uniform sampling frequency of 1 Hz.

6.3 Extracting the first pass signal

Focussing on first-pass kinetics, it is needed to extract from the vascular and tissue TACs the contribution due to the first passage only. As far as the hepatic monocompartmental model is concerned, the CA vascular kinetics can be reasonably applied on tissue haemodynamics, yet more in the absence of altered vascular pathways (e.g., due to angiogenesis or diseases), yielding a CA recirculation flooding the tissue after the maximum CA concentration is reached. Therefore, the main problem is to find out enough TAC samples expectedly belonging to the first-pass phase, that could be successively used in a parametric fitting model to extract a complete first-pass signal from the real tissue TAC. In practice, while the left bound of the interval is known (i.e., the first acquired sample t_0), what lacks is the last first-pass interval sample.

Based on the considerations above, I have found reasonable choosing the time instant halfway between the peak time (t_p) and the washout time (t_w), when the outflow is maximum. Hence, I have first performed a signal denoising for each patient through a smoothing spline [210], computed over all the acquired n samples,

setting the smoothing parameter $\lambda = 0.7$ [211], also on the basis of preliminary tests. Then I have computed the derivative on these smoothed signals, thus achieving t_p (when it has 0-value) and t_w (when it has its minimum). It is worth noting that if different values of λ can affect the goodness of fit, the effects on t_p and t_w are almost negligible. Finally, for each TAC $[t_0, (t_p + t_w)/2]$ represents the time interval whose samples are considered as belonging to the first-pass phase.

Figure 6.2 (a) shows a real TAC with a superimposed smoothing spline, where

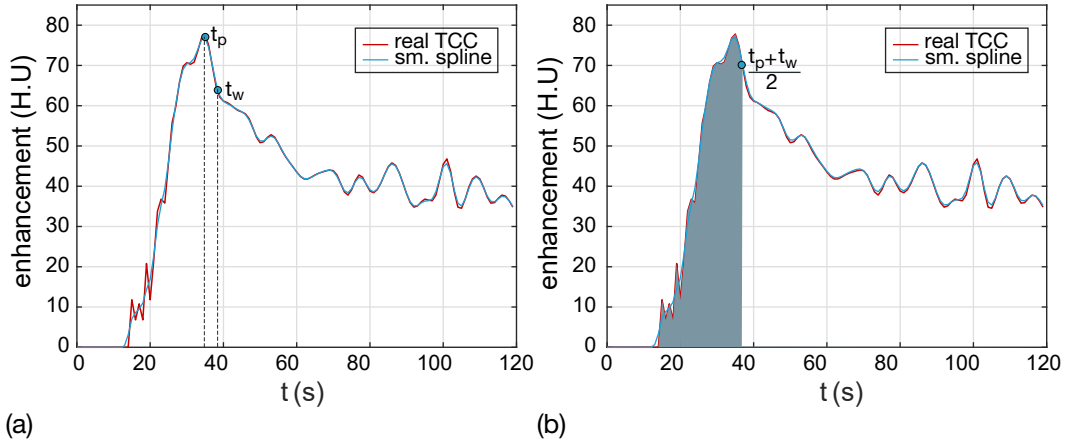


Fig. 6.2: The last time instant of the first pass phase of CA is assumed halfway between t_p and t_w , when the peak value and the minimum derivative of the TCC occur (a). Hence, the first pass signal is extracted from the samples included within the interval $[t_p, \frac{t_p+t_w}{2}]$ highlighted in (b).

the time instants t_p and t_w are highlighted, when the peak value and the minimum derivative occur, respectively. Hence, in Figure 6.2 (b), by assuming the end of the first pass halfway between t_p and t_w , the portion of the tissue TAC due to the CA first pass is emphasized by the coloured area.

6.4 Modelling the first pass through signal fitting

A fitting procedure is performed against the real TACs between $[t_p, \frac{t_p+t_w}{2}]$ to extract and model the signals referred to the first passage of CA. In particular, vascular and tissue signals have been fitted by adopting two widely used parametric models, the LN and GV functions, respectively. The fitting procedure has been performed using the Interior Point [212], a constrained nonlinear optimization algorithm implemented in the *fmincon* function of Matlab[®] (R2018b v.9.5, The MathWorks, Natick, MA, USA).

According to the theoretical description provided in Chapter 4, Sect. 4.4.1, LN model is expressed through four parameters, AUC, t_m , RD , and s . An initial es-

timate of the two parameters AUC and t_m has been achieved for each patient by considering the maximum value of the signal multiplied by ten and the time instant in which the maximum value occurs, as reported in Eqs. 6.3 and 6.4.

$$AUC = TAC|_{max} \cdot 10 \quad (6.3)$$

$$t_m = t|_{TAC_{max}} \quad (6.4)$$

In particular, the multiplication by ten in Equation 6.3 accounts for a minimum dispersion of the vascular TACs. In fact, non-instantaneous injections cause dispersed TACs, whose maximum value underestimates the area under the curve of the impulse function, which would represent an instantaneous injection. Instead, for RD and s , patient-independent initial estimates have been used. A preliminary analysis based on the shape of arterial and portal vein signals allowed detecting the most likely couple of these two parameters, to represent a tall, narrow, and slightly right-skewed aortic signal, and a short, large, and strongly right-skewed portal vein signal. In this regard, Figure 6.3 shows the variations of LN functions for five dif-

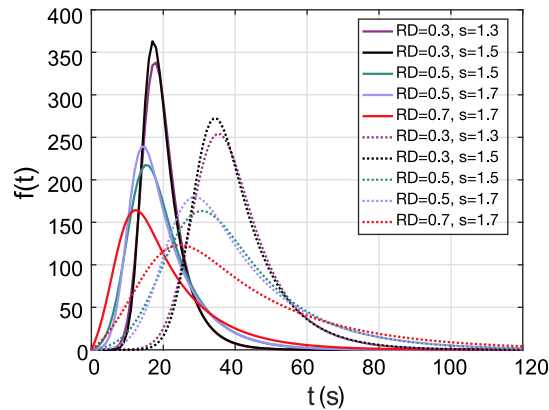


Fig. 6.3: LN models for five different values of the couple RD and s , with AUC and t_m kept fixed at $AUC=4000$ and $t_m=20$ for the aortic signal and $AUC=6000$ and $t_m=40$ for the portal vein, respectively.

ferent values of the couple RD and s , with AUC and t_m kept fixed. In particular, $AUC=4000$ and $t_m=20$ for the artery, whilst $AUC=6000$ and $t_m=40$ for the portal vein. Then, based on the different LN shapes achieved, $RD=0.3$ and $s=1.5$ have been chosen as initial estimate of the minimization procedure for fitting the aorta, whilst $RD=0.7$ and $s=1.7$ have been used for the portal vein.

As introduced in Chapter 4, Sect. 4.4.3, GV model has four main parameters, that is the global scale K , shape and scale factors, α and β , respectively, and the location time t_0 . Patient-independent initial estimates have been fed to the minimization algorithm for all the four parameters. A preliminary analysis, based on

the shape of tissue signals, maximum enhancement values and initial time instants of tissue enhancements, allows choosing as the initial estimates the values $K=50$, $\alpha=0.5$, $\beta=20$, and $t_0=25$ s. In particular, the choice of K refers to the average of maximum enhancement values achieved on tissue signals. Accordingly, the value of β has been increased by one order magnitude with respect to the β values generally used when GV model describes a probability density function (as in Figure 4.16). Then, the small α value has been chosen to allow the complete decrease of the curve towards the baseline value within the acquisition time interval. The estimate of t_0 corresponds to t_m value of LN model observed, on average, on the aortic vascular input, considered as the minimum required for having the tissue enhancement. For LN and GV parameters, all possible positive values have been explored by the minimization algorithm. Finally, Figure 6.4 shows the signals extracted from $C_A(t)$,

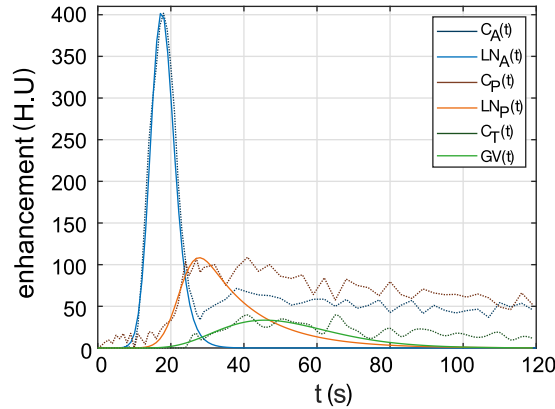


Fig. 6.4: LN and GV fitting of $C_A(t)$, $C_P(t)$, and $C_T(t)$ referring to the first pass signals.

$C_P(t)$, and a representative $C_T(t)$, and their first pass fitting signals superimposed, $LN_A(t)$, $LN_P(t)$, and $GV(t)$, respectively. Hereinafter, $C_A(t)$, $C_P(t)$, and $C_T(t)$ refer to the first pass fitting signals.

6.5 Computation of perfusion parameters

If one worked with a single input model, extracting the first pass signals would be enough to compute BF values referred only to first pass through DV, and compare them with those arising from MS. Instead, in the dual input hepatic model, both MS and DV require $C_{TA}(t)$ and $C_{TP}(t)$ to compute aBF and pBF, respectively. Hereinafter, aBF_x, pBF_x, and BF_x stand for the parameters computed with MS or DV depending on x=MS or x=DV.

In order to avoid approximating the two contributions according to Eqs. 4.22 and 4.23, I have decided to compute BF_{MS} analytically, employing the same signals

as DV, this also expectedly improving reproducibility. From a mathematical point of view, the two contributions $C_{T_A}(t)$ and $C_{T_P}(t)$ arise from Eqs. 4.68 and 4.69. Accordingly, in order to compute $C_{T_A}(t)$ and $C_{T_P}(t)$, both HPI and $R^*(t)$ are needed. In fact, one of the open issues mentioned for DV (Table. 6.3) is how to solve the procedure having two unknown parameters.

To this purpose, I have implemented a two-stage procedure, made of an initialization (Init) and a computation (Compute) block, outlined in Figure 6.5. The

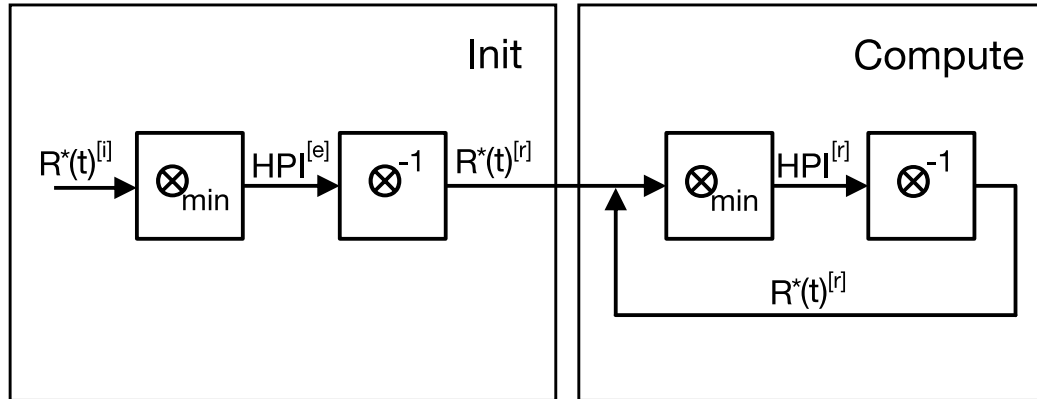


Fig. 6.5: Two-stage procedure made of initialization (Init) and computation (Compute) blocks. Initially, $\text{HPI}^{[e]}$ is estimated through convolution (\otimes_{\min}), stemming from an ideal model of $R^*(t)$ (i.e. $R^*(t)^{[i]}$), minimized against $C_T(t)$. Then, by exploiting $\text{HPI}^{[e]}$, $R^*(t)^{[r]}$ is first achieved via deconvolution (\otimes^{-1}), then fed to the Compute block, where the estimates of $\text{HPI}^{[r]}$ and $R^*(t)^{[r]}$ are iteratively refined until convergence (i.e., until the mean residuals computed between two subsequent estimates of $R^*(t)^{[r]}$ reach a plateau).

first, sequential, block aims at providing a very preliminary estimate of HPI and $R^*(t)$, used to initialize the second block, whose purpose is iteratively refining HPI and $R^*(t)$, until convergence is reached. In particular, I have started in the first block by estimating the voxel-based HPI values ($\text{HPI}^{[e]}$) *via* convolution (\otimes_{\min}), from Equation 4.67, which is directly minimized against $C_T(t)$ using an *ideal* model of $R^*(t)$ (i.e., $R^*(t)^{[i]}$). Then, $\text{HPI}^{[e]}$ is employed to deconvolve Equation 4.67, now achieving voxel-based estimations of $R^*(t)$ on *real* data (i.e., $R^*(t)^{[r]}$), which is fed to the second stage to achieve an early estimate of the *real* HPI (i.e., $\text{HPI}^{[r]}$). The subsequent refinements of $R^*(t)^{[r]}$ and $\text{HPI}^{[r]}$ are iteratively performed minimizing the mean residuals of $R^*(t)^{[r]}$ ($\mu[R^*(t)^{[r]}]$) computed between two subsequent estimates. The mean curve computed over all the included patients (Figure 6.6 (a)) shows a **L**-like curve shape, similar in each patient, with a plateau for $\mu[R^*(t)^{[r]}]$ starting at the light blue point, the fourth iteration (that is, referring to differences between estimates at $i=4$ and $i=3$). As one can infer by the very low standard deviations, this occurs for all patients and because there are no real benefits to wait for convergence, I have chosen to stop the process and taking $R^*(t)^{[r]}$ at $i=3$. This choice has been

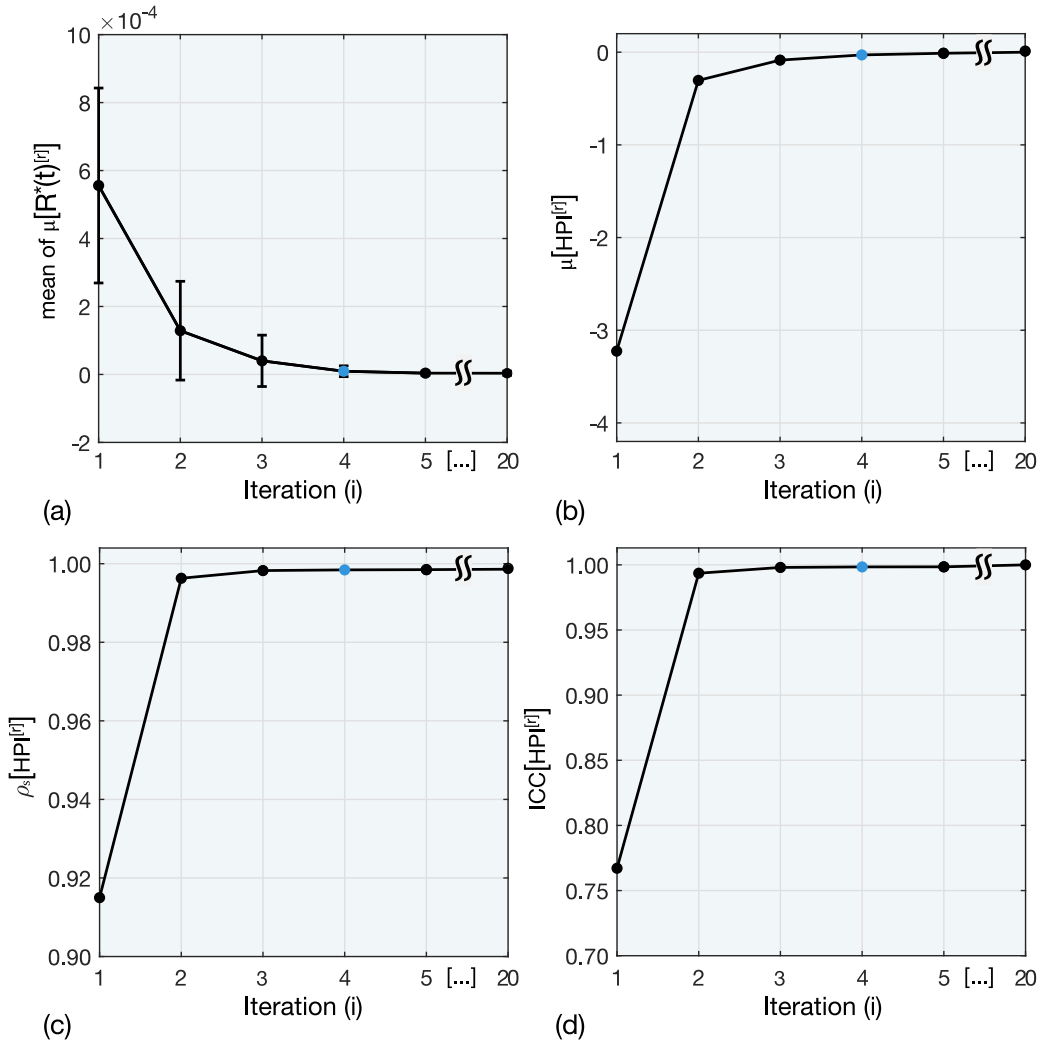


Fig. 6.6: (a) The mean curve of $\mu[R^*(t)]^{[r]}$ referring to whole patients, with the light blue point highlighting the iteration ($i = 4$) at which a plateau starts; for a sample patient (ID37, Centre 1) (b) $\mu[HPI]^{[r]}$, (c) $\rho_s[HPI]^{[r]}$ and (d) $ICC[HPI]^{[r]}$ are reported.

also supported by the concomitant best $HPI^{[r]}$, as one can see for a representative sample patient (ID37, Centre 1) in Figure 6.6 (b), reporting the evolution of the mean residuals, and by the Spearman coefficient (ρ_s) and the Intraclass Correlation Coefficient (ICC) in Figure 6.6 (c,d), respectively.

Finally, by deconvolving $C_{T_A}(t)$ and $C_{T_P}(t)$ with the corresponding input functions, $C_A(t)$ and $C_P(t)$ respectively, we can compute voxel-based aBF_{DV} and pBF_{DV} values, subsequently summed up to yield the total BF_{DV} value (Equation 4.20). Similarly, MS is independently applied to $C_{T_A}(t)$ and $C_{T_P}(t)$ to compute aBF_{MS} and pBF_{MS} *via* central finite-differences, then summed up to achieve BF_{MS} .

6.6 The *generalized* MS

Once computed the two tissue contributions $C_{TA}(t)$ and $C_{TP}(t)$, MS and DV reproducibility has been further investigated by analysing the modelling assumptions of the two methods, in order to detect any theoretical inconsistency hampering reproducibility.

In particular, I have focused on the MS theoretical formulation which is grounded on a strong modelling assumption. In this regard, it is worth noting that the first experimental measurement of tissue perfusion by dynamic CT is reported in the study by Miles [213] on cardiac chambers. In that case, in addition to the hypothesis of no-outflow, the experiment implicitly relies on being no dispersion between the input and the organ, since the radiolabelled microspheres are injected directly into the left ventricle. Hence, the latter hypothesis implies that the maximum gradient of the time-activity curve of the tissue occurs at the same time of the maximum value of the input curve.

In a real case, there are two possible factors deviating from theory: (a) a minimum venous outflow and (b) the dispersion between CA concentration in the aorta and that one in the vessel directly supplying the organ, that is often a closer small artery. In particular, depending on dispersion, the true vascular input could have a slightly lower peak's amplitude than aorta and its peak may be delayed. Hence, the second condition implies that the peak of the aortic input may occur before the maximum gradient of the tissue CA enhancement.

Therefore, I have hypothesised that the denominator in the classical formulation of MS, which I recall in Equation 6.5:

$$aBF \approx \frac{\left. \frac{dC_{TA}(t)}{dt} \right|_{max}}{C_A(t)|_{max}} \quad (6.5)$$

could not be necessarily the peak of the aortic function. In fact, if the peak time does not coincide with the time instant of the tissue maximum slope, Equation 6.5 will provide us with an underestimated aBF value. Hence, I have thought that each value of $C_A(t)$ could be plausible in order to attain a more accurate aBF estimate through the MS. Then, I have relaxed the constraint in Equation 6.5, by replacing its denominator with any time-dependent sample of $C_A(t)$, thus attaining the *generalized* MS, gMS, of Equation 6.6:

$$aBF \approx \frac{\left. \frac{dC_{TA}(t)}{dt} \right|_{max}}{C_A(t)|_{t \in [t_a, t_{max}]}} \quad (6.6)$$

where t_a represents the beginning of the ascent phase of $C_A(t)$ and t_{max} refers to its peak's time-instant.

Finally, the voxel-based gMS is computed on each $C_{T_A}(t)$.

6.7 Reproducibility assessment

Reproducibility of MS and DV values has been assessed in two stages, being both based on voxel-based comparisons.

As explained before, the methods developed have allowed me to separate the two contributions, $C_{T_A}(t)$ and $C_{T_P}(t)$, and compute aBF and pBF values on the same signals separately through MS and DV, this expecting to increase precision of measurements and reducing the variance accordingly. In addition, this expected to improve the concordance of clinical evaluations arising from the two methods. First, BF values through MS and DV have been compared to assess the agreement of measurements, assessed through the linear correlation of BF values, which is expected to yield strong and lowly dispersed correlations, although with numerical differences.

Second, because of the hypothesis formulated in Sect. 6.6 (MS cannot represent the same status of the system with respect to DV, due to a non-negligible transport delay between input and tissue) I have assessed whether and to what extent gMS, an extended formulation of MS accounting for transport delays, could represent the same status of DV. Accordingly, by employing gMS and DV, I assessed the numerical equivalence of perfusion parameters, referring to aBF values. For each voxel-based $C_{T_A}(t)$, aBF values computed while varying the denominator in Equation 6.6 have been compared with those arising from DV, searching for the time-instant minimizing the differences between the two measures, namely the “equivalence” times. The distributions of equivalence times have been evaluated through the interquartile range (IQR). Figure 6.7 (a) shows an arterial input function, where the time interval $[t_a, t_{max}]$ at the denominator of Equation 6.6, is highlighted. Then, a representative equivalence time-instant is highlighted and the corresponding distribution of equivalence times is shown in Figure 6.7 (b), achieved from the equivalence times of all voxel within the analysed ROI. In addition, the medians of these distributions have been also used to compare aBF achieved with gMS and DV, meaning that each median have been used as the value at the denominator of Equation 6.6, thus mimicking the classical MS formulation.

Finally, in order to allow comparing all the medians with each other, and removing the inter-patient variability in a simple manner, each temporal distribution has been normalized to a common physiological time reference. Therefore, instead of using scan beginning, that is an absolute reference, I have arbitrarily used the arterial wash-in (WI_a) (that is the maximum slope of aorta) time which, practically, has been subtracted from the time distribution. Hence, the wash-in time normalization

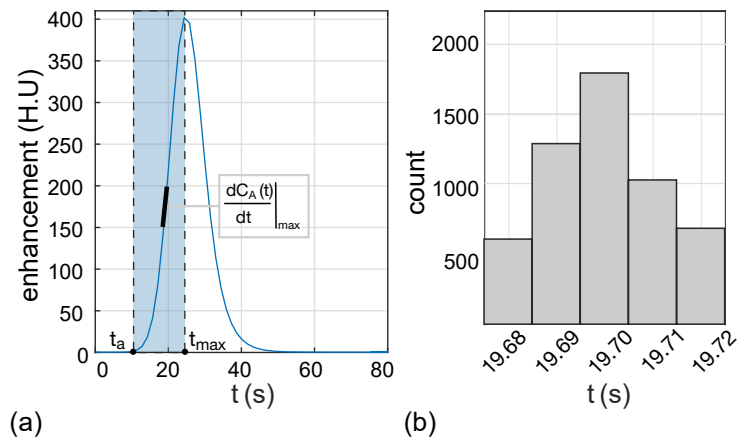


Fig. 6.7: A representative arterial input function in which it is highlighted the time interval $[t_a, t_{max}]$ where the equivalence time is searched out (a). Moreover, it is shown the arterial wash-in time used as normalization factor for gMS computation. The distribution of equivalence times achieved from all ROI voxels (b).

factor is also highlighted on the arterial function in Figure 6.7 (a).

Chapter 7

Experimental results: voxel-based reproducibility of BF

This Chapter presents the experimental results referring to the methodology developed to pursue MS and DV reproducibility, presented in Chapter 6. After a preliminary study focused on the analysis of fitting errors (Sect. 7.1), the Chapter presents results achieved on a dataset of 75 CTP examinations of PIXEL. First, this Chapter shows how the methods developed for solving computational open issues and improving precision of BF measurements have effectively reduced the variance and yielded the voxel-based agreement of BF measurements computed independently through MS and DV (Sect. 7.2). Yet more, this Chapter presents the proof of concept that MS, in its classical formulation, cannot represent the same status of the system described by DV, and the cause has to be attributed to transport delays between inputs and tissue. In this regard, the voxel-based aBF numerical equivalence obtained through gMS and DV is reported (Sect. 7.3). By supporting the idea that non-negligible transport delays are the cause of non-reproducibility of MS and DV, a case study is reported (Sect. 7.4), where, in the absence of any delay between input and tissue, the equivalence of MS and DV is achieved. In the end, a clinical application is presented (Sect. 7.5), where perfusion parameters have been used to compare, referring to SARAH study, the effects at one-month follow up of Sorafenib and selective internal radiotherapy (SIRT) treatments in advanced non-operable HCC.

7.1 Analysis of the effects of CTP fitting errors on BF values

In Chapter 6, tissue fitting through the GV model has been addressed. Since the computing methods for perfusion analysis, including MS and DV, are based on specific signal descriptors, such as maximum derivative, area under the curve, and

time to peak, the reliability of estimates strongly depends on the quality of the modelled TACs. Of course, performing a correct estimate of TACs represents the first necessary stage for the reliability of all the subsequent computational steps. Therefore, in the following I present a preliminary study aimed at investigating how TAC fitting errors may affect final perfusion values. To this end, first, the distributions of fitting residuals have been analysed in significant signal portions. Then, the implications of GV fitting errors upon BF values computed with MS (BF_{MS}) and DV (BF_{DV}) have been considered. The study has been carried out on 21 patients randomly chosen from the Centre 1 of the PIXEL study.

In order to assess to what extent GV fits real signals, I have performed the analysis of residuals by means of the histograms of voxel-based averaged residuals (μ_ϵ), thus evaluating statistical parameters as median value (M) and Median Absolute Deviation (MAD). Moreover, I have investigated the local distribution of the percentage errors. In particular, I have focused on error affection on different phases of the signals, separated as follows:

- **Ascent phase**
Between the first non-zero value of the real signal and one sample before the maximum value
- **Peak phase**
Evaluated in a range of ± 5 samples, with peak as the central one
- **Descent phase (or end phase)**
Between the peak and the end of the first pass, computed according to Sect. 6.3

This choice has been suggested by the high significance of these components in computing perfusion parameters. In particular, the ascent phase retains a great relevance for the MS method, since it holds the maximum derivative of the signal. Instead, the peak phase results crucial because it strongly affects the solution recovered by the DV method (whose maximum value estimates the BF, as in Sect. 4.5.3). Then, the descent phase fits the first-pass enhancement of CTP signals, as stated above. In order to assess the Gaussianity of the fitting process, a one-sample two-tail t -test has been performed on each histogram as well. In addition, also the third moment has been computed, to check for possible asymmetries. To the purpose of assessing the implications of the fitting procedure over perfusion parameters, the relation between μ_ϵ and BF values, arising from both MS and DV methods applied on GV fittings, has been inspected visually by means of residual plots. Then, I have split up the BF range (for both MS and DV) into 10 groups holding the same number of values, thus permitting to evaluate differences in terms of standard deviation of residuals among the groups.

7.1.1 Analysis of residuals

I have first considered the analyses of residuals (ϵ) for a single-patient, and only later on investigated on whether the attained results were generalizable, by considering the fitting errors globally occurring for the 21 examinations. Figures 7.1 and 7.2 (a) show the histograms of voxel-based errors of the ROI over the whole

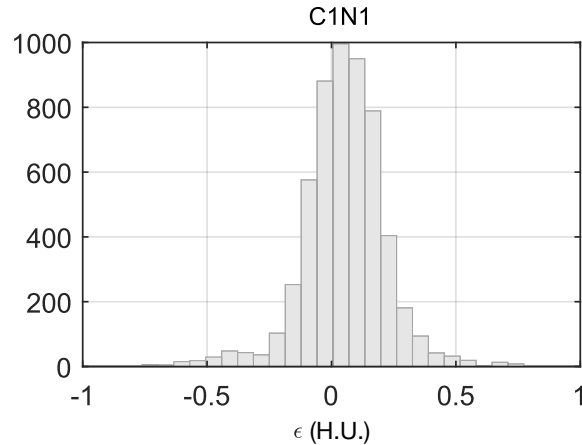


Fig. 7.1: Histogram of mean residuals (μ_ϵ) computed between GV fitting and real signals for patient C1N1 over the whole time interval.

time interval considered, by referring to a single-case (patient C1N1) and to all patients, respectively. As one can see, all fitting errors show a normal distribution, as confirmed by t -test (p-value $\sim 10^{-5}$).

Actually, the histogram in Figure 7.1 just shows a slight left skewness ($3.48 \cdot 10^{-1}$), that is absorbed by the global histogram (Figure 7.2 (a)), whose skewness value ($1.3 \cdot 10^{-3}$) is compliant with Gaussianity, this confirming single examinations as being instances of a Gaussian process. Yet more, this also indirectly supports the hypothesis that the patient population considered is statistically representative.

However, the histogram in Figure 7.1 has non-null mean ($\mu = -0.0190$) with standard deviation (σ) equal to 0.1683. Nevertheless, by investigating the local degree of dispersion, shown in Figure 7.3, voxel-based percentage errors result homogeneously distributed in a narrow range ($[-4.63 \div 3.93]\%$), this permitting to exclude the presence of local abnormalities in the performed fitting procedure. As regard the histogram of all examinations (Figure 7.2 (a)), both μ and σ increase in the absolute values ($\mu_{tot} = -0.454$ and $\sigma_{tot} = 0.3140$) thus emphasizing single-cases with possible higher absolute fitting errors. Figures 7.2(b), (c) and (d) depict the histograms of residuals referring to the ascent, peak and end phase of the signal, respectively. By considering the median residuals, one can see that in Figure 7.2 (b)

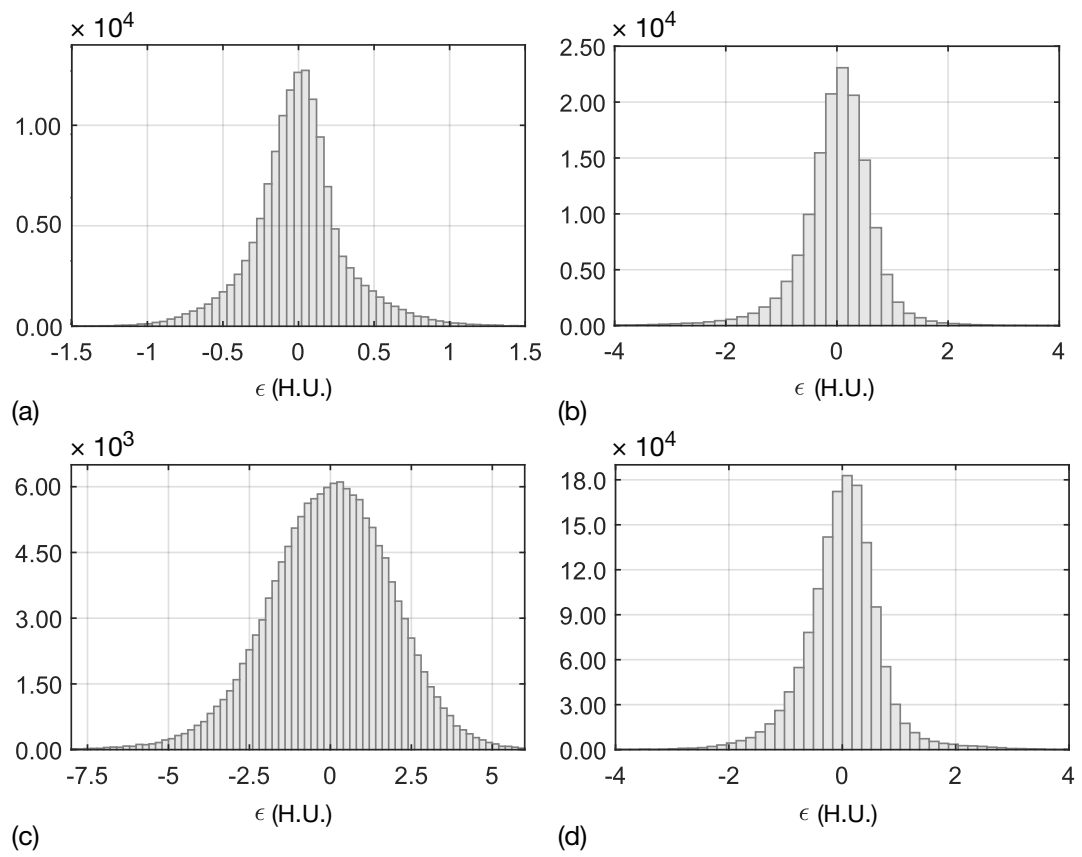


Fig. 7.2: (a) Histogram of mean residuals (μ_ϵ) computed between GV fittings and real signals for all the 21 patients and histograms referred to the (b) ascent, (c) peak, (d) descent phase, respectively.

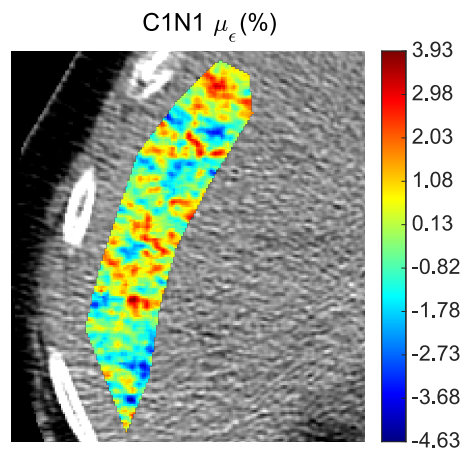


Fig. 7.3: Map of percentage errors of patient C1N1.

($\mu_{\text{ascent}} = -0.2304$) and (c) ($\mu_{\text{peak}} = -0.3449$) absolute values are bigger than in (d) ($\mu_{\text{end}} = -0.1163$). This confirms both the ascent and the peak phases as the most critical ones for the fitting procedure where, as expected, residuals are higher in (b) because of the rapid increments of the HU values, and in (c) due to the non-monotone behaviour of the signal within the interval. As regards (c), its criticality is also confirmed by the worst σ ($\sigma_{\text{peak}} = 1.8872$) among the three phases, while $\sigma_{\text{ascent}} = 0.6660$ is now even lower than $\sigma_{\text{end}} = 0.6990$. Figure 7.4 shows three different GV curve families, referred to as many patients. As one can see, there is much more variability in the peak or the end phase between GV curves of the same ROI rather than in the ascent phase, where signals show a higher similarity. This could motivate errors being more spread in those phases ((b) and (c)).

Globally, all histograms report negative M values, thus hinting at an underestimation of the real signals. I have realized that, often, this occurs in case of very noisy real signals, that are also greatly frequent. In fact, if referring to the HU mean value of the $C(t)$, acquisition's artefacts more often introduce over-enhanced values rather than under-enhanced ones. In order to clarify this aspect, Figure 7.5 (a) shows an example of GV curve that fits a weakly noisy $C_T(t)$. The fitting has frame-by-frame low residuals, without any singularity. On the contrary, as one can see in Figure 7.5 (b), the GV model seems not to be suitable to take into account strong and rapid variations of the real signal, whether these are localised in only one or a very few number of samples, although in those cases residuals are high. Nevertheless, errors' distributions reported in the separated phases of Figures 7.2 (b), (c) and (d) account for different numbers of evaluated samples. In fact, the ascent phase generally retains a greater number of samples and, moreover, it is the most compliant with the voxel-based signal features. On the contrary, the peak and the descent phase are studied by necessarily (and somewhat arbitrarily) defining the interval they refer to, that usually has a lower number of samples.

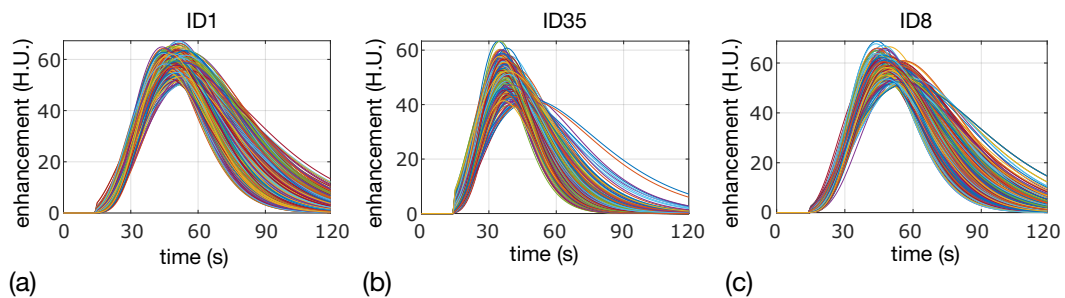


Fig. 7.4: Meaningful GV fittings for the ROI of three patients (from left to right, ID1, ID35, ID8).

7.1.2 Implications on BF values

On the basis of the achievements arising from the analysis of residuals regarding the fitting procedure, I have focused on relationships between errors and BF_{MS} and BF_{DV} . In particular, the Gaussianity of errors distribution shown in Figure 7.1 (a), still confirmed in the global evaluation reported in Figure 7.2 (a), allows considering BF values, attained for all the 21 patients involved in the study, in a globally manner. Therefore, Figures 7.6 (a) and (b) show the residual plots of voxel-based averaged errors over BF values for MS and DV, respectively. One can see that the residuals of Figure 7.2 (a), falling in the range $[-1.5 \div 1.5]$ HU, are uniformly distributed over the BF ranges of both MS ($[44 \div 218]$ [ml/min/100g]) and DV ($[45 \div 215]$ [ml/min/100g]). In addition, even an early visual evaluation of the plots, Figure 7.6 allows realizing that μ_ϵ arising from the fitting procedure result independent from BF values. Nonetheless, from Tab. 7.1 it is possible appreciating that the ten groups of 14165 BF values each show a comparable σ , this hinting the homoscedasticity of the classes. The residual plots in Figure 7.6 also confirm that fitting errors yield a quite good degree of symmetry with respect to 0 HU (the black horizontal line).

The setup fitting procedure yields residuals normally distributed, with percentage errors spatially homogeneous. Each voxel-based fitting curve shows the same uncertainty, thus limiting the risk of local anomalies or spatio-dependent erroneous values regarding the BF distributions that, besides, are independent from fitting errors. In practice, the achievements confirm that a GV fitting, properly parametrizes, can give the possibility to independent methods to compute perfusion parameters, laying in similar ranges, which could potentially converge towards similar values.

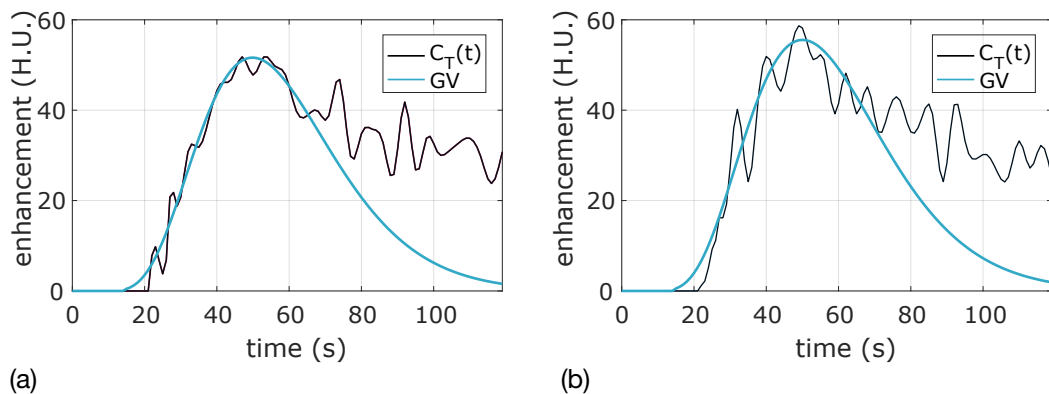


Fig. 7.5: (a) GV fitting of a weakly noisy $C(t)$, (b) GV fitting of a noisy $C(t)$.

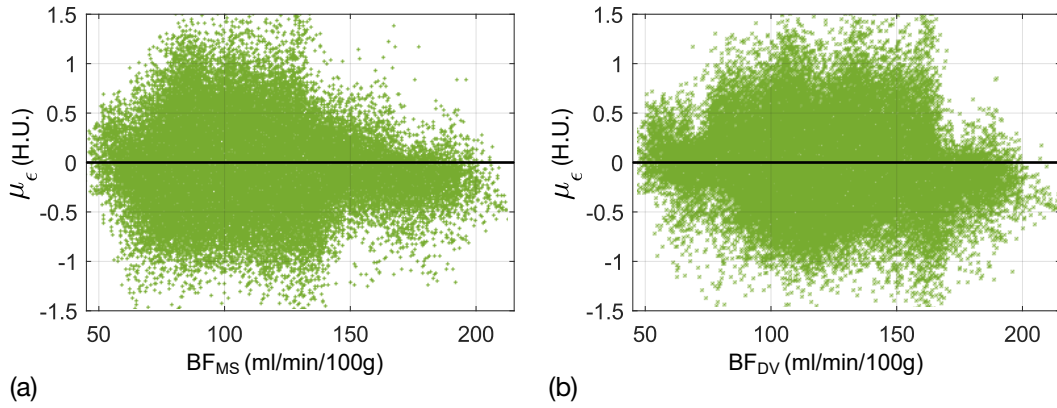


Fig. 7.6: Residual plot of voxel-based averaged μ_ϵ over BF_{MS} (a) and BF_{DV} (b).

Table 7.1: Table of σ values of μ_ϵ referred to ten different groups of BF values computed with MS and DV.

σ	μ_ϵ (MS)	μ_ϵ (DV)
σ_1	0.298	0.283
σ_2	0.343	0.277
σ_3	0.346	0.295
σ_4	0.308	0.304
σ_5	0.302	0.295
σ_6	0.291	0.290
σ_7	0.281	0.294
σ_8	0.307	0.306
σ_9	0.322	0.320
σ_{10}	0.246	0.362

7.2 Reproducibility of BF in the multicentre PIXEL study

As seen during the first preliminary study presented above and also mentioned in Chapter 6, Sect. 6.1, the methods developed for the reproducibility of perfusion parameters through MS and DV mainly address the PIXEL study. In particular, in the following I present the results of the reproducibility of BF values at single voxel-level through MS and DV achieved on a dataset of 75 patients, coming through the Centres 1 and 16 of the PIXEL study. In particular, fifty-four patients have been randomly chosen from Centre 1 and all patients sharing the same protocols of Centre 1 have been selected from the Centre 16.

In the following, after a brief discussion of the methodological results regarding the fitting procedure set up for vascular and tissue signals, using LN and GV parametric models (Sect. 6.4), I present the voxel-based reproducibility of BF values,

by considering also a comparative analysis of my results with the state of the art. The methods developed for extracting and modelling the first pass, separating tissue components due to the dual inputs on which computing independently MS and DV, are expected to improve precision of measurements and allow a fair voxel-based comparison of perfusion values. Accordingly, from the numerical side, the voxel-based agreement of BF values is addressed. Moreover, from the theoretical side, the main achievement of this study concerns the questioning of the MS validity in its classical formulation and the proof of concept that an alternative formulation of MS can lead the two methods to effectively represent the same status of the system, and MS and DV to be substantially equivalent, accordingly.

7.2.1 Analysis of fitting parameters

The implementation of the fitting procedure set up for vascular and tissue signals, through LN and GV parametric models, has been addressed in Sect. 6.4. In this regard, Table 7.2 resumes the parameters arising from the minimization procedure

Table 7.2: Summary of LN parameters minimized against real vascular signals of aorta and porta. All parameters but t_m are in arbitrary unit. t_m is measured in s .

		m	σ	min	max
AUC	Aorta	5412	1580	2550	9600
	Porta	4910	1627	2350	10700
t_m	Aorta	22.6	4.2	15.5	34.2
	Porta	49.1	9.5	32.0	85
RD	Aorta	0.3	0.1	0.2	0.4
	Porta	0.5	0.1	0.4	0.6
s	Aorta	1.3	0.3	0.4	1.7
	Porta	1.8	0.1	1.5	2.0

for vascular signals, that is aorta and portal vein. In particular, mean (m), standard deviation (σ), minimum (min), and maximum (max) values are provided for the four LN parameters AUC, t_m , RD , and s . I recall that patient-based initial estimates have been used for AUC and t_m , whilst patient-independent initial values have been adopted for RD and s , being equal to $RD=0.3$ and $s=1.5$ for the aorta, and $RD=0.7$ and $s=1.7$ for the porta. Results shown in Table 7.2 substantially confirm the need for patient-based initial estimates for the two parameters AUC and t_m , both showing high variability either in aorta or portal vein, as justified by the coefficient of variation, CV , of both parameters. In particular, as regards AUC, $CV=29\%$ for the aorta, and $CV=33\%$ for the portal vein. In addition, even the ranges are spread over wide distributions, as reported by the range width (rw) which is $AUC_{rw}=7050$ for the aorta and $AUC_{rw}=8350$ for the porta. Moreover, data in Table 7.2 highlight

slight differences between aorta and portal vein if referring to AUC, thus suggesting that when the enhancement of aorta and portal vein is caught, the amount of CA into the two vessels is almost comparable. However, a delayed transport is highlighted in the portal vein where, as one can see in Table 7.2, the t_m of the portal vein is on average twice the one of the aorta, with the same $CV=19\%$ for both the vessels. The delayed transport in the portal vein is also confirmed by the ranges of t_m , where one can notice that the max of t_m of the aorta is 2.2 s after the min of t_m of the portal vein.

As regards RD , the initial value $RD=0.3$ for the aorta has been confirmed on average after the minimization procedure, whilst the value $RD=0.7$ used as initial estimate for the portal vein has overestimated the mean value of RD after the minimization equals to $RD_m=0.5$. In addition, the values of RD suitable for fitting the two vascular signals result limited to three discrete values, as also shown by the ranges and the values of σ , reported in Table 7.2. Finally, the values of s used as initial estimates for the two vessels, over- and under- estimates the values achieved after the minimization procedure, for the aorta and portal vein, respectively. In fact, $s=1.3$ is slightly lower than $s=1$, initially considered for the aorta, whilst $s=1.8$ results slightly greater than $s=1.5$ used as initial seed for the porta. Moreover, if compared to RD , the wider ranges of the s parameter achieved for both aorta and portal vein show a great variability, especially as regards the aorta, where $s_{rw}=1.3$, whilst $s_{rw}=0.5$ for the portal vein. As a concluding remark, I recall that RD and s represent respectively the dispersion of the curve, which mainly affects the fitting of the positive slope of the signal, and its skewness, which influences the fitting in the descending phase of the curve. Accordingly, it is worth noting that, based on these parameters, a great variability of vascular signals is highlighted during the descent phase rather than during the ascent one, since s shows a wide range of possible values, whilst the range of RD is really limited.

Table 7.3 reports the parameters arising from the minimization procedure of

Table 7.3: Resume of GV parameters minimized against real tissue signals. All parameters but t_0 are in arbitrary units. t_0 is measured in s .

	m	σ	min	max
K	$1.08 \cdot 10^{-3}$	$7.57 \cdot 10^{-3}$	$2.71 \cdot 10^{-6}$	$1.52 \cdot 10^{-2}$
α	5.26	0.73	3.50	6.85
β	8.96	2.76	4.29	15.4
t_0	7.80	0.68	4.82	12.9

tissue signals, achieved employing the GV model. In particular, m and σ refer to the averaged values arising from all voxel-based solutions achieved for each patient. Similarly, min , and max values are referred to the extreme values of m achieved for

all patients. Actually, the values of all four GV parameters, K , α , β , and t_0 , result very different from the initial estimates which have been fed to the minimization algorithm in a patient-independent approach, that have been $K=50$, $\alpha=0.5$, $\beta=20$, and $t_0=25$ s. This reasonably means that the initial estimates identified in a heuristic manner, represented sub-optimal solutions for the parameters, corresponding to local minimum values. However, one can appreciate from data reported in Table 7.3, that the minimization algorithm exploited for estimating fitting parameters converges towards more reliable distributions, also spread over quite narrow ranges for all parameters but K . In fact K , which is a global scale factor, shows a wide variability both in terms of dispersion ($\sigma=7.57 \cdot 10^{-3}$) and range width ($K_{rw}=0.15$). This means that K is the GV parameter which mainly accounts for the voxel-based variability of real signals. One should consider that GV fitting, differently from LN fitting, refers to each voxel-based signal separately. Accordingly, the wide diversity of tissue signals, even emphasized from a greater – if compared to ROI-averaged vascular signals – noise component corrupting them, leads to such a large range of K values. A similar behaviour is shown also by β , which is the local scale factor in the GV formulation. In fact, although less variable than K , β shows $CV_\beta=30\%$, thus representing a similar tendency to catch the variability of voxel-based signals, as also reported by the quite large range, $\beta_{rw}=11.1$. Meanwhile, the other three parameters, α and t_0 are much more similar among different voxels, as shown by the very low CVs, which are $CV_\alpha=14\%$ and $CV_{t_0}=9\%$, respectively. Moreover, α is characterized by even a smaller range ($\alpha_{rw}=3.35$), which means that it is the most stable in identifying the family of tissue signal curves, as expected because it is the shape parameter. Finally, despite of a low CV value, t_0 shows a quite large range with $rw=8.08$ s, thus representing the very low variability of CA appearance time instants among voxels of the same patient and the large variability among different patients.

7.2.2 Voxel-based reproducibility of MS and DV

When considering the assessment of agreement between MS and DV, one should notice that few works exist reporting BF values achieved with both MS and DV, and all of these only refer to aggregate data, related to the entire cohort. This is the first research reporting a patient-wise comparison of voxel-based BF values achieved with both MS and DV, besides a cohort analysis to enable a comparison with the state of the art.

For each patient of Centres 1 and 16, voxel-based BF values achieved via MS and DV have been compared through the Pearson correlation index (ρ), split into five contiguous classes with increasing correlation, in order to permit a more accurate comparison between Centres. In addition, for each patient, M and MAD, m and

σ , and CV are also computed for MS and DV separately. In order to assess the semantic consistency of the results achieved by both methods, I have considered their spatial distributions through colorimetric maps, investigating the similarity between patterns through the normalized cross-correlation (NCC). Moreover, the voxel-based agreement of BF_{MS} and BF_{DV} has been assessed with scatter plots, thus measuring the standard deviation of the secondary component obtained through the Principle Component Analysis (PCA).

As regards the cohort analysis, the correlation of all mean BF (BF_m) values achieved via MS and DV is computed at group (G) level, where “group” is meant as the set of patients of either Centres 1 or 16 or both (1&16) and assessed through Spearman (ρ_s), Pearson (ρ_G), and ICC indexes. To this purpose, when addressing the comparison with the literature, these correlations (c) between MS and DV have been considered good or very good if $0.80 \leq c < 0.90$ or $c \geq 0.90$, respectively. On BF_m and BF_M distributions, M_G , MAD_G , m_G , σ_G , and $CV_G = \sigma_G/m_G$ have been assessed. Finally, a comparison between MS and DV is carried out considering the absolute percentage differences of M_G (Δ_M) and m_G (Δ_m). Afterwards, we compared our results with other studies’, considering all the published works between 2013 and 2019, retrieved from PubMed database including the keywords: “functional CT, perfusion CT, CT-perfusion, deconvolution, maximum slope, CT-based, dynamic contrast-enhanced computed tomography, dynamic contrast-enhanced CT” and excluding: “brain, cerebral, artery, coronary, stroke, cardi, dynamic contrast-enhanced MRI, dynamic contrast enhanced MRI”. Finally, 18 works are considered, dealing with different organs and glands, including liver (8), kidney (3), pancreas (3), lung (2), oesophagus (1), lymph nodes (1).

I start presenting the patient-wise MS and DV voxel-based correlations, followed by a comparison of the aggregate results with other studies.

Table 7.4 resumes the Pearson correlations between BF values computed at voxel

Table 7.4: Correlation (ρ) of BF between MS and DV in Centres 1 and 16

Centre	Total	Patients					CV_{MS}	CV_{DV}
		$\rho=0.99$	$\rho=0.98$	$\rho=0.97$	$\rho=0.96$	$0.90 \leq \rho \leq 0.95$		
1	54	37	5	6	3	3	11.6%	10.7%
16	21	15	4	-	1	1	11.3%	11.0%

level with MS and DV for each patient of Centres 1 and 16, where ρ values are partitioned into 5 contiguous classes. Correlations are excellent, with 95% of patients with $\rho \geq 0.96$ in Centre 1 as well as in Centre 16. These values are yet more significant in the light of the very low mean CV s of all patients for MS (CV_{MS}) and DV (CV_{DV}), suggestive of BF distributions with narrow ranges, with maximum $CV_{MS}=11.6\%$ and minimum $CV_{DV}=10.7\%$ values in Centre 1, for MS and DV, respectively. It is also

worth noting that CV_{DV} values are also lower than CV_{MS} ones, this confirming the better precision of DV. Figure 7.7 (a,b) show the colormaps achieved by MS and

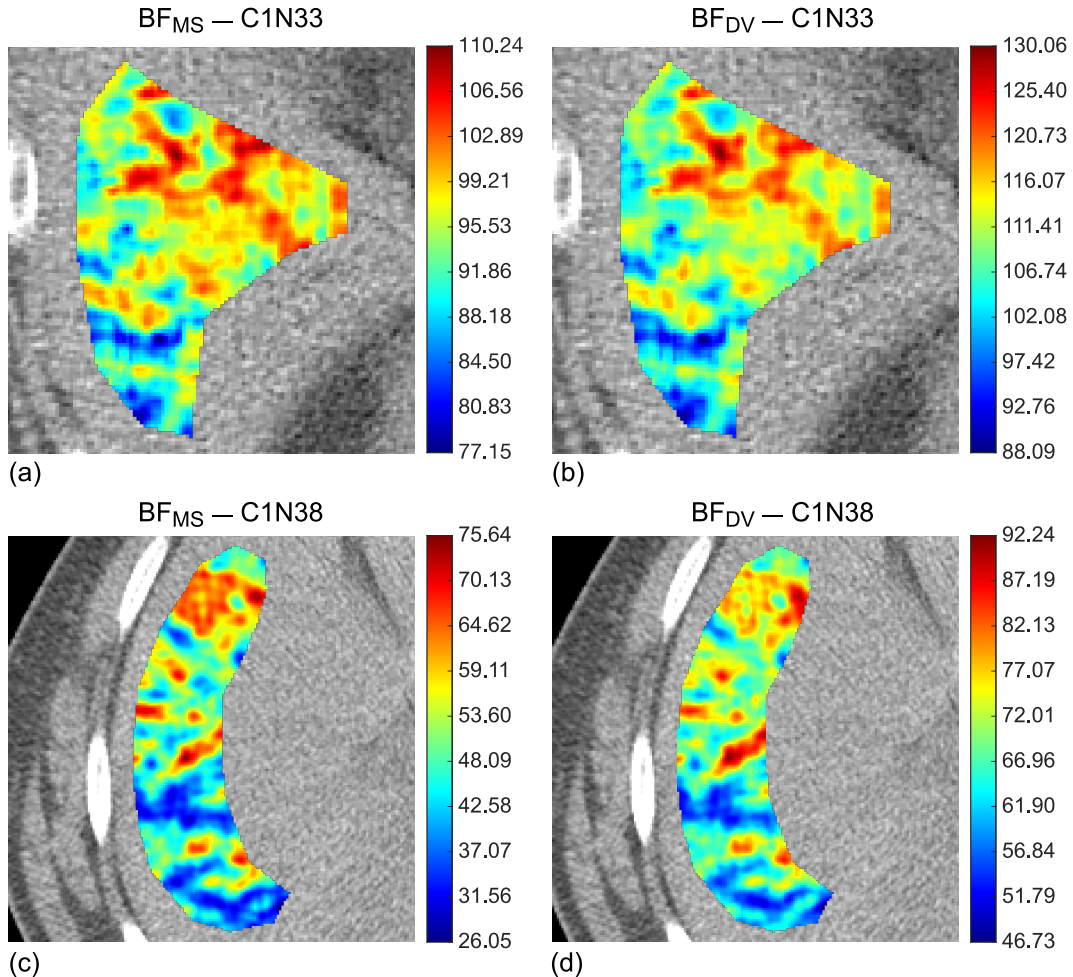


Fig. 7.7: Colorimetric maps of BF_{MS} (a,c) and BF_{DV} (b,d) for one among the best examinations (C1N33), which shows $\rho = 0.99$, and a representative worse case (C1N38), with $\rho = 0.97$.

DV for one among the best examinations (C1N33), as one can see also from the scatter plot of Figure 7.8 (a), which yields $\rho = 0.99$, and the results of NCC in Figure 7.9 (a), where $NCC_{max}=0.9997$. The dispersion of the scatter plot is very low, with $\sigma=0.4105$, this reflecting the high homogeneity between patterns even at a local level, as confirmed by the symmetry of the peak and the quite anisotropic dispersion of NCC values. Similarly, Figure 7.7 (c,d) reports the colorimetric maps of BF by MS and DV for a representative patient (C1N38) having a worse correlation, as also shown by the scatter plot in Figure 7.8 (b) where $\rho=0.97$, and the NCC in Figure 7.9 (b) where $NCC_{max}=0.9954$. While the latter is really a high value for NCC, the isotropic distribution of NCC values hints at a much worse match,

Table 7.5: Report of DV- and MS-based BF values in the most recent research studies, addressing healthy tissue (H), primary cancer (C), or metastases (m) in different organs and glands.

Year, Authors	#Pats.	Organ/gland, H/C/m	Correlation		$M_G \pm MAD_G$		$m_G \pm \sigma_G$		CV _G		Δ (%)		
			ρ_G	ρ_s	MS	DV	MS	DV	MS	DV	Δ_M	Δ_m	
[214]	(2013, Djuric-S et al.)	35	Oes., C	0.55	0.59	25.4	74.8	28.6 ± 12.3	78.5 ± 28.0	0.43	0.36	66.0	63.6
[215]	(2013, van Elmpt et al.)	33	Lun., m	0.91	-	-	-	36.5 ± 20.8	74.7 ± 47.7	0.57	0.64	-	51.1
[132]	(2015, Kaufman et al.)	79	Liv., C	0.86	-	-	-	37.8	68.3	-	-	-	44.7
[216]	(2016, Schneeweiß et al.)	48	Pan., C	0.89	0.62	-	-	20.4 ± 9.7	36.9 ± 16.0	0.48	0.43	-	44.7
[217]	(2017, Kurucay et al.)	36	Liv., C	-	-	-	-	-	97.3 ± 45.1	-	0.46	-	-
[218]	(2017, Fischer et al.)	20	Liv., C	-	-	-	-	-	48.3 ± 15.8	-	0.33	-	-
[219]	(2017, Tamandl et al.)	16	Liv., C	-	-	-	-	-	38.5	-	-	-	-
[220]	(2018, Deniffel et al.)	35	Kid., C	-	0.85	-	-	134.8 ± 61.9	-	0.46	-	-	-
[221]	(2018, Aslan et al.)	73	Pan., H	-	-	-	-	-	118.6 ± 36.4	-	0.31	-	-
[222]	(2018, Kaufman et al.)	28	Liv., C	-	-	-	-	-	103.6 ± 36.8	-	0.36	-	-
[223]	(2018, Horger et al.)	23	Lim., C	-	-	-	-	41.4 ± 18.8	75.2 ± 36.3	0.45	0.48	-	45.0
[224]	(2018, Mains et al.)	69	Kid., m	-	0.81	-	466.3 ± 213.4	211.9 ± 80.6	-	-	-	54.6	-
[225]	(2018, Mulé et al.)	16	Liv., H	-	-	-	-	-	118.3 ± 92.9	-	0.79	-	-
[226]	(2018, Nakamura et al.)	36	Liv., H	-	-	-	-	-	149.6	-	-	-	-
[151]	(2019, Fan et al.)	10	Kid., C	-	-	-	148.5	-	125.4 ± 70.5	-	0.56	-	-
[227]	(2019, Andersen et al.)	39	Liv., C	-	-	-	-	-	123.9	-	-	-	-
[130]	(2019, Kovač et al.)	44	Pan., C	-	-	-	-	-	-	0.53	-	-	-
[228]	(2019, Wang et al.)	39	Lun., C	-	-	-	-	24.0 ± 12.8	-	-	-	-	-
our work-Centre 1		54	Liv., H	0.97	0.96	95.1 ± 18.2	109.8 ± 19.9	94.3 ± 23.4	111.3 ± 25.3	0.25	0.23	13.4	15.3
our work-Centre 16		21	Liv., H	0.99	0.98	90.9 ± 16.1	102.4 ± 16.6	90.0 ± 19.7	102.1 ± 20.7	0.22	0.20	11.2	11.9
our work-Centres 1&16		75	Liv., H	0.97	0.96	94.0 ± 17.6	108.7 ± 19.1	93.1 ± 22.4	108.7 ± 24.3	0.24	0.22	13.5	14.4

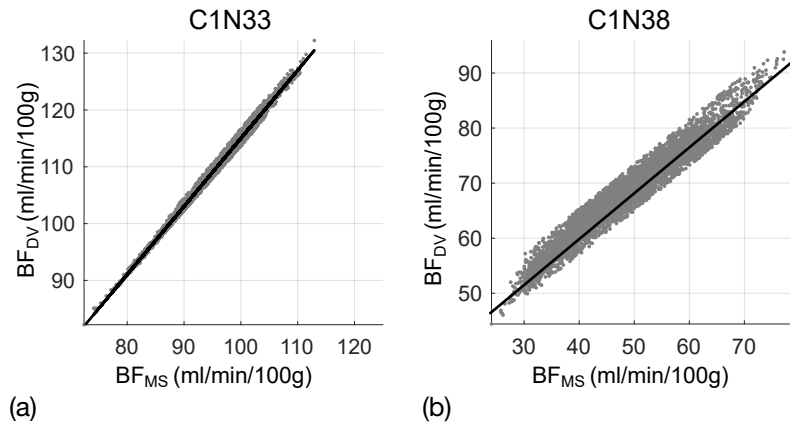


Fig. 7.8: Linear regression of BF distribution for the examination C1N33(a) of Figure 7.7 (a,b) and the case C1N38(b) of Figure 7.7 (c,d). The two scatter plots have $\rho=0.99$ and $\sigma=0.4105$ in (a) and $\rho=0.97$ and $\sigma=1.4843$ in (b) showing an excellent and discrete agreement, respectively.

which emphasizes small local inhomogeneities between patterns, also confirmed by a much higher dispersion degree of the scatter plot ($\sigma=1.4843$). Nevertheless, both the maximum values are achieved at the image centre.

Table 7.5 reports the outcomes of our study (Centres 1, 16, 1&16) and of the most recent literature addressing healthy tissue (H), primary cancer (C), or metastases (m) in different organs and glands. These studies were all single Centres, except for [214], and perfusion parameters were always computed with vendor's Software. The results reported perfusion parameters, correlations, and absolute percentage differences of median and mean BF values achieved with MS and DV, referred to the whole cohorts. As one can see, most of parameters are not computed ('-' points out not available values) and this regards not only voxel-based, but group-wise analyses as well, where the only parameters reported are those deriving from mean (μ_G , σ_G , CV_G), while median-derived parameters are almost never computed. As a matter of fact, this element itself hints at a lack of accurate comparative studies, making my work the most analytical one. Six works reported at least M_G or m_G values for both MS and DV. Apart from the older work in [214], reporting correlation values lower than 0.60, almost all the other correlation indexes are good (>0.80) or very good (>0.90), but they are never coupled with low differences between MS and DV BF values, this suggesting at least relevant systematic errors between MS and DV computations. This happens in [224] ($\rho_s=0.81$, $\Delta_m=54.6\%$), [132] ($\rho_s=0.86$, $\Delta_m=44.7\%$), [216] ($\rho_s=0.89$, $\Delta_m=44.7\%$), and even in [215], where $\rho_G=0.91$, the highest ρ_G value of all the comparing studies considered, derives from $\Delta_m=51.1\%$, probably due to a linear correlation having a slope much higher than 1. Also the work in [220] shows very good correlations ($\rho_s=0.85$, $ICC = 0.83$) but, besides

reporting by far the highest σ_G in MS computations, not any (absolute percentage) difference is given, nor ρ_G , this probably suggesting that neither voxel-based nor global BF values were comparable with DV ones.

As regards my results, Figure 7.10 highlights the correlation of BF values computed with MS and DV on the patients of Centres 1 (a) and Centre 16 (b). ρ_G , ρ_s , and ICC coefficients are very high for Centre 1 (0.97, 0.96, 0.78) and excellent for Centre 16 (0.99, 0.98, 0.84), and such an agreement is confirmed (even slightly improved for ICC) by the multicentre analysis of 1&16 (Figure 7.6 (c)), with 0.97, 0.96, 0.79, respectively. Analogously, as regards m and q values, we can see that increasing the number of patients by adding to Centre 1 those of Centre 16 does not improve the slope, but it improves the bias, from $q=12.24$ to $q=10.48$. Actually, ICC coefficients for Centres 1 and 1&16 are lightly lower than those reported in [220], but it is worth noting that our BF values are associated to the highest precision, as confirmed by the lowest CV_G s, when referring to either MS ($CV_G=0.25$ and $CV_G=0.24$, for Centres 1 and 1&16, respectively) and DV ($CV_G=0.23$ for Centre 1 and $CV_G=0.22$ for 1&16). The high precision is confirmed also in Centre 16, which yields the best CV_G values, for DV (0.20) and MS (0.22), and the lowest $\Delta_M=11.2\%$ and $\Delta_m=11.9\%$. In addition, also for Centres 1 and 1&16 the percentage median and mean differences between MS- and DV-based BF values are incomparably lower than those reported in the studies considered in Table 7.5, with the “worst” $\Delta_M=13.5\%$ and $\Delta_m=15.3\%$ occurring for Centre 1&16 and 1, respectively. These low differences between MS- and DV-based BF values are possible thanks to the almost unitary slope and the quite low bias, as confirmed by the intercept (q) values shown in Figure 7.6. The last remarks arise from a comprehensive view of

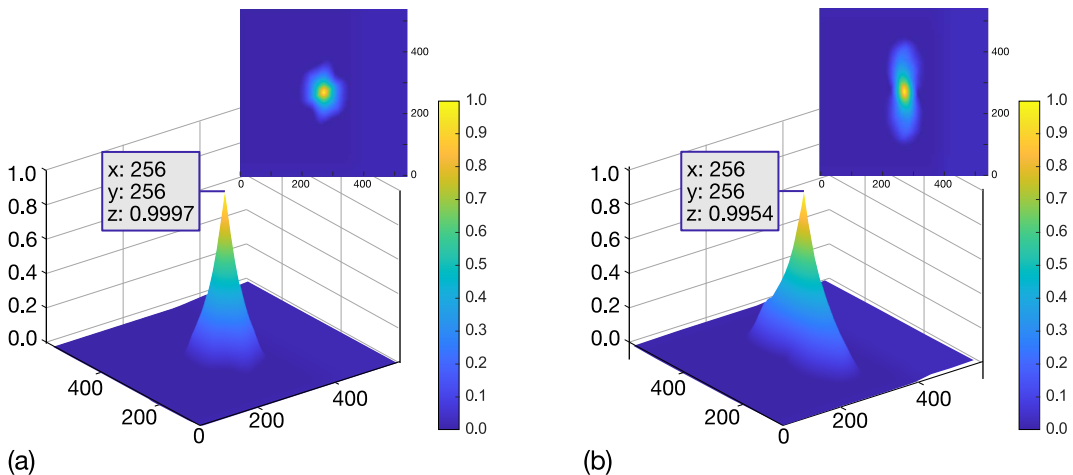


Fig. 7.9: NCC of the maps in Figure 7.7 for the patients C1N33 (a) and C1N38 (b), respectively, where $NCC_{max}=0.9997$ (a) and $NCC_{max}=0.9954$ (b), both occurring at image centre.

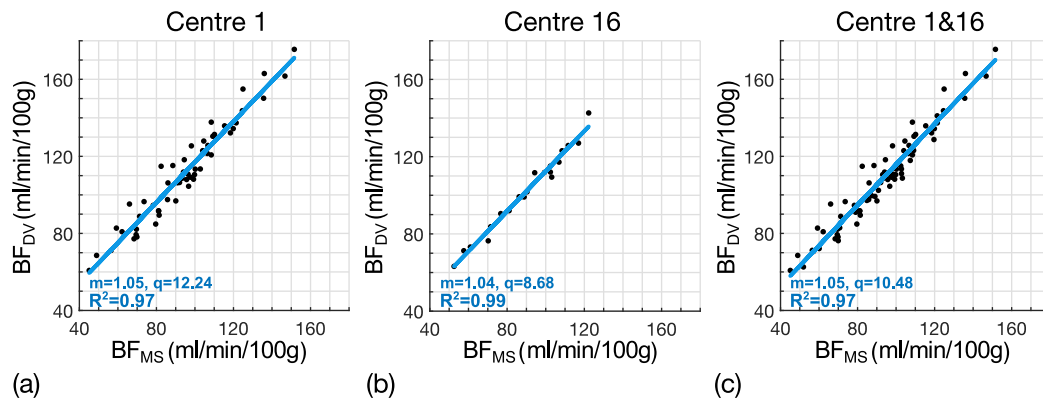


Fig. 7.10: Scatter plots of median BF values computed with MS (x -axis) and DV (y -axis) in Centre 1 (a), 16 (b), and 1&16 (c), respectively.

Table 7.5. It is clear that there are very few studies directly addressing the problem of reproducibility of BF values, whether these are computed with either MS or DV, and when the percentage differences were reported, these were around 50% or even more. This is independent of the organ and its healthy status - 13 works deal with primary cancer, two with metastasis and three works only address healthy tissue. All CV_G referring to MS and DV computation are much higher than ours and it is worth noting that the second best $CV_G=0.33$ [218] and the worst $CV_G=0.79$ [225] refer to liver cancer and healthy tissue, respectively. This suggests that the lowest CV_G values of our results do not depend on the healthy status of liver and, more in general, on the organ, but it can be ascribed to the precision and the accuracy of our CTp parameters computation methods, whose results are emphasized by Table 7.4.

7.3 The equivalence times of MS and DV

As explained in Sects. 6.6, 6.7, besides the agreement of BF measurements, I have investigated the theoretical aspects underlying the MS method, to go beyond the “simple” concordance of BF values and achieve their numerical equivalence. To this end, I have wondered whether the two methods were really representing the same status of the system, and so the equivalence was basically reasonable. Accordingly, by questioning the validity of classical MS assumptions, I have formalized the gMS (Sect. 6.6), which addresses potential transport delays from the vascular input to the tissue being studied, by varying the time reference in which the vascular input is considered to compute the BF through the MS. In particular, I have searched that time-instant when the vascular input signal can allow – if employed in the MS formulation – the equivalence with BF computed through DV, thus mimicking a reduced amount of vascular input which reaches the system in case of a delayed

delivery. Hence, the distributions of so called equivalence times has been recovered, that is where MS and DV yield exactly the same values, and the median of these distributions has been adopted to compute the gMS, thus analysing whether a time of equivalence between the two methods existed and where it was localized with respect to a common temporal reference, which was the WI_a . In particular, the methodology has been setup for the equivalence of aBF. Accordingly, the following results refer to the 75 patients of the PIXEL study, for which I have already presented the agreement of total BF. The distributions of equivalence times have been assessed through σ , IQR, and rw in order to measure their dispersion. Then, the localization of the median equivalence time has been analysed with respect to meaningful references on vascular and tissue signals, by considering the time distances with respect to the time instant of WI_a , the peak of the aortic signal ($PEAK_a$), and the maximum slope of the tissue (MS_t). Moreover, for each patient, the linear correlation between the gMS computed in the median of the equivalence time distribution and DV has been measured through the R^2 index with the aim of proving the numerical equivalence if achieving unit slope and null intercept.

All the equivalence time distributions of DV and gMS result Gaussian-like distributions, as that one shown in Figure 7.11 for the patient C1N26, which has the me-

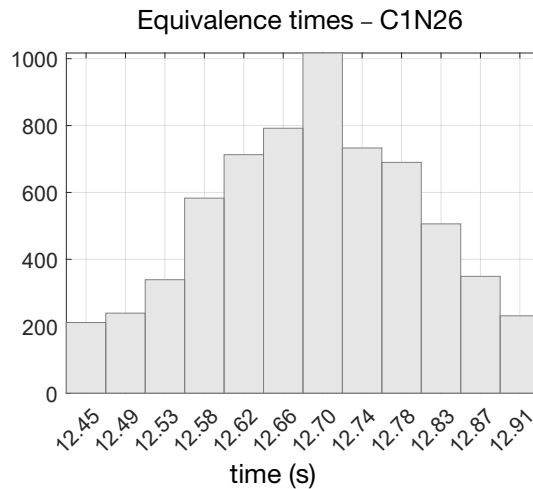


Fig. 7.11: Distribution of equivalence times for patient C1N26, which yields the median σ value, $\sigma = 0.11s$. As one can notice, equivalence times are normally distributed.

dian σ value of the dataset equals to $\sigma=0.11$ s (IQR=0.16 s, $rw=0.50$ s). Moreover, the same Gaussian-like distributions are achieved for patients C1N38 and C16N12, yielding one of the most and least spread histograms, respectively, with $\sigma_{C1N38}=0.29$ s and $\sigma_{C16N12}=0.02$ s, and reported in Figure 7.12. In addition, these patients represent nearly the worst and the best case, if considering their IQRs, IQR=0.39 s for C1N38 and IQR=0.01 s for C16N12, with $rw_{C1N38}=1.22$ s and $rw_{C16N12}=0.16$

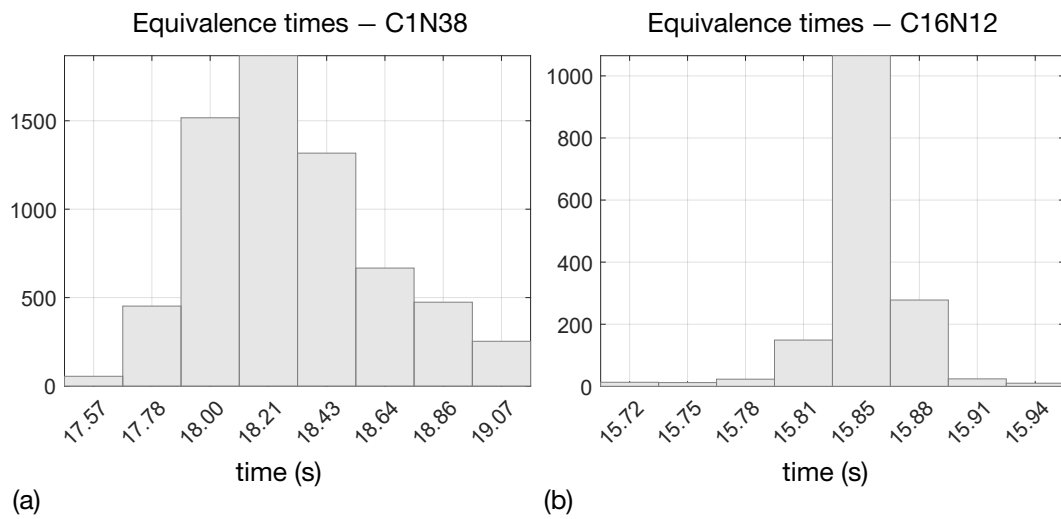


Fig. 7.12: Distributions of equivalence times for two patients yielding nearly the most and least spread equivalence time distributions, C1N38 (a) with $\sigma = 0.29s$, $IQR=0.39s$, $rw = 1.22s$, and C16N12 (b) with $\sigma = 0.02s$, $IQR= 0.01s$, and $rw = 0.16s$

s, respectively. Moreover, as one can notice from Figure 7.13 reporting the spatial

Map of equivalence times – C1N26

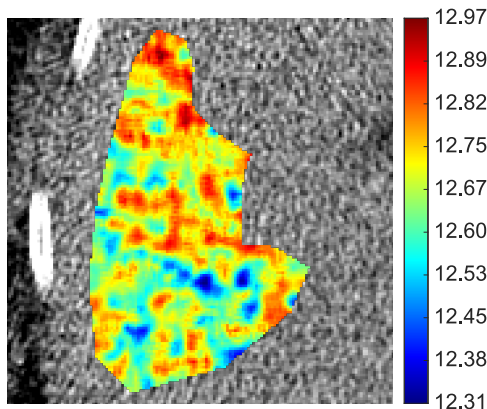


Fig. 7.13: Spatial distribution of equivalence times for patient C1N26 (the one reported in Figure 7.11) which results quite homogeneous if considering the small range depicted, without any local abnormality.

distribution of equivalence times for patient C1N26 (the one in Figure 7.11), these time of equivalence are quite homogeneously distributed over the ROI, thus allowing excluding local abnormalities in transport delays.

Figure 7.14 (a) reports the distribution of IQRs for all patients. Despite the greatest $IQR=0.78s$ (C1N20), also corresponding to the worst $rw=1.97s$, one can notice that most of distributions are quite narrow, as proved by the 81% of patients

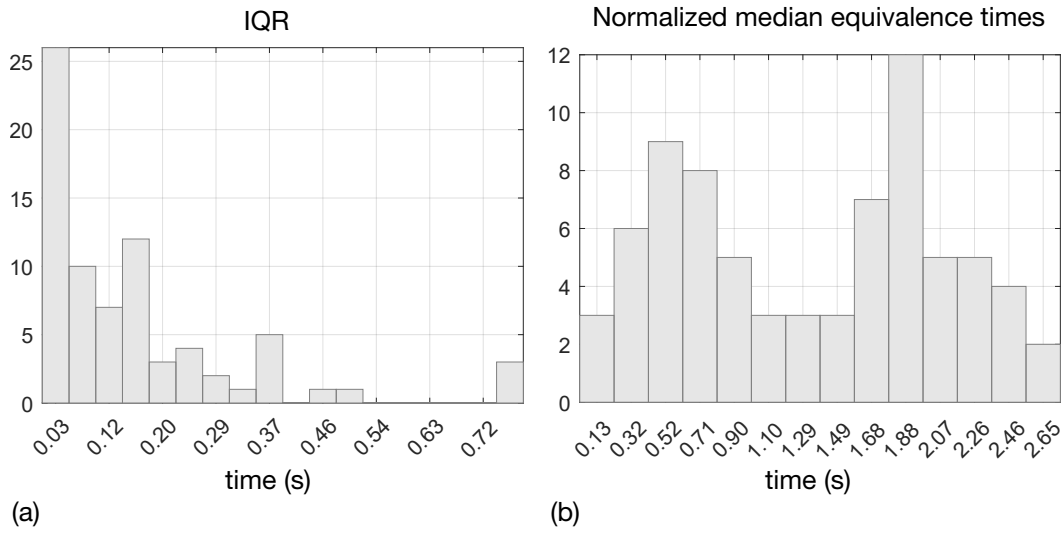


Fig. 7.14: Histograms of the IQRs (a) of the equivalence time distributions and normalized median values (b) with respect to WI_a .

having $IQR \leq 0.25$ s. As above explained, for each equivalence time distribution, the median time has been considered in the gMS formulation, thus employing the value of the vascular input in that time-instant rather than in the peak one. Figure 7.14 (b) shows the distributions of these medians with respect to the chosen normalization factor which is WI_a , the wash-in time of the aorta, corresponding to the maximum slope of $C_A(t)$. The histogram is clearly bimodal, where each distribution is Gaussian-like.

As one can see in Figure 7.14 (b), the median of equivalence time distributions are localized between $[0.04 \div 2.59]$ s after WI_a , this corresponding to the ascent portion of the aortic signal comprised between the WI_a and $PEAK_a$. It is also worth noting that this is the region of the aortic signal most prone to errors due to the steepest slope. Under this consideration, the IQR histogram in Figure 7.14 (a), which reports such narrow distributions, is yet more meaningful. Hence, by considering the bimodal behaviour in Figure 7.14 (b), it basically means that for the first sub-population of patients, where the normalized median is nearly lower than 1.22 s, the aBF computed with gMS and DV are equal when the arterial input in a time instant close to the WI_a is considered in the gMS formulation, on average, 0.61 s after WI_a . Instead, for the second sub-population of patients, the normalized median is nearly greater than 1.22 s, on average 1.88 s after the WI_a , closer to $PEAK_a$. Therefore, I have hypothesized that the reasons of such bimodal behaviour in Figure 7.14 (b) could be related to the tissue enhancement. In fact, since the gMS has been thought to account for any time delay hampering the MS in its classical formulation to represent the same status of the system depicted by the DV, I have analysed the relationships between MS_t and

PEAK_a. Accordingly, Figure 7.15 (a) reports the histogram of the time distance

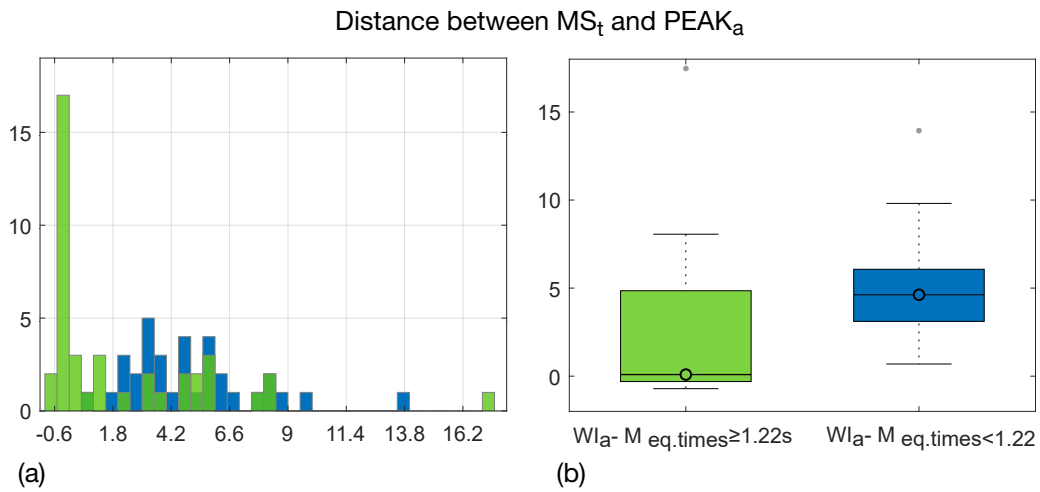


Fig. 7.15: Analysis of the distances between MS_t and PEAK_a in the two sub-populations highlighted in Figure 7.14 (b) with respect to the cut-off value of 1.22 s. The green bars (corresponding to data in Figure 7.14 (b) greater than 1.22 s) shows a minimal delay between MS_t and PEAK_a. The blue bars (corresponding to data in Figure 7.14 (b) lower than 1.22 s) shows MS_t centred around 4.96 s after PEAK_a (a). The statistical separations of these two groups is confirmed with $p\text{-value} \sim 10^{-5}$.

between MS_t and PEAK_a for the first sub-population of Figure 7.14 (b), that is the green bars corresponding to normalized medians centred around 0.61 s, and the second sub-population represented by the blue bars, having normalized medians centred around 1.88 s. As one can see from Figure 7.15 (a), the two populations of Figure 7.14 (b) are even separated by considering the localization of MS_t with respect to PEAK_a. Moreover, the separation is statistically confirmed at Wilcoxon rank-sum test with $p\text{-value} \sim 10^{-5}$, as also shown in the boxplots in Figure 7.15 (b). Basically, this means that the green-bar-group has MS_t very close to PEAK_a, and requires the arterial input value to be more distant from the WI_a (on average after 1.88 s), whilst, the blue-bar-group shows MS_t around 4.96 s after PEAK_a and requires the arterial input very close to WI_a (after 0.61 s on average). This seems to confirm the hypothesis that a delayed delivery between arterial input and tissue compartment is the primary cause of the underestimation of aBF achieved when using the MS in its classical formulation. In fact, Figure 7.15 (a) means that when a short delay between PEAK_a and MS_t exists, one performs a little underestimation with classical MS (using the PEAK_a as the reference at the denominator of MS formulation) and the correction factor searched via gMS is localized quite close on the left of PEAK_a. Instead, when there is a greater delay (around 4.96 s) between PEAK_a and MS_t, the classical MS formulation substantially underestimates the aBF value, and the correction factor searched via gMS is localized far before PEAK_a, around 0.61 s

after WI_a , and could be even 100 [H.U.] lower than $PEAK_a$.

By computing voxel-based aBF values with DV and gMS for each equivalence time, I have attained perfectly identical results for all the examinations, proved by linear regressions with $R^2 = 1$, null intercept ($\sim 10^{-3}$) and unitary angular coefficient (with 10^{-5} accuracy). Even applying gMS with one median value only, results are very good with $R^2 \geq 0.98$ in the 93% of patients, as shown in the histogram of Figure 7.16 (a). In addition, the distributions of slopes (Figure 7.16 (b))

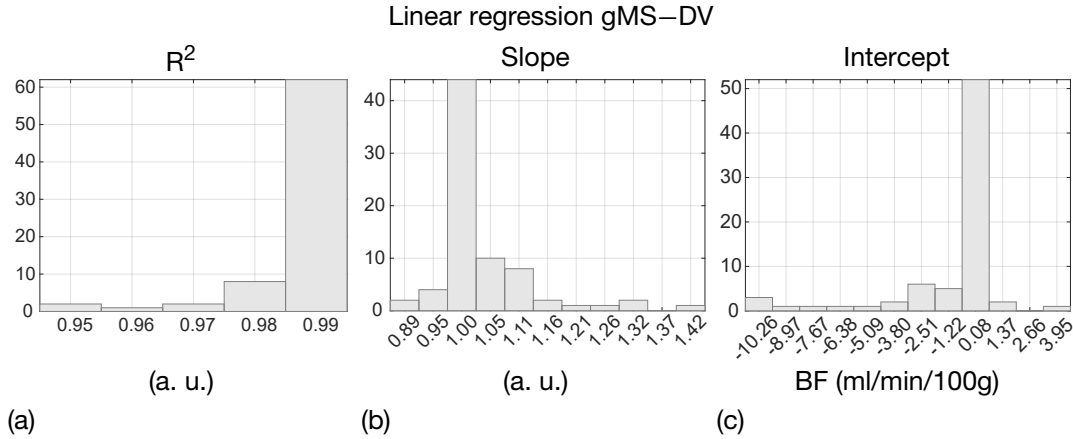


Fig. 7.16: Linear regression between gMS computed in the median of the equivalence times and DV. The histograms of R^2 (a), slopes (b), and intercepts (c) are reported.

and intercepts (Figure 7.16 (c)) are centred around 1.001 and -0.016, respectively, with $IQR=0.05$ for the slopes and $IQR=0.74$ for the intercepts, corresponding to the range $[0.99 \div 1.05]$ for the former and $[-7.4 \cdot 10^{-1} \div 1.2 \cdot 10^{-3}]$ ml/min/100g for the latter. Actually, when considering the gMS computed with respect to the median of the equivalence times, there are some outliers showing higher values of slope and intercept, that also include patients yielding $R^2 < 0.98$.

Colorimetric maps in Figure 7.17 report for the patients C1N38 (a,b) and C16N12 (c,d) the spatial distributions of aBF values. As already shown with the histograms in Figure 7.12, these patient have nearly the most and least spread histograms of equivalence times. Accordingly, for patient C1N38, employing the median of equivalence times in the gMS formulation is expected to worsen the linear correlation between gMS and DV, since the median is less representative of the whole distribution in case of more spread values, which means less representative of the voxel-based state of the system. In fact, the patient C1N38 yields one of the worst value of slope and intercept, being 1.29 and -10.90 ml/min/100g, respectively. Meanwhile, in case of very narrow distributions, as occurs for C16N12, an excellent correlation is confirmed when assuming the median of equivalence times in gMS, with slope equals to one and intercept $\sim 10^5$. However, at a visual-based assessment of colorimetric maps,

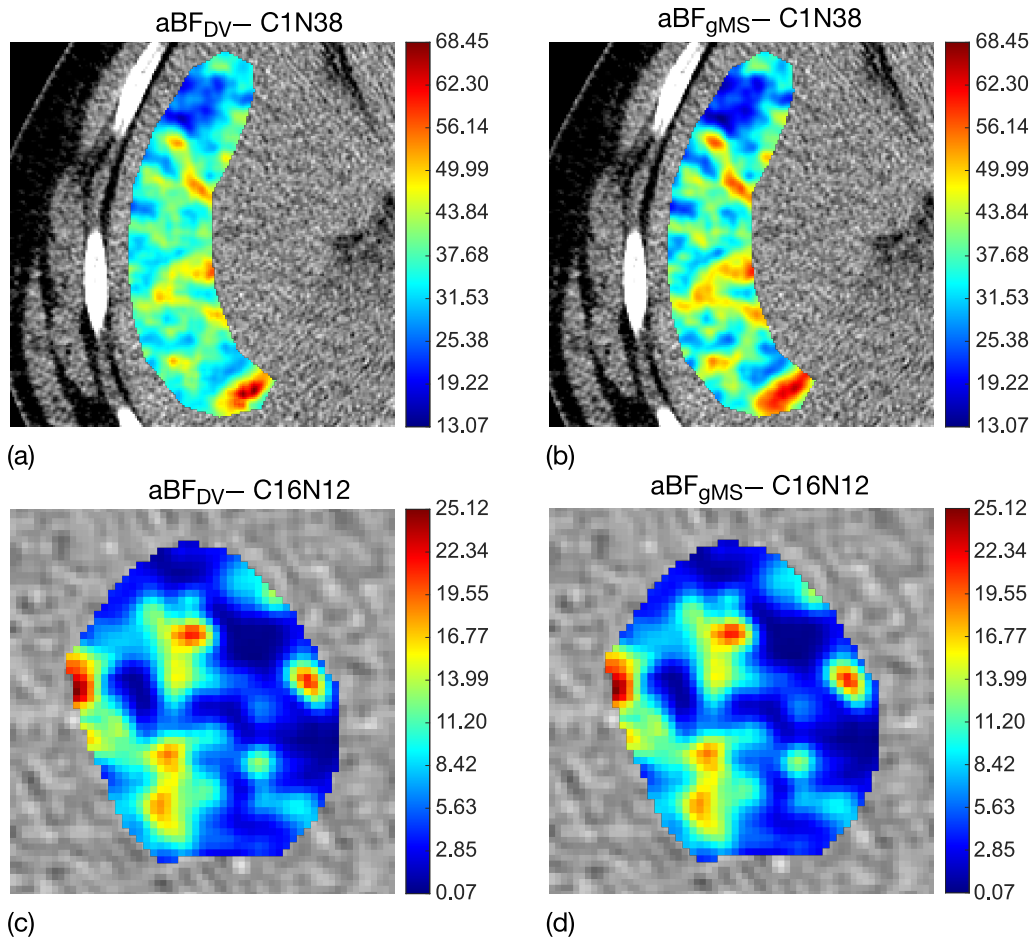


Fig. 7.17: Colorimetric maps of aBF values computed with DV and gMS for patients C1N38 and C16N12, whose equivalence time distributions, reported in Figure 7.12 are nearly the most and least spread of the dataset. Even with slope and intercept departing from one and zero, respectively, thus being corresponding to 1.29 and -10.90 ml/min/100g, respectively, patient C1N38 shows very similar aBF maps (a). Instead, the maps of C16N12 (b) are clearly identical, since having very narrow distributions of equivalence times and perfect linear correlation.

it is worth noting that the areas of high and low perfusion are still matching even in case of worse correlation (C1N38), with aBF values ranging within very similar ranges, that is $[13.97 \div 65.28]$ ml/min/100g for DV and $[15.62 \div 68.45]$ ml/min/100g for gMS. Of course, aBF colorimetric maps are indistinguishable for C16N12, where even the aBF ranges through DV and gMS are equal.

As a concluding remark, the equivalence achieved with gMS and DV is expected to improve the accurateness of aBF values also computed in normal liver. Accordingly, the range of aBF values with both methods in agreement results $[2.89 \div 39.74]$ ml/min/100g for DV and $[2.94 \div 37.07]$ ml/min/100g for gMS, with median values equal to 15.18 ml/min/100g and 15.16 ml/min/100g, respectively.

7.4 A theoretical proof by a single-input model case study

As mentioned in Chapter 6, I have also had the opportunity to work on PROSPeCT, a British clinical trial which enrolled 447 patients with the primary objective of improving the prediction of metastatic disease in patients with colorectal cancer through detecting a hybrid biomarker prognostic of disease free survival, by using jointly clinical parameters and findings from CTP imaging. The availability of PROSPeCT data has been possible in the context of the collaboration – although focused on a different project – between the CVG and Prof. Vicky Goh from the King’s College, London, which led the PROSPeCT trial. My study on reproducibility of MS and DV is based on the hypothesis that MS and DV represent different system’s status due to CA transport delays. In particular, by questioning the validity of MS assumptions, I have formulated the gMS to account for transport delays occurring in a dual input model between the aortic signal which is considered the reference vascular input in place of the hepatic artery, generally hidden in hepatic CTP examinations, and the tissue compartment. Accordingly, the availability of even just two cases of colorectal cancer CTP examinations belonging to the PROSPeCT study, has represented the opportunity to test on a single input model, the hypotheses regarding transport delays hampering the MS in achieving the correct estimates of BF values. In fact, for physiological reasons, the colon-rectum is expected to have minimum delay between input and tissue. Therefore, the MS in its classical formulation and DV are expected to converge towards a unique solution. In case of the colon-rectum, the mesenteric artery is the primary feeding vessel. However, it cannot often be employed for CTP analysis since being very small in CTP images. Therefore, the aortic artery is used in place of the mesenteric one, as it happens for liver perfusion studies. However, differently from what happens with liver, the point where the aortic vessel is captured through colorectal CTP images should minimize any potential delay in reaching the tissue compartment.

Figure 7.18 shows the CTP images of the two CTP examinations, ID12001 and ID12006, which have been analysed. In particular, the two CTP examinations have undergone the same computational workflow set up for the hepatic perfusion study, obviously adapted in case of a single input model. Accordingly, as shown in Figure 7.18, two ROIs outline the aorta – in red – for patients ID12006 (a) and ID12001 (c), and as many ROIs have been drawn on the colorectal cancer (Figure 7.18 (b,c)) – in yellow. Similarly to what explained in Chapter 6, vascular and tissue signals have been fitted through LN and GV models, respectively. MS and DV were applied on each voxel of the colorectal cancer ROIs. The agreement between MS and DV has been first visually assessed through histograms and colorimetric maps of BF values. In addition, to assess to what extent the two measurements were equivalent, a linear regression of BF_{MS} and BF_{DV} has been performed, computing R^2 index,

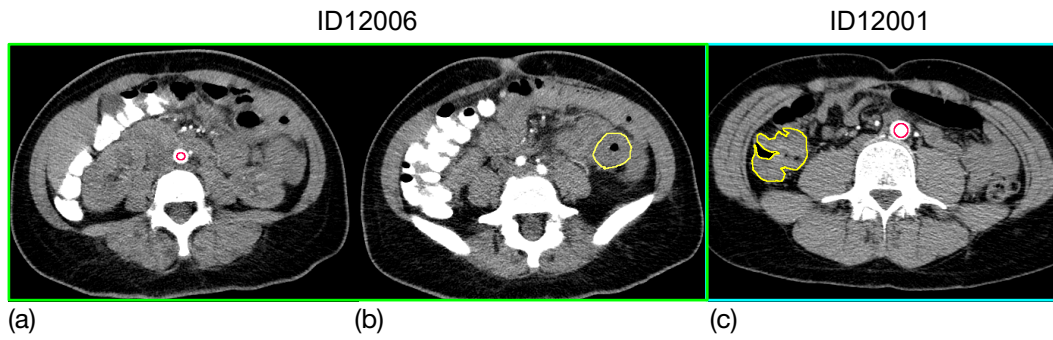


Fig. 7.18: CTP examinations for the perfusion study of colorectal cancer for patients ID12006 (a) and ID12001 (b,c) with the aortic ROIs superimposed in red (a,c), and the tumour ROIs (b,c) in yellow.

slope (m), and intercept (q). Finally, the statistical equivalence of measurements have been measured through Wilcoxon rank-sum test at $\alpha=0.05$ significance level.

Figure 7.19 shows the histograms of BF_{MS} (Figure 7.19 (a)) and BF_{DV} (Fig-

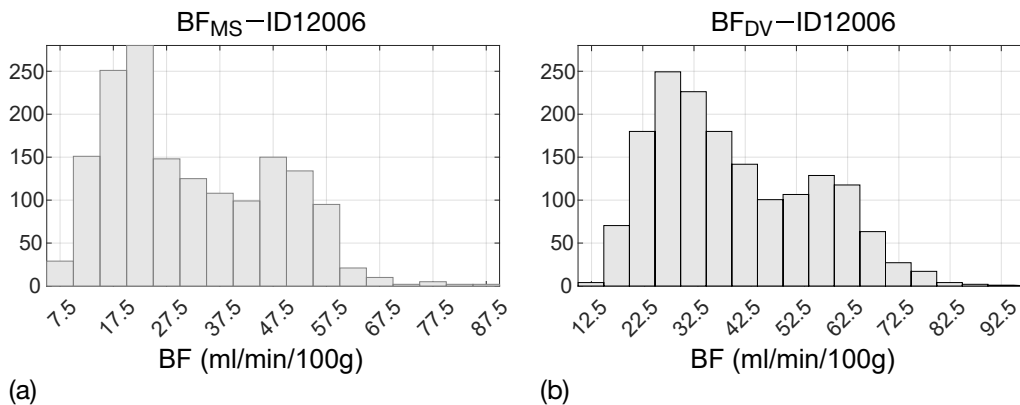


Fig. 7.19: Histograms of BF_{MS} (a) and BF_{DV} (b) achieved for patient ID12006

ure 7.19 (b)) computed for the patient ID12006, where one can appreciate very similar shapes of distributions, although slightly different ranges. In fact, BF_{MS} falls within $[7.26 \div 86.19]$ ml/min/100g, whilst BF_{DV} ranges within $[16.10 \div 94.77]$ ml/min/100g. Despite these differences, the colorimetric maps in Figure 7.20 highlight very similar spatial distributions of low- and high- perfused areas, although with more emphasized high-perfusion patches in DV map.

Actually, the linear regression in Figure 7.21 shows a slight dispersion of BF values, with $R^2 = 0.98$, $m=0.996$, very close to unit, and a residual bias, $q=7.829$ ml/min/100g, which is the major cause thwarting the exact numerical equivalence of BF_{MS} and BF_{DV} . Accordingly, Wilcoxon rank-sum test yields a non-significant p -value ($p=0.32$), meaning equivalence of BF_{MS} and BF_{DV} measurements, if com-

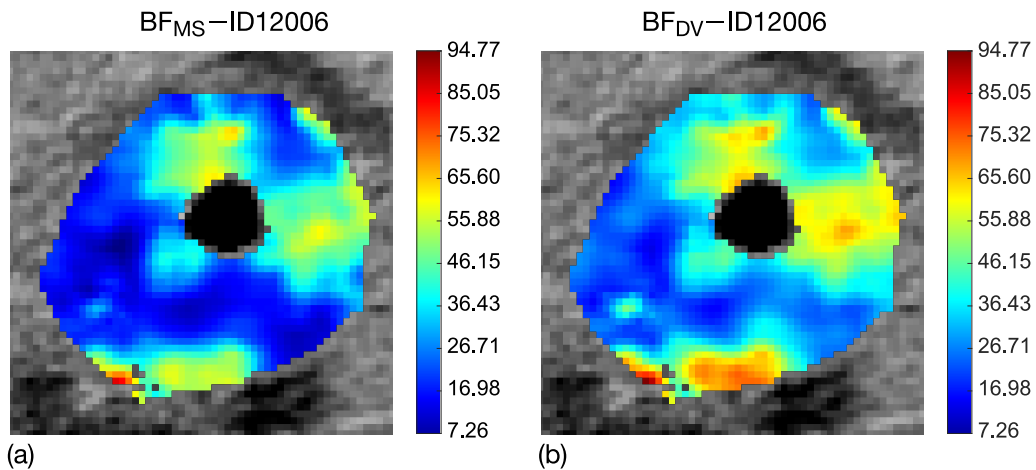


Fig. 7.20: Colorimetric maps of BF_{MS} (a) and BF_{DV} (b) for patient ID12006

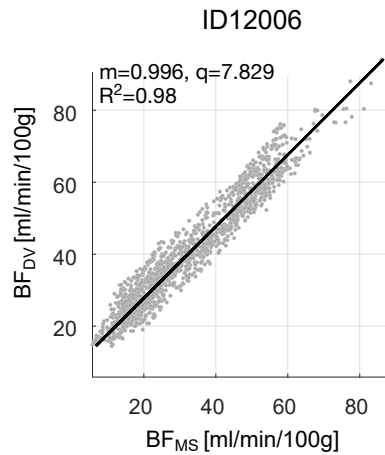


Fig. 7.21: Linear regression of BF_{MS} and BF_{DV} for patient ID12006 with $R^2=0.98$, $m=0.996$, and $q=7.829$.

paring BF_{MS} and BF_{DV} unless 8 BF units (8 ml/min/100g) equivalent to the bias obtained in the linear regression. Basically, this means that even without a statistical equivalence, the two measurements of BF values with MS and DV can substantially be clinically equivalent, being the difference of BF values less than the 10%. In fact, the almost unit slope of linear regression (Figure 7.21) confirms that MS and DV are representing the same status of the system, although with a residual numerical bias.

Figure 7.22 reports the histograms of BF_{MS} (a) and BF_{DV} (b) referring to patient ID12001.

As one can see just from the distributions, this patient yields much more sim-

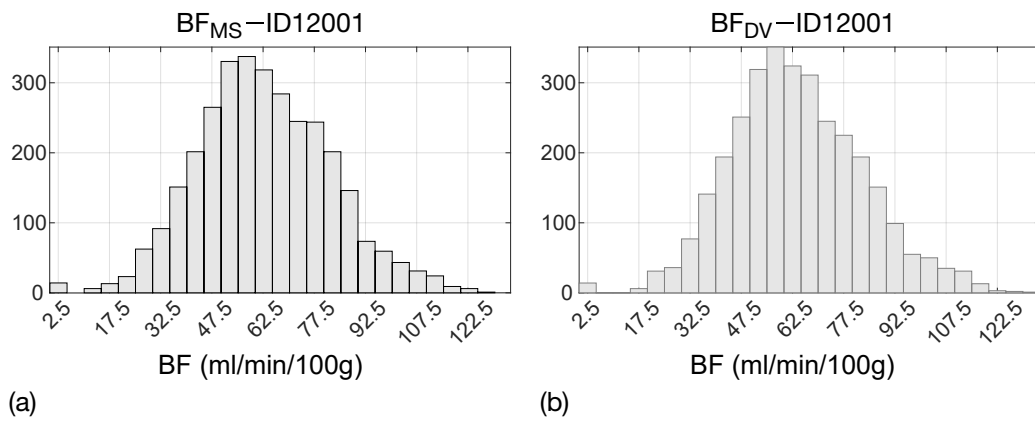


Fig. 7.22: Histograms of BF_{MS} (a) and BF_{DV} (b) achieved for patient ID12001

ilar BF values than patient ID12006, with BF_{MS} ranging within $[17.61 \div 120.76]$ ml/min/100g and BF_{DV} belonging to $[20.74 \div 123.99]$ ml/min/100g. In addition, the excellent agreement is confirmed through the colorimetric maps reported in Figure 7.23. The spatial distribution of BF_{MS} and BF_{DV} are basically indistin-

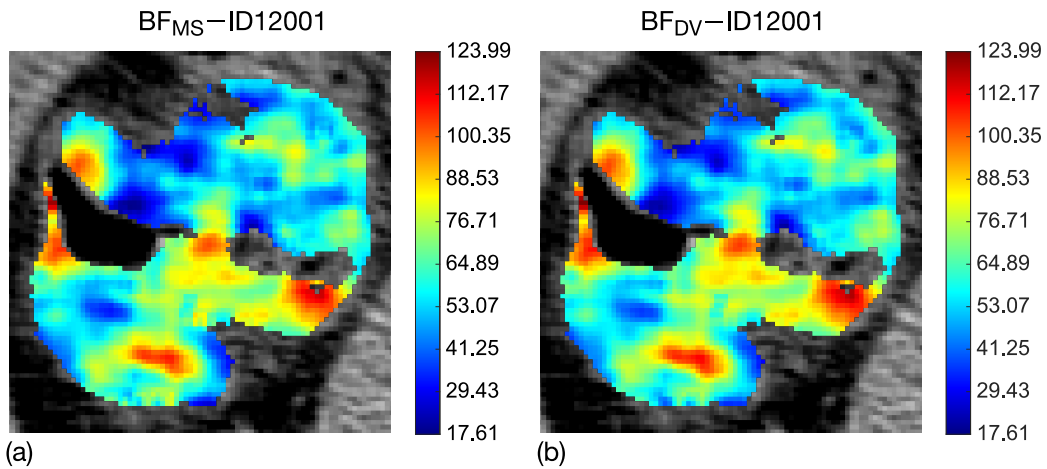


Fig. 7.23: Colorimetric maps of BF_{MS} (a) and BF_{DV} (b) for patient ID12001

guishable, with identical patterns of low- and high- perfused tumour regions. The equivalence of BF measurements is even proved by the linear regression of BF_{MS} and BF_{DV} , shown through the scatter plot in Figure 7.24, which reports $R^2 = 0.99$, with $m=1.017$, and $q=-0.035$. Of course, this case confirms the clinical equivalence of the BF measurements achieved independently with MS and DV. In addition, it is yet more meaningful if considering that the statistical equivalence of BF values is attained as well, as confirmed by the Wilcoxon rank-sum test yielding $p=0.06$.

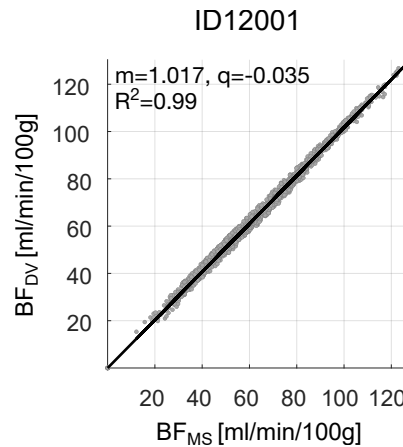


Fig. 7.24: Linear regression of BF_{MS} and BF_{DV} for patient ID12001 with $R^2=0.99$, $m=1.017$, and $q=-0.035$.

7.5 Assessing perfusion after Sorafenib and SIRT treatments

As explained throughout this Chapter, the methodological workflow set up for perfusion analysis has allowed reaching three main goals. First, it has enabled the formulation of the modelling conditions under which two independent perfusion methods can represent the same status of the system. Second, it has shown how to yield reproducible BF values on a dataset of 75 patients undergoing dual-input hepatic CTP analysis. Third, in a case study involving two CTP examinations of colorectal cancer, modelled through a single input compartment, the theoretical hypothesis have been proved. Hence, in this last Section, I present the preliminary results of a clinical application study, in which perfusion parameters have been employed to investigate the effects of two different oncologic treatments. In particular, this data refer to SARA, currently the most important CTP perfusion study which have been conducted from 2011 to 2017, in order to compare efficacy and safety of Sorafenib, the recommended treatment for patients with advanced HCC, and selective internal radiotherapy (SIRT) performed with yttrium-90 (^{90}Y) resin microspheres. SARA study enrolled 25 French Centres and 467 patients which have been randomly assigned to Sorafenib (400 mg twice daily) or SIRT (2-5 weeks after randomization) treatments. Initially, 222 patients were assigned to Sorafenib group and 237 ones to SIRT group. During the course of the study, some patients have been shifted from one group to the other one due to clinical reasons related to patient health status. Ultimately, 239 patients have been treated with Sorafenib, and 228 with SIRT. Included patients have undergone baseline CT and CTP examinations, a first follow-up one month later, and every three months subsequently. At each follow-up, both CT and CTP examinations have been repeated. Findings

of SARAH study, exclusively based on the analysis of CT examinations and clinical parameters, were published in 2017 [4] and report overall survival not significantly differing between the two treatments.

In this preliminary study, I have analysed for the first time the CTP examinations of SARAH study, limited to Centre 1 (the same Centre 1 as PIXEL). Centre 1 has totally enrolled fifty-four patients, twenty-six treated with Sorafenib and twenty-eight undergoing SIRT. Due to the unavailability of some CTP examinations, presence of imaging artefacts, non-visible portal vein in CTP images, some patient were excluded from perfusion analysis of Centre 1. Ultimately, I have considered sixteen patients belonging to Sorafenib group, of which fifteen with one month-follow up (1m-Fup) and one-only with three month-follow up (3m-Fup), and eighteen patients of SIRT group, of which fourteen with 1m-Fup and four with 3m-Fup. Perfusion analysis has been performed adopting the same methodological workflow used for the PIXEL study. In this case, perfusion measurements of aBF, pBF, BF, and HPI have been performed with the DV method only to achieve a better precision of HPI estimates. In each Centre, perfusion measurements have been compared between baseline and Fup, by considering the signed variation of median and MAD values. Absolute variations of perfusion parameters greater than one percentage unit for HPI and BF unit for aBF, pBF, and BF, have been considered for detecting an increase or decrease of the perfusion parameters. Otherwise, parameters have been considered stable.

From the clinical side, Sorafenib is an antiangiogenic drug and, as such, it is expected to have the primary impact on tumour perfusion, much more than SIRT can do, especially at 1m-Fup. Accordingly, it can improve patient prognosis and overall survival. At the same time, SIRT is the treatment holding the highest expectations within the medical community in order to improve advanced HCC prognosis even making the surgery possible. Accordingly, the analysis of tumour perfusion may result more accurate in characterizing the effects of the two treatments since focusing on dynamic tumour features related to angiogenesis, known to be the early hallmark of tumour variations (as explained in Chapter 3).

Table 7.6 resumes the variations of perfusion parameters in the Sorafenib and SIRT groups, by considering the three possible cases of Base>Fup, Base<Fup representing the decrease and increase of perfusion parameters, respectively, and Base≈Fup indicating stable parameters. In addition, the variations $|\Delta|_{\text{Base}>\text{Fup}}$ and $|\Delta|_{\text{Base}<\text{Fup}}$ are reported on average. For each patient the variation of $M \pm \text{MAD}$ of all perfusion parameters separately is also shown in Figures 7.25 and 7.26 for Sorafenib and SIRT groups, respectively.

As one can notice from Table 7.6, for almost all parameters the absolute variations of parameters between baseline and Fup are quite similar, in both Sorafenib and SIRT groups. A partial exception is represented by the BF of SIRT group, whose

Table 7.6: Resume of variations of the four perfusion parameters, aBF, pBF, BF, and HPI, for Sorafenib and SIRT groups.

		Base>Fup	$ \Delta _{\text{Base>Fup}}$	Base<Fup	$ \Delta _{\text{Base<Fup}}$	Base \approx Fup
Sorafenib	aBF	12	4 \pm 0	2	4 \pm 3	2
	pBF	12	19 \pm 4	4	22 \pm 6	–
	BF	13	23 \pm 5	3	20 \pm 3	–
	HPI	6	7 \pm 0 (%)	8	11 \pm 0 (%)	2
SIRT	aBF	5	7 \pm 1	12	9 \pm 2	2
	pBF	6	26 \pm 4	11	25 \pm 7	–
	BF	5	51 \pm 6	11	38 \pm 5	2
	HPI	10	4 \pm 0 (%)	7	7 \pm 1 (%)	1

decrease in 5 patients is slight higher (51 \pm 6 ml/min/100g) than its increase (38 \pm 5 ml/min/100g) referred to 11 patients. Sorafenib yields in a high number of patients the decrease of aBF (12 patients), pBF (12 patients), and BF (13 patients) values, whilst the opposite tendency is obtained for SIRT group, where the highest number of patients registers an increase of perfusion parameters, 12 patients for aBF, and 11 ones for pBF and BF. Based on the characteristics of angiogenesis, and according to the dual input physiology of the liver, therapy-responder patients should show a decrease of the HPI, following the regularization of the vascular network, major altered in the arterial circulation. Although the Sorafenib group shows a reduction of most of perfusion parameters, the HPI does not show any relevant regularization, since almost the same number of patients show either a decrease (6 patients) or an increase (8 patients) of HPI values. Basically, the reduction of aBF, observed in patients treated with Sorafenib is not enough to cause the consequent reduction of HPI. In fact, the decrease of aBF is limited on average to 4 \pm 0 ml/min/100g. Moreover, in the two patients showing stable HPI values (C1N11 and C1N26), the aBF and pBF values slight decrease proportionally (Figure 7.25). On the contrary, in the two patients showing stable aBF values (C1N17 and C1N25), pBF values strongly vary, thus leading the HPI to increase in C1N25 and decrease in C1N17 (Figure 7.25 (a)). In addition, one can notice that the variation of aBF, pBF and BF values in Sorafenib group are quite proportional with each other, thus leading the graph Δ_{BF} in Figure 7.25 (b) to be inversely correlated with Δ_{HPI} in Figure 7.25 (a).

As said above, SIRT group shows an opposite tendency to Sorafenib for all parameters. Actually, the absolute variation of aBF values is nearly twice than that occurring in Sorafenib group in both increase (9 \pm 2 ml/min/100g) and decrease (7 \pm 1 ml/min/100g). The strong reduction of BF observed in SIRT group causes a high increase of HPI on average, equals to the 7 \pm 1 %. By comparing the increases of aBF, pBF, and BF values in SIRT and Sorafenib groups, one can notice a high variation (e.g., 38 \pm 5 ml/min/100g against 20 \pm 3 ml/min/100g for BF) in SIRT, also

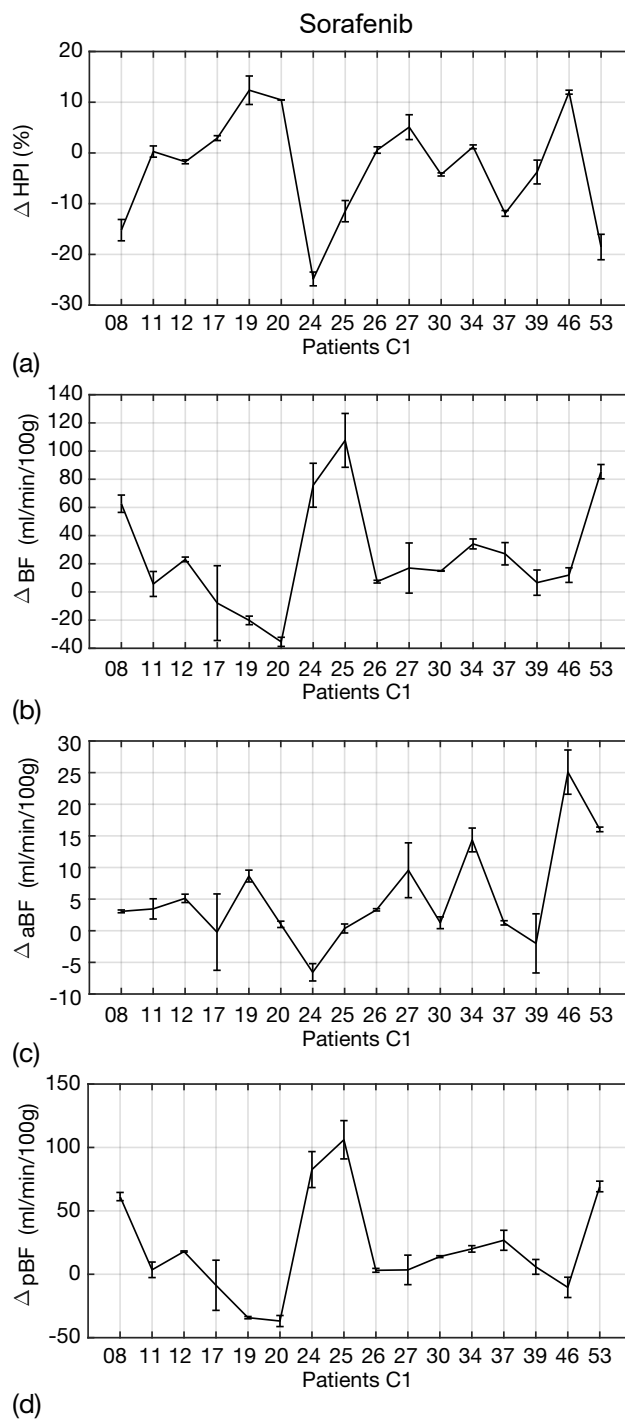


Fig. 7.25: Variations in Sorafenib group of $M \pm MAD$ values of HPI (a), BF (b), aBF (c), and pBF (d).

occurring in more patients, than in Sorafenib group. This leads the HPI in the SIRT group to show a decrease, although limited to 4 ± 1 %, for the highest number

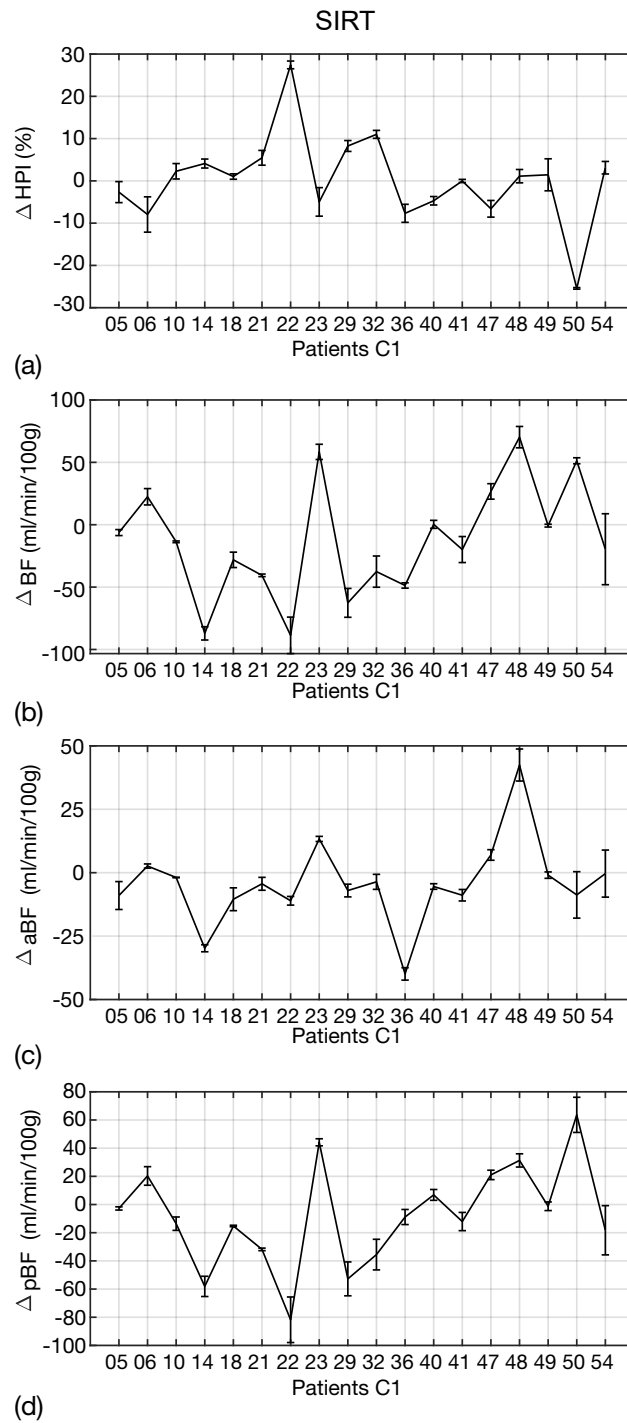


Fig. 7.26: Variations in Sorafenib group of $M \pm MAD$ values of HPI (a), BF (b), aBF (c), and pBF (d).

of patients (10 patients), while just 7 patients show an increased HPI. Finally, the only patient, C1N41 showing a stable HPI shows proportional variations of the BF

contributions. The patient C1N49 shows stable both BF and aBF values, with a very small variation of pBF, whilst C1N49 shows a stable BF value due to the balanced decrease and increase of aBF and pBF, respectively.

Overall, these preliminary findings show a greater impact of Sorafenib in regularizing perfusion parameters, although not yet visible in HPI values at 1m-Fup. Of course, a long-term analysis is needed to really verify the efficacy of Sorafenib treatment and its correlation to overall survival, as well as to confirm the lower performance of SIRT group. As a concluding remark, it is worth noting that perfusion analysis is confirmed, once more, to provide very useful information in characterising dynamic tumour properties, and improving the comprehension of treatment effects on tumour vascular network.

Chapter 8

Radiomics: methodological and clinical studies

This Chapter presents my contributions in the field of radiomics, starting with a methodological study (Sect. 8.1) aiming at assessing the reproducibility of FO and second order texture features on RCC imaged with CT. In particular, this study has been conducted during the three-month period which I have spent at the Department of Radiology, Addenbrooke's Hospital, University of Cambridge. Then, being aware of major advantages and limits of radiomic features widely adopted in the literature, I propose a novel approach to compute imaging features (Sect. 8.2), which have been exploited in clinical radiomic studies presented later on. In particular, Sect. 8.3 shows a discrimination study, carried out in collaboration with the Department of Radiology, S. Orsola-Malpighi Hospital, University of Bologna, to assess neoadjuvant chemo-radiotherapy response in LARC. Sect. 8.4 reports two studies focusing on prediction of PCa aggressiveness, developed in partnership with Diagnostic Imaging Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori "D. Amadori" – IRST s.r.l. (hereinafter IRCCS IRST). The first is a preliminary staging of PCa into four risk levels, followed by a classification study to predict csPCa.

8.1 A reproducibility study of first order and texture features

This study has been carried out during the three-month period which I spent as visiting researcher at the Department of Radiology, Addenbrooke's Hospital, University of Cambridge, under the supervision of Prof. Evis Sala. In particular, between February and April 2020, I have frequented the Radiogenomic and Quantitative Imaging Group, and I was involved in a radiomic study on RCC. Actually, due to their internal policy limiting the access to clinical data to short-term external researchers, my contribution has been focused on methodological aspects.

As already discussed in Chapter 5, the number of radiomic studies has exploded during the last years and the reproducibility of radiomic features is still an open issue (Sect. 5.11). The plethora of software packages available for radiomic analyses has highlighted the urgent need for standardisation of methodology and measurements [229], [183]. Many factors are known to induce variability in radiomic features including noise [170], heterogeneous voxel size [230], and ROI segmentation [231], besides tumour phenotype [232]. Despite its importance, only few studies directly address robustness and reproducibility of radiomic features. Accordingly, since I had to focus on RCC radiomics, I have considered to perform a reproducibility study of the most adopted radiomic features. To the best of my knowledge, this is the first work assessing robustness of FO and 2D and 3D second order texture features in CT imaging of RCC and contralateral normal kidney (CK), by addressing three types of perturbations induced by Added White Gaussian Noise (AWGN, hereinafter N), different voxel-size (V) and varying ROI (R). In addition, I have performed a comparative analysis to select the best interpolation method to be preliminarily applied, if needed, before any feature extraction procedure. Finally, results can provide practical operating guidelines to choose the proper voxel size in case of datasets with heterogeneous in-plane resolutions and to aggregate information derived from GLCMs, this contributing to improve standardisation of radiomic studies.

8.1.1 Data preparation

Patient images

This study has included 98 patients with RCC imaged at a single institution. CT acquisition parameters are provided in Table 8.1. Images have been acquired with Siemens SOMATOM Definition AS/AS+ CT scanners, with iterative reconstruction kernel I30f\3. Scan resolution ranges from square voxel spacing $v_s = 0.541$ mm to $v_s = 0.957$ mm, with $v_z = 5$ mm-slice thickness. Mean values and ranges of tube voltage and exposure are 109 [100,140] KVp and 166 [137,535] mAs, respectively. Automatic tube voltage selection (CARE kV) and current modulation (CARE Dose) have been employed to optimize the dose to patients, resulting in a mean and range of 109 [100,140] KVp and 166 [137,535] mAs, respectively. Image series of the corticomedullary and nephrographic phase, acquired at 35 and 100 seconds after the administration of the intravenous contrast agent (Omnipaque 300 mg I/ml, GE Healthcare) have been included for 28 and 70 patients, respectively. This retrospective study has been approved by the Health Research Authority (HRA), University of Cambridge and Cambridge Research and Development (R&D) department that waived the written informed consent.

Parameter	CT
Number of scans	98
Scanners	Siemens SOMATOM AS/AS+
Tube voltage (kVp)	109 [100,140]
Exposure (mAs)	166 [137,535]
Reconstruction kernel	l30f\3
Voxel spacing ([x,y]-axes; mm)	0.740 [0.541,0.957]
Voxel spacing (z-axis; mm)	5
Image noise (σ ; HU)	4 [2.9, 5.9]
SNR (dB)	42 [37, 45]

Table 8.1: CT image acquisition parameters

Segmentation

ROIs of RCC and CK have been semi-automatically outlined using the Microsoft Radiomics Tool (Version 1.0.30558.1, project InnerEye, <https://www.microsoft.com/en-us/research/project/medical-image-analysis/>) by a medical doctor and clinical researcher with three years experience in renal imaging. The structures have been segmented in all slices at the original scan resolution. In particular, CK and RCC have been segmented on the first and last slice where they were visible and contours for every other slice have been interpolated using an image intensity based distance metric. Manual corrections have been applied to sub-optimally segmented slices, leading to an iterative re-calculation of the remaining interpolated slices. The segmentation of CK has included the renal cortex and medulla but not the the collecting system and hilar fat.

8.1.2 Image processing

Image processing for feature robustness analysis has been performed according to the workflow reported in Figure 8.1, where the main steps are outlined. For each block of the flowchart more details are provided in the corresponding subsequent Section. After CT image acquisition and segmentation, an image interpolation procedure has been needed to standardize the different resolutions of the dataset. To this purpose, three well-known interpolation methods have been analysed and compared, to find the method best preserving the statistical properties of the original images, to be used for all the subsequent steps. The comparison has been carried out by exploiting the Enhancement Measurement Error-Image Quality (EME-IQ) [233] score that measures local image contrast, amongst the most important information cues, so that higher EME-IQ values indicate the interpolation method better preserving sharpness and edges. After that, three different perturbations have been

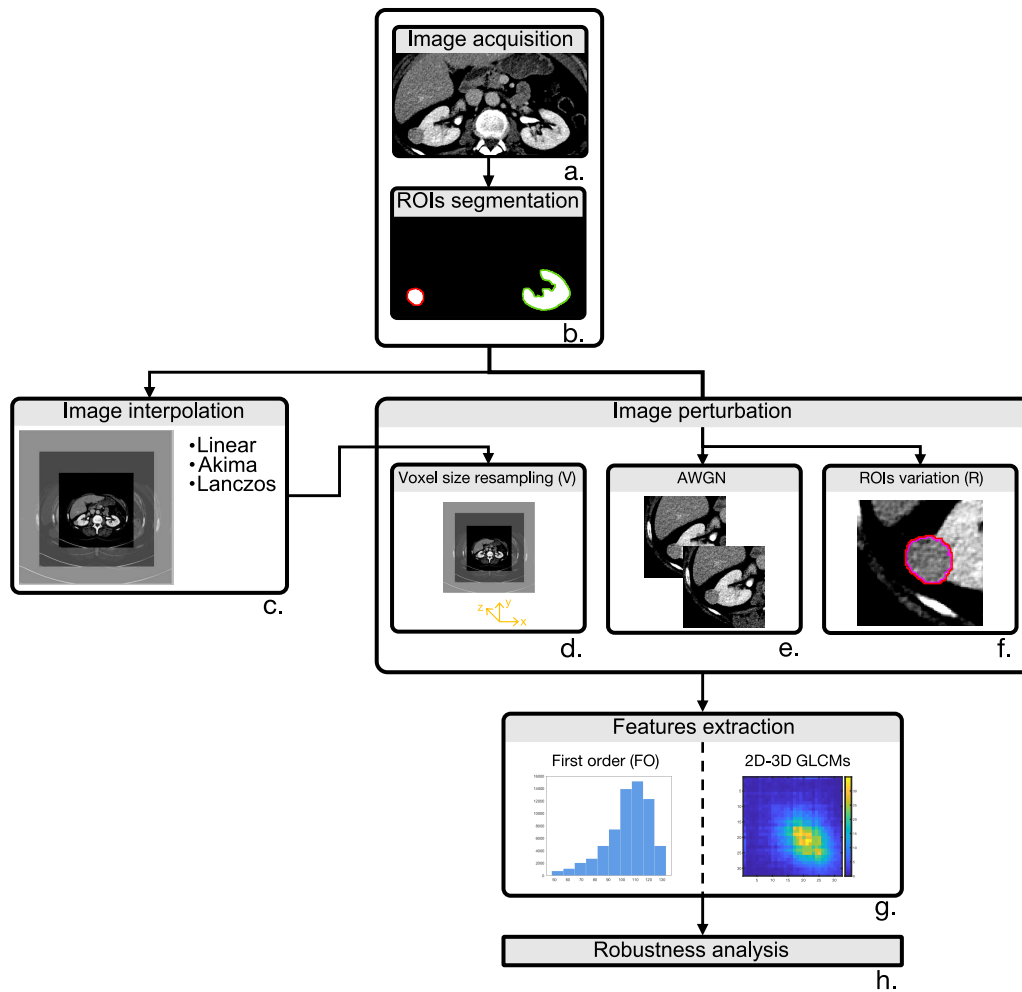


Fig. 8.1: Workflow of CT image processing for feature robustness analysis. First, CT images have been acquired (a) and ROIs segmented (b). Then, one out of three resampling methods has been chosen based on the highest EME-IQ score (c), then CT images have been perturbed by changing resolution (d), noise addition (e), and ROI variation (f). Finally, FO, and 2D and 3D texture features based on GLCMs (g) have been extracted from original and perturbed images and robustness analysis has been performed (h).

applied to whole images and ROIs. The first perturbation considered has been the AWGN, added to the original CT images using the same standard deviation (σ) as the original images. As the second perturbation, the original CT slices have been resampled and interpolated along the z directions. Finally, to simulate inter-reader variability, ROI enlargement and shrinking have been considered [234]. Lastly, 3D FO and 2D-3D texture features (from GLCMs) have been computed on the original and perturbed CT images, and their reproducibility has been studied.

Image interpolation

Different scan voxel size is a well-known source of variability for radiomic features in a heterogeneous image dataset that should be taken into account when performing radiomic studies [230]. Hence, applying a resampling procedure is needed, to achieve one voxel size [200]. However, any resampling method relies on interpolation techniques that may potentially alter latent image properties, thus affecting the reproducibility of the features extracted from the resampled image and from the original [191]. In radiomic analyses, the interpolation method must be conceived for quantitative imaging and preserve the original properties of the CT GL distributions, rather than to simply yield visually pleasant images. In this work, I have analysed and compared three well-known interpolation methods, that is, linear interpolation, largely employed even in most radiomic studies [200] thanks to its low complexity and computational cost, Akima cubic Hermite spline [235] and Lanczos [236] interpolation, the latter mostly used in computer vision and quantitative imaging. The comparison has been performed by resampling the original CT images at the best ($v_s^b = 0.541$) and worst ($v_s^w = 0.957$) original voxel size, and rounding the GLs to the nearest integer. When employed for visualization purposes, the different interpolation methods are assessed through a forward-backward process, which compares the quality of original and restored image [237]. Here, I have measured directly the quality of the interpolated images and adopted the EME-IQ score, a Non-Reference IQ measure that quantifies the level of local contrast [233]. For each patient, EME-IQ scores have been computed and averaged on all CT slices, in both upsampling and downsampling. Then, the three interpolation methods have been ranked according to their EME-IQ score. I have chosen the method that resulted the best score for the highest number of patients for both upsampling and downsampling, and adopted in the subsequent steps of V image perturbation.

Image perturbation

Additive noise (AWGN)

CT images are known to be mostly affected by quantum noise, arising from the effects of the variability of electronic density of tissue voxels [238], statistically represented by a random Gaussian process [239]. Therefore, I have perturbed CT images by AWGN where, for each patient, σ is given by the average of the standard deviation of each slice, estimated according to the method proposed by Ikeda *et al.* [240].

Changing voxel size

Original CT images consisted of anisotropic voxels, with different in-plane resolutions, but one slice spacing that was on average one order of magnitude bigger. As regards slice resolution, I have investigated three different resampling strategies, that are: (i) upsampling the whole dataset to v_s^b , (ii) downsampling to v_s^w , (iii) resampling at the median resolution ($v_s^M = 0.741$ mm). Although working with isotropic voxels would be advisable, resampling to the z -axis resolution for isotropy would introduce an unrecoverable signal loss. Therefore, besides keeping the original scan resolution ($v_z = 5$ mm), I have limited the highest resolution to $v_z = 1$ mm, exploring intermediate values, with 1-mm steps. In total, combining three voxel sizes with five slice thicknesses, I have tested 15 different voxel sizes.

Segmentation perturbation

One of the causes affecting the clinical reliability of radiomic features as predictive or prognostic biomarkers is the lack of reproducibility of quantitative measurements, depending on the variability of intra- and inter-observer ROI segmentation [181]. Similarly to the work in [234], I have simulated such a variability, considering volume variations up to 20%, by ROI enlargement and shrinking. Actually, while dilating ROIs means including different tissues, making a mistake by defect implies missing some tissue of the same type. Consequently, ROIs have been shrunk by 10%, 15% and 20% or dilated by 10%. This procedure has been carried out through binary morphological dilation and erosion, with a 3×3 pixel square structuring element, according to a pixel-based random contourization procedure. In this regard, the percentage of pixels exceeding the desired volume variation, has been randomly restored to the original value.

8.1.3 Feature extraction

In this study, I have included 13 FO and 19 GLCM features computed in both 2D (GLCM2D) and 3D (GLCM3D), since no agreement exists yet on how to aggregate GLCM information to extract single representative features [183]. Hence, GLCM2D have computed in four directions, $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, and GLCM3D have been extended in 13 directions [241], with five odd distances, from $\delta = 1$ to $\delta = 9$. The features have been extracted after intensity-based outlier removal performed on CT images at the 2.5% threshold at both left and right tails of GL distributions. Based on a preliminary analysis of our CT dataset, the commonly used choice [202], [242] of 32 quantization levels has been adopted for the GLCM computation. GLCMs have been also symmetrized and direction-weighted. In GLCM2D, features have

been first computed on each slice and then averaged. In all, 108 radiomic features have been computed on the original and perturbed CT images, for RCC and CK separately. Table 8.2 lists all FO and GLCM features.

First order	GLCM2D-3D
mean (m)	autocorrelation ($autoc$)
median (M)	correlation ($corr$)
skewness (s)	cluster prominence ($cprom$)
maximum value (max)	homogeneity ($homom$)
m of last decile ($m90th$)	maximum probability ($maxpr$)
M of last decile ($M90th$)	contrast ($contr$)
standard deviation (std)	cluster shade ($cshade$)
M absolute deviation (MAD)	variance ($sosvh$)
interquartile range (iqr)	dissimilarity ($dissi$)
local coefficient of variation (ICV)	energy($energ$)
uniformity (u)	entropy ($entro$)
entropy (e)	difference variance ($dvarh$)
kurtosis (k)	difference entropy ($denth$)
	information measure of $corr$ ($inf1h$)
	inverse difference normalized ($indnc$)
	inverse difference moment normalized ($idmnc$)
	sum average ($savgh$)
	sum variance ($svarh$)
	sum $entro$ ($senth$)

Table 8.2: List of FO (n=13) and GLCM2D-3D (n=19) features.

8.1.4 Robustness analysis

All the extracted features have been analysed for both RCC and CK and robustness has been assessed using the ICC(1,1) with 95% confidence interval (CI) [234]. Radiomic features have been considered as being robust (r) if ICC 95% CI ≥ 0.90 , non-robust (nr) if CI < 0.90 , and with indeterminate robustness (i) otherwise (i.e., with 0.90 strictly included in CI). In total, 29 perturbations have been assessed, one coming through N, 24 combinations of V, and four from R, as detailed in Table 8.3.

First, the robustness of all radiomic features together has been investigated against all perturbations, to have an overview of features behaviour depending on the tissue phenotype (i.e., RCC or CK) only. Mean percentage of r , nr , and i features have been reported for each perturbation type. Moreover, the proportional contribution given by each feature class to the global robustness has been investigated, together with the contribution of the single features. In practice, robustness has been assessed (i) for all feature classes (i.e., FO, GLCM2D, GLCM3D) against

all image perturbations, (ii) for each feature class against each perturbation type (i.e., N, V, R), and (iii) for each feature against all image perturbations.

N°	Type	Perturbation	Description
1	N	N	AWGN
15		$v_s - v_z$	15 voxel-sizes by combining v_s^b, v_s^M, v_s^w with $v_z=[1\div 5]$
1	V	V	Global assessment of the all 15 voxel-sizes
1		v_s^b	Resolution v_s^b kept fixed and $v_z=[1\div 5]$
1		v_s^M	Resolution v_s^M kept fixed and $v_z=[1\div 5]$
1	V	v_s^w	Resolution v_s^w kept fixed and $v_z=[1\div 5]$
1		Z1	Resolution $v_z = 1$ kept fixed for all v_s^b, v_s^M, v_s^w
1		Z2	Resolution $v_z = 2$ kept fixed for all v_s^b, v_s^M, v_s^w
1		Z3	Resolution $v_z = 3$ kept fixed for all v_s^b, v_s^M, v_s^w
1		Z4	Resolution $v_z = 4$ kept fixed for all v_s^b, v_s^M, v_s^w
1		Z5	Resolution $v_z = 5$ kept fixed for all v_s^b, v_s^M, v_s^w
4	R	R+10	Dilation, volume variation equals to +10%
		R-10	Erosion, volume variation equals to -10%
		R-15	Erosion, volume variation equals to -15%
		R-20	Erosion, volume variation equals to -20%

Table 8.3: Descriptions of the 29 perturbations assessed

Finally, this study has assessed the real need for having GLCMs computed at multiple δ distances, because of their known high correlation. To this end, I have performed this analysis by adopting the voxel size resulting as the most reliable from analyses at step (i). In conclusion, the correlation of features computed at all δ has been measured through the linear Pearson coefficient (ρ) and the statistical significance of the differences has been assessed by the ANOVA test ($p\text{-value} \leq 0.001$).

8.1.5 Experimental results

Assessing the robustness of radiomic features has become necessary to determine feature reproducibility before translating predictive and prognostic radiomic biomarkers into clinical practice. Feature robustness depends on the tumour phenotype and is not generalizable [232], hence, this study focuses on the need for analysing feature robustness on RCC in CT. In addition, the analysis is extended to CK to determine which features might be robust enough to assess, for instance, diffuse renal diseases. As first, this study offers an in-depth analysis of three known interpolation methods aiming at supporting researchers in choosing the most appropriate one when resampling CT images. Then, it analyses the robustness of radiomic features against some of the most frequent sources of variability, which are noise, heterogeneous scan voxel size and varying segmentation.

Figure 8.2 reports the comparison of Linear, Akima and Lanczos interpolation

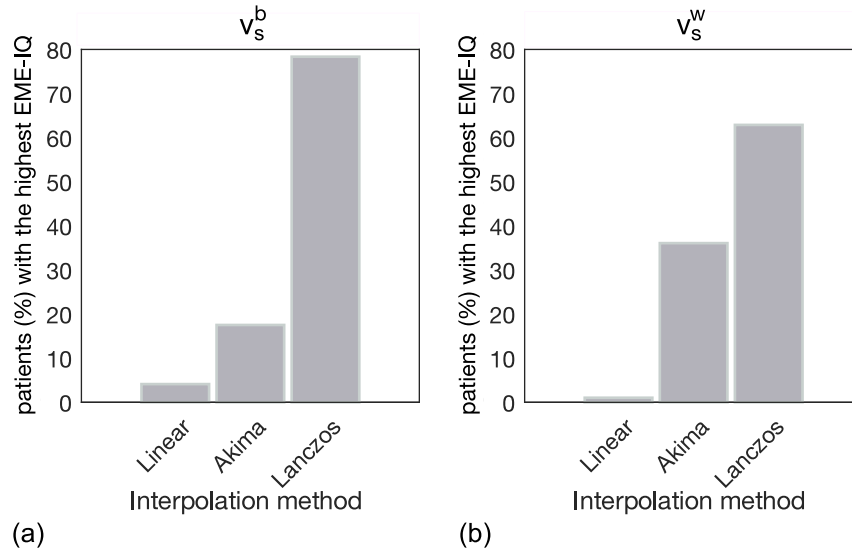


Fig. 8.2: Comparison of linear, Akima, and Lanczos interpolation methods, based on the percentage of patients receiving the highest EME-IQ score, for both upsampling (a) and downsampling (b) respectively at the best (v_s^b) and the worst (v_s^w) resolutions of the CT image dataset. The Lanczos method yielded the highest EME-IQ score in 78% and 63% of patients, respectively.

methods based on the EME-IQ score. Results show that Lanczos interpolation outperforms the other methods in both upsampling and downsampling procedures. In fact, when resampling at v_s^b (Figure 8.2 (a)), the linear method achieves the best result for only 4% of patients, Akima for 18% and Lanczos for 78% of patients. Moreover, this ranking is confirmed when resampling at v_s^w (Figure 8.2 (b)), where linear and Akima methods reach 1% and 36%, respectively, whilst Lanczos still proved to be the best one for 63% of patients. As one can notice, results emphasize the poor performance of the linear interpolation, although being probably the method most widely employed in radiomic studies, and even suggested by the IBSI [183]. More specifically, if on the one hand resampling based on linear interpolation improves visual images' perception, on the other hand it smooths tissue edges and texture variation, thus limiting quantitative information [243]. By comparing this study with the state of the art, I have to say that some authors are aware of the importance of the interpolation methods, which may influence feature robustness. For instance, Whybra *et al.* [244] carry out a comparison of feature robustness, after linear and spline interpolations. The authors conclude that the two methods are equivalent since no difference exist in terms of feature stability, albeit in the presence of large numerical variations. However, although those features may result as being reproducible, this does not ensure that the features are correctly representing

the original CT image information. For this purpose, based on my finding I can recommend a preliminary analysis to assess that the resampling procedure does not affect the properties of the GLs distribution.

It is also worth noting that the upsampling procedure, although adding artificial information, improves the original image quality – where the mean EME-IQ score is 2.87 ranging between [1.36,6.72] – with all methods considered, and according to their rank. In fact, EME-IQ=2.99 (with EME-IQ=[1.61÷6.48]) for linear, EME-IQ=3.09 (with EME-IQ=[1.68÷6.71]) for Akima, and EME-IQ=3.12 (with EME-IQ=[1.70÷6.80]) for Lanczos. Even more relevant, when downsampling, while linear interpolation degrades – with EME-IQ=2.71 [1.53÷5.40] – the original image quality, Lanczos, which performs the best, preserves the EME-IQ score of the original image, reporting EME-IQ=2.85 [1.65÷5.64], even similarly to Akima (EME-IQ=2.84 [1.63÷5.66]).

Therefore, this results has allowed me to chose Lanczos as the reference interpolation method.

Figure 8.3 reports the percentage of r , nr , and i features against each perturbation for RCC (Figure 8.3 (a)) and CK (Figure 8.3 (b)), respectively. As one can notice, there are many more r features in RCC than in CK. In fact, on average against all perturbations, RCC has 65.6% of r features, 18.0% nr , and 16.4% i (Figure 8.3 (a)), whilst in CK the 39.0% is r , 42.9% nr , and the remaining 18.0% i (Figure 8.3 (b)). Nevertheless, both RCC and CK show an excellent robustness against N perturbation, as reported by N-RCC:100% and N-CK:99.6% (with the remaining 0.4% of i features). Moreover, this agrees with the outcome of Zwanenburg *et al.* [234], which similarly found that the highest percentage of robust features is against N perturbations.

Instead, substantial differences of r features between RCC and CK are found under R and V perturbations. In fact, results show that against R- perturbations, while in RCC r features are never lower than 70%, being 94.6%, 76.9%, and 70.4% for R-10, R-15, and R-20, respectively, in CK a very low percentage of feature is reproducible, being 8.9%, 4.9%, 2.5% for R-10, R-15, and R-20, respectively. In addition, as regards R+10, RCC shows 19% of r features, whilst CK has only 4% of them. As expected, r features have a much worse performance against R+10 perturbation, this suggesting that when segmenting it is always better performing a “safe” contouring, that is, underestimating rather than overestimating the ROI.

As regards V perturbations, whilst in RCC the percentage of r and nr features is 73.0% and 14.5%, in CK values are lower, with r -CK: 50.8% and nr -CK: 30.3%. Despite this difference, RCC and CK show a common behaviour against all perturbations at fixed v_z values (i.e., Z1÷Z5), both having a low percentage of r features (on average, 56.7% for RCC and 40.8% for CK) if compared with those at fixed v_s resolutions (i.e., v_s^b , v_s^M , v_s^w), that is on average 68.6% for RCC and 48.1% for

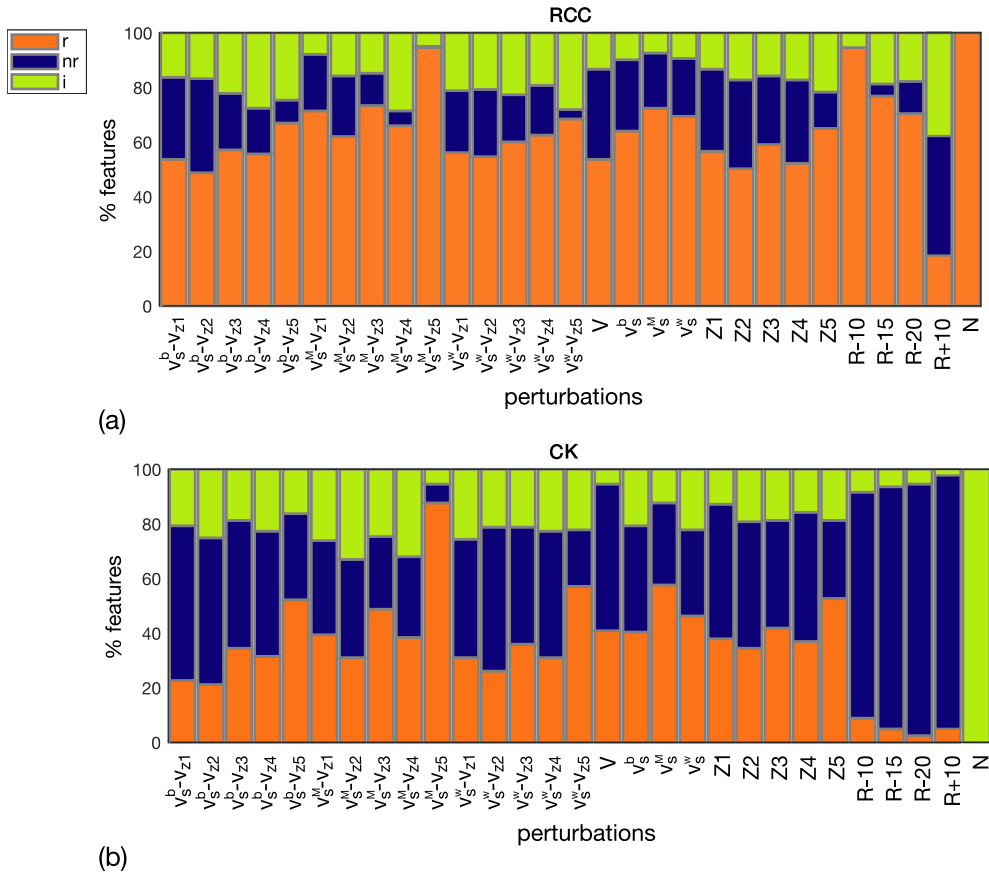


Fig. 8.3: Overall percentage of robust (r), non-robust (nr), and with indeterminate robustness (i) features against image perturbation for RCC (a) and CK (b), separately.

CK. In addition, if considering v_z , most of the 15 combinations have more than 60% of r features in RCC, mainly referred to v_s^M coupled with multiple v_z values, with the highest percentage (94.6%) achieved with v_s^M -Z5, that is, without interpolating along the z -direction. This couple is also the best in CK, with 87.7% of r features. One should consider that when resampling a heterogeneous CT dataset, the goal is to minimize interpolation artefacts. Accordingly, my results show that choosing the median resolution (v_s^M) answers this purpose, with a greater effect in CK rather than in RCC. This suggests that resampling at the median voxel size is strongly recommended. In addition, focussing on the different v_s - v_z couples, v_s^M performs at best when no interpolation in the z -direction is carried out between slices. This is somewhat expected, since the large difference between v_s (higher) and v_z (lower) voxel sizes makes the interpolation along the z -direction introduce a low reliable signal, if compared with the information in the original CT slices. Accordingly, while resampling along z -axis should be carefully evaluated, especially in case of a large slice spacing, preservation of the original v_z resolution could be in most cases the

best choice.

Figure 8.4 focuses on feature robustness of each feature class per perturbation

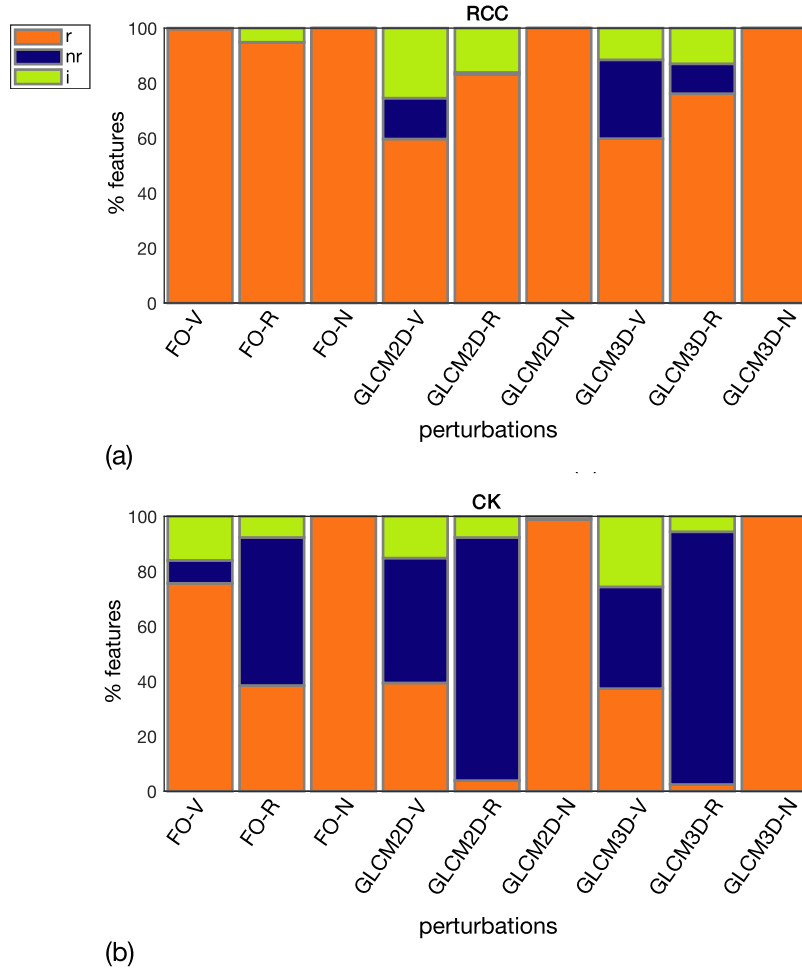


Fig. 8.4: Overall percentage of robust (*r*), non-robust (*nr*), and with indeterminate robustness (*i*) feature robustness reported per each feature class (i.e., FO, GLCM2D, and GLCM3D) for RCC (a) and CK (b).

type (values are also reported in Table 8.4). Our robustness analysis finds the FO features as being definitely the most reproducible ones, with FO features yielding the highest percentage of *r* features, in both RCC and CK, for V (RCC: 100.0%, CK: 75.6%), R (RCC: 94.9%, CK: 38.5%), and N (RCC and CK: 100.0%) perturbation types. This also confirms what is reported in the review of the most recent research works regarding feature repeatability and reproducibility by Traverso *et al.* [191]. Instead, GLCM2D and GLCM3D features achieve comparable results in both RCC and CK. In fact, the percentage of *r* features averaged over all perturbations is GLCM2D: 80.9% and GLCM3D: 78.6% in RCC, and GLCM2D:47.4% and

		FO			GLCM2D			GLCM3D		
		N	V	R	N	V	R	N	V	R
RCC	<i>r</i> (%)	100	100	94.9	100	59.7	83.2	100	59.8	76.1
	<i>nr</i> (%)	0	0	0	0	14.9	0.7	0	28.7	10.9
	<i>i</i> (%)	0	0	5.1	0	25.4	16.1	0	11.5	13
K	<i>r</i> (%)	100	75.6	38.5	98.9	39.4	3.9	100	37.3	2.5
	<i>nr</i> (%)	0	8.3	53.8	0	45.4	88.4	0	37.0	91.9
	<i>i</i> (%)	0	16.1	7.7	1.1	15.2	7.7	0	25.7	5.6

Table 8.4: Feature robustness against image perturbations referred to Figure 8.4.

GLCM3D: 46.4% in CK. Robustness against V and R is always higher in RCC than in CK. In particular, R perturbation shows the greatest difference, with GLCM2D-R: 83.2% and GLCM3D-R: 76.1% in RCC and GLCM2D-R: 3.9% and GLCM3D-R: 2.5% in CK.

All the FO features are *r* features in RCC in at least 60% of perturbations, and 9 of them are confirmed in CK too, this meaning that all *r* features in CK (9/13) are robust in RCC as well. Besides the well-established statistical descriptors (e.g. *m*, *M*, etc.), there are both *lcv* and *e*, two common indicators for measuring local heterogeneity or irregularity, that is also one of the changing properties of normal tissues while shifting into tumour ones [61]. The remaining 4/13 FO features result robust in RCC only, thus showing a higher specificity for tumour tissues, which could be useful to face specific tumour-related clinical questions. As far as second order features are concerned, 48 GLCM2D and 53 GLCM3D result *r* features in RCC in at least 60% of perturbations, and 44 of them are in both classes. In CK, 28 GLCM2D and 24 GLCM3D result *r* features in at least 60% of perturbations and 21 of them are shared. Similarly to FO features, all GLCM features (both 2D and 3D) that prove to be robust in CK, are robust in RCC too (such as *autoc*, *entro*, *savgh*, *sentro*, *sosvh*, *svarh*), while other features, measuring local asymmetries of GLCMs, are more tumour-specific (such as *cprom*, *cshad*, *energ*, *inf1h*, and *maxpr*).

This research also investigates the well-known phenomenon of the high correlation of GLCM-based texture features computed at different distances, to see whether and to what extent using higher distances is worth. Even this study shows that GLCM features computed at multiple δ are often highly correlated, with $\rho \geq 0.90$ for all the selected features. Moreover, ANOVA tests yield *p*-values > 0.03 for almost all δ . Figure 8.5 shows the boxplots of a representative feature, GLCM2D-*sosvh*, computed in RCC at multiple δ , where no relevant difference can be detected and ANOVA test confirms their statistical equivalence (*p*-value = 0.97). Actually, the equivalence between distances weakens as they shorten. For instance, the features *entro* and *sentro* representing local tissue heterogeneity show significant differences

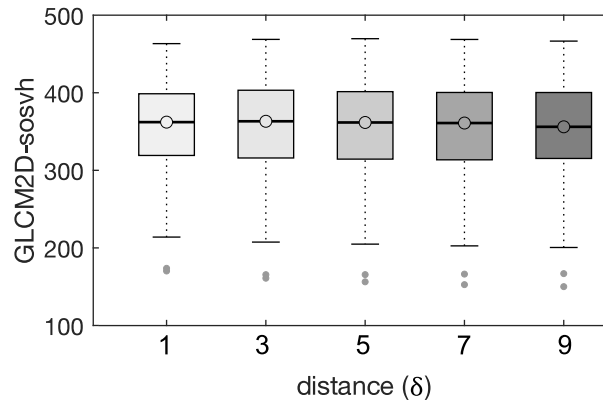


Fig. 8.5: Boxplots of GLCM2D-sosvh computed from $\delta = 1$ to $\delta = 9$ did not show any relevant difference and confirmed no statistical significance at ANOVA test (p -value=0.97).

between $\delta = 1$ and $\delta = 3$ in both RCC and CK (p -value $\leq 10^{-6}$), and between $\delta = 3$ and $\delta = 5$ in CK only. Analogously, the features *cprom*, *maxpr*, and *inf1h* in RCC are statistically equivalent for $\delta \geq 3$, and different from $\delta = 1$ (p -value $\leq 10^{-4}$). In practice, almost all features show to be equivalent when computed at distances from 3 to 9. This evidence is yet more relevant if considering that even in CK, having really wide ROIs, distances higher than $\delta = 3$ are most of times equivalent. This could suggest that computing textures at distance $\delta = 3$ should be generally enough, this permitting to make features selection simpler and more effective, beside reducing computational burden.

As a general remark, the recent literature lacks of comparative studies between 2D and 3D texture feature robustness, and even when features are compared on the basis of their capabilities (e.g., predictive ability, and so on) the outcomes are controversial [245]. Our results show that the overall robustness of GLCM features computed in 2D or 3D is similar. However, our findings show a higher number of r features for GLCM2D and, at the same time, a higher number of perturbations not affecting robustness of GLCM2D features. Therefore, GLCM2D texture features should be preferred.

8.2 A novel approach for radiomic feature computation

Beside the methodological radiomic study which I have presented above, my main contribution to radiomics also includes a novel approach to extract imaging features. In particular, I propose a new set of imaging features stemming from a deep analysis of the open issues of the imaging features currently used in the

literature (discussed in Chapter 5). Accordingly, this Section aims at (i) introducing the reasons to set up a novel approach to compute radiomic features (Sect. 8.2.1), describing (ii) the two step process to achieve them, based on a local computation of FO features (Sect. 8.2.2) and (iii) describing how to derive the whole set of global descriptors which summarises the properties of tissue being investigated through a vector of eighty-four features (Sect. 8.2.2).

Then, this new set of imaging features has been applied in different clinical studies carried out on LARC and PCa – presented in Sects. 8.3,8.4 – with the aim of testing their capability to identify potential prognostic and predictive IBs.

8.2.1 Why new imaging features

In Chapter 5 it has been provided a theoretical explanation of the imaging features widely adopted in radiomic study applications (Sect. ??). To summarize the main advantages and disadvantages of the different features classes, I recall that both shape and FO features have low computational complexity and allow an easy clinical interpretation, but the former limit image analysis to morphological assessments, also strongly dependent from ROI segmentation, whilst the latter, allowing only global assessments without any spatial reference, may yield a limited additional information content. Then, by considering more complex imaging features, including texture features (of second and higher orders) and wavelet transforms, the information content extracted from images substantially increases through the investigation of spatial relationships among neighbouring voxels. However, the computational cost rises accordingly and these more complex features can be very sensitive to image acquisition modalities and parameters, thus increasing the risk of poorly reproducible and robust measurements. In addition, by analysing images employing GLCMs, GLRLMs, or wavelet transforms, the clinical interpretation of results is more difficult, especially for radiologists and oncologists to whom these tools are primarily addressed.

Accordingly, I propose a new set of imaging features aiming at owing the main advantages of the feature classes explained so far. In particular, the novel imaging features have been thought to have an easy tool in terms of both computational complexity and clinical interpretation of the outcomes, which could also provide robust measurement, reducing at most any dependence from technical imaging parameters, while preserving the spatial information content of images. In fact, this is known to be very relevant in characterizing tissue patterns and detecting the tumour structural properties.

The computation of the new set of imaging features is based on a two stage procedure. The first stage is the local computation of FO features, which allows common FO features to account also for the spatial reference with the aim of catching

the variability of image value at local level, thus investigating tissue heterogeneity locally. The second stage is the extraction of global descriptors characterizing, with different metrics, various aspects of local heterogeneity. This Section addresses these two computational stages, respectively.

8.2.2 Local computation of parametric feature maps

As explained in Sect. ??, FO features substantially rely on histogram based measurements, thus performing a global assessment of image values. However, one should notice that the global histogram of an image is not representative of its details. For instance, Figure 8.6 shows four different grey scale images having exactly the

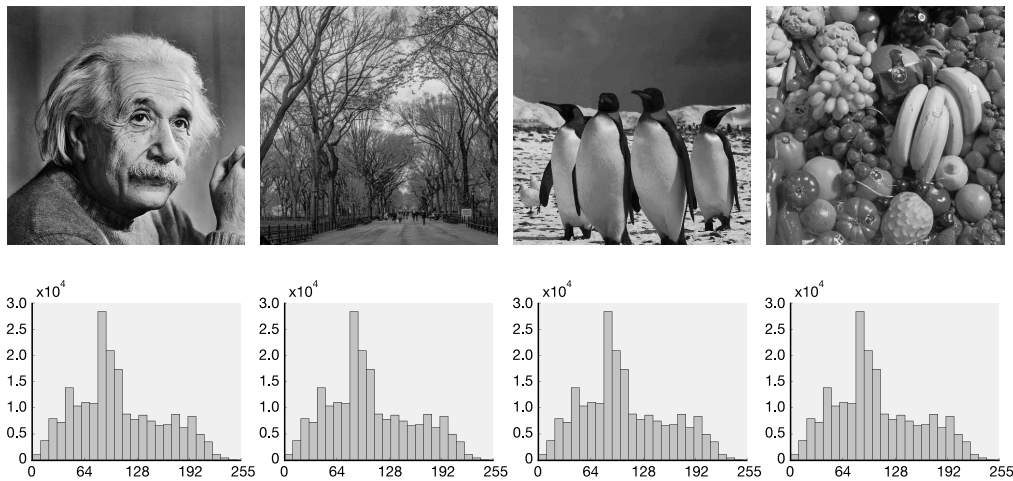


Fig. 8.6: The four different gray-scale images have the same histogram with $m \pm \sigma = 105 \pm 52$, thus proving that the global histogram is not representative of local image details.

same histogram, where m and σ are 105 and 52, respectively. In addition, Figure 8.7 highlights, in one of the images of Figure 8.6, three pixels with the same grey value in which a surrounding square window of 15 pixels side length has been investigated by way of example through the histograms reported on the right side of the figure. One can notice that, at local level, the three pixels with same grey value equal to 146 are inserted into very different contexts, and shapes and mean values of local histogram can depict these local differences, that represent different image details. Yet more, Figure 8.8 reports the maps of σ values computed locally by using the same representative square window of 15 pixels side for each of the images in Figure 8.6. As one can see, even if the global σ value is identical for the four images, the local values are very different as well as the mean of local σ (σ_m) values, that is $\sigma_m = 13$, $\sigma_m = 27$, $\sigma_m = 15$, $\sigma_m = 17$, respectively from left to right. It is also worth noting that the colorimetric maps in Figure 8.8 highlight the areas with a greater image contrast,

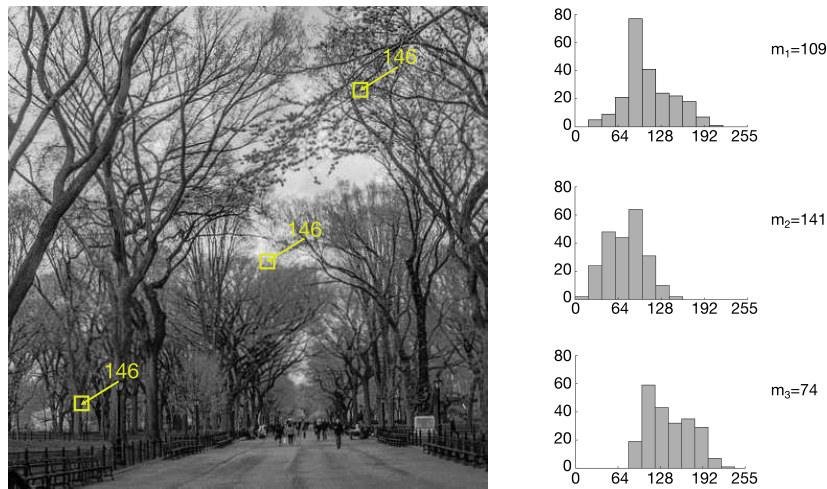


Fig. 8.7: Three local image windows centred on three different pixels having the same grey value but showing a different neighbourhood as represented by the histograms of each window, having three different shapes and mean values.

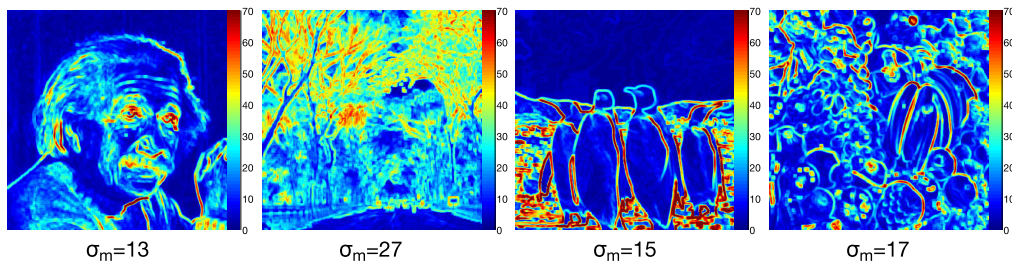


Fig. 8.8: σ values computed locally with a window of 15 pixels side length. The four images have the same global σ ($\sigma=52$ as in Figure 8.6), but very different local σ values. Yet more, σ_m for each image is different, $\sigma_m=13$, $\sigma_m=27$, $\sigma_m=15$, $\sigma_m=17$, respectively from left to right.

thus emphasizing structures and edges.

By referring to medical image analysis, for instance of MRI or CT, I propose to adopt a similar approach based on FO imaging features computed locally, that is, by considering a small local window centred on each pixel of the image ROI. For instance, Figure 8.9 shows the process of computing a local map of m values considering a ROI outlining the prostate gland in a T_2 -weighted MR image slice. Hence, for each pixel within the ROI, a FO feature (e.g., the mean m) is computed by considering the pixels belonging to the square window centred on the pixel itself, further represented through a colorimetric map for visualization purposes. In this way, the local computation of image features leads to three main achievements. First, each global metric computed on parametric maps provides a more representative

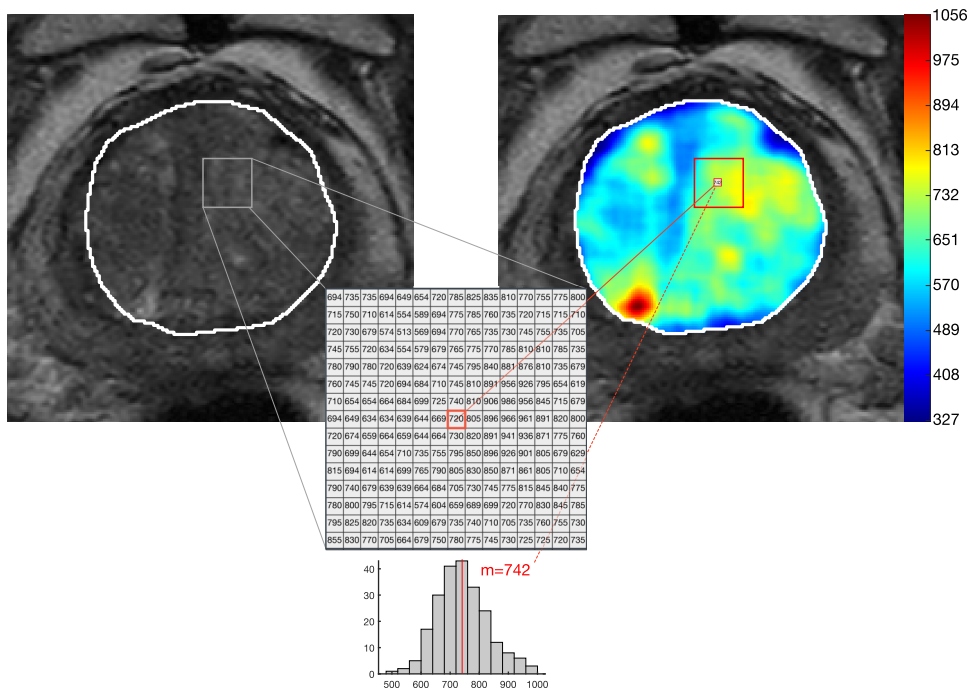
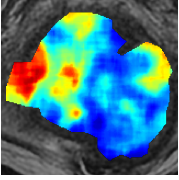
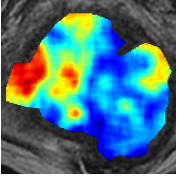
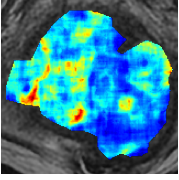
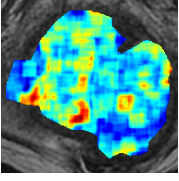
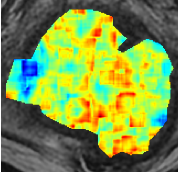
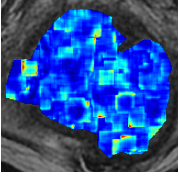
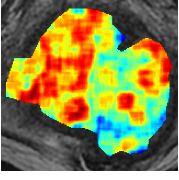


Fig. 8.9: Computation of local features in a T_2 -weighted MR image by adopting a local window centred on each pixel within the prostate ROI, thus achieving parametric maps of local FO features. In this representative case, the local m is considered.

feature value (as in Figure 8.8, where the σ_m of each feature is different to the global σ value computed directly on images in Figure 8.6). Second, this process allows achieving parametric maps through which emphasizing slightly visible image details and sharpening structures and edges. Third, parametric maps of local features can even highlight, depending on the metric considered, latent properties of images through measures of local variabilities and asymmetries of image values, aiming at catching the heterogeneity of structures at underlying layers.

In this Thesis, I have focused on seven metrics, among the most widely adopted measures of dispersion, in order to investigate the extent of variability of image values. The local computation of the seven metrics allowed achieving as many parametric maps, which enable, in their turn, a visual assessment of FO features. This is very helpful to improve the confidence of radiologists for the tool, since radiologists are allowed to see the variations of local features in correspondence to the morphological references. In particular, the seven metrics considered are reported in Table 8.5, each coupled with a parametric map computed on a T_2 -weighted MR image of a rectal cancer ROI. For all feature formulas reported in Table 8.5, let us define the domain $\Omega:LW$, where LW is the local window and N is the number of pixels in Ω , so that $x_i \in \Omega$, $i = 1, \dots, N$.

Table 8.5: Seven parametric maps of local first order features

Parametric map	Formula
 <p>377.00 357.44 337.89 318.33 298.78 279.22 259.67 240.11 220.56 201.00</p>	<p>Median (M)</p> $M = x_{[\frac{f_c}{2}]}, f_c = \frac{\sum_i N p_{x_i}}{2}$
 <p>369.45 351.63 333.82 316.00 298.19 280.38 262.56 244.75 226.93 209.12</p>	<p>Mean (m)</p> $m = \frac{1}{N} \sum_i x_i$
 <p>113.75 103.08 92.42 81.75 71.08 60.42 49.75 39.08 28.42 17.75</p>	<p>Interquartile range (IQR)</p> $IQR = M_{upper\ half} - M_{lower\ half}$
 <p>0.23 0.21 0.19 0.18 0.16 0.14 0.12 0.10 0.09 0.07</p>	<p>Coefficient of Variation (CV)</p> $CV = \frac{\frac{1}{N} \sum_i x_i}{\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}}$
 <p>1.74 1.36 0.97 0.59 0.20 -0.18 -0.57 -0.95 -1.34 -1.72</p>	<p>Skewness (s)</p> $s = \frac{\frac{1}{N} \sum_i (x_i - \bar{x})^3}{\left(\frac{1}{N} \sum_i (x_i - \bar{x})^2\right)^{3/2}}$
 <p>7.19 6.58 5.97 5.36 4.75 4.14 3.52 2.91 2.30 1.69</p>	<p>Kurtosis (k)</p> $k = \frac{\frac{1}{N} \sum_i (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_i (x_i - \bar{x})^2\right)^2} - 3$
 <p>4.43 4.22 4.00 3.79 3.58 3.36 3.15 2.93 2.72 2.51</p>	<p>Entropy (e)</p> $e = - \sum_i p_{x_i} \log p_{x_i}$

The major factor determining the capability of parametric maps of increasing the information content of original images relies on the choice of the window size, through which performing the local computation. In the applications which will be presented in Sects. 8.3 and 8.4, the window size has been selected for investigating a tissue unit ranging between 0.5 mm and 10 mm side length, generally corresponding to the smallest detectable tumour size through mpMRI images. Accordingly, a variable number of pixels is used for each patient, adaptively depending on the spatial acquisition resolution, without any need of pre-fixed values for the window size.

Actually, the image analysis based on small local windows recalls the basic concept of local image processing relying on the convolution of a kernel function with an image. However, while convolution is a linear operator and can be exploited, for instance, for the computation of m at local level, yielding the same results as the approach that I propose, a convolution kernel cannot apply to the computation of non-linear functions, such as M , s , k , CV , etc.

Finally, as regards the FO features computed locally, M and m metrics have been considered to measure the primal – even visible – content of image on two different filtered images, the former which preserves contrast and edges, the latter which smooths local variations and attenuates spurious components. Instead, the other five metrics, computed locally, aim at measuring different aspects of the heterogeneity of the analysed tissue. The maps of IQR and CV represent two different ways to emphasize structure boundaries. The IQR is a direct measure of image value dispersion. Since homogeneous structural components are reasonable associated to similar GLs in the image, higher values of IQR characterize transition zones among adjacent tissue components. Instead, the CV map is an inverse measure of heterogeneity, since the σ is at the denominator of the formula. Accordingly, CV assumes higher values in homogeneous structures, like edges, characterized by small σ values. Accordingly, CV maps, although indirectly, may result in a similar visual effect of IQR maps, but, depending on the application, each of them can have a better or worst specificity in tissue characterization. Then, the maps of s and k are two direct measures of Gaussianity of distributions. Departures from Gaussianity, above all if measured at local level, can unravel sub-populations of image values which may hint at semantic heterogeneities associated to structural or functional tissue alterations. This is valid for both changes of benign tissue towards malignancy, leading to an increase of heterogeneity, and recovery towards homogeneity, for instance of tumour tissues responding to therapies, whose heterogeneity decreases. Finally, maps of e directly measure tissue heterogeneity. The logarithmic transform of image content, realized with the computation of entropy, emphasizes the more informative areas, with the highest variability of image values in the smallest region, and flattens those regions with a lower variability of values. Accordingly, it allows measuring objec-

tively the intrinsic heterogeneity of the image itself.

8.2.3 A new set of imaging features

To summarize quantitatively the information retained by the seven colorimetric maps and measure specific properties of them, twelve global statistical descriptors have been computed accordingly, based on the histogram of the feature values within the local map. Actually, in medical imaging applications, several image slices are assessed contemporaneously for each patient. Therefore, to aggregate information from all patient slices, one global histogram is derived from the parametric maps of all slices of each patient. Then, twelve global descriptors are computed on each of the seven different parametric maps, this allowing achieving eighty-four radiomic features. Figure 8.10 shows the process, referred to one patient, through which the new set of eighty-four imaging features is computed. Hence, given the acquired images and the ROIs segmented on multiple image slices, the seven local parametric maps (Table. 8.5) are computed for each segmented slice according to the procedure reported in Sect. 8.2.2. Then, for each parametric map, twelve statistical descriptors are computed on a global histogram referring to all slices. The pseudo-code for the computation of the new set of imaging features as Algorithm 1. Besides the

Algorithm 1: Pseudo-code for the computation of the set of new imaging features

```

define a local window LW;
define a set of  $m$  features;
define a set of  $k$  descriptors;

read  $imm, ROI$ ;
padding of  $imm$ ;
foreach feature  $m$  do
     $PM[m] \leftarrow ROI$ ;
    foreach pixel  $p \in ROI$  do
         $PM[m](p) \leftarrow m(LW)$ ;
    end foreach
    foreach descriptor  $k$  do
         $ret[k] \leftarrow k(PM)$ ;
    end foreach
    return  $ret$ ;
end foreach
return  $PM$ ;

```

seven features computed locally, the twelve global descriptors also include the five descriptors reported in Table 8.6 and are defined in the domain $\Omega : PM$ where PM is the parametric map.

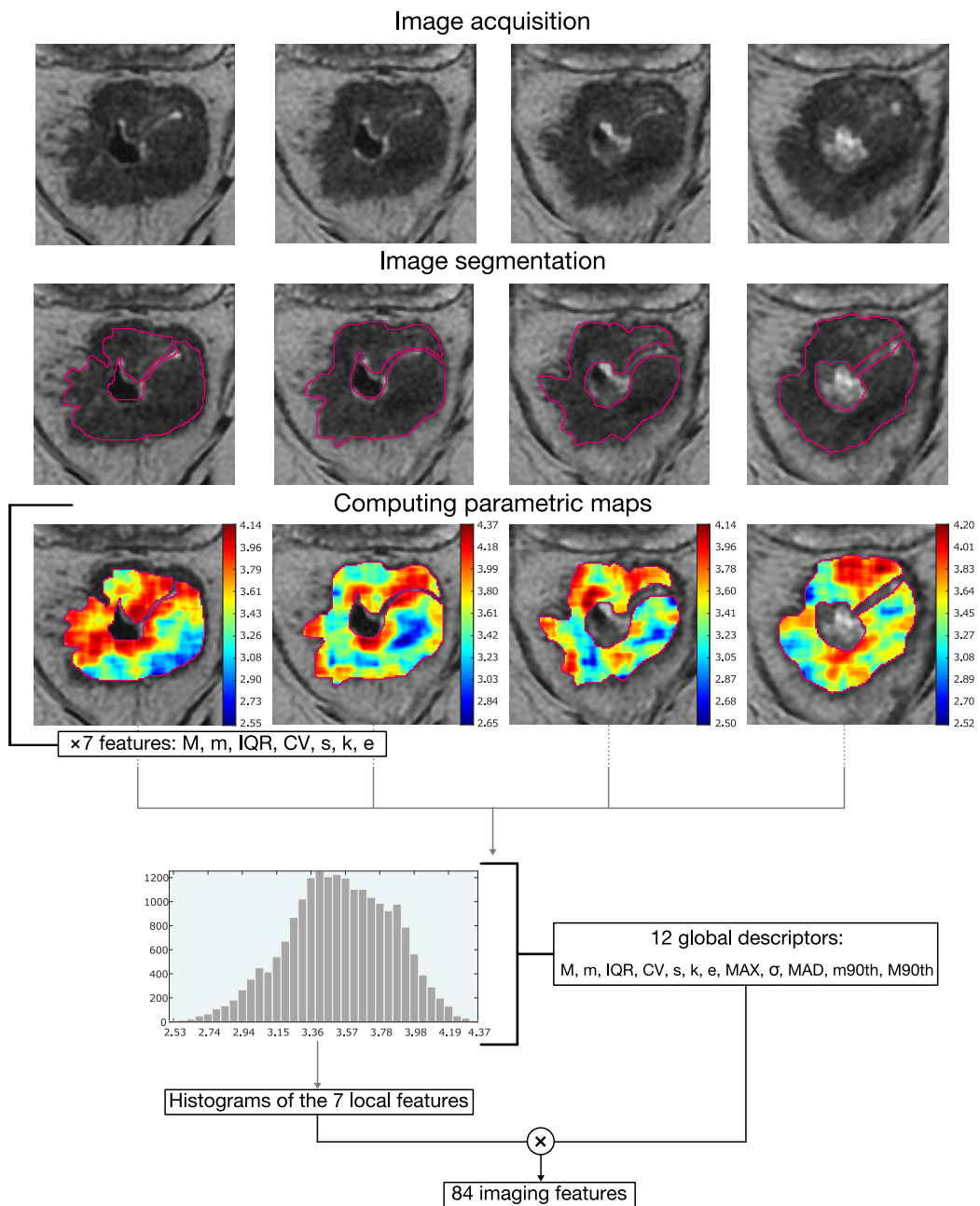


Fig. 8.10: Workflow for computing the new set of imaging features for each patient. Seven parametric maps are derived from as many features computed locally within image ROIs. Then, from the histograms of the seven features referred to all ROI slices, twelve global descriptors are derived accordingly, thus finally achieving 84 imaging features.

Table 8.6: Five more statistical descriptors computed on the seven parametric maps, beyond the seven features computed even locally and above reported.

Statistical descriptor	Formula
Maximum value (MAX)	$\max_{x_i}(\Omega)$
Standard deviation (σ)	$\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}$
Median absolute deviation (MAD)	$M(x_i - M(\Omega))$
Mean over the 90th percentile (m_{90th})	$m(x \in x_i \geq x_{90th})$
Median over the 90th percentile (M_{90th})	$M(x \in x_i \geq x_{90th})$

8.3 Discriminate nCRT responder from non-responder in Rectal Cancer

This clinical study has been conducted in collaboration with the Department of Radiology of the S. Orsola-Malpighi Hospital, University of Bologna, led by Prof. R. Golfieri.

Colorectal cancer is the third most common cancer and the second leading cause of oncologic-related mortality in the world [246]. It is more common among men than women, although the incidence is rapidly increasing in the female population due to the standardization of lifestyle habits, and 3–4 times more common in developed than in developing countries [247]. LARC is defined as a tumour penetrating through the whole bowel wall (stages T3/T4) and/or characterized by the involvement of regional lymph nodes (N1/N2), without any distant metastases (M0). The standard of care treatment for patients affected with LARC currently involves nCRT, followed by total excision of the mesorectum (TME) [248]. MRI plays an important role, together with endorectal ultrasound, both in the primary staging and in the restaging of LARC after nCRT [249]. According to the ESMO guidelines [250], accurate imaging of the tumour and lymph nodes using high-quality MRI is essential to determine the local staging of rectal cancer, which is a critical marker to decide whether performing nCRT according to the staging system. Recent evidence suggests that 15-27% of patients will achieve pathological complete response (PCR) to nCRT, before the surgery, suggesting that a ‘wait and watch’ approach could be the best choice for these patients, avoiding all the surgical complications. On the other hand, the percentage of patients who do not achieve tumour regression after nCRT, defined as non-responders (NRs) patients, is reported to be between 7 and 30% [251],[252]. Moreover, the potential side effects of nCRT can be very serious,

varying between hematologic, gastrointestinal and dermatologic effects, incontinence or sexual dysfunction [253]). Finally, 14–27% of patients with LARC who receive this regimen can develop acute or long-term grade 3–4 toxic effects as reported by Sauer et al [254]. In this scenario, the early identification of NR and PCR patients before the beginning of nCRT could be of great value in order to avoid ineffective treatment and to develop a more tailored strategy of care, such as a primary surgical intervention or an intensified treatment regimen. Based on the concept that radiomic features extracted from routinely acquired medical images might help the identification of predictive IBs, as explained in Chapter 5, this study aims at assessing the role of radiomic features extracted from pre-therapy baseline T_2 -weighted MR images in predicting the pathological response of patients undergoing nCRT, thus distinguishing the group of patients with poor or minimal response from those obtaining moderate response or PCR, based on the tumour regression grade (TRG) assigned after surgical resection.

Histopathological reference standard

The clinical reference standard to determine nCRT response is achieved at histopathological report. Following the TRG staging system according to AJCC [255], nCRT response is assessed through the following four groups:

- TRG 0: no viable cancer cells (i.e., PCR)
- TRG 1: single or small groups of tumor cells (i.e., moderate response)
- TRG 2: residual cancer outgrown by fibrosis (i.e., minimal response)
- TRG 3: minimal or no tumor cells killed (i.e., poor response)

8.3.1 Study population

This is a retrospective study enrolling initially eighty-five patients diagnosed with LARC, who have undergone MRI for primary staging, from January 2018 to December 2019. The inclusion criteria have been (i) patients diagnosed with LARC in S. Orsola-Malpighi Hospital according to ESGAR-ESGE guidelines undergoing, (ii) primary staging pre-treatment MR, and (iii) treated with long-course neoadjuvant CRT, (iv) followed by TME. The exclusion criteria have been: (i) patients who did not complete the standard nCRT, (ii) patients who did not undergo surgical resection in our institution, (iii) patients without available TRG information on the histopathological report, and (iv) patients with incomplete staging MRI or imaging artefacts.

As reported in Figure 8.11 showing the flowchart of patient enrolment, from the initial cohort of eighty-five patients, they have been excluded eleven patients who

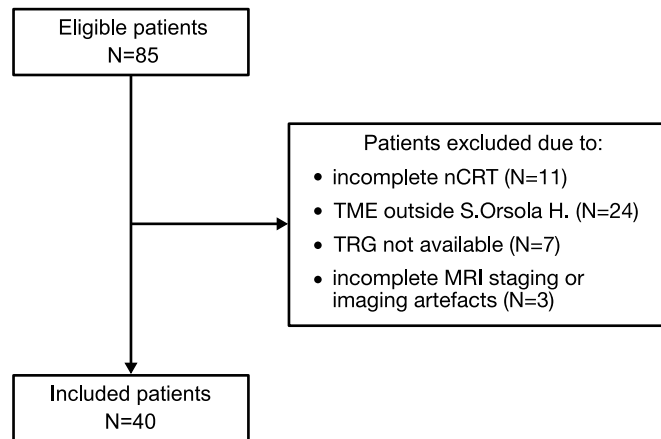


Fig. 8.11: Flowchart of patient enrolment in radiomic LARC study

did not complete the standard nCRT, twenty-four patients who have not undergone surgical resection in S. Orsola-Malpighi Hospital, seven patients with TRG not available, and three patients with incomplete MRI staging or MR images corrupted by artefacts. Forty patients have been finally enrolled, 14 (35%) women and 26 (65%) men, with a median age of 65 years (interquartile range, IQR=14.5 years).

Based on the histopathological report, patients have been finally clustered into two groups: (1) patients with TRG=[0,1] (TRG0-1), who have obtained moderate response or PCR, and (2) patients with TRG=[2,3] (TRG2-3), who have obtained a poor or minimal response.

The Institutional Review Board has approved the study and waived the requirement for written informed consent (No 842/20202/Oss/AOUBo).

8.3.2 Image protocol acquisition

All patients underwent pelvic MRI scan performed with 1.5 T MRI (GE Healthcare) following ESGAR guidelines [256] for rectal cancer primary staging. The protocol included three T_2 -weighted fast spin echo sequences in the sagittal, oblique-axial high-resolution and oblique-coronal high-resolution planes. DWI images were also obtained in axial planes using Echo-Planar Imaging sequences at three b-values ($b=0, 600, \text{ and } 1000 \text{ s/mm}^2$) and restriction of diffusion was quantified by the ADC value. Bowel cleansing was performed with two days of low-fibre diet and oral administration of Macrogol-Na-K 14 gr the day before the MRI study.

8.3.3 Data preparation and feature generation

All oblique-axial high-resolution T_2 -weighted fast spin echo and DWI images have been retrieved from the Picture Archiving and Communication System (PACS, Carestream, Canada) for image segmentation. A radiologist with more than 15 years of experience has performed all the segmentations by manually outlining the lesion on each consecutive slide, excluding the intestinal lumen, employing a designated, free open source software package for visualization and medical image computing (ImageJ, version 1.52a, available at <https://imagej.nih.gov/ij> [257] in a two-step process: (i) the entire tumour region has been highlighted on 3 orthogonal planes of the space, (ii) a manual segmentation has been performed on the 3 mm-thickness axial plane in all slices of the tumour's site, by outlining a ROI on each of them. The radiologist has been blinded to the histopathological results.

Eighty-four radiomic features have been generated from tumour ROIs according to Sect 8.2.3.

8.3.4 Statistical analysis and feature selection

A statistical univariate analysis has been performed on the original eighty-four radiomic features to select the one with the highest discriminative capability for the two groups TRG0-1 (TP) and TRG2-3 (TN). It is well known that to have the highest generalization capability the simplest discrimination model has to be adopted. In addition, the ratio $r = N/l$, with N being the number of samples of the smallest class and l the number of features finally selected, must be kept as the highest as possible. To this aim, to prevent overfitting, one RF only has been considered for discrimination [161]. One-tail Wilcoxon rank-sum test at $\alpha=0.05$ significance level has been used to assess the statistical significance of the separation between the two groups. The overall selection procedure has been carried out in three steps. First, radiomic features unable to statistically differentiate TRG0-1 and TRG2-3 have been discarded. Second, the radiomic features coming through the first step have been ranked according to the p -value after Bonferroni correction. Third, the feature showing the lowest p -value has been selected and its discriminative capability assessed through the ROC, by computing its AUC. To determine the best cut-off for the selected radiomic feature, the YI has been computed and the couple of values for specificity (SP) and sensitivity (SE) has been achieved, accordingly. The values of FP, FN, PPV, and NPV are also used to discuss performances. Separation between TRG0-1 and TRG2-3 has also been assessed by computing M values and IQR for each group.

8.3.5 Experimental results

At present, there is a growing interest in the medical community towards radiomics applied in the study of colorectal cancer. Nevertheless, there is a number of studies whose specific aim is the possibility of predicting the response to neoadjuvant therapy through the radiomic analysis, which can prospectively tailor medical care based on tumour profiling [22]. Also, this might allow identifying the responsive patients and providing them with targeted therapies, whilst distinguishing the non-responsive ones who can be immediately addressed to surgery, avoiding toxicity and collateral effects of the neoadjuvant therapy. In this work, radiomic features extracted from pre-therapy baseline MR images have been analysed to discriminate TRG0-1 and TRG2-3 according to the AJCC classification, which indicates the response to nCRT carried out subsequently. Eighteen radiomic features come through the first feature selection step ($p\text{-value} < 0.05$), whilst the local entropy of the skewness (s_e), a local measure of tumour heterogeneity, results the most discriminative radiomic feature to separate TRG0-1 and TRG2-3, with $p \sim 10^{-5}$, much lower than the significance threshold of $\alpha = 3 \cdot 10^{-3}$, considering the Bonferroni correction of $\alpha = 0.05$. Moreover, being the smallest class (TRG0-1) of 15 patients, the final selection of a single radiomic feature leads the ratio value to be $r = 15$, thus strongly preserving the generalizability of the model.

Figure 8.12 reports the ROC curve of s_e , yielding $AUC = 0.90$ (95% CI, 0.73-0.96), with $YI = 0.68$ (cut-off $s_e = 3.93$) corresponding to the couple $SE = PPV = 80\%$ and $SP = NPV = 88\%$. In particular, as one can see from the waterfall plot in Figure 8.12 (b), showing s_e for the two groups, TRG0-1 with green bars and TRG2-3 with blue ones – the cut-off value has been subtracted for visualization purposes – the separation is achieved with 3 FPs and 3 FNs. Accordingly, the couple $SE = PPV$, $SP = NPV$ represents a very good trade-off between need for containing the risk of overtreatment and detection of the responder patients. It goes without saying that the cut-off of s_e currently set at $s_e = 3.93$ can be adjusted to ensure that no patient who needs nCRT is excluded from therapy, allowing the achievement of a number of FN equals to 0 (i.e., sensitivity of 100%). Consequently, this would lead to lose the specificity of the calculation and to rise the number of FPs, with a consequent increase of nCRT-induced over-treatment.

The two populations of TRG0-1 and TRG2-3 have very different characteristics that the radiomic analysis performed on pre-therapy MR images has allowed highlighting. In particular, the distance between patients with different grades of nCRT response relies on parametric skewness maps at baseline, which results spread over the ranges [2.82-5.26] for TRG0-1 and [2.88-8.74] for TRG2-3, respectively. As one can appreciate through the boxplots in Figure 8.12 (c), the skewness maps at baseline show different heterogeneity values. In fact, the first group (TRG0-1) iden-

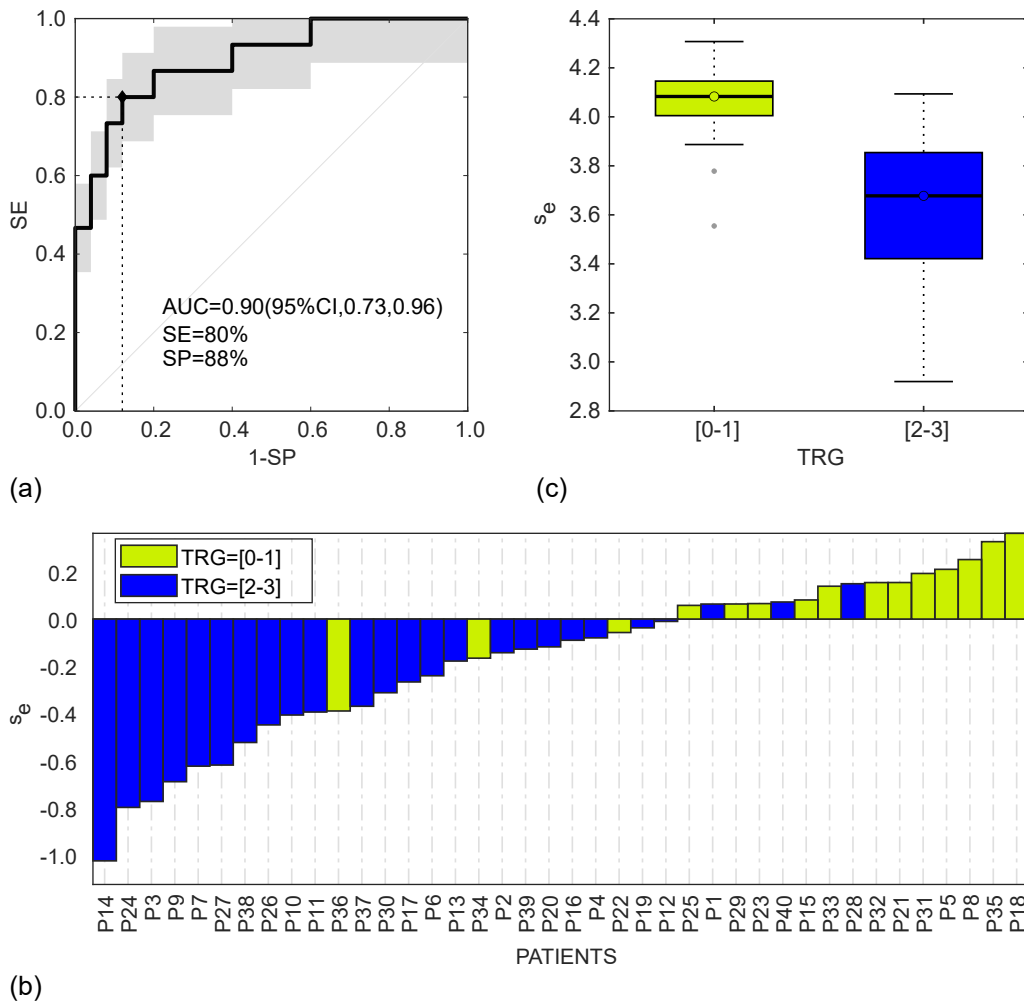


Fig. 8.12: Discrimination results between TRG0-1 and TRG2-3. In particular, (a) shows the ROC curve yielding AUC=0.90 (95% CI, 0.73-0.96) with SE=80% and SP=88%, (b) shows the waterfall plot of s_e where the two groups are highlighted with green (TRG0-1) and blue (TRG2-3) bars, and (c) reports the boxplots of the two groups (p -value $\sim 10^{-5}$).

tifying patients that will completely respond to the therapy (TRG 0) and those with single cells or small groups of tumour cells (TRG 1), is characterized by a greater heterogeneity of skewness maps, as reported by s_e -M=4.08, even with a very low dispersion of values, as shown by IQR=0.15 [4.00÷4.15]. Instead, the second group (TRG2-3) of patients who will respond poorly (TRG 2) or show minimal response (TRG 3) have a low heterogeneity of skewness values (s_e -M=3.68), also spread over a wider range, with IQR=0.46 [3.40÷3.86].

As regards a plausible clinical interpretation of our feature, first it is worth noting that this is not the heterogeneity of tumour tissue, but the heterogeneity of the skewness, computed on the local tissue regions. To clarify, Figure 8.13 reports the

skewness maps of four patients, with TRG from 0 to 3, highlighting the evolution of local skewness variability. In fact, while in the maps of TRG0 and TRG1 the asymmetries are localized, as it can be seen by the wide red and blue regions, these tend to shade in the skewness map of TRG2 and yet more in TRG3, showing quite uniform speckles. Accordingly, s_e , the entropy computed on the skewness maps, catches the higher heterogeneity of TRG0 and TRG1 in Figure 8.13 (a,b), with respect to that of TRG2 and TRG3 in Figure 8.13 (c,d). It is worth noting that the skewness represents the asymmetry of local T_2 -weighted values, that is a departure from normality, and in case of an early alteration stage, it begins in subregions of the ROI, each with a different evolving status (i.e., local skewness value), this yielding a heterogeneous picture of the ROI in TRG 0-1. Instead, in a

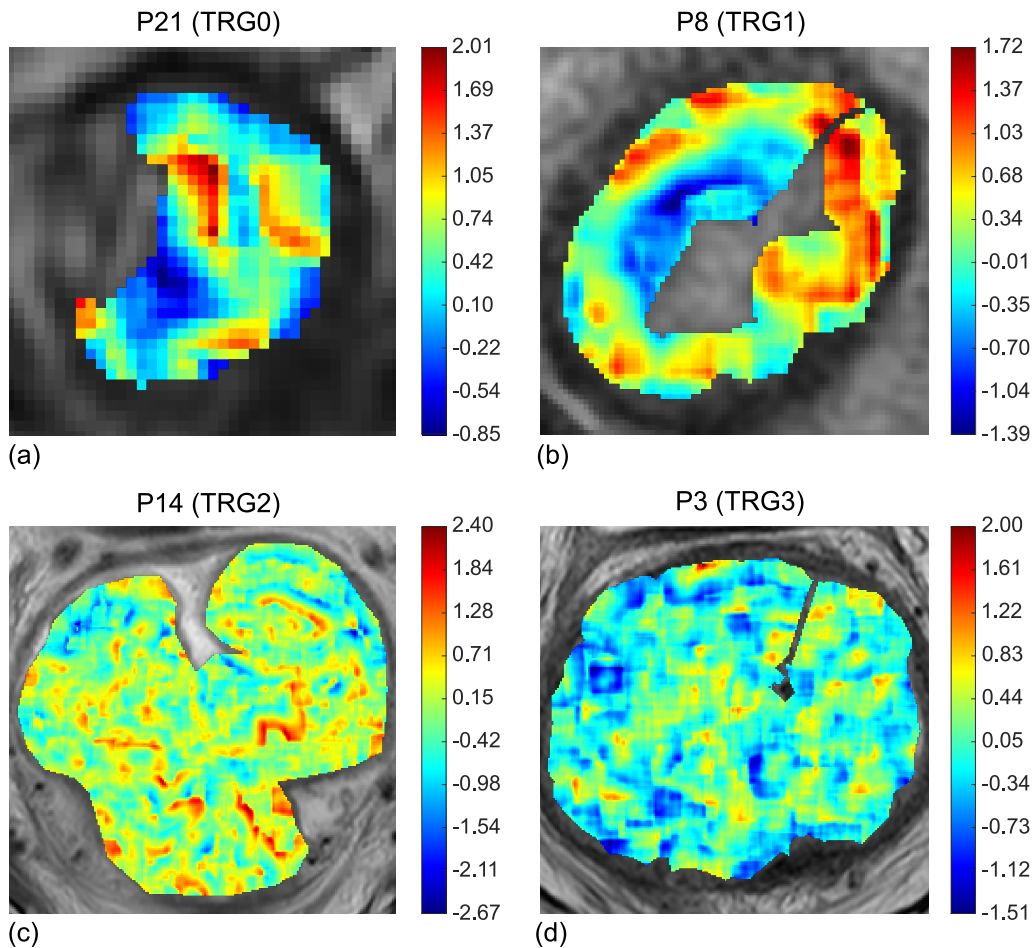


Fig. 8.13: Colorimetric maps of local skewness of four patients with TRG 0 (a), TRG 1 (b), TRG 2 (c), and TRG 3 (d) highlighting the evolution of local skewness variability. While localized asymmetries are present in the maps of TRG0-1, quite uniform speckles are shown in TRG2-3. Accordingly, s_e catches the higher heterogeneity of (a,b) maps and the lower ones in (c,d) maps.

later evolution stage, the extension of the skewed regions dramatically shrinks, this hinting at a greater similitude between voxels of the ROIs, represented by much less heterogeneous TRG2-3. The lower the baseline tumour aggressiveness, the higher the heterogeneity (contrast) in skewness maps, the greater the benefit from nCRT therapy.

By comparing these results with the state of the art, one can notice that several authors ([252],[258],[259],[260],[261],[262],[263],[264]) have recently addressed the prediction of nCRT response based on clinical assessments [258],[261] or different TRG [252],[259],[260],[262],[263] staging systems. However, the first relevant difference with this work is that in all these studies the ratio r is far smaller than ours is ($r=15$). The higher ratio ($r=9.4$) can be found in the work of Cusumano et al [259], that however achieves a worse separation ($AUC=0.77$), whilst the remaining have the ratio value ranging from $r=5.5$ [252] to even $r=0.87$ [258], that is, more features than patients. Nevertheless, if sometimes the results are a little better than ours ($AUC=0.92$ [260] and $AUC=0.94$ [258],[261],[263]) some others are even much worse ($AUC=0.72$ [252], $AUC=0.75$ [264], $AUC=0.82$ [262]). Despite most of these studies perform an external validation of the predictive model on a holdout test-set, their very low r makes the generalizability of their models questionable.

As a concluding remark, these findings suggest that radiomics can really help physicians in choosing the best therapy based on quantitative imaging data, which have remained hidden so far, and have never been considered in the choice of the optimal therapeutic process of the patient. In addition, the new set of imaging features shows very good performance, although limited to a discrimination study, and holds great potentialities in developing predictive IBs of nCRT response on a wider dataset. Finally, the preliminary findings showing such a marked differentiation between TRG0-1 and TRG2-3, which indicate that in the future the staging of patients according to the four single TRG stages could be even possible.

8.4 Exploiting radiomics in Prostate Cancer (PCa)

In this Section, I will introduce two clinical radiomic studies carried out in collaboration with the Department of Diagnostic Imaging of IRCCS IRST, led by Dr. Domenico Barone. In particular, these studies refer to applications of radiomics and quantitative imaging techniques in the study of PCa.

According to the last statistics, PCa is the second most diagnosed cancer worldwide and the fifth most common cause of cancer death among men [246]. This strongly impacts on the clinical management in terms of costs and resources, also based on the PCa stage at the diagnosis, that could suggest different clinical path-

ways [265]. Non-clinically significant PCa (ncsPca) diagnosis assigned at biopsy, with a Gleason Score (GS) equals to 3+3 (GS=3+3), generally corresponds to a favorable prognosis, with a high life expectancy at 10 years from diagnosis, and a low risk of biochemical recurrence. In most of cases, ncsPCa do not require curative treatments, thus being admitted to active surveillance (AS) protocols [266]. Instead, PCa with a $GS \geq 3+4$ or higher, namely clinically significant PCa (csPCa), may be submitted to radiotherapy or even radical prostatectomy (RP) and have a worse clinical outcome [195]. PCa aggressiveness is conventionally assessed through biopsy, that can be totally random or aimed at the most supposed malignant areas, whether it is TRUS- or MRI-guided. Hence, tissue samples are analysed in terms of structural changes and glandular patterns which may describe different cell differentiations, correlated to tumour aggressiveness [267].

Frequently, biopsy outcomes are reported to differ from those obtained after RP [268], and even between closely repeated examinations [269]. Moreover, notable side effects are experienced from men undergoing biopsy, including bleeding, pain, and infection [270]. Therefore, the availability of non-invasive imaging approaches for distinguishing ncsPCa from csPCa is a very attractive prospective to both increase the detection rate of csPCa and spare patients from unnecessary biopsies and overtreatment [271]. mpMRI is employed in the clinical routine, primarily for PCa diagnosis. The current guidelines of PI-RADS v2.1 attribute a key role to the morphological T_2 -weighted sequence and the DWI to obtain functional information regarding variations of tissue diffusivity [56]. In fact, from a theoretical point of view, DWI measures the (water) diffusion properties within tumour tissues, where the structural diffusion of the biological components and the hyper-cellularization processes obstruct the motion of the water molecules [272]. Water restriction yields a high DWI signal, detecting tumour changes towards malignancy, progressively more emphasized in high b-value sequences ($b \geq 1000 \text{ s/mm}^2$), to detriment of the benign glandular components, where any morphological reference is lost. However, high b-value DWI sequences can be altered by anatomical components or T_2 -shining artefacts, particularly in the 1.5T systems. Therefore, in the clinical practice, a definite clinical confirmation for PCa diagnosis is conveyed by the ADC maps. As explained in Chapter 2, ADC maps derive from a normalization process of two, or more, DWI sequences acquired at different b-values. Originally, high signals in DW images are converted into low signals in ADC maps, which recover the information related to the apparent diffusion of water's molecules, thus losing the specific measures contained in the native DWI sequences, which arise directly from the tissue properties. Nevertheless, the ADC normalization process also allows removing the misleading high signals in DWI and, consequently, distinguishing the tumour boundaries more clearly, keeping the morphological information of the gland [272], albeit motion artefacts can remain. Moreover, ADC maps improve SNR of DWI

sequences and enhance image quality. Hence, they are currently the most effective sequences for PCa detection and localization. Accordingly, ADC has been predominantly exploited in quantitative imaging too ([273],[274]), eventually combined with T2-weighted and native DWI sequences ([275]), and on several occasions, ADC metrics have proved to correlate with the GS successfully. It is not surprising that good results have been achieved by previous studies in classifying ncsPCa and csCPa ([276],[274]). More recently, the advent of 3T-MRI systems has allowed enhancing the quality of DW images and improving SNR of high b-value sequences, thus reducing noise and artefacts. Consequently, with these systems, the motivations for which the ADC has been preferred to DWI, decay. The present technology allows exploiting the native DWI sequences at their best, both in the clinical practice and, above all, in the quantitative imaging.

Thanks to the collaboration with IRCCS IRST and the availability of a 3T-MR system, the two clinical studies addressed in the following sections have represented the opportunity to restore the main role of DWI in quantitative imaging applications, for patient staging and characterizing cancer aggressiveness, letting ADC have a key role in qualitative and semi-quantitative assessment for detection.

PCa grading system

The clinical reference standard to determine PCa aggressiveness and prognosis is represented by GS, assigned at histopathological or biopsy examination.

GS grading systems classify the microscopic tumour appearance into five different patterns:

- Pattern 1: well differentiated carcinoma, close to normal prostate tissue
- Pattern 2: well differentiated carcinoma, with increased stroma.
- Pattern 3: moderately differentiated carcinoma
- Pattern 4: poorly differentiated carcinoma
- Pattern 5: anaplastic carcinoma

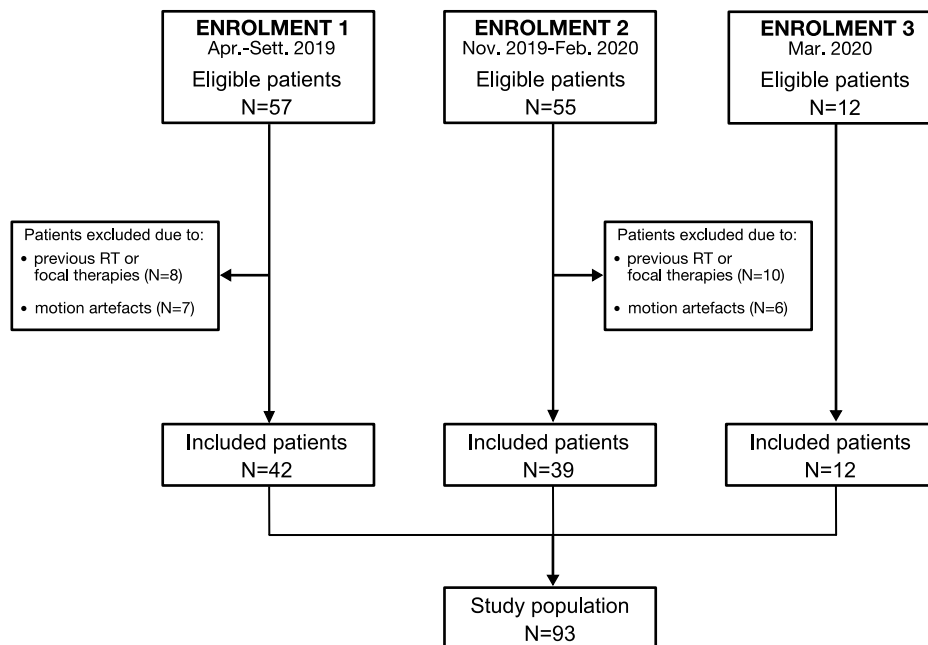
Accordingly, the GS is given by the sum of the two most recurrent patterns arising from tissue inspection. Besides the GS grading system, the ISUP grading system has been conceptualized, also to the aim of distinguishing GS=3+4 from GS=4+3, which both yield GS=7, but have been reported as having a different prognosis. Table 8.7 summarizes the correspondence between Gleason and ISUP score systems.

Table 8.7: Correspondence between GS and ISUP score systems

GS	GS _{sum}	ISUP
3+3	6	1
3+4	7	2
4+3	7	3
4+4	8	4
4+5	9	5
5+4	9	5
5+5	10	5

8.4.1 Study population

The two retrospective studies on PCa refer to three distinct and subsequent processes of patient enrolment, as shown in Figure 8.14. Accordingly, a first preliminary

**Fig. 8.14:** Flowchart of patient enrolment processes in the PCa radiomic study

study has been conducted on the dataset collected after the first enrolment process. Based on the achieved results, a second enrolment process has been carried out, thus extending the sample size of ncsPCa and csPCa. The extended dataset has been employed to conduct a second study. In particular, data analysis for conducting that second study has begun at the end of the second enrolment process. At that

time, a third enrolment process has been further carried out to enrich the sample size used in the second study.

The patient cohort selected through the first enrolment consisted of fifty-seven consecutive patients who have undergone prostate 3T-mpMRI at IRCCS IRST between April and September 2019. The records have been retrieved from the PACS. Eligibility criteria have been: (i) clinical suspicion of PCa (high level of PSA, abnormal digital rectal examination, family history of PCa), (ii) 3T-mpMRI performed at IRST IRCCS, (iii) TRUS biopsy performed as a part of standard-of-care or due to recruitment into other clinical trials at IRCCS IRST. Exclusion criteria have been: (i) absence of previous RT or focal therapies, (ii) motion artefacts. As reported in Figure 8.14, from the first enrolment, eight patients have been excluded due to previous RT or focal therapies, and seven ones because of imaging motion artefacts. In the end, forty-two patients have been included.

Then, a second cohort of patients has been retrieved from the PACS between November 2019 and February 2020, by modifying the eligibility criteria as follows: (i) clinical suspicion of PCa (high level of PSA, abnormal digital rectal examination, family history of PCa), (ii) 3T-mpMRI performed at IRST IRCCS, (iii) TRUS biopsy performed as part of standard-of-care or due to recruitment into other clinical trials at IRCCS IRST, yielding GS at least equals to six (i.e., GS=3+3). Meanwhile, exclusion criteria have been kept fixed. Hence, according to Figure 8.14, from an initial cohort of fifty-five patients, sixteen have been excluded due to previous RT or focal therapies (n=10) and imaging motion artefacts (n=6), thus including thirty-nine patients.

Lastly, the third enrolment process carried out on March 2020, by adopting the same eligibility and exclusion criteria of the second one, has allowed including twelve more patients. In all, the study population includes ninety-three patients, and their clinical characteristics are reported in Table 8.8.

According to mpMRI and biopsy results, patients have been split into seven different groups (G). Besides patients with confirmed PCa having positive both mpMRI (PI-RADS \geq 3) and biopsy (GS \geq 6), split in their turn into five groups according to the five ISUP grade groups (arising from the GS assigned at biopsy), study population has included patients with mpMRI and biopsy both negative (G-NN), reported as unconfirmed PCa, and patients with undetermined clinical outcome due to positive mpMRI and negative biopsy (G-PN).

Both the retrospective studies are IRB approved and the written informed consent has been waived.

Table 8.8: Study population of radiomic PCa study enrolling ninety-three patients split into seven groups, of which five with confirmed PCa (26 ncsPCa and 50 csPCa), one with undetermined clinical outcome showing positive mpMRI and negative biopsy, and one with unconfirmed PCa (mpMRI and biopsy both negative).

mpMRI (P/N)	Biopsy (P/N)	GS	ISUP	Group (G)	PCa	Pats.
N	N	neg.	–	NN	–	7
P	N	neg.	–	PN	–	10
P	P	3+3	1	1	ncs	26
P	P	3+4	2	2	cs	22
P	P	4+3	3	3	cs	14
P	P	4+4	4	4	cs	8
P	P	4+5	5	5	cs	4
P	P	5+4	5	5	cs	2

8.4.2 Objectives of the studies

Two distinct studies have been conducted adopting a radiomic approach for identifying potential IBs of PCa. As already introduced above in Sect. 8.4, both of them exploit one-only DWI sequence, DWI_{b2000} , considered as the one with the highest prognostic potentiality in evaluating PCa aggressiveness. It is also worth noting that employing a single DWI sequence can reduce variability of measures, patient motion artefacts, thus contributing to improve standardization.

The first study, whose experimental results are reported in Sect. 8.4.4, aims at investigating whether radiomic features can stratify patients in four classes of progressive PCa risk level. In particular, if the clinical suspicion is unconfirmed at biopsy (i.e., negative biopsy), two distinct classes of risk of *developing PCa* are considered, depending on it is unconfirmed even at mpMRI or not, that are G-NN and G-PN respectively. On the contrary, if PCa is confirmed (biopsy is positive), again two distinct classes of risk are considered, where the risk is that of *increasing PCa aggressiveness*, from ncsPCa to csPCa, and worsening of prognosis, accordingly. Table 8.9 reports the population involved in this preliminary staging study. In particular, this study refers to the dataset collected after the first patient enrolment process, described above. As a consequence, the population sample, limited to forty-two patients, has allowed exclusively a discrimination study.

Thanks to the second and the third enrolment processes, a wider population of ncsPCa and csPCa has been available. Accordingly, the second study aims at assessing the predictive role of high b-value DWI in classifying ncsPCa and csPCa, through measuring the local tumour heterogeneity in DWI_{b2000} sequence. Consequently, this study, whose experimental results are presented in Sect. 8.4.5, has

Table 8.9: Study population of the preliminary study aiming at staging four progressing PCa risk levels. Accordingly, this study includes the four groups G-NN (n=7), G-PN (n=10), ncsPCa (n=10), and csPCa (n=15).

GS	Group (G)	PCa	Pats.	Total pats.
neg.	NN	–	7	7
neg.	PN	–	10	10
3+3	1	ncs	10	10
3+4	2	cs	5	15
4+3	3	cs	6	
4+4	4	cs	3	
4+5	5	cs	1	

involved the population reported in Table 8.10. In particular, Tables 8.11 and 8.12

Table 8.10: Study population of the classification study for ncsPCa (b=26) and csPCa (n=50)

GS	Group (G)	PCa	Pats.	Total pats.
3+3	1	ncs	26	26
3+4	2	cs	22	50
4+3	3	cs	14	
4+4	4	cs	8	
4+5	5	cs	4	
5+4	5	cs	2	

describe the population of ncsPCa and csPCa collected at the end of the second enrolment process (Table 8.11) – including the ncsPCa and csPCa already enrolled during the first process – and the population enrolled during the third process (Table 8.12), respectively.

Table 8.11: Dataset of ncsPCa and csPCa after the first and second enrolment process

GS	Group (G)	PCa	Pats.	Total pats.
3+3	1	ncs	6	24
3+4	2	cs	19	40
4+3	3	cs	10	
4+4	4	cs	5	
4+5	5	cs	4	
5+4	5	cs	2	

Table 8.12: Dataset of ncsPCa and csPCa enrolled during the third process

GS	Group (G)	PCa	Pats.	Total pats.
3+3	1	ncs	2	2
3+4	2	cs	3	
4+3	3	cs	4	10
4+4	4	cs	3	

8.4.3 Data preparation and feature generation

Images have been acquired with a 3T multi-coil Ingenia MRI system (Philips). mpMRI protocol has included T_2 -weighted, DWI with nine b-values, ADC maps computed exploiting the nine b-weighted DWI sequences, and DCE-MRI sequences. Patient preparation has required fasting 6 h before the examination, bowel preparation to be performed 2 h before the examination, and emptying of the bladder. To reduce peristaltic motion, 1 ml of scopolamine-butylbromide (Buscopan, Boehringer Ingelheim, Ingelheim, Germany) has been administered in a slow bolus infusion at 20 mg/ml, diluted in 10 ml of saline solution.

MRI examinations have been analysed by a radiologist with eight-year experience in urogenital pathologies. Axial T_2 -weighted, DWI, DCE sequences, and ADC maps are considered contemporary for reporting and each detected lesion is assigned a PI-RADS score [PI-RADS V2.1]. Prostate ROIs have been manually outlined on axial T_2 -weighted images and co-aligned on the parallel DWI_{b2000} sequences using Aliza Medical Imaging 1.98.18 (Bonn, Germany - <https://www.aliza-dicom-viewer.com/>). Meanwhile, all PCa lesions having at least a PI-RADS score equal to three have been segmented on DWI_{b2000} sequences. PCa ROIs have been outlined slice by slice along the most emphasized internal boundaries detected on DWI_{b2000} , also using cognitive fusion of all available MRI sequences, especially useful in case that motion artefacts, intestinal air, hip prosthesis artefacts are present. While PCa lesions in the peripheral zone (PZ) have been segmented directly on DWI sequences, for central and transitional zone lesions ROIs have been outlined on DWI_{b2000} and refined using the cognitive fusion of parallel axial T_2 -weighted.

By recalling the biopsy approach in sampling prostate tissue, the set of eighty-four imaging (Sect. 8.2) has been extracted from PCa ROIs or from those patients with positive mpMRI (G-PN, and $G=[1\div 5]$), whilst they have been computed on the whole prostate ROI for patients with negative mpMRI (G-NN group).

8.4.4 Preliminary staging of four progressive PCa risk levels

As introduced above, this study performs a preliminary discrimination of four progressive PCa risk levels, G-NN, G-PN, G-1 (i.e., ncsPCa), and G-[2÷5] (i.e., csPCa) based on the dataset reported in Table 8.9. Accordingly, this Section reports data analysis and experimental results.

RFs selection and reproducibility analysis

Kruskal-Wallis (p -value <0.001) has been first performed for the multiple comparison of the four groups (i.e., G-NN, G-PN, ncsPCa, and csPCa) after considering Bonferroni correction (p -value $<10^{-5}$). Hence, the radiomic features yielding the lowest p -value has been selected. Then, the one-tail Wilcoxon rank-sum test has been employed to assess the pairwise separation of the four groups. Finally, the potential role in patient stratification has been assessed by the Spearman index ($p<0.001$) through the rank correlation of the selected radiomic feature with the four classes. In order to assess the reproducibility of the selected radiomic feature, I have considered to perturb the segmentation performed by the radiologist. To this aim, prostate and lesions ROIs have been dilated and eroded by means of binary morphological operations, to lead heavy averaged changes in segmentation of about $\pm 25\%$ of the volumes. In particular, prostate ROIs of each slice have been isotropically dilated and eroded with a 5×5 square structuring element, while for lesion ROIs, a much smaller 3×3 square structuring element has been used. Hence, the mean absolute variations between the radiomic feature extracted from the dilated/eroded volumes and the original ones have been computed, and their reproducibility has been assessed through the ICC.

Experimental results

One out of eighty-four radiomic features has been selected, the mean of local coefficient of variation (CV_{L-m}), providing a measure of the local variability of DWI_{b2000} values. Figure 8.15 shows the boxplots of the four groups, with the clear separation confirmed by the Kruskal-Wallis multiple comparison ($p\sim 10^{-6}$). Moreover, Wilcoxon confirms the pairwise separation between G-NN and G-PN (p -value=0.005), G-PN and G-1 (p -value=0.026) and, of course, G-1 and G-[2÷5] (p -value=0.008). The reported median values for CV_{L-m} are in increasing order, 10.7%, 13.2%, 15.8% and 20.1%, respectively, meaning that CV_{L-m} can stratify the 4 groups according to their expected risk level, this confirmed by the strong Spearman correlation $\rho_s=0.81$ ($p\sim 10^{-6}$). Moreover, this Spearman correlation, $\rho_s=0.81$, results still valid when the separated seven groups are considered (i.e., when csPCa

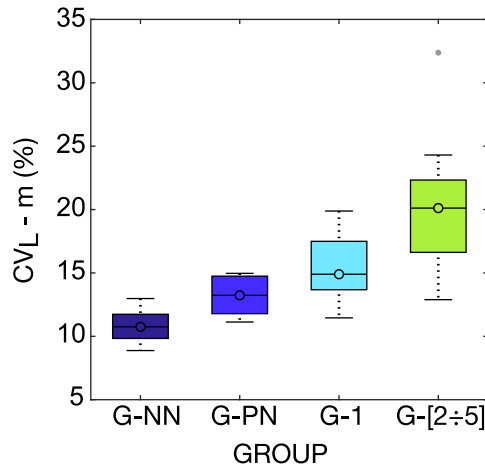


Fig. 8.15: Boxplot of the stratification of four progressive PCa risk levels, G-NN, G-PN, ncsPCa, and csPCa performed by CV_{L-m} , with $p \sim 10^{-6}$ at Kruskal-Wallis.

group is split into four ones according to either GS or ISUP groups, as in Table. 8.9). The selected radiomic feature, CV_{L-m} , also shows a great reproducibility, with $ICC=0.99$ for both dilatation and erosion, with $\Delta=3.54\%$ for dilatation and $\Delta=1.57\%$ for erosion.

In the end, these preliminary results show that one radiomic feature based on local variability of high b-value DWI allows representing the known inhomogeneity of tumour habitat and complexity of tissue microstructure worsened by reduced intracellular space. CV_{L-m} has a high propensity in risk stratification, as established by differentiating G-NN, G-PN, G-1, and G-[2÷5]. In fact, this radiomic feature naturally sorts the median values of these four groups, with a very low overlapping between boxplots (Figure 8.15). This finding strongly supports the promising role of CV_{L-m} in characterizing the progressiveness of tissue heterogeneity in groups of patients having a potential increasing risk of developing PCa or having increasingly aggressive PCa.

As a matter of fact, the relevance of CV_{L-m} to represent the heterogeneity of the MRI imaged values is present in [277], when it is used to assess the heterogeneity of images acquired at different resolutions. Then, CV_{L-m} has also been employed on lung and liver to detect abnormalities within locally structured patterns [278]. It is worth noting that the high value of the Spearman correlation among CV_{L-m} and the seven groups (Table. 8.9) also emphasizes the promising role of the radiomic feature in depicting local changes hinting at increasing malignancy. To the best of my knowledge, this is the first study reporting such a risk stratification for PCa, including patients from unconfirmed PCa (G-NN) up to csPCa. Results achieved so far, albeit preliminary, are a proof of what has motivated my research, that is,

choosing multiple b-values would hardly lead to reproducible parameters and results. Above all, results confirm that DWI_{b2000} retains the highest sensitivity for measuring imaging features of PCa aggressiveness. As a concluding remark, these preliminary results have paved the way to a more focused study on classifying ncsPCa and csPCa, which is presented in Sect. 8.4.5.

8.4.5 Prediction of clinically significant PCa

According to the objective of the second study introduced in Sect. 8.4.3, this study regards the patient cohort of ncsPCa and csPCa reported in Table. 8.10. In the following, the development of the csPCa predictive model will be explained in detail, and experimental results will be discussed.

Predictive model

A radiomic model has been built to recognize csPCa (TPs), distinguishing them from ncsPCa (TNs), according to the process outlined in Figure 8.16. All radiomic

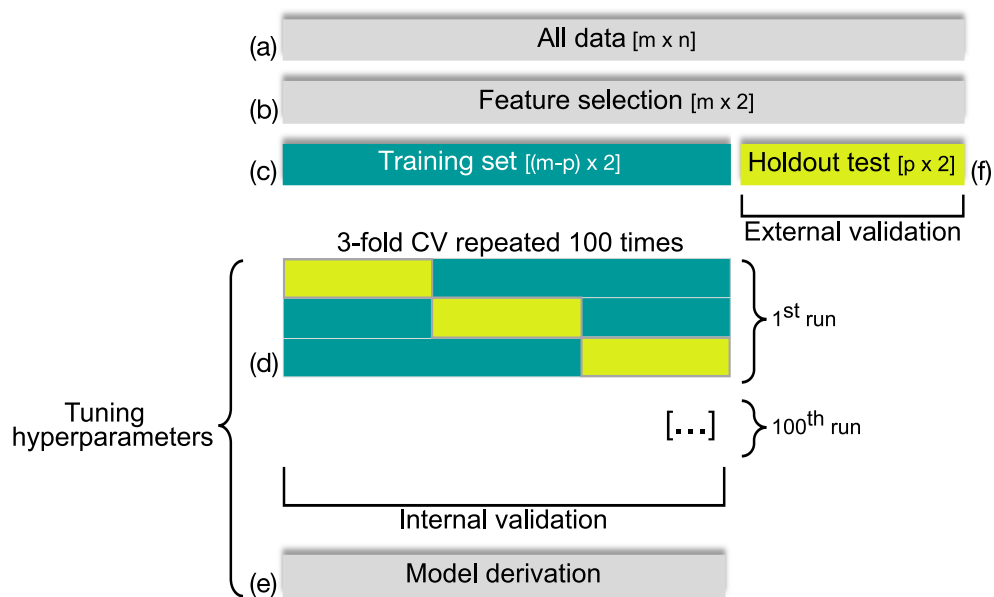


Fig. 8.16: Predictive model building

features (Figure 8.16 (a)) have been normalized and standardized, and redundant and irrelevant radiomic features have been removed through LASSO, with the optimal tuning parameter (λ) selected using 10-fold CV (Figure 8.16 (b)) and the minimum CV error rule. To prevent overfitting, two only radiomic features have been considered from the subset of the ones selected by LASSO, using a two-stage

approach. First, the couples with a high Pearson correlation ($\rho \geq 0.15$) have been discarded. Second, the most discriminant couple of radiomic features (i.e., yielding the lowest p -value according to the Wilcoxon rank-sum test) has been selected from those surviving the previous step.

The entire dataset available at the end of the second enrolment process, reported in Table 8.11, has been split into training and (holdout) test set, made of 48 and 16 patients, respectively (Figure 8.17). The training set consisted of 30 TP and 18 TN,

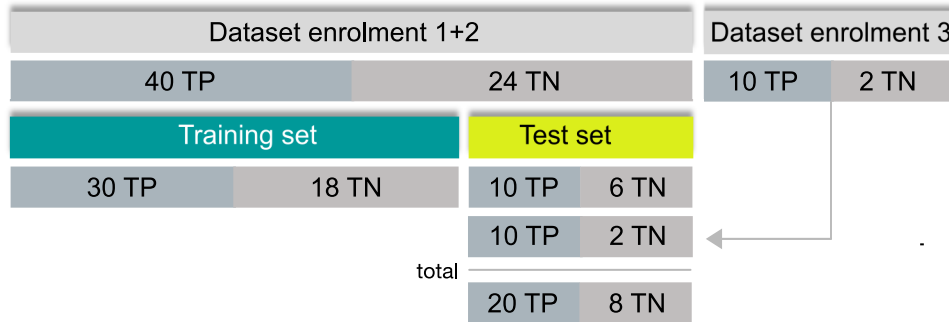


Fig. 8.17: Splitting of training and test set from the entire dataset

whilst the test set comprised 10 TP and 6 TN. Then, the dataset available after the third patient enrolment process, and reported in Table 8.12, has been integrated in the holdout test set, once the training phase of the predictive model was concluded. Accordingly, the holdout test set included at last 20 TP and 8 TN.

To preserve the representativeness of the training set without degrading the generalization performance, the training set has been derived from the entire dataset (sixty-four patients of Table 8.11) to include the patients candidate to represent the SVs of a SVM classifier, according to the method described in [279], based on their distance from the separating hyperplane. In particular, the SVM margin (SVM_m) has been computed and patients have been ranked into three classes (Table 8.13), based on their SVM_m . Class 1 corresponds to $SVM_m > 1$, patients correctly classified ($n=40$). Class 2 corresponds to $0 \leq SVM_m \leq 1$, patients falling within the decision surface. Class 3 corresponds to $SVM_m < 0$, patients misclassified ($n=14$). Hence, the training and test set have been selected to balance the distribution of patients belonging to each class, as reported in Table 8.14.

Then, the SVM classifier with linear kernel has been trained on the training set (Figure 8.16 (c)) with a repeated 3-fold CV (as the one explained in Chapter 5, Sect. 5.10, Figure 5.11 (d)), as reported in Figure 8.16 (d), for tuning the SVM hyperparameters, that is, the kernel scale and the global misclassification cost, C . The cost C is then scaled by the weight of the error occurring in each class, which corresponds to its own prior probability. Each fold is made of sixteen patients, 10

Table 8.13: Derivation of the training set based on the SVM margin criterion

SVM_m	Class	Patients
$SVM_m > 1$	1	40
$0 \leq SVM_m \leq 1$	2	11
$SVM_m < 0$	3	14

Table 8.14: Selection of training and test set to balance the number of patients belonging to the three classes defined according to their SVM_m value.

	Class 1	Class 2	Class 3
Training set	32	7	10
Test set	8	4	4

TP and 6 TN. To prevent any spurious solution, an internal validation procedure is performed by one hundred repetitions of 3-fold CV. The model selection procedure has been carried out into five steps. For each round, the ROC curve and the corresponding AUC are computed for training and validation sets, and the models most prone to overfitting, yielding a AUC on the validation set higher than that on the training one, are discarded (step 1). Based on radiologist's indication, missing a csPCa has been considered much worse than yielding a false alarm to ncsPCa, therefore I have computed the F_2 -score on the validation set and used it to select the best SVM model (step 2), thus achieving at most one-hundred competing models, corresponding to as many runs of CV. Among the models coming through the first two steps, an early selection is carried out by analysing their performance on the training sets, discarding the models with a very low C parameter ($C < 1$), more prone to overfitting (step 3) and with F_2 -score < 0.80 (step 4). At the end, the model showing the highest F_2 -score on the validation set (Figure 8.16 (e)) is selected as the ultimate predictive model, to be externally validated on the holdout test set (step five). The performance of the SVM classifier are assessed through AUC, and SE, SP, and I, measured at the Youden cut-off. PPV and FDR are computed accordingly.

Experimental results

Biopsy examination is presently the reference clinical tool for distinguishing csPCs from ncsPCa, starting different clinical paths, that is, curative treatments or active surveillance, watchful waiting, and observation, respectively [280]. mpMRI has having a more and more crucial role in the pre-biopsy patient management,

to prevent patients to undergo unnecessary operations [269] which are known to cause side effects in about the 30% of men, 1% of which requires hospitalization for observation [270]. A radiomic and quantitative mpMRI-based imaging approach is frequently adopted in PCa study with the aim of enriching the radiological assessment of medical images providing additive information referred to tumour aggressiveness and prognosis, for instance, to distinguish csPCa from ncsPCa, prior to biopsy. However, a “considerable overlap between csPCa and ncsPCa in mpMRI parameter values” is known [273] and it represents the major limitation for mpMRI to replace the biopsy in patient staging [273]. At present ADC is considered, among the mpMRI sequences, as the most promising tool available for quantitative image analysis. In particular, the ADC images have been very successful in the clinical routine, mainly for two reasons. From the one hand, they allow reconstructing the diffusion-weighted information achieving a SNR much higher than that of native DWI. From the other hand, they allow preserving the morphology, especially if compared to high b values, and annulling the artefacts of DWI images, like the T_2 shine artefact, which are known to mislead the assessments of suspicious malignant areas. Consequently, the ADC sequences have become the reference ones for confirming diagnosis of PCa and, as such, they have been largely employed even to extract information as regards PCa prognosis.

In fact, if considering the scientific works published in the last five years (from PubMed database) which implement a predictive model of csPCa (independently of the lesion zone) and reported in Table 8.15, one can see that, except for [281], most

Table 8.15: Comparison of our study with the state of the art

Year, Authors	mpMRI sEquation	Feature	AUC	SE	SP	I
[269] 2015, Fehr <i>et al.</i>	T_2 , ADC	18 rad. feat.	0.83	–	–	–
[273] 2017, Barbieri <i>et al.</i>	ADC, IVIM	$ADC_{m(b[0-900])}$	0.79	0.85	0.74	0.59
[276] 2018, Bonekamp <i>et al.</i>	T_2 , ADC	10 rad. feat.	0.88	0.97	0.58	0.55
[281] 2019, Cristel <i>et al.</i>	DCE-MRI	k_{trans}	0.75	0.95	0.61	0.56
[282] 2019, Min <i>et al.</i>	T_2 , ADC, DWI_{b1500}	9 rad. feat.	0.82	0.84	0.73	0.57
[274] 2020, Hiremath <i>et al.</i>	ADC	$ADC_{m(b[0-1300])}$	0.85	0.77	0.81	0.58
[275] 2020, Zhang <i>et al.</i>	T_2 , ADC, DWI	10 rad. feat.	0.81	0.80	0.73	0.53
2020, our work	DWI_{b2000}	2 rad. feat.	0.84	0.90	0.75	0.65

of these works utilize the ADC sequence [274], sometimes coupled with T_2 -weighted ([269], [275], [276], [282]), whilst one work only couples ADC with IVIM parametric maps [273]. However, also in this last case, the best result reported refers to the mean valued of ADC maps, ADC_m .

Actually, with the coming of the 3T MR systems, which allow obtaining DW images with an intrinsically higher SNR, there is no longer any real need to limit the

quantitative analysis of tissue diffusivity to ADC sequences only. On the contrary, our results confirm our hypothesis that high b-value DWI (e.g., DWI_{b2000}) images can embody information regarding tissue heterogeneity and tumour functional properties with specificity and sensibility higher than ADC maps can do.

High b-value DWI has already shown to increase both reader's sensitivity [283] and radiomic accuracy in distinguishing PCa from non-cancerous lesions [284], although a limited success is reported in recognizing csPCa and ncsPCa so far. The authors in [273] even state that DWI sequences are not yet feasible for reliable clinical indications of tumour prognosis and, moreover, they cannot bring any added value with respect to the ADC sequence in identifying csPCa.

In this study, I have developed a predictive model exclusively based on DWI_{b2000} . In particular, as regards feature selection, LASSO yields ten radiomic features, in correspondence of the optimal $\lambda=0.0325$, whose coefficients are reported in Figure 8.18 (a) according to their rank. The correlation coefficients computed be-

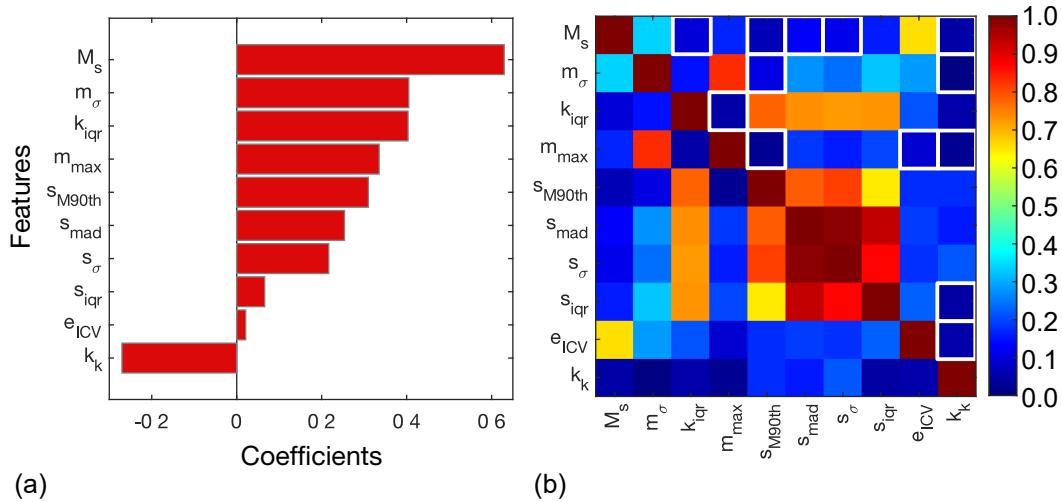


Fig. 8.18: Ranking of radiomic features selected by LASSO (a) and their correlation matrix (b), where the uncorrelated couples are highlighted with white squares.

tween all the couples of radiomic features are resumed in the matrix shown in Figure 8.18 (b), where the white squares highlight the thirteen uncorrelated couples. Hence, I have selected the most discriminant couple of radiomic features having a p -value $\sim 10^{-7}$. In particular, that couple is composed by the standard deviation of the mean, m_σ , and the median of the last decile of the skewness, s_{M90th} , whose LASSO coefficients are 0.405 and 0.310, respectively, corresponding to the second and the fifth radiomic features in Figure 8.18 (a). Accordingly, one can notice that the two features give information regarding the heterogeneity and the degree of asymmetry among DWI_{b2000} image values.

Figure 8.19 (a) shows the ROC curve achieved by the couple m_σ - s_{M90th} on the

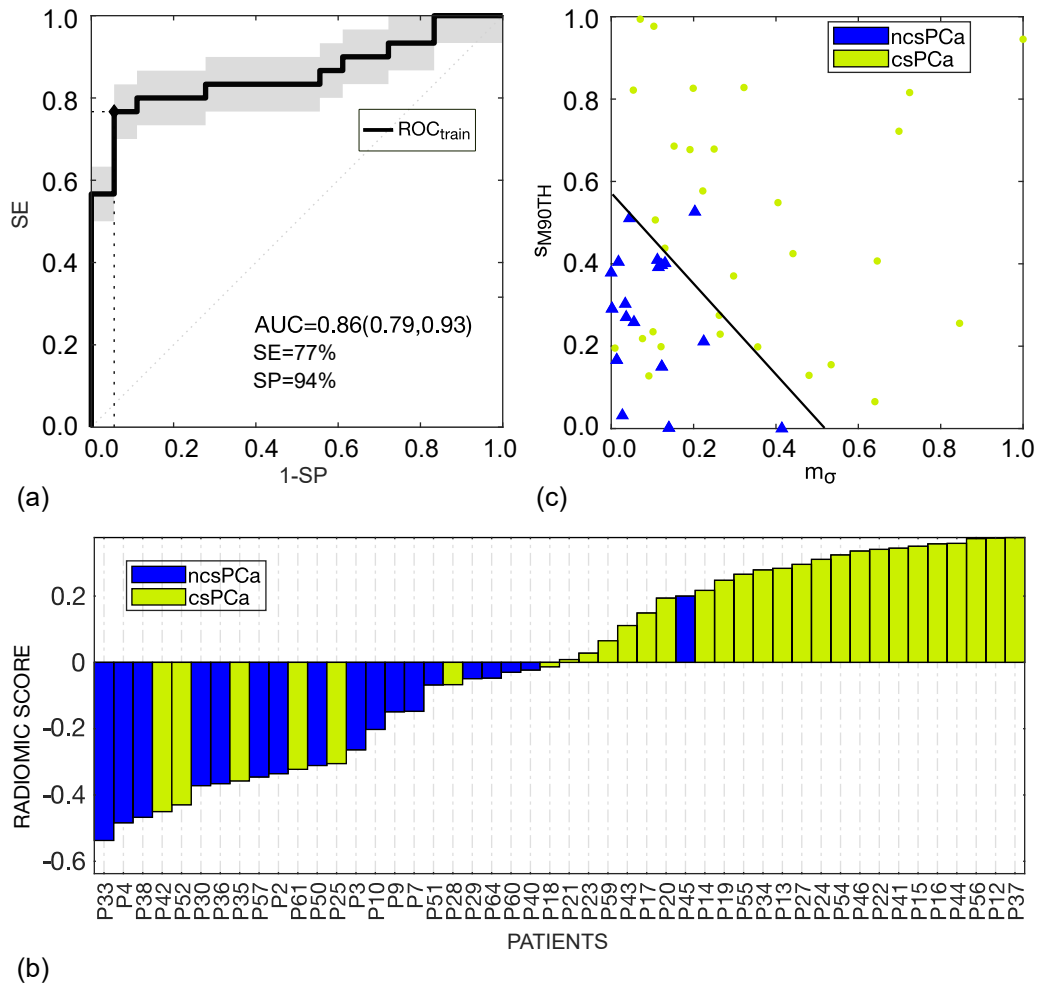


Fig. 8.19: Performance of the predictive model on the training set. The ROC curve (a) yields $AUC=0.86$ (95% CI, 0.79-0.93), and $SE=77\%$, $SP=94\%$, at the Youden cut-off ($I=0.71$). The waterfall plot (b) with blue (ncsPCa) and green (csPCa) bars shows 1 FP and 7 FN.

training set, where csPCa are predicted with $AUC=0.86$ (95% CI, 0.79-0.93), and $SE=77\%$, $SP=94\%$, at the Youden cut-off ($I=0.71$). Such a good separation is also shown in the waterfall plot reported in Figure 8.19 (b), where blue and green bars highlight ncsPCa and csPCa, respectively. As one can notice, even through the scatter plot in Figure 8.19 (c), where the SVM separation hyperplane is highlighted in black, the separation yields on the training set seven FN and one only FP, thus leading $FDR=4\%$, $PPV=96\%$, and $F_2\text{-score}=80\%$ accordingly. Similarly, Figure 8.20 (a) reports the ROC curve achieved on the holdout test set which quite confirms the performances of the couple m_σ - s_{M90th} already obtained on the training set, with a close value of $AUC=0.84$ (95% CI, 0.63,0.90), and $SE=90\%$, $SP=75\%$ at Youden cut-off ($I=0.65$). Hence, in Figure 8.20 (b), one can appreciate the waterfall plot

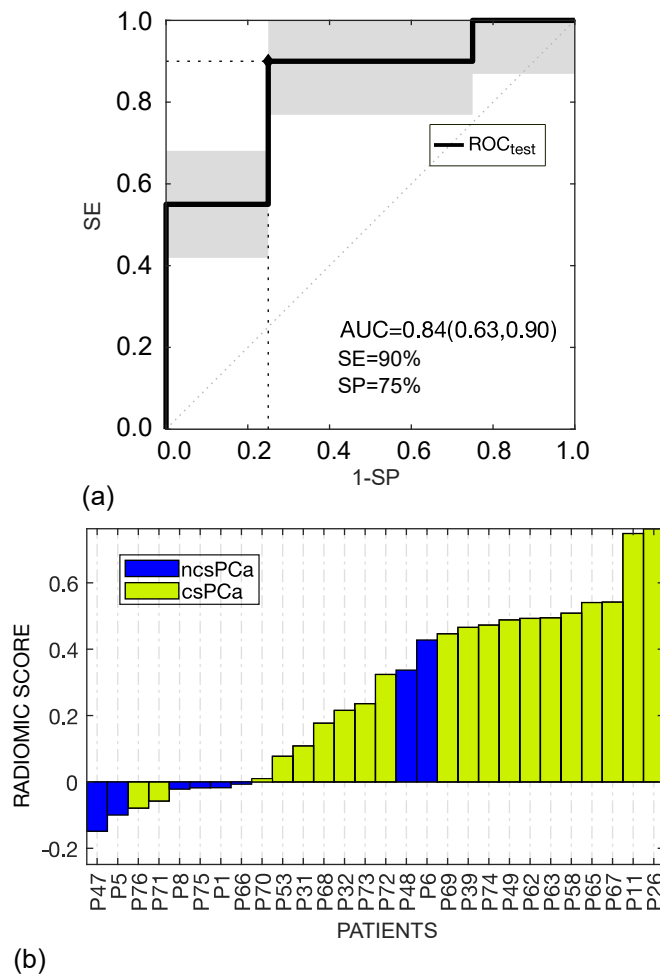


Fig. 8.20: Performance of the predictive model on the holdout test set. The ROC curve (a) yields $AUC=0.84$ (95% CI, 0.63, 0.90), and $SE=90\%$, $SP=75\%$ at Youden cut-off ($I=0.65$). The waterfall plot (b) with blue (ncsPCa) and green (csPCa) bars shows 2 FP and 2 FN.

referred to the test set, where prediction of csPCa is achieved with two FP and two FN. Accordingly, $FDR=10\%$ and $PPV=90\%$, and $F_2\text{-score}=90\%$. Then, it is worth noting by considering the performance of the predictive model on both training and test sets that it reaches to main goals. On the one hand, it substantially improves (with those high values of PPV) the prediction of csPCa with respect to the clinical mpMRI used in triage pre-biopsy setting which reaches at most $PPV=51\%$ [271]. On the other hand, the radiomic model kept bounded the risk of overtreatment, as reported by the low FDR values, thus confirming the high potential role of radiomic MRI in the clinical decision making. In fact, overtreatment of ncsPCa is reported as being the major side-effect of the high-sensitivity tests used for revealing the tumour malignancy degree [285]. A further confirmation of such a marked separation achieved between ncsPCa and csPCa arises from the boxplots shows in Figure 8.21

for the training (Figure 8.21 (a)) and test (Figure 8.21 (b)) sets, respectively. In

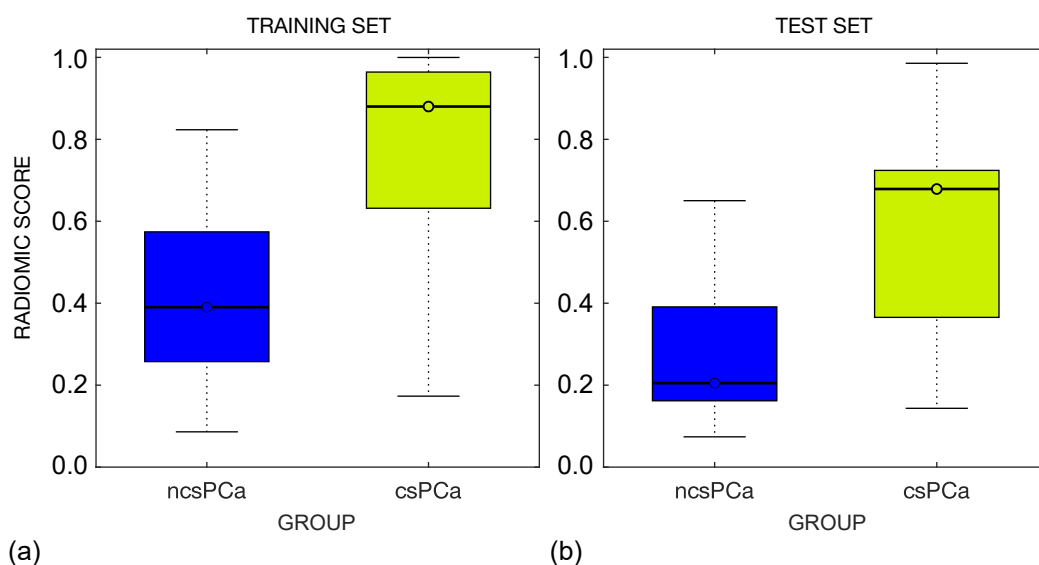


Fig. 8.21: Boxplot of ncsPCa and csPCa in training (a) and test (b) sets.

fact, the two groups are separated with a p -value $\sim 10^{-5}$ in the training set, with the M of the two groups being equal to 0.39 (ncsPCa) and 0.88 (csPCa). Similarly, on the test set, p -value $= 7 \cdot 10^{-3}$, and the M values of the radiomic score are 0.20 (ncsPCa) and 0.68 (csPCa).

One can refer to Table 8.15 to compare the performance of this predictive model with the state of the art. One can see that the work of [273], where the classification is performed exclusively with ADC_m , computed between $b=0$ and $b=900 \text{ s/mm}^2$, reports nearly the worst values of AUC (AUC=0.79), with $I=0.59$. Analogously [281], the only work using the DCE-MRI, reaches at most AUC=0.75, the worst considered, with $I=0.56$, substantially confirming the direction of the present guidelines PI-RADS v2.1, where “DCE-MRI has become secondary to DWI and T_2 -weighted images”, also considering that prostate DWI allows an easier image acquisition and processing if compared with other functional MR techniques [283]. Actually, two of the works considered, the first one employing ADC_m [274] and the second one a radiomic signature where 7 out of 10 radiomic features are extracted by the ADC map [276], achieve quite high AUC values. In fact, AUC=0.85 in [274] e AUC=0.88 in [276], even though with low I values, $I=0.58$ and $I=55$, respectively, quite lower than ours ($I=0.65$). Two works only include some native DWI sequences for extracting the radiomic signature, with $b=1500 \text{ s/mm}^2$ in [282] and $b=0, 1000 \text{ s/mm}^2$ in [275]. However, although the work in [282] reports a good AUC=0.82 value, but $I=0.57$, only one of the nine features composing the signature is extracted from the DWI sequence, besides being the second-last by importance, while in [286], where the

signature is made by 5 radiomic features, out of ten, extracted from DWI, a quite high $AUC=0.81$ value is coupled with the worst I result ($I=0.53$). Finally, [269] seems to achieve a result quite similar to ours in terms of $AUC=0.83$, but not any other metric is provided for a deep comparison.

On the whole, it seems that ADC, although being largely employed, cannot offer the performance of DWI in detecting csPCa. This is due to the ADC parametric maps – as introduced above – which arise from a normalization procedure between DWI images at different b-values. In fact, normalization implicitly acts as a low-pass (averaging) filtering of the local value differences between adjacent structures, thus weakening the native information conveyed by the original DWI sequences. In many works, DWI has been reported as “the best monoparametric component of prostate MRI assessment” [276], where “quantitative analysis at high b-value DWI” (from $b=1000$ to $b=2000$) “suggests” the highest sensitivity of DWI in both detecting PCa [283] and staging high grade diseases [287], but it has had a limited diffusion in radiomic studies so far.

At the end, I agree that visual-based tumour detection and segmentation can be performed with a much higher accuracy on the ADC sequences, and these should remain the reference tool for visual assessments and ultimate confirmation of cancer diagnosis. Nonetheless, our results and some literature strongly suggest that they cannot be the best tool for quantitative imaging, since the information extracted is far beyond what even expert eyes can visually detect.

Accordingly, the native DWI information can have a higher specificity, from a quantitative point of view in detecting/catching the cellular differentiation degree needed to distinguish csPCa from ncsPCa. [276] reports that the good performance of the radiomic model and of the ADC_m are equivalent. Also based on our results, this suggests that a radiomic analysis carried out on DWI images rather than on ADC maps can yield a marked advantage, whether the original information is either visual or semi-quantitative.

One final consideration is worth to be reported. Often, the signal restriction in ADC has been attributed to the iper-cellularity process associated, in its turn, at a progression in terms of tumour aggressiveness. Actually, the work of [272] shows as the ADC signal restriction is only weakly correlated to the main cell metrics (nuclear count, nuclear area), but the stronger correlation is reported with the variation of gland component volumes (epithelium, stroma, and lumen). The tumour progression attributed to a higher GS, results to be associated with an increasing volume of low-diffusivity epithelial cells and decreasing volumes of high-diffusivity stroma and lumen space. Accordingly, GS grade definitions relies on changes of tissue architecture, which makes the tumour progressively more heterogeneous and less differentiated as malignancy increases. Then, it is worth noting that our two radiomic features extracted from DWI_{b2000} are two direct measures of tissue asym-

metry and local variability in tissue diffusivity. DWI_{b2000} seems to catch with a high specificity the asymmetry gradients found between the local propriety of tissue diffusivity, following the disproportion between the gland components.

Chapter 9

Conclusions

In this Thesis I have exploited principles of quantitative imaging to improve reproducibility of CTP and reliability of radiomic MRI analysis to the aim of developing reliable biomarkers and favouring their translation into clinical routine. Both dynamic imaging, such as CTP, and machine learning techniques applied to MR image analysis have aroused great expectations in the medical community, since they allow investigating those cancer properties, including angiogenesis and heterogeneity, where the major interests of pharmaceutical and clinical fields converge. In fact, a deep comprehension of specific cancer features and mutual interactions occurring between cancer and its surrounding habitat should advantage the development of efficacious drugs targeted against different tumour subtypes, thus promoting the role of precision medicine in cancer care. However, although for very different reasons, both CTP and radiomics applied to MRI have encountered methodological and technical limitations, which have hampered their clinical applicability so far.

On the one hand, the entering of CTP in the clinical practice has been hindered by the difficulty of the different perfusion computing methods to yield reproducible results, and the huge patient's radiation dose exposure required from CTP acquisitions. On the other hand, the expansion of the machine learning techniques applied to MRI analysis by research groups with different degrees of expertise have led to an increased variability of results arising from non-standardized procedures. Yet more, the radiomic analytic approach often relying on a huge amount of features with poor interpretability have weakened the clinical meaning of radiomic outcomes and damped the enthusiasm of clinicians for the real benefit of machine learning in clinical practice.

In this Thesis, I have adopted common principles for improving CTP and MRI radiomics by analysing technical and modelling aspects at single-voxel level, which has allowed reducing variance of measurements while increasing their accuracy and precision. Moreover, in several occasions, my approach has highlighted relevant clinical information benefiting the comprehension of both physiological and pathological

tissue properties.

In particular, in CTP analysis, the methods developed for extracting the first pass tracer kinetics from extended CTP signals (i.e., including first pass and re-circulating phases), have removed the computational sources of variance, and have allowed me to perform a fair comparison between two of the most widely adopted perfusion methods, MS and DV. Therefore, I have discovered that the causes of unreproduceability between MS and DV has to be attributed to theoretical aspects. In fact, the reproducibility of DV and MS as it was originally formulated, has not to be expected because the methods are representing two different status of the system. In fact, MS in its classical formulation neglects any vascular transport delay between vascular inputs and tissue, which instead have resulted the major cause of such non-reproducibility. Yet more, I have formulated an alternative MS, the gMS, to allow MS to represent the same temporal status of DV, thus achieving the equivalence of measurements through gMS and DV. I have had an earlier proof in hepatic perfusion, where a substantial delay in tissue modelling is introduced by considering the aortic artery as one of the vascular inputs, in place of the hepatic one (not detectable in CTP sequences). In addition, I have obtained a proof of concept on a CTP case study of colorectal cancer where no delays were expected between input and tissue, and MS and DV are resulted equivalent.

Future works will inquire on whether pathological conditions, such as cancers at different stages, could affect the vascular transport inducing varying delays. Yet more, it could be worth investigating whether the efficacy of anti-angiogenic drugs induces the tissue recovery from a delayed transport.

In radiomic MRI analysis, I have developed a two-stage strategy to characterize each voxel based on its neighbourhood. First, I have computed locally FO features, thus enabling the assessment of their spatial distributions through colorimetric maps. Second, information retained by multiple parametric maps corresponding to different cancer image slices has been summarized by simple global statistical descriptors, which have allowed achieving robust and reliable measurements referred to the entire cancer volume. This radiomic analytic approach has allowed investigating local tissue properties, providing insights into cancer heterogeneity and how it varies under the effects of treatments. Yet more, this investigating strategy is based on FO features computed locally, which are easily interpretable and allows correlating the spatial variations of the features to the morphology of underlying anatomical structures. This is really helpful in favouring the clinical interpretation of features by radiologists. It is worth noting that, even the global descriptors computed on local parametric feature maps, resulted good estimators of local variability. In fact, when this strategy of imaging feature computation have been applied on the studies on LARC and PCa respectively, it has allowed achieving marked separations between different classes, employing a very low number of features. Ultimately, one should

notice that this radiomic analytic approach is ready to be validated on a larger patient population, meanwhile it could be already used in prospective studies to support radiologists in the clinical decision making.

Future works will aim at deepening to what extent these imaging features can describe tissue heterogeneity and how they correlate with biological and histopathological findings to allow a unified framework of cancer investigation.

Appendix

CTp software development and management

Data processing for perfusion analysis, from data collection to computation of perfusion parameters, has a large and complex workflow, and it is fundamental to keep track of each step for reproducibility and repeatability of measurements. Working with multicentre studies introduces an additional complexity, because a large amount of data needs to be handle contemporaneously, and this may easily lead a young researcher to get lost in data storing, managing, and analysis.

Me too, when starting coping with PIXEL data, had some difficulties. How organizing the code? How controlling the results correctness? How preventing possible errors when setting parameters required by the functions? How storing results? How storing information for recovering them rapidly in each moment? When addressing a scientific research, the theory can be handle by many easily, while its implementation is the crucial point, which can even jeopardize the reliability of methodology itself. In several cases, the discordance between the theory and its scientific implementation yield results non-compliant with the scientific methodology. Therefore, I decided to spent a couple of months to design, define, and implement a prototype of software architecture for CTp analysis according to the principles of modularity and reusability. Not being neither a computer scientists or a computer engineer, I had to study fundamentals of software development and management and learn how to implement them for scientific projects. The software has been implemented in MatLab[®].

The main topics addressed when defining the architecture of the CTp analysis software are the following:

- **Software organization**

Its meaning is twofold. First, it refers to the design of the data storage architecture based on hierarchical structures. Second, the definition of the relationships which may occur among main code, procedures, and functions, calling each other during software execution.

- **Software maintenance**

All implementation strategies should preserve the logic structure of the software over time. This means that any software update must never require radical changes of the original logic structure. Accordingly, software has to be designed to be flexible, to adapt to any new requirement, and safety, for avoiding errors or loss of data.

- **Data recovering**

This issue is strictly related to data collection. The most effective way for storing information allows recovering them exhaustively in the shortest possible time.

- **Association of results – generation mode**

Most importantly, the store of each data processing outcome, including figures and graphs, has to include its generation mode, the software procedure used and its running parameters, in order to reproduce that outcome anytime.

Figure 9.1 shows the chart of CTP analysis software architecture, comprising both the structure of repositories and code organization. Contents are organized employing a tree-structure split into multi-levels.

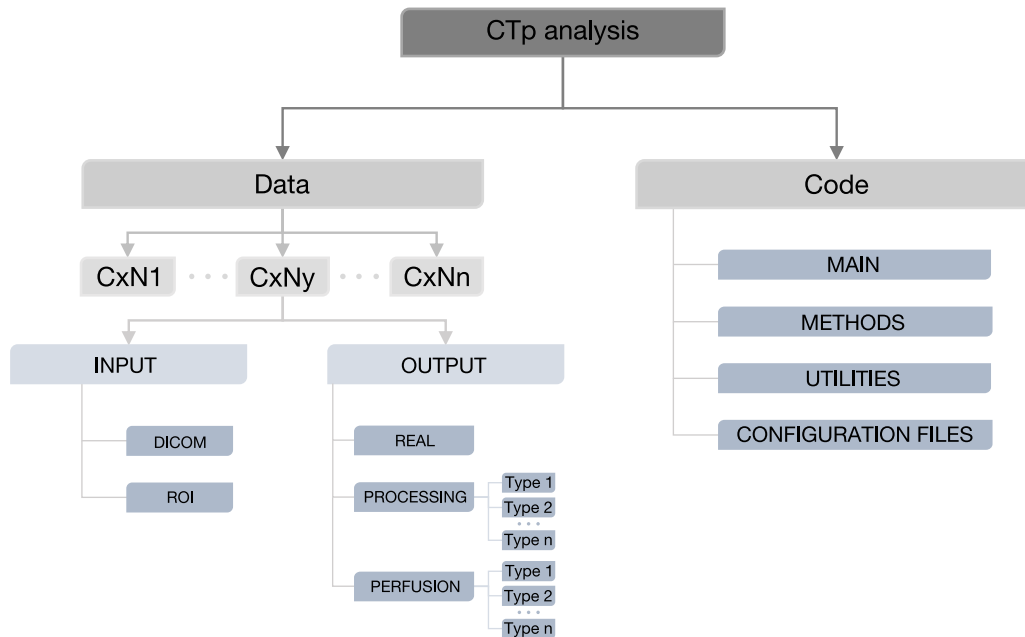


Fig. 9.1: Chart representing the architecture of the prototypic software for CTP analysis. It has a tree-structure, split into multiple semantically coherent levels.

The folder DATA is split at the first level in multiple folders, referred to as CxNy, corresponding to the generic Centre (C) x , Patient Number (N) y . Each CxNy folder contains the INPUT and OUTPUT folders. The former includes the input data, DICOM and ROI files into two separate folders, respectively, whilst the latter

collects the outcomes of preprocessing and processing procedures. In particular, the repository OUTPUT contains in its turn three sub-folders. The one named REAL collects the **.mat* files related to TCCs extracted from CTP image sequences, from vascular input and tissue ROIs, separately. The folder PROCESSING contains different folders, each referred to as Type n, corresponding to different processing functions, like signal fitting or interpolation. For instance, this folder currently includes Interp, FittingA, FittingP, and FittingT, corresponding to the resampling of tissue TCCs, and the fitting of aortic (FittingA) and portal (FittingP) vascular inputs or tissue (FittingT), respectively. Similarly, the folder PERFUSION collects the outcomes arising from different computational methods for perfusion parameters (e.g. MS, DV, etc.).

Instead, the CODE folder is organized into four different sub-levels. MAIN contains the main **.m* file, controlling the execution of the whole software. All functions and procedures called by the *main.m* file are contained into METHODS and UTILITIES depending on whether their employment is specific or transversal across different purposes, respectively. Finally, the folder CONFIGURATION FILES collects several **.ini* files, each reporting the running parameters for each procedure called by the main code during each specific run. Hereinafter, the term function will be adopted to refer to the methods included in the CTP software for signal processing and computation of perfusion parameters.

In particular, the execution of the main code is organized as follows, also described through the flowchart in Figure 9.2.

As first, the software executes an initialization procedure, where the user is asked to select the modality “write” or “read” for the configuration file by activating and deactivating the corresponding flags. If the “write” modality is selected, the user has to provide all running parameters for all the available procedures or just for those of interest for the specific purpose. Actually, a new configuration file can be written directly through the initialization procedure in MatLab or through any text editor by following the template defined in the initialization procedure. Otherwise, if the “read” modality is chosen, the user has to indicate the name of the configuration file to be loaded from the CONFIGURATION FILES folder.

Each configuration file contains an initial general section with a series of flags, one for each possible executable methods, which is activated or deactivated for allowing its execution or not. Then, there are as many specific sections as the number of implemented procedures, where all running parameters can be set.

Two different types of configuration files are adopted, global and local ones. Generally, the user creates a new global configuration file only if changes regard multiple methods, otherwise, variations of just execution parameters are stored into local configuration files. In this regards, the automatic process of generation of local configuration file will be clarified after explaining the whole process of execution of

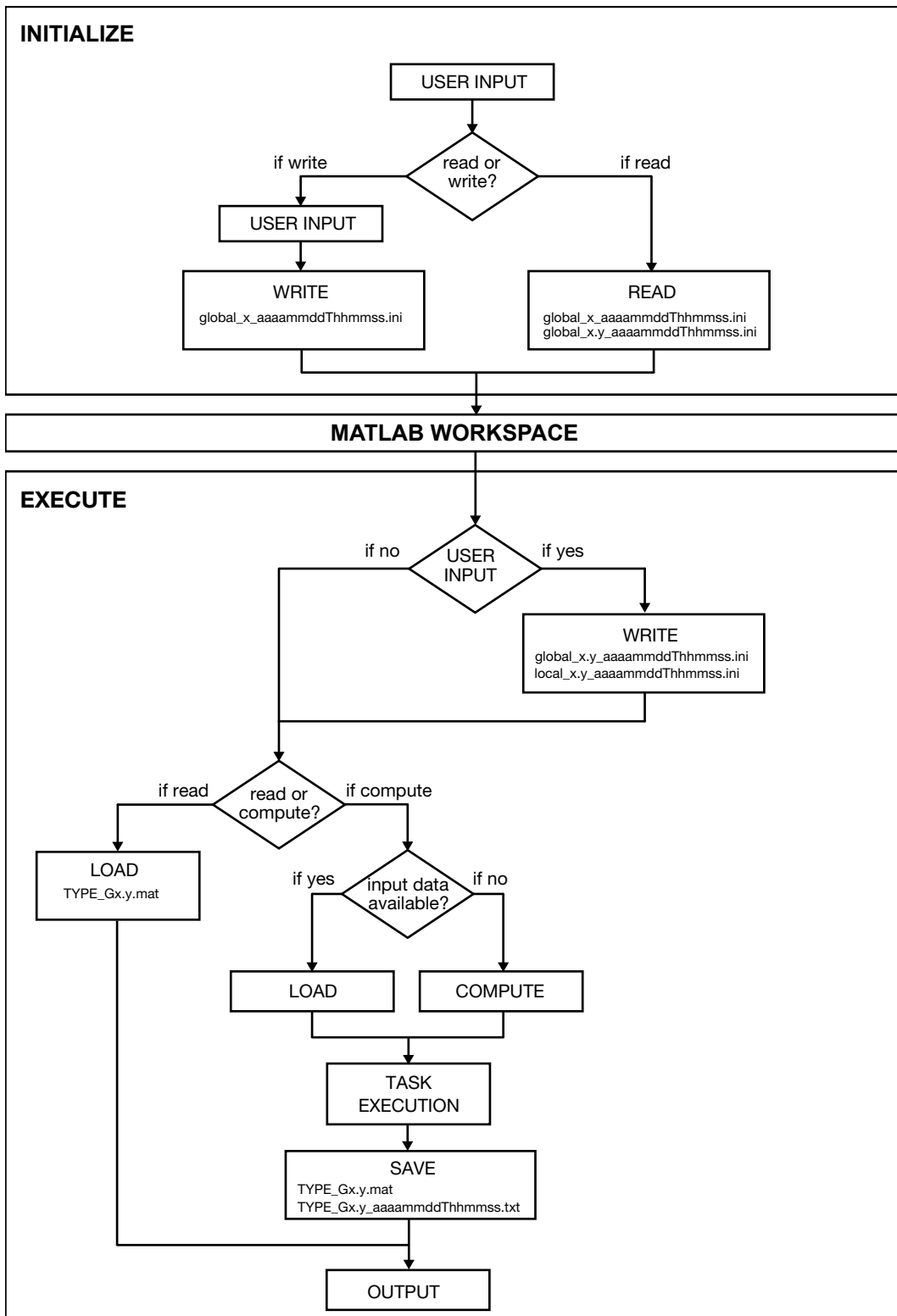


Fig. 9.2: The flowchart represents the execution of the main code, made of “initialize” and “execute” blocks. Service and execution parameters are defined during the initialization procedure and loaded into the MatLab workspace. After initialization, the block “execute” uses the parameters available in the Matlab workspace to execute each selected method.

the main code.

Once completed the initialization procedure, all the selected methods are subsequently executed. Normally, each function works with the parameters specified in the configuration file. In addition, the user can provide locally some modifications of the parameters, directly in input to the function itself. At the end of the execution of the main code, if local modifications of function parameters have occurred, they are automatically collected and stored in a local configuration file. At the start of execution, each function reads its relative execution parameters from the MatLab workspace and verifies whether local changes have occurred, eventually updating parameters. Running parameters of each function contain a list of flag defining the execution modality of the function itself, filenames of input data required for execution, and specific running parameters. Hence, a function can be executed through two different modalities, “read” or “compute”. In the first case, the function just loads the file according to the filename provided and returns the content of the file as the output. Instead, if the “compute” modality is selected, the function loads the input data from the file specified as function parameter or, if not found, it calls the relative functions to generate it. Then, when input data are available within the function MatLab environment, the function task is executed by calling specific functions. The output of the function is both saved as a **.mat* file in its corresponding folder path, semantically coherent (see the chart above) and returned as the output in the principal MatLab workspace, to be used by the functions subsequently executed. For favouring reproducibility of measurements, the output of the function includes the configuration filename relative to its generation. Moreover, each function saves an execution report file as a **.txt* file in the same folder of the output file, which includes date and time of execution, reference configuration file, execution parameters, execution messages or warnings collected during the execution and time statistics.

Finally, a standard for naming configuration and output files has been conceived. Configuration files are named as follows:

```
global_x_aaaammddThhmmss.ini
```

where **x** is an integer number starting from 0. If local configuration files are generated, two configuration files are then saved, the local ones containing the specific portion of the global configuration file that has been modified and the new global configuration file, after the local update. Accordingly, the two files are named as follows:

```
global_x.y_aaaammddThhmmss.ini
```

```
local_x.y_aaaammddThhmmss.ini
```

where **x.y** are two integer numbers starting from 0 and 1, respectively. Finally, the output and report files saved by functions are named as follows, respectively:

TYPE_Gx.y.mat

TYPE_Gx.y_aaaammddThhmmss.txt

where TYPE refers to the specific function acronym which generated the files, whilst x.y identifies the configuration file used for the execution.

List of Abbreviations

Abbreviation	Definition
AATH	Adiabatic Approximation Tissue Homogeneity
ACC	Accuracy
ACR	American College of Radiology
ADC	Apparent Diffusion Coefficient
AS	Active Surveillance
AUC	Area Under the Curve
BF	Blood Flow
BN	Bayesian Network
BSA	Body Surface Area
BV	Blood Volume
BW	Body Weight
CA	Contrast agent
CEA	Carcinoembryonic Antigen
CI	Confidence Interval
CK	Contralateral normal Kidney
C-RADS	Colon RADS
CRC	Colorectal Cancer
csPCa	clinically significant Prostate Cancer
CT	Computed Tomography
CTDI	CT Dose Index
CV	Coefficient of Variation
CV	Cross Validation
2CX	Two Compartments Exchange
D	Diffusion Coefficient
DAG	Directed Acyclic Graph
DCE	Dynamic Contrast-enhanced
DICOM	Digital Imaging and Communication in Medicine
DLP	Dose Length Product
DP	Distributed Parameter
DT	Decision Tree

DV	Deconvolution
DWI	Diffusion Weighted Imaging
EES	Extracellular Space
EME-IQ	Enhancement Measurement Error-Image Quality
FBP	Filtered Backprojection
FDA	Food and Drug Administration
FDG	fluoro-D-glucose
FDR	False Discovery Rate
FID	Free Induction Decay
FLT	3'-fluoro-3'deoxythymidine
FNR	False Negative Rate
FO	First Order
FOR	False Omission Rate
FOV	Field of View
FPR	False Positive Rate
FWHM	Full Width Half Maximum
GL	Grey Level
GLCM	Grey Level Co-occurrence matrix
GLRLM	Grey Level Run Length matrix
GS	Gleason Score
GV	Gamma Variate
HCC	Hepatocarcinoma
HPI	Hepatic Perfusion Index
HU	Hounsfield Unit
IB	Imaging Biomarker
IBSI	Imaging Biomarker Standardization Initiative
IDT	Indicator Dilution Theory
IQR	Interquartile Range
LARC	Locally Advanced Rectal Cancer
LBM	Lean Body Mass
LI-RADS	Liver RADS
LN	Lagged Normal
mpMRI	multi-parametric Magnetic Resonance Imaging
MR	Magnetic Resonance
MS	Maximum Slope
MTC	Medullary Thyroid Carcinoma
MTT	Mean Transit Time
MVD	Microvascular Density
nCRT	Neoadjuvant Chemo-Radiotherapy
ncsPCa	non-clinically significant Prostate Cancer

NET	Neuroendocrine Tumour
NMR	Nuclear Magnetic Resonance
NN	Neural Network
NPV	Negative Predictive Value
NR	Non-Responder
PACS	Picture Archiving and Communication System
PCa	Prostate Cancer
PCR	Pathological Complete Response
PET	Positron Emission Tomography
PI	Perfusion Index
PI-RADS	Prostate Reporting and Data System
PLS	Blood Plasma Space
PPV	Positive Predictive Value
PS	Permeability Surface
PSA	Prostate Specific Antigen
PSF	Point Spread Function
PSMA	Prostate Specific Membrane Antigen
PZ	Peripheral Zone
QIBA	Quantitative Imaging Biomarkers Alliance
R	Responder
RADS	Reporting and Data System
RCC	Renal Cell Carcinoma
RD	Relative Dispersion
RF	Radio-frequency
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RQS	Radiomics Quality Score
SE	Sensitivity
SIRT	Selective Internal Radiotherapy
SNR	Signal to Noise Ratio
SP	Specificity
SUV	Standard Uptake Value
SVM	Support Vector Machine
TAC	Time Attenuation Curve
TCC	Time Concentration Curve
TH	Tissue Homogeneity
TIC	Time Intensity Curve
TME	Total Mesorectal Excision
TN	True Negative
TP	True Positive

TZ	Transitional Zone
VEGF	Vascular Endothelial Growth Factor
WL	Level of HU window
WW	Width of HU window

List of Figures

1.1	Thesis content flowchart	4
2.1	HU windows for lung and abdomen CT images	14
2.2	HU windows abdomen CT images with incorrect WW and WL values	14
2.3	Axial T_2 -weighted and DWI with $b=0$ s/mm ² images	21
2.4	DWI images acquired with four different b-values	22
2.5	ADC parametric map	24
2.6	DCE-CT imaging protocol	25
3.1	Angiogenesis	31
3.2	Normal and tumour vasculature	33
4.1	DCE-CT signal extraction	46
4.2	DCE-CT hepatic vascular inputs	47
4.3	CTp denoising	48
4.4	DCE-CT TCCs	49
4.5	Single and dual input models	49
4.6	CA kinetics	52
4.7	Compartmental models	53
4.8	Mono and multicompartmental models	54
4.9	Well-mixed and plug-flow systems	54
4.10	LN fitting model	57
4.11	LN fitting with varying RD and s	58
4.12	LN fitting of vascular TCC	58
4.13	Sigmoid model	59
4.14	Sigmoidal fitting of TCC	60
4.15	GV model with varying t_0	61
4.16	GV model with varying α and β	61
4.17	GV model as the ratio of α and β varies	62
4.18	GV fitting of tissue TCC	63
4.19	Fick's principle	64
4.20	MS on a single input model	65
4.21	MS on a dual input model	66

4.22	Matterne model	67
4.23	2CX model	68
4.24	TH model	69
4.25	AATH model	70
4.26	DP model	70
4.27	$h(t)$, $H(t)$, and $R(t)$	74
5.1	Machine learning workflow	83
5.2	Hypothesis testing region	90
5.3	ROC curve	94
5.4	TPR(T) and TNR(T)	95
5.5	Radiomic workflow	96
5.6	Lasso and ridge constraint regions	110
5.7	Constraint regions for different L_n norms	111
5.8	Decision hyperplane	113
5.9	Maximum margin SVM	114
5.10	SVM for non separable classes	115
5.11	k-fold CV	118
6.1	Dual input monocompartmental hepatic model	126
6.2	Extracting the first pass signal	128
6.3	LN models for fitting aortic and portal vein as RD and s vary	129
6.4	Fitting of the first pass portion of TCCs	130
6.5	Two-stage procedure for solving DV	131
6.6	Convergence of the iterative process	132
6.7	gMS graphical representation	135
7.1	Histogram of mean residuals (μ_ϵ) for patient C1N1	139
7.2	Histograms of mean residuals (μ_ϵ) during the three signal phases	140
7.3	Map of percentage errors of patient C1N1.	140
7.4	Examples of GV fittings	141
7.5	GV fitting of differently noisy signals	142
7.6	Residual plot of μ_ϵ against BF_{MS} and BF_{DV}	143
7.7	Voxel-based comparison of BF_{MS} and BF_{DV} through colorimetric maps	148
7.8	Voxel-based comparison of BF_{MS} and BF_{DV} through scatter plots	150
7.9	Voxel-based comparison of BF_{MS} and BF_{DV} through NCC	151
7.10	Scatter plots of BF_{MS} against BF_{DV} in Centres 1 and 16	152
7.11	Equivalence times for patient C1N26	153
7.12	Equivalence times for patients C1N38 and C16N12	154
7.13	Spatial distribution of equivalence times for C1N26	154
7.14	Histograms of IQRs and normalized median equivalence times	155

7.15	Distance between MS_t and $PEAK_a$	156
7.16	Linear regression between gMS and DV	157
7.17	Colorimetic maps of aBF values for patients C1N38 and C16N12	158
7.18	CTp examinations of colorectal cancer	160
7.19	Histograms of BF_{MS} and BF_{DV} for patient ID12006	160
7.20	Colorimetic maps of BF_{MS} and BF_{DV} for patient ID12006	161
7.21	Scatter plot of BF_{MS} and BF_{DV} for patient ID12006	161
7.22	Histograms of BF_{MS} and BF_{DV} for patient ID12001	162
7.23	Colorimetic maps of BF_{MS} and BF_{DV} for patient ID12001	162
7.24	Scatter plot of BF_{MS} and BF_{DV} for patient ID12001	163
7.25	Variation of perfusion parameters in Sorafenib group	166
7.26	Variation of perfusion parameters in SIRT group	167
8.1	Workflow of CT image processing for feature robustness analysis	172
8.2	Comparison of linear, Akima, and Lanczos interpolation methods	177
8.3	Overall feature robustness against image perturbation for RCC and CK	179
8.4	Overall feature robustness per feature class for RCC and CK	180
8.5	Boxplots of $GLCM2D-sosvh$ computed from $\delta = 1$ to $\delta = 9$	182
8.6	Four different images have the same histogram	184
8.7	Differences in image details	185
8.8	Local maps of σ	185
8.9	Local computation of parametric feature maps	186
8.10	New set of imaging features	190
8.11	Flowchart of patient enrolment in radiomic LARC study	193
8.12	Discrimination results between TRG0-1 and TRG2-3	196
8.13	Local skewness for different TRG	197
8.14	Flowchart of patient enrolment processes in the PCa radiomic study	201
8.15	Boxplot of four progressive PCa risk levels	207
8.16	Predictive model building	208
8.17	Splitting of training and test set from the entire dataset	209
8.18	Feature selection process	212
8.19	Performance of the predictive model on the training set	213
8.20	Performance of the predictive model on the holdout test set	214
8.21	Boxplot of ncsPCa and csPCa in training and test sets	215
9.1	CTp analysis software architecture	224
9.2	Flowchart of execution of the main code	226

List of Tables

5.1	Errors occurring in hypothesis testing	91
5.2	Confusion matrix	93
5.3	Metrics derived from a two-by-two confusion matrix	93
5.4	Overview of the main characteristics of the most adopted radiomic features, where D stands for diagnosis and P for both prognosis and prediction	99
5.5	Four unique directions of 2D GLCM	102
5.6	Thirteen unique directions of 3D GLCM	103
6.1	PIXEL Centres and enrolled patients	124
6.2	Patients of PIXEL selected for perfusion analysis	125
6.3	Open issues of MS and DV in liver perfusion	126
7.1	σ values of μ_ϵ of ten BF values for MS and DV	143
7.2	LN fitting parameters of aorta and portal vein	144
7.3	GV fitting parameters of tissue signals	145
7.4	Correlation of BF_{MS} and BF_{DV} in Centres 1 and 16	147
7.5	DV- and MS-based BF values in the recent literature	149
7.6	Variation of perfusion parameters between Sorafenib and SIRT group	165
8.1	CT image acquisition parameters	171
8.2	List of FO (n=13) and GLCM2D-3D (n=19) features.	175
8.3	Descriptions of the 29 perturbations assessed	176
8.4	Feature robustness against image perturbations referred to Figure 8.4.	181
8.5	Seven parametric maps of local first order features	187
8.6	Some global statistical descriptors	191
8.7	Correspondence between GS and ISUP score systems	201
8.8	Study population of radiomic PCa study	203
8.9	Study population of PCa staging study	204
8.10	Study population of the classification study for ncsPCa and csPCa	204
8.11	Dataset of ncsPCa and csPCa after the first and second enrolment process	204
8.12	Dataset of ncsPCa and csPCa enrolled during the third process	205
8.13	Derivation of the training set from the entire one	210
8.14	Balanced training and test sets	210

8.15 Comparison of our study with the state of the art 211

Bibliography

- [1] R. Gunderman and T. Rickett. “Engineering and radiology: implications for education”. In: *Acad. Radiol.* 19.8 (2012), pp. 1041–1042 (cit. on p. 1).
- [2] D. Hanahan and R. Weinberg. “Hallmark of cancer: the next generation”. In: *Cell* 144.5 (2011), pp. 646–674 (cit. on pp. 1, 29, 30, 32, 35).
- [3] PIXEL. *Programme Hospitalier de Recherche Clinique (PHRC) 2007 n° AOM 07228-France and Assistance-Publique Hôpitaux de Paris (APHP)* (cit. on pp. 3, 78).
- [4] V. Vilgrain, H. Pereira, E. Assenat, et al. “Efficacy and safety of selective internal radiotherapy with yttrium-90 resin microspheres compared with sorafenib in locally advanced and inoperable hepatocellular carcinoma (SARAH): an open-label randomised controlled phase 3 trial”. In: *Lancet Oncol.* 18 (2017), pp. 1624–1626 (cit. on pp. 3, 78, 164).
- [5] S. Histed et al. “Review of functional/anatomic imaging in oncology”. In: *Nucl. Med. Commun.* 33.4 (2012), pp. 349–361 (cit. on p. 9).
- [6] Y. Cao. “The promise of dynamic contrast-enhanced imaging in radiation therapy”. In: *Semin. Radiat. Oncol.* 21 (2011), pp. 147–156 (cit. on p. 9).
- [7] E. Shefer et al. “State of the art of CT detectors and sources: a literature review”. In: *Curr. Radiol. Rep.* 1 (2013), pp. 76–91 (cit. on p. 10).
- [8] M. Prokop et al. *Spiral and multislice. Computed Tomography of the body*. Thieme, 2003. Chap. 1 (cit. on pp. 11, 15).
- [9] W. Kalender. *Computed Tomography. Fundamentals, systems, technology, image quality, applications*. Wiley, 2011. Chap. 1 (cit. on pp. 12, 15, 16).
- [10] S. Carmignato, W. Dewulf, and R. Leach. *Industrial X-Ray computed tomography*. Springer, 2018. Chap. 2 (cit. on p. 12).
- [11] D. Platte et al. “Radiation dosimetry for wide-beam CT scanners: recommendations of a working party of the Institute of Physics and Engineering in Medicine”. In: *Br. J. Radiol.* 86.1027 (2013), p. 20130089 (cit. on p. 15).
- [12] M. Salvatori et al. “Radiation dose in nuclear medicine: the hybrid imaging”. In: *Radiol. Med.* 124.8 (2019), pp. 768–776 (cit. on p. 15).
- [13] T. Yankeelov, D. Pickens, and R. Price. *Quantitative MRI in cancer*. Taylor & Francis Group, 2012. Chap. 1 (cit. on pp. 18, 20, 22, 23, 26).

- [14] V. Grover et al. “Magnetic resonance imaging: principles and techniques: lessons for clinicians”. In: *J Clin. Exp. Hepatol.* 5.3 (2015), pp. 246–255 (cit. on p. 18).
- [15] E. V. Reeth et al. “Super-resolution in magnetic resonance imaging: a review”. In: *Concept Magn. Reson. A* 40A.6 (2012), pp. 306–325 (cit. on p. 18).
- [16] T. Yankeelov, D. Pickens, and R. Price. *Quantitative MRI in cancer*. Taylor & Francis Group, 2012. Chap. 3 (cit. on p. 19).
- [17] C. Westbrook. *MRI at a glance*. Blackwell Science, 2002. Chap. 1 (cit. on p. 20).
- [18] A. Phadani et al. “Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations”. In: *Neoplasia* 11.2 (2009), pp. 102–125 (cit. on p. 21).
- [19] N. Bhatt et al. “Role of diffusion-weighted imaging in head and neck lesions: pictorial review”. In: *Neuroradiol. J* 30.4 (2017), pp. 356–369 (cit. on pp. 21, 23).
- [20] M. Morone et al. “Whole-body MRI: current applications in oncology”. In: *AJR Am. J. Roentgenol.* 209 (2017), W336–W349 (cit. on p. 23).
- [21] Y. Sun et al. “Automatic stratification of prostate tumour aggressiveness using multiparametric MRI: a horizontal comparison of texture features”. In: *Acta Oncol.* 58.8 (2019), pp. 1118–1126 (cit. on p. 23).
- [22] G. Jayson et al. “Antiangiogenic therapy in oncology: current status and future directions”. In: *Lancet* 388 (2016), pp. 518–529 (cit. on p. 24).
- [23] K. Miles et al. “Current status and guidelines for the assessment of tumour vascular support with dynamic contrast-enhanced computed tomography”. In: *Eur. Radiol.* 22.7 (2012), pp. 1430–1441 (cit. on pp. 24, 45).
- [24] C. Cuenod and D. Balvay. “Perfusion and vascular permeability: basic concepts and measurements in DCE-CT ad DCE-MRI”. In: *Diagn. Interv. Imaging* 94 (2013), pp. 1187–1204 (cit. on pp. 24, 35, 71).
- [25] Y. Gordon et al. “Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion”. In: *Cardiovasc. Diagn. Ther.* 4.2 (2014), pp. 147–164 (cit. on p. 26).
- [26] J. Winfield et al. “DCE-MRI, DW-MRI, and MRS in cancer: challenges and advantages of implementing qualitative and quantitative multi-parametric imaging in the clinic”. In: *Magn. Reson. Imaging* 25.6 (2016), pp. 245–254 (cit. on p. 26).
- [27] A. Othman et al. “Radiation dose reduction in perfusion CT imaging of the brain: a review of the literature”. In: *J. Neuroradiology* 43 (2016), pp. 1–5 (cit. on p. 26).

- [28] D. Muenzel et al. “Dynamic CT perfusion imaging of the myocardium: a technical note on improvement of image quality”. In: *PLoS one* 8.10 (2013), e75263 (cit. on pp. 26, 47).
- [29] C. Yang et al. “Comparison of quantitative parameters in cervix cancer measured by dynamic contrast-enhanced MRI and CT”. In: *Magn. Reson. Med.* 63.6 (2010), pp. 1601–1609 (cit. on p. 26).
- [30] G. Brix et al. “Dynamic contrast-enhanced CT studies. Balancing patient exposure and image noise”. In: *Invest. Radiol.* 46 (2011), pp. 64–70 (cit. on p. 26).
- [31] T. Yankeelov and J. Gore. “Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples”. In: *Curr. Med. Imaging Rev.* 3.2 (2009), pp. 91–107 (cit. on p. 27).
- [32] J. O’Connor et al. “Imaging intratumour heterogeneity: role in therapy response, resistance, and clinical outcome”. In: *Clin. Cancer Res.* 21.2 (2015), pp. 249–257 (cit. on pp. 29, 36, 37).
- [33] N. Pavlova and C. Thompson. “The emerging hallmarks of cancer metabolism”. In: *Cell Metab.* 23.1 (2016), pp. 27–47 (cit. on p. 30).
- [34] L. Boroughs and R. DeBerardinis. “Metabolic pathways promoting cancer cell survival and growth”. In: *Nat. Cell Biol.* 17.4 (2015), pp. 351–359 (cit. on p. 30).
- [35] M. Heiden and R. DeBerardinis. “Understanding the intersections between metabolism and cancer biology”. In: *Cell* 168.4 (2017), pp. 657–669 (cit. on p. 30).
- [36] P. Porporato et al. “Mitochondrial metabolism and cancer”. In: *Cell Res.* 28.3 (2017), pp. 265–280 (cit. on p. 30).
- [37] A. Elmi et al. “Cell-proliferation imaging for monitoring response to CDK4/6 inhibition combined with endocrine-therapy in breast cancer: comparison of [¹⁸F]FLT and [¹⁸F]ISO-1 PET/CT”. In: *Clin. Cancer Res.* 25.10 (2019), pp. 3063–3073 (cit. on p. 30).
- [38] J. Horvat et al. “Diffusion-Weighted Imaging (DWI) with Apparent Diffusion Coefficient (ADC) mapping as a quantitative imaging biomarker for prediction of immunohistochemical receptor status, proliferation rate, and molecular subtypes of breast cancer”. In: *J. Magn. Reson. Imaging* 50.3 (2019), pp. 836–846 (cit. on p. 30).
- [39] A. Chung, J. Lee, and N. Ferrara. “Targeting the tumour vasculature: insights from physiological angiogenesis”. In: *Nat. Rev. Cancer* 10.7 (2010), pp. 505–514 (cit. on p. 31).
- [40] B. Lopes-Bastos, W. Jiang, and J. Cai. “Tumour-endothelial cell communications: important and indispensable mediators of tumour angiogenesis”. In: *Anticancer Res.* 36.3 (2016), pp. 1119–1126 (cit. on p. 31).

- [41] L. Yadav et al. “Tumour angiogenesis and angiogenic inhibitors: a review”. In: *J. Clin. Diagn. Res.* 9.6 (2015), XE01–XE05 (cit. on p. 32).
- [42] L. Welsh and M. Welsh. “VEFGA and tumour angiogenesis”. In: *J. Intern. Med.* 273.2 (2019), pp. 114–127 (cit. on p. 32).
- [43] I. Dregely et al. “Imaging biomarkers in oncology: basics and application to MRI”. In: *J. Magn. Res. Imaging* 48.1 (2018), pp. 13–26 (cit. on pp. 32, 37, 38).
- [44] R. Jain. “Antiangiogenesis strategies revisited: from starving tumors to alleviating hypoxia”. In: *Cancer Cell* 26.5 (2014), pp. 605–622 (cit. on p. 33).
- [45] H. Yang et al. “Perfusion computed tomography evaluation of angiogenesis in liver cancer”. In: *Eur. Radiol.* 20 (2010), pp. 1424–1430 (cit. on p. 33).
- [46] V. Goh and A. Padhani. “Imaging tumor angiogenesis: functional assessment using MDCT or MRI?” In: *Abdom. Imaging* 31 (2006), pp. 194–199 (cit. on p. 33).
- [47] R. Jain. “Perfusion CT imaging of brain tumors: an overview”. In: *AJNR Am. J. Neuroradiol.* 32 (2011), pp. 1570–1577 (cit. on p. 34).
- [48] D. Ippolito et al. “Diagnostic value of dynamic contrast-enhanced CT with perfusion imaging in the quantitative assessment of tumor response to sorafenib in patients with advanced hepatocellular carcinoma: A feasibility study”. In: *Eur. J. Radiol.* 90 (2017), pp. 34–41 (cit. on p. 34).
- [49] A. Uusaro, E. Ruokonen, and J. Takala. “Estimation of splanchnic blood flow by the Fick principle in man and problems in the use of indocyanine green”. In: *Cardiovasc. Res.* 30.1 (1995), pp. 106–112 (cit. on p. 34).
- [50] G. Brix et al. “Tracer kinetic modelling of tumour angiogenesis based on dynamic contrast-enhanced CT and MRI measurements”. In: *Eur. J. Nucl. Med. Mol. Imaging* 37.1 (2010), pp. 30–51 (cit. on pp. 34, 47).
- [51] W. Thaiss et al. “Iodine concentration as a perfusion surrogate marker in oncology: Further elucidation of the underlying mechanisms using Volume Perfusion CT with 80 kVp”. In: *Eur. Radiol.* 26.9 (2016), pp. 2929–2936 (cit. on pp. 35, 48).
- [52] S. Zhou et al. “Correlations between computed tomography perfusion enhancement parameters and lymph node metastasis in non-small cell lung cancer”. In: *Int. J. Clin. Exp. Med.* 9.7 (2016), pp. 12896–12903 (cit. on p. 35).
- [53] H. Sun et al. “Assessment of tumor grade and angiogenesis in colorectal cancer”. In: *Acad. Radiol.* 21.6 (2014), P750–757 (cit. on p. 35).
- [54] S. Kim, A. Kamaya, and J. Willmann. “CT perfusion of the liver: principles and application in oncology”. In: *Radiology* 272.2 (2014), pp. 322–344 (cit. on pp. 35, 51, 79).

- [55] K. Miles. “Perfusion CT for the assessment of tumour vascularity: which protocol?” In: *Br. J. Radiol.* 76 (2003), S36–42 (cit. on pp. 35, 50).
- [56] *ACR (American College of Radiology) PI-RADS Prostate imaging - reporting and data system, version 2.1.* 2019 (cit. on pp. 35, 199).
- [57] K. Cyll et al. “Tumour heterogeneity poses a significant challenge to cancer biomarker research”. In: *Br. J. Cancer* 117 (2017), pp. 367–375 (cit. on p. 36).
- [58] I. Dagogo-Jack and A. Shaw. “Tumour heterogeneity and resistance to cancer therapies”. In: *Nat. Rev. Clin. Oncol.* 15.81-94 (2018) (cit. on p. 36).
- [59] P. Bedard et al. “Tumour heterogeneity in the clinic”. In: *Nature* 501.7467 (2013), pp. 355–364 (cit. on pp. 36, 37).
- [60] L. Gay, A. Baker, and T. Graham. “Tumour cell heterogeneity.” In: *Version 1. F1000Res.* 5.F1000 Faculty Rev-238 (2016). DOI: 10.12688/f1000research.7210.1 (cit. on pp. 36, 37).
- [61] H. Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. In: *Nat. Commun.* 5.4006 (2014). DOI: 10.1038/ncomms5006 (cit. on pp. 37, 105, 181).
- [62] NCBI. *Food and Drug Administration & National Institutes of Health. BEST (Biomarkers, EndpointS, and other tools).* From: <http://www.ncbi.nlm.nih.gov/books/NBK326791>. Accessed: 2020-09-10 (cit. on p. 37).
- [63] C. Oldenhuis et al. “Prognostic versus predictive value of biomarkers in oncology”. In: *Eur. J. Cancer* 44 (2008), pp. 946–953 (cit. on p. 37).
- [64] L. Fass. “Imaging and cancer: a review”. In: *Mol. Oncol.* 2.2 (2008), pp. 115–152 (cit. on p. 37).
- [65] S. Yip and H. Aerts. “Applications and limitations of radiomics”. In: *Phys. Med. Biol.* 61.13 (2016), R150–66 (cit. on p. 37).
- [66] P. Lambin et al. “Radiomics: the bridge between medical imaging and personalized medicine”. In: *Nat. Rev. Clin. Oncol.* 14.12 (2017), pp. 749–762 (cit. on p. 38).
- [67] K. Doi. “Computer-aided diagnosis in medical imaging: historical review, current status and future potential”. In: *Comput. Med. Imaging Graph.* 31.4-5 (2007), pp. 198–211 (cit. on p. 38).
- [68] M. Avanzo, J. Stancanella, and I. E. Naqa. “Beyond images: the promise of radiomics”. In: *Phys. Med.* 38 (2017), pp. 122–139 (cit. on pp. 38, 95, 97, 100).
- [69] QIBA. *Quantitative Imaging Biomarkers Alliance (QIBA).* From: <https://www.rsna.org/QIBA>. Accessed: 2020-09-10 (cit. on p. 38).
- [70] S. Rizzo et al. “Radiomics: the facts and the challenge of image analysis”. In: *Eur. Radiol. Exp.* 2.1 (2018), p. 36. DOI: 10.1186/s41747-018-0068-z (cit. on pp. 38, 95, 99, 100).
- [71] J. Aronsons. “Biomarkers and surrogate endpoints”. In: *Br. J. Clin. Pharmacol.* 59.5 (2005), pp. 491–494 (cit. on p. 39).

- [72] FDA-NIH Biomarker Working Group. *BEST (Biomarkers, Endpoints, and other Tools) Resource [Internet]. Silver Spring (MD): Food and Drug Administration (US); 2016-Diagnostic Biomarker*. From: <https://www.ncbi.nlm.nih.gov/books/NBK402285/> Co-published by National Institutes of Health (US), Bethesda (MD) (cit. on p. 39).
- [73] S. Ruberg and L. Shen. “Personalized medicine: four perspectives of tailored medicine”. In: *Stat. Biopharm. Res.* 7.3 (2015), pp. 214–229 (cit. on p. 39).
- [74] K. Sechidis et al. “Distinguishing prognostic and predictive biomarkers: an information theoretic approach”. In: *Bioinformatics* 34.19 (2018), pp. 3365–3376 (cit. on p. 39).
- [75] S. Sethi et al. “Clinical advances in molecular biomarkers for cancer diagnosis and therapy”. In: *Int. J. Mol. Sci.* 14.7 (2013), pp. 14771–14784 (cit. on p. 39).
- [76] Y. Wang. “Development of cancer diagnostics - from biomarkers to clinical tests”. In: *Transl. Cancer Res.* 4.3 (2015), pp. 270–279 (cit. on p. 40).
- [77] P. Maruvada et al. “Biomarkers in molecular medicine: cancer detection and diagnosis”. In: *Biotechniques* 38.4S (2005), pp. 9–15 (cit. on p. 40).
- [78] J. P. O’Connor et al. “Imaging biomarker roadmap for cancer studies”. In: *Nat. Rev. Clin. Oncol.* 14.3 (2017), pp. 169–186 (cit. on p. 40).
- [79] S. Edge and C. Compton. “The american joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM”. In: *Ann. Surg. Oncol.* 17 (2010), pp. 1471–1474 (cit. on p. 40).
- [80] J. Brierley, M. Gospodarowicz, and C. Wittekind. *TNM classification of malignant tumours*. Wiley, 2017. Chap. 1 (cit. on p. 41).
- [81] J. An, K. Unsdorfer, and J. Weinreb. “BI-RADS, C-RADS, CAD-RADS, LI-RADS, Lung-RADS, NI-RADS, O-RADS, PI-RADS, TI-RADS: reporting and data systems”. In: *Radiographics* 39 (2019), pp. 1435–1436 (cit. on p. 41).
- [82] J. Barentsz et al. “Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use”. In: *Eur. Urol.* 69.1 (2016), pp. 41–49 (cit. on p. 41).
- [83] S. Gupta. “Role of image-guided percutaneous needle biopsy in cancer staging”. In: *Semin. Roentgenol.* 41.2 (2006), pp. 78–90 (cit. on p. 41).
- [84] J. Epstein et al. “The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason Grading of Prostatic Carcinoma: definition of grading patterns and proposal for a new grading system”. In: *Am. J. Surg. Pathol.* 40.2 (2016), pp. 244–252 (cit. on p. 41).
- [85] F. Bladou et al. “Transrectal ultrasound-guided biopsy for prostate cancer detection: systematic and/or magnetic resonance imaging-targeted”. In: *Can. Urol. Assoc. J.* 11.9 (2017), E330–E337 (cit. on p. 41).

- [86] N. Mottet et al. “EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent”. In: *Eur. Urol.* 71.4 (2017), pp. 618–629 (cit. on p. 41).
- [87] M. Nishino. “Tumor response assessment for precision cancer therapy: response evaluation criteria in solid tumors and beyond”. In: *Am. Soc. Clin. Oncol. Educ. Book* 38 (2018), pp. 1019–1029 (cit. on p. 42).
- [88] S. Grimaldi, M. Terroir, and C. Caramella. “Advances in oncological treatment: limitations of RECIST 1.1 criteria”. In: *Q. J. Nucl. Med. Mol. Im.* 62.2 (2018), pp. 129–139 (cit. on p. 42).
- [89] R. Abramson et al. “Pitfalls in RECIST data extraction for clinical trials: beyond the basics”. In: *Acad. Radiol.* 22.6 (2015), pp. 779–786 (cit. on p. 43).
- [90] G. Petralia et al. “CT perfusion in oncology: how to do it”. In: *Cancer Imaging* 10 (2010), pp. 8–19 (cit. on p. 45).
- [91] D. Ippolito et al. “Dynamic computed tomography perfusion imaging: complementary diagnostic tool in hepocellular carcinoma assessment from diagnosis to treatment follow-up”. In: *Acad. Radiol.* (2019), pp. 1–11. DOI: 10.1016/j.acra.2019.02.010 (cit. on pp. 46, 47, 125).
- [92] K. Miles and M. Griffiths. “Perfusion CT: a worthwhile enhancement?” In: *Br. J. Radiol.* 76.904 (2003), pp. 220–231 (cit. on p. 46).
- [93] A. Bevilacqua et al. “Automatic detection of misleading blood flow values in CT perfusion studies of lung cancer”. In: *Biomed. Signal Process. Control* 26 (2016), pp. 109–116 (cit. on pp. 47, 59, 79, 127).
- [94] T. S. Koh et al. “Dynamic contrast-enhanced CT imaging of hepatocellular carcinoma in cirrhosis: feasibility of a prolonged dual-phase imaging protocol with tracer kinetics modeling”. In: *Eur. Radiol.* 19 (2009), pp. 1184–1196 (cit. on p. 47).
- [95] A. Bevilacqua et al. “A novel approach for semi-quantitative assessment of reliability of blood flow values in DCE-CT perfusion”. In: *Biomed. Signal Process. Control* 31 (2017), pp. 257–264 (cit. on pp. 47, 55).
- [96] M. Kapanen, J. Halavaara, and A. Häkkinen. “Comparison of liver perfusion parameters studied with conventional extravascular and experimental intravascular CT contrast agents”. In: *Acad. Radiol.* 14.8 (2009), pp. 951–958 (cit. on pp. 48, 51).
- [97] T. Yeung et al. “Improving Quantitative CT Perfusion Parameter Measurements Using Principal Component Analysis”. In: *Acad. Radiol.* 21.5 (2014), pp. 624–632 (cit. on p. 48).
- [98] A. Bevilacqua and S. Malavasi. “A novel algorithm to detect the baseline value of a time signal in Dynamic Contrast Enhanced-Computed Tomography”. In: *Proceedings of the IEEE 16th International Symposium on Cir-*

- cuits and Systems (ISCAS)*. DOI : 10.1109/ISCAS.2018.8351281. IEEE. 2018, pp. 1–4 (cit. on p. 48).
- [99] C. Thng et al. “Perfusion magnetic resonance imaging of the liver”. In: *World J. Gastroenterol.* 16.13 (2010), pp. 1598–1609 (cit. on p. 48).
- [100] Y. Ohno et al. “Dynamic contrast-enhanced perfusion area detector CT for non-small cell lung cancer patients: Influence of mathematical models on early prediction capabilities for treatment response and recurrence after chemoradiotherapy”. In: *Eur. J. Radiol.* 85 (2016), pp. 176–186 (cit. on p. 49).
- [101] X. Yuan et al. “Lung cancer perfusion: can we measure pulmonary and bronchial circulation simultaneously?” In: *Eur. Radiol.* 22 (2012), pp. 1665–1671 (cit. on p. 49).
- [102] E. Talakić et al. “CT perfusion imaging of the liver and the spleen in patients with cirrhosis: Is there a correlation between perfusion and portal venous hypertension?” In: *Eur. Radiol.* 27.10 (2017), pp. 4173–4180 (cit. on p. 50).
- [103] X. Yuan et al. “Differentiation of malignant and benign pulmonary nodules with first-pass dual-input perfusion CT”. In: *Eur. Radiol.* 23 (2013), pp. 2469–2474 (cit. on p. 50).
- [104] C. Ng et al. “Effect of dual vascular input functions on CT perfusion parameter values and reproducibility in Liver Tumors and Normal liver”. In: *J. Comput. Assist. Tomogr.* 36.4 (2012), pp. 388–393 (cit. on p. 50).
- [105] B. Chen and T. T.-F. Shih. “DCE-MRI in hepatocellular carcinoma-clinical and therapeutic image biomarker”. In: *World J. Gastroenterol.* 20.12 (2014), pp. 3125–3134 (cit. on p. 50).
- [106] G. Petralia et al. “Perfusion computed tomography in patients with hepatocellular carcinoma treated with thalidomide: initial experience”. In: *J. Comput. Assist. Tomogr.* 35 (2011), pp. 195–201 (cit. on p. 50).
- [107] S. Lee et al. “Dual-input tracer kinetic modelling of dynamic contrast enhanced MRI in thoracic malignancies”. In: *J. Appl. Clin. Med. Phys.* 20.11 (2019), pp. 169–188 (cit. on p. 50).
- [108] L. Faggioni, E. Neri, and C. Bartolozzi. “CT Perfusion of head and neck tumors: how we do it”. In: *AJR Am. J. Roentgenol.* 194 (2010), pp. 62–69 (cit. on pp. 51–53).
- [109] A. Razek et al. “CT Perfusion of head and neck cancer”. In: *Eur. J. Radiol.* 83 (2014), pp. 537–533 (cit. on pp. 51, 52).
- [110] P. Sahbaee, P. Segars, D. Marin, et al. “Determination of contrast media administration to achieve a targeted contrast enhancement in computed tomography”. In: *J. Med. Imaging* 3.1 (2016), p. 013501 (cit. on p. 51).
- [111] K. Bae. “Intravenous contrast medium administration and scan timing at CT: consideration and approaches”. In: *Radiology* 256.1 (2010), pp. 32–62 (cit. on p. 51).

- [112] B. Romain et al. “Parameter estimation of perfusion models in dynamic contrast-enhanced imaging: a unified framework for model comparison”. In: *Med. Image Anal.* 35 (2017), pp. 360–374 (cit. on pp. 54, 79).
- [113] M. Ingrisich and S. Sourbron. “Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer”. In: *J. Pharmacokinet. Pharmacodyn.* 40 (2013), pp. 281–300 (cit. on pp. 55, 68, 71).
- [114] A. Fieselmann et al. “Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details”. In: *Int. J. Biomed. Imaging* 2011 (2011), pp. 1–20 (cit. on pp. 55, 75).
- [115] T. Koh et al. “Assessment of tumor blood flow distribution by dynamic contrast-enhanced CT”. In: *IEEE Trans. Med. Imaging* 32.8 (2013), pp. 1504–1514 (cit. on p. 55).
- [116] A. Akhbardeh et al. “A multi-model framework to estimate perfusion parameters using contrast-enhanced ultrasound imaging”. In: *Med. Phys.* 46.2 (2018), pp. 590–600 (cit. on pp. 55, 57).
- [117] B. Bassingthwaighe, F. Ackerman, and E. Wood. “Application of the Lagged Normal Density Curve as a Model for Arterial Dilution Curve”. In: *Circ. Res.* 18.4 (1966), pp. 398–415 (cit. on p. 56).
- [118] M. Bindschadler et al. “Comparison of blood flow models and acquisitions for quantitative myocardial perfusion estimation from dynamic CT”. In: *Phys. Med. Biol.* 59 (2014), pp. 1533–1556 (cit. on p. 57).
- [119] S. Goutelle et al. “The Hill equation: a review of its capabilities in pharmacological modelling”. In: *Fundam. Clin. Pharmacol.* 22 (2008), pp. 633–648 (cit. on p. 58).
- [120] M. Mischi, J. den Boer, and H. Korsten. “On the physical and stochastic representation of an indicator dilution curve as a gamma variate”. In: *Physiol. Meas.* 29 (2008), pp. 281–294 (cit. on p. 60).
- [121] M. Wagner et al. “A model based algorithm for perfusion estimation in interventional C-arm CT systems”. In: *Med. Phys.* 40.3 (2013), p. 031916. DOI: 10.1118/1.4790467 (cit. on p. 60).
- [122] K. Scherer et al. “Dynamic quantitative iodine myocardial perfusion imaging with dual-layer CT using a porcine model”. In: *Sci. Rep.* 9 (2019), p. 16046. DOI: 10.1038/s41598-019-52458-1 (cit. on p. 60).
- [123] T. Behrenbeck et al. “Early changes in myocardial microcirculation in asymptomatic hypercholesterolemic subjects: as detected by perfusion CT”. In: *Ann. Biomed. Eng.* 42.3 (2014), pp. 515–525 (cit. on p. 60).
- [124] A. Gill et al. “A semi-automatic method for the extraction of the portal venous input function in quantitative dynamic contrast enhanced CT of the liver”. In: *Br. J. Radiol.* 90.20160875 (2017) (cit. on p. 60).

- [125] M. Madsen. “A simplified formulation of the gamma variate function”. In: *Phys. Med. Biol.* 37.7 (1992), pp. 1597–1600 (cit. on p. 60).
- [126] R. Davenport. “The derivation of the gamma-variate relationship for tracer dilution curves”. In: *J. Nucl. Med.* 24.10 (1983), pp. 945–948 (cit. on p. 60).
- [127] M. Visscher and J. Johnson. “The Fick principle: analysis of potential errors in its conventional application”. In: *J. Appl. Physiol.* 5.10 (1953), pp. 635–638 (cit. on p. 64).
- [128] K. A. Miles et al. “Application of CT in the investigation of angiogenesis in oncology”. In: *Acad. Radiol.* 7 (2000), pp. 840–850 (cit. on p. 65).
- [129] T. Kanda et al. “CT hepatic perfusion measurement: comparison of three analytic methods?” In: *Eur. J. Radiol.* 81.9 (2012), pp. 2075–2079 (cit. on p. 66).
- [130] D. J. Kovač et al. “CT perfusion and diffusion-weighted MR imaging of pancreatic adenocarcinoma: can we predict tumor grade using functional parameters?” In: *Acta Radiol.* 60.9 (2019), pp. 1065–1073 (cit. on pp. 67, 149).
- [131] G. Brix et al. “Estimation of tissue perfusion by dynamic contrast-enhanced imaging: simulation-based evaluation of the steepest slope method”. In: *Eur. Radiol.* 20.9 (2010), pp. 2166–2175 (cit. on p. 67).
- [132] S. Kaufman et al. “Characterization of hepatocellular carcinoma (HCC) lesions using a novel CT-based volume perfusion (VPCT) technique.” In: *Eur. J. Radiol.* 84 (2015), pp. 1029–1035 (cit. on pp. 67, 149, 150).
- [133] R. Materne et al. “Non-invasive quantification of liver perfusion with dynamic computed tomography and a dual-input one-compartmental model”. In: *Clin. Sci.* 99.6 (2000), pp. 517–525 (cit. on p. 67).
- [134] K. S. Lawrence and T. Lee. “An adiabatic approximation to the tissue homogeneity model for water exchange in the brain: I. Theoretical derivation”. In: *J. Cereb. Blood Flow Metab.* 18.12 (1998), pp. 1365–1377 (cit. on p. 70).
- [135] L. D. Cheong et al. “Functional Imaging: Dynamic Contrast-Enhanced CT using a Distributed-Parameter physiologic model for accessing stroke and intracranial tumor”. In: *Proceedings of IEEE Eng Med Biol Soc.* Vol. 2006. 10.1109/IEMBS.2005.1616402. IEEE. 2005, pp. 294–297 (cit. on p. 70).
- [136] H. Chen et al. “Extended graphical model for analysis of dynamic contrast-enhanced MRI”. In: *Magn. Reson. Med.* 66.3 (2011), pp. 868–878 (cit. on p. 71).
- [137] T. Ichihara et al. “Quantitative analysis of first-pass contrast-enhanced myocardial perfusion MRI using a Patlak plot method and blood saturation correction”. In: *Magn. Reson. Med.* 62.2 (2009), pp. 373–383 (cit. on p. 72).
- [138] R. Gacía-Figueiras et al. “CT Perfusion in oncologic imaging: a useful tool?” In: *AJR Am. J. Roentgenol.* 200 (2013), pp. 8–19 (cit. on p. 72).

- [139] P. Meier and K. L. Zierler. “One the theory of the indicator-dilution method for measurements of blood flow and volume”. In: *J. Appl. Physiol.* 6.12 (1954), pp. 731–744 (cit. on p. 72).
- [140] K. L. Zierler. “Equations for measuring blood flow by external monitoring of radio isotopes”. In: *Circ. Res.* 16 (1965), pp. 309–321 (cit. on p. 73).
- [141] P. Hansen. “Analysis of discrete ill-posed problems by means of the L-curve”. In: *SIAM Rev.* 34.4 (1992), pp. 561–580 (cit. on p. 76).
- [142] H.-J. Wittsack et al. “CT-perfusion imaging of the human brain: advanced deconvolution analysis using circulant singular value decomposition”. In: *Comput. Med. Imag. Grap.* 32 (2008), pp. 67–77 (cit. on p. 76).
- [143] V. S. Kadimesetty et al. “Convolutional neural network based robust denoising of low-dose Computed Tomography perfusion maps”. In: *TRPMS* (2018). DOI = 10.1109/TRPMS.2018.2860788 (cit. on p. 76).
- [144] M. Pizzolato, T. Boutlelier, and R. Deriche. “Perfusion deconvolution in DSC-MRI with dispersion-compliant bases”. In: *Med. Image Anal.* 36 (2017), pp. 197–215 (cit. on p. 77).
- [145] L. He et al. “A spatio-temporal deconvolution method to improve perfusion CT quantification”. In: *TMI* 29.5 (2010), pp. 1182–1191 (cit. on p. 77).
- [146] L. Antonelli et al. “Integrating imaging and omics data: a review”. In: *Biomed. Signal Process. Control* 52 (2019), pp. 264–280 (cit. on p. 78).
- [147] B. Bevilacqua, S. Malavasi, and V. Vilgrain. “Liver CT perfusion: which is the relevant delay that reduces radiation dose and maintains diagnostic accuracy?” In: *Eur. Radiol.* 29.12 (2019), pp. 6650–6658 (cit. on p. 78).
- [148] M. Bogowicz et al. “Perfusion CT radiomics as potential prognostic biomarker in head and neck squamous cell carcinoma”. In: *Acta Oncol.* (2019), pp. 1–6. DOI: 10.1080/0284186X.2019.1629013 (cit. on p. 78).
- [149] C. Huang et al. “Diagnostic performance of perfusion computed tomography for differentiating lung cancer from benign lesions: a meta analysis”. In: *Med. Sci. Monit.* 25 (2019), pp. 3485–3494 (cit. on p. 78).
- [150] D. H. Lee et al. “Prediciton of treatment outcome of chemotherapy using perfusion computed tomography in patients with unresectable advanced gastric cancer”. In: *Korean J. Radiol.* 20.4 (2019), pp. 589–598 (cit. on p. 78).
- [151] A. Fan et al. “Early changes in CT perfusion parameters: primary renal carcinoma versus metastases after treatment with targeted therapy”. In: *Cancers* 11.608 (2019), pp. 1–12. DOI: 10.3390/cancers11050608 (cit. on pp. 78, 149).
- [152] PROSPECT. *Information Services Division Scotland. Cancer Clinical Trials Service*. From: <http://www.isdscotland.org/Products-and-Services/Cancer-Clinical-Trials-Service/PROSPeCT.asp>. Accessed: 2013-10-03 (cit. on p. 78).

- [153] L. L. Chu, R. J. Knebel, A. D. Shay, et al. “CT perfusion imaging of lung cancer: benefit of motion correction for blood flow estimates”. In: *Eur. Radiol.* 28 (2018), pp. 5069–5075 (cit. on p. 78).
- [154] J. Li et al. “Influence of tube voltage, tube current and newer iterative reconstruction algorithms in CT perfusion imaging in rabbit liver VX2 tumors”. In: *Diagn. Interv. Radiol.* 26 (2017), pp. 264–270 (cit. on p. 78).
- [155] E. Bretas et al. “Is liver perfusion CT reproducible? A study on intra- and interobserver agreement of normal hepatic haemodynamic parameters obtained with two different software packages”. In: *Br. J. Radiol.* 90.1078 (2017), p. 20170214. DOI: 10.1259/bjr.20170214 (cit. on p. 78).
- [156] S Bouix et al. “On evaluating brain tissue classifiers without a ground truth”. In: *NeuroImage* 36 (2012), pp. 1207–1224 (cit. on p. 79).
- [157] T. S. Koh, Q. S. Ng, C. H. Thng, et al. “Primary colorectal cancer: use of kinetic modeling of dynamic contrast-enhanced CT data to predict clinical outcome”. In: *Radiology* 267.1 (2013), pp. 145–154 (cit. on p. 79).
- [158] P. Lambin et al. “Radiomics: extracting more information from medical images using advanced feature analysis”. In: *Eur J Cancer* 48.4 (2012), pp. 441–446 (cit. on pp. 82, 97).
- [159] A. Rajkomar, J. Dean, and I. Kohane. “Machine learning in medicine”. In: *N. Engl. J. Med.* 380 (2019), pp. 1347–1358 (cit. on pp. 82, 89).
- [160] G. Carleo et al. “Machine learning and the physical sciences”. In: *Rev. Mod. Phys.* 91.4 (2019), 045002(39) (cit. on p. 83).
- [161] S. Theodoridis and K. Koutroumbas. *Pattern recognition. 2nd edition*. Elsevier, 2009. Chap. 5 (cit. on pp. 83, 85, 86, 88, 92, 107, 108, 116, 194).
- [162] A. Singh, N. Thakur, and A. Sharma. “An evaluation of filter and wrapper methods for feature selection in categorical clustering”. In: *Proceedings of the 6th International Symposium on Intelligent Data Analysis (IDA 2005)*. 2005, pp. 440–451 (cit. on p. 86).
- [163] R. Duda, P. Hart, and D. Stork. *Pattern classification. 2nd edition*. Wiley, 1997. Chap. 2 (cit. on p. 86).
- [164] C. Bishop. *Pattern recognition and machine learning*. Springer, 2006. Chap. 3 (cit. on pp. 86, 110, 116).
- [165] I. Maglogiannis et al. *Emerging artificial intelligence applications in computer engineering: real word AI systems with applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press, 2007. Chap. 1 (cit. on pp. 87, 88).
- [166] A. Singh, N. Thakur, and A. Sharma. “A review of supervised machine learning algorithms”. In: *Proceedings of the 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE. 2016 (cit. on p. 87).

- [167] J. Ding, V. Tarokh, and Y. Yang. “Model selection techniques. An overview”. In: *IEEE Signal Process. Mag.* 35.6 (2018), pp. 16–34 (cit. on p. 89).
- [168] J. Zar. *Biostatistical analysis. 5th edition*. Pearson, 2009. Chap. 6 (cit. on pp. 90–92).
- [169] T. Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognit. Lett.* 27 (2006), pp. 861–874 (cit. on p. 94).
- [170] D. Mackin et al. “Effect of tube current on computed tomography radiomic features”. In: *Scientific Reports* 8 (2018), pp. 2354–2363 (cit. on pp. 97, 119, 170).
- [171] R. Da-Ano, D. Visvikis, and M. Hatt. “Harmonization strategies for multicenter radiomics investigation”. In: *Phys. Med. Biol.* (2020). DOI: 10.1088/1361-6560/aba798 (cit. on pp. 97, 98).
- [172] P. Mukherjee et al. “CT-based radiomic signatures for predicting histopathologic features in head and neck squamous cell carcinoma”. In: *Radiology: Imaging cancer* 2.3 (2020). DOI: 10.1148/rycan.2020190039 (cit. on p. 97).
- [173] G. Wu et al. “Diagnosis of invasive lung adenocarcinoma based on chest CT radiomic features of part-solid pulmonary nodules”. In: *Radiology* (2020), pp. 1–8. DOI: 10.1148/radiol.2020192431 (cit. on pp. 97, 100).
- [174] D. Leithner et al. “Radiomic signatures derived from diffusion-weighted imaging for the assessment of breast cancer receptor status and molecular subtypes”. In: *Mol. Imaging Biol.* 22.2 (2020), pp. 453–461 (cit. on p. 97).
- [175] L. Gong et al. “Noninvasive prediction of high-grade prostate cancer via biparametric MRI radiomics”. In: *J. Magn. Res. Imaging* (2020). DOI: 10.1002/jmri.27132 (cit. on p. 97).
- [176] S. Hectors et al. “MRI radiomics features predict immuno-oncological characteristics of hepatocellular carcinoma”. In: *Eur. Radiol.* 30.7 (2020), pp. 3759–3769 (cit. on p. 97).
- [177] G. Cook et al. “Challenges and promises of PET radiomics”. In: *Int. J. Radiat. Oncol. Biol. Phys.* 102.4 (2018), pp. 1083–1089 (cit. on p. 97).
- [178] Z. Liu et al. “The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges”. In: *Theranostics* 9.5 (2019), pp. 1303–1322 (cit. on pp. 97, 98).
- [179] M. Elter and A. Horsch. “CADx of mammographic masses and clustered microcalcifications: a review”. In: *Med. Phys.* 36.6 (2009), pp. 2052–2068 (cit. on p. 98).
- [180] Y. Tan, L. Schwartz, and B. Zhao. “Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field”. In: *Med. Phys.* 40.4 (2013), p. 043502. DOI: 10.1118/1.4793409 (cit. on p. 98).
- [181] C. Parmar et al. “Robust radiomics feature quantification using semiautomatic volumetric segmentation”. In: *PLoS One* 9.7 (2014), e102107. DOI:

- <https://doi.org/10.1371/journal.pone.0102107> (cit. on pp. 98, 119, 174).
- [182] C. Chowdhary and D. Acharjya. “Segmentation and feature extraction in medical imaging: a systematic review”. In: *Procedia Comput. Sci.* 167 (2020), pp. 26–36 (cit. on p. 98).
- [183] A. Zwanenburg, M. Vallières, M. Abdalah, et al. “The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping”. In: *Radiology* 295 (2020), pp. 328–338 (cit. on pp. 99, 100, 103, 119, 170, 174, 177).
- [184] R. Haralick, K. Shanmugam, and I. Dinstein. “Texture features for image classification”. In: *IEEE Trans. Syst., Man, Cybern. B. Cybern.* 3 (1973), pp. 610–621 (cit. on pp. 99, 101, 102).
- [185] M. Galloway. “Texture analysis using gray level run lengths”. In: *Comput. Gr. Image Process.* 4 (1975), pp. 172–179 (cit. on pp. 99, 104).
- [186] M. Soufi, H. Arimura, and N. Nagami. “Identification of optimal mother wavelets in survival prediction of lung cancer patients using wavelet decomposition based radiomic features”. In: *Med. Phys.* 45.11 (2018), pp. 5116–5128 (cit. on pp. 99, 104–106).
- [187] D. Patiño, A. Ceballos-Arroyo, J. Rodriguez-Rodriguez, et al. “Melanoma detection on dermoscopic images using superpixels segmentation and shape-based features”. In: *Proceedings of the 15th International Symposium on Medical Information Processing and Analysis*. Vol. 11330. SIPAIM. 2020 (cit. on p. 99).
- [188] R. Javed, M. Rahim, T. Saba, et al. “A comparative study of features selection for skin lesion detection from dermoscopic images”. In: *Netw. Model. Anal. Health Inform. Bioinform.* 9 (2020). DOI: 10.1007/s13721-019-0209-1 (cit. on p. 99).
- [189] W. Lorensen and H. Cline. “Marching cubes: a high resolution 3D surface construction algorithm”. In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques. August 1987*. 10.1145/37402.37422. SIGGRAPH. 1987, pp. 163–169 (cit. on p. 100).
- [190] M. Kolossváry et al. “Cardiac computed tomography radiomics. A comprehensive review on radiomic techniques”. In: *J. Thorac. Imaging* (2017). DOI: 10.1097/RTI.0000000000000268 (cit. on pp. 101, 103, 105).
- [191] A. Traverso, L. Wee, A. Dekker, et al. “Repeatability and reproducibility of radiomic features: a systematic review”. In: *Int. J. Radiat* 102 (2018), pp. 1143–1159 (cit. on pp. 101, 173, 180).
- [192] J. Foy et al. “Variation in algorithm implementation across radiomics software”. In: *J. Med. Imaging* 5.4 (2018), p. 044505. DOI: 10.1117/1.JMI.5.4.044505 (cit. on pp. 101, 103, 120).

- [193] B. Zhao et al. “Reproducibility of radiomics for deciphering tumor phenotype with imaging”. In: *Sci. Rep.* 6 (2018), p. 23428. DOI: 10.1038/srep23428 (cit. on pp. 103, 104, 106).
- [194] S. Livens and G. V. de Wouwer. “Wavelets for texture analysis: an overview.” In: *Proceedings of the 6th international conference on Image Processing and its Applications*. 10.1049/cp:19970958. IPA 1997. 1997 (cit. on p. 105).
- [195] A. Chaddad, P. Daniel, and T. Niazi. “Radiomics evaluation of histological heterogeneity using multiscale textures derived from 3D wavelet transformation of multispectral images”. In: *Front. Oncol.* 8.96 (2018). DOI: 10.3389/fonc.2018.00096 (cit. on pp. 106, 199).
- [196] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning. 2nd edition*. Springer, 2001. Chap. 3 (cit. on pp. 109, 111).
- [197] J. Nalepa and M. Kawulok. “Selecting training sets for support vector machines: a review”. In: *Artif. Intell. Rev.* 52 (2019), pp. 857–900 (cit. on p. 112).
- [198] A. Vabalas et al. “Machine learning algorithm validation with a limited sample size”. In: *PLoS One* 14.11 (2019), e0224365. DOI: 10.1371/journal.pone.0224365 (cit. on p. 119).
- [199] F. Orlhac et al. “Validation of a method to compensate multicenter effects affecting CT radiomics”. In: *Radiology* 291 (2019), pp. 53–59 (cit. on p. 119).
- [200] M. Shafiq-ul-Hassan, G. Zhang, and K. Latifi. “Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels”. In: *Med. Phys.* 44 (2017), pp. 1050–1062 (cit. on pp. 119, 173).
- [201] Y. Balagurunathan, V. Kumar, Y. Gu, et al. “Test-retest reproducibility analysis of lung CT image features”. In: *J. Digit. Imaging* 27 (2014), pp. 805–823 (cit. on p. 119).
- [202] P. Hu, J. Wang, H. Zhong, et al. “Reproducibility with repeat CT in radiomics study for rectal cancer”. In: *Oncotarget* 7 (2016), pp. 71440–71446 (cit. on pp. 119, 174).
- [203] S. Sanduleanu et al. “Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score”. In: *Radiother. Oncol.* 127.3 (2018), pp. 349–360 (cit. on p. 120).
- [204] Q. Li et al. “A fully-automatic multiparametric radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme”. In: *Sci. Rep.* 7 (2017), p. 1433. DOI: 10.1038/s41598-017-14753-7 (cit. on p. 120).
- [205] C. Kuhl and D. Truhm. “The long route to standardized radiomics: unraveling the knot from the end”. In: *Radiology* 295 (2020), pp. 339–341 (cit. on p. 120).
- [206] E. Klotz et al. “Technical prerequisites and imaging protocols for CT perfusion imaging in oncology”. In: *Eur. J. Radiol.* 84 (2015), pp. 2359–2367 (cit. on p. 122).

- [207] A.-A. Konstas et al. “Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, Part 2: technical implementations”. In: *Am. J. Neuroradiol.* 30 (2009), pp. 885–892 (cit. on p. 125).
- [208] A. Bevilacqua et al. “Quantitative Assessment of Effects of Motion Compensation for Liver and Lung Tumors in CT Perfusion”. In: *Acad. Radiol.* 21.11 (2014), pp. 1416–1426 (cit. on p. 127).
- [209] S. Malavasi et al. “The effects of baseline length in Computed Tomography perfusion of liver”. In: *Biomed. Signal Process. Control* 62.102135 (2020). DOI: 10.1016/j.bspc.2020.102135 (cit. on p. 127).
- [210] C. de Boor. *A practical guide to splines (Revised edition)*. Vol. 27. Springer-Verlag New York, 2001, pp. 207–214 (cit. on p. 127).
- [211] J. Lee. “Robust smoothing: smoothing parameter selection and applications to fluorescence spectroscopy”. In: *Comput. Stat. Data Anal.* 54.12 (2010), pp. 3131–3143 (cit. on p. 128).
- [212] A. Forsgren, P. Gill, and M. Wright. “Interior Methods for Nonlinear Optimization”. In: *SIAM* 44.4 (2002), pp. 525–597 (cit. on p. 128).
- [213] K. Miles. “Measurement of tissue perfusion by dynamic computed tomography”. In: *Br. J. Radiol.* 64.761 (1991), pp. 409–412 (cit. on p. 133).
- [214] A. Djuric-Stefanovic et al. “Comparison between the deconvolution and maximum slope 64-MDCT perfusion analysis of the esophageal cancer: is conversion possible?” In: *Eur. J. Radiol.* 82 (2013), pp. 1716–1723 (cit. on pp. 149, 150).
- [215] W. van Elmpt et al. “Characterization of tumor heterogeneity using dynamic contrast enhanced CT and FDG-PET in non-small cell lung cancer”. In: *Radiother. Oncol.* 109 (2013), pp. 65–70 (cit. on pp. 149, 150).
- [216] S. Schneeweiß et al. “CT-perfusion measurements in pancreatic carcinoma with different kinetic models: is there a chance for tumour grading based on functional parameters?” In: *Cancer Imaging* 16 (2016), pp. 43–50 (cit. on pp. 149, 150).
- [217] M. Kurucay, C. Kloth, S. Kaufmann, et al. “Multiparametric imaging for detection and characterization of hepatocellular carcinoma using gadoxetic acid-enhanced MRI and perfusion-CT: which parameters work best?” In: *Cancer Imaging* 2017.17:19 (2017). DOI: 10.1186/s40644-017-0121-9 (cit. on p. 149).
- [218] M. A. Fischer, H. P. Marquez, S. Gordic, et al. “Arterio-portal shunts in the cirrhotic liver: perfusion computed tomography for distinction of arterialized pseudolesions from hepatocellular carcinoma”. In: *Eur. Radiol.* 27 (2017), pp. 1074–1080 (cit. on pp. 149, 152).
- [219] D. Tamandl, F. Waneck, W. Sieghart, et al. “Early response evaluation using CT-perfusion one day after transarterial chemoembolization for HCC predicts

- treatment response and long-term disease control”. In: *Eur. J. Radiol.* 90 (2017), pp. 73–80 (cit. on p. 149).
- [220] D. Deniffel et al. “Computed Tomography perfusion measurements in renal lesions obtained by Bayesian estimation, advanced Singular-Value Decomposition deconvolution, Maximum Slope, and Patlak models.” In: *Invest. Radiol.* 00.00 (2018), pp. 1–9. DOI: 10.1097/RLI.0000000000000477 (cit. on pp. 149–151).
- [221] S. Aslan, M. S. Nural, I. Camlidag, et al. “Efficacy of perfusion CT in differentiating of pancreatic ductal adenocarcinoma from mass-forming chronic pancreatitis and characterization of isoattenuating pancreatic lesions”. In: *Abdom. Radiol.* (2018). DOI: 10.1007/s00261-018-1776-9 (cit. on p. 149).
- [222] S. Kaufmann, W. M. Thaiss, M. Schulze, et al. “Prognostic value of perfusion CT in hepatocellular carcinoma treatment with sorafenib: comparison with mRECIST in longitudinal follow-up”. In: *Acta Radiol.* 59.7 (2018), pp. 765–772 (cit. on p. 149).
- [223] M. Horger, P. Fallier-Becker, W. M. Thaiss, et al. “Is there a direct correlation between microvascular wall structure and k-Trans values obtained from perfusion CT measurements in lymphomas?” In: *Acad. Radiol.* (2018) (cit. on p. 149).
- [224] J. R. Mains, F. Donskov, E. M. Pedersen, et al. “Use of patient outcome endpoints to identify the best functional CT imaging parameters in metastatic renal cell carcinoma patients”. In: *Br. J. Radiol.* 91.20160795 (2018) (cit. on pp. 149, 150).
- [225] S. Mulé, F. Pigneur, R. Quelever, et al. “Can dual-energy CT replace perfusion CT for the functional evaluation of advanced hepatocellular carcinoma?” In: *Eur. Radiol.* 28 (2018), pp. 1977–1985 (cit. on pp. 149, 152).
- [226] Y. Nakamura, T. Kawaoka, T. Higaki, et al. “Hepatocellular carcinoma treated with sorafenib: arterial tumour perfusion in dynamic contrast-enhanced CT as early imaging biomarkers for survival”. In: *Eur. J. Radiol.* 98 (2018), pp. 41–49 (cit. on p. 149).
- [227] I. R. Andersen, K. Thorup, B. N. Jepsen, et al. “Dynamic contrast-enhanced computed tomography in the treatment evaluation of patients with colorectal liver metastases treated with ablation: a feasibility study”. In: *Acta Radiol.* 60.8 (2018), pp. 936–948. DOI: 10.1177/0284185118806661 (cit. on p. 149).
- [228] M. Wang et al. “Correlation study between dual source CT perfusion imaging and the microvascular composition of solitary pulmonary nodules”. In: *Lung Cancer* 130 (2019), pp. 115–120 (cit. on p. 149).
- [229] R. Larue et al. “Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures”. In: *Br. J. Radiol.* 90 (2017), p. 20160665 (cit. on p. 170).

- [230] M. Shafiq-ul-Hassan et al. “Voxel size and gray level normalization of CT radiomic features in lung cancer”. In: *Scientific Reports* 8 (2018), p. 10545 (cit. on pp. 170, 173).
- [231] I. Fotina et al. “Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy”. In: *Strahlentherapie Onkol.* 188 (2012), pp. 160–167 (cit. on p. 170).
- [232] J. van Timmeren et al. “Test-retest data for radiomics feature stability analysis: generalizable or study-specific?” In: *Tomography* 2 (2016), pp. 361–365 (cit. on pp. 170, 176).
- [233] K. Panetta, A. Samani, and S. Agaian. “Choosing the optimal spatial domain measure of enhancement for mammogram images”. In: *International Journal of Biomedical Imaging* (2014), Article ID 937849, 8 pages. DOI: <https://doi.org/10.1155/2014/937849> (cit. on pp. 171, 173).
- [234] A. Zwanenburg et al. “Assessing robustness of radiomic features by image perturbation”. In: *Scientific Reports* 9 (2019), pp. 614–623 (cit. on pp. 172, 174, 175, 178).
- [235] R. Song et al. “Heart Rate Estimation from Facial Videos Using a Spatiotemporal Representation with Convolutional Neural Networks”. In: *IEEE Trans. Instrum. Meas.* (2020), pp. 1–1. DOI: <https://doi.org/10.1109/TIM.2020.2984168> (cit. on p. 173).
- [236] T. Moraes et al. “Medical image interpolation based on 3D Lanczos filtering”. In: *Comput. Methods Biomech. Biomed. Eng. Imaging. Vis.* (2019). DOI: <https://doi.org/10.1080/21681163.2019.1683469> (cit. on p. 173).
- [237] K. Umehara, J. Ota, and T. Ishida. “Application of super-resolution convolutional neural network for enhancing image resolution in chest CT”. In: *Journal of Digital Imaging* 31 (2018), pp. 441–450 (cit. on p. 173).
- [238] G. Andria, F. Attivissimo, and A. M. L. Lanzolla. “A statistical approach for MR and CT images comparison”. In: *Measurement* 46 (2013), pp. 57–65 (cit. on p. 173).
- [239] H. Koyuncu and R. Ceylan. “Elimination of white Gaussian noise in arterial phase CT images to bring adrenal tumours into the forefront”. In: *Computerized Medical Imaging and Graphics* 65 (2018), pp. 46–57 (cit. on p. 173).
- [240] M. Ikeda et al. “A method for estimating noise variance of CT image”. In: *Computerized Medical Imaging and Graphics* 34 (2010), pp. 642–650 (cit. on p. 173).
- [241] V. Kovalev et al. “Three-dimensional texture analysis of MRI brain datasets”. In: *IEEE Transaction on medical imaging* 20 (2001), pp. 424–433 (cit. on p. 174).

- [242] H. Jin and J. Kim. “Evaluation of feature robustness against technical parameters in CT radiomics: verification of phantom study with patient dataset”. In: *J. Signal Process. Sys.* 92 (2020), pp. 277–287 (cit. on p. 174).
- [243] J. Tan et al. “Sharpness preserved sinogram synthesis using convolutional neural network for sparse-view CT imaging”. In: *Proc. SPIE 10949, Medical Imaging 2019: Image processing, 109490E (15 March 2019)* (2019). DOI: <https://doi.org/10.1117/12.2512894> (cit. on p. 177).
- [244] P. Whybra et al. “Assessing radiomic feature robustness to interpolation in ^{18}F -FGD PET imaging”. In: *Scientific Reports* 9 (2019), pp. 9649–9659 (cit. on p. 177).
- [245] Y. Liu et al. “Tumor heterogeneity assessed by texture analysis on contrast-enhanced CT in lung adenocarcinoma: association with pathologic grade”. In: *Oncotarget* 8 (2017), pp. 53664–53674 (cit. on p. 182).
- [246] F. Bray et al. “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. In: *CA Cancer J. Clin.* 68.6 (2018), pp. 394–424 (cit. on pp. 191, 198).
- [247] P. Rawla, T. Sunkara, and A. Barsouk. “Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors”. In: *Gastroenterology Rev.* 14.2 (2019), pp. 89–103 (cit. on p. 191).
- [248] E. Lawson et al. “Advances in the management of rectal cancer”. In: *Curr. Prob. Surg.* 56.11 (2019), p. 100648 (cit. on p. 191).
- [249] M. Gersak et al. “Endoscopic ultrasound for the characterization and staging of rectal cancer. Current state of the method. Technological advances and perspectives”. In: *Med. Ultrason.* 17.2 (2019), pp. 227–234 (cit. on p. 191).
- [250] R. Labianca et al. “Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up”. In: *Ann. Oncol.* 24.6 (2013), pp. vi64–72 (cit. on p. 191).
- [251] C. De Cecco et al. “Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance”. In: *Invest. Radiol.* 50.4 (2015), pp. 239–245 (cit. on p. 191).
- [252] V. Giannini et al. “Predicting locally advanced rectal cancer response to neoadjuvant therapy with ^{18}F -FDG PET and MRI radiomics features”. In: *Eur. J. Nucl. Med. Mol. Imaging* 46.4 (2019), pp. 878–888 (cit. on pp. 191, 198).
- [253] J. Herman et al. “The quality-of-life effects of neoadjuvant chemoradiation in locally advanced rectal cancer”. In: *Int. J. Radiat. Oncol. Biol. Phys.* 85.1 (2013), e15–9 (cit. on p. 192).
- [254] R. Sauer et al. “Preoperative versus postoperative chemoradiotherapy for rectal cancer”. In: *N. Eng. J. Med.* 351 (2004), pp. 1731–1740 (cit. on p. 192).

- [255] S. Edge and C. Compton. “The American Joint Committee on Cancer: the 7th edition of the AJCC Cancer Staging manual and the future of TNM”. In: *Ann. Surg. Oncol.* 17 (2010), pp. 1471–1474 (cit. on p. 192).
- [256] R. Betts-Tan et al. “Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting”. In: *Eur. Radiol.* 28.4 (2018), pp. 1465–1475 (cit. on p. 193).
- [257] M. Abramoff, P. Magalhaes, and S. Ram. “Image Processing with ImageJ”. In: *Biophotonics Int.* 11.7 (2004), pp. 36–42 (cit. on p. 194).
- [258] Z. Liu et al. “Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer”. In: *Clin. Cancer Res.* 23.23 (2017), pp. 7253–7262 (cit. on p. 198).
- [259] D. Cusumano et al. “Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer”. In: *Radiol. Med.* 123.4 (2018), pp. 286–295 (cit. on p. 198).
- [260] Y. Li et al. “Predicting pathological complete response by comparing MRI-based radiomics pre- and postneoadjuvant radiotherapy for locally advanced rectal cancer”. In: *Cancer Med.* 8.17 (2019), pp. 7244–7252 (cit. on p. 198).
- [261] Y. Cui et al. “Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer”. In: *Eur. Radiol.* 29.3 (2019), pp. 1211–1220 (cit. on p. 198).
- [262] X. Zhou et al. “Radiomics-Based pretherapeutic prediction of non-response to neoadjuvant therapy in locally advanced rectal cancer”. In: *Ann. Surg. Oncol.* 26.6 (2019), pp. 1676–1684 (cit. on p. 198).
- [263] B. Petresce et al. “Pre-treatment T2-WI based radiomics features for prediction of locally Advanced rectal cancer non-response to neoadjuvant chemoradiotherapy: a preliminary study”. In: *Cancer (Basel)* 12.7 (2020), p. 1894 (cit. on p. 198).
- [264] I. Petkovska et al. “Clinical utility of radiomics at baseline rectal MRI to predict complete response of rectal cancer after chemoradiation therapy”. In: *Abdom. Radiol. (NY)* 45.11 (2020), pp. 3608–3617 (cit. on p. 198).
- [265] T. McClintock et al. “Prostate cancer management costs vary by disease stage at presentation”. In: *Prostate Cancer Prostatic Dis.* 23.4 (2020), pp. 564–566 (cit. on p. 199).
- [266] F. Drost et al. “Can active surveillance reduce the harms of overdiagnosis prostate cancer? A reflection of real life clinical practice in the PRIAS study”. In: *Transl. Androl. Urol.* 7.1 (2018), pp. 98–105 (cit. on p. 199).

- [267] M. Bjurlin et al. “Multiparametric MRI and targeted prostate biopsy: improvements in cancer detection, localization, and risk assessment”. In: *Cent. European J. Urol.* 69 (2016), pp. 9–18 (cit. on p. 199).
- [268] J. Epstein et al. “Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades”. In: *Eur.Urol.* 61.5 (2012), pp. 1019–1024 (cit. on p. 199).
- [269] D. Fehr et al. “Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images”. In: *Proc. Natl. Acad. Sci. USA* 112.46 (2015), E6265–6273 (cit. on pp. 199, 211, 216).
- [270] D. Rosario et al. “Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study”. In: *BMJ* 344 (2012), p. d7894. DOI: 10.1136/bmj.d7894 (cit. on pp. 199, 211).
- [271] H. Ahmed et al. “Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study”. In: *Lancet* 389.10071 (2017), pp. 815–822 (cit. on pp. 199, 214).
- [272] A. Chatterjee et al. “Changes in Epithelium, Stroma, and Lumen Space correlate more strongly with Gleason pattern and are stronger predictors of prostate ADC changes than cellularity metrics”. In: *Radiology* 277.3 (2015), pp. 751–762 (cit. on pp. 199, 216).
- [273] S. Barbieri et al. “Differentiation of prostate cancer lesions with high and with low Gleason score by diffusion-weighted MRI”. In: *Eur. Radiol.* 27.4 (2017), pp. 1547–1555 (cit. on pp. 200, 211, 212, 215).
- [274] A. Hiremath et al. “Test-retest repeatability of a deep learning architecture in detecting and segmenting clinically significant prostate cancer on apparent diffusion coefficient (ADC) maps”. In: *Eur. Radiol.* (2020). DOI: 10.1007/s00330-020-07065-4 (cit. on pp. 200, 211, 215).
- [275] Y. Zhang et al. “Development of a novel, multi-parametric, MRI-based radiomic nomogram for differentiating between clinically significant and insignificant Prostate Cancer”. In: *Front. Oncol.* 10 (2020), p. 888 (cit. on pp. 200, 211, 215).
- [276] D. Bonekamp et al. “Radiomic machine learning for characterization of prostate lesions with MRI: Comparison to ADC values”. In: *AJR Am. J. Roentgenol.* 289.1 (2018), pp. 128–137 (cit. on pp. 200, 211, 215, 216).
- [277] S. Hurrel et al. “Optimized b-value selection for the discrimination of prostate cancer grades, including the cribriform pattern, using diffusion weighted imaging”. In: *J Med Imaging (Bellingham)* 5.1 (2018), p. 011004 (cit. on p. 207).

- [278] A. Gibaldi et al. “Effects of guided random sampling of TCCs on blood flow values in CT perfusion studies of lung tumors”. In: *Acad. Radiol.* 22.1 (2015), pp. 58–69 (cit. on p. 207).
- [279] M. Li, F. Chen, and J. Kou. “Candidate vectors selection for training support vector machines”. In: *Proceedings of the 3rd international conference on Natural Computation*. 10.1109/ICNC.2007.292. ICNC 2007. 2007 (cit. on p. 209).
- [280] A. Schulman et al. “The Contemporary Role of Multiparametric Magnetic Resonance Imaging in Active Surveillance for Prostate Cancer”. In: *Curr. Urol. Rep.* 18.7 (2017), p. 52. DOI: 10.1007/s11934-017-0699-2 (cit. on p. 210).
- [281] G. Cristel et al. “Can DCE-MRI reduce the number of PI-RADS v.2 false positive findings? Role of quantitative pharmacokinetic parameters in prostate lesions characterization”. In: *Eur. J. Radiol.* 118 (2019), pp. 51–57 (cit. on pp. 211, 215).
- [282] X. Min et al. “MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: Cross-validation of a machine learning method”. In: *Eur. J. Radiol.* 115 (2019), pp. 16–21 (cit. on pp. 211, 215).
- [283] A. Rosenkrantz et al. “Transition zone prostate cancer: revisiting the role of multiparametric MRI at 3 T”. In: *AJR Am. J. Roentgenol.* 204.3 (2015), W266–72 (cit. on pp. 212, 215, 216).
- [284] G. Litjens et al. “Computer-extracted features can distinguish noncancerous confounding disease from prostatic adenocarcinoma at multiparametric MR Imaging”. In: *Radiology* 278.1 (2016), pp. 135–145 (cit. on p. 212).
- [285] S. Loeb et al. “Overdiagnosis and overtreatment of prostate cancer”. In: *Eur. Urol.* 65.6 (2014), pp. 1046–1055 (cit. on p. 214).
- [286] M. Zhang et al. “Value of increasing biopsy cores per target with cognitive MRI-targeted transrectal US prostate biopsy”. In: *Radiology* 291.1 (2016), pp. 83–89 (cit. on p. 215).
- [287] K. Grant et al. “Comparison of calculated and acquired high b value diffusion-weighted imaging in prostate cancer”. In: *Abdom. Imaging* 40.3 (2015), pp. 578–586 (cit. on p. 216).

Acknowledgements

This work has been accomplished thanks to the contribution of many. Firstly, I would like to thank both my PhD supervisor, Prof. Alessandro Bevilacqua, and my co-supervisor, Prof. Mauro Ursino. I am really grateful to Prof. A. Bevilacqua for placing his trust in me. I really appreciate the time he has devoted to me, the day-by-day guidance he has given me, the scientific training he has delivered to me and the way he helped me grow both as a professional and as a person.

I would also like to thank Prof. Valérie Vilgrain, leading the Dept. of Radiology of Beaujon Hospital (Paris, FR), for involving me in the PIXEL and SARAH projects, alongside Prof. Vicky Goh, leading the Dept. of Radiology of King's College (London, UK), whom I would like to thank for providing me with PROSPeCT data. Prof. Evis Sala also has my gratitude for hosting me at the Radiogenomics and Quantitative Imaging Group of the Dept. of Radiology of the Addenbrooke's Hospital (Cambridge, UK), and Dr. Leonardo Rundo for building this connection. My gratitude extends to Dr. Domenico Barone heading the Diagnostic Imaging Unit of IRCCS-IRST (Meldola, IT), Prof. Giampaolo Gavelli, and Dr. Fabio Ferroni to whom the CVG is tied by a consolidated cooperative relationship. I am thankful for their proactiveness and precious suggestions. Furthermore, I am grateful to Prof. Rita Golfieri leading the Dept. of Radiology of S.Orsola-Malpighi Hospital (Bologna, IT) and her entire working team, in particular Dr. Francesca Coppola and Dr. Silvia Lo Monaco. I thank all physicians of the team of Prof. R. Golfieri for the enthusiasm with which they have welcomed our collaboration and encouraged our research.

On a personal note, I am grateful to my parents, my sister, and Gianvito, not only for their continuous support and their patience, but also for the enthusiasm and joy which we have shared together celebrating my successes.

I am thankful to Barbara, Danila, and Loretta who have always been there for me. And finally I thank Francesco, who has supported and encouraged me on both my good and my bad days, continually turning the bad days into better ones.

THESIS END

Margherita Mottola

E-mail: margherita.mottola@unibo.it

Research group website: <https://cvg.deis.unibo.it/>

Personal website: <https://www.unibo.it/sitoweb/margherita.mottola>

(Curriculum Vitae available on the personal website)

Office address:

Via Toffano 2/2, 40125, Bologna, Italy

Home address:

Via Torleone 23, 40125, Bologna, Italy