UNIVERSITÀ DEGLI STUDI DI MILANO

DEPARTMENT OF PHYSICS

PHD SCHOOL IN
PHYSICS, ASTROPHYSICS AND APPLIED PHYSICS
CYCLE XXXIII

NOVEL PHANTOMS FOR ROBUST MRI-BASED RADIOMICS IN ONCOLOGY

Disciplinary Scientific Sector FIS/07

PhD Thesis of:

Linda Bianchini

Director of the School:
Prof. Matteo Paris

Supervisors of the Thesis:
Prof. Alessandro Lascialfari
Prof. Ivan Veronese

A.Y. 2019-2020

Commission of the final examination:

Professor Angelo Bifone, Università degli Studi di Torino

Professor Pasquina Marzola, Università degli Studi di Verona

Professor Eric Thiaudiere, Université de Bordeaux

Final examination:

10^{th} November 2020

Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

Cover illustration:

Axial T_2-weighted MR image of the pelvic phantom designed in this thesis work to support radiomic studies in oncology.

MIUR subjects:

FIS/07

To Sara

"Bene, gatto. Ci siamo riusciti" disse sospirando.
"Sì, sull'orlo del baratro ha capito la cosa più importante" miagolò Zorba.
"Ah sì? E cosa ha capito?" chiese l'umano.
"Che vola solo chi osa farlo" miagolò Zorba.

Storia di una gabbianella e del gatto che le insegnò a volare

Luis Sepúlveda

Contents

Introduction

Motivation and key findings

Radiomics is a technique of medical image analysis that has started to be developed in the last years, with focus in oncology. The term refers to the extraction of quantitative features from standard-of-care images, including Computed Tomography (CT), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI), that are part of the diagnostic process and follow-up in oncologic patients. Thousands of "synthetic biomarkers" can be calculated starting from the grey level intensities in each voxel enclosed within the tumoural region, providing a description of the tumour heterogeneity and texture. These biomarkers, called *radiomic features*, have been investigated as they can be predictive of specific clinical outcomes, e.g. the choice of a treatment strategy, the prognosis or patients' stratification, and thus they can be a useful support for the process of decision-making in healthcare.
The application of radiomics in MRI is particularly challenging, given the number of variables included in the process of image acquisition, reconstruction, processing and interpretation. Recently, the literature has been enriched with many promising studies on MRI-based radiomics applied to various anatomical districts and cancer diseases. However, the majority of the published studies lack of standardisation in the radiomic process and the following analysis. The harmonisation of the radiomic workflow, from the image acquisition to the building of predictive models for clinical support, is one of the goal of present radiomics. The validation of the findings on extended clinical database is another important gap in the current state of MRI radiomics, and it is necessary for bringing it into daily clinical practice in the near future. In addition, the investigation on the quality of the radiomic features extracted from the medical images is of major importance, to ensure the robustness and generalisability of the derived predictive models. Part of the open issues in MRI-based radiomics might be fulfilled with phantom studies, which offer the possibility for multiple acquisitions in a controlled experimental setup. Nevertheless, the phantoms already available in medical centres for quality controls in MRI are not suitable for radiomic investigations, as they lack the heterogeneity typical of the investigated tumours.

The present thesis work was performed in this framework, with the aim to provide useful indications towards the extraction of robust radiomic features with a well-defined and optimised protocol for image handling and analysis. The work was carried out in collaboration with two cancer centres: the European Institute of Oncology (IEO, Milan) and the Champalimaud Centre for the Unknown (CCU, Lisbon), with the clinical interest focused on malignancies in the female pelvis (e.g. carcinoma of the cervix) and, for the last part of the thesis and the ongoing studies, on breast cancer. The primary objective of the thesis was the fabrication of *ad hoc* MRI radiomic phantoms that could mimic the radiomic properties of a set of real patients and that could be used for the assessment of the radiomic features robustness in clinical-like scenarios. A secondary objective was the evaluation of the radiomic features repeatability and reproducibility in the context of diagnostic MRI for female pelvic cancer. As a third objective, this thesis proposed to assess the impact of the choice of a specific software for the features extraction on the feature values themselves.
The described objectives have resulted in the fabrication of two innovative phantoms for MRI-based radiomics. The first phantom reproduces the abdomen of a patient affected by a pelvic tumour and was built with materials and compositions capable of mimicking the relaxation times and texture properties of human tissues as observed in a set of patients. Such phantom was exploited in a multicentric investigation on the stability of the radiomic features, considering three different MRI scanners. The considered scenarios included a study on the repeatability of the features with scanner-induced variations and repositioning-induced variations. The features reproducibility when changing the scanner type, the field strength or the acquisition parameters was evaluated as well. The results showed that a consistent percentage of features were not repeatable or not reproducible, suggesting that clinical studies performed on inhomogeneous dataset should be conducted with awareness and caution, after an acquainted assessment of the features stability in the specific scenario through phantom studies. Moreover, this investigation allowed the establishment of a well-defined procedure to identify repeatable, reproducible and informative radiomic features. Another contribution of this work in the field of MRI-based radiomics consisted in the development of a second phantom. The first prototype of a breast phantom, including an insert to mimic the signal and radiomic properties of a real tumour, was designed coupling a research on innovative materials and the 3D-printing technology. In addition, the analysis of different radiomic software allowed to identify the main sources of discrepancies which cause the features to differ if extracted with different tools, hindering the robustness and generalisation of results.
An overview of the thesis content and structure is given in the next Section.

Thesis overview

In Chapter 1 the physical principles of Nuclear Magnetic Resonance (NMR) will be summarised, both with classical and quantum approaches. The Bloch equations describing the temporal evolution of the magnetisation will be recalled, with focus on the relaxation times T_1 and T_2. In addition, the main techniques for relaxometry experiments will be

explained. The translation of the NMR phenomenon into the MRI technique will be presented, with particular attention to the imaging parameters affecting the MR signal and the way to obtain images with different contrast. The concepts of field gradient and k space will be illustrated, introducing the slice selection, the frequency and the phase encoding.
Chapter 2 will focus on radiomics. The fundamentals of this technique will be covered, including the rationale of its application on medical images of cancer patients and its potentiality. The radiomic workflow will be illustrated, in the context of the big data analytics and with connections to genomics. Special attention will be given to the texture analysis, including a mathematical description of the main categories of texture features considered in this work. The challenges and open issues of radiomics will be explained, with focus on the radiomic analysis of MR images. The need for methodological studies for the harmonisation and strengthening of the radiomic analysis will be highlighted, with particular emphasis on the potentiality of radiomic phantom studies. A review of the literature on the major results of MRI-based radiomics in pelvic and breast cancer will be discussed, including a description of the state of the art of both clinical and methodological studies.
In Chapter 3 the results concerning the comparison of three freely available radiomic software, namely *IBEX*, *LIFEx* and *PyRadiomics*, will be illustrated. The investigation aimed at evaluating the compatibility of the value of the radiomic features when extracted from the same set of MR images with different packages and similar initial conditions. The main properties of the three tools will be presented and the procedure for the image acquisition and segmentation, the radiomic feature selection and extraction will be explained in detail. Finally, the major sources of discrepancies among the results obtained with the three software will be identified and discussed.
Chapter 4 will deal with the results on the development of a dedicated pelvic phantom for MRI-based radiomic analysis and the radiomic experiments performed on such phantom. The Chapter will be divided in two parts. The first part will cover the realisation of the phantom, from the initial design concept to the production and validation, including the measurement of the relaxation times T_1 and T_2 *in vivo* on a set of patients affected by pelvic cancer. In the second part, the results of the radiomic experiments performed on the phantom will be illustrated. The tests were carried out on three different MRI scanners of two medical centres and aimed at evaluating the repeatability (in the case of both 2D and 3D acquisitions) and the reproducibility of the radiomic features in a clinical-like scenario. The reproducibility of the features will be evaluated in a set of possible scenarios, including different magnetic field strengths, different vendors or variable imaging parameters. The ability of the radiomic features to distinguish different textures will be briefly discussed in relation to the size of the region of interest from which the features were extracted. In addition, the dependency of the texture features on the shape features will be assessed, to obtain useful information on the quality of the extracted features.
In Chapter 5 the preliminary results on the realisation of a phantom to support MRI radiomics applied to breast cancer will be presented. The phantom design and production in its first prototype will be illustrated, with particular attention on the insert reproduc-

ing the MR signal and texture of a real tumour. The preliminary validation and the schedule for the following prototype will be presented, along with the radiomic experiments to be performed on the breast phantom in the near future and its application to two identified clinical scenarios.

CHAPTER 1

Nuclear magnetic resonance and its application in medical imaging

This Chapter has two main purposes. Firstly, the physical principles of Nuclear Magnetic Resonance (NMR) will be introduced, including the definition of the typical relaxation times T_1 and T_2 and the description of the main experimental techniques used to measure them in this thesis work. Secondly, the basic concepts in Magnetic Resonance Imaging (MRI) will be summarised, focusing on the transition from an NMR experiment to an MR image reconstruction.

1.1 Fundamentals of Nuclear Magnetic Resonance

The NMR phenomenon was discovered by Purcell [1] and Bloch [2] in 1946 and studies the temporal evolution of the nuclear magnetisation in a set of nuclei, due to the hyperfine interactions in which the nuclei are involved. For the purpose of this Chapter, the summary presented in this Section will be focused on the interaction between the nuclear magnetic moments of the nuclei and an external magnetic field [3].

1.1.1 The precessional motion: classical approach

Each of the probe nuclei, used in an NMR experiment, has a nuclear magnetic moment defined as:

$$\vec{\mu} = \gamma \hbar \vec{I} \tag{1.1}$$

where γ is the *gyromagnetic ratio*, equals to $2\pi \cdot 42.576$ MHz $\cdot$ T^{-1} for protons, and $\vec{I}$ is the nuclear spin. From a macroscopic point of view, the observed spin population generates a resultant magnetisation, given by:

$$\vec{M} = \sum_i \vec{\mu}_i \tag{1.2}$$

Let $\vec{H}_0$ be an external magnetic field oriented along the $\hat{z}$ axis of the laboratory frame S. In the absence of an external field, the total magnetisation is $\vec{M} = 0$, because the magnetic moments are randomly oriented. After the application of $\vec{H}_0$, the magnetic

moments are preferably aligned along the external field direction, thus a non-null resultant magnetisation is observed. Following the classical approach, the temporal evolution of the magnetisation, when $\vec{H}_0$ is applied, is given by the equation of motion:

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{H}_0 \tag{1.3}$$

Equation 1.3 describes a clockwise precessional motion (Figure 1.1) of the vector $\vec{M}$ about $\vec{H}_0$, with an angular velocity given by the *Larmor law*:

$$\vec{\omega}_0 = -\gamma \vec{H}_0 \tag{1.4}$$

Let S' be a new reference frame, with axis $\hat{z}' \equiv \hat{z} \parallel \vec{H}_0$ and $\hat{x}'$ and $\hat{y}'$ axes rotating in the

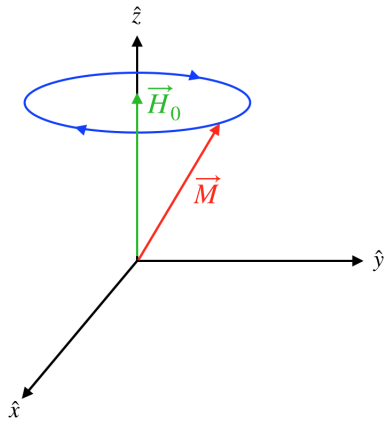


Figure 1.1: Clockwise precession of the magnetisation $\vec{M}$ in the presence of an applied external field in the laboratory reference frame S.

plane with an angular velocity $\vec{\omega}$ (Figure 1.2). The equation of motion in S' is then:

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \left(\vec{H}_0 + \frac{\vec{\omega}}{\gamma} \right) \tag{1.5}$$

which means that in the rotating frame S', the magnetisation has a precessional motion about an effective magnetic field

$$\vec{H}_{eff} = \vec{H}_0 + \frac{\vec{\omega}}{\gamma} \tag{1.6}$$

If $\vec{\omega} = \vec{\omega}_0 = -\gamma \vec{H}_0$, then $\frac{d\vec{M}}{dt} = 0$. This means that if S' is rotating at the Larmor frequency, the magnetisation $\vec{M}$ is constant in the rotating reference frame, enabling us to study the interactions of the nuclei, beyond the precessional motion.

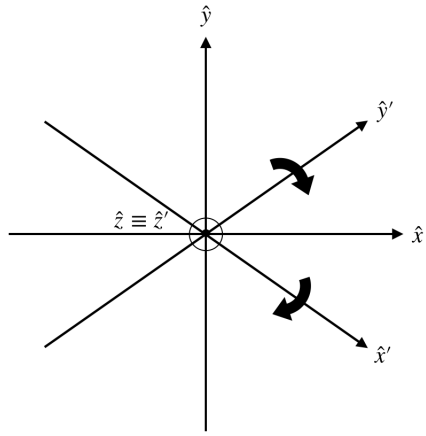


Figure 1.2: Comparison of the laboratory reference frame S and the rotating reference frame S' with $\hat{z}'$ along the external field direction and $\hat{x}'$ and $\hat{y}'$ rotating in the $\hat{x}$-$\hat{y}$ plane at $\vec{\omega}$.

1.1.2 The NMR signal

Let $\vec{H}_1$ be a magnetic field oriented along the $\hat{z}'$ axis of the frame S', which we now consider to be rotating at $\vec{\omega} = \vec{\omega}_0$ (*resonance condition*). $\vec{H}_0$ is a constant static field, on the contrary $\vec{H}_1$ is an oscillating radiofrequency (rf) field much weaker than $\vec{H}_0$ and perpendicular to it [4]. The oscillating field is generated by a coil in which an alternating current flows. Under these conditions, in S':

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{H}_1 \tag{1.7}$$

Equation 1.7 describes a precessional motion of $\vec{M}$ about $\vec{H}_1$ with angular velocity $\vec{\omega}_1 = \gamma \vec{H}_1$. The motion is represented in Figure 1.3. At $t = 0$ the magnetisation $\vec{M}$ is oriented along $\hat{z}'$ and for $t > 0$, $\vec{M}$ starts to rotate around $\vec{H}_1$. Thus, by applying $\vec{H}_1$, it is possible to guide the magnetisation out of the $\hat{z}'$ direction by an angle α, which is called *flip angle* and is given by:

$$\alpha = \gamma H_1 \delta t \tag{1.8}$$

where δt is the interval in which the rf field $\vec{H}_1$ is applied. If δt is long enough to produce $\alpha = \frac{\pi}{2}$, the rf field application is referred to as a $90°$ *pulse*. Similarly, if $\alpha = \pi$, a $180°$ *pulse* was applied.
After the application of the oscillating field $\vec{H}_1$, in the reference frame S the magnetisation $\vec{M}$ describes a spiral motion around the $\hat{z}$ axis. Due to the electromagnetic induction, the transverse component M_{xy} of the magnetisation generates an electric signal detected by the receiver coil and referred to as *Free Induction Decay* (FID).

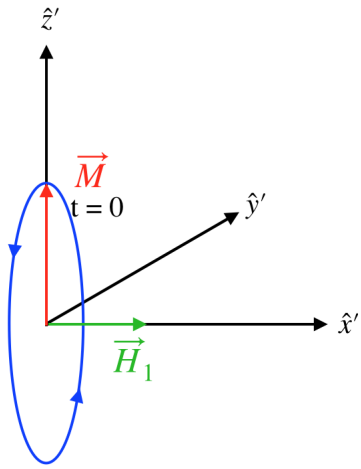


Figure 1.3: Precessional motion of the magnetisation about the radiofrequency field $\vec{H}_1$ in the reference frame S'. At $t = 0$ the magnetisation $\vec{M}$ is oriented along $\hat{z}'$ and for $t > 0$, $\vec{M}$ starts to rotate around $\vec{H}_1$.

1.1.3 Quantum approach

From the quantum mechanics point of view, the interaction of the spin population with an external field $\vec{H}_0 = H_0\hat{z}$ is described by the Zeeman hamiltonian:

$$\mathcal{H} = -\gamma\hbar\vec{I}\cdot\vec{H}_0 = -\gamma\hbar I_z H_0 \tag{1.9}$$

For the nuclei with $I = \frac{1}{2}$, like protons, the Zeeman effect generates two energy levels in the spin population, given by the energy eigenvalues:

$$E_\uparrow = -\frac{1}{2}\hbar\omega_0 \tag{1.10}$$

and

$$E_\downarrow = +\frac{1}{2}\hbar\omega_0 \tag{1.11}$$

as represented in Figure 1.4.
The number $N_\uparrow$ of spin parallel and $N_\downarrow$ anti-parallel to the external field in the two energy levels can be calculated according to the *Boltzmann statistics*:

$$\frac{N_\uparrow}{N_\downarrow} = e^{\left(\frac{\Delta E}{K_B T}\right)} \tag{1.12}$$

where ΔE is the difference in the energy levels, given by

$$\Delta E = E_\downarrow - E_\uparrow = \hbar\omega_0, \tag{1.13}$$

K_B is the Boltzmann constant, and T is the absolute temperature. At room temperature, for a 1 T magnetic field, $\frac{N_\uparrow}{N_\downarrow}$ = 1.000007 [5]. This indicates a small excess of protons in the

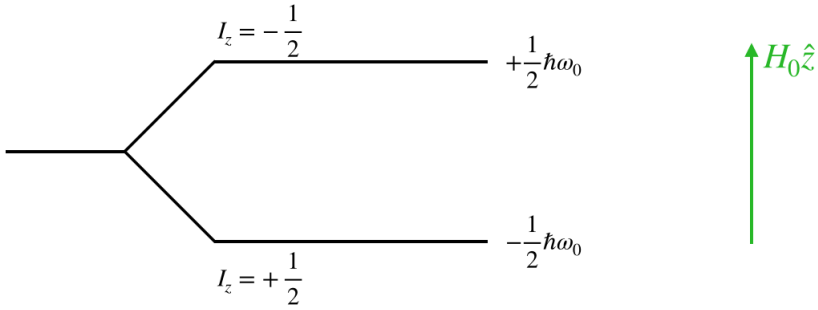


Figure 1.4: The Zeeman energy levels for a proton $\left(I = \frac{1}{2}\right)$ when an external magnetic field $\vec{H}_0$ is applied. The Zeeman effect generates two energy levels in the spin population, given by the energy eigenvalues $E_\uparrow$ and $E_\downarrow$.

lower energy level and, consequently, it is possible to observe a resultant magnetisation $\vec{M} \neq 0$.

In quantum terms, the application of the rf field $\vec{H}_1$ generates an excited states in the energy levels of the system. For example, after the application of a $90°$ pulse, the magnetisation is in the $\hat{x}$-$\hat{y}$ plane and $M_z\left(\frac{\pi}{2}\right) = 0$. This implies that the mean statistical value of I_z is null, i.e. the two Zeeman energy levels are equipopulated. From Equation 1.12, if $N_\uparrow/N_\downarrow = 1$, than $T \to \infty$. In other words, the $90°$ pulse causes the temperature of the spin system to rise, bringing them in an excited state. After the rf pulse, the system evolves to a condition of *thermal equilibrium* through transitions from the excited state to the ground level.

1.1.4 Bloch equations

To describe the temporal evolution of the magnetisation after the application of an rf field, in the 1940s Bloch modified Equation 1.3, by adding two phenomenological terms. The temporal evolution of the longitudinal component $\vec{M}_z = M_z\hat{z}$ and the transverse component $\vec{M}_{xy} = M_x\hat{x} + M_y\hat{y}$ is given respectively by Equations 1.14 and 1.15, referred to as *Bloch equations*.

$$\frac{dM_z}{dt} = \gamma\left(\vec{M} \times \vec{H}_0\right)_z + \frac{M_0 - M_z}{T_1} \tag{1.14}$$

$$\frac{dM_{xy}}{dt} = \gamma\left(\vec{M} \times \vec{H}_0\right)_{xy} + \frac{M_{xy}}{T_2} \tag{1.15}$$

In Equation 1.14, M_0 is the magnetisation value at equilibrium and T_1 is called *spin-lattice relaxation time* and it is the typical time describing the regrowth of $\vec{M}_z$ to the equilibrium value, after the perturbation introduced by the rf field. The solution for Equation 1.14 is shown in Figure 1.5 and is given by:

$$M_z(t) = M_0\left(1 - e^{-\frac{t}{T_1}}\right) \tag{1.16}$$

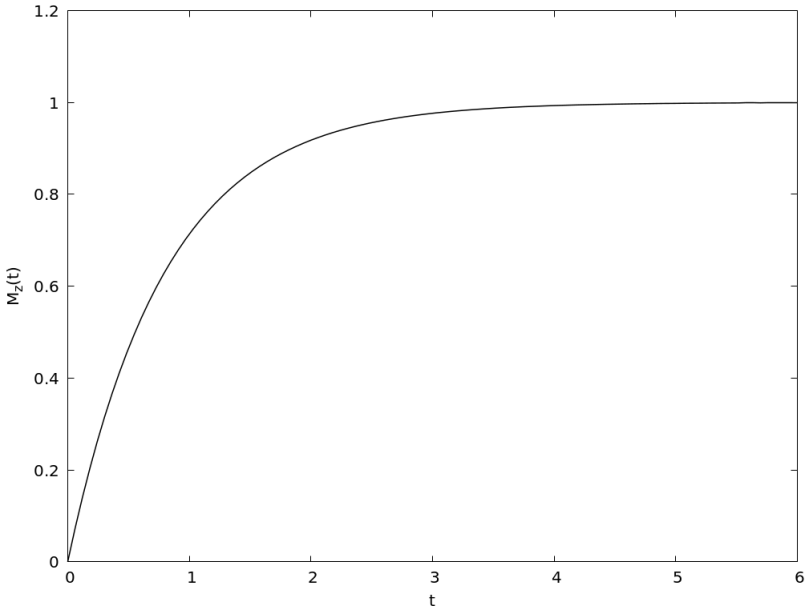


Figure 1.5: The regrowth of the longitudinal magnetisation $M_z\,(t)$ with the typical time T_1 according to the first Bloch equation.

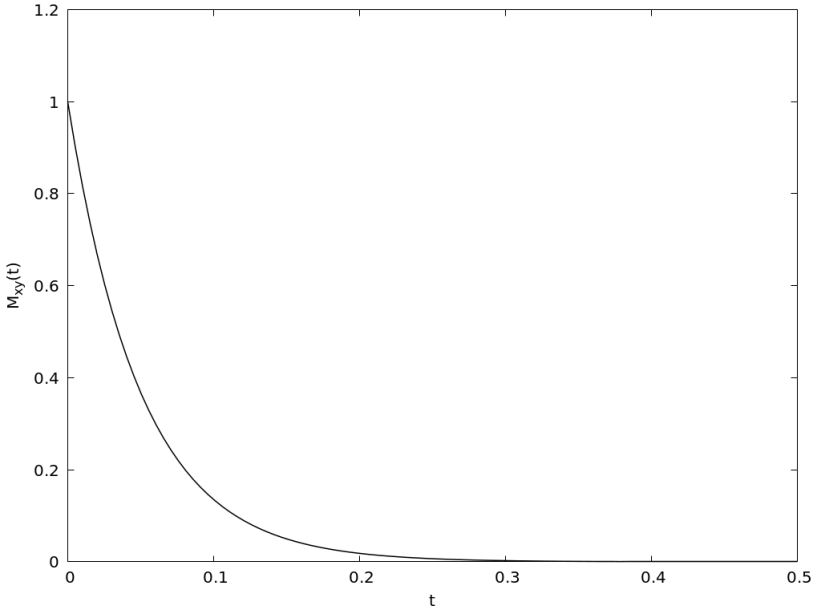


Figure 1.6: The decay of the transverse magnetisation $M_{xy}\,(t)$ with the typical time T_2 according to the second Bloch equation.

In Equation 1.15, Bloch introduced the parameter T_2, called *spin-spin relaxation time*, to describe the exponential decay of $\vec{M}_{xy}$, as shown by the plot in Figure 1.6. The solution of the second Bloch equation is:

$$M_{xy}(t) = M_0 e^{-\frac{t}{T_2}} \tag{1.17}$$

The decay is correlated to variations or inhomogeneities in the local magnetic field experienced by each magnetic moment in the system, causing local precessional frequencies slightly different from the Larmor one. This process is referred to as *dephasing* and causes the decay of the resultant transverse magnetisation.

1.1.5 Relaxometry

By the means of an NMR spectrometer, relaxometry experiments aim at measuring the relaxation times T_1 and T_2 of a sample, using different excitation sequences, i.e. characteristic series of rf pulses. In this Section, the experimental techniques used for this work will be described.

Measurement of the transverse relaxation time T_2

Equation 1.17 has been obtained under the assumption that the static field $\vec{H}_0$ is homogeneous on all the volume occupied by the sample. From an experimental point of view, this condition is not fulfilled and the dephasing effect of the transverse components is enhanced. As a consequence, the transverse magnetisation M_{xy} decays with a time constant T_2^*, which is given by both the spin-spin interaction (T_2) and the dephasing caused by the field inhomogeneities (T_2') :

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'} \tag{1.18}$$

To recover the effects of the field inhomogeneities, Hahn proposed the *Spin Echo* (SE) sequence [6]. This sequence includes a first $\pi/2$ pulse, followed by a π pulse after a time interval τ_{echo}. The second pulse has a refocusing effect and at time $2\tau_{echo}$ it is possible to observe the so-called signal of spin echo. The echo signal results to be attenuated, according to Equation 1.17:

$$M_{xy}(2\tau_{echo}) = M_0 e^{-\frac{2\tau_{echo}}{T_2}} \tag{1.19}$$

Equation 1.19 holds when the T_2^* effects are negligible. In case of diffusion inside the sample (especially in liquid samples), the spins dephasing is recovered only partially, causing a loss of signal. To limit the effect of diffusion in T_2 measurements, the *Carr-Purcell-Meiboom-Gill* (CPMG) sequence [7] [8] was developed. As shown in Figure 1.7, after the refocusing pulse of the SE sequence, the CPMG sequence includes a train of π pulses, each separated of $2\tau_{CP}$. As a result, for $t = 2n\tau_{CP}$, an echo signal is generated and the M_{xy} decay is reconstructed. It can be shown that the relaxation of the transverse magnetisation, when measured with a CPMG sequence, is given by:

$$M_{xy}(t) = M_0 e^{-\frac{t}{T_2}} e^{-\gamma^2 G^2 D (2\tau_{CP})^2 \frac{t}{12n^2}} \tag{1.20}$$

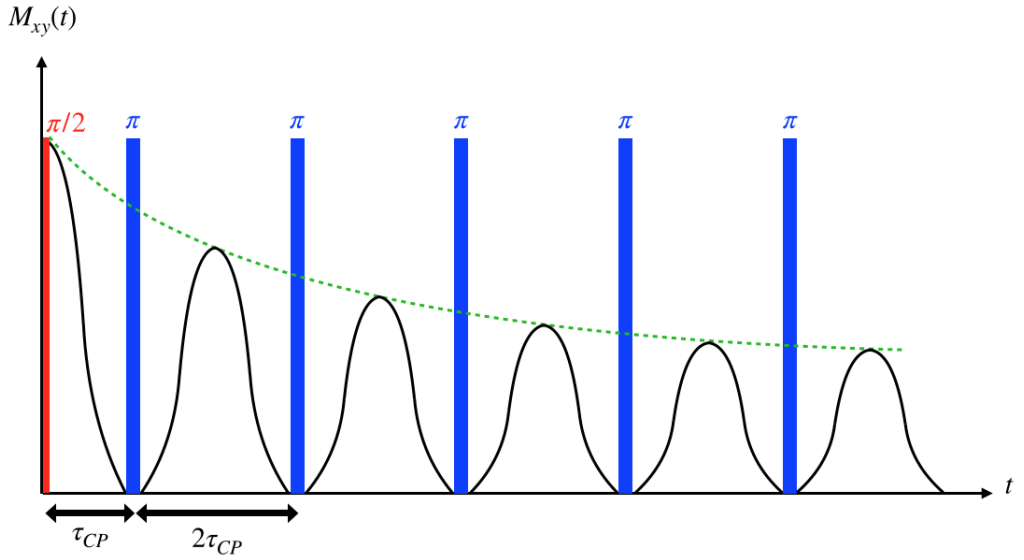


Figure 1.7: CPMG sequence. This sequence includes a first $\pi/2$ pulse, followed by a train of π pulses. The signal attenuation (green line) follows an exponential decay with a typical time T_2 (see Equation 1.17).

where G is a field gradient taken as a model for the inhomogeneities, D is the diffusion coefficient typical of the sample and n is the number of refocusing π pulses. Equation 1.20 holds when the T_2^* effects are negligible and shows that the effect of diffusion may be neglected by increasing n.
The CPMG sequence was used in this work to measure the transverse relaxation time of $MnCl_2$ solutions, during the process of the pelvic phantom development. The experimental results will be given in Section 4.1.2.

Measurement of the longitudinal relaxation time T_1

The longitudinal relaxation time is usually measured with a *Saturation Recovery* (SR) or an *Inversion Recovery* (IR) sequence. Both sequences were tested in this work on some samples, providing compatible results. Thus, the SR technique was chosen for two main reasons: (i) the total acquisition time is less for the SR than for the IR; (ii) the electronics of the spectrometer is more precise on shorter pulses ($\pi/2$) than on longer ones (π). In fact, an SR sequence usually comprises two pulses: an *excitation pulse* and a *reading pulse*, to read the signal. The excitation is performed with a $\pi/2$ pulse, to bring the magnetisation in the xy plane. After a certain time interval t, a $\pi/2$ reading pulse or a SE sequence is used to read the signal and to detect the magnetisation restored along the $\hat{z}$ axis. The magnetisation M_z can be obtained by varying the interval t between the excitation and the reading pulses and inverting Equation 1.16.
In the relaxometry experiment performed to measure the longitudinal relaxation time of

$MnCl_2$ solutions (see Section 4.1.2), the excitation of the system was carried out through a series of three $\pi/2$ pulses, aiming at saturating the absorption line. The scheme of the sequences used in this work to measure T_1 and T_2 is shown in Figure 1.8.

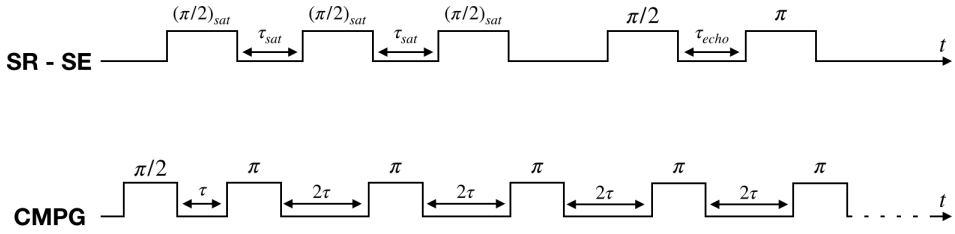


Figure 1.8: Diagrams of pulse sequences used to measure T_1 (SR-SE) and T_2 (CPMG) in this thesis work.

1.2 Physical principles of Magnetic Resonance Imaging

This Section will describe how the NMR technique can be exploited to obtain medical images. Magnetic Resonance Imaging (MRI) was introduced in the 1980s and nowadays is increasingly being used in healthcare for its excellent soft tissue contrast and the non-involvement of ionizing radiation.

1.2.1 MRI signal and contrast images

When an MR image is acquired, the signal detected can be expressed as:

$$S(t) \propto \rho\left({}^1H\right) e^{\frac{-TE}{T_2^*}} \left(1 - e^{\frac{-TR}{T_1}}\right) e^{-bD} \tag{1.21}$$

where $\rho\left({}^1H\right)$ is the proton density in the sample, TE is the *time of echo*, defined as the time interval between the first pulse in a sequence and the the signal acquisition, TR is the *repetition time*, defined as the time interval between the first pulse of the sequence and the the same pulse of the following repeated sequence, b is the diffusion weighting factor (which depends on the gradient amplitudes and timings) and D is the diffusion coefficient. $\rho\left({}^1H\right)$, T_1, T_2^* and D are intrinsic parameters (i.e. they depend on the sample - or tissue - under investigation), on the contrary b, TE and TR can be controlled by the experiment operator. If Equation 1.21 describes the signal in each point of the sample, it is possible to obtain different contrast images of the sample, by varying the operator-dependent parameters. This technique is exploited in clinics to enhance different tissues or lesions both for diagnostic and follow-up purposes.
Neglecting the diffusion term in Equation 1.21, a *spin density-weighted* image can be achieved by minimising the dependence on T_1 and T_2^*, selecting TE much shorter than the shortest T_2^* in the sample and TR much longer than the longest T_1:

$$e^{\frac{-TE}{T_2^*}} \to 1 \Leftrightarrow TE \ll T_2^* \tag{1.22}$$

$$e^{\frac{-TR}{T_1}} \rightarrow 0 \Leftrightarrow TR \gg T_1 \tag{1.23}$$

If both conditions 1.22 and 1.23 are satisfied, then from expression 1.21 one has $S(t) \propto \rho\left({}^1H\right)$ and the corresponding image is mainly based on the differences in proton density in each tissue involved.
To obtained a T_1*-weighted (T_1-w) image*, TE has to be chosen much shorter than the shortest T_2^* in the sample (condition 1.22) and TR has to be on the order of the T_1 values in the sample:

$$TR \simeq T_1 \tag{1.24}$$

In a similar way, a T_2^**-weighted (T_2^*-w) image* can be achieved with TR much longer than the longest T_1 in the sample (condition 1.23) and TE on the order of the T_2^* values in the sample:

$$TE \simeq T_2^* \tag{1.25}$$

An example of different image contrast on the brain is given in Figure 1.9.

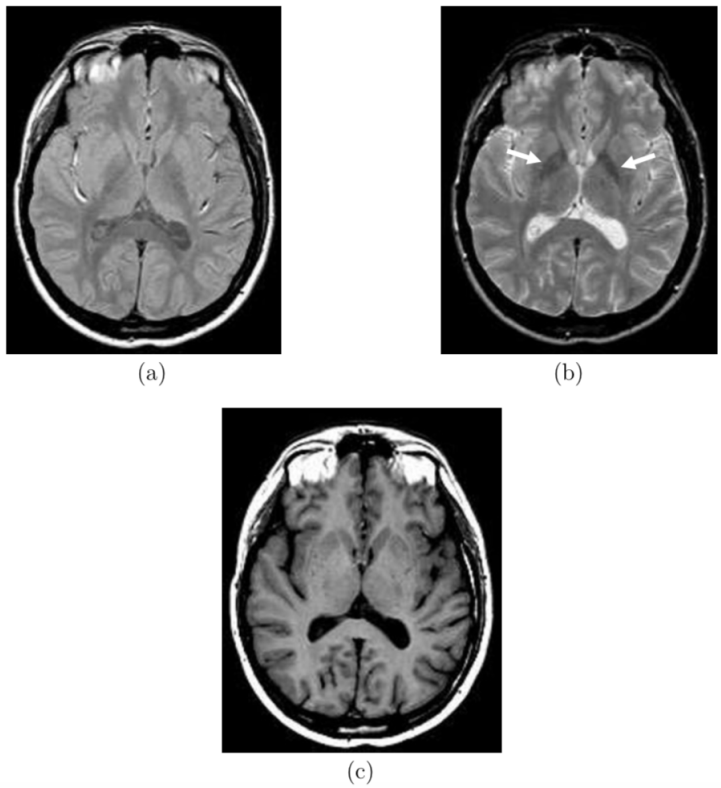


Figure 1.9: (a) Spin density-weighted image, (b) T_2-weighted image and (c) T_1-weighted image of the brain, obtained with a SE sequence with variable parameters. Reprinted from [4], with permission from John Wiley and Sons.

1.2.2 1D imaging Equation and frequency encoding

In the previous Section, the possibility of generating MR images with different weighting has been illustrated. The transition from the NMR signal to an image will be summarised in this Section, pointing out the major physical principles behind it.
The NMR is a volumetric technique, since the detected signal is given by all the resonant nuclei (protons for the purpose of this description) in the sample. In order to generate an image of the sample, it is necessary to distinguish the spatial origin of the signal inside it, i.e. to spatially identify each proton. This task can be achieved by applying magnetic field *gradients* along each direction x, y and z, to make the field linearly variable and, as a consequence, to make the resonance frequency of the protons unique point by point of the sample.
Considering a 1D problem, let $G_z(t)$ be a field gradient applied along the direction of the static field $\vec{H}_0 = H_0\hat{z}$. The resulting field is given by:

$$H_z\left(z,t\right) = H_0 + zG_z(t) \tag{1.26}$$

where

$$G_z(t) \equiv \frac{\partial H_z}{\partial z} \tag{1.27}$$

Correspondingly, the resulting angular frequency of the protons is:

$$\omega\left(z,t\right) \equiv \omega_0 + \omega_G\left(z,t\right) \tag{1.28}$$

with

$$\omega_G\left(z,t\right) = \gamma z G_z(t) \tag{1.29}$$

Equation 1.29 allows to link the precessional frequency of the spins with their positions along the field direction and it is referred to as *frequency encoding*.
The application of the gradient induces a dephasing in the spins frequency. The accumulated phase at time t can be expressed as:

$$\phi_G\left(z,t\right) = -\int_0^t dt'\omega_G\left(z,t'\right) = -\gamma z\int_0^t dt' G_z\left(z,t'\right) = -\gamma z G_z t \tag{1.30}$$

Consequently, the signal $S(t)$ in the reference frame rotating at $\omega = \omega_0$ can be written as:

$$S(t) = \int dz\rho(z)e^{i\phi_G(z,t)} \tag{1.31}$$

where

$$\rho(z) \equiv \int\int dx\, dy\, \rho\left(\vec{r}\right) \tag{1.32}$$

is the effective spin density along z and $\vec{r}$ is the spatial variable.
Let $k(t)$ be the *spatial frequency* variable, defined as:

$$k(t) = \frac{\gamma}{2\pi}\int_0^t dt' G(t') \tag{1.33}$$

The signal in Equation 1.31 can thus be expressed as:

$$S\left(k(t)\right) = \int dz \rho(z) e^{-i2\pi kz} \tag{1.34}$$

Equation 1.34 is referred to as *1D imaging equation* and it contains the basic concept for the images reconstruction in MRI. In other words, when a linear field gradient is applied, the signal $S(k)$ is the Fourier transform (FT) of the spin density of the sample. Being $S(k)$ and $\rho(z)$ two variables linked by the FT, the spin density can be obtained with the inverse FT of 1.34:

$$\rho\left(z\right) = \int dk S(k) e^{+i2\pi kz} \tag{1.35}$$

In conclusion, the MRI technique consists in acquiring a signal $S(k)$ in the domain of the spatial frequencies (which intrinsically depends on time and allows the identification of the spins position in the sample) and taking the FT of this signal to obtain the effective density of spins, i.e. the image.

1.2.3 Slice selection and phase encoding

In order to unequivocally identify the spatial position of each resonant proton in the sample, in an MRI scanner three pairs of gradient coils are placed each in a direction of the set $(\hat{x}, \hat{y}, \hat{z})$. Usually the frequency encoding, introduced in the previous Section, is chosen as the $\hat{x}$ direction and it is also called *readout gradient*.
The gradient along $\hat{z}$ is usually referred to as *slice selection gradient* and allows to uniformly excite a slice of the sample Δz, so that after the pulse all the spins in the slice have the same phase and the precessional frequency $f = \omega/2\pi$ is linear with the position along the $\hat{z}$ axis:

$$f(z) = f_0 + \frac{\gamma}{2\pi} G_z z \tag{1.36}$$

where $f_0 = \frac{\gamma}{2\pi} H_0$ is the Larmor frequency at $z = 0$. An rf pulse has a certain bandwidth in the frequency domain $BW_{rf} \equiv \Delta f$. To excite a slice of thickness $\left(z_0 - \frac{\Delta z}{2}; z_0 + \frac{\Delta z}{2}\right)$, the rf pulse has to be:

$$BW_{rf} = \frac{\gamma}{2\pi} G_z \Delta z \tag{1.37}$$

as shown in Figure 1.10. The profile of the selective rf pulse is a sinc function in the time domain. By taking the FT of this profile, it is possible to show that in the frequency domain the pulse is a boxcar function.
After identifying the slice to be imaged with the gradient along $\hat{z}$, a linear gradient is applied along $\hat{y}$. It is usually referred to as *phase encoding gradient* and introduces a dephasing in the spins according to their position along $\hat{y}$ axis (as seen in Equation 1.30). Finally, a readout gradient along $\hat{x}$ introduces an additional dephasing, allowing the spatial identification of each spin in the matrix. The effect of the phase encoding and the the readout gradients on the spins in a 3x3 matrix is shown in the central picture and in the right picture of Figure 1.11, respectively.

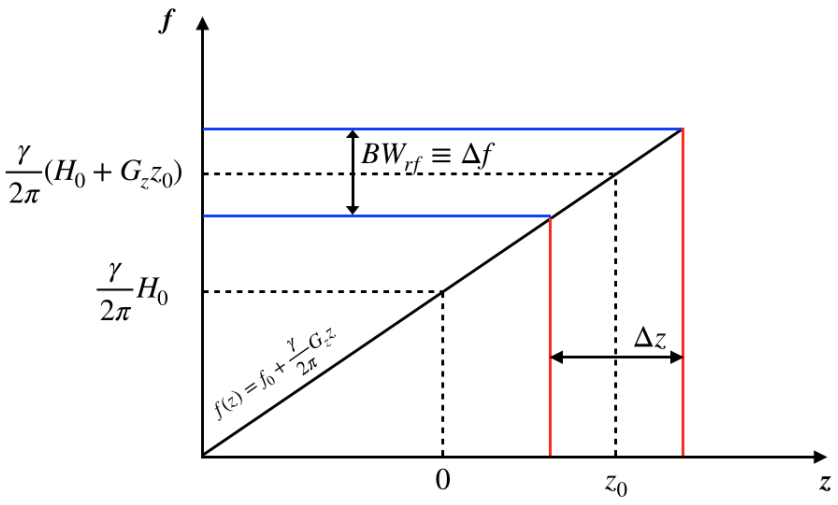


Figure 1.10: Slice selection gradient along the $\hat{z}$ axis. A magnetic field gradient generates a linear relation between the spatial location of the excited slice and the Larmor frequency.

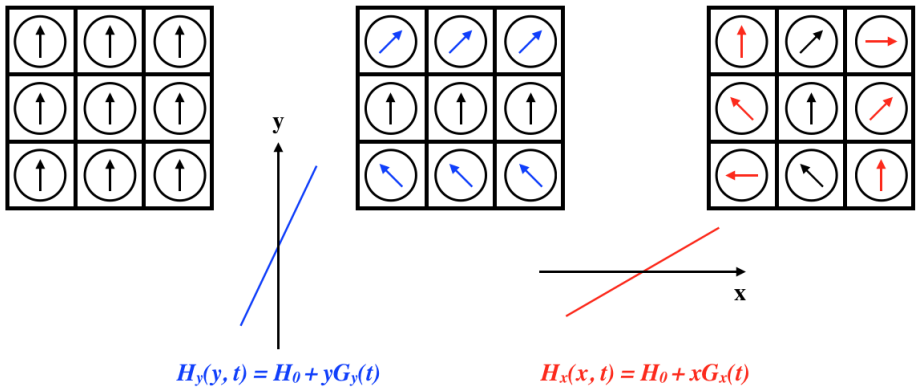


Figure 1.11: The application of the phase encoding gradient G_y (central image) to a matrix 3x3 of spins (left image) introduces a dephasing in the spins precession according to their position along the $\hat{y}$ axis. The application of the frequency encoding gradient G_x (right image) causes an additional dephasing along $\hat{x}$ axis, allowing to spatially identify the position of each spin unequivocally.

CHAPTER 2

State of the art in MRI radiomics

The aim of this Chapter is to offer an overview of the current status of the radiomic analysis in MRI of the pelvis and breast, that represent the main anatomical sites of interest for the present work. In the first part of the Chapter, the basic concepts in radiomics will be introduced, from its rationale and definition to the advantages that this discipline can provide. The workflow of the radiomic analysis will be explained, focusing on its main operational phases. The challenges and open issues of this ongoing technique will be highlighted. In the second part of the Chapter, the most recent results of the radiomic analysis in both pelvic and breast MRI will be discussed.

2.1 Fundamentals of radiomics

2.1.1 Rationale and potentiality

A tumour is a complex and dynamic ecosystem. Its growth involves different spatial and temporal scales, from the molecular to the whole-organ level. When a tumour originates, the non-linear interactions happening at the genomic level result in intra-tumoural heterogeneity affecting its physiology and anatomy. Tumours with higher heterogeneity are thought to be associated with a worse prognosis [9], thus complicating clinical workflows, from diagnosis to treatment selection and response assessment. On this basis, medical imaging plays a key role in the investigation of spatial variations in the tumoural tissue and its surrounding (Figure 2.1).
Specifically, quantitative imaging of the tumour anatomy has been proposed as a means to investigate the underlying gene expression patterns in order to infer information on the tumour functional and progression mechanisms [10].
In this frame, the practice of converting medical images into minable high-dimensional data in order to support clinical decisions has emerged and it is referred to as *radiomics* [11]. To go beyond the visual interpretation made by the physician, each digital image is exploited for its quantitative nature. In fact, a Computed Tomography (CT), a Magnetic Resonance (MR) or a Positron Emission Tomography (PET) image can be read as a matrix of pixels (or voxel for 3D acquisitions), each associated with a number representing a certain grey level intensity. Mathematical formulas applied on these numbers allow the extraction of many parameters, called *radiomic features*, describing the tumour in terms of its intensity, shape, size, volume and texture properties. The radiomic features can be

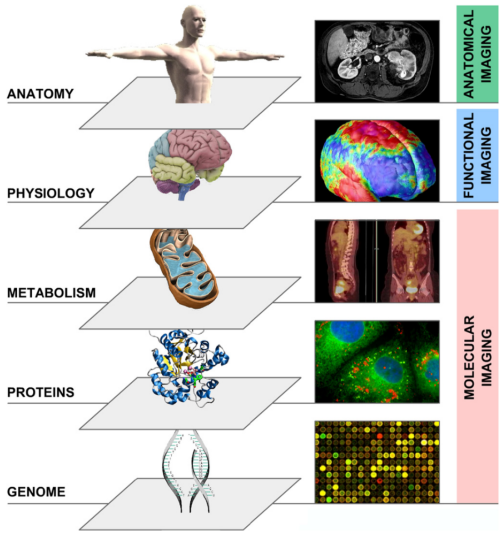


Figure 2.1: Different modalities of medical imaging can investigate biological systems at increasing levels of organisation, form the genome to the anatomy. Reprinted from [9], with permission from Elsevier.

correlated with specific clinical outcomes and can be helpful in early assessment of therapy response, recurrence probability and/or prognosis, patients stratification and choice of treatment strategy. In this sense, the radiomic features can be interpreted as "synthetic biomarkers", that can be integrated with other -omics data, including genomics, transcriptomics, proteomics and metabolomics for a more comprehensive understanding of the pathophysiology of the tumour. The mining of these data, i.e. the procedure of recognising relationships and correlations among them, covers a pivotal role in the most recent innovations in medical imaging analysis, towards the achievement of a predictive, preventive and personalised medicine [12]. Specifically, the correlation between radiomic and genomic data has been called *radiogenomics*. In presence of demonstrated correlation between radiomic and genomic data, the extraction of radiomic features from medical imaging could potentially allow to assess gene expression patterns in a non-invasive way (Figure 2.2).

The very big number of parameters involved puts the radiomics discipline in the framework of the *Big data* analytics, which is becoming more and more popular in many different fields, including healthcare, thanks to the collection of huge amount of data and the availability of fast and efficient computational tools for handling them. Big data have been defined through the five "Vs" [13]: Volume, referring to the considerable size of the datasets required for the investigation; Variety, which refers to the collection of data from different sources; Velocity in the data generation and collection; Veracity, which deals with the quality of the data collected and Value, which refers to the potentiality of the Big data analysis to actively support the decison-making process.

The strength of radiomics lies in its application to standard-of-care medical images, that

are included in the usual diagnostic workflow in oncology and are easily available on the PACS (Picture Archiving and Communication System). Another advantage is that the features extracted from the clinical images are able to describe the whole tumour heterogeneity. As explained, deep knowledge of the tumour spatial organization is essential in all the clinical phases of the pathology assessment and treatment. In this sense, radiomics represents a powerful non-invasive way to inspect the tumour heterogeneity, as a valid alternative to an invasive biopsy. Moreover, it allows a description of the whole tumour instead of a partial one, as offered by biopsy [14].

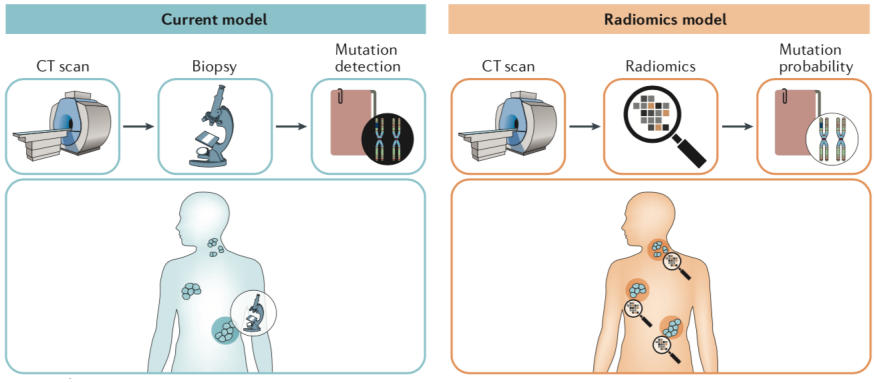


Figure 2.2: Comparison between the bioptic and the radiomic paradigm. The radiomic approach may represent a non-invasive alternative to biopsy for the assessment of the tumour molecular properties, based on the existing link between the macroscopic aspect of the imaged tissue and the underlying gene expression patterns. Reprinted from [14], with permission from Springer Nature.

2.1.2 Workflow

The identification of a clinical scenario of interest for the desired radiomic study, including the choice of an imaging protocol and the definition of a predictive target, precedes the operational phases of the radiomic analysis [14]. After that, the radiomic workflow is composed of four main steps, illustrated in Figure 2.3. In this Section, each phase will be analysed and discussed.

(i) Image acquisition and reconstruction

First of all, high-quality images of the patients are acquired and reconstructed. The typical imaging modalities involved in oncology are CT, PET, MRI and ultrasound, depending on the specific pathology and tumour site. The examination can be performed at different stages of the disease assessment and treatment, including diagnosis, treatment planning and follow-up [15]. Focusing on MRI, it has been less extensively investigated than CT or PET for use in radiomics. The main reason lies in the complexity of MRI, where the numerous variables involved are optimised in light of the clinical needs and equipment capacities, resulting in a wide variety of MRI protocols. Moreover, the lack of a standard intensity scale for MR images necessitates the use of additional pre-

processing, such as intensity normalisation [16]. Eventual inhomogeneity of the magnetic field could represent an additional issue.

(ii) Segmentation

On each image, one or more Volumes Of Interest (VOIs) are selected with a manual contouring by expert physicians or with computer-based segmentation tools, identifying the regions considered to be relevant for the radiomic analysis. The VOI typically includes the whole tumour volume or some of its subregions. Manual segmentation remains the most popular technique, but it is affected by intra- and inter-operator variability and it is very time-consuming when dealing with the large datasets required for a radiomic analysis. On the other hand, although many semi-automatic or automatic software for contouring are available, there is a lack of standardisation in their application [17]. In addition, the majority of the semi-automatic tools requires a manual correction by the operator in complex cases. The automatic packages have been developed for a limited number of districts, mainly the ones with high contrast, and the majority are for research only. These systems are rarely integrated in the radiomic software.

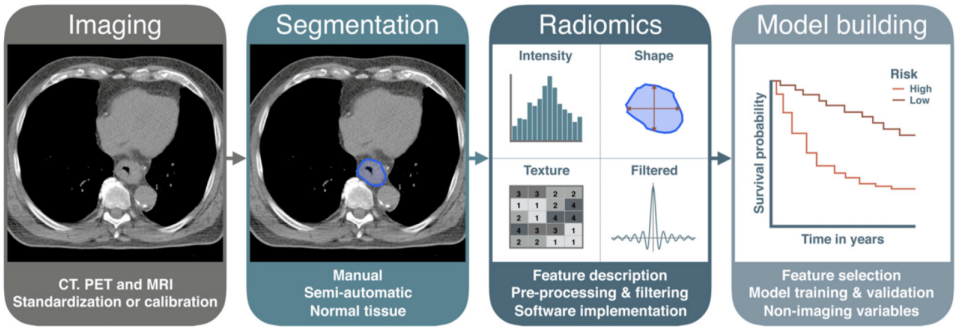


Figure 2.3: The radiomics analysis includes four main phases. (i) The patients undergo a medical imaging examination. (ii) The physician identifies on each image the Volume Of Interest (VOI) relevant for the radiomic analysis. (iii) The radiomic features are extracted from the VOI through a dedicated software. (iv) The radiomic features are correlated with other clinical data of the patients to build predictive models for the treatment outcomes or for prognostic purposes. Figure from [15] [1].

(iii) Feature extraction

Consequently, a very big amount of quantitative parameters are extracted from the selected volumes through a dedicated software. The radiomic features aim at quantifying different properties of the tumour, including the shape, the distribution of voxel values,

[1]The reproduction of one figure from an article in *British Journal of Radiology* is allowed in a PhD thesis providing the original source article is cited. For additional information visit this link.

and the texture, assessing the intensity heterogeneity, the correlation of intensity levels between neighbour pixels and the directionality of intensity distribution. Specifically, the radiomic features can be classified into four main categories, described in detail below: morphological, histogram-based (or first-order statistics), textural (or second-order statistics) and filtering features [18], [19], [20], [21].

- *Morphological features*

  Morphological features characterise the three-dimensional shape and size of the VOI. Features describing to what degree the tumour is spherical, rounded or elongated are included, along with parameters indicating the volume, the surface area and the minimum/maximum diameter.

- *Histogram-based features (first order statistics)*

  The features of this group aim at describing the frequency histogram (or first order histogram) of the intensity values corresponding to the voxels included in the VOI. The histogram is obtained after discretising the original intensity values in the VOI into intensity bins. The features of the first order statistics do not provide information regarding the relative position of the intensity levels over the VOI. Typical features of this category describe the shape of this histogram and include mean, median and minimum/maximum value, skewness (asymmetry), kurtosis (flatness), range and percentiles. Histogram-based features with a less intuitive meaning include: energy, measuring the magnitude of voxel values in the VOI; entropy, evaluating the randomness of the intensity values, and uniformity, describing the homogeneity of the intensity distribution.

- *Textural features (higher order statistics)*

  The higher order analyses give information about the spatial inter-voxel relationships, including textural features, which identify specific patterns in the intensity distribution. Textural features are extracted with mathematical formulas from specific matrixes, calculated from the original image considering only the voxels included in the VOI. The main matrixes considered in the present work are described below.

  Grey Level Co-Occurence Matrix

  Let N_g be the total number of grey levels in which the VOI intensity values have been discretised, $i = 1, ..., N_g$ and $j = 1, ..., N_g$. The index i will appear with the same meaning also in the definition of the other matrixes below. The Grey Level Co-Occurence Matrix (GLCM) is a N_g x N_g matrix in which the $(i, j)^{th}$ element represents the number of times the combination of levels i and j occurs in two voxels at Chebyshev distance [2] δ along a predefined direction θ in the VOI. Figure 2.4

[2]The Chebyshev distance between two points x and y, with coordinates x_i and y_i respectively, is defined

shows an example of the calculation of a GLCM for $\delta = 1$, i.e. considering adjacent voxels, and $\theta = 0°$, i.e. reading the rows of the image starting from the top left corner and going in horizontal direction. The image given in the example will be also used as a reference image for the calculation of the following matrixes.

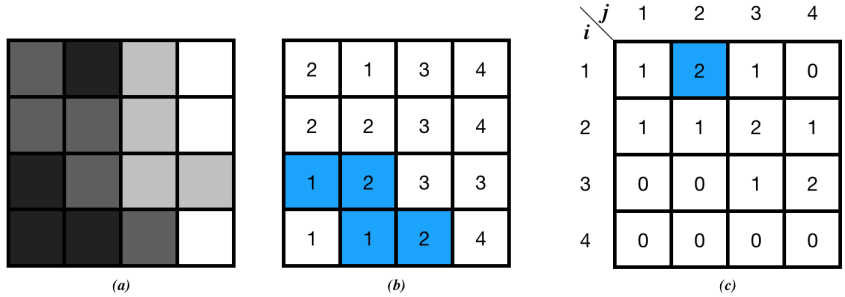


Figure 2.4: GLCM calculation. (a) 4 x 4 pixel given image. (b) The image discretised into four grey levels. (c) GLCM of the given image ($\delta = 1$; $\theta = 0°$). The occurrences of the couple (i, j) of adjacent voxels is counted and allocated in the co-occurrence matrix. The blue element of the GLCM indicates the number of times the combination $(i, j) = (1, 2)$ appears in the discretised image, along horizontal direction.

Grey Level Run Length Matrix

The Grey Level Run Length Matrix (GLRLM) quantifies the run lengths in the VOI intensity mask. A *run* is defined as a sequence of consecutive voxels along a certain direction, all with the same grey level. Let N_r be the maximum length of a run observed in the image, $r = 1, ..., N_r$ and $i = 1, ..., N_g$. The GLRLM is a N_g x N_r matrix in which the $(i, r)^{th}$ element represents the number of times the run of grey level i and length r occurs in the image along a direction θ. Figure 2.5 shows the GLRLM for $\theta = 0°$ obtained from the 2D image considered as reference example.

Grey Level Size Zone Matrix

In 2D, a pixel is defined as *connected* to one of its eight first neighbour pixels if they show the same grey level. Similarly, in 3D it is considered a 26-connectedness. A *zone* is defined as a set of connected pixels or voxels. The Grey Level Size Zone Matrix (GLSZM) describes the number of zones in the VOI. Let N_z be the maximum zone size of any group of connected voxels observed in the image, $z = 1, ..., N_z$ and $i = 1, ..., N_g$. The GLSZM is a N_g x N_z matrix in which the $(i, z)^{th}$ element represents the number of times the zone of grey level i and size z occurs in the image. Figure 2.6 shows the GLSZM obtained from the 2D reference image.

as $\delta(x, y) := \max_i (|x_i - y_i|)$.

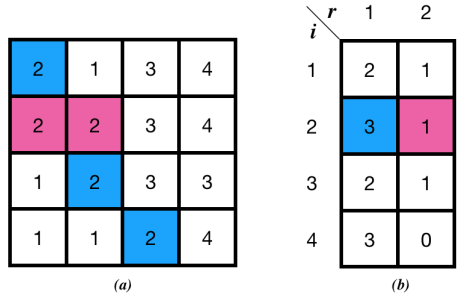


Figure 2.5: GLRLM calculation. (a) The reference image discretised into four grey levels. (b) GLRLM of the given image ($\theta = 0°$). The blue cell of the matrix indicates the number of runs with grey level 1 and run length 1 (single pixels). The pink cell gives the number of runs with grey level 1 and run length 2 (two adjacent pixels of grey level 2).

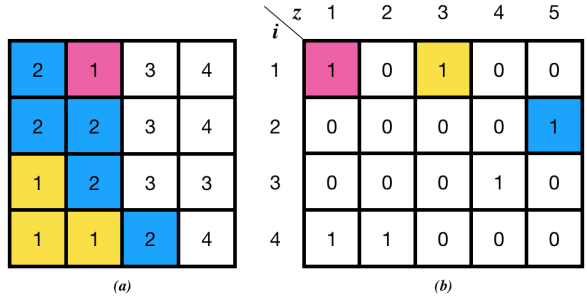


Figure 2.6: GLSZM calculation. (a) The reference image discretised into four grey levels. (b) GLSZM of the given image. The pink cell of the matrix indicates the number of zones of the image with grey level 1 and zone size 1 (a pixel with grey level 1 with no connected pixels with grey level 1). The yellow cell gives the number of zones with grey level 1 and zone size 3 (three connected pixels with grey level 1). The blue cell gives the number of zones with grey level 2 and zone size 5 (five connected pixels with grey level 2).

Neighbourhood Grey Tone Difference Matrix

Let $A_{d,v}$ be the discretised grey level of a voxel at position $\vec{v} = (v_x, v_y, v_z)$ in the VOI. The average grey level within a neighbourhood centred at (v_x, v_y, v_z) and within a Chebyshev distance δ is

$$\bar{A}_v = \frac{1}{W} \sum_{m_z=-\delta}^{\delta} \sum_{m_y=-\delta}^{\delta} \sum_{m_x=-\delta}^{\delta} A_d \left(v_x + m_x, v_y + m_y, v_z + m_z\right) \tag{2.1}$$

with $(m_x, m_y, m_z) \neq (0, 0, 0)$ to exclude the central voxel of reference and W is the size of the neighbourhood and it is defined as $W = (2\delta + 1)^3 - 1$ in 3D and $W = (2\delta + 1)^2 - 1$ in 2D (i.e. for $\delta = 1$ in a 2D image a pixel is surrounded by 8 pixels).
The element s_i in the Neighbourhood Grey Tone Difference Matrix (NGTDM) in-

dicates the sum of grey level differences of voxels with grey level $i = 1, ..., N_g$ and the average grey level $\bar{A}_v$ of neighbouring voxels within a distance δ:

$$s_i = \sum_{k}^{N} \left| i - \bar{A}_v \right| \tag{2.2}$$

with the conditions that the grey level $A_{d,z}$ of voxel v is equal to i and the voxel has a valid neighbourhood (i.e. for $\delta = 1$ it is surrounded by 26 voxels in 3D; in 2D and for $\delta = 1$, a pixel has a valid neighbourhood when it is surrounded by 8 pixels). In Equation 2.2, N is the number of voxels in the VOI. An example of the procedure to obtain the main quantities involved in calculating the NGTDM-based features is shown in Figure 2.7. In this example, the voxels with a valid neighbourhood are the four central pixels, of grey levels 2 and 3. Thus the s_1 and s_4 entries of the NGTDM are zero. To calculate s_2, the two pixels with grey level 2 has to be considered separately to calculate the average grey value of their surrounding pixels. For example, the 2 in the blue cell has an average value of the neighbourhood given by Equation 2.1: (2+1+3+3+3+2+1+2)/8 = 17/8. A similar calculation for the pink 2 gives 15/8. Thus, from Equation 2.2, $s_2 = |2 - 17/8| + |2 - 15/8| = 0.250$. In the same way, $s_3 = |3 - 22/8| + |3 - 21/8| = 0.625$. The definition of the NGTDM

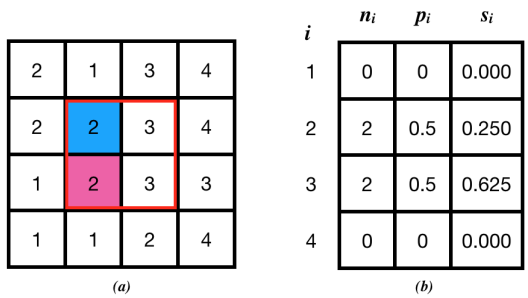


Figure 2.7: NGTDM calculation. (a) The reference image discretised into four grey levels. The N_n pixels with a valid neighbourhood (and thus considered for the matrix calculation) are located within the red box. (b) Parameters for the NGTDM calculation from the given image. n_i is the number of voxels of grey level i and with a valid neighbourhood. $p_i = n_i/N_n$ is the grey level probability. s_i is a matrix element of the NGTDM.

in the literature was given originally by Amadasun and King for 2D image [22]. The identification of the pixels with a valid neighbourhood in order to build the NGTDM causes some limitations for the radiomic analysis, since the voxels on the edge of the VOI mask have to be disregarded. To include them in the analysis, a valid neighbourhood is considered to exist if there is at least one neighbouring voxel included in the VOI mask [18]. This deviation from the original matrix definition holds also for the Grey Level Dependence Matrix.

Grey Level Dependence Matrix

The Grey Level Dependence Matrix (GLDM) quantifies the coarseness of the voxel intensities within the VOI. As for the NGTDM matrix, the concept of neighbourhood (in its radiomic acception, as introduced in the previous matrix) of a pixel or voxel is involved, along with the concept of dependency. The grey levels of a central voxel v at position $\vec{v}$ and a neighbouring voxel w at position $\vec{v}+\vec{w}$ are defined as *dependent* if $|A_d(\vec{v}) - A_d(\vec{v}+\vec{w})| \leq \alpha$, where $\alpha \in \mathbb{Z}$ and $\alpha \geq 0$. Let N_k be the maximum dependence observed in VOI intensity mask and $i = 1, ..., N_g$. The GLDM is a N_g x N_k matrix in which the $(i,k)^{th}$ element represents the number of neighbourhoods with a center voxel of grey level i and a neighbouring voxel dependence k, for a certain value of α and fixed Chebyshev distance δ. Figure 2.8 shows the GLDM obtained from the 2D image considered as reference example, for $\delta = 1$ and $\alpha = 0$. The condition $\delta = 1$ means to consider, in a 2D image, adjacent voxels in all possible directions. The condition $\alpha = 0$ means that two voxels are dependent if they have the same grey level.

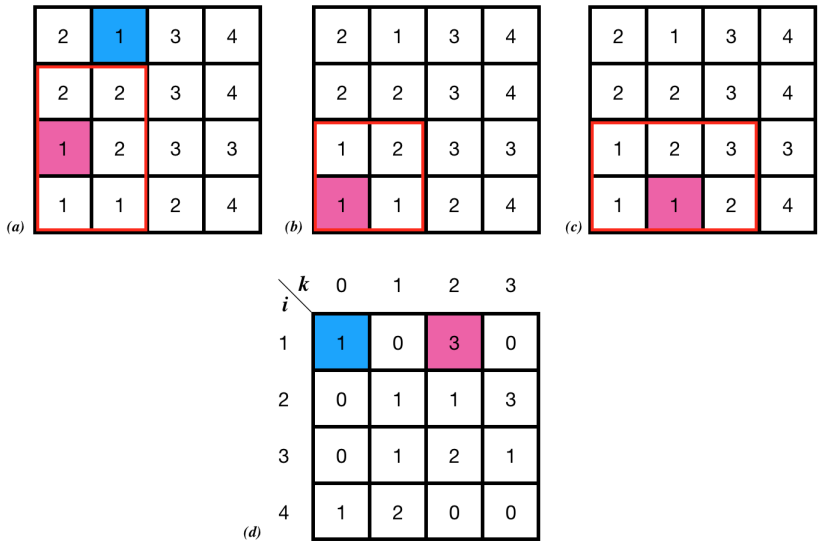


Figure 2.8: GLDM calculation. (a) - (c) The reference image discretised into four grey levels. (d) GLDM of the given image for $\delta = 1$ and $\alpha = 0$. The blue cell of the matrix indicates the number of times a pixel with grey level 1 appears in the image surrounded by no pixels of the same intensity (dependence 0, as highligthed in blue in image (a)). The pink cell indicates how many times in the image a pixel with grey level 1 has in its neighbourhood (represented by the red box) 2 pixels of grey level 1 (dependence 2). Each of the 3 pixels of the image contributing to the count in the pink cell of the matrix is shown respectively in (a), (b) and (c).

- *Filtering approaches*

The filtering features are extracted from images obtained by applying a transformation on the original images. The main filters applied in this study are listed below (and described in detail in the PyRadiomics documentation and in [23]).

Exponential: when applying this filter, the intensities in the resulting image are the exponential of the original intensities.

Laplacian of Gaussian: the filtered image is obtained by convolving the original image with the second derivative (Laplacian) of a Gaussian kernel, enhancing the areas of grey level change (e.g. edges of a defined region) in the original image.

Local Binary Pattern: applying this transformation, a reference pixel/voxel is compared with its neighbours. If an adjacent pixel has a grey level greater than the reference pixel, it is assigned the value 1 and it is considered 0 otherwise. Once all the neighbouring pixels have been assigned a value (0 or 1), a binary number is obtained and assigned to the reference pixel. For example, the reference pixel in a 2D Local Binary Pattern image is represented by a binary number with 8 digits, with each digit indicating one of the 8 surrounding pixels.

Logarithm: when applying this filter, the intensities in the resulting image are the logarithm of the absolute original intensities + 1.

Square: the intensities in the resulting image are the square of the original intensities.

Square Root: the intensities in the resulting image are the square root of the original intensities.

Wavelet: Wavelet transform decomposes the original image in low and high frequency images, which can be separately analysed. Such decomposition, schematised in Figure 2.9, allows the decoupling of textural properties of the input image.

After the application of the Exponential, Logarithm, Square and Square Root filters, the intensity range is rescaled to match the original image range and negative original values are stored as negative also in the filtered image.
Many dedicated software packages, both open-source and commercial, are available for image handling and features extraction, differing for the number of features extracted (ranging from a few tens to more than one thousand), and the possibility to perform one or more image pre-processing. The main tools used in this work are IBEX [24], LIFEx [25] and PyRadiomics [26], all freely available. One of the goal of this thesis is to compare the algorithms implementation across these radiomics software. This analysis will be covered in Chapter 3, where a detailed description of each software will be included.

***(iv)* Model building**

After the extraction, a selection of the most informative features is performed to exclude redundant features and to avoid overfitting, by reducing the problem dimensionality.

[3]The reproduction of the image is allowed from the license available at this link.

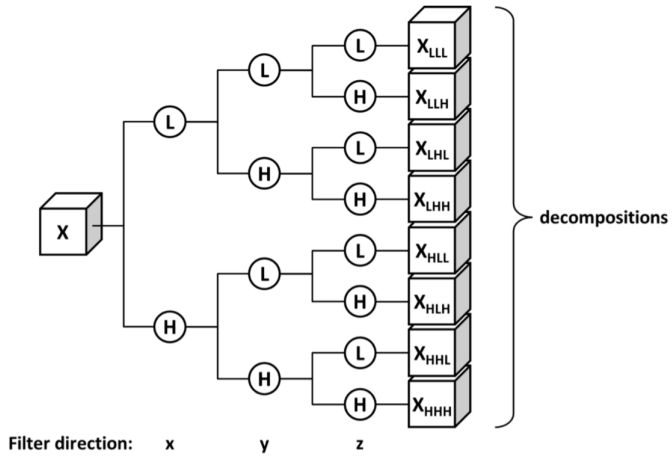


Figure 2.9: Schematic of 3D level Wavelet decomposition. By directional low-pass (L) and high-pass (H) filtering, the original image X is replicated into 8 decompositions. Figure from [19], Supplementary Material.[3]

An accepted rule is to have at least ten patients/subjects for each feature used in the model [27]. The set of the identified features is integrated in a database, along with the results of clinical and/or genomic analyses. These data are analysed, with the final goal of building predictive models for the clinical outcome of interest, for example the response to a specific treatment or prognosis among others. The model building can be performed with various techniques, depending on the nature of the initial clinical question to investigate, on the quality of the available data and on the size of the available sample. If the outcome to be predicted is a continuous variable, regression methods, including Linear, Cox and Lasso, can be exploited. If the target is a discrete variable, classification methods, including Support Vector Machines and Random Forest, can be used [28]. Both traditional statistical analysis and machine learning methods can be exploited to carry out the analysis. Correction methods for the false discovery rate are commonly used.

One of the requirement for a robust model lies in a proper data management. The data should be divided into three groups: the training, the validation and the testing sets. The training set is used to train the model; the validation set is dedicated to the optimisation of the model parameters; the testing set is used to assess the performance of the optimised model on different data, involved neither in the training nor in the validation procedure [29]. According to the TRIPOD guidelines [30], external data are preferable to internal ones for the testing investigation.

2.1.3 Challenges and open issues

Early clinical demonstrations of radiomics potential in MRI have been seen for various pathologies and anatomical districts. An insight into the main results obtained for pelvis and breast studies will be provided in the next Sections. In the literature, the amount of publications associated with the keyword 'radiomics' has increased exponentially in the last years, with a great variability of the procedures followed in the radiomic workflow. This variability makes comparison between studies challenging and hinders generalisation of results between centers [15], [31]. Thus, before bringing radiomics into daily clinical practice, a process of methods standardisation, benchmarking, and validation is essential [32].
The main challenges that radiomics is currently facing are summarised in [33] and are related to each phase of the radiomic approach, from the image acquisition to the study reporting. In this Section, the main open issues in each step of the radiomic analysis are illustrated.

The **image acquisition** protocols, especially in MRI, vary among centres and are optimised on each scanner, according to the clinical question under investigation. To understand the sensitivity of radiomic models to differences between scanners and scanning protocols, a comprehensive investigation on the influence of different image acquisition settings (involving variations in MR sequences parameters, scanner, field strength, vendor and/or centre) on radiomic features is warranted. Establishing a well-defined, standard procedure for supporting radiomic analyses is also advisable in view of prospective clinical studies [34]. More in detail, repeatability and reproducibility studies of the radiomic features needs to be performed for a more complete and deep understanding of the features trend in different settings. *Repeatability* is "the magnitude of measurement error under a set of repeatable conditions", including the same experimental setup and procedure, same system and operator, "over a reasonably short interval" [35]. The repeatability of the radiomic features is evaluated through a test-retest analysis, a comparison between the features value extracted from images acquired in the same conditions within a short time (in this work typically 10-15 minutes). *Reproducibility* demonstrates the ability of a quantitative imaging biomarker "to obtain the same measurement when made on the same experimental unit under different experimental conditions" [35]. For example, a reproducibility experiment can be performed extracting the features from images acquired on the same scanner with the same scanning protocol, but varying one of the imaging parameter in a certain range, during repeated acquisitions. Accepted metrics to evaluate the repeatability and the reproducibility measures include the Intraclass Correlation Coefficient (ICC) and the Concordance Correlation Coefficient (CCC), respectively [36]. The repeatability and reproducibility of radiomic features have been extensively investigated during this thesis project and will be covered in Chapter 4.

The radiomic features can also be influenced by the **segmentation** method [37], underlying the need of a study on the impact of intra- and inter-observer variability (when using manual and semi-automatic segmentation) on the radiomic features stability.

Another source of influence on the radiomic models robustness is represented by the choice of a specific radiomics software for the **image processing** and **features extraction**. Before extracting the features, the MR images necessitate the normalisation and discretisation of their intensities. Both these techniques are far from being universal, with plenty of possible solutions to be applied.
Some of the radiomics software offer options for the *normalisation* process. Otherwise, the user has to implement an in-house code or to rely on other specific tools for normalisation. This is crucial for the radiomic features stability, as it has been shown that the choice of image intensity normalisation technique has a strong impact on the reproducibility of radiomic features, evaluated on T_2-weighted MR images of the pelvic region [38].
The main intensity *discretisation* techniques are the fixed bin-count and the fixed bin-width. These methods allow to establish the number of bins or the bin-width of each bin for the discretisation of voxel intensity, respectively. In the literature, there is not a consensus about the optimum method to choose. In addition, it has been demonstrated that the choice of a certain discretisation method can influence the radiomic features reproducibility [39].
Since the beginning of radiomics, the availability of many *software* for the features extraction - many of which are in-house codes developed on purpose - has made the comparison and generalisation of results challenging. In fact, each software follows its own libraries for features definitions and formulas and there has not been for long a shared list of accepted and well-defined features of reference. This issue has been partially solved by the efforts of the Imaging Biomarker Standardisation Initiative (IBSI) [18], which has the goal to provide the features nomenclature and their mathematical definitions, along with benchmark data sets.

Regarding the **model building** phase, one of the main challenges is the size of the sample used in a study. Many radiomics papers involved few patients with respect to the number of features on which the model is based. A limited dataset decreases the predictive power of a model and increases the risk of overfitting of the data. Another crucial point is the necessity of a more extensive validation of the models, possibly through external data, as previously discussed. This task can be reached also by promoting data sharing and accessibility towards a multi-disciplinary collaboration [40], [41]. To quantitatively evaluate the robustness and reliability of a radiomic study, the Radiomics Quality Score (RQS) has been proposed [14]. The RQS is composed of sixteen criteria that, if fulfilled by the radiomic study under evaluation, allow to assign a number of points to the study. The higher the score, the higher the quality of the study. The RQS can help in the standardisation of the reporting guidelines for radiomic studies; an example of its application during the different phases of the radiomic analysis is shown in Figure 2.10.

As pointed out by Lambin et al. [14], some of the illustrated open issues can be fulfilled through *phantom studies*. The phantoms provide many advantages, including a controlled experimental setting and the possibility for repeated measurements. In addi-

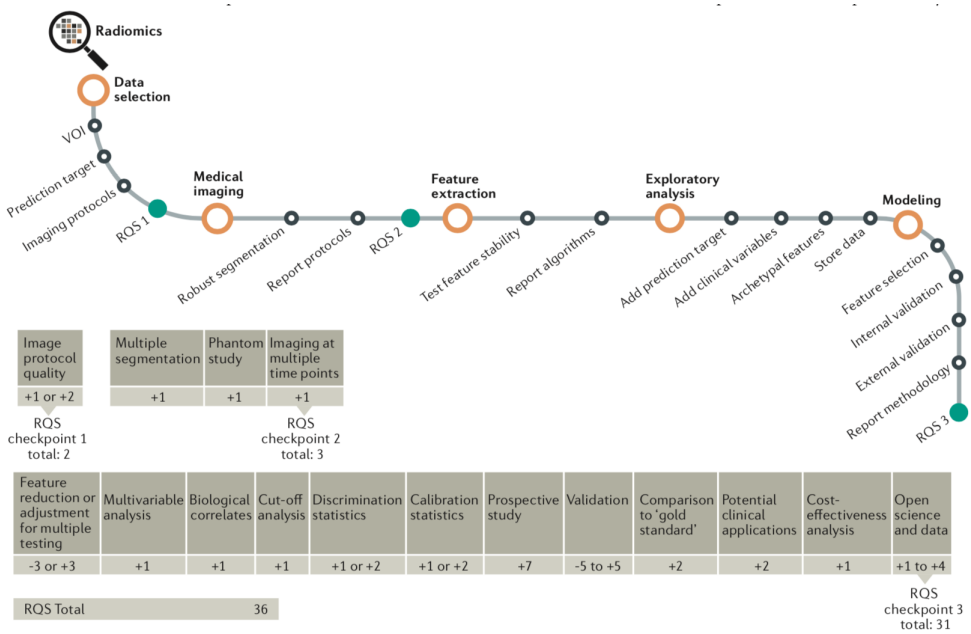


Figure 2.10: Application of the Radiomics Quality Score (RQS) to evaluate the quality of a radiomic study. The RQS consists in a set of criteria to be met by a radiomic paper. The higher the number of fulfilled criteria, the higher the points assigned to the study. Reprinted from [14], with permission from Springer Nature.

tion, the use of a phantom, in which biological processes are not present (rectal/bladder filling, peristalsis, breathing, tissue diffusion and perfusion), allows the assessment of the reproducibility issues caused by different imaging scenarios (e.g. comparing two different scanner types or vendors). The methodological studies performed on phantoms can allow the identification of repeatable and reproducible features and can serve as an initial feature selection method that ensures the development of reliable biomarkers and models.

The phantoms currently available for quality control testing of MR scanners, however, are not suitable for radiomic investigation. In fact, they tend to lack the heterogeneity that serves to differentiate radiomic features between tissues. Phantoms designed for routine quality assurance usually consist of a container filled with homogeneous aqueous solutions in which are embedded objects of specific shape and variable size (commonly plastic homogeneous inserts) to assess individual aspects of scanner performance in terms of geometry, distortion, spatial resolution, contrast and contrast resolution [42]. In particular situations, other solutions have been investigated, including jelly-like materials to provide tissue-equivalent phantoms [43], phantom created with fruits to simulate different texture [44] and anthropomorphic and multimodal phantoms that approximate *in vivo* anatomy and imaging appearance [45]. Some efforts have also pursued digital phantoms, mainly for brain applications. Yang, Ford et al. [46], [47] made use of a digital MR phantom to confirm that the radiomic features are influenced by the image

acquisition process and reconstruction algorithm. They also underlined the need for a physical phantom for experimental validation of their preliminary results. Although the available MRI phantoms are increasingly specialized across a widening range of applications, to date there is still none optimised for radiomic purposes. Physical phantoms specifically dedicated to the evaluation of radiomic features reliability and robustness in CT [48] and PET [49] have recently been reported. An overview of the existing designs is given in [50]. At the beginning of this thesis work, a physical MRI phantom dedicated to radiomic analysis had not been designed yet. To meet this demand, two MRI phantoms (simulating pelvic and breast cancer respectively, for radiomic purposes) have been designed for methodological studies to support clinical decisions. A detailed description of the phantoms realisation will be given in Chapters 4 and 5.

2.2 MRI-derived radiomics in pelvic oncology

This Section has the aim to provide an overview of the main MR radiomic studies on pelvic cancer in the last years. The literature search was conducted in two ways: firstly, focusing on the clinical studies; secondly, pointing out methodological studies. The Section structure, being divided into two parts, follows the literature search.

2.2.1 Clinical studies

The clinical applications of MRI-based radiomic models in pelvic oncology have been summarised in some recent reviews, and can be classified according to the cancer type and location.

- *Prostate cancer*

  The radiomic features in prostate cancer (PCa) are extracted from T_2-weighted, diffusion-weigthed and dynamic contrast-enhanced images, included in the usual clinical protocols for PCa diagnosis [51]. The radiomic models have been investigated at the histopatological level, for their application to tumour characterisation. The association with tumour biology (e.g. cancer aggressiveness or classification) has also been tested and compared with the traditional Gleason score system [52]. Other studies used radiomics for diagnostic and prognostic purposes, and to predict the response to a certain treatment. These applications include the prediction of extracapsular extension and the correlation with biochemical recurrence after different treatments (e.g. radiotherapy) [29]. The first radiogenomic investigations have shown evidence of an association between the radiomic features extracted from the tumour volume and the correlated genetic variations.

- *Gynaecological cancer*

  Promising results have been seen in gynaecological malignancies, including cervical, endometrial and ovarian cancer. Preoperative radiomic features were seen to

be predictive of lymph-vascular space invasion [53] and of the presence of lymph node metastasis [54] in cervical cancer patients. The potential role of radiomics in the prediction of recurrence in advanced cervical cancer patients treated with concurrent chemoradiotherapy has also been evidenced [55].

- *Colorectal cancer*

  MRI-based radiomics of patients with rectal cancer has shown potential both to predict the tumour response to therapy (including neoadjuvant chemoradiotherapy) [56], [57], [29], [58], [59], [60], and to non-invasively evaluate tumour differentiation and other biological properties [61]. An interesting approach by Cusumano et al. [62] consisted in the identification of tumour subpopulations in rectal cancer patients and their characterisation through fractal analysis, with the aim to predict the pathological response after therapy. Jalil et al. [63] used texture analysis on T_2-weighted images to predict long-term survival of patients affected by advanced rectal cancer and treated with chemoradiotherapy.

- *Bladder cancer*

  MRI-based radiomics in bladder cancer has been investigated for tumour classification and grading, prognostic purposes or to predict the response to a selected therapy [64]. Some radiomic features were found to be correlated with the tumour biology and aggressiveness and were useful for the tumour detection, by differentiating the normal bladder wall from cancerous tissue. Radiogenomic studies showed that the radiomic features can be correlated with some genetic mutations correlated with bladder cancer [65].

2.2.2 Methodological studies

Gourtsoyianni et al. [66] studied the repeatability of MRI features extracted from primary rectal cancer. They found that the repeatability was higher for first-order and fractal features than for second and other higher-order statistical features. Fiset et al. [67] studied the repeatability and reproducibility of MRI-based radiomic features on cervical cancer. The reproducibility was assessed in two settings: (i) between diagnostic MRI and simulation MRI; (ii) among three operators who performed the tumour segmentation. The (ii) setting offered the highest number of reproducible features. Both setting (i) and test-retest studies showed low performance in terms of features reproducibility and repeatability, respectively. Only 5.6% of features showed an excellent repeatability and reproducibility (ICC$\geq$0.9) in all the three scenarios. More recently, Schwier and colleagues [68] investigated the repeatability of radiomic features on small prostate tumors, assessing various normalisation schemes, image pre-filtering, and bin-widths for image

discretisation. Despite this thorough inspection, they did not find consistent improvements in repeatability across the different approaches.
Several clinical studies were conducted with variable acquisition parameters, known to affect the tissue contrast, while the influence of such changes in radiomic features was investigated to some extent by Mayerhoefer et al. [69] and Chirra et al. [70]. In the first study, the authors investigated the sensitivity of texture features to acquisition parameters (including TE ranging from 20 to 125 ms and TR in the range 900-4500 ms) at different spatial resolution on T_2-weighted images acquired on a 3 T scanner. They found that the imaging parameters influenced the values of features, with this influence increasing with spatial resolution. In the second study, prostate T_2-weighted images of 147 patients from four different sites were used to assess cross-site reproducibility by performing multivariate cross-validation and assessing preparation-induced instability. The results showed that most of the Haralick features were reproducible in over 99% of all cross-site comparisons.

2.3 MRI-derived radiomics in breast oncology

Following an approach similar to the previous Section, representative studies in the recent literature about MR-based radiomics in breast cancer are presented here.

2.3.1 Clinical studies

Some reviews have systematically collected the main findings in MRI-based radiomics of breast cancer (BC) in the last decade. Valdora et al. [71] reviewed the studies of radiomics in BC, assigning to each article the RQS to evaluate their quality. Codari et. al [72] and, more recently, Sheth et al. [73] summarised the applications of artificial intelligence (including some radiomics methodologies) for breast MRI. Chitalia et al. [74] explained the use of texture analysis in breast MRI and Reig et al. [75] presented machine learning applications in breast MRI. Following the approach by Reig and colleagues, the MRI-based radiomic studies on BC can be classified according to the initial clinical question under investigation, reporting some illustrative results.

- *Lesion classification (benign vs. malignant)*

  Radiomic features were tested with promising results for discriminating malignant from benign lesions [76]. Bickelhaupt et al. [77] studied this aspect on images acquired with an unenhanced diffusion-weighted protocol, finding a good discriminating power. In [78], the radiomic features were investigated for lesion classification and for their association with breast tissue biology.

- *Lymph node status and tumour aggressiveness*

  Liu et al. [79] studied the prediction of sentinel lymph node metastasis in BC with radiomic features extracted from the primary tumour region on the dynamic

contrast-enhanced images. Dong and colleagues showed that textural features extracted from anatomical and functional MR images improved the prediction of the presence of sentinel lymph node metastasis [80]. The tumour aggressiveness is also associated to Ki-67 status. The correlation between radiomic features and Ki-67 expression has been investigated in various studies, like [81].

- *Predicting prognosis and recurrence*

  The role of radiomics in predicting the risk of BC recurrence was studied, for example, in [82]. Dietzel et al. [83] found that radiomic analysis of breast cancer vascularization improved the survival prediction in primary breast cancer. In another study, Koh et al. [84] evaluated the radiomic features as prognostic parameters to predict recurrence in triple-negative BC.

- *Radiogenomics*

  The radiomic features extracted from MR images of BC patients showed good performance in discriminating BC subtypes [85], [86]. The discriminating power could be improved including features extracted from the surrounding parenchyma, as shown in [87]. A study on the The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA) datasets pointed out a correlation between radiogenomic features and tumour staging, lymph node metastasis and status of different receptors involved in the pathology [88]. The same datasets were used to extract multiomics data to study the association between quantitative MRI tumour characteristics and underlying molecular properties (e.g. gene expression patterns) [89]. Figure 2.11 shows the main correlations found by the authors. Radiogenomic features were also studied on dynamic contrast enhanced MRI of BC patients [90].

- *Response to therapy*

  Reig et al. [91] reviewed the role of MRI to evaluate the response to neoadjuvant therapy in BC, including radiomics and radiogenomics studied. In their study, Braman et al. [92] showed that combining intratumoural and peritumoural radiomics on dynamic contrast-enhanced images it was possible to predict the response to neoadjuvant chemotherapy.

2.3.2 Methodological studies

Few studies investigated the robustness of radiomic features in BC and they were mainly on mammography. Robinson et al. [93], for example, focused on the reproducibility of

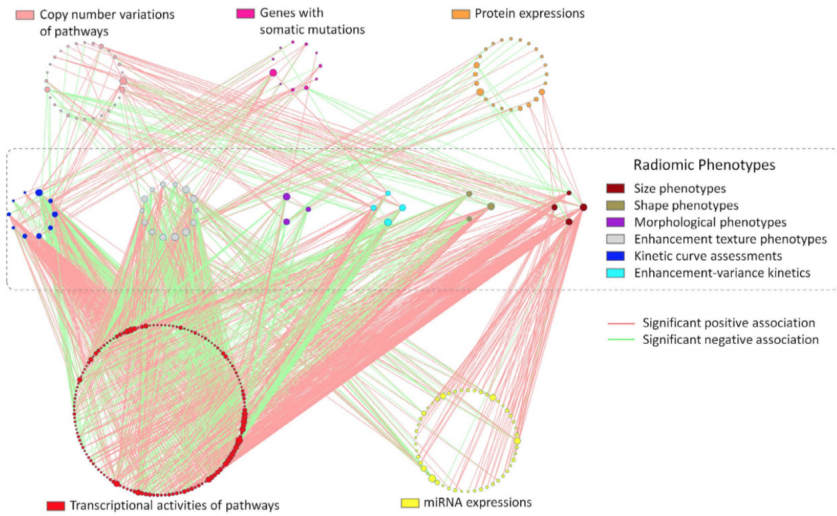


Figure 2.11: Nodes schematic showing the associations between MRI-based radiomic phenotypes of 91 breast invasive carcinomas and their corresponding molecular profiles. The node size indicates the amount of connections of a node with respect to other nodes in the category. Figure from [89]. [4]

features considering mammograms acquired on two systems of different vendors. Regarding MRI, Duron et al. [39] studied the reproducibility of MRI-based radiomic features when extracted from 30 breast lesions, varying the grey-level discretisation method and evaluating the inter-observer variability. Saha and colleagues [94] investigated the impact of the scanner vendor, the magnetic field strength and the slice thickness on both breast lesions and fibroglandular tissue. The research was conducted on MR images of 272 BC patients and showed that 38.9% to 58.2% of the features were highly affected by the choice of one of the three variables considered. In addition, they showed that the features extracted from the fibroglandular tissue seemed to be the most sensitive to the scanner parameters.

CHAPTER 3

Comparison of radiomics software

Since the raising of the radiomics discipline, many software packages have been developed for the radiomic features extraction and handling. The available tools often differ on the mathematical definition and the computation of the features, suggesting the need for a comparison towards a standardisation process. Recently, the Image Biomarker Standardisation Initiative (IBSI) [18] has started to address this issue, providing a list of reference features (both regarding the nomenclature and formulas) and benchmark datasets. In this framework, one of this thesis objectives was to compare the features computed with three freely available radiomics software on the same set of MR images, also in light of the IBSI recommendations. The results of this investigation are reported in this Chapter.

3.1 Software packages

In this Section, the main properties of the three tools used for the radiomic features extraction will be illustrated.

3.1.1 IBEX

IBEX (Imaging Biomarker EXplorer) [24] was developed in MATLAB and c/c++ environments and allows both DICOM and Pinnacle image format as input. It runs on Windows only and it is available in two versions: one stand-alone (version 1.0β, used in this work) and another open-source, both freely downloadable.
The opening screen of *IBEX* is shown in Figure 3.1. The workflow starts with importing the patients' images and visualising them in the dataset tool, where it is also possible to add, draw and modify VOIs on the images. The next step is the creation of a feature set, selecting from an available list the radiomic features to be extracted from the VOIs. In the feature set window, a preprocess module (e.g. a filter) can be added and the specific parameters for each feature class can be tuned starting from the default settings. Examples of these parameters are the number of bins or of grey levels for the histogram calculation or the directions for the GLCM and the GLRLM matrixes computation. Some settings relate to all the features in a certain category, others are specific for a single feature. Finally, in the results section the feature set and the dataset can be selected for the features extraction and the results saving. The output is an XLS file with each row corresponding

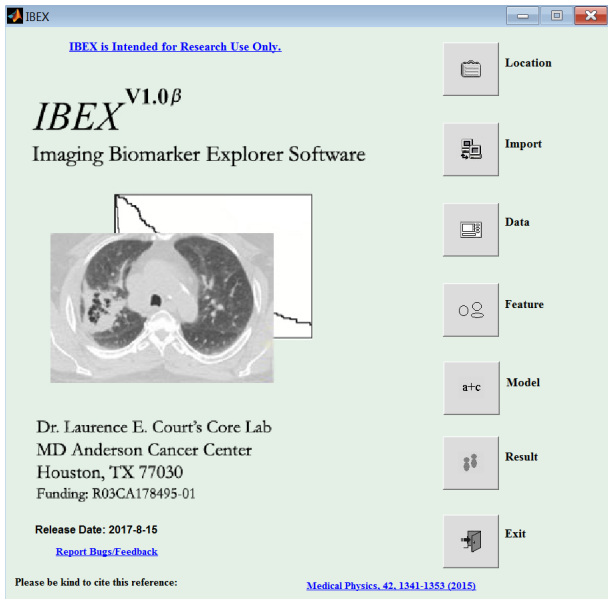


Figure 3.1: *IBEX* opening screen, which allows to import patients' images and masks, create datasets and feature sets and give instruction for the results calculation.

to a patient and each column representing a radiomic feature. The list of radiomic feature categories available in *IBEX* is shown in Table 3.1. A total of about 400 features can be extracted with this software. The matrixes Grey Level Co-Occurrence Matrix, Grey Level Run Length Matrix and Neighbor Intensity Difference can be calculated both in 2D (category "25") and 3D (category "3"). In the first case, the matrix is computed slice by slice and then averaged on all the slices in the selected mask.

3.1.2 LIFEx

LIFEx [25] is a Java freeware, developed for Windows, Linux and Mac systems. The supported image formats include DICOM and non-DICOM files (e.g. NIFTI). For this work, *LIFEx* version 4.90 was exploited. It offers a similar workflow as *IBEX*, starting with the import of images and masks. The VOIs can also be directly drawn in the tool, by a manual or a semi-automatic segmentation, including region growth and threshold techniques. Before extracting the features, the user can set the parameters regarding the spatial resampling (to tune the voxel dimension), the intensity discretisation (number of grey levels and bin size) and the intensity rescaling (absolute or relative). The available features in *LIFEx* are listed in Table 3.1. The total number of features that can be calculated with this software is about 60 (depending on the image type: MR, PET, CT). The output consists in a CSV or XLS file, similarly to *IBEX*.

Table 3.1: List of the feature categories available in the three software under investigation (*IBEX* version 1.0β, *LIFEx* version 4.90 and *PyRadiomics* version 2.2.0). The number of available features in each software is indicated in the last row of the Table.

IBEX	***LIFEx***	***PyRadiomics***
Shape	First Order Features - Shape	Shape
Intensity Histogram	First Order Features - Histogram	First Order Statistics
Grey Level Co-Occurrence Matrix 25 Grey Level Co-Occurrence Matrix 3	Grey Level Co-Occurrence Matrix	Grey Level Co-Occurrence Matrix
Grey Level Run Length Matrix 25	Grey Level Run Length Matrix	Grey Level Run Length Matrix
Neighbor Intensity Difference 25 Neighbor Intensity Difference 3	Neighborhood Grey Level Different Matrix	Neighbouring Grey Tone Difference Matrix
	Grey Level Zone Length Matrix	Grey Level Size Zone Matrix
Gradient Orient Histogram		
Intensity Direct		
Intensity Histogram Gauss Fit		
	Conventional Indices	
		Grey Level Dependence Matrix
~ 400	~ 60	~ 1000

3.1.3 PyRadiomics

PyRadiomics [26] has been implemented in Python and it can be used in two ways, through a front-end interface in 3D Slicer [95] or through a back-end interface. In this thesis, the second option has been chosen and *PyRadiomics* version 2.2.0 was used. The first step is the creation of an input file, i.e. a text file indicating the images and the corresponding masks of interest for the features extraction. Then, the settings for the features extraction have to be indicated in a parameter file (in YALM format), including bin count or bin width, eventual image normalisation, image filters and feature classes. The list of the available feature categories in *PyRadiomics* is reported in Table 3.1. The number of features that can be calculated with this software is about 80 per each image type (original or filtered image), with a total up to a thousand or more (depending for example on the choice of a 2D or 3D extraction and the number of wavelet decomposition).

3.2 Methods

3.2.1 Image acquisition

For this study, a pelvic phantom expressly assembled for radiomic studies was used. The phantom design and its construction has represented one of the tasks of this thesis work and will be described in Chapter 4. Two versions of the phantom were prepared: the first one reproducing the relaxation times T_1 of the pelvis as derived from a set of representative patients and the second one mimicking the relaxation times T_2. MR images of the phantom were acquired at the IEO (European Institute of Oncology, IRCCS, Milan) on a 1.5 T scanner (Optima MR450w, General Electric) with a 24 channels anterior body array coil and a combined 24 channels spinal coil. All images were acquired on the axial plane. The acquisitions were performed following the clinical protocol for pelvic diagnostic imaging used at the Institute, that includes the following sequences:

- T_1-weighted (T_1-w): 354 ms repetition time (TR), 9 ms echo time (TE), 5 mm slice

thickness, 5 mm slice spacing, 320x320 mm^2 field of view (FoV), 320 x 320 acquisition matrix (AM);

- T_2-weighted (T_2-w): 729 ms TR, 141 ms TE, 5 mm slice thickness, 5.5 mm slice spacing, 320x320 mm^2 FoV, 320 x 224 AM.

The first version of the phantom was acquired with the T_1-w sequence, the second one was imaged with the T_2-w acquisition. After the first image series, the phantom (each version) was repositioned and a second acquisition was acquired with an identical protocol.

3.2.2 VOIs

The VOIs were drawn on the images of the first acquisition with the tool 3D Slicer [95] version 4.10.1. The pelvic-shaped phantom includes (in each of its version) four cylindrical inserts, each with a specific texture. The inserts aim to mimic the tumour in a patient affected by a pelvic cancer and thus represent the object of interest for the radiomic analysis. For each insert, four circular contours were drawn, for a total of sixteen segmentations per each images series. In this way, the feature extraction was performed from volumes or areas segmented on the same insert/tissue but of different sizes. The masks were segmented on 3 consecutive slices on the T_2 dedicated phantom (identifying sixteen VOIs), while on the T_1 phantom it was possible to draw the masks only on a single slice, due to the different preparation of the inserts (identifying sixteen ROIs). The biggest VOI follows the edge of the insert (diameter 24 mm) and the others are concentric and with diameters 75% of the main one (diameter 18 mm), 50% (diameter 12 mm) and 25% (diameter 6 mm). The ROIs are shown in Figure 3.2, where the T_1-w phantom images and masks have been loaded in *LIFEx*. The segmentations were exported in NIFTI format to be used as input masks for *LIFEx* and *PyRadiomics*. The NIFTI files were then converted to MAT files to create the input masks for *IBEX*, which does not support the NIFTI format.

3.2.3 Radiomic features selection

Analysing Table 3.1, it is evident that each software offers its own features to be extracted from the VOIs and only some of them have their corresponding in the other software. In fact, some of the features may have different names in other platforms or they may have the same name but a different mathematical definition. In order to compare the feature values when extracted with different software, it is necessary to identify and select for the analysis only the features shared among the three packages. With this aim, the documentation of each software was studied and compared with the others. The definition of *IBEX* features was not as easily accessible as in the other software, since to trace back the formulas it was necessary to collect and consult the references cited by *IBEX* developers. On the contrary, both *LIFEx* and *PyRadiomics* offer a list of the available features with the corresponding definition. The lists can be consulted at this address for *LIFEx* and at this address for *PyRadiomics*.

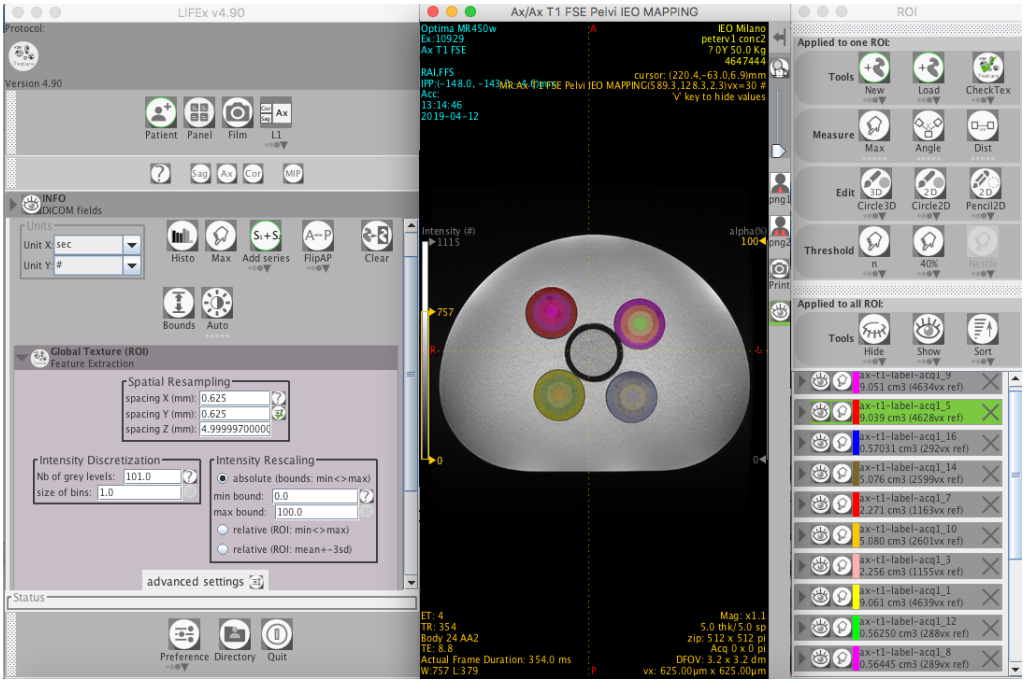


Figure 3.2: The phantom image and corresponding VOIs (in different colours) imported in *LIFEx*.

The feature definitions were also compared with the nomenclature and formulas proposed by the IBSI, which is becoming the reference in the development of radiomics.

3.2.4 Preprocessing and settings

Intensity rescaling and discretisation

The various options offered by the three packages in terms of preprocessing (resampling, discretisation, rescaling) were analysed and compared. An intensity rescaling method shared among the tools was not identified. This consideration, along with the necessity to make the results comparable, led to the decision of rescaling all the images in the range of intensities $[0, 100]$ before using them as input to *IBEX*, *LIFEx* and *PyRadiomics*. The images were rescaled using the filter *Rescale Intensity* provided by *Simple ITK* [96]. The intensity range was then discretised, fixing the bin-width to 1. In this way, 101 bins were obtained. The choice of the bin-width was driven by the suggestion to have a number of bins in the range $[30, 130]$, that has shown good performance and reproducibility in the literature [97].

Software settings

With the aim to perform the feature extraction on equal terms in each software, the following settings were set up. The settings mainly regard the intensity discretisation (intensity range $[0, 100]$, bin-width 1) and the offset for the GLCM (i.e. the distance between

neighbouring voxels), that was set to 1 (i.e. only adjacent voxels are considered for the matrix calculation). The extraction was performed on a 2D level (and not on 3D level), since the image voxel was not isotropic and the downsampling/upsampling to isotropic voxels could introduce a bias. In fact, the downsampling may cause a loss of information on the acquisition plane. On the contrary, the upsampling may artificially create new voxel data that may have an impact on the features [18].

- *IBEX*
  Intensity histogram: NBins 101, Range Fix 0;
  GLCM25: AdaptLimitLevel 1, Offset 1, Symmetric 1;
  GLRLM25: GrayLimits 0 100, NumLevels 101;
  NID25: AdaptLimitLevel 1.

- *LIFEx*
  Intensity Discretization: Nb of grey levels = 101.0, Size of bins = 1.0;
  Intensity Rescaling: absolute, min bound 0.0, max bound 100.0;
  Advanced: GLCM: 1.

- *PyRadiomics*
  Normalize: false; Distances: [1];
  PreCrop: true;
  Force2D: true;
  Force2Dimension: 0;
  GeometryTolerance: 1.e+4;
  BinWidth: 1;
  VoxelArrayShift: 0;
  Label: 1;
  ImageType: Original: ;
  FeatureClass: shape, firstorder, glcm, glrlm, ngtdm.

3.2.5 Statistical analysis

The statistical analysis was performed following the approach proposed by Foy et al. [98], who considered four radiomics software for a comparison on the algorithm implementation. All the analysis were carried out with RStudio version 1.2.1335. Firstly, the Shapiro-Wilk test [99] was used to assess the normality of the data. The null hypothesis was that the data followed a normal distribution. The significance was assessed at $\alpha = 0.05$. Since the majority of the data showed a significant deviation from a normal distribution, the non-parametric Friedman test [100] was performed to detect differences among the feature value when extracted with different software. The null hypothesis was that a feature was the same across the radiomics software. As before, the considered significance level was $\alpha = 0.05$. The Bonferroni correction was applied to take into account the multiple comparisons ($p < 0.00208$). In addition, the correlation among the three feature values (one for each software) was evaluated both pairwise, with the Concordance Correlation Coefficient (CCC) [101] to enhance eventual similarities and

correlations between two software, and globally, through the Overall Concordance Correlation Coefficient (OCCC) [102], in order to identify the features showing the highest stability across the packages.

3.3 Results

3.3.1 Radiomic features selection

Only twenty-four radiomic features were common in all the three packages and they are shown in Table 3.2. For completeness, the mathematical definition of these features is reported in Appendix A. In the *IBEX* column, the suffix 25 refers to the matrixes computed in 2D slice by slice. In the same column, the suffix 333 refers to the feature obtained as the mean of the values calculated for each default angle separately. In fact, the output file of *IBEX* includes both the feature calculated from each angle separately and the mean value, which is the one of interest for the comparison, since the other software give in the output only the mean value. By definition, four directions are included for the computation of the GLCM and the GLRLM in 2D (eight neighbour pixels). This holds for all the packages for the GLCM but it is true for the GLRLM for *LIFEx* and *PyRadiomics* only. On the contrary, *IBEX* supports only two directions for the GLRLM computation ($0°$ and $90°$). In addition, the definition of the feature "(excess) discretised intensity kurtosis" is the same in the three packages but it differs from the IBSI definition by a factor three, as in Equation A.2.2. In IBSI, the Fisher correction is applied to the kurtosis definition.

3.3.2 Software comparison

Table 3.3 shows the results of the Shapiro-Wilk normality test performed on the features extracted with the three software from both the T_1-w and T_2-w images. When a feature showed exactly the same value on all the sixteen VOIs and for both the repeated acquisitions, the test could not be performed. The missing values in Table 3.3 correspond to these cases. The test showed that the distribution of the feature values was significantly different from a normal distribution. Thus, the Friedman test was chosen to assess differences among the features obtained with the three software. The results of the test performed on each feature are shown in Table 3.4, along with the corresponding OCCC values. The results are listed for both the T_1-w and T_2-w image series. The values of the features obtained with the different packages are shown in Figure 3.3 (morphological features), Figure 3.4 (intensity histogram-based features), Figure 3.5 (GLCM-based features), Figure 3.6 (GLRLM-based features) and Figure 3.7 (NGTDM-based features). In each plot, the thirty-two lines represent the sixteen values of the feature (one for each VOI) in the first acquisition and the sixteen values of the feature in the second acquisition (after phantom repositioning).
All features, except "(Excess) discretised intensity kurtosis" and "Discetised intensity skewness", showed significant differences among the packages. Nine (eight) out of twenty-four features showed an OCCC > 0.726 in the T_1-w (T_2-w) experiment, suggesting a correlation among the features, despite the significant differences in their absolute values among the packages. The features showing a correlation in both T_1-w and

Table 3.2: List of radiomic features shared among the three packages. The corresponding feature nomenclature proposed by the IBSI is reported. According to the categories proposed by IBSI (reference manual v9, accessible here), the light cyan-highlighted rows belong to *Morphological features*, the light green to *Intensity Histogram features*, the pink to *Grey Level Co-Occurence based features*, the orange to *Grey Level Run Length based features* and the violet to *Neighbourhood Grey Tone Difference based features*. From the *PyRadiomics* features the root "original_" was omitted in each feature, as it refers to the fact that original (and not filtered images) were considered for this analysis.

N.	IBSI name	*IBEX* name	*LIFEx* name	*PyRadiomics* name
1	Sphericity	Shape-Sphericity	SHAPE_Sphericity	shape_Sphericity
2	Volume (voxel counting)	Shape-NumberOfVoxel	SHAPE_Volume(#vx)	diagnostic_Mask_VoxelNum
3	(Excess) discretised intensity kurtosis	InstensityHistogram-Kurtosis	HISTO_Kurtosis	firstorder_Kurtosis
4	Discetised intensity skewness	InstensityHistogram-Skewness	HISTO_Skewness	firstorder_Skewness
5	Inverse difference	GrayLevelCooccurenceMatrix25-333-1Homogeneity	GLCM_Homogeneity	glcm_Id
6	Angular second moment	GrayLevelCooccurenceMatrix25-333-1Energy	GLCM_Energy	glcm_JointEnergy
7	Contrast	GrayLevelCooccurenceMatrix25-333-1Contrast	GLCM_Contrast	glcm_Contrast
8	Correlation	GrayLevelCooccurenceMatrix25-333-1Correlation	GLCM_Correlation	glcm_Correlation
9	Joint entropy	GrayLevelCooccurenceMatrix25-333-1Entropy	GLCM_Entropy_log2	glcm_JointEntropy
10	Difference average	GrayLevelCooccurenceMatrix25-333-1Dissimilarity	GLCM_Dissimilarity	glcm_DifferenceAverage
11	Short run emphasis	GrayLevelRunLengthMatrix25-333ShortRunEmphasis	GLRLM_SRE	glrlm_ShortRunEmphasis
12	Long run emphasis	GrayLevelRunLengthMatrix25-333LongRunEmphasis	GLRLM_LRE	glrlm_LongRunEmphasis
13	Low grey level run emphasis	GrayLevelRunLengthMatrix25-333LowGrayLevelRunEmpha	GLRLM_LGRE	glrlm_LowGrayLevelRunEmphasis
14	High grey level run emphasis	GrayLevelRunLengthMatrix25-333HighGrayLevelRunEmpha	GLRLM_HGRE	glrlm_HighGrayLevelRunEmphasis
15	Short run low grey level emphasis	GrayLevelRunLengthMatrix25-333ShortRunLowGrayLevelEmpha	GLRLM_SRLGE	glrlm_ShortRunLowGrayLevelEmphasis
16	Short run high grey level emphasis	GrayLevelRunLengthMatrix25-333ShortRunHighGrayLevelEmpha	GLRLM_SRHGE	glrlm_ShortRunHighGrayLevelEmphasis
17	Long run low grey level emphasis	GrayLevelRunLengthMatrix25-333LongRunLowGrayLevelEmpha	GLRLM_LRLGE	glrlm_LongRunLowGrayLevelEmphasis
18	Long run high grey level emphasis	GrayLevelRunLengthMatrix25-333LongRunHighGrayLevelEmpha	GLRLM_LRHGE	glrlm_LongRunHighGrayLevelEmphasis
19	Grey level non-uniformity	GrayLevelRunLengthMatrix25-333GrayLevelNonuniformity	GLRLM_GLNU	glrlm_GrayLevelNonUniformity
20	Run length non-uniformity	GrayLevelRunLengthMatrix25-333RunLengthNonuniformity	GLRLM_RLNU	glrlm_RunLengthNonUniformity
21	Run percentage	GrayLevelRunLengthMatrix25-333RunPercentage	GLRLM_RP	glrlm_RunPercentage
22	Coarseness	NeighborIntensityDifference25-Coarseness	NGLDM_Coarseness	ngtdm_Coarseness
23	Contrast	NeighborIntensityDifference25-Contrast	NGLDM_Contrast	ngtdm_Contrast
24	Busyness	NeighborIntensityDifference25-Busyness	NGLDM_Busyness	ngtdm_Busyness

T_2-w series included "Volume (voxel counting)" within the shape category, "(Excess) discretised intensity kurtosis" within the intensity histogram group, and "Inverse difference", "Angular second moment", "Contrast", "Joint Entropy" and "Difference Average", which all belong to the GLCM-based features.
Regarding the shape features, surprisingly "Sphericity" showed significant differences among software and an OCCC close to zero. This is an unexpected result, since the formula indicated by the packages is the same (see Equation A.1.1). The calculation of this feature involves the surface area and the volume of the VOI. Thus, a possible explanation for the disagreement could reside in the choice of the volume (number of voxels in the VOI, number of voxels multiplied by the volume of a voxel, mesh volume) or in the way the area is obtained. In fact, the surface area is calculated through a triangulation process. *LIFEx* states that the calculation of the surface area is based on the Delaunay triangulation, but it is not clear is this is the same method used by the other software. Concerning the feature "Volume (voxel counting)", the Friedman test showed $p < 0.0002$. The feature had the same values in *PyRadiomics* and *LIFEx*, but it showed some differences in the *IBEX* output. This can be due to a different method for voxel counting at the edges of the VOI. Despite this small deviations, the OCCC for this feature show that the values from the software are highly correlated.
The feature "Discetised intensity skewness" showed an OCCC close to zero for the T_2-w acquisition, despite the Friedman test indicating a perfect agreement among the codes. This unexpected result is due to an outlier value of the feature when extracted from *IBEX* in only one of the VOIs, probably caused by a bug in the code. In fact, after removing the outlier, it results OCCC = 0.981. Similarly, "Contrast (NGTDM)" showed an OCCC close to zero only for the T_1-w acquisition. Also in this case, an outlier was identified in the data from *IBEX*. As expected, by removing the outlier, it can be obtained OCCC = 0.960. Once such corrections were applied, the Friedman test and the OCCC evaluation indicated that the results of the features comparison are independent from the MR sequence used to acquire the images under investigation. Nevertheless, the feature "Correlation" showed OCCC = 1.000 when the feature was extracted from the T_1-w images, but OCCC = 0.301 for the T_2-w acquisition, suggesting that this feature could depend on the sequence used to acquire the images.

Table 3.3: Results of the Shapiro-Wilk normality test. The test statistic W and the corresponding p-value are reported, for the features extracted from both the T_1-w and T_2-w images. The test was performed on the feature values extracted from the three software. For each feature in the table, the first row shows the results of the test on the *LIFEx* value (L), the second row the results on the *PyRadiomics* value (P) and the third row the results on the *IBEX* value (I). The missing results correspond to the feature that had the same value in all the VOIs and for both image acquisitions. In the Table, p-values ≥ 0.05 were highlighted in red.

Feature		T_1-w		T_2-w	
		W	p-value	W	p-value
Sphericity	L	-	-	0.80	< 0.05
	P	0.88	< 0.05	0.83	< 0.05
	I	0.91	< 0.05	0.91	< 0.05
Volume (voxel counting)	L	0.83	< 0.05	0.83	< 0.05
	P	0.83	< 0.05	0.83	< 0.05

	I	0.83	< 0.05	0.83	< 0.05
(Excess) discretised intensity kurtosis	L	0.89	< 0.05	0.91	< 0.05
	P	0.89	< 0.05	0.91	< 0.05
	I	0.89	< 0.05	0.93	< 0.05
Discetised intensity skewness	L	0.91	< 0.05	0.97	0.40
	P	0.91	< 0.05	0.97	0.40
	I	0.91	< 0.05	0.18	< 0.05
Inverse difference	L	0.97	0.43	0.87	< 0.05
	P	0.97	0.43	0.85	< 0.05
	I	0.97	0.44	0.84	< 0.05
Angular second moment	L	0.92	< 0.05	0.71	< 0.05
	P	0.92	< 0.05	0.70	< 0.05
	I	0.92	< 0.05	0.70	< 0.05
Contrast	L	0.86	< 0.05	0.93	< 0.05
	P	0.86	< 0.05	0.92	< 0.05
	I	0.86	< 0.05	0.91	< 0.05
Correlation	L	0.90	< 0.05	0.98	0.79
	P	0.90	< 0.05	0.93	< 0.05
	I	0.17	< 0.05	0.87	< 0.05
Joint entropy	L	0.92	< 0.05	0.90	< 0.05
	P	0.92	< 0.05	0.89	< 0.05
	I	0.91	< 0.05	0.86	< 0.05
Difference average	L	0.95	0.19	0.95	0.17
	P	0.95	0.19	0.93	0.05
	I	0.95	0.19	0.91	< 0.05
Short run emphasis	L	0.97	0.39	0.79	< 0.05
	P	0.97	0.39	0.77	< 0.05
	I	0.75	< 0.05	0.71	< 0.05
Long run emphasis	L	0.98	0.86	0.74	< 0.05
	P	0.98	0.86	0.72	< 0.05
	I	0.84	< 0.05	0.83	< 0.05
Low grey level run emphasis	L	0.78	< 0.05	0.91	< 0.05
	P	0.73	< 0.05	0.88	< 0.05
	I	-	-	-	-
High grey level run emphasis	L	0.99	0.96	0.93	< 0.05
	P	0.98	0.82	0.94	0.08
	I	-	-	-	-
Short run low grey level emphasis	L	0.84	< 0.05	0.92	< 0.05
	P	0.76	< 0.05	0.90	< 0.05
	I	0.37	< 0.05	0.47	< 0.05
Short run high grey level emphasis	L	0.97	0.61	0.93	< 0.05
	P	0.98	0.74	0.94	0.06
	I	0.75	< 0.05	0.71	< 0.05
Long run low grey level emphasis	L	0.67	< 0.05	0.89	< 0.05
	P	0.65	< 0.05	0.81	< 0.05
	I	0.84	< 0.05	0.83	< 0.05
Long run high grey level emphasis	L	0.99	0.95	0.93	< 0.05
	P	0.97	0.60	0.96	0.29
	I	0.84	< 0.05	0.83	< 0.05
Grey level non-uniformity	L	0.91	< 0.05	0.88	< 0.05
	P	0.91	< 0.05	0.91	< 0.05
	I	0.76	< 0.05	0.76	< 0.05
Run length non-uniformity	L	0.86	< 0.05	0.88	< 0.05
	P	0.86	< 0.05	0.86	< 0.05

	I	0.74	< 0.05	0.68	< 0.05
Run percentage	L	0.97	0.42	0.80	< 0.05
	P	0.97	0.42	0.78	< 0.05
	I	0.76	< 0.05	0.76	< 0.05
Coarseness	L	0.75	< 0.05	0.72	< 0.05
	P	0.75	< 0.05	0.73	< 0.05
	I	0.94	0.05	0.18	< 0.05
Contrast (NGTDM)	L	0.81	< 0.05	0.91	< 0.05
	P	0.81	< 0.05	0.94	0.10
	I	0.17	< 0.05	0.94	0.06
Busyness	L	0.96	0.25	0.77	< 0.05
	P	0.92	< 0.05	0.75	< 0.05
	I	0.27	< 0.05	0.27	< 0.05

Table 3.4: Results of the Friedman test (with Bonferroni correction) performed among the feature values when extracted with the three software and OCCC values assessing agreement among packages.

Category	Feature	T_1-w		T_2-w	
		p-value	OCCC	p-value	OCCC
Morphological	Sphericity	< 0.002	< 0.001	< 0.002	0.019
	Volume (voxel counting)	< 0.002	0.999	< 0.002	0.999
Intensity histogram	(Excess) discretised intensity kurtosis	1.000	1.000	1.000	0.947
	Discetised intensity skewness	1.000	1.000	1.000	< 0.001
GLCM based	Inverse difference	< 0.002	1.000	< 0.002	0.855
	Angular second moment	< 0.002	0.997	< 0.002	0.925
	Contrast	< 0.002	1.000	< 0.002	0.726
	Correlation	< 0.002	1.000	< 0.002	0.301
	Joint entropy	< 0.002	0.987	< 0.002	0.947
	Difference average	< 0.002	1.000	< 0.002	0.813
GLRLM based	Short run emphasis	< 0.002	0.001	< 0.002	0.004
	Long run emphasis	< 0.002	< 0.001	< 0.002	< 0.001
	Low grey level run emphasis	< 0.002	0.004	< 0.002	< 0.001
	High grey level run emphasis	< 0.002	< 0.001	< 0.002	< 0.001
	Short run low grey level emphasis	< 0.002	0.018	< 0.002	0.049
	Short run high grey level emphasis	< 0.002	< 0.001	< 0.002	0.135
	Long run low grey level emphasis	< 0.002	< 0.001	< 0.002	< 0.001
	Long run high grey level emphasis	< 0.002	< 0.001	< 0.002	< 0.001
	Grey level non-uniformity	< 0.002	0.257	< 0.002	0.289
	Run length non-uniformity	< 0.002	0.272	< 0.002	0.276
	Run percentage	< 0.002	0.003	< 0.002	0.007
NGTDM based	Coarseness	< 0.002	< 0.001	< 0.002	< 0.001
	Contrast (NGTDM)	< 0.002	< 0.001	< 0.002	0.756
	Busyness	< 0.002	0.146	< 0.002	0.309

The CCC coefficient was calculated pairwise for each feature, to compare two software at a time. The resulting CCC values are shown in Table 3.5, for both the T_1-w and the T_2-w images. Regarding the T_1-w sequence, the best agreement was obtained by comparing *PyRadiomics* and *LIFEx*, with sixteen out of twenty-four radiomic features

Table 3.5: Pairwise CCC values assessing agreement between two packages. Software P, I and L stand for *PyRadiomics*, *IBEX* and *LIFEx*, respectively. The missing results correspond to the feature that had the same value in all the VOIs, not allowing the calculation of the CCC.

Feature	P vs. I		L vs. I		P vs. L	
	T_1-w	T_2-w	T_1-w	T_2-w	T_1-w	T_2-w
Sphericity	0.142	0.038	-	0.006	-	0.041
Volume (voxel counting)	0.999	0.999	0.999	0.999	1.000	1.000
(Excess) discretised intensity kurtosis	1.000	0.921	1.000	0.921	1.000	1.000
Discetised intensity skewness	1.000	0.972	1.000	0.972	1.000	1.000
Inverse difference	1.000	0.949	1.000	0.830	1.000	0.775
Angular second moment	0.996	0.930	0.996	0.955	1.000	0.895
Contrast	1.000	0.838	1.000	0.693	1.000	0.685
Correlation	1.000	0.928	1.000	0.187	1.000	0.223
Joint entropy	0.980	0.927	0.980	0.953	1.000	0.960
Difference average	1.000	0.903	1.000	0.784	1.000	0.760
Short run emphasis	< 0.001	< 0.00	< 0.001	< 0.001	1.000	0.572
Long run emphasis	< 0.001	< 0.001	< 0.001	< 0.001	1.000	0.529
Low grey level run emphasis	-	-	-	-	0.046	0.141
High grey level run emphasis	-	-	-	-	0.169	0.436
Short run low grey level emphasis	0.002	0.004	0.008	0.007	0.038	0.137
Short run high grey level emphasis	0.188	0.004	0.040	0.013	0.159	0.445
Long run low grey level emphasis	0.309	0.018	0.011	< 0.001	0.081	0.141
Long run high grey level emphasis	< 0.001	< 0.001	< 0.001	< 0.001	0.257	0.373
Grey level non-uniformity	< 0.001	< 0.001	< 0.001	< 0.001	1.000	0.969
Run length non-uniformity	< 0.001	< 0.001	< 0.001	< 0.001	1.000	0.984
Run percentage	0.001	< 0.001	0.001	< 0.001	1.000	0.556
Coarseness	< 0.001	< 0.001	< 0.001	< 0.001	1.000	0.817
Contrast (NGTDM)	0.936	0.884	0.936	0.718	1.000	0.714
Busyness	0.001	0.002	0.004	0.002	0.575	0.966

showing CCC > 0.7. For the same acquisition, the comparison between *PyRadiomics* and *IBEX* gave ten out of twenty-four features showing CCC > 0.7. The same holds for *LIFEx* if compared to *IBEX*. Similarly, on the T_2-w images the best agreement is between *PyRadiomics* and *LIFEx*, but the number of features showing CCC > 0.7 decreased to twelve. This could suggest a dependence of the results on the choice of the sequence for image acquisition. This hypothesis can be further investigated with the following analysis of the feature distributions across the packages, focusing on the comparison of the feature distribution when changing the MR sequence.

Morphological features

The feature "Sphericity" (Figure 3.3 a) could be calculated only with *PyRadiomics* and *IBEX* on the T_1-w images, since the VOI was drawn only on one slice and *LIFEx* support the calculation only on a number of slices greater than one. The results are shown in a different colour on the basis of the corresponding VOI size (blue = diameter 24 mm, red

= diameter 18 mm, green = diameter 12 mm, black = diameter 6 mm). The plot showed a trend inversion when comparing the results from *PyRadiomics* and *IBEX*: in *PyRadiomics*, the highest sphericity matched the smallest VOI and the lowest sphericity matched the biggest VOI. In *IBEX*, the trend was the other way around. The results from the T_2-w images (b) showed a more linear trend, despite the values being significantly different among software. Being the sphericity a 3D feature, the OCCC value close to zero in (a) could be associated with the software not functioning properly with a VOI drawn on a single slice. However, this is not true for (b), where the VOI was identified on three consecutive slices. The OCCC was still close to zero in the T_2-w images, suggesting that this feature (surprisingly, a morphological one) was not stable in this experiment.
Figure 3.3 (c) and (d) show the values of the feature "Volume (voxel counting)" on the T_1-w and T_2-w images, respectively. The data were presented based on VOI size, as in the previous feature. The trend was stable in both acquisitions, but the OCCC was not exactly one, being the *IBEX* values slightly smaller than the others. This could be associated with a round down counting of the number of voxels at the edges of the VOI.

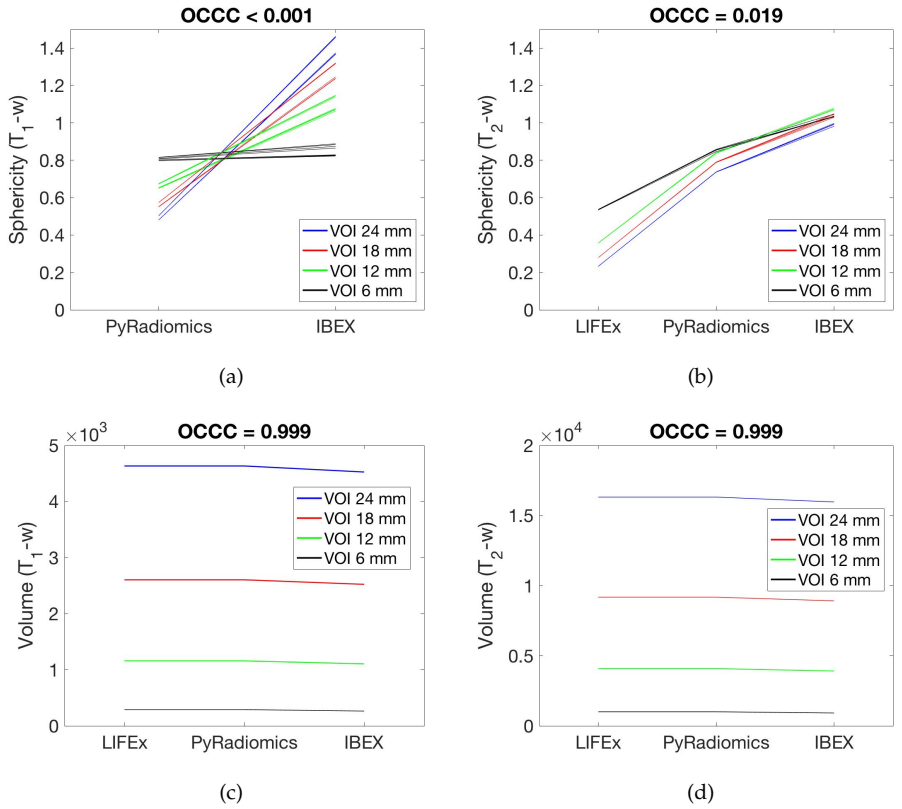


Figure 3.3: Distribution of the morphological features. (a-b) Sphericity; (c-d) Volume (voxel counting). The plots on the left (right) show the features extracted from the T_1-w (T_2-w) images.

Intensity histogram features

The feature "(Excess) discretised intensity kurtosis" showed a complete agreement (OCCC = 1) among packages when extracted from the T_1-w images, as shown in Figure 3.4 (a). The results were presented based on acquisition: blue lines are the sixteen values in the corresponding VOIs in the first acquisition, and the red lines refer to the data of the second acquisition (after phantom repositioning). A shift of the values between the two acquisitions was observed. In (b) the results of the extraction from the T_2-w images are reported. The plot shows that, in the second acquisition, some of the *IBEX* values diverge as compared to the other two software, causing the OCCC coefficient to decrease at 0.947. A similar behaviour was observed for the feature "Discetised intensity skewness", as shown in 3.4 (c) and (d). These results suggest that in the T_2-w acquisition, the phantom repositioning affected the stability of the histogram-based features in *IBEX*, in contrast with the T_1-w acquisition.

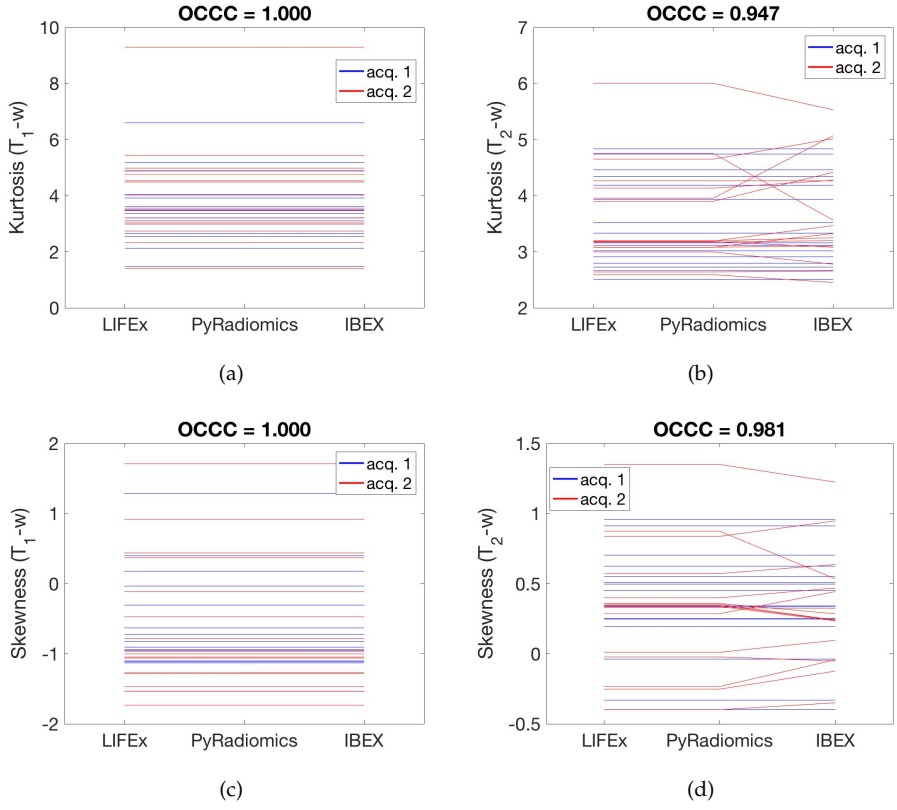


Figure 3.4: Distribution of the intensity histogram-based features. (a-b) (Excess) discretised intensity kurtosis; (c-d) Discetised intensity skewness. The plots on the left (right) show the features extracted from the T_1-w (T_2-w) images.

GLCM-based features

The values of the GLCM-based features are shown in Figure 3.5. This class of features, being a descriptor of the VOI texture, is of particular interest for the radiomic analysis. In all plots, the data were grouped by insert. Inserts 3 had a finer texture, inserts 1 and 4 had a medium texture and insert 2 was characterised by a coarser texture. A detailed description of the inserts design and content will be provided in the next Chapter. The data from the first and second acquisition were represented by a solid and a dashed line, respectively.
Subfigures (a) and (b) show the feature "Inverse difference". The agreement among software was maximum for the T_1-w images (OCCC = 1.000), while it decreased to OCCC = 0.855 for the T_2-w images. In (b), the feature had the same value in *PyRadiomics* and *IBEX* for the first acquisition, with divergence increasing for the second acquisition. In both (a) and (b), differences were observed between the feature value at the first acquisition (solid lines) and after phantom repositioning (dashed lines). In (b), the feature allowed to identify the inserts, having different values according to the different texture. In particular, the feature values from inserts 1 and 4 overlapped as expected, since the two inserts had a similar texture. This behaviour was not observed in the T_1 plot, probably due to the VOI being drawn only on one slice of the phantom.
Similar considerations hold for the features "Angular second moment" and "Contrast". As shown in the plot (f), insert 3 exhibited the lowest contrast value, as expected. In fact, insert 3 had the finer texture, resulting in a small variation of grey levels in neighbouring voxels and a low contrast.
The feature "Correlation" showed a perfect agreement among software (OCCC = 1.000) when extracted from the T_1-w images, but a scarce agreement on the T_2-w images, with the best accordance between the *PyRadiomics* and *IBEX* packages. Similar considerations can be applied to the feature "Difference average". As regards "Joint entropy", in both plots (i) and (j) an overestimation by *IBEX* was observed with respect to the results of *PyRadiomics* and *LIFEx*.
All the GLCM features considered in this analysis, except for "Correlation", were able to discriminate the texture of the different inserts, with particularly good performance in distinguishing insert 3 from the others, when extracted from the T_2-w images.

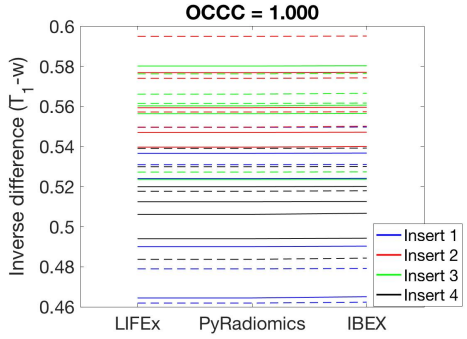
GLRLM-based features

Figure 3.6 shows the distribution of the GLRLM-based radiomic features. The comparison is shown only for the features calculated with *PyRadiomics* and *LIFEx*, since the values from *IBEX* were in disagreement in all the cases and not even comparable for the order of magnitude, as confirmed by the pairwise CCC values in Table 3.5.
Three features ("Short run emphasis", "Long run emphasis", and "Run percentage") showed a perfect agreement between the software when extracted from the T_1-w images. They showed a more modest agreement on the T_2-w images, where they were able to clearly discriminate the different texture (plots (b), (d), and (v)). As expected, the values in inserts 1 and 4, which had a similar content, overlapped.
The majority of the GLRLM-based features had no agreement between the values ex-

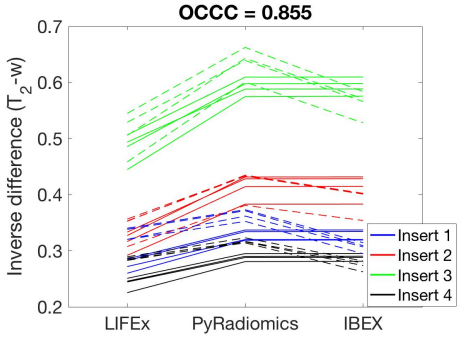
tracted with *PyRadiomics* and *LIFEx*, neither on the T_1-w images nor on the T_2-w images. These features are: "Low grey level run emphasis", "High grey level run emphasis", "Short run low grey level emphasis", "Short run high grey level emphasis", "Long run low grey level emphasis", and "Long run high grey level emphasis".
The features "Grey level non-uniformity" and "Run length non-uniformity" showed a maximum CCC on the T_1-w acquisitions (plots (q) and (s)) as well as a very good agreement on the T_2-w images (plots (r) and (t)). However, they showed a scarse ability to distinguish the four inserts.

NGTDM-based features

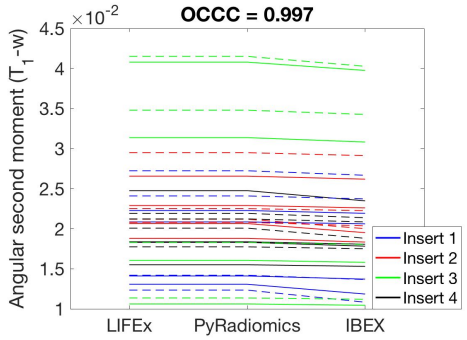
The values of the NGTDM-based radiomic features considered in this analysis are shown in Figure 3.7. As before, in all plots the data was grouped by insert and the values from the first and second acquisition were represented by a solid and a dashed line, respectively.
For the feature "Coarseness" only the results from *PyRadiomics* and *LIFEx* are represented, as the feature values in *IBEX* were out of the same order of magnitude and showed no agreement with the other software. Both plots (a) and (b) demonstrated a perfect agreement between *PyRadiomics* and *LIFEx* on the T_1-w images and a good agreement on the T_2-w images, as confirmed by the corresponding CCC values.
The feature "Contrast" had the maximum agreement among all packages in the category of the NGTDM-based features. The comparison resulted in an OCCC > 0.7 in both the T_1 and the T_2 assessment.
The values of the feature "Busyness" were comparable if considering *PyRadiomics* and *LIFEx*, but in disagreement with *IBEX*, resulting in low OCCC in both (e) and (f).


OCCC = 1.000
Inverse difference (T1-w)
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46
LIFEx
PyRadiomics
IBEX
Insert 1
Insert 2
Insert 3
Insert 4

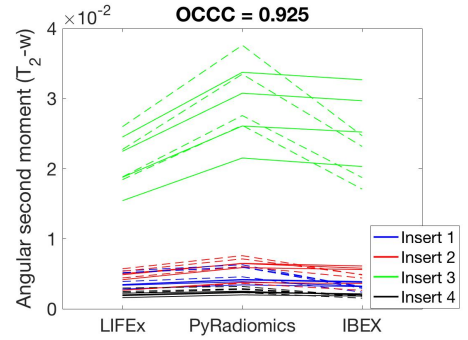
(a)


OCCC = 0.855
Inverse difference (T2-w)
0.7
0.6
0.5
0.4
0.3
0.2
LIFEx
PyRadiomics
IBEX
Insert 1
Insert 2
Insert 3
Insert 4

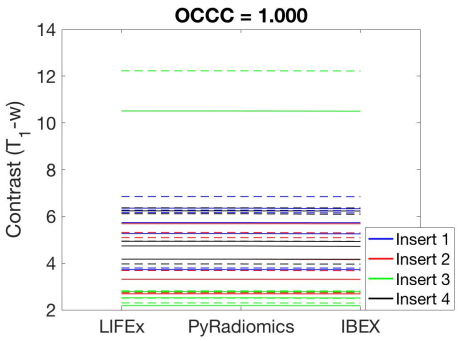
(b)


OCCC = 0.997
×10-2
Angular second moment (T1-w)
4.5
4
3.5
3
2.5
2
1.5
1
LIFEx
PyRadiomics
IBEX
Insert 1
Insert 2
Insert 3
Insert 4

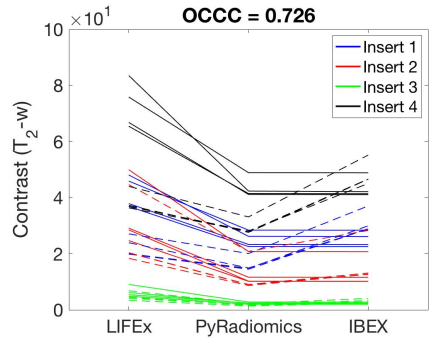
(c)


OCCC = 0.925
×10-2
Angular second moment (T2-w)
4
3
2
1
0
LIFEx
PyRadiomics
IBEX
Insert 1
Insert 2
Insert 3
Insert 4

(d)


OCCC = 1.000
Contrast (T1-w)
14
12
10
8
6
4
2
LIFEx
PyRadiomics
IBEX
Insert 1
Insert 2
Insert 3
Insert 4

(e)


OCCC = 0.726
×101
Contrast (T2-w)
10
8
6
4
2
0
LIFEx
PyRadiomics
IBEX
Insert 1
Insert 2
Insert 3
Insert 4

(f)

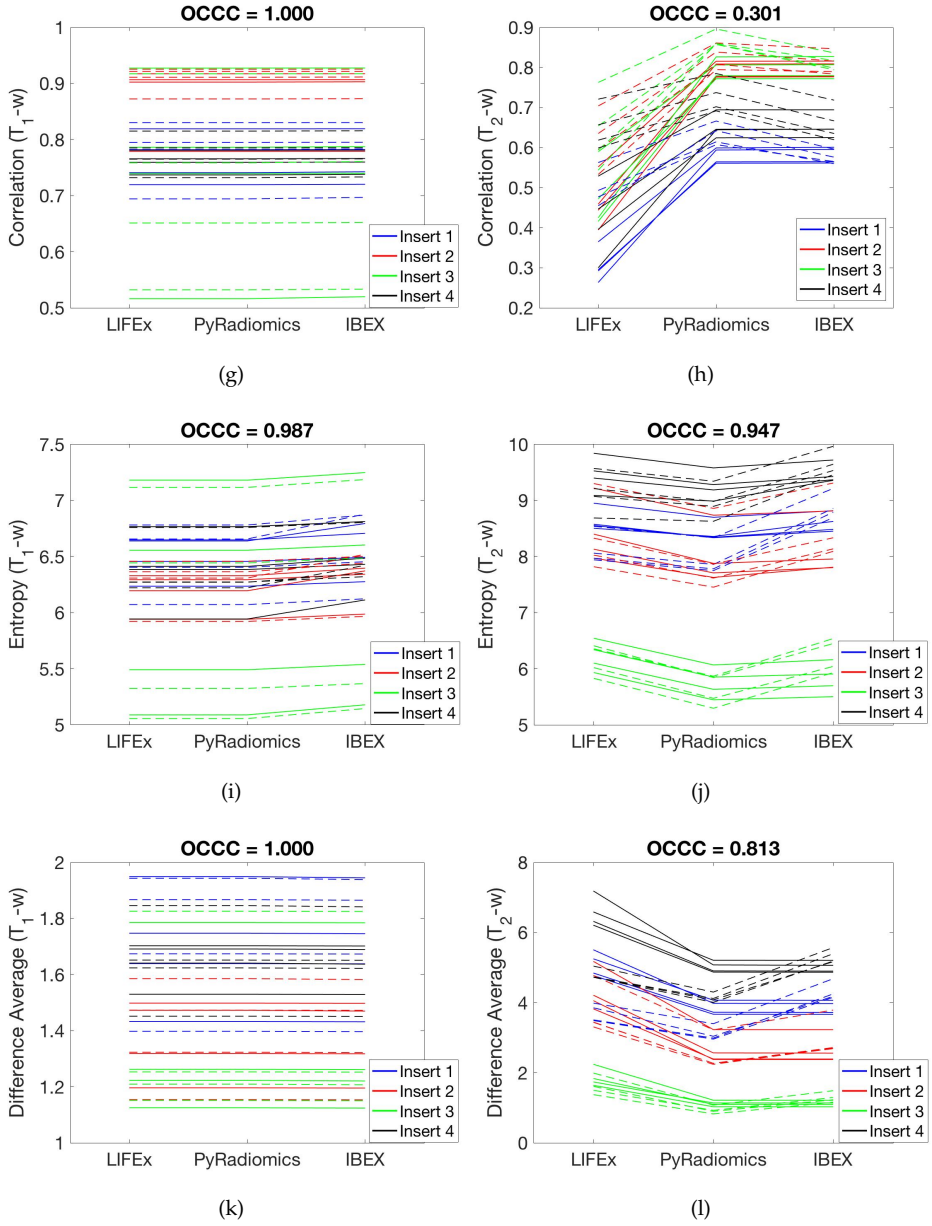
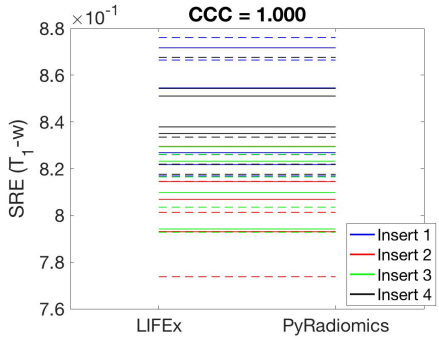
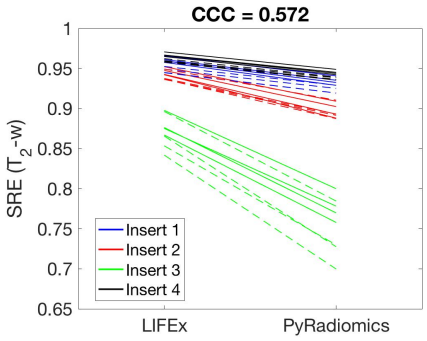


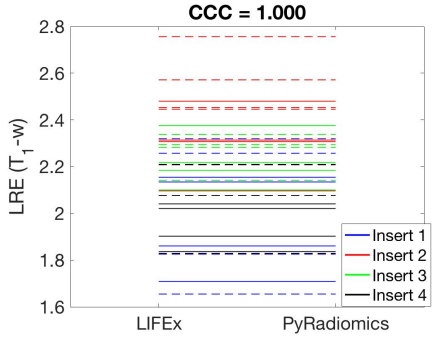
Figure 3.5: Distribution of the GLCM-based features. (a-b) Inverse Difference; (c-d) Angular second moment; (e-f) Contrast; (g-h) Correlation; (i-j) Joint entropy; (k-l) Difference average. The plots on the left (right) show the features extracted from the T_1-w (T_2-w) images. The results of the first and second acquisition are represented as a solid and dashed line, respectively.


CCC = 1.000
×10^-1
SRE (T1-w)
8.8
8.6
8.4
8.2
8
7.8
7.6
LIFEx
PyRadiomics
Insert 1
Insert 2
Insert 3
Insert 4

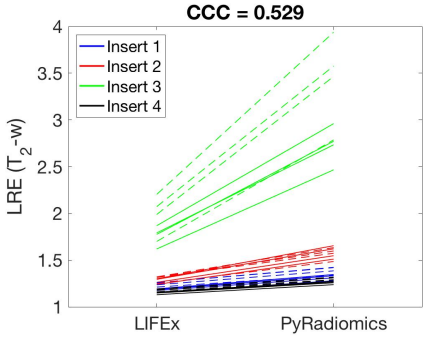
(a)


CCC = 0.572
SRE (T2-w)
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
LIFEx
PyRadiomics
Insert 1
Insert 2
Insert 3
Insert 4

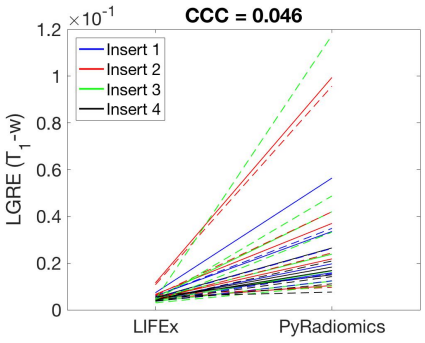
(b)


CCC = 1.000
LRE (T1-w)
2.8
2.6
2.4
2.2
2
1.8
1.6
LIFEx
PyRadiomics
Insert 1
Insert 2
Insert 3
Insert 4

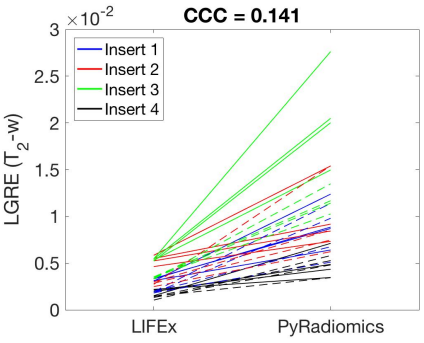
(c)


CCC = 0.529
LRE (T2-w)
4
3.5
3
2.5
2
1.5
1
LIFEx
PyRadiomics
Insert 1
Insert 2
Insert 3
Insert 4

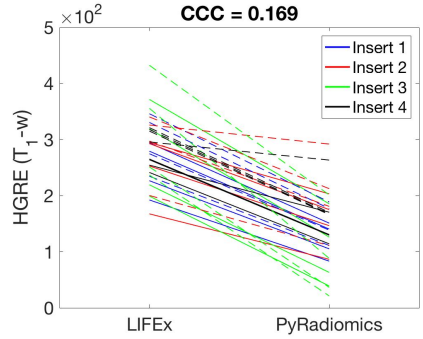
(d)


CCC = 0.046
×10^-1
LGRE (T1-w)
1.2
1
0.8
0.6
0.4
0.2
0
LIFEx
PyRadiomics
Insert 1
Insert 2
Insert 3
Insert 4

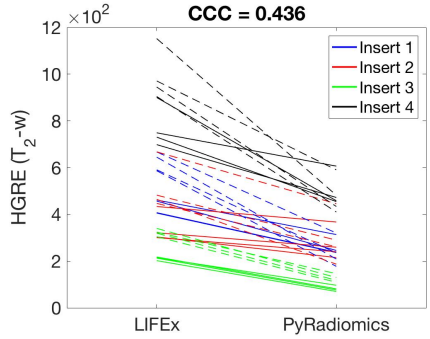
(e)


CCC = 0.141
×10^-2
LGRE (T2-w)
3
2.5
2
1.5
1
0.5
0
LIFEx
PyRadiomics
Insert 1
Insert 2
Insert 3
Insert 4

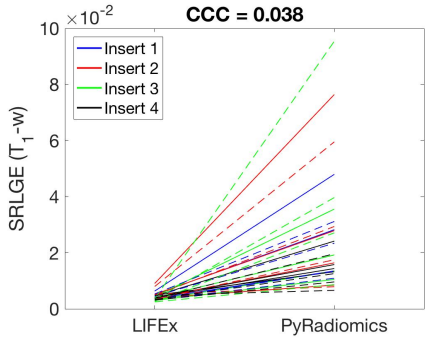
(f)


CCC = 0.169
×10^2
HGRE (T1-w)
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

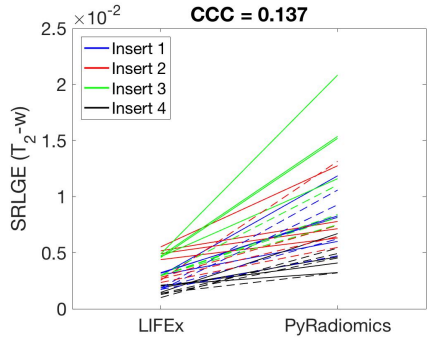
(g)


CCC = 0.436
×10^2
HGRE (T2-w)
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

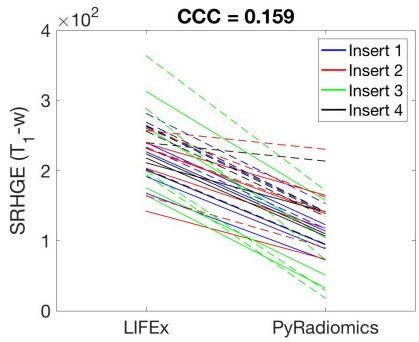
(h)


CCC = 0.038
×10^-2
SRLGE (T1-w)
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

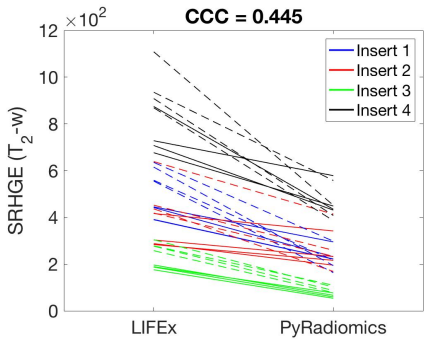
(i)


CCC = 0.137
×10^-2
SRLGE (T2-w)
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

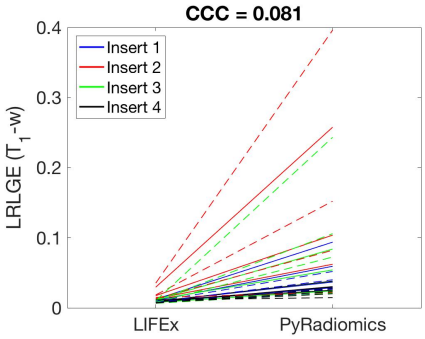
(j)


CCC = 0.159
×10^2
SRHGE (T1-w)
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

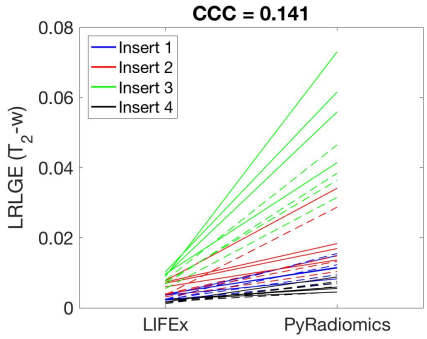
(k)


CCC = 0.445
×10^2
SRHGE (T2-w)
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

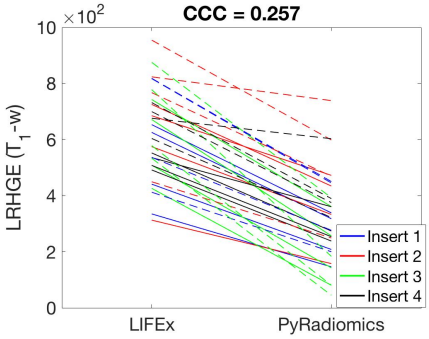
(l)


CCC = 0.081
LRLGE (T1-w)
0.4
0.3
0.2
0.1
0
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

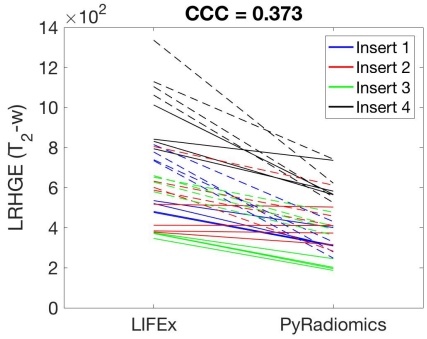
(m)


CCC = 0.141
LRLGE (T2-w)
0.08
0.06
0.04
0.02
0
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

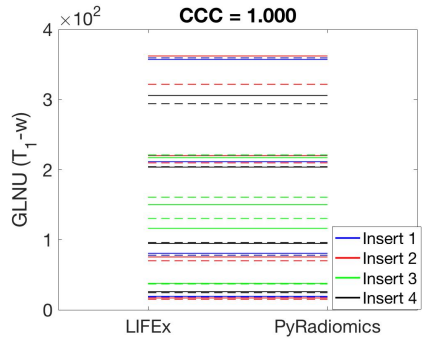
(n)


CCC = 0.257
×10^2
LRHGE (T1-w)
10
8
6
4
2
0
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

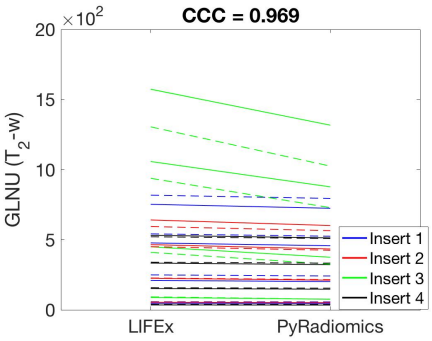
(o)


CCC = 0.373
×10^2
LRHGE (T2-w)
14
12
10
8
6
4
2
0
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

(p)


CCC = 1.000
×10^2
GLNU (T1-w)
4
3
2
1
0
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

(q)


CCC = 0.969
×10^2
GLNU (T2-w)
20
15
10
5
0
Insert 1
Insert 2
Insert 3
Insert 4
LIFEx
PyRadiomics

(r)

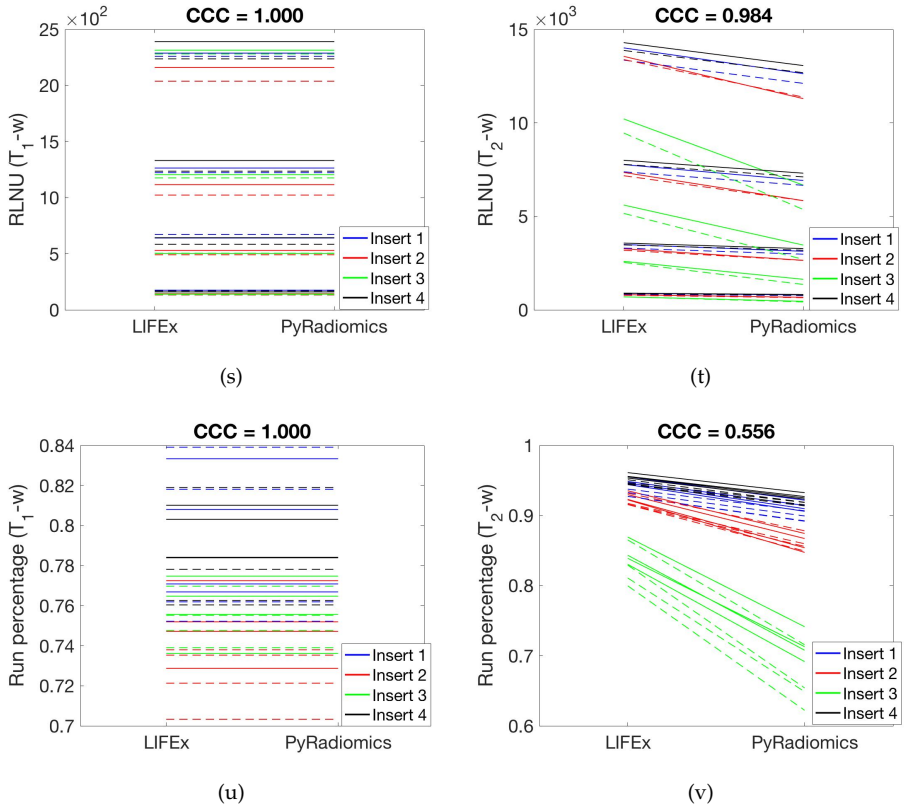


Figure 3.6: Distribution of the GLRLM-based features. (a-b) Short run emphasis; (c-d) Long run emphasis; (e-f) Low grey level run emphasis; (g-h) High grey level run emphasis; (i-j) Short run low grey level emphasis; (k-l) Short run high grey level emphasis; (m-n) Long run low grey level emphasis; (o-p) Long run high grey level emphasis; (q-r) Grey level non-uniformity; (s-t) Run length non-uniformity; (u-v) Run percentage. The plots on the left (right) show the features extracted from the T_1-w (T_2-w) images. The results of the first and second acquisition are represented as a solid and dashed line, respectively.

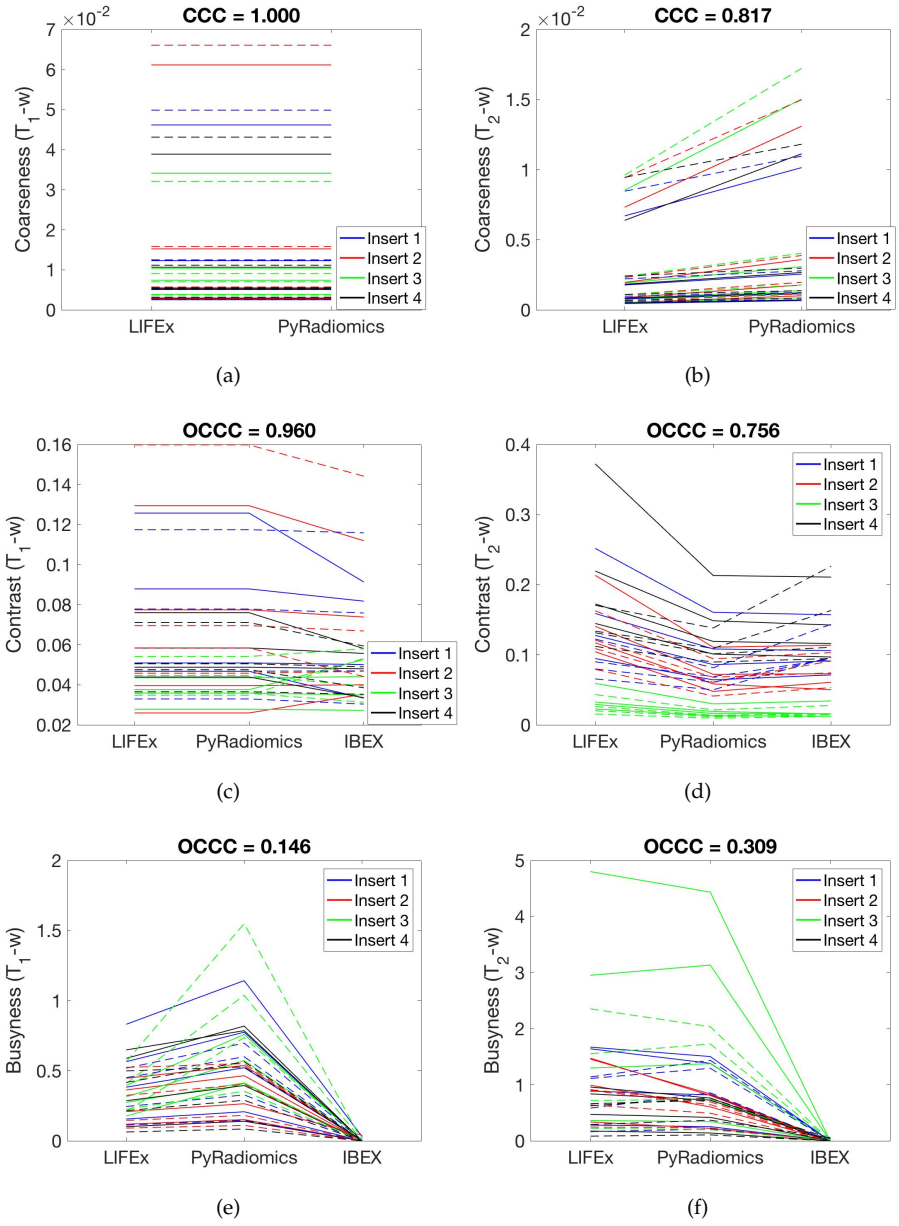


Figure 3.7: Distribution of the NGTDM-based features. (a-b) Coarseness; (c-d) Contrast (NGTDM); (e-f) Busyness. The plots on the left (right) show the features extracted from the T_1-w (T_2-w) images. The results of the first and second acquisition are represented as a solid and dashed line, respectively.

3.4 Discussion

In this Chapter, the results about the comparison among three software for the extraction of radiomic features were presented.
Firstly, the features shared (same mathematical definition) among the packages were identified to ensure comparable results. Only twenty-four features (among the tens or hundreds available in the different software) were selected for this analysis, pointing out how a standardisation is needed but still not reached in the field. The IBSI proposed a list of the features with nomenclature and formulae, but many software are still not in accordance with them. Inevitably, this lack of standardisation hinders the generalisability of the results of a radiomic study, also considering the availability of an increasing number of radiomic software, many of which developed in-house.

The analysis demonstrated that the majority of the features had significant differences among their values when extracted with the three tools, despite implementing the same mathematical formula. In fact, the Friedman test showed that only the histogram-based features were the same across the radiomics software. Surprisingly, even the considered shape features showed instability in this assessment. Despite the mathematical formula being the same, the algorithm implementation was software-specific and could differ for the counting of the number of voxels at the edge of the VOI or for the calculation of the surface area (by using different triangulation systems), generating variations in the calculation of the shape features. The results of the Friedman test and the OCCC analysis were in almost all the cases independent from the MR sequence used to acquire the images (T_1-w and T_2-w acquisitions).
Although the significant differences among the absolute values, many features showed a significant correlation (OCCC > 0.7) among the packages. As proposed by Foy et al. [98], this suggests that the variations in the software implementation are sufficiently high to result in significant differences among a feature value but, at the same time, sufficiently small respect to the value of the feature itself to result in a high OCCC. Among the texture features, the GLCM-based features showed the highest correlation and stability among the packages, along with the feature "Contrast" of the NGTDM category.
The majority of the GLRLM-based features showed poor correlation. To explain this result, two main sources of error in the software comparison were identified. The first one is that in *PyRadiomics* and *LIFEx* the GLRLM is calculated along four directions, but only two directions are available for the calculation in *IBEX* (0 and 90°). The second reason lies in the different calculation of the GLRLM indices performed in *PyRadiomics* and *LIFEx*. More in detail, in *PyRadiomics* the grey levels of the GLRLM are cropped between the minimum and maximum level observed in the image and the run length is cropped to the maximum run length. This results in a shift of the matrix indexes *(i, r)*, that does not occur in the calculation made by *LIFEx*. As a matter of fact, in *LIFEx* documentation, the reader is warned about this difference and invited to perform with care an eventual comparison of the texture results among software. From one side, this warning helps the researchers in radiomics to take extra care in the choice of the radiomics software for a specific study and to know that the results will probably be dependent on the software

used. From another side, the warning encloses an open issue: if the results of a radiomic analysis are dependent on the software used, how can the results be generalisable? Secondly, if the radiomic features are not robust, how can the corresponding correlations with the clinical outcomes be reliable? In this framework, the IBSI initiative should help in the harmonisation and standardisation of radiomics, providing a benchmark for the features definition and the feature values if extracted from a digital phantom of reference. However, if different software give comparable features on the digital phantom, it is not immediate that this result will hold on all clinical images, as it could be texture-dependent.
In general, apart from variations in the software algorithms, another possible explanation for the differences obtained with the three software could be an incorrect setting of the initial parameters and offset in the tools. It is not straightforward to understand how to adjust the parameters to ensure an extraction in exactly the same conditions among the packages. Thus, the user cannot be sure to be operating in equal terms. Evidently, a standardisation in the protocols for the feature extraction, also with respect to the setting of the initial conditions, is a goal in the near future radiomics.
Given the previous considerations, an ideal solution in perspective of a well-established radiomics could be the convergence of the scientific community towards a "universal radiomic software". Alternatively, a comprehensive investigation on the comparison of the most diffused radiomics software in the clinical studies should be carried out with the aim to identify a set of calibration functions among them.

In order to assess the agreement between each software pair, the CCC was calculated as well. The results showed that *PyRadiomics* and *LIFEx* had the best agreement, especially on the T_1-w series. In fact, it seemed that the repositioning of the phantom (and the consequent co-registration of the images between the two acquisitions) had an impact on the agreement of the two software results extracted from the T_2-w images. This result is particularly significant, as some clinical studies selected for a radiomic analysis include multiple acquisitions (also with different imaging type, e.g. PET and CT) and the co-registration of the images is a common practice.
In addition to the previous reasons, the fact that *IBEX* showed less agreement with the other packages could be related also with the conversion of the VOIs mask into MAT files which was necessary only for this tool, given the limited file type supported. Moreover, it was noticed that for some features and specific VOIs, *IBEX* provided values that were evidently misaligned with the other values and had to be treated as outliers. However, it is possible that not all these exceptions were identified given the amount of data, influencing the final comparison.
Another consideration, rising from the plots showing the distribution of each feature in the three software, is about the ability of the radiomic features to discriminate different textures. As explained, the inserts included in the phantom under investigation were prepared to simulate different textures in a certain range. The majority of the GLCM-based features extracted from the T_2-w images showed a good performance in distinguishing the inserts, especially regarding insert 3, that had the finer texture. Only three out of eleven GLRLM-based features showed this ability.

Summarising, the analyses presented in this Chapter show that it is difficult to compare different radiomic software, as they offer different features to be extracted and different user-dependent parameters. Further discrepancies can emerge due to the implementation of different software algorithms. The major sources of divergences identified in this study are shown in Figure 3.6. The texture features showing the best agreement among all the considered packages were the GLCM-based features, which revealed also a good ability in distinguishing different textures and thus were identified as the most robust features in this analysis.

Table 3.6: Major sources of differences in the values of radiomic features extracted with the three packages considered in this study.

Identified sources of divergences
Software-specific algorithm implementation
Customised directions for matrixes calculation (e.g. GLRLM)
User-independent settings for matrixes calculation
Incorrect adjustment of user-dependent parameters to operate the software in identical initial conditions
Conversion of mask files in other format to make them readable from all packages

In the next Chapter, the extraction of the radiomic features will be performed with the software *PyRadiomics*, which is widely used in many scientific papers and turned out to be the closest to the IBSI recommendations.

CHAPTER 4

PETER PHAN: the pelvis phantom

This Chapter consists in two main Sections. The first one provides a detailed description of the phantom creation, from the design to the production and validation. The phantom, called PETER PHAN (PElvis TExtuRe PHANtom), was developed to simulate, for radiomic purposes, the pelvis of female patients affected by pelvic malignancies. The investigations performed on the MR images of the phantom aimed at the optimisation of the radiomic workflow, in support of clinical studies. The second Section shows the results of the radiomic experiments performed on the phantom itself, including the assessment of the radiomic features repeatability and reproducibility.

4.1 Phantom design and realisation

This Section has been adapted from [103].
The image acquisition were performed at the IEO (European Institute of Oncology, IRCCS, Milan).

The characteristics desired of a phantom suitable for supporting clinical MRI-based radiomic studies of tumours located in female pelvis include: being pelvis-shaped, constructed of MR-safe materials, providing relaxation times similar to human tissues and containing regions of different textures, similar to those found *in vivo* for the tumoural tissue. As described in the following paragraphs, these design criteria were addressed through a series of experimental steps regarding:

- measurements of *in vivo* longitudinal relaxation time (T_1) and transverse relaxation time (T_2) of pelvic tissues on both patients and healthy volunteers;
- choices of materials and formulations to reproduce the relaxation times measured *in vivo*;
- investigation of means to incorporate different textures into the phantom;
- validation of the assembled phantom through comparison of relaxation times and texture with the *in vivo* reference values.

4.1.1 *In vivo* T_1 and T_2 mapping

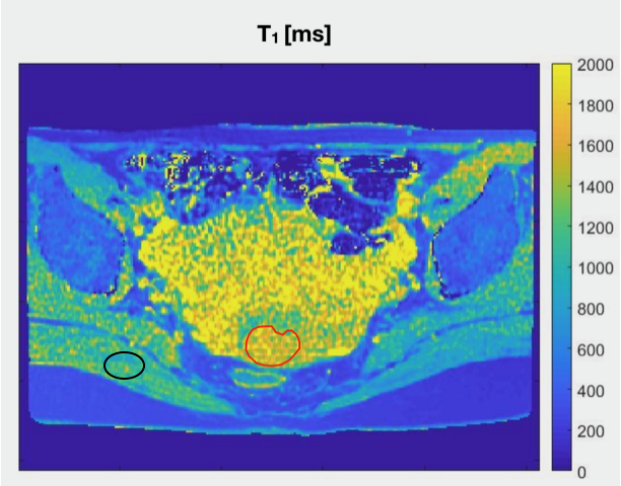
19 healthy volunteers (22-53 yrs, mean 35) and 5 patients with a pelvic tumour (29-57 yrs, mean 47) were imaged with dedicated MR sequences for *in vivo* assessment of T_1 and T_2 relaxation times of healthy tissue (muscle) and pelvic tumour. Written informed consent to undergo the examination, and to the use of anonymised clinical and imaging data for scientific and/or educational purposes, was obtained from all subjects involved. Clinical characteristics of patients are listed in Table 4.1. Inclusion criterion for patients was the presence of a lesion visible on MR images, with a minimum tumour diameter 18 mm.

Table 4.1: Characteristics of the patient cohort.

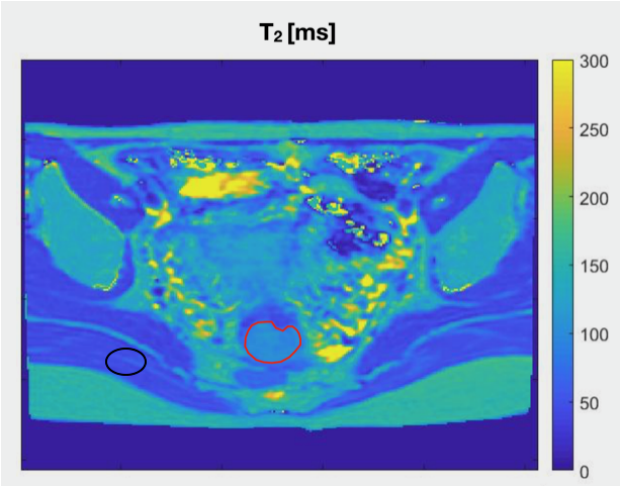
Patient	Age (years)	Pathology	Tumour site
1	57	Vulvar cancer	Inguinal lymph nodes
2	47	Carcinoma of cervix	Uterine cervix
3	50	Adenocarcinoma of the rectum	Rectum
4	29	Carcinoma of cervix	Uterine cervix
5	54	Submucous myoma of the uterus	Endometrium

Image acquisition was carried out using a 1.5 T MR scanner (Optima MR450W, General Electric Healthcare, Waukesha, USA) with a 24 channels anterior body phased array coil and a 24 channels spine array coil.
Maps of T_1 can be obtained with various approaches, including Inversion Recovery (IR) pulse sequence [104] and novel techniques like Magnetic Resonance Fingerprinting [105], although the last one is far from being clinically validated and confirmed by studies in a sizeable number of different centers. In this study *in vivo* T_1 mapping made use of the variable flip angle (VFA) spoiled gradient recalled echo (SPGR) technique [106]. Three axial scans were acquired, each with a different flip angle (5, 10 or 20°), keeping the other parameters fixed: repetition time (TR) 6.7 ms, echo time (TE) 4.2 ms, slice thickness 3 mm, slice spacing 3 mm, field of view (FoV) 220x220 mm^2.
T_2 mapping was based on a dedicated 2D multi-echo spin-echo pulse sequence which ran under research mode (granted by GE Healthcare) and could read up to 16 echoes. The sequence parameters were: TR 5000 ms, min TE 7.5 ms, max TE 119.8 ms, echo spacing 7.5 ms, slice thickness 5 mm, slice spacing 6.5 mm, FoV 220x220 mm^2.
After the acquisition, the DICOM images were anonymised and exported for processing. The VFA data were processed with the NOVIFAST [107] algorithm (version 1.0.0.3, University of Antwerp, Belgium) to calculate T_1 maps. T_2 maps were generated by fitting a mono-exponential curve to the 16-echo spin-echo data on a voxel-by-voxel basis, using a MATLAB (version R2018b, The MathWorks, Inc., USA) script developed in-house.
Volumes of interest (VOIs) in the gluteus maximus muscle were drawn manually on the T_1 and T_2 maps of all subjects. In addition, an expert radiologist drew a VOI for the tumour on the axial T_2-w images obtained as part of the routine clinical protocols. The tumour VOI was then transferred onto the relaxation time maps. For each VOI, the subject-wise mean values of T_1 and T_2 were extracted.

The range of T_1 and T_2 values obtained for the tissues of interest are summarised in the upper part of Table 4.2 and examples of the *in vivo* T_1 and T_2 maps are given in Figure 4.1. The values reported are the mean over all the patients and the minimum and maximum values.



(a)



(b)

Figure 4.1: (a) T_1 and (b) T_2 maps of a patient with cervical cancer. The tumour site VOI is highlighted with a red contour and the muscle VOI contour in black.

4.1.2 Relaxometry of $MnCl_2$ solutions

Conventional MR phantoms are based on solutions of paramagnetic contrast agents, which are easily soluble in water. The solutions obtained are stable in time and mimic

Table 4.2: Relaxation times of human tissues measured at 1.5 T *in vivo* (upper) and of background and inserts measured in PETER PHAN V1 and V2 (lower). *mean (range)

		T_1 [ms]*	**T_2 [ms]***
in vivo mapping	Muscle	1128 (806 - 1378)	51 (40 - 65)
	Tumour	1637 (1396 - 2121)	94 (79 - 101)
Phantom validation	$MnCl_2$	1277 (768 - 2120)	57 (55 - 59)
	INS1	1366 (6 - 1999)	137 (106 - 183)
	INS2	1345 (22 - 2000)	123 (101 - 192)
	INS3	1525 (16 - 1996)	117 (101 - 150)
	INS4	1447 (19 - 2000)	120 (89 - 167)

the typical relaxation times of human tissues well [108]. The paramagnetic ion manganese chloride ($MnCl_2$) was chosen for this study due to its availability and low cost. A Fourier Transform spectrometer (Apollo, Tecmag, Houston, USA) for pulsed Nuclear Magnetic Resonance (NMR) was used to measure T_1 and T_2 of water solutions with increasing $MnCl_2$ concentration (range 0.1 - 1.2 mM, in steps of 0.1 mM). The measurements were performed at room temperature (25°) with an operating proton resonance frequency of 63.86 MHz, corresponding to a 1.5 T static magnetic field, the same used for T_1 and T_2 mapping *in vivo*. T_1 (longitudinal relaxation time) measurements were performed with a Saturation Recovery (SR) Spin-Echo (SE) sequence, composed of three saturation $\pi/2$ pulses - aimed at saturating the absorption line - followed by an SE sequence [6] for reading the signal. T_2 (transverse relaxation time) was evaluated with a Carr-Purcell-Meiboom-Gill (CPMG) sequence [7], [8]. The physical principles of these sequences were explained in Chapter 1. The time between $\pi/2$ pulse and the first pulse, referred to as τ, was chosen as short as possible to reduce the effects of diffusion and a high number of echoes was used in order to follow the entire relaxation process. Sequences diagrams and parameters are reported in Figure 1.8 and Table 4.3. The NMR

Table 4.3: Parameters set on the NMR spectrometer for the SR - SE and CPMG sequences.

SR - SE					**CPMG**		
$(\pi/2)_{sat}$	τ_{sat}	$\pi/2$	τ_{echo}	π	$\pi/2$	τ	π
5.3 μs	200 ms	5.1 μs	1 ms	10.2 μs	5.1 μs	50 μs	10.2 μs

raw data were analysed with the in-house QtNMR software and with OriginPro (version 8.1, OriginLab Corporation, Northampton, USA) software. The signal from the SR-SE experiment, performed on each sample with different $MnCl_2$ concentration, describes the temporal evolution of the longitudinal nuclear magnetisation $M_z(t)$, as in Equation 1.16. From this equation, the following quantity can be defined:

$$e^{-\frac{t}{T_1}} = \frac{M_0 - M_z(t)}{M_0} \equiv Y \tag{4.1}$$

that was fitted with a mono-exponential function to evaluate the spin-lattice relaxation time T_1.

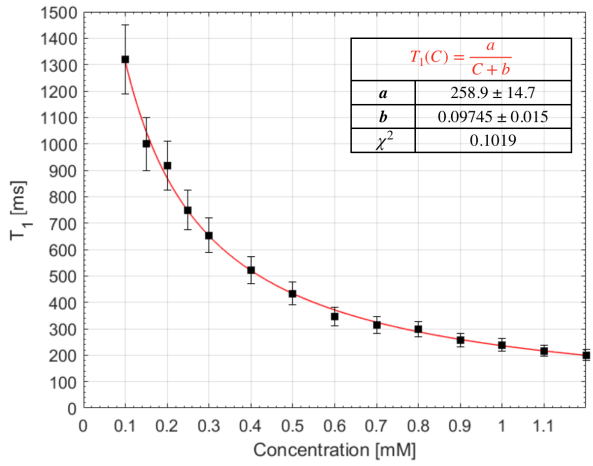
Similarly, the signal form the CPMG (Equation 1.17) describes the temporal evolution for the transverse magnetization M_{xy}. A mono-exponential fit function was used to evaluate T_2 from the experimental data. Combining the errors due to electronics and software in the data collection phase with the errors on model fitting, the maximum total a priori error was estimated to be 10% of the measured values.
The relaxation times obtained from the samples were plotted as a function of the $MnCl_2$ concentration in order to build two calibration curves, one for T_1 and one for T_2. The two set of experimental points were fitted with the function:

$$T_i\left(C\right) = \frac{a}{C+b} \tag{4.2}$$

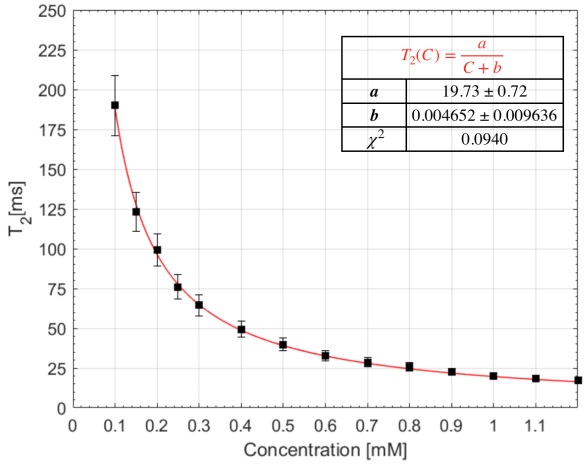
with $i = (1, 2)$, C as the concentration and (a, b) as fitting parameters. The optimal fit for the experimental data was obtained with a MATLAB script developed in-house. The concentration - relaxation time curves were used for the selection of concentrations that produce relaxation times matching the values measured for tissues in the previous phase of the work.
The obtained calibration curves, linking the concentration of the $MnCl_2$ solution with its relaxation times, are reported in Figure 4.2. The chi-square values confirm that the data are well described by the hypothesized model.

4.1.3 Phantom design and assembly

To approximate the shape of the pelvis, the NEMA IEC Body Phantom Set^{TM} (Spectrum Corporation, Durham, USA) was used as an abdomen-shaped container made of MR compatible materials. This phantom is routinely used for quality assurance on radionuclide imaging devices and is constructed entirely of solid plastics. It consists of an external plastic container, that serves as a background compartment, and an empty internal cylinder exploited to assemble PETER PHAN.
Two versions of the phantom were prepared: version one (V1) in which the T_1 relaxation time of tissues were reproduced, and version two (V2), with matching of T_2 relaxation times. In each version, the background compartment of the phantom was filled with a solution of $MnCl_2$ to reproduce the relevant (T_1 or T_2) relaxation time of muscle. To reproduce the lesion, which represents the target for radiomic studies, four cylindrical inserts (48 mm diameter, 72 mm height) were placed inside the phantom. The role of the inserts was twofold: to mimic the MR average signal of a tumour and to provide a selection of textures for identifying a possible simulant for tumours. These paired demands were fulfilled by filling the inserts with a mixture of polystyrene spheres and agar gel, using the recipe introduced for the PSAG phantoms [109] as a starting point. The four inserts for each phantom were prepared with a 0.1% water solution of agar (Sigma-Aldrich, St. Louis, USA) which was heated to $100°$ and poured over the spheres, adding 1 ml of 0.1% NaN_3 per 1 L of agar as antiseptic. The first and second inserts (INS1-INS2) were filled with the medium-sized (3-4 mm) spheres, INS3 contained only small (1 mm) spheres and INS4 was prepared with a mixture of spheres of small, medium and large (7-8 mm) diameter spheres.
The difference among the inserts of phantom V1 and V2 was in the amount of spheres



(a)



(b)

Figure 4.2: (a) T_1 and (b) T_2 relaxation times for solutions with different $MnCl_2$ concentrations. Black dots are the experimental data. The red line represents the fitted model. The χ^2 values quantify the goodness of the fit.

filling the inserts. The number of spheres per unit volume affected the T_1 and T_2 values of the spheres and gel mixture, so it was varied experimentally until the contrast between the T_1 (or T_2) values of the inserts and the background compartment solution (the last reproducing the muscle T_1 or T_2) matched the contrast between tumour tissue and muscle observed in patients' images. This procedure was chosen due to the fact that a direct measurement of T_1 and T_2 of the mixture in the NMR spectrometer was not feasible. The four inserts were fixed to the main NEMA cylinder in a central position inside

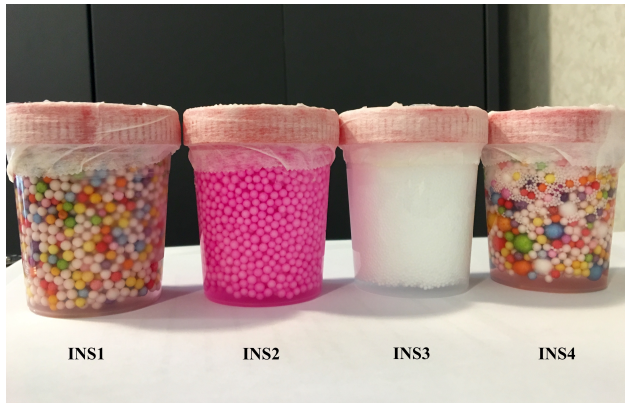
the phantom, via a purpose-built plastic support.
To test the long-term stability of the phantom inserts, six months after their production PETER PHAN V1 and V2 were imaged on the same MR scanner with the same imaging parameters of T_1-w and T_2-w sequences described in the following paragraph. Between evaluations, the inserts were kept refrigerated at 4 °C. The visual inspection showed no deterioration of the object structures, confirmed by the comparison of the phantom images at the time of the main experiment and after six months. Additional evaluations on the phantom stability for longer periods of time are ongoing.
Pictures in Figure 4.3 show the inserts used to simulate lesions with different texture (V2) and the assembled PETER PHAN. Based on the fitted model of T_1 relaxation as a function of $MnCl_2$ concentration, a 0.15 mM $MnCl_2$ solution was chosen to fill the background of PETER PHAN V1 to reproduce the muscle T_1. Similarly, the phantom V2 was prepared with a 0.4 mM solution of $MnCl_2$ to reproduce the muscle T_2.

4.1.4 Validation

The validation process included two verifications: (i) check the agreement between the relaxation times obtained in the phantom and the *in vivo* relaxation times; (ii) verify the compatibility of the textures offered by the phantom inserts with the ones observed in the patients' lesions.

Relaxation times

To meet the first task, T_1 and T_2 mapping acquisitions were performed on PETER PHAN V1 and V2 respectively, in order to verify that the phantom adequately reproduced the range of relaxation times of muscle and tumour tissue. The relaxation time mapping acquisitions were performed with the same MRI sequences and settings as used above for *in vivo* T_1 and T_2 mapping. T_1 and T_2 maps of PETER PHAN V1 and V2 can be seen in Figure 4.4. The mean relaxation time values inside the four inserts and in the surrounding background are summarised in the lower part of Table 4.2. For the background solution representing the muscle, the values are the average of the results obtained on four identical ROIs identified in different positions on the fluid-filled region to assess the homogeneity of the relaxation time maps. Comparing the phantom with patients' results, the agreement between muscle and $MnCl_2$ is within 13% for T_1 mean values and 12% for T_2 mean values. The measurements performed on the inserts showed the following absolute percentage variations (T_1, T_2) with respect to the tumour mean value: INS1 (17%, 46%), INS2 (18%, 31%), INS3 (7%, 24%) and INS4 (12%, 28%).
Additionally, PETER PHAN V1 and V2 were imaged with the routine T_1-w and T_2-w sequences used for clinical pelvis diagnostic imaging. More in detail, the acquisition of PETER PHAN V1 was performed with a series of axial T_1-w images, with TR 354 ms, TE 8.8 ms, slice thickness 5 mm, spacing 5 mm, 512x512 matrix, whereas PETER PHAN V2 was acquired with an axial T_2-w sequence, with TR 3723 ms, TE 107 ms, slice thickness 5 mm, spacing 6 mm, matrix 512x512. Such images were visually compared with the clinical images of patients listed in Table 4.1, acquired on the same scanner with the following parameters: TR 541 ms, TE 8.1 ms, slice thickness 5.5 mm, spacing 5 mm, matrix



(a)

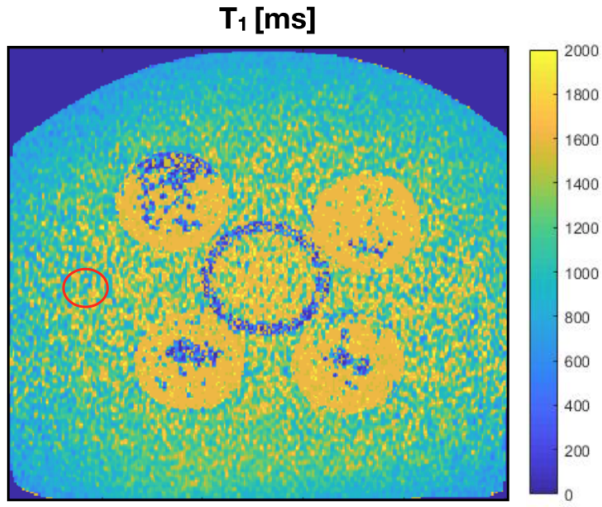
(b)

Figure 4.3: (a) Inserts prepared for PETER PHAN V2, representing tumours with various textures. (b) Axial view of PETER PHAN assembled.

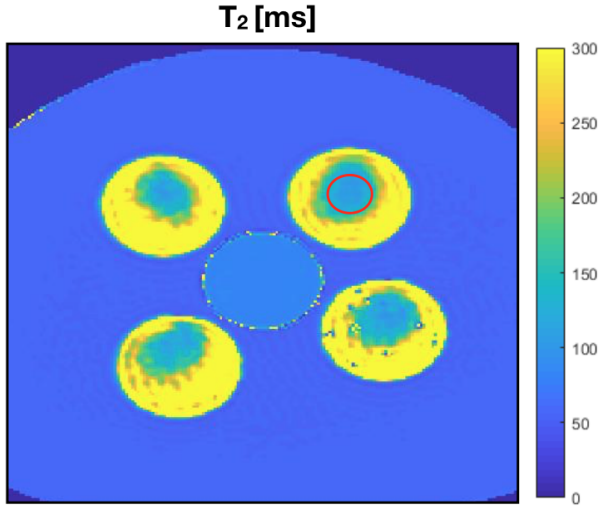
512x512 for the T_1-w acquisition; TR 7771 ms, TE 105 ms, slice thickness 5 mm, spacing 5.5 mm, matrix 512x512 for the T_2-w series. Figure 4.5 offers a comparison of images of the phantom and a patient obtained with the clinical T_1-w and T_2-w sequences. These results showed that the inserts well reproduced a typical lesion in terms of its relaxation times.

Texture properties

The second verification consisted in a quantitative comparison of the texture features in the patients' lesions and the phantom inserts. PETER PHAN images acquired with the T_1-w and T_2-w sequences of the clinical protocol described in the previous paragraph were exploited for a preliminary radiomic assessment. Sixteen cylindrical VOIs of four



(a)



(b)

Figure 4.4: (a) T_1 map of PETER PHAN V1. An example of VOI in the background compartment is highlighted with a red contour. (b) T_2 map of PETER PHAN V2. An example of VOI drawn inside one of the inserts is highlighted with a red contour.

sizes were drawn on the four phantom inserts for the first acquisition images (the same procedure for the VOIs identification was explained in detail in Chapter 3) using 3D Slicer [95] (version 4.10.1, Brigham and Women's Hospital, Harvard University, Boston, USA) and considering three consecutive slices. *PyRadiomics* software [26] (version 2.2.0, Brigham and Women's Hospital, Harvard Medical School, Boston, USA) was exploited for images normalisation and to extract the radiomic features. The normalisation func-

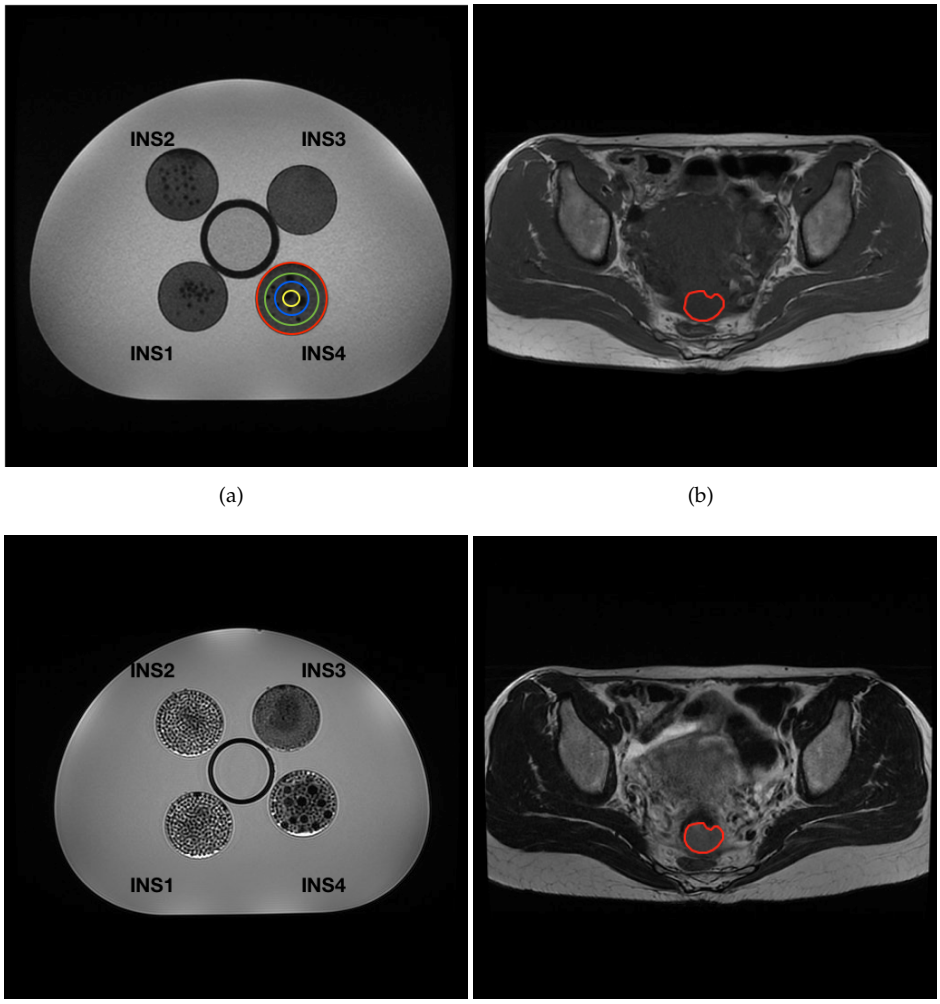


Figure 4.5: (a-b) T_1-w image of PETER PHAN V1, representing tumours with various textures, and of a patient with cervical cancer. In (a) selected VOIs of four sizes identified on the phantom inserts for radiomic analysis are shown (yellow d = 12 mm; blue d = 24 mm; green d = 36 mm, red d = 48 mm). In (b) the tumour site in highlighted in red. (c-d) T_2-w image of PETER PHAN V2 and of the same patient.

tion was

$$f(x) = \frac{s\,(x - \mu_x)}{\sigma_x} \tag{4.3}$$

where x is the original intensity, $f(x)$ the normalised intensity, μ_x and σ_x the mean and the standard deviation of the image intensity values, s is a scaling factor which was set to 100 in this study. The radiomic features extracted from each VOI belonged to the categories: Shape, First Order, GLCM, GLRLM, GLSZM, NGTDM, GLDM. The extraction

was performed on both original images and images filtered with the following kernels: Laplacian of Gaussian, Wavelets, Exponential, Logarithm, Square and Square Root. The filter width σ used for the Gaussian kernel was set to 6.0 mm and one level was used for the wavelet decompositions. The fixed-bin width was chosen as grey level discretisation technique. The bin width was optimised for each extraction to obtain a number of bins in the range (30-130). Its value was set to 2 and 8 for the T_1-w and T_2-w images of the phantom respectively.
Following an approach similar to the one proposed by Samei et al. [110], the range of the features values in the phantom and in a set of patients was compared. The procedure described for extracting the features from the phantom images was used to extract the radiomic features from the tumour volume on the patients' images (both T_1-w and T_2-w acquisitions. Patients listed in Table 4.1), with 3 as bin width. A statistical test of compatibility between the two samples (the values extracted from the phantom images and those obtained from the patients' images), was performed to test the hypothesis that the two populations have equal means.
After verifying that the data were not normally distributed using the Shapiro-Wilk test [99], the non-parametric Mann-Whitney U test [111] was performed to detect whether the features extracted from the phantom and the patients had the same distribution. The null hypothesis was that the distribution of a feature was the same in both samples. Out of 944 features, the Mann-Whitney U test confirmed the null hypothesis in 456 (T_1-w) and 359 (T_1-w) features. In these cases, the inserts were representative of the patients' lesions in terms of its texture properties.
As an example, the distribution of five representative features of different classes are represented as a box plot in Figure 4.6 (PETER PHAN V1) and Figure 4.7 (PETER PHAN V2). These features are 'Energy' (first-order), 'RunLengthNonUniformity' (GLRLM), 'GrayLevelNonUniformity' (GLSZM), 'SizeZoneNonUniformity' (GLSZM) and 'DependenceNonUniformity' (GLDM). As shown in the plots and confirmed by the Mann-Whitney U test, the radiomic features in the phantom inserts were in agreement with the features extracted from the patients' lesions.

4.1.5 Discussion

In developing PETER PHAN, the first phase of the work was the measurement of the relaxation times of both muscle (normal tissue) and tumoural tissue in the case of female patients affected by a pelvic tumour. The maps performed on patients and healthy volunteers indicate a muscle (normal tissue) T_1 to be in the range 806-1378 ms, and T_2 in the range 40-65 ms. As mapping of T_1 and T_2 is not commonly included in the clinical protocols, T_1 and T_2 reference values are limited and not easily available. However, a comparison can be made with the values published by Kato et al. [112], that for muscle are in the range 797-1206 ms for T_1 and 31-47 ms for T_2. For the tumour the results obtained in this work are in the range of 1396-2121 ms for T_1 and 79-101 ms for T_2. The values reported in the same reference for miscellaneous tumours and fibrosarcoma are in the range 1011-1083 ms for T_1 and 65-87 ms for T_2. The range indicated do not overlap with the results of the present study and the reason can reside in the different nature of

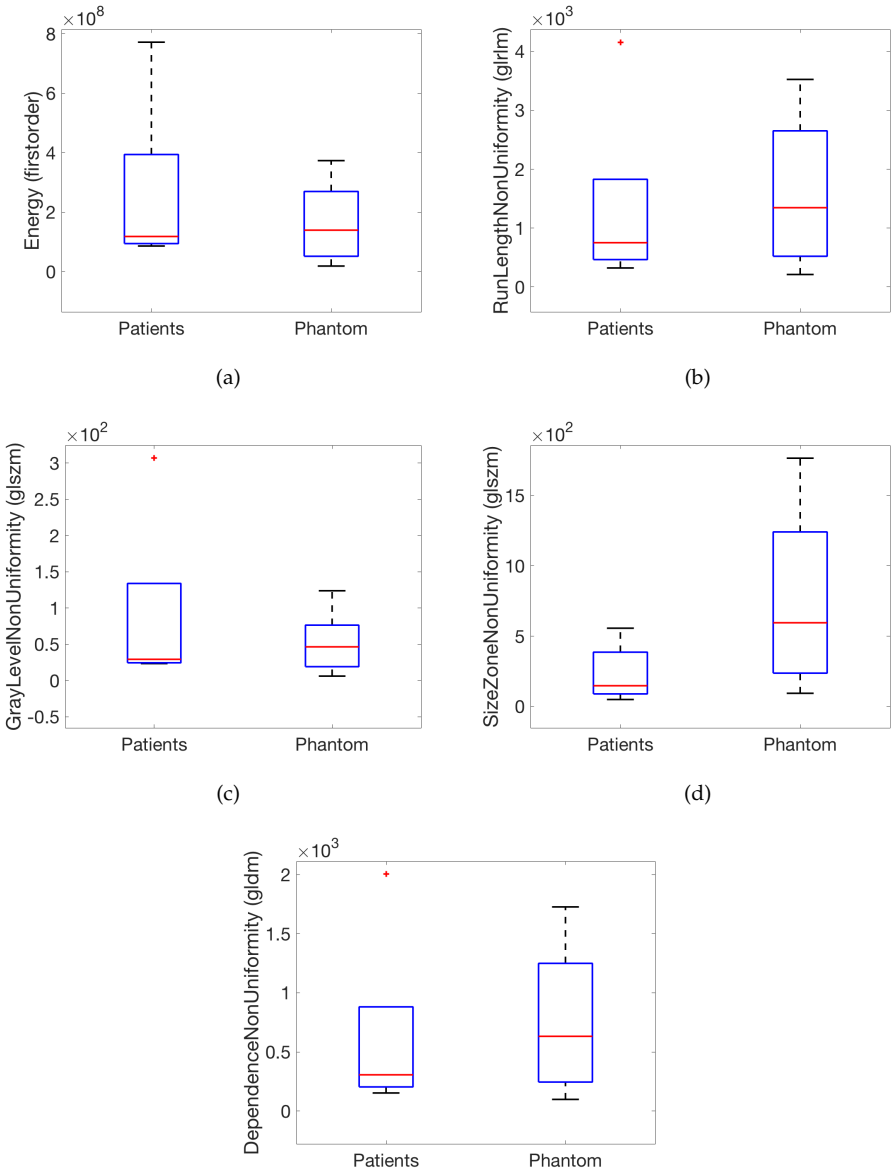


Figure 4.6: Each plot (a-e) shows the comparison of the selected radiomic features values extracted (left) from the tumour site of the patients and (right) from the sixteen VOIs identified on the phantom inserts. All the images were acquired with a T_1-w sequence typically included in the clinical protocol for pelvis diagnostic imaging. On each box, the red line indicates the median, and the bottom and top edges of the blue box indicate the 25^{th} and 75^{th} percentiles, respectively. The black whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually in red using the '+' symbol.

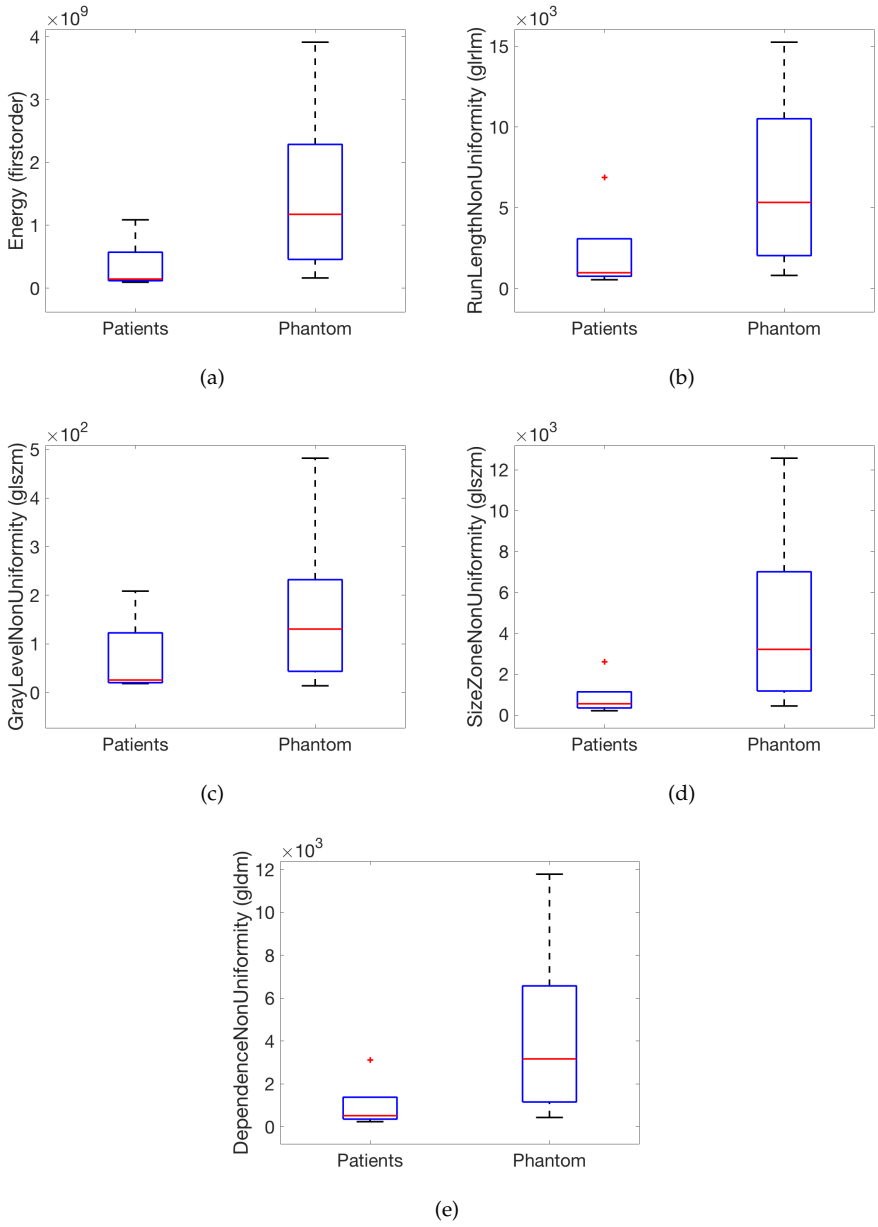


Figure 4.7: Each plot (a-e) shows the comparison of the selected radiomic features values extracted (left) from the tumour site of the patients and (right) from the sixteen VOIs identified on the phantom inserts. All the images were acquired with a T_2-w sequence typically included in the clinical protocol for pelvis diagnostic imaging. On each box, the red line indicates the median, and the bottom and top edges of the blue box indicate the 25^{th} and 75^{th} percentiles, respectively. The black whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually in red using the '+' symbol.

the tumours considered for the analysis. Precisely because this work was specifically focused on pelvis tumours, it was decided to measure the relaxation times on patients to obtain the reference values for the phantom development.

The $MnCl_2$ solution concentrations chosen to simulate the muscle tissue resulted in T_1 and T_2 values within 13% and 12% of the corresponding average *in vivo* values. In addition, inserts with different textures were incorporated to mimic both the mean signal and the texture observed for lesion during *in vivo* clinical imaging. INS3, consisting of 1 mm diameter polystyrene beads and agar, was the closest in reproducing both these properties, as can be also seen visually in both T_1-w and T_2-w images shown in Figure 4.5. Although INS1, INS2 and INS4 were characterised by textures less similar to the clinical images considered in this study, they were introduced with the aim to test the radiomic features ability to distinguish different textures. Moreover, they could be useful for reproducing more heterogeneous lesions observed in populations different from the one considered for this study.

Selecting five representative radiomic features, it was proved that they showed the same order of magnitude when extracted from the phantom inserts and the tumour volume of the patients considered. Despite the number of patients used for this assessment was limited, this first quantitative evaluation confirmed that both phantom versions (V1 and V2) adequately reproduce, for the purpose of a radiomic study, not only the average signal of healthy tissue and tumour, but also the texture seen in a female pelvic tumour. Summing up, a phantom that mimics the female pelvis and lesions to inspect radiomic properties was successfully created.

The radiomic investigation performed on the present version of PETER PHAN, as well as its future developments, may allow the exploration of the complex binomial radiomics-MRI in pelvis imaging, contributing to the achievement of a robust methodology to be included in the clinical research. PETER PHAN could be easily assembled and acquired on different scanners and/or at different centres, in order to compare the radiomic response in relation to the scanner vendor, scanner type and magnetic field strength. The radiomic features extracted from the phantom images will be studied not only for their repeatability but also reproducibility, potential dependence on the MR sequence parameters and on the process of image acquisition in its entirety. Different approaches in image processing will be compared (algorithms for image normalisation or correction for magnetic field inhomogeneity) and different options for the choice of parameters in the features extraction phase will be tested. The possibility to probe the relationship between the VOI volume and the features stability is offered along with the chance to assess the features ability to discriminate between different textures. Parts of these tasks were addressed in this work thesis and will be presented in the second part of this Chapter. PETER PHAN represents also a test object to be used for a comparison in the performance of platforms available for the extraction of the radiomic features, as seen in the previous Chapter.
Despite it was originally thought for reproducing the female pelvis, the phantom can

easily be extended to the application in male pelvis studies, upon dedicated T_1 and T_2 measurement of male pelvis tumours, or more generally to other cancers occurring in the pelvis.

A limitation of the phantom is that it was separately optimized for T_1 and T_2. The combined use of two paramagnetic agents may allow the simultaneous matching of T_1 and T_2, which would simplify the use of the phantom. Modifying the relaxation times of the agar gels with similar relaxation agents may also improve the modest agreement with *in vivo* tumour values, seen for T_2 values of interest. Instead, it was decided to match image contrast by adjusting the concentration of the polystyrene spheres.
Future versions of the phantom could include inserts with combinations of other sizes of spheres, in order to widen the range of textures reproduced and cover the variability seen in patients. In operational terms, starting from the partial overlap shown in Figure 4.6, a new edition of PETER PHAN V1 will aim at reaching a complete overlap of the texture features range offered by the phantom inserts with the range seen in the patients' lesions. Furthermore, the number of patients considered for this evaluation will be increased to make the phantom more representative of the patients' population. From this perspective, also PETER PHAN V2, already showing an optimum overlap with the patients considered, could be optimised.
Additionally, in a recent paper, R. Rai et al. [113] presented a work about an MR texture phantom produced with 3D printing technology, which could be considered for further development and refinement on PETER PHAN inserts, guaranteeing the possibility to realise a variety of tumour-like shapes and textures, in a more reproducible way. The employment of this strategy requires additional investigation on the relaxation times of the materials used as printer ink, in order to develop a potential material (not yet available) that properly reproduces the relaxation of the tumoural tissue of our interest as achieved, despite in a hand-crafted fashion, in PETER PHAN.

4.2 Radiomics experiments on PETER PHAN

This Section has been adapted from [114].

In this Section, the results of the radiomics experiments performed on PETER PHAN are presented. This part of the thesis work was carried out in collaboration with the European Institute of Oncology (Milan) and the Computational Center for the Unknown, Champalimaud Foundation (Lisbon).

4.2.1 Study design

In a clinical study, the database used for a radiomic investigation can include images acquired on different scanners and/or with different imaging protocols. The robustness of the radiomic features in some of these possible scenarios was investigated.
In this multicenter study (schematized in Figure 4.8), three scanners of two manufactures and two magnetic field strengths were used to assess repeatability and reproducibility

of radiomic features. Phantom images were acquired in a test-retest study to investigate the repeatability of the features on each scanner. The acquisitions were performed with a 2D T_2-w sequence, optimised on each scanner for pelvic imaging. Between the two repeated scans, the phantom was either kept fixed or repositioned, to evaluate also the repeatability after repositioning. Additionally, repeatability was investigated on 3D features extracted from images acquired using a 3D T_2-w sequence on scanner B, to interpret and strengthen the result obtained for the repeatability of features on each scanner for 2D acquisitions.

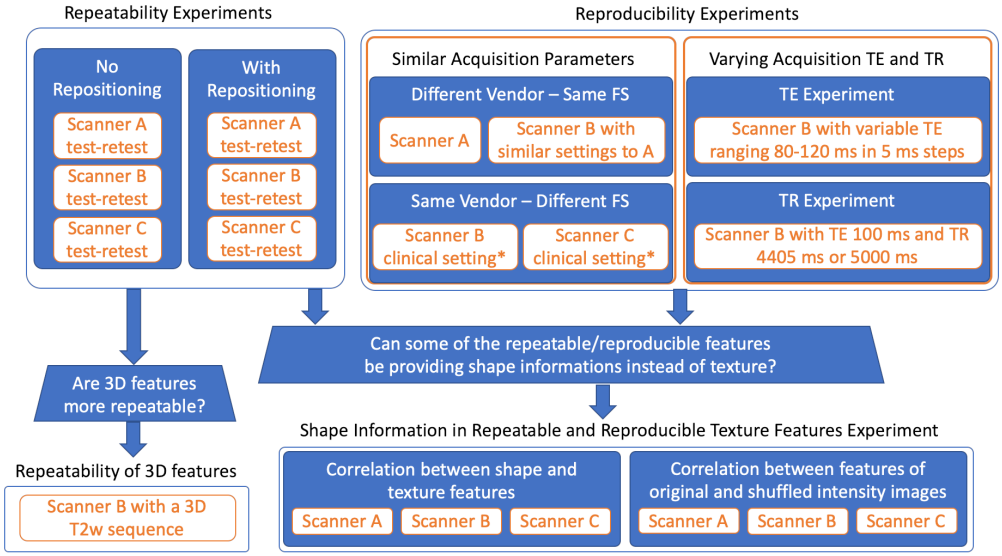


Figure 4.8: Study design schematic showing repeatability and reproducibility experiments and additional experiments executed to bring understanding to the repeatability and reproducibility results. * - clinical T_2-w MRI sequence was shared between scanner B and C. FS refers to the magnetic Field Strength.

The reproducibility of the radiomic features was evaluated in three different scenarios: (1) reproducibility at fixed imaging parameters on two scanners of same field strength, but different manufacturers; (2) reproducibility at fixed imaging parameters on two scanners of the same manufacturer, but different field strengths (1.5 T and 3 T); (3) reproducibility with varying TE and TR imaging parameters on a fixed scanner. Study (1) was performed on scanner A (1.5 T Optima MR450W, General Electric Healthcare, Waukesha, USA) and scanner B (1.5 T Ingenia, Philips Healthcare, Best, the Netherlands). The sequence parameters used for pelvic diagnostic imaging on scanner A were replicated on scanner B, and the phantom was imaged on both scanners. The replication of the same sequence in scanners of different vendors was not exact, as MR sequences are vendor-specific, and not all the parameters are accessible to users. However, considering that some radiomics retrospective studies made use of images acquired on different scanners, a similar situation in ideal conditions (in phantom study) was mimicked to quantify the

robustness of features in this scenario and to give useful indications for prospective studies. Study (2) was carried out on scanner B and C (3 T Ingenia, Philips Healthcare, Best, the Netherlands). In this case, it was possible to replicate the same sequence parameters on both scanners. Doubling the magnetic field strength, the MR signals - and, thus, the images obtained - are different, as they are influenced by the relaxation times T_1 and T_2, which depend on the field strength. As a consequence, the radiomic features were expected to vary as well. However, this scenario could be part of a clinical radiomic study and would necessitate the assessment of the reproducibility of features between field strengths. Study (3) was conducted on scanner B (which offered more accessibility due to scheduling and technical reasons), varying TE or TR in the range commonly used when imaging patients for pelvic investigations.
Besides these assessment, the correlation between texture and shape features was evaluated, to identify whether the excellent performance in terms of repeatability and reproducibility might be ascribable to high correlation with shape information rather than robust quantification of a texture property.

4.2.2 Methods

2D repeatability and reproducibility between scanners

Images acquisition and segmentation T_2-w images of the phantom were acquired on scanners A, B, and C, using different pelvic imaging setups. The parameters of the MR sequences are listed in Table 4.4. To study the repeatability on scanners B and C, and the reproducibility between these scanners, the filling solution was replaced with oil (Spectrasyn 4 phantom oil, Philips Heathcare, Best, the Netherlands - T_1=230 ms and T_2=194 ms at 3T) that has a lower dielectric constant than water allowing to avoid dielectric artifacts intensified at higher fields. To discuss the reproducibility results, contrast-to-noise ratio (CNR) was calculated using equation 4.4 [115], [116], where the mean signal intensity of A, μ_{SI_A}, was measured in the green region of interest (ROI) in Figure 4.9 (insert content), the mean signal intensity of B, μ_{SI_B}, was measured in the yellow ROI (main phantom compartment) and the standard deviation of the signal intensity of background air, $\sigma_{SI_{air}}$, was measured in the red ROI. CNR values are reported in Table 4.4.

$$CNR = \frac{\mu_{SI_A} - \mu_{SI_B}}{\sigma_{SI_{air}}} \tag{4.4}$$

To test the short-term repeatability of the radiomic features, the acquisition was repeated twice on each scanner without changing the setup nor moving the phantom. Then, the phantom was removed and repositioned, and the acquisition was repeated. Sixteen cylindrical VOIs of four sizes were drawn on the phantom inserts for the first acquisition images using 3D Slicer [95] version 4.10.1, with the same procedure explained in the first part of the Chapter for the phantom validation measurements. Each VOI was drawn on a total of three consecutive slices. The VOIs for the images acquired after phantom repositioning were obtained by applying to the original VOIs segmentation the rigid transformation that allowed to align the initial images with the images of the repositioned phantom.

Table 4.4: Scanners properties and imaging sequence parameters. TR = Repetition Time; TE = Echo Time; CNR = Contrast-to-Noise ratio; FoV = Field of View.

Scanner	Field strength [T]	Vendor	Experiment	Acquisition	TR [ms]	TE [ms]	CNR	Slice thickness [mm]	Slice spacing [mm]	Pixel spacing [mm]	FoV [mm x mm]	Averages
A	1.5	GE	Repeatability A/ Reproducibility A vs. B	2D	4763	109	64.90	5	5.5	0.6 x 0.6	320 x 320	1
B	1.5	Philips	Reproducibility A vs. B	2D	4700	110	24.71	5	6	0.6 x 0.6	320 x 320	1
			Repeatability B 3D	3D	1050	160	455.52	1	1	1 x 1	320 x 320	2
			Reproducibility TE/TR	2D	5000	80	24.47	5	6	0.6 x 0.6	320 x 320	1
						85	23.92					
						90	24.67					
						95	22.83					
						100	19.16					
						105	14.89					
						110	12.15					
						115	12.51					
						120	12.28					
					4405	100	21.84					1
			Reproducibility B 2D/ Reproducibility B vs. C	2D	3750	90	58.86	5	5	0.6 x 0.6	340 x 340	1
C	3	Philips	Repeatability C/ Reproducibility B vs. C	2D	3750	90	176.55	5	5	0.6 x 0.6	340 x 340	1

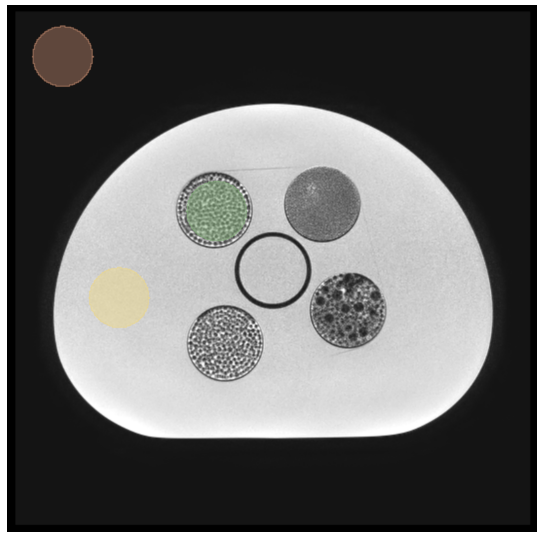


Figure 4.9: Representation of ROIs placement for calculation of contrast-to-noise ratio. Green ROI was used to measure the mean signal intensity of region A, the yellow ROI was used to measure the mean signal intensity of region B, and the red ROI was used to determine the standard deviation of the signal intensity of background air region.

Extraction of radiomic features The package *PyRadiomics* [26] version 2.2.0 was used to normalise the images and to extract 2D radiomic features (included categories: Shape, First Order, GLCM, GLRLM, GLSZM, NGTDM, and GLDM) from each VOI, on both original and filtered images (Laplacian of Gaussian - LoG -, Wavelet, Square, Square Root, Logarithm, and Exponential). The set of parameters used for the extraction were the same described in paragraph 4.1.4. *PyRadiomics* normalisation was applied by setting the images mean signal intensity to 300 and standard deviation to 100. This normalisation was used to reduce the effect of different MR image intensity ranges on different scanners due to the lack of a standard intensity scale in MRI. The bin-width was optimized for each extraction in order to obtain a number of bins in the range 30 to 130, as explained earlier. In Table 4.5, the main settings used for the feature extraction are listed. The corresponding parameter files are available in the dedicated GitHub repository. In Tables 4.6 and 4.7, the list of features extracted from each category is reported.

Repeatability and reproducibility assessment The repeatability of the radiomic features was quantified by computing the Intraclass Correlation Coefficient (ICC) [35], [117], [36]. ICC (2, 1), equation 4.5, was calculated pairwise (between repeated acquisitions) for each radiomic feature to test repeatability, with and without phantom repositioning.

$$ICC(2,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} \tag{4.5}$$

The suffix (2, 1) indicates the ICC form computed considering two-way random effects for absolute agreement and single rater/measurement. In equation 4.5, MS_R corre-

Table 4.5: Settings in the parameter files used for the radiomic features extraction with the package *PyRadiomics*.

Setting	**Value**
normalize	true
normalizeScale	100
preCrop	true
force2D	true (2D extraction) false (3D extraction)
force2Ddimension	0 (2D extraction only)
geometryTolerance	1.e+4
binwidth	10 (scanner A) 5 (scanners B and C)
voxelArrayShift	300
imageType	Original: LoG: 'sigma': [6] Wavelet: 'level': 2 Square: SquareRoot: Logarithm: Exponential:
featureClass	Shape: firstorder: glcm: glrlm: glszm: ngtdm: gldm:

sponds to the mean square for VOIs, MS_E corresponds to the mean square for error, MS_C corresponds to the mean square for repeated measures, n is the number of VOIs, and k is the number of repeated acquisitions.
To assess reproducibility, the concordance correlation coefficient (CCC) [101] was calculated pairwise for the features extracted from the images acquired on two different scanners (scanner A vs. B, scanner B vs. C), with the same imaging sequence. In equation 4.6, σ_1^2 and σ_2^2 are the variances of a feature for each acquisition, μ_1 and μ_2 are the feature means, and ρ_{12} is the correlation coefficient between the acquisitions.

$$CCC = \frac{2\sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} \tag{4.6}$$

The features were classified into four groups based on the ICC or CCC values (Table 4.8). The most robust features for the particular scenario considered (across three scanners, including two manufacturers and two magnetic field strengths) were identified by intersecting the sets of features showing both excellent repeatability and excellent repro-

Table 4.6: List of radiomic features per category extracted with the package *PyRadiomics*. Shape features were extracted from the original images only, whilst the features of the other categories were extracted from both the original and filtered images. The definition of the features listed in this table is available on the *PyRadiomics* website.

Shape	**First Order**	**GLCM**	**GLRLM**
Elongation	10Percentile	Autocorrelation	GrayLevelNonUniformity
Flatness	90Percentile	ClusterProminence	GrayLevelNonUniformityNormalized
LeastAxisLength	Energy	ClusterShade	GrayLevelVariance
MajorAxisLength	Entropy	ClusterTendency	HighGrayLevelRunEmphasis
Maximum2DDiameterColumn	InterquartileRange	Contrast	LongRunEmphasis
Maximum2DDiameterRow	Kurtosis	Correlation	LongRunHighGrayLevelEmphasis
Maximum2DDiameterSlice	Maximum	DifferenceAverage	LongRunLowGrayLevelEmphasis
Maximum3DDiameter	MeanAbsoluteDeviation	DifferenceEntropy	LowGrayLevelRunEmphasis
MeshVolume	Mean	DifferenceVariance	RunEntropy
MinorAxisLength	Median	Id	RunLengthNonUniformity
Sphericity	Minimum	Idm	RunLengthNonUniformityNormalized
SurfaceArea	Range	Idmn	RunPercentage
SurfaceVolumeRatio	RobustMeanAbsoluteDeviation	Idn	RunVariance
VoxelVolume	RootMeanSquared	Imc1	ShortRunEmphasis
	Skewness	Imc2	ShortRunHighGrayLevelEmphasis
	TotalEnergy	InverseVariance	ShortRunLowGrayLevelEmphasis
	Uniformity	JointAverage	
	Variance	JointEnergy	
		JointEntropy	
		MCC	
		MaximumProbability	
		SumAverage	
		SumEntropy	
		SumSquares	

Table 4.7: List of radiomic features per category extracted with the package *PyRadiomics* (continued). Shape features were extracted from the original images only, whilst the features of the other categories were extracted from both the original and filtered images. The definition of the features listed in this table is available on the *PyRadiomics* website.

GLSZM	**NGTDM**	**GLDM**
GrayLevelNonUniformity	Busyness	DependenceEntropy
GrayLevelNonUniformityNormalized	Coarseness	DependenceNonUniformity
GrayLevelVariance	Complexity	DependenceNonUniformityNormalized
HighGrayLevelZoneEmphasis	Contrast	DependenceVariance
LargeAreaEmphasis	Strength	GrayLevelNonUniformity
LargeAreaHighGrayLevelEmphasis		GrayLevelVariance
LargeAreaLowGrayLevelEmphasis		HighGrayLevelEmphasis
LowGrayLevelZoneEmphasis		LargeDependenceEmphasis
SizeZoneNonUniformity		LargeDependenceHighGrayLevelEmphasis
SizeZoneNonUniformityNormalized		LargeDependenceLowGrayLevelEmphasis
SmallAreaEmphasis		LowGrayLevelEmphasis
SmallAreaHighGrayLevelEmphasis		SmallDependenceEmphasis
SmallAreaLowGrayLevelEmphasis		SmallDependenceHighGrayLevelEmphasis
ZoneEntropy		SmallDependenceLowGrayLevelEmphasis
ZonePercentage		
ZoneVariance		

ducibility.

Table 4.8: Classes of stability of radiomic features.

Repeatability/Reproducibility	**ICC or CCC value**
Excellent	ICC or CCC > 0.9
Good	$0.75 <$ ICC or CCC ≤ 0.9
Moderate	$0.5 <$ ICC or CCC ≤ 0.75
Poor	ICC or CCC ≤ 0.5

On this last set of features, a study on their ability to discriminate between different textures was performed. This ability was investigated for different VOI sizes, to test the hypothesis that a texture descriptor may lose its meaning below a certain threshold for the tumour volume (or the VOI volume identified on the inserts in this study). This hypothesis was advanced by Brooks and Grigsby [118] [119], who found that some texture quantifiers depended on the tumour volume on PET images, being unreliable when considering very small tumours. In addition, the dependence on the VOI volume can be intrinsic to the mathematical definitions of some texture features, as underlined by some recent studies [120] [121]. To remove the intrinsic volume dependence on the set of features selected for the texture discrimination study, the corrected algorithms proposed by these authors were applied *a posteriori* to the extracted feature values. After that, the trend of the selected features as a function of the VOI volume was studied. The investigation was carried out on three datasets, namely the test-retest images (without phantom repositioning) acquired on each scanner (A, B and C) for the previous repeatability assessment. Besides the VOIs already described, the radiomic features were extracted from two additional VOIs (per each insert), one smaller and one bigger than the previous ones. Both VOIs were centered in each insert, one with a 6 mm diameter and the other with 60 mm diameter. The former was added to mimic a very small tumour, whilst the latter simulated the case of a VOI including a portion of surrounding tissue in addition to the tumour volume, mimicking an imprecise delineation of the VOI. For this analysis, only INS1, INS3 and INS4 were considered. INS2 was disregarded as it had a texture similar to INS1. All the VOIs considered for the texture discrimination study are shown in Figure 4.10. For each selected feature, the Repeatability Coefficient (CR) [122] was calculated with the formula

$$CR = 1.96\sqrt{\frac{\sum_{i=1}^{n}\left(d_{2,i} - d_{1,i}\right)^2}{n}} \tag{4.7}$$

where $n = 18$ is the number of available VOIs, d_1 is the vector of the feature values in the 18 VOIs on the test images and d_2 is the vector of the feature values in the 18 VOIs on the retest images. CR represents the value below which the absolute difference between each feature repeated measure may be expected to lie with a probability of 95%. This metric can be used to assess the ability to differentiate two different inserts, e.g., if, for a certain feature and a certain VOI size, the absolute difference between the values from insert x and y is greater than 2*CR, the feature is considered to be able to differentiate these two inserts.

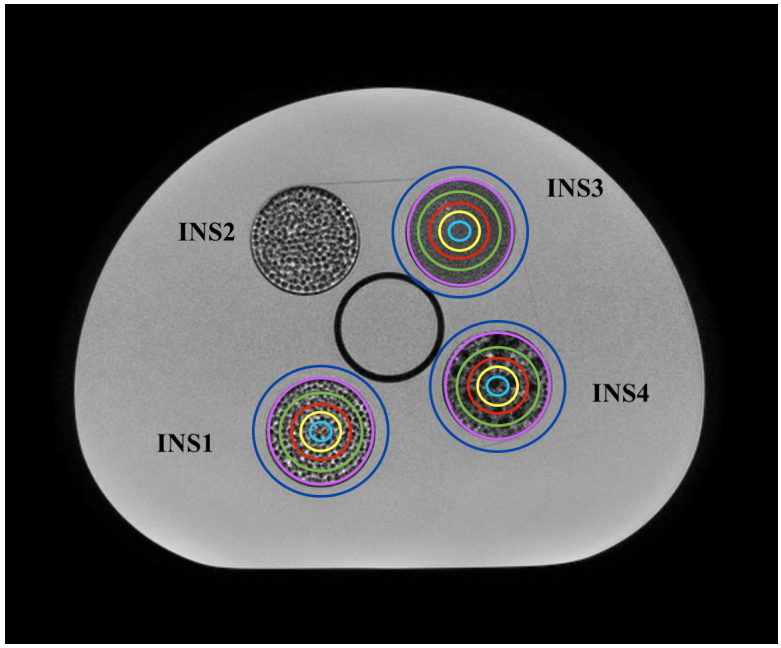


Figure 4.10: VOIs considered for the features extraction in the texture discrimination assessment. The VOIs were drawn on three consecutive slices and had the following volumes: blue 42.3 cm^3, pink 27.1 cm^3, green 15.3 cm^3, 6.8 red cm^3, yellow 1.7 cm^3 and cyan 0.4 cm^3.

Reproducibility under varying TE or TR

In this experiment, the features reproducibility under variation of the TE and TR parameters was investigated. T_2-w images of PETER PHAN were acquired on scanner B using different TE and TR values (Table 4.4), varied in the range of the usual setting for clinical diagnostic imaging. First, the values of TE were varied between 80 and 120 ms, with a step size of 5 ms for a fixed TR of 5000 ms. Then, to investigate the influence of TR in a T_2-w sequence, the Philips scanner mode of TR range (TR: 4000 ms to 6000 ms), which selects the optimal TR within the predefined range, was used to acquire an image with TE of 100 ms and an automatically selected TR of 4405 ms. The phantom inserts were segmented, and the radiomic features extracted as previously described in the other experiments. CCC values were calculated pairwise between the features extracted from images obtained with all possible combinations of TEs and for the images acquired with TR/TE as 5000/100 and 4405/100 to test the features reproducibility.

Repeatability at 3D level

The repeatability and reproducibility of the radiomic features may be affected by the type of feature extraction performed (2D or 3D level). The majority of clinical MRI acquisitions are 2D, and the downsampling to isotropic voxels - to allow the features extraction in 3D - will likely cause a considerable loss of information in the plane of acquisition. In opposition, the upsampling to isotropic voxel will artificially create new

voxel data that may have an impact on the features [18]. Thanks to several technological advances, the acquisition of 3D sequences is becoming a real possibility; therefore, it is interesting to investigate whether 3D features have a better performance in terms of rotation invariance as compared to 2D features, which would result in improved repeatability after phantom repositioning. To investigate this issue, the rotational invariance of 1316 3D features was assessed. The features were extracted from images acquired with a 3D T_2-w sequence on scanner B with an isotropic voxel 1 mm x 1 mm x 1 mm. The imaging parameters are listed in Table 4.4. The difference in the number of extracted features between 2D and 3D is a consequence of the increased number of wavelet decompositions in 3D feature extraction. PETER PHAN was scanned with this sequence twice, repositioning the phantom between the acquisitions. The radiomic features were extracted from the 16 VOIs obtained with the procedure explained in the experiment to assess the 2D repeatability. The extraction was performed at 3D level, disabling the "force2D" function on the *PyRadiomics* parameters file (available in the GitHub repository). The evaluation of the rotational invariance of the features was assessed through the ICC (2, 1), calculated pairwise for each feature, comparing its value in the first and the second acquisition.

Assessment of shape information in non-shape radiomic features

Besides repeatability and reproducibility issues, a recent study showed other vulnerabilities of radiomic features by assessing the performance of previously defined models on the same datasets but where images had their voxel intensities randomly shuffled and found that 3 out of 3 non-shape features constituting the model were actually capturing the tumour volume [123]. Following this result, an investigation on the non-shape features was performed, aiming at identifying the radiomic features that were robust because of their high correlation with shape features, but were not extracting the informative content their mathematical definition was expected to quantify. This behaviour could be due to an intrinsic dependence on the volume in the mathematical definition (as explained before) or, after the correction of this dependence, to an inadequate image quality or an inadequate sampling of the intensity distribution. To this purpose, two consecutive analyses were carried out.
Firstly, the subset of non-shape features that were highly correlated with shape were identified, through a pairwise-correlation analysis between shape and non-shape radiomic features, evaluated with the Spearman correlation coefficient. Non-shape features showing Spearman's correlation above 0.8 were considered as highly correlated with shape features. Other shape metrics besides volume [24] were included, since characteristics like maximum diameter [124], roundness [125] and spiculation [126], [127], among others, are important oncological diagnostic and prognostic factors.
Secondly, starting from the first acquisitions performed on the three scanners (for scanner B the same sequence used for the repeatability study was considered), three additional sets of images were created by shuffling the voxel intensities of each original image. The radiomic features were extracted from the shuffled intensities datasets, following the procedure described above for the segmentation and the calculation of features.

An example of an original image and the corresponding image with randomly shuffled intensities is shown in Figure 4.11. ICC (2, 1) was calculated between the original features and the corresponding features extracted from images that had their intensities randomly shuffled. Features showing ICC > 0.9 were considered texture uninformative since they could not distinguish the original image from the shuffled-intensities one in radiomic terms.
Intersecting the results of the two analyses, it was possible to identify a set of features to exclude *a priori* from an eventual radiomic model since they were both highly correlated with shape and not carrying texture information.

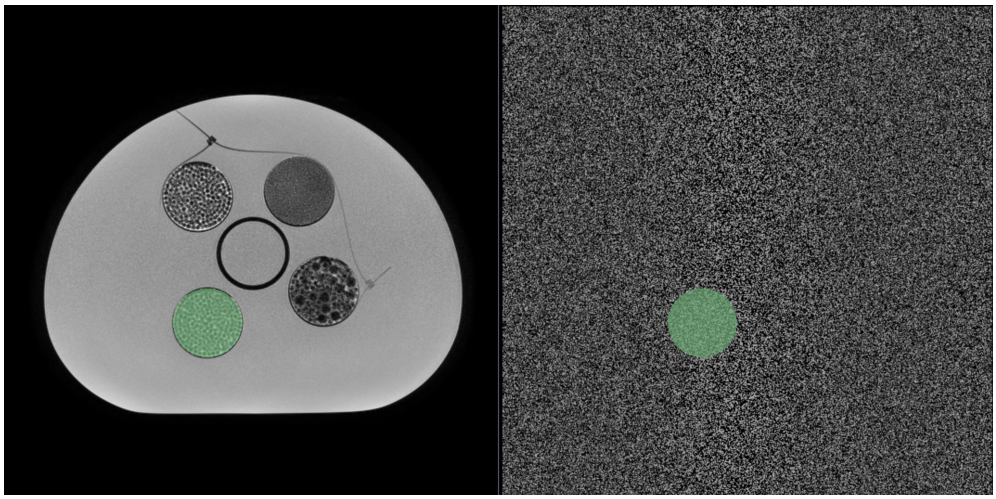


Figure 4.11: Example of original axial image (left) and of corresponding randomly shuffled intensities axial image (right). The green overlay exemplifies one of the 16 cylindrical volumes of interest delineated for the extraction of features.

4.2.3 Results

2D repeatability and reproducibility between scanners

Repeatability assessment A total of 944 features were extracted. Repeatability results for the three scanners, with and without phantom repositioning, are shown in Figure 4.12. Without repositioning, on scanner A, 910 (96.4%) features showed excellent, 29 (3.1%) good, 3 (0.3%) moderate, and 2 (0.2%) poor repeatability. Similar results were obtained on scanner B, with 869 (92.1%), 58 (6.1%), 8 (0.8%), and 9 (1.0%) features showing respectively excellent, good, moderate, and poor repeatability. The features extracted from the images acquired on scanner C showed less repeatability, with the percentage of features with excellent ICC decreasing to 754 (79.9%), and the percentage of features showing ICC ≤ 0.9 increasing to 190 (20.1%). When considering phantom repositioning, a consistent reduction of repeatability was evident across all the scanners. The numbers were 740 (78.4%), 138 (14.6%), 49 (5.2%), 17 (1.8%) (scanner A), 806 (85.4%), 76 (8.1%),

42 (4.4%), 20 (2.1%) (scanner B) and 106 (11.2%), 199 (21.1%), 215 (22.8%), 462 (44.9%) (scanner C) features showing excellent, good, moderate, and poor repeatability.

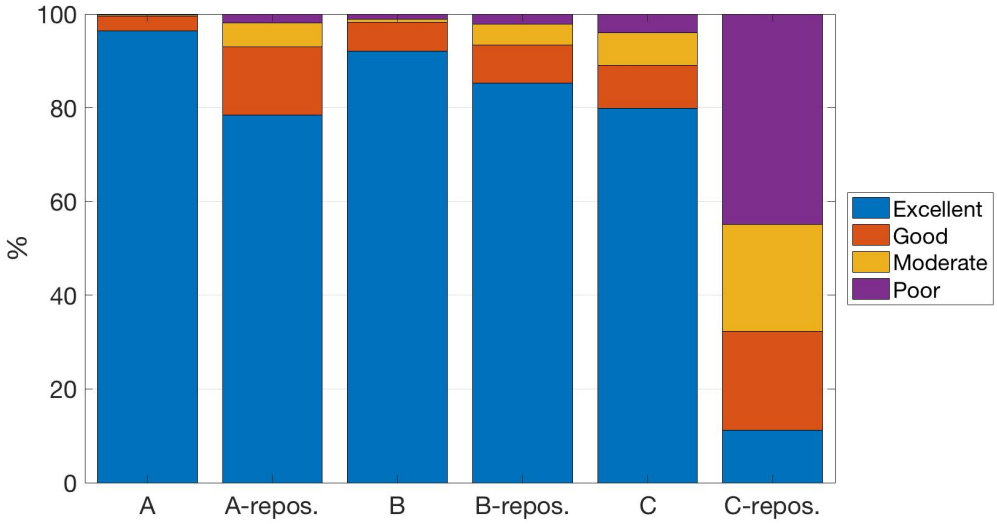


Figure 4.12: Repeatability of radiomic features. The results are reported for repeatability without phantom repositioning on scanners A, B, and C, and with phantom repositioning (repos.).

Reproducibility assessment Reproducibility was assessed both in terms of variations between the features extracted from scanners of different manufacturers and equal magnetic field strength and acquisition parameters (A vs. B, Figure 4.13) and between features extracted from scanners of the same manufacturer but different magnetic field strengths (B vs. C, Figure 4.13). The CCC values of A vs. B and B vs. C, and corresponding confidence intervals are available in the GitHub repository.
The analysis of the features extracted from the images acquired on scanners A and B showed that 830 (87.9%) features had poor reproducibility. Only 43 (4.6%) features exhibited excellent reproducibility; 29 (3.1%) showed good, and 42 (4.4%) moderate reproducibility.
In terms of reproducibility of the features extracted from images obtained with the same sequence parameters on scanners B (1.5 T) and C (3 T), 147 (15.6%), 295 (31.2%), 321 (34.0%), and 181 (19.2%) of them exhibited excellent, good, moderate, and poor reproducibility.

Overall repeatability and reproducibility In order to identify the most stable radiomic features in this study, the ones showing both excellent repeatability and reproducibility were selected. A total of 31 (3.3%) features showed excellent repeatability (ICC > 0.9) and reproducibility (CCC > 0.9) across all different scenarios studied (repeatability: with and without phantom repositioning; reproducibility: across manufacturers and field

strengths). Apart from shape features, expected to be independent of the experiment settings [1], the 20 (2.1%) remaining features are listed in Table 4.9.

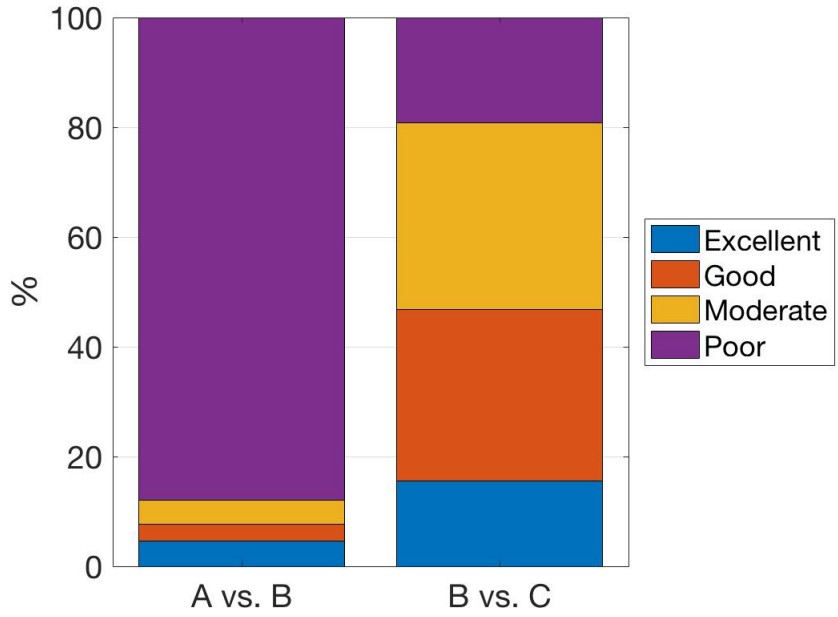


Figure 4.13: Reproducibility of radiomics features for scanners of equal magnetic field strength (1.5 T) and acquisition parameters, but different manufacturers (A vs. B) and for scanners of the same manufacturer but different magnetic field strengths - 1.5 T and 3 T - (B vs. C).

Texture discrimination assessement The analysis of the features ability to distinguish between different textures was performed on the features shown in Table 4.9. Firstly, the features for which the literature evidenced an intrinsic dependence of the feature value on the number of voxels N_v in the VOI were corrected. Specifically, the features "Energy", "Total Energy" and "Run Length Non Uniformity" were divided by N_v, according to the correction proposed by [120]. The feature "Coarseness" was multiplied by N_v, as suggested by [121]. Secondly, the absolute difference of the value of each corrected feature in two different inserts was calculated and compared with the corresponding CR value. The number of features showing absolute differences (between the feature value when extracted from two different inserts) greater than 2*CR are listed in Table 4.10. For the images acquired on scanner A, the majority of the features considered were able to distinguish inserts 1-3 and 1-4, independently of the VOI size. The same holds for scanner B and inserts 1-3. On the contrary, on scanner A the number of features distinguishing inserts 3-4 decreased by decreasing the VOI size. Only 7 out of 20 features extracted

[1]The reproducibility of the shape features shape_Flatness, shape_LeastAxisLength, and shape_Sphericity being 3D features, can be affected by differences in the through-place spacing.

Table 4.9: List of features showing excellent repeatability with and without phantom repositioning and excellent repeatability across manufacturers and magnetic field strengths (Shape features, expected to be independent of the experiment settings, are not reported in this Table). Each feature is indicated in the form: ImageType_Class_FeatureName.

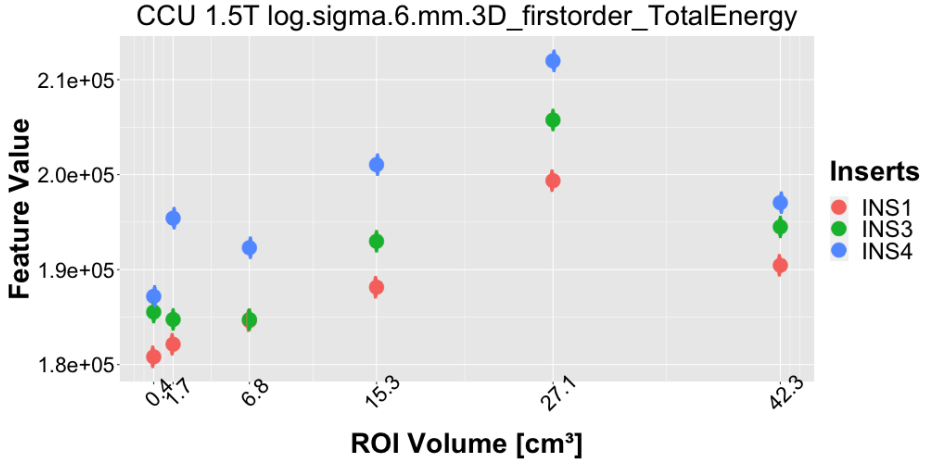
Repeatable and reproducible features
log.sigma.6.mm.3D_firstorder_TotalEnergy
exponential_firstorder_Energy
exponential_firstorder_TotalEnergy
logarithm_glrlm_RunLengthNonUniformity
original_glrlm_RunLengthNonUniformity
square_firstorder_Energy
square_firstorder_TotalEnergy
squareroot_glrlm_RunLengthNonUniformity
wavelet.HH_firstorder_Energy
wavelet.HH_firstorder_TotalEnergy
wavelet.HH_ngtdm_Coarseness
wavelet.HL_firstorder_Energy
wavelet.HL_firstorder_TotalEnergy
wavelet.HL_glrlm_RunLengthNonUniformity
wavelet.HL_ngtdm_Coarseness
wavelet.LH_firstorder_Energy
wavelet.LH_firstorder_TotalEnergy
wavelet.LH_glrlm_RunLengthNonUniformity
wavelet.LH_ngtdm_Coarseness
wavelet.LL_glrlm_RunLengthNonUniformity

from the images acquired on scanner B showed ability to distinguish inserts 1-4 in the biggest VOI. These results suggests that for small VOIs or VOIs including background some texture feature may stop to be informative on the VOI texture, not allowing to distinguish two different tissue (e.g tumoural and normal tissues). This preliminary result seems to confirm the dependence of some imaging biomarkers on the VOI volume, as found by other authors on PET images [118] [119], at least for some of the scanners used in this evaluation. In fact, a similar result was obtained for scanner C, inserts 1-4. For the same scanner, the number of features able to differentiate inserts 1-3 decreased when considering the largest VOI (42.3 cm^3), while it is almost constant for the others VOI sizes. The general trend seems to suggest a greater difficulty of the features in discriminating inserts 3-4 than the other couples. This could indicate that the radiomic feature ability of quantifying a texture property and differentiating different textures depends on the specific textures involved. Evidently, a feature will differentiate more easily two textures which present very different properties than two similar textures. The plots of the 20 features considered in this assessment as a function of the VOI size are shown in Figure 4.14. The features in the plots were extracted from the images acquired on scanner B, which was taken as an example. The error bars were calculated as feature value $\pm$ CR.

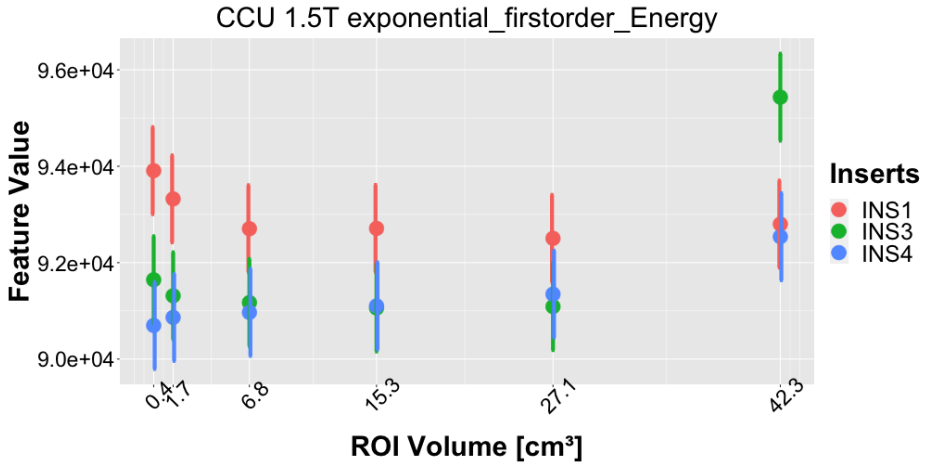
The feature "original_RunLengthNonUniformity" (e) showed overlap among the three inserts when extracted from the smallest VOIs (0.4 and 1.7 cm^3) and the biggest one (42.3 cm^3). The overlap between inserts 3-4 is more evident. The features "Square_Energy" (f), "square_TotalEnergy" (g), "wavelet.HH_Energy" (i), "wavelet.HH_TotalEnergy" (j) and "wavelet.LL_RunLengthNonUniformity" (t) were not able to differentiate the inserts at VOI size 42.3 cm^3. The features "Wavelet.HL_Energy" (l), "wavelet.HL_TotalEnergy" (m), "wavelet.LH_Energy" (p), "wavelet.LH_TotalEnergy" (q) and "wavelet.LH_RunLengthNonUniformity" (r) showed overlap for inserts 3-4 at VOI size 0.4 cm^3.
Summarising, from the plots it seems that below 0.5 cm^3 some of the considered radiomic features were not able to distinguish two similar (but different) textures. The features which did not differentiate two distinct textures, independently of the VOI size, were: exponential_Energy, exponential_TotalEnergy, squareroot_RunLengthNonUniformity, wavelet.HL_Coarseness and wavelet.LH_Coarseness.

Table 4.10: Numbers of features (out of 20) able to differentiate two inserts with different textures. The evaluation was performed on the features in Table 4.9, which were the most repeatable and reproducible. The results are presented for each scanner (A, B and C) and each couple of inserts considered in this experiment (inserts 1, 3 and 4).

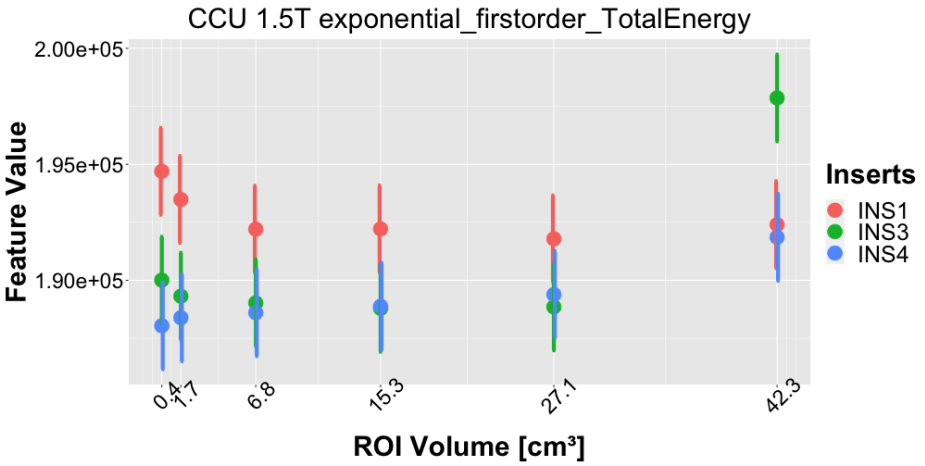
VOI size [cm^3]	**1 vs. 3**			**1 vs. 4**			**3 vs. 4**		
	A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
42.3	18	16	10	16	7	5	20	8	5
27.1	17	14	15	17	12	11	20	10	4
15.3	17	15	14	18	12	12	12	8	5
6.8	17	14	14	17	12	10	13	10	6
1.7	17	16	16	19	15	12	10	10	10
0.4	19	16	15	18	18	11	11	7	5


CCU 1.5T log.sigma.6.mm.3D_firstorder_TotalEnergy
Feature Value
2.1e+05
2.0e+05
1.9e+05
1.8e+05
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

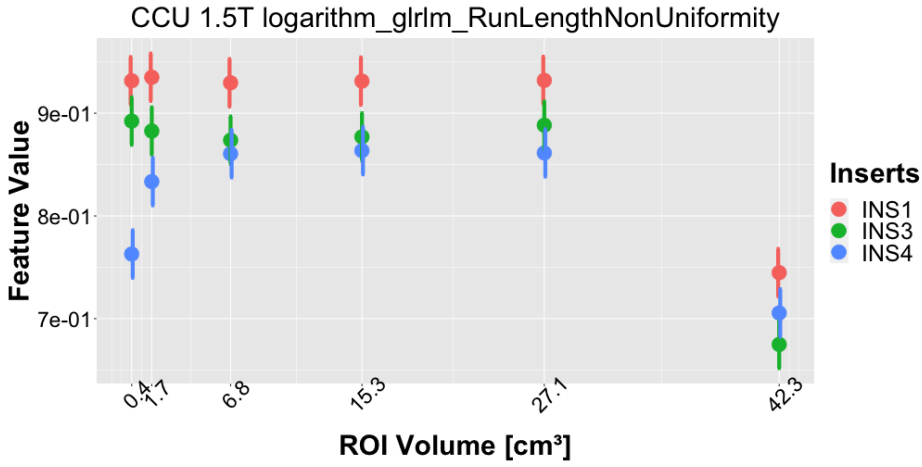
(a)


CCU 1.5T exponential_firstorder_Energy
Feature Value
9.6e+04
9.4e+04
9.2e+04
9.0e+04
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

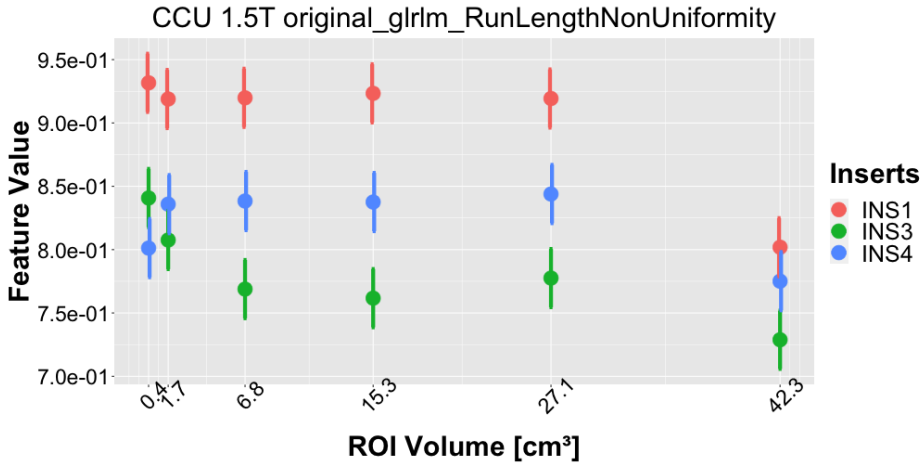
(b)


CCU 1.5T exponential_firstorder_TotalEnergy
Feature Value
2.00e+05
1.95e+05
1.90e+05
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

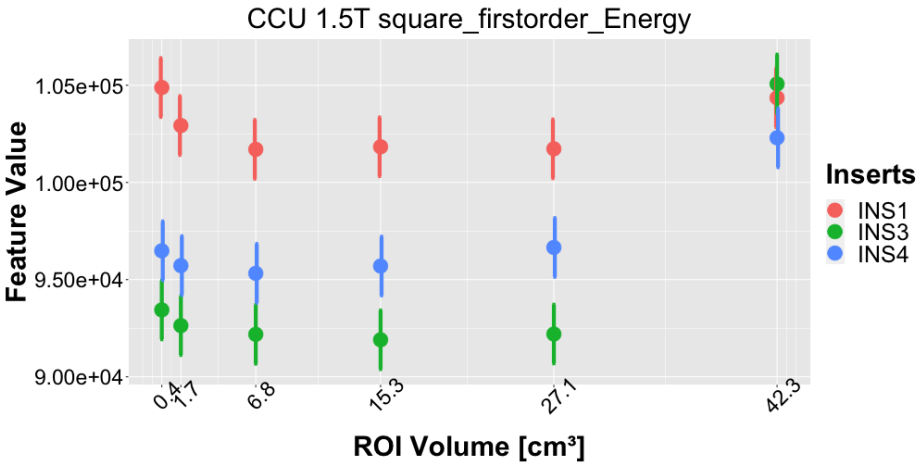
(c)


CCU 1.5T logarithm_glrlm_RunLengthNonUniformity
Feature Value
9e-01
8e-01
7e-01
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

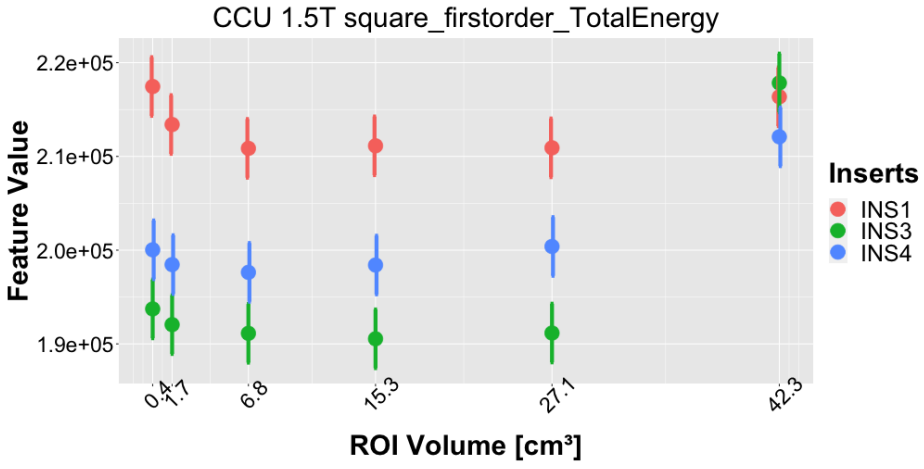
(d)


CCU 1.5T original_glrlm_RunLengthNonUniformity
Feature Value
9.5e-01
9.0e-01
8.5e-01
8.0e-01
7.5e-01
7.0e-01
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

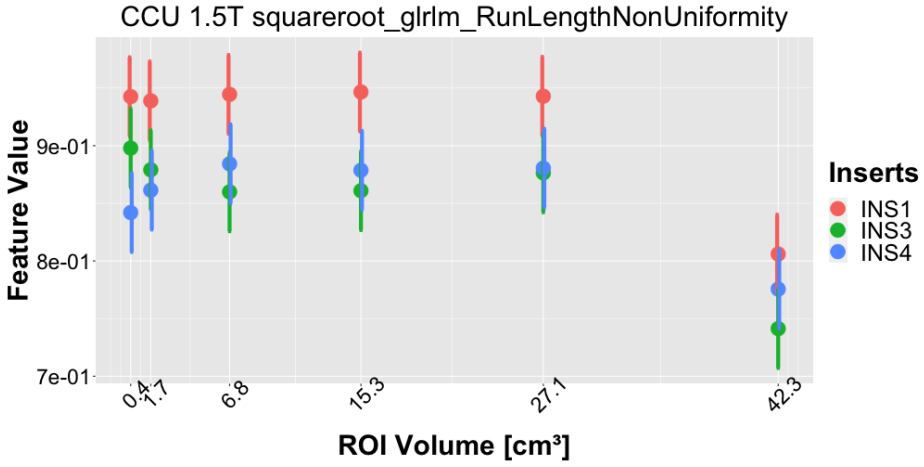
(e)


CCU 1.5T square_firstorder_Energy
Feature Value
1.05e+05
1.00e+05
9.50e+04
9.00e+04
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

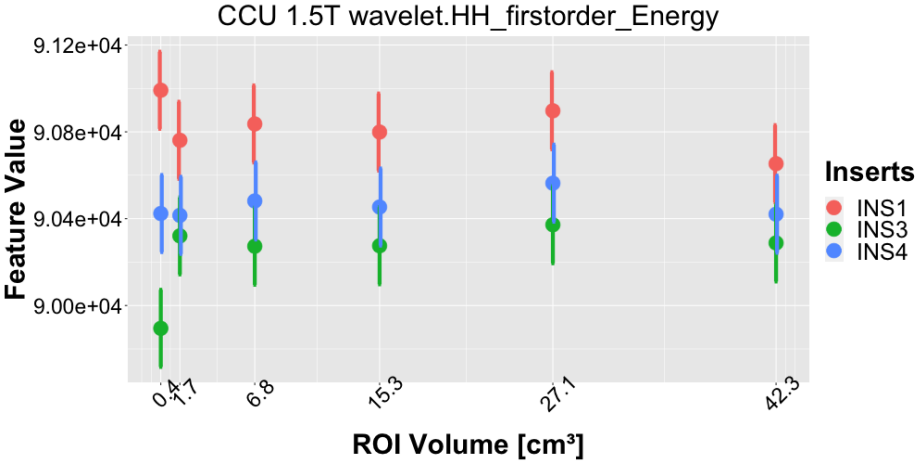
(f)


CCU 1.5T square_firstorder_TotalEnergy
Feature Value
2.2e+05
2.1e+05
2.0e+05
1.9e+05
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

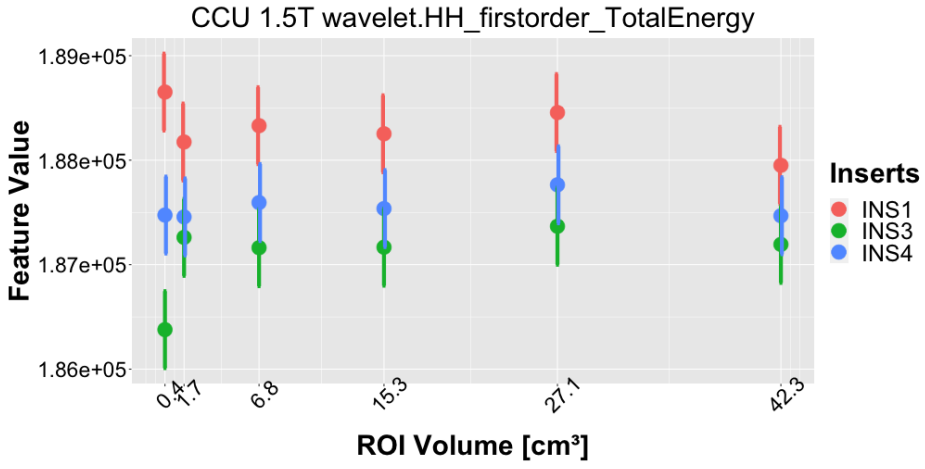
(g)


CCU 1.5T squareroot_glrlm_RunLengthNonUniformity
Feature Value
9e-01
8e-01
7e-01
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

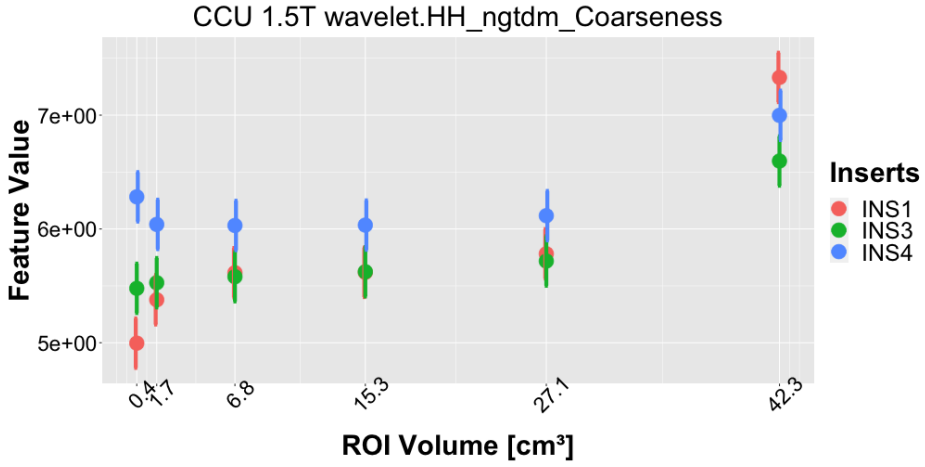
(h)


CCU 1.5T wavelet.HH_firstorder_Energy
Feature Value
9.12e+04
9.08e+04
9.04e+04
9.00e+04
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

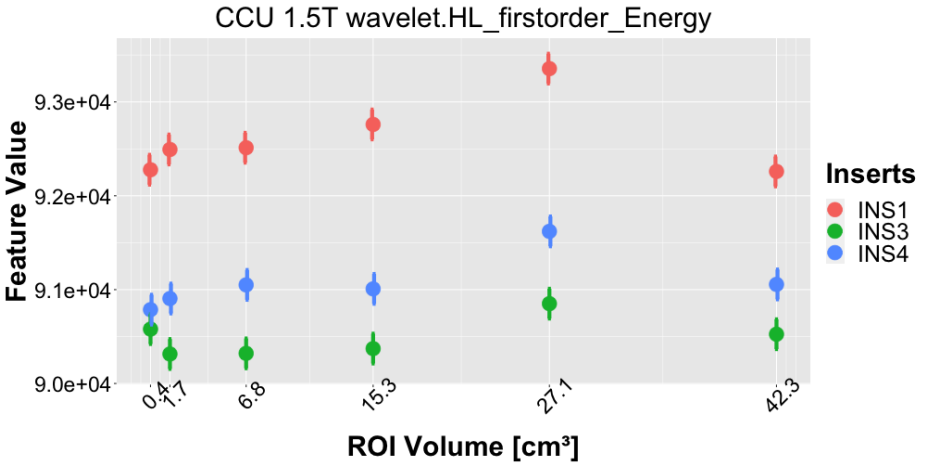
(i)


CCU 1.5T wavelet.HH_firstorder_TotalEnergy
Feature Value
1.89e+05
1.88e+05
1.87e+05
1.86e+05
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

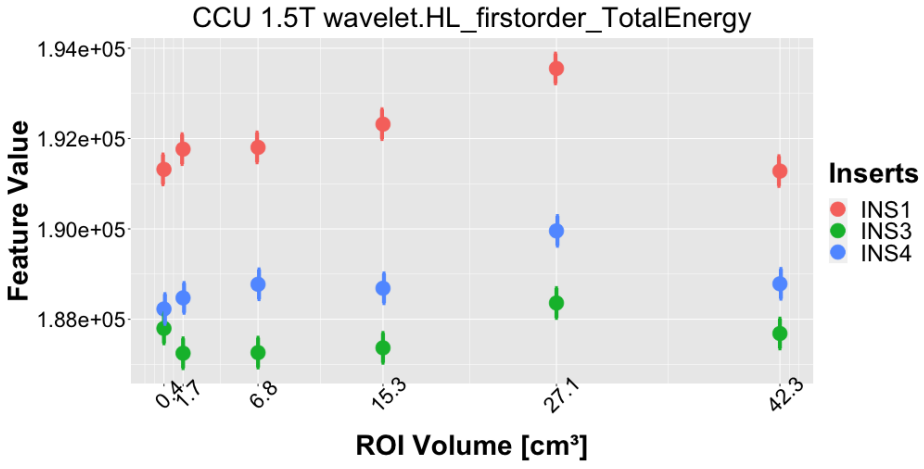
(j)


CCU 1.5T wavelet.HH_ngtdm_Coarseness
Feature Value
7e+00
6e+00
5e+00
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

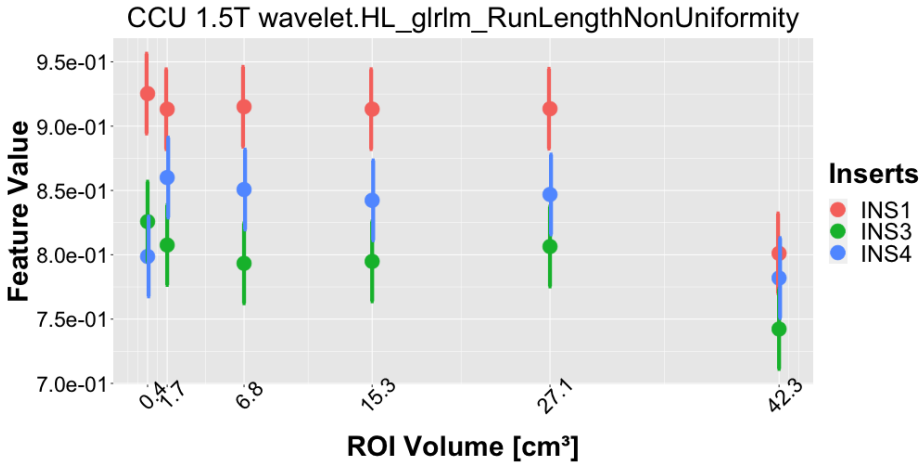
(k)


CCU 1.5T wavelet.HL_firstorder_Energy
Feature Value
9.3e+04
9.2e+04
9.1e+04
9.0e+04
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

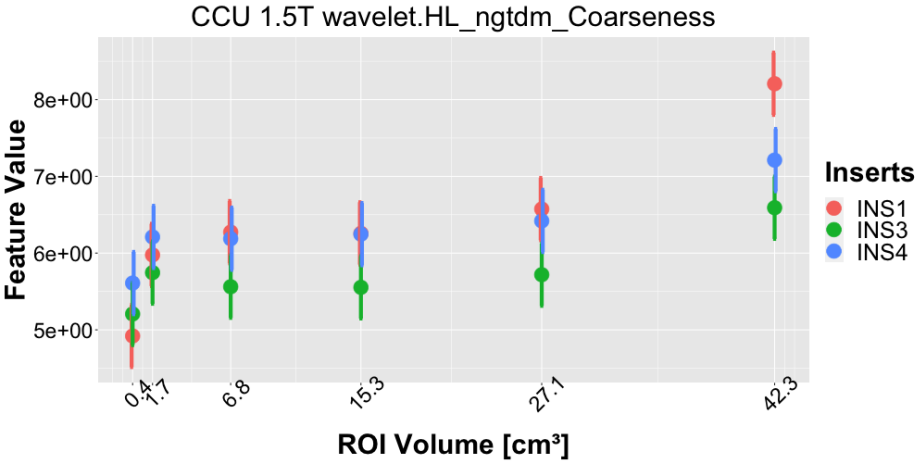
(l)


CCU 1.5T wavelet.HL_firstorder_TotalEnergy
Feature Value
1.94e+05
1.92e+05
1.90e+05
1.88e+05
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

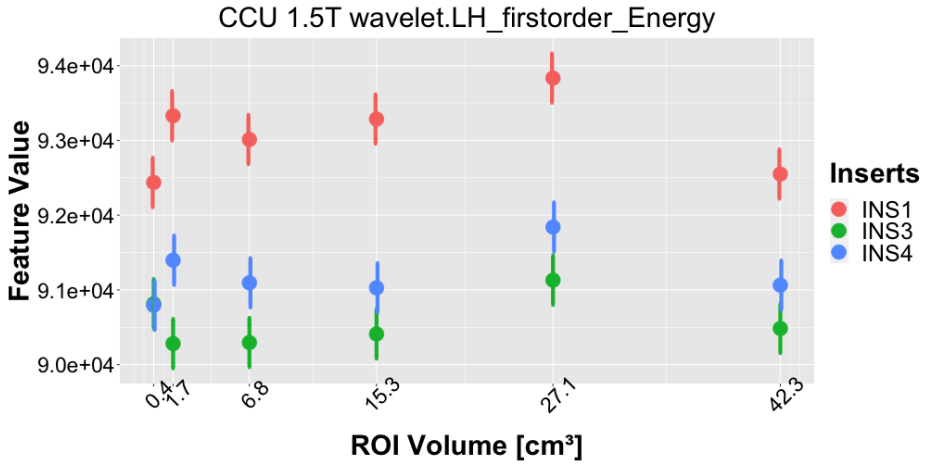
(m)


CCU 1.5T wavelet.HL_glrlm_RunLengthNonUniformity
Feature Value
9.5e-01
9.0e-01
8.5e-01
8.0e-01
7.5e-01
7.0e-01
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

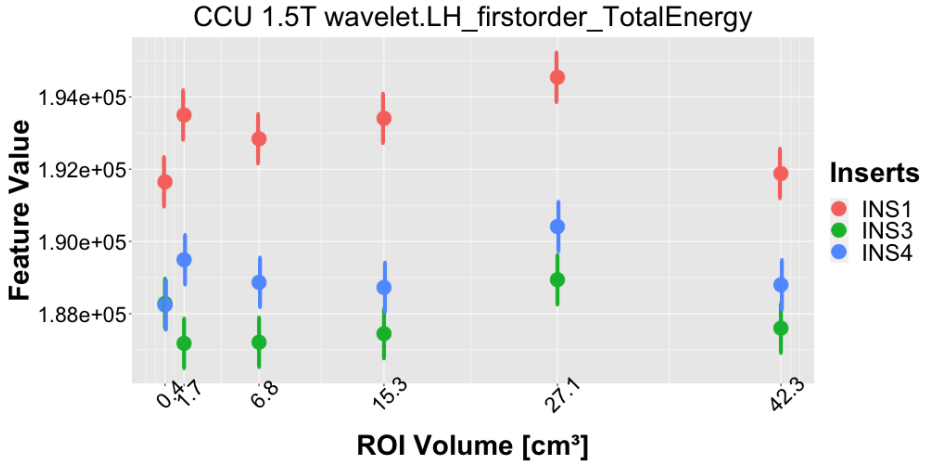
(n)


CCU 1.5T wavelet.HL_ngtdm_Coarseness
Feature Value
8e+00
7e+00
6e+00
5e+00
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

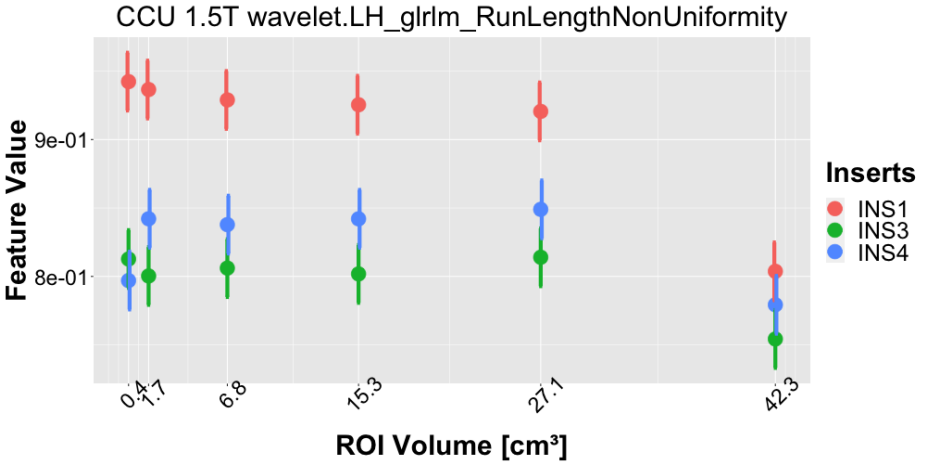
(o)


CCU 1.5T wavelet.LH_firstorder_Energy
Feature Value
9.4e+04
9.3e+04
9.2e+04
9.1e+04
9.0e+04
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

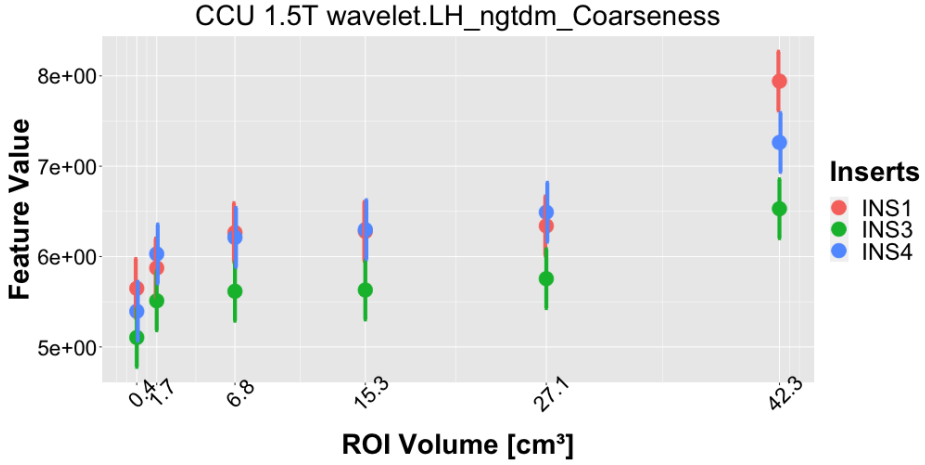
(p)


CCU 1.5T wavelet.LH_firstorder_TotalEnergy
Feature Value
1.94e+05
1.92e+05
1.90e+05
1.88e+05
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

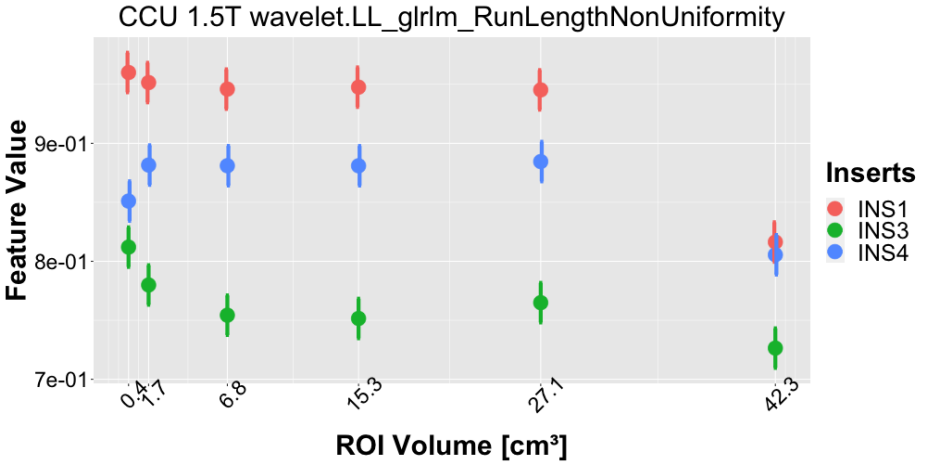
(q)


CCU 1.5T wavelet.LH_glrlm_RunLengthNonUniformity
Feature Value
9e-01
8e-01
0.4
1.7
6.8
15.3
27.1
42.3
ROI Volume [cm³]
Inserts
INS1
INS3
INS4

(r)



(s)



(t)

Figure 4.14: Distribution of the 20 most repeatable and reproducible texture features as a function of the VOI size, for the phantom inserts 1, 3 and 4. The features were extracted from the images acquired on scanner B. The results are presented as feature value $\pm$ CR.

Reproducibility under varying TE or TR

Figure 4.15 shows the percentage of radiomic features presenting (A) excellent, (B) good, (C) moderate, and (D) poor reproducibility between different TE values (according to the CCC thresholds in Table 4.8). Focusing on Figure 4.15(a), 82.4% to 94.9% of features showed excellent absolute agreement between TEs 5 ms apart for TEs ranging between 80-120 ms (first diagonal on the left). Increasing the TE interval between two experiments, the percentage of features showing excellent reproducibility decreased progressively. When considering a TE interval of 40 ms, the reproducible features were 19.8% of the total.
When changing from a TR of 5000 ms to a TR of 4404 ms, out of the 944 features, 856 (90.7%) showed excellent, 64 (6.8%) good, 22 (2.3%) moderate, and 2 (0.2%) poor reproducibility.

Repeatability at 3D level

Out of the 1316 3D features, 446 (33.9%) showed excellent, 394 (29.9%) good, 359 (27.3%) moderate, and 117 (8.9%) poor repeatability. Although we cannot directly compare with the 2D repeatable features, we observe that only a relatively small number of 3D features showed excellent repeatability after repositioning and, therefore, rotational invariance. The 13 features showing rotational invariance in more than 80% of the image types (original or filtered images) are presented in Table 4.11. Of these, 2 features were First Order, 4 GLCM-based, 2 GLRLM-based, 2 GLSZM-based, 1 NGTDM-based, and 2 GLDM-based.

Table 4.11: Repeatable 3D features. List of 3D radiomic features showing excellent repeatability in more than 80% of the image filters.

Feature name	Number of features (%)
firstorder_Energy	14 (100)
firstorder_TotalEnergy	14 (100)
glcm_Correlation	12 (86)
glcm_Idn	12 (86)
glcm_Imc1	13 (93)
glcm_Imc2	12 (86)
gldm_DependenceNonUniformity	14 (100)
gldm_GrayLevelNonUniformity	12 (86)
glrlm_GrayLevelNonUniformity	13 (93)
glrlm_RunLengthNonUniformity	13 (93)
glszm_GrayLevelNonUniformity	14 (100)
glszm_SizeZoneNonUniformity	12 (86)
ngtdm_Coarseness	14 (100)

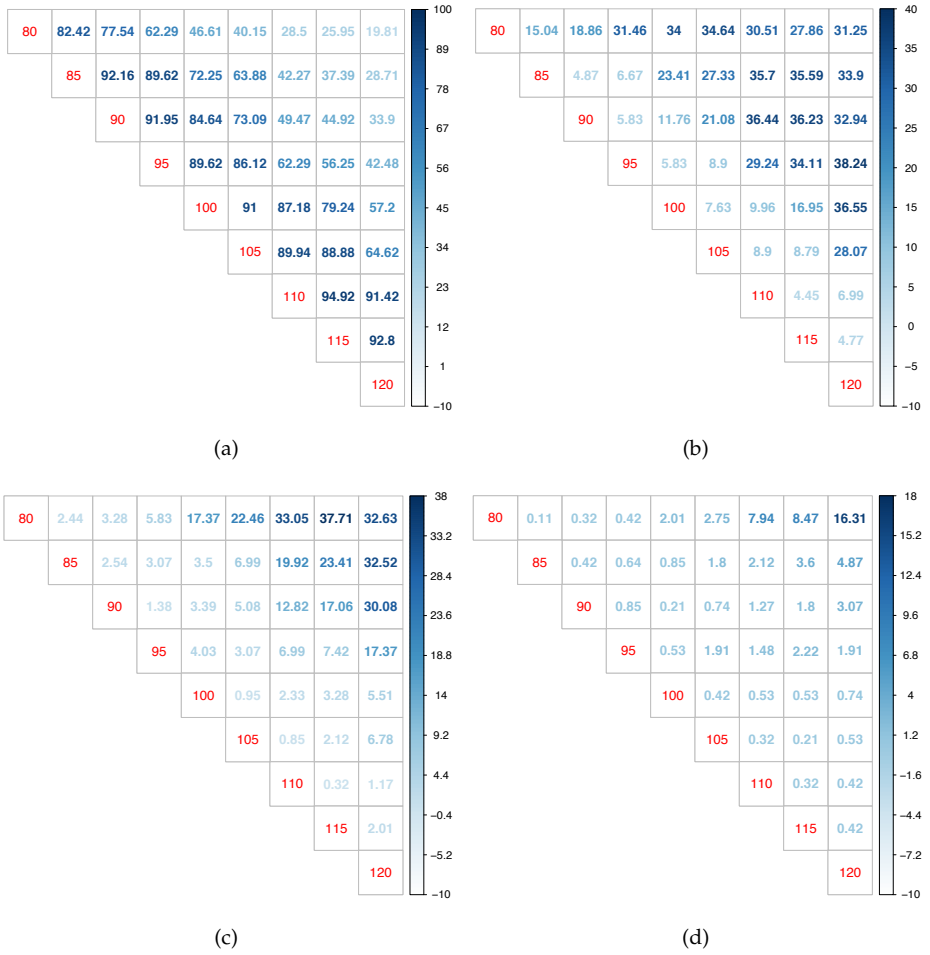


Figure 4.15: Reproducibility of radiomic features with varying TEs. Reproducibility was assessed for all possible combinations of TEs between 80 ms and 120 ms in 5 ms intervals (as indicated in red). (a) Percentage of features with excellent reproducibility; (b) Percentage of features with good reproducibility; (c) Percentage of features with moderate reproducibility; (d) Percentage of features with poor reproducibility.

Assessment of shape information in non-shape radiomic features

In this analysis, two subsets of radiomic features were identified: (i) the subset of non-shape features that were highly correlated with shape; (ii) the subset of features showing a high correlation between their value when extracted from the original image and when extracted from the shuffled-intensities image.
In set (i), out of 930 features, 155 (16.7%), 144 (15.5%), and 158 (17.0%) non-shape features were highly correlated with shape features in scanners A, B, and C, respectively. The discrepancy across scanners in non-shape features showing dependence with shape may be due to differences in contrast between scanners. A total of 99 (10.6%) features were

shared among all scanners.
The subset (ii) was made up of 27 (2.9%), 28 (3.0%), and 27 (2.9%) features for scanners A, B, and C, respectively.
When intersecting the features in subsets (i) and (ii) common to all the scanners, 19 (2.0%) features were obtained. In this way, we identified the set of features nominally belonging to texture but providing only shape information, listed in Table 4.12.

Table 4.12: List of texture features showing, across all scanners, high correlation with shape features (Spearman correlation coefficient > 0.8) and uninformative about texture. Each feature is indicated in the form: ImageType_Class_FeatureName.

Uninformative texture features
log.sigma.6.mm.3D_firstorder_Energy
log.sigma.6.mm.3D_firstorder_TotalEnergy
log.sigma.6.mm.3D_ngtdm_Coarseness
wavelet.LH_firstorder_Energy
wavelet.LH_firstorder_TotalEnergy
wavelet.LH_glrlm_RunLengthNonUniformity
wavelet.HL_firstorder_Energy
wavelet.HL_firstorder_TotalEnergy
wavelet.HL_glrlm_RunLengthNonUniformity
wavelet.HH_firstorder_Energy
wavelet.HH_firstorder_TotalEnergy
wavelet.HH_glrlm_RunLengthNonUniformity
wavelet.HH_ngtdm_Coarseness
wavelet.LL_glrlm_RunLengthNonUniformity
square_firstorder_Energy
square_firstorder_TotalEnergy
square_glszm_GrayLevelNonUniformity
exponential_firstorder_Energy
exponential_firstorder_TotalEnergy

4.2.4 Discussion

Clinically meaningful radiomic signatures, providing not only high performance, but also good generalisability, should be constructed using high-quality, reliable features. Therefore, it is essential to understand and incorporate the factors affecting the reliability of such features. This study focused on several aspects that may influence the values of the radiomic features, as it was designed to provide information on: repeatability; reproducibility at fixed imaging parameters; shape information in non-shape features; influence of acquisition parameters (TE and TR).

Two types of 2D **repeatability** were investigated, offering measures on scanner-induced variations (without phantom repositioning) and repositioning-induced variations across

different scanners. Without repositioning, it is possible to observe that 92.1% to 96.4% of the features showed excellent repeatability in both 1.5 T scanners. The percentage decreased to 79.9% when considering the features extracted from the images acquired at 3 T. The decrease in repeatability at 3 T may be due to the artifacts more frequently affecting images acquired at a higher field strength. Instead, when considering repositioning, a reduction in the number of features with excellent repeatability was observed: 78.4% of features showing excellent repeatability on scanner A, 85.4% on scanner B, and 11.2% on scanner C. Possible causes for the relatively small reduction observed with scanners A and B may be the rotational invariance of some features and some degree of misalignment of the acquisition after repositioning. As for the larger reduction observed with scanner C, after visual inspection of the images, it was possible to observe that the chemical shift artifact was considerably larger than with scanners A and B, as expected, being proportional to the magnetic field strength. Additionally, as chemical shift artifacts occur in the frequency encoding direction due to the coexistence of water and lipid protons in a voxel, a slight change in the positioning of the phantom will be translated into a change in these artifacts.
The repeatability in 3D MRI acquisition was assessed as well. Only 33.9% of 3D radiomic features showed excellent repeatability after repositioning. As these features were extracted from isotropic images, this represents an unexpected result. In fact, the percentage of 2D features showing excellent repeatability with phantom repositioning for scanner B was much higher (85.4%). This seems to suggest that a 3D extraction from isotropic voxels does not increase the repeatability performance if compared with a 2D extraction. A possible explanation may be that the 3D feature extraction on 3D acquisitions, by being able to offer more detail in the through-plane, may result in more unstable features, as the planning after repositioning does not ensure perfect alignment between acquisitions. In fact, a coarser spacing may not affect the performance of radiomic models and may even improve them as features become less susceptible to noise, repositioning, and other artifacts. These considerations justify the stability observed on the 2D features extracted from the 2D acquisitions.

The **reproducibility** of the features was assessed by comparing radiomic features extracted from images acquired on scanners of different manufacturers with equal magnetic field strength and acquisition parameters, and scanners of the same manufacturer and different magnetic field strengths. In the first case, only 4.6% of the features showed excellent reproducibility. A similar trend was found in the second case, where 15.6% of the features exhibited excellent reproducibility, and most of the features, 53.2%, had moderated or poor reproducibility. The results showed a higher reproducibility when comparing two scanners of field strength 1.5 T and 3 T from the same manufacturer (even though the relaxation times T_1 and T_2 affecting the MR signal are dependent on the field strength), than comparing two scanners of 1.5 T with same acquisition parameters but from different vendors. Regarding reproducibility between different vendors at fixed imaging parameters, we observed that the images acquired on the two scanners exhibited different CNR values. Despite the basic principles of the sequence being the same, these differences can be caused by various sources, that cannot be controlled by

the user. A first factor is the type of systems, including the digital versus analog and the location of the analog-to-digital converter. As a matter of fact, Philips converts the signal to digital directly on the coil while GE makes it on the magnet. Other factors consist of different preparation and calibration phases, including distinct power optimizations, frequency determination, shimming, and coils tune.
Despite not being focused on radiomic features, previous studies on the inter-scanner and intra-scanner variability, considering both 1.5 T and 3 T images, on a set of organ-specific measures [128] and proton density fat function measurements [129] showed that disagreement could be seen when comparing scanners of different vendors. Given this, this study allowed to highlight that even in the presence of unavoidable differences between images acquired with two scanners of different manufacturers (quantified for example by CNR), a subset of features turned out to be reproducible.
In this illustrative study, a controlled pelvic imaging scenario was considered, resulting in only 31 radiomic features (3.3% of the total number of features extracted) showing excellent robustness in the two repeatability settings and the two reproducibility settings. Of these, 11 were shape features. The other 20 features, after correction for an intrinsic dependence on the VOI volume, were analysed for their ability to differentiate the different textures offered by the phantom inserts. This preliminary investigation showed that some texture features depend on the VOI size. In particular, for very small VOIs (below 0.5 cm^3) the some features were not able to distinguish two inserts. A similar result was obtained on VOIs drawn around the inserts, including a portion of background. The interesting results obtained will be the basis for an investigation extended to all the available radiomic features rather than a selected subset. In fact, it is crucial to understand which is the volume threshold below which the features lose their descriptive power in terms of texture properties. The scenarios of clinical studies may be different from the one herein considered, but an analogous procedure could be applied, focusing on the phantom inserts that better simulate the texture properties of the tissues under investigation. This is of particular importance in clinical studies involving very small tumours (like lymphnode metastasis or lung nodules).

As demonstrated in the study by Welch and colleagues [123], some features may exhibit dependencies on volume. However, these dependencies may be extended to other shape features that may also contain diagnostic and prognostic information. Therefore, it is crucial to understand if non-shape features may be repeatable and reproducible due to **shape information** they may contain. A total of 19 non-shape features common to all three scanners appeared to contain only shape information or were heavily dependent on shape. These features showed both a high correlation with shape features and excellent agreement between features extracted from the original images and images with randomly shuffled intensities. Excluding the identified uninformative features (Table 4.12) from the list of the features resulted to be highly repeatable and reproducible (Table 4.9), only five features remained:

- log.sigma.6.mm.3D_Energy;
- log.sigma.6.mm.3D_Coarseness;

- wavelet.LH_RunLengthNonUniformity;
- wavelet.HH_RunLengthNonUniformity;
- square_GrayLevelNonUniformity.

All the results discussed so far used images acquired with fixed MR sequence parameters. However, the need for a large quantity of data to implement radiomic models may lead to the creation of inhomogeneous databases as used in several MRI clinical radiomic studies. This inhomogeneity, also due to the optimisation of imaging parameters on each scanner, might affect the values of radiomic features and, consequently, the performance of models [16], [130]. The preliminary investigation on the **influence of different TE and TR** within this study demonstrated that the majority of the features showed excellent reproducibility when the difference in TE was 5 ms. However, the reproducibility decreased as the TE interval increased, as seen in Figure 4.15. This corroborated the expectations for a T_2-w MRI sequence, as a different TE changes the weighting in T_2 and CNR, which will lead to variations of T_2-w signal intensity depending on the underlying tissue and pathophysiology. As texture depends on contrast, but since TE-induced differences are non-linear, the dependence on contrast is not removed during image normalization. Similarly, when assessing changes in TR, it was observed that 90.7% of the features showed excellent reproducibility between acquisitions with a TR of 5000 ms and with a TR of 4405 ms.
These results suggest that radiomic studies should be conducted with standardised imaging protocols or making use of the features showing excellent reproducibility in the interval of TEs and TRs used, ensuring this way that the observed results are not associated with differences in the acquisition parameters. Following this consideration, multicenter retrospective studies (involving different scanners with variable protocols) should be coupled with methodological studies to understand the relationship between the variability range of the radiomic features and the variation of each sequence parameter separately. Moreover, given the differences observed between Philips and GE technologies, the investigation of the TE/TR impact on scanners from different vendors will be considered as a natural extension of the present study in future developments.

A few previous studies have assessed the stability of radiomic features in different MRI settings. Mayerhoefer and colleagues [69] investigated the sensitivity of texture features to acquisition parameters (including TE ranging from 20 to 125 ms and TR in the range 900-4500 ms) on T_2-w images acquired on a 3 T scanner. They found that the imaging parameters influenced the values of features, with this influence increasing with spatial resolution. Although different spatial resolutions were not compared in this study, the present investigation performed on a 1.5 T scanner extends the findings of those authors to a lower field strength, while making use of a phantom design for patient imaging, as it was proved that the texture features values are dependent on the TE and TR parameters. In another study by Chirra et al. [131], prostate T_2-w MRI images of 147 patients from four different sites were used to assess cross-site reproducibility by performing multivariate cross-validation and assessing preparation-induced instability. The authors found that most of the Haralick features were reproducible in over 99% of all cross-site

comparisons. However, that study uses different patient populations on each site and assesses non-tumoral regions under the assumption that these should have a similar texture. Besides, part of their pre-processing involves image upsampling, in some cases by a factor of $\sim$11 times in the image through-plane. Such choices make the comparison with the results obtained in this thesis difficult.

The limitations of the experiments described in this Section are highlighted below. Firstly, the use of a phantom cannot include all the effects that exist in real clinical scenarios, such as patient motion, rectal/bladder filling, peristalsis, breathing, tissue diffusion and perfusion, intra-patient tissue variability. Undeniably, a phantom, as such, cannot be exhaustively representative of all tumours (with substantial differences in terms of dimension, internal structure, texture, shape, etc.), so it cannot identify all the trustable and robust features for patient image analysis. On the other hand, contrarily to the clinical reality, the use of a phantom allows repeating as many acquisitions as desired, to compare results, and to assess the influence of many parameters. In particular, the use of an inhomogeneous phantom allowed the implementation of a procedure for the identification of features that may not be trustworthy and robust. Furthermore, the use of a phantom, in which biological processes are not present, allows the assessment of the reproducibility issues caused by different system types, e.g., digital vs. analog, coils used, and other sequence parameters, which are translated into distinct image properties like the contrast-to-noise ratio that may have an impact on the values of radiomic features as well. Having this in mind, the experiments conducted in this study provide excellent baseline assessments, under very controlled environments, of the stability of features, and allow avoiding rough misleading results that could be derived without an acquainted input selection before the statistical analysis.
Another limitation was the use of data acquired on a restricted number of scanners (two field strengths, two manufacturers) and the focus on the investigation of the radiomics stability on T_2-w MRI images, as part of the clinical diagnostic protocol for pelvic imaging. Conversely, this can represent the starting point for further extension. Thus, the imaging data, parameter files, and analysis scripts are made available in order to allow other researchers to perform similar analyses and incorporate new experimental paths. Following these results, it might be interesting to make use of the phantom in an extended multicenter study using different clinically optimized sequences, for a more comprehensive investigation.

In conclusion, this study investigated the robustness of 2D and 3D radiomic features extracted from T_2-w images of a pelvic phantom created for MR radiomic analyses. The methodological investigation quantified the stability and quality of radiomic features in different MRI settings, enlightening important issues towards robust and reliable models. Based on a workflow designed to test repeatability and reproducibility, features showing the highest performance were identified. Importantly, many of these repeatable and reproducible features turned out to be inadequate for radiomic analysis, e.g., being non-informative or affected by the image acquisition process. This, or similar, procedure should be applied to strengthen and support each clinical radiomic study.

CHAPTER 5

BREATH: the breast phantom

The aim of this Chapter is to show the preliminary results of the feasibility study on a breast phantom to support radiomic analysis on MR images of patients affected by breast cancer. The design of the phantom, called BREATH (BREAst Texture pHantom), is presented in its preliminary version, along with the search of materials and validation. The potential of BREATH will be explained in the context of ongoing activities and future developments, which represent a natural extension of this thesis.
This part of the thesis work was performed in collaboration with the European Institute of Oncology (IEO, Milan), the Champalimaud Centre for the Unknown (CCU, Lisbon) and the Interdisciplinary Centre for Nanostructured Materials and Interfaces (CIMaINa, Milan).

5.1 Phantom design and realisation (first prototype)

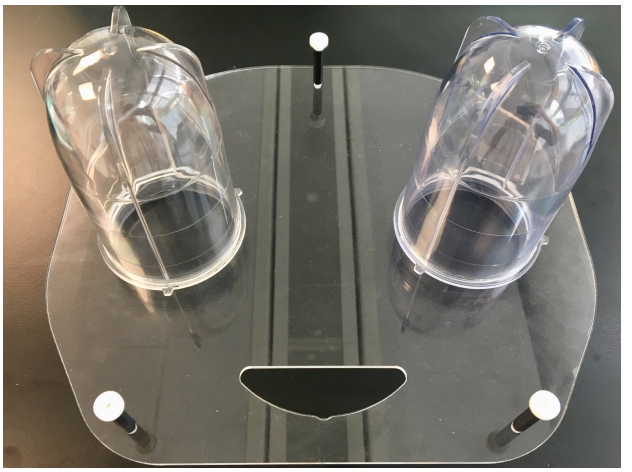
The main focus of BREATH, as radiomic phantom, was to reproduce the texture and MR signal of the tumour and surrounding tissues observed in a representative set of patients with breast cancer. The pelvic phantom, implemented earlier, provided a model for the development of the breast phantom. In addition, in designing BREATH attempts have been made to overcome the limitations observed in PETER PHAN, mainly with the aim to create more reproducible inserts and facilitate the usage of the phantom by a wider number of operators and/or centres. In this Section, the operative steps for the phantom design and realisation will be covered, starting from the external structure to the internal compartments and finally the description of the insert reproducing the tumour.

5.1.1 The shell

A breast-shaped shell has been produced as external container of the phantom. It consists of two main parts: the support base and two jars, which mimic the breast. The support base and the jars, shown in Figure 5.1, are separated modules, that can be assembled as an interlocking system. The phantom is of simple assembling and disassembly for an easy transport. Plastic jars of different size and shape can be used to fit the clinical scenario of interest. They were easily found in a home good store and were chosen as containers with a lid, to facilitate the incorporation of the artificial tumour, that will be described later, and the filling with a liquid solution. The support base was built

on purpose in plexiglass, as this material is MR-safe and does not interfere with the MR signal by causing artifacts on the image. Moreover, it is easily processable, cheap and light. A handle was carved in the base, to hang it when not used, and plastic screws were exploited to help a stable positioning of the phantom inside one of the available breast coils, as shown in Figure 5.2 (a). The positioning of the phantom and dedicated coil on the bed of the MR scanner is shown in Figure 5.2 (b).

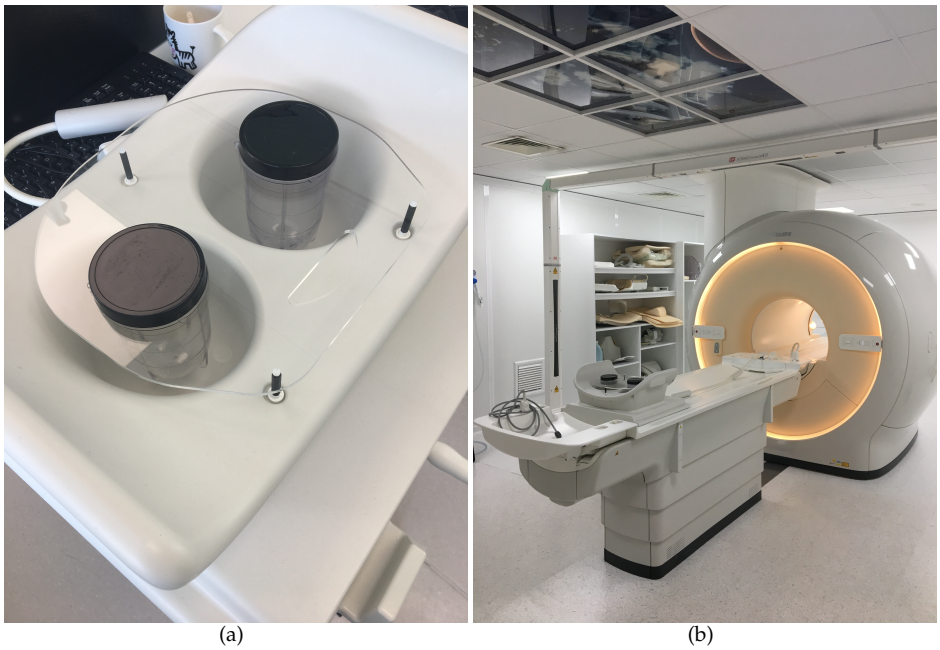
(a) (b) (c)



(d)

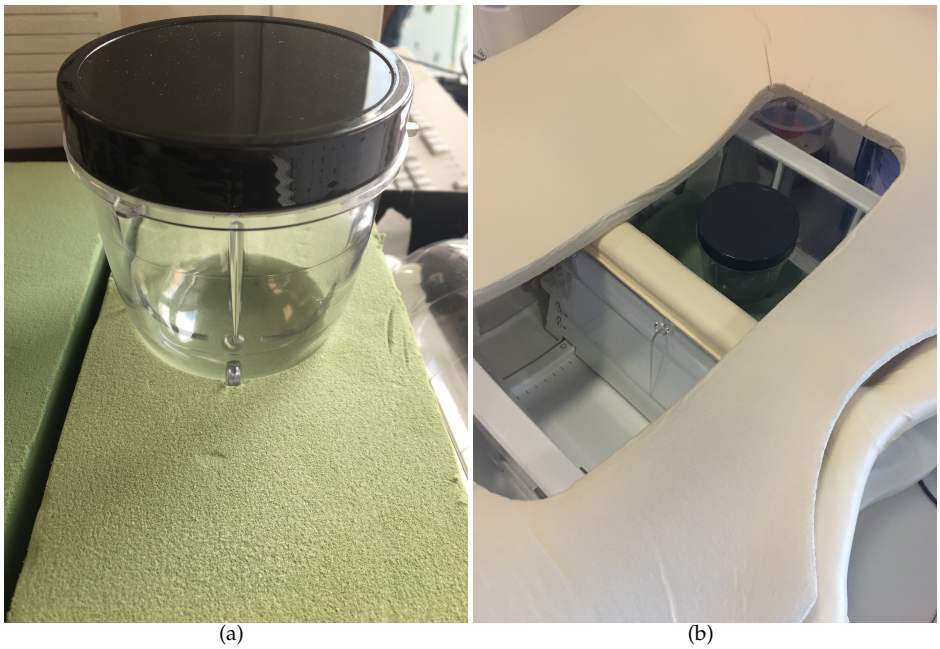
Figure 5.1: (a-b-c) Jars of different sizes to mimic the breast shape. (d) Jars and support base assembled to build the shell of BREATH.

With the aim of a future multicentric study with BREATH, another solution was developed if the support base did not fit other coil models. Instead of using the plexiglass support, the jars could be inserted in a rigid sponge which can be easily modeled to fit the coil and at the same time can guarantee the mechanical stability of the jars during the image acquisition. This alternative setup is shown in Figure 5.3.



(a) (b)

Figure 5.2: (a) The phantom shell positioned in the dedicated breast coil. (b) Overview of the phantom shell positioned inside the coil on the bed of the MR scanner.



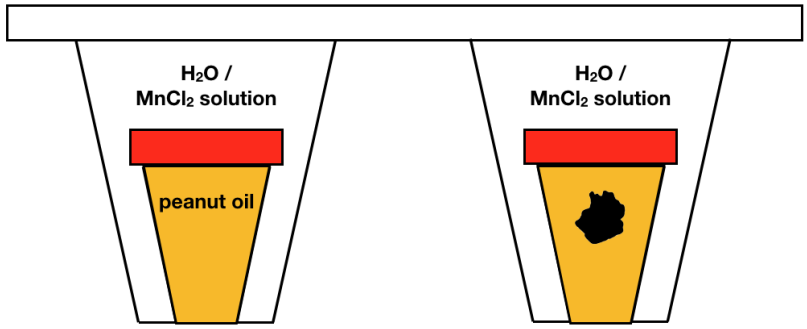
(a) (b)

Figure 5.3: Alternative setup for a different breast coil. The green sponge can easily be cut to fit the coil and can keep the jars still during the experiments.

5.1.2 The phantom compartments

Each jar hosted a smaller cup (48 mm diameter, 72 mm height), creating two separated compartments, as shown in Figure 5.4. The smaller cups were filled with peanut oil, which reproduces the adipose tissue of breast, giving a similar MR signal [132], [133]. The external compartment was filled with water in the preliminary version of the phantom. In the future versions, it could be filled with a solution of $MnCl_2$ or other ions, to control the relaxation times of the solution and tune it for a specific setting.

(a)



(b)

Figure 5.4: BREATH assembled. (a) Picture of the jars before inserting them in the coil. (b) Sketch of the setup. The artificial tumour is represented in black. The peanut oil mimics the adipose tissue surrounding the tumour.

An artificial tumour, described below, was embedded in one of the jar, to simulate the texture and signal of a set of representative patients with breast cancer for radiomic purposes.

5.1.3 The artificial tumour

In this Section, the procedure for the realisation of the artificial tumour will be illustrated. The general approach was the following: (i) starting from a set of selected patients with breast cancer, a representative model of a tumour was identified; (ii) the model was exploited to create a tumour-shaped virtual object; (iii) on the basis of the virtual model, the shell of the artificial tumour was 3D-printed; (iv) a test material was injected inside the shell and expanded to create an artificial heterogeneous tissue. In the next paragraphs, these experimental steps and choices will be explained in detail.

The geometrical choices The shape and volume of the artificial tumour were selected to be representative of a retrospective database of breast cancer patients, candidate to a radiomic study at the IEO. The dataset (database 1, Table 5.1) included 84 patients affected by non-metastatic breast cancer, not previously treated, with a preatment MRI scan followed by NAT (neoadjuvant therapy = chemotherapy, hormonotherapy and/or anti-HER2 therapy) and surgery. The patients underwent a basal MRI scan, including

Table 5.1: Datasets of patients and healthy volunteers involved in the breast phantom study.

Database	Number of subjects	Age (years)	Inclusion criteria
1 (NAT)	84	27-79 (mean 48)	(i) confirmed non-metastatic breast cancer (ii) pre-treatment MRI + NAT + surgery
2 (mapping *in vivo*)	11 (healthy volunteers)	26-54 (mean 38)	-
2 (mapping *in vivo*)	6 (patients)	25-65 (mean 46)	presence of a breast lesion (diagnostic MRI for suspected cancer)
3 (surgery)	183	29-84 (mean 52)	(i) histologically-confirmed breast cancer (ii) pre-operative MRI + surgery

a T_2-w morphologic sequence, a diffusion-weighted sequence and a dynamic T_1-w sequence, with the injection of a paramagnetic contrast agent (DOTAREM). Written informed consent to undergo the examination, and to the use of anonymised clinical and imaging data for scientific and/or educational purposes, was obtained from all subjects involved. After the acquisition, the DICOM images were anonymised and exported for processing. An expert radiologist segmented the tumoural region on the images of the first arterial phase after the injection of the contrast agent (dynamic sequence). The imaging parameter of this sequence were: 7.39 ms TR, 3.44 ms TE, 1.4 mm slice thickness, 0.7 mm slice spacing, 350x350 mm^2 FoV, 352 x 352 AM. The VOIs were segmented with a semi-automatic threshold technique on ITK-SNAP [134], as shown in Figure 5.5.
The analysis of the obtained VOIs across the considered patients showed a wide variety of shapes and volumes. Some of the different shapes are represented in Figure 5.6 and included both complex structures and more simple volumes (rounded/spherical or elongated). The surfaces were often irregular and granular. In some cases the VOI included non-continuous structures. The analysed tumours had volumes in the range 0.3-400 cm^3. Apart from few massive tumours, $\sim$74% of the VOIs had a volume within 15 cm^3, as shown in the histogram 5.7. These results, along with the experimental choice to select a sufficiently simple shape for the preliminary phantom version, led to the identification

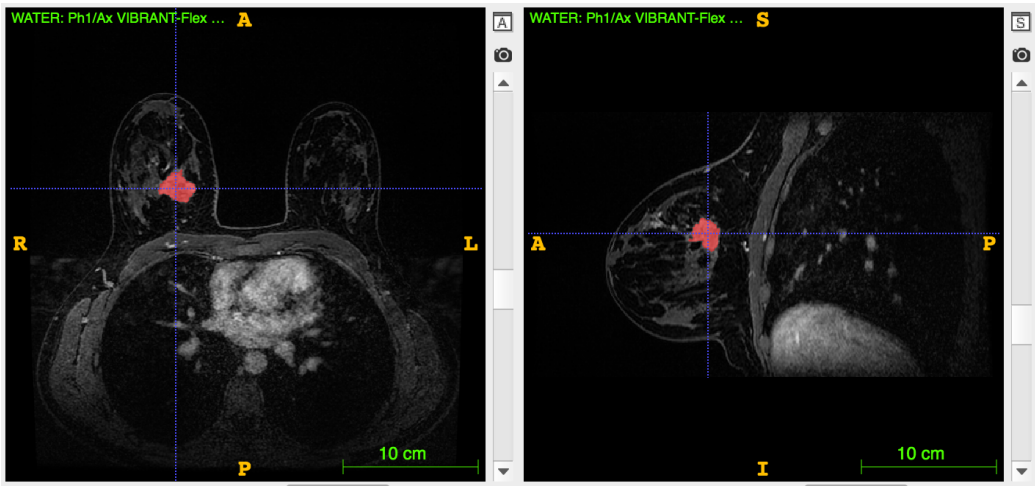


Figure 5.5: Example of a T_1-w image (axial view on the left and sagittal view on the right) of a patient with breast cancer and the corresponding tumoural region (in red) segmented on ITK-SNAP. The segmentation was performed on all the slices which included the tumour. This Figure shows only one slice.

of a representative shape and volume to be mimicked in the phantom. The selected tumour volume is shown in Figure 5.8. It has a spherical shape, an irregular surface and a 6 cm^3 volume. The .stl file of the volume has been extracted from ITK-SNAP to serve as input for a 3D printer used to print the mould for the external shell of the artificial tumour, as explained later.

The choice of materials After the identification of the shape and volume of the tumour model, a research was carried out to identify the materials to exploit for the production of the artificial tumour. The aim of the artificial tumour is to mimic the MR signal (in terms of intensity and relaxometry properties) and the texture of the selected model tumour (Figure 5.8) on the images obtained with the T_1-w sequence (first arterial phase post-injection), which is the one involved in the future radiomic analysis of the considered patients' database.
For these purposes, the intensity histogram of the real tumour (used as a model) has been extracted from the images obtained with the selected sequence (Figure 5.9). In addition, on a set of volunteers and selected patients with a breast malignancy, an *in vivo* T_1 and T_2 mapping was performed, using a procedure similar to the one used for the pelvic phantom, to obtain the range of relaxation times to mimic in the phantom.
11 healthy volunteers (26-54 yrs, mean 38) and 6 patients with a breast tumour (25-65 yrs, mean 46) (database 2, Table 5.1)were imaged with dedicated MR sequences for *in vivo* assessment of T_1 and T_2 relaxation times of healthy tissue (fat) and breast tumour. Written informed consent to undergo the examination, and to the use of anonymised clinical and imaging data for scientific and/or educational purposes, was obtained from all subjects involved. Image acquisition was carried out using a 1.5 T MR scanner (Optima MR450W,

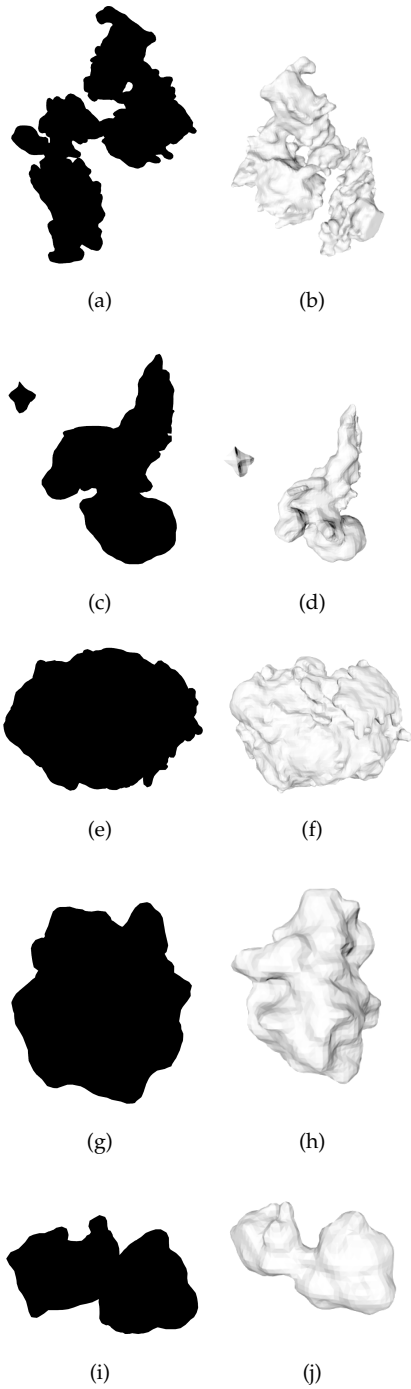


Figure 5.6: Examples of representative tumour shape in 2D (right) and 3D (left) for some of the breast cancer patients in the database considered for this study.

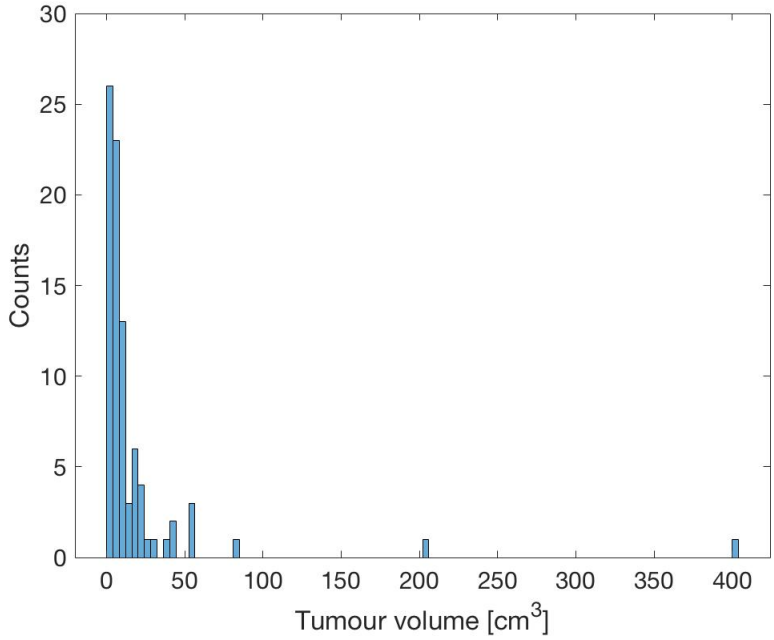
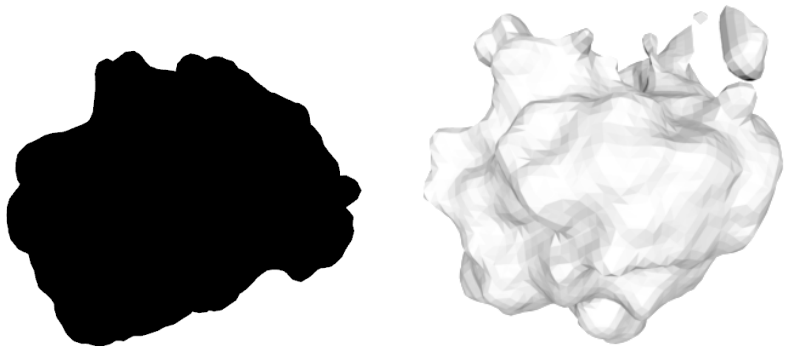


Figure 5.7: Distribution of the tumour volume in the set of breast cancer patients considered in this study.



(a) (b)

Figure 5.8: The tumour of the patient in 2D (right) and 3D (left) selected as model for the artificial tumour to be included in BREATH.

General Electric Healthcare, Waukesha, USA). *In vivo* T_1 mapping made use of the variable flip angle (VFA) spoiled gradient recalled echo (SPGR) technique [106]. Three axial scans were acquired, each with a different flip angle (5, 10 or 20°), keeping the other parameters fixed: repetition time (TR) 6.6 ms, echo time (TE) 4.2 ms, slice thickness 3

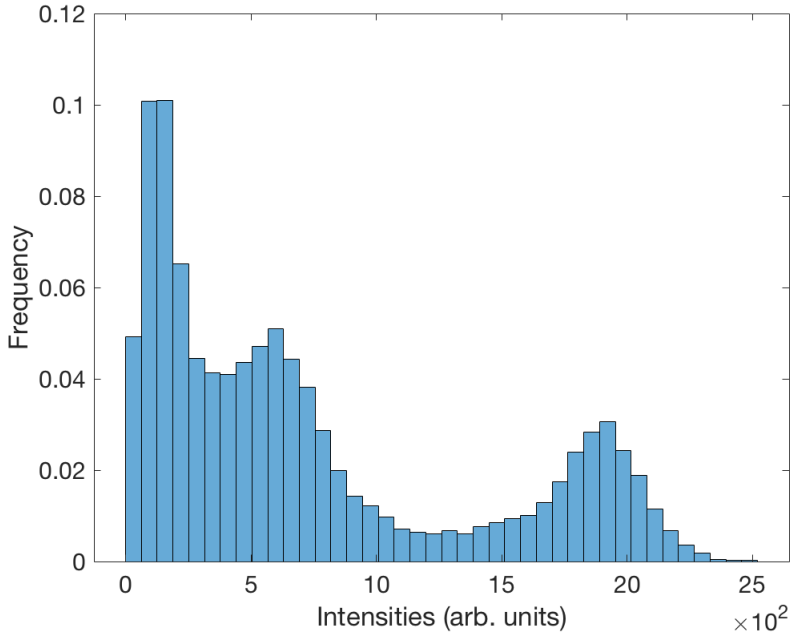
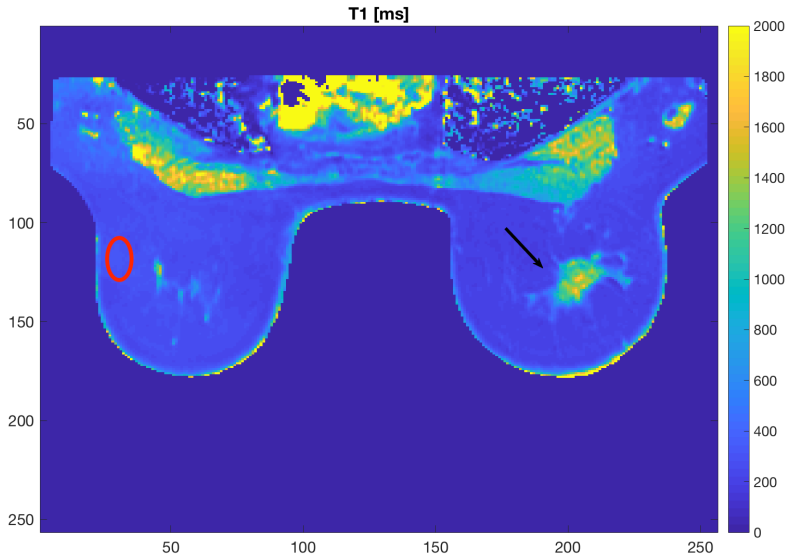


Figure 5.9: Normalised intensity histogram of the tumour volume selected as a model for the artificial tumour. The y axis is the frequency of number of voxels in the tumour volume.

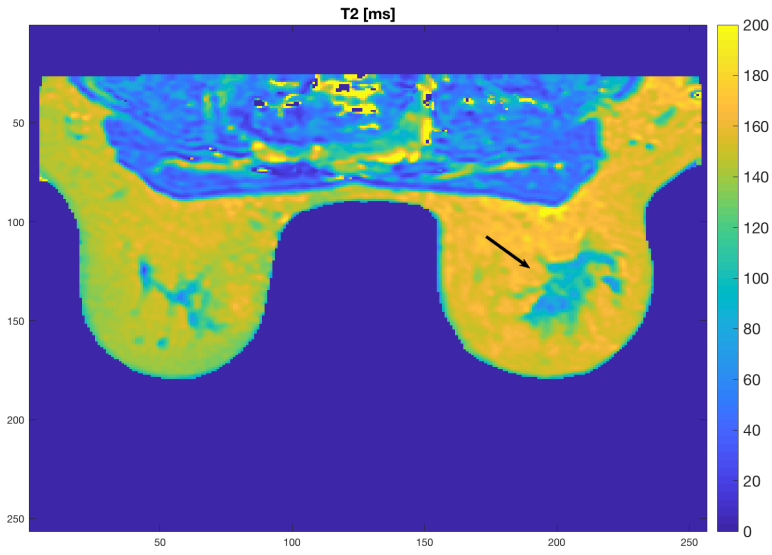
mm, slice spacing 3 mm, field of view (FoV) 256x256 mm^2. T_2 mapping was based on a dedicated 2D multi-echo spin-echo pulse sequence which ran under research mode and could read up to 16 echoes. The sequence parameters were: TR 5000 ms, min TE 7.5 ms, max TE 119.8 ms, echo spacing 7.5 ms, slice thickness 4 mm, slice spacing 5.5 mm, FoV 256x256 mm^2.

After the acquisition, the DICOM images were anonymised and exported for processing. The VFA data were processed with the NOVIFAST [107] algorithm to calculate T_1 maps. T_2 maps were generated by fitting a mono-exponential curve to the 16-echo spin-echo data on a voxel-by-voxel basis, using a MATLAB (version R2018b) script developed in-house. Volumes of interest (VOIs) in the adipose tissue were drawn manually on the T_1 and T_2 maps of all subjects. In addition, an expert radiologist drew a VOI for the tumour on the axial T_1-w images obtained as part of the routine clinical protocols. The tumour VOI was then transferred onto the relaxation time maps. For each VOI, the subject-wise mean values of T_1 and T_2 were extracted. The range of T_1 and T_2 values obtained for the tissues of interest are summarised in Table 5.2 and examples of the *in vivo* T_1 and T_2 maps are given in Figure 5.10. The values reported are the mean over all the patients and the minimum and maximum values.

BREATH aimed at reproducing a patient's breast when scanned with a T_1-w dynamic sequence. In this terms, the artificial tumour should mimic the T_1 range found *in vivo* and have a distribution of intensities similar to the histogram 5.9. As regards the adipose tissue, it has been simulated with peanut oil, which has similar relaxometry properties.



(a)



(b)

Figure 5.10: (a) T_1 and (b) T_2 maps of a patient with breast cancer. The tumour site is highlighted with a black arrow and the fat ROI contour in red.

Table 5.2: Relaxation times of human tissues measured *in vivo* at 1.5 T. *mean (range)

	T_1 [ms]*	T_2 [ms]*
Adipose tissue	251 (211 - 339)	151 (134 - 164)
Tumour	1322 (981 - 1712)	86 (64 - 112)

In addition, the artificial tumour should have a texture similar to the model tumour. Attempts have been made to fulfill all these requirements by designing the artificial tumour as an external plastic shell, with the same shape of the model tumour, filled with a mixture of two components: (i) a matrix of expanded epoxy resin soaked in (ii) a solution of $MnCl_2$ to control the MR signal (Figure 5.11). To produce the tumour phantom a manufacturing approach combining 3D printing using photo-reactive resins and molding of soft polymers was adopted. This implied the additive fabrication of a 3D mould provided with a cavity reproducing the morphology of the tumour model, surrounded by a shell having the same surface profile. This structure was designed starting from the solid model represented in Figure 5.8 using a dedicated software. After printing, a polymeric foam was casted into the manufactured part to completely fill the mould after the expansion, resulting into a porous material with a characteristic morphological heterogeneity typical of that of the malignancy. The experimental details of the manufacturing procedure are reported here below.

The tumour-shaped shell

The tumor mould was 3D-printed using Stereolithography (SLA). This technique relies on the use of a focused laser beam to selectively polymerise a photo-reactive resin, fabricating a solid object additively layer-by-layer. The machine used was a Form 2 (FormLabs), equipped with a 405 nm laser (nominal laser spot diameter around 140 μm). The photopolymer employed was a translucent resin (FormLabs Clear), which solidifies into a tough and stiff plastic material after processing. The virtual model of the mould was generated by building a 2 mm thick shell around the tumor model reported in Figure 5.8 using the software MeshMixer (Autodesk). The design of this hollow structure (the tumour-shaped cell) was further modified by adding a fluidic port, fitting commercial syringes, for the injection of the test material and the liquid solution. The refined model was then printed with the Form 2 using the machine dedicated software, optimising the orientation of the printed part to favor the process.

The internal compartment

The fabricated mould was subsequently filled using the test material (an expanded epoxy resin from ProChima) according to a custom-made protocol developed empirically. This hydrophobic compound was prepared by mixing the prepolymer and catalyzer in a ratio equal to 1:1 vol/vol. The obtained viscous solution presents a creaming time of about 30 seconds and expands four to five times its original volume. After injection (the volume injected was 3 ml, around half of the mould volume), the foaming process took place into

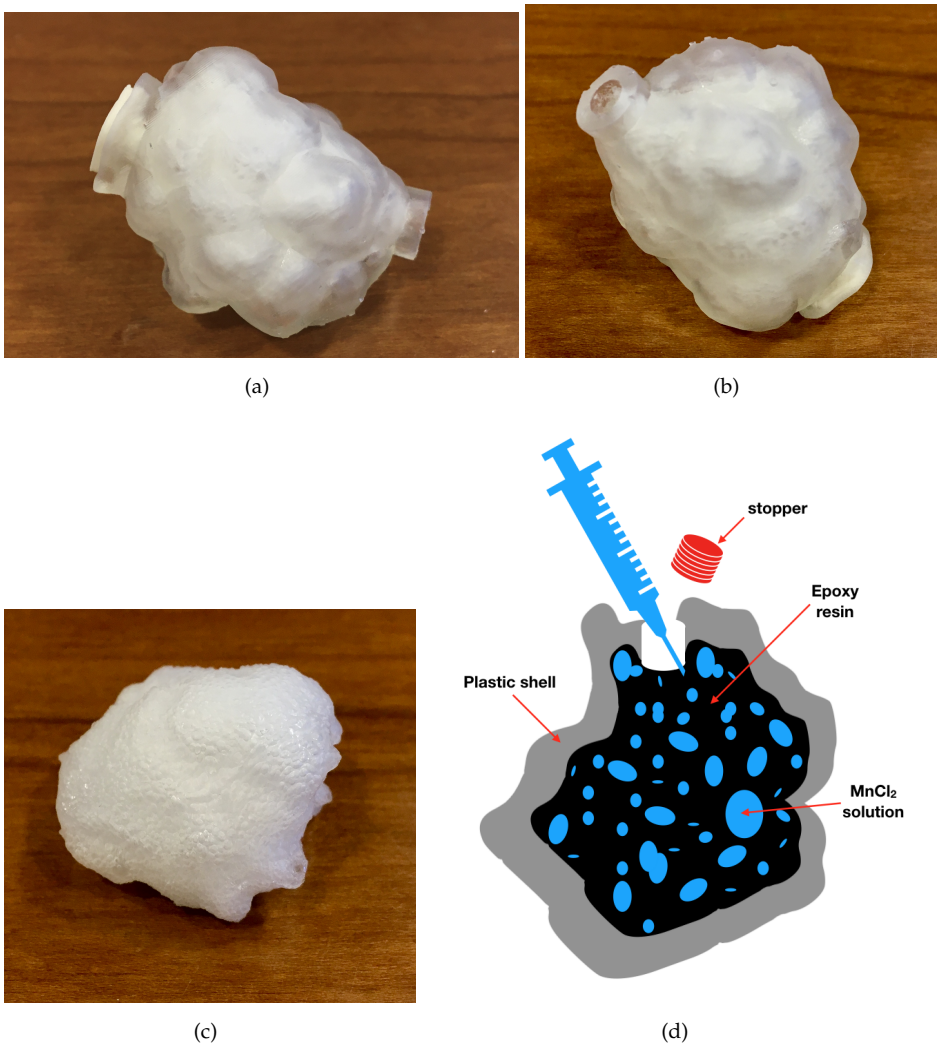


Figure 5.11: The first prototype of the artificial tumour. Pictures of the artificial tumour (a-b), which consists of two main parts: the external shell with a small cavity to allow the injection of the $MnCl_2$ solution and (c) the epoxy resin which fills the internal part. (d) Sketch of the main components.

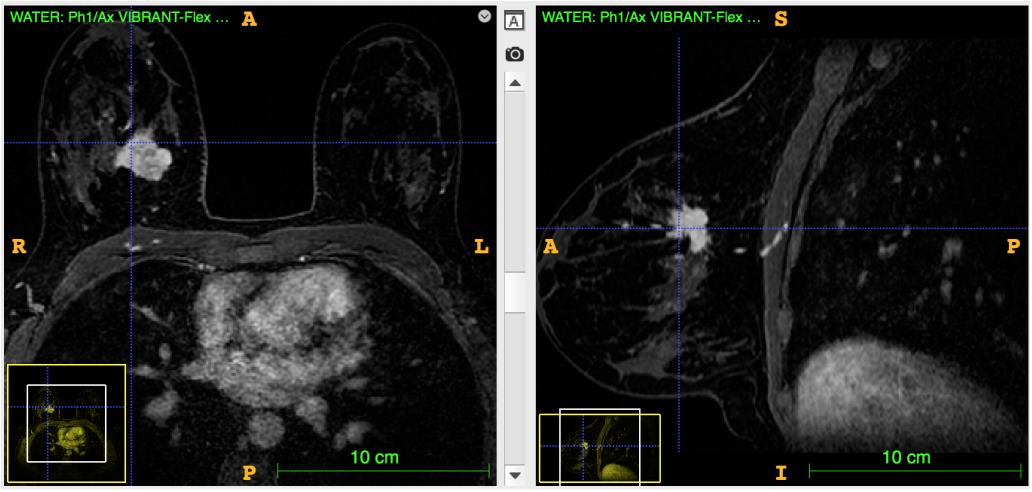
a vacuum chamber to enhance the material porosity. The expanded resin fully occupied the volume encased by the shell. Complete polymerization of the material took place in 60 minutes.

The epoxy resin serves as a matrix of porous material which creates a fine heterogeneous texture on the T_1-w image. This specific material for the realisation of the artificial tumour has been identified after a test on 28 different materials, described in Appendix B. After the foaming, an aqueous solution of $MnCl_2$ was injected inside the shell through the port, which then was closed with a stopper, to isolate the internal artificial tumour

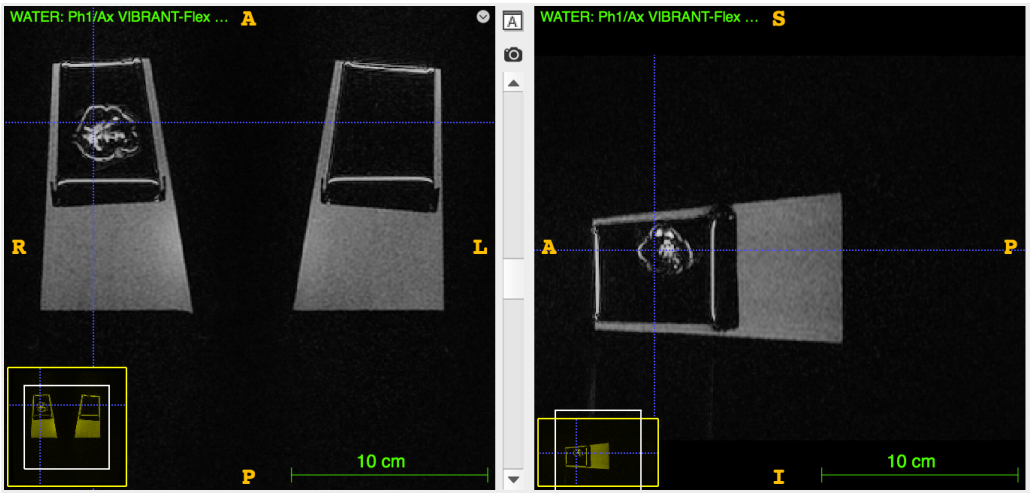
from the peanut oil in which will be immersed. The first prototype of the artificial tumour was filled with a 0.1 mM $MnCl_2$ solution. The ion concentration was established on the basis of the calibration curves described in Section 4.1.2 to obtain the spin-lattice relaxation time of the tumour T_1 = 1322 ms, as in Table 5.2.

5.2 Preliminary validation and ongoing studies

The first assembled prototype of BREATH was positioned inside a phased array breast coil (8 channel) with the usual setup of breast imaging. It was scanned with the T_1-w dynamic sequence with the following imaging parameters: 7.39 ms TR, 3.44 ms TE, 1.4 mm slice thickness, 0.7 mm slice spacing, 350x350 mm^2 FoV, 352 x 352 AM. A comparison between the obtained images of the phantom and the images of the reference patient (affected by the tumour taken as a model for the phantom) are shown in Figure 5.12.
From a visual inspection of the images, the artificial tumour is able to offer inhomogeneous subregions with different grey level intensities, as desired. However, comparing the artificial and the real tumour on the image, it is evident that the artificial case shows more areas with low MR signal (black) than the real case. This observation can be quantified comparing the normalised intensity histograms of the two VOIs, shown in Figure 5.13.
The histogram has been extracted from the VOI segmented with ITK-SNAP. The ratio between the counts in the first peak (red arrow) and the second one (black arrow) is $\sim$3 for the real tumour and $\sim$15 for the artificial case, confirming that the artificial object does not create enough voxels with brighter intensities on the MR image. The reason for this result was the difficulties encountered during the injection of the $MnCl_2$ solution in the matrix of resin. In fact, the resin was very thick and the syringe needle struggled to go through the sponge. It was not possible to fill the internal part of the insert that were most distant from the cavity in which the syringe was inserted. Consequently, the ion solution was concentrated in the middle of the insert and closer to the cavity, that in fact resulted to be the brighter zones in the MR images.
To face this problem, the realisation of a second prototype of the artificial tumour is ongoing. The new prototype will have more than one cavity on the plastic surface, to facilitate the injection from several sides. Moreover, a reduced quantity of resin will be poured in the shell, with the final expected result of a less dense porose matrix.

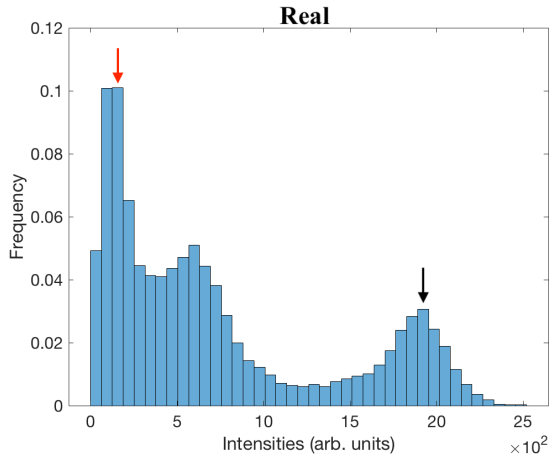


(a)

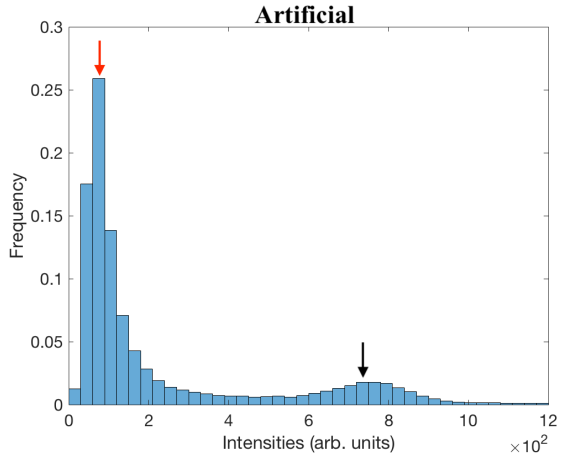


(b)

Figure 5.12: T_1-w image of (a) the reference breast cancer patient and (b) the first prototype of the BREATH phantom. The images are axial (left) and sagittal (right).



(a)



(b)

Figure 5.13: Normalised intensity histogram of (a) the real tumour and (b) the artificial tumour. The y axis is the frequency of number of voxels. The arrows indicate the peaks considered for the analysis.

5.3 Future planning

5.3.1 Validation

After the realisation of the new version of the artificial tumour, the second prototype of BREATH will be reassembled and will undergo a validation process, similar to the one applied in the case of the pelvic phantom, including:

- comparison of the artificial and real tumour in terms of intensity histogram on the MR images acquired with the T_1-w dynamic sequence;
- acquisition of relaxation times maps of the phantom and comparison with the relaxation times found *in vivo* on the healthy volunteers and patients (database 2, Table 5.1), 17 subjects);
- extraction of the radiomic features from the artificial tumour and evaluation of compatibility with the features extracted from the patients' images (database 1, Table 5.1), 84 subjects).

If the second version of BREATH will give positive results during the validation process (i.e. good agreement with the texture properties and MR signal of the patients) it will be exploited for the optimisation of the radiomic analysis in breast MRI, as described below. Otherwise, other versions of the artificial tumour will be planned to make the phantom the most representative of a real patient for radiomic purposes.

5.3.2 Repeatability and reproducibility of the radiomic features

The validated BREATH version will be scanned on different MR scanners, to test the repeatability of the radiomic features in fixed experimental conditions. The experimental design will be similar to the one planned for PETER PHAN (Figure 4.8). The repeatability will be tested on each scanner involved, including the three scanners already used for the pelvic phantom experiments and possibly additional scanners. The breast phantom will be scanned in a test-retest experiment with the identified MR sequence for the specific clinical scenario which the phantom study will support. The impact of the phantom repositioning between acquisitions will be investigated.
In parallel, the reproducibility of the radiomic features varying the scanner or the imaging parameters will be studied. In particular, the parameters will be varied in the range identified on the clinical images in the database of interest. The set of the most unstable features (non-repeatable and/or non-reproducible) will be identified and the selected features will be excluded *a priori* from the analysis on the clinical database.
In addition, a study on the ability of the radiomic features to properly quantify the texture under a certain volume threshold will be performed. In fact, some features may lose their meaning (according to their mathematical definition) when extracted from small volume lesions. This property will be investigated on the breast phantom and could be useful to guide the feature selection process and/or prove useful criteria for patient enrollment when designing clinical studies.

5.3.3 Application to two clinical scenarios

The methodological studies on BREATH (described in the previous paragraph) will be coupled with two planned clinical studied, currently ongoing at the IEO.
The first study is on database 1 (Table 5.1), described in 5.1.3. The primary aim of the study is to investigate the association between the radiomics biomarkers extracted from the breast lesion and the pathological response to the neoadjuvant therapy. The secondary aim is to study the association between the radiomic features and the biological profile of the tumour, obtained from the surgical specimen. A tertiary endpoint of the study is the investigation of the prognostic value of the radiomic parameters when extracted from the MRI executed before and, when available, after the neadjuvant treatment. The radiomic features will be extracted from the images acquired with a T_1 dynamic sequence (first arterial phase after the contrast agent injection).
The second study in on database 3 (Table 5.1), which is a retrospective dataset including 183 breast cancer patients who underwent a breast MRI before surgery. The primary aim is to investigate the correlation between the breast tumour subtypes (in terms of the expression of ER, PR, and HER2 receptors and evaluation of Ki-67) and the radiomic features extracted from the lesion segmented on the MR images (diffusion contrast enhanced sequence with contrast agent injection). The secondary aim is to verify if, in presence of secondary lesions, the radiomic features would be able to differentiate the patients in two subgroups: (i) the patients in which all the lesions have the same biological and histological parameters and (ii) the patients in which the biological and histological parameters of the principal lesion are different from the parameters of the satellite lesions.
In both these studies, a parallel investigation on the breast phantom will ensure the robustness of the radiomic analysis, serving as first feature selection method and providing indications on the volume threshold under which some features may be unstable or unreliable.

Conclusions and future perspectives

In this thesis work, the application of radiomics to MRI was investigated under various methodological aspects. A dedicated phantom to support MRI-based radiomics on pelvic cancer was developed and validated. The phantom, called PETER PHAN (PElvis TExtuRe PHANtom), included inserts mimicking the texture and the MR signal (in terms of the relaxation times T_1 and T_2) of a set of identified patients, overcoming in this way the limitations of already existing homogeneous phantoms for MRI. PETER PHAN was the first MRI phantom dedicated to radiomic analysis of the pelvis, meeting the demand of the present MRI-based radiomics [14]. The procedure for assembling the phantom was published to extend its use in the radiomics community, towards the harmonisation and standardisation of the procedure [31].

The phantom was exploited for an initial analysis on three available radiomic software, namely *IBEX*, *LIFEx*, and *PyRadiomics*, that were tested to assess the impact of the choice of a specific package on the value of the radiomic features extracted from T_1-w and T_2-w images. Only 24 radiomic features were found to be defined the same (considering both nomenclature and mathematical formulas) in the three packages. Despite the same nominal definition, consistent discrepancies were found in the results obtained with different software and they were attributable to different identified sources of error, including software-specific algorithm implementation, customised settings for texture matrixes calculation (especially the Grey Level Run Length Matrix), variability in the user-dependent parameters (for image preprocessing and discretisation) and incompatibility in the readable file format. The findings extend the study performed by Foy et al. [98] on the variation in algorithm implementation across others radiomic software. The results suggested that not all the available radiomic software show agreement on the mathematical definition and calculation of the radiomic features. As a consequence, the choice of a radiomic package should be conducted with awareness and guided by the IBSI indications, being constantly updated. This preliminary investigation was useful also to identify a single tool (*PyRadiomics*) that was exploited for all the following radiomic analysis.

Afterwards, PETER PHAN was used in a multicentric study to assess the radiomic features repeatability when extracted from T_2-w images acquired on three different scanners (two vendors -GE and Philips- , two magnetic field strengths -1.5 T and 3 T-). The

results showed a decrease in the number of highly repeatable features when the phantom was repositioned before the second acquisition in a test-retest study. The percentage of repeatable features (ICC > 0.9) after phantom repositioning varied from 11.2% to 85.4% depending on the considered scanner. The reproducibility of the features at fixed imaging parameters and varying the field strength or the vendor was evaluated as well, with a consistent number of features resulting not reproducible. Based on the specific considered scenario, 19.2% to 87.9% of the features showed poor reproducibility, i.e. CCC ≤ 0.5. The features identified as most stable in this study included Energy and Total Energy among the first order features, the Run Length Non Uniformity of the GLRLM-based features and the Coarseness of the NGTDM-based features. The variation of the Time of Echo (TE) or the Repetition Time (TR) in the range of clinical imaging when acquiring T_2-w images of the phantom resulted to have an impact on the reproducibility of the radiomic features. A decrease in the features reproducibility was particularly evident when increasing the gap in TE, in the range 80-120 ms, of two compared T_2-w images. All these observations highlighted the need for a phantom study to be coupled with a specific clinical scenario of interest, in order to identify and exclude the non-repeatable or non-reproducible radiomic features before building a prediction model. This thesis explored few clinical scenarios, providing a well-defined workflow for the identification of non-robust radomic features, that can be adapted for specific clinical needs.
The dependence of the texture features on shape features was investigated in a preliminary study on selected MR images, leading to identify a subset of 20 features nominally belonging to the texture features but not carrying the informative content they were expected to quantify. In addition, an investigation on the dependence of the feature values on the dimension of the volume of interest was carried out. The preliminary results suggested that some features tended to loose their ability to distinguish different textures below a volume threshold of 0.5 cm^3. Both these last investigations, despite limited to a specific scenario, allowed to identify a set of low-quality texture features that should be excluded *a priori* in a clinical study. The specific identified features may not be universal, but a similar procedure can be adapted in other scenarios.
The last part of the thesis work was dedicated to the development of a breast phantom specifically designed for MRI-based radiomics, not previously described in the literature. The potential and novelty of such phantom include the closeness to an *in vivo* situation, reached by fabricating an object which mimics the geometry, MR signal and texture properties of a real breast with a tumour. These characteristics optimise and push forward the attempts already made by exploiting fruits as test objects to study the radiomic features robusteness in MRI [135]. Moreover, with respect to previous fruit phantoms, it can be easily shared among medical centres to guarantee reproducibility and promote a comprehensive multicentric investigation on an identical object. The first prototype of the phantom, called BREATH (BREAst Texture pHantom), is already available and made use of a breakthrough material, namely expanded epoxy resin, to mimic the texture of a breast tumour, encased in a 3D-printed mould based on a virtual model obtained from an MRI scan of a real tumour. This phantom will be used for the optimisation of radiomic protocols in breast MRI, following an approach similar to the one used with the pelvic phantom.

Future perspectives

Regarding the pelvic phantom, the extension of the repeatability and reproducibility studies to different scenarios will be considered, including the acquisition of images on others MRI scanners and the variation of other imaging parameters besides TE and TR. A comprehensive investigation on the impact of the dimension of the volume of interest on the radiomic features value is planned, along with an extended study on the dependence of the texture features on the shape properties. All these studies might be performed also considering pulse sequences different from the ones already tested. The software comparison could be extended considering other tools, possibly commercial packages that should adhere more closely to the IBSI indications for the harmonisation of the features extraction.
Concerning the breast phantom, a new prototype of the artificial tumour is being tested and the assembled phantom in its second version will be validated in the near future. BREATH will be acquired on different scanners, with a schedule similar to the study design presented for the radiomic experiments on PETER PHAN. Comparable analysis will be carried out to identify the most repeatable and reproducible high-quality radiomic features. These methodological studies will be performed in support of two planned radiomic analyses, involving two distinct database of breast cancer patients, that have been already identified.

Appendices

APPENDIX A

Definition of selected radiomic features

This Appendix provides the definitions of the radiomic features used in Chapter 3 for the investigation on the compatibility of the feature values when extracted from different software packages. The nomenclature and formulas are recommended by the IBSI [18] in the reference manual version v9 (accessible at this address).

A.1 Morphological features

Morphological features describe the shape and geometrical properties of the Volume of Interest (VOI). Let V be the volume of the VOI, A the surface area of the VOI, N_v the number of voxels in the VOI and V_k the volume of voxel k. On this basis, the following features can be defined.

A.1.1 Sphericity

Sphericity express the degree of roundness of the considered VOI.

$$Sphericity = \frac{\left(36\pi V^2\right)^{\frac{1}{3}}}{A} \tag{A.1.1}$$

A.1.2 Volume (voxel counting)

This feature indicates how many voxels compose the VOI.

$$Volume(vox.count) = \sum_{k=1}^{N_v} V_k \tag{A.1.2}$$

A.2 Intensity histogram features

The features of this subset quantitatively describe the intensity histogram obtained after discretising the intensity values of the voxels in the VOI into intensity bins. Let $\vec{X}_d =$

$(X_{d,1}, X_{d,2}, ..., X_{d,N_v})$ be the set of N_g grey levels in which the VOI intensties have been discretised. The mean discretised intensity μ is given by:

$$\mu = \frac{1}{N_v} \sum_{k=1}^{N_v} X_{d,k} \tag{A.2.1}$$

Thus, the following features can be introduced.

A.2.1 (Excess) discretised intensity kurtosis

This feature, along with *skewness* described below, aims at quantifying the shape of the intensity histogram. In particular, *kurtosis* refers to the "peakedness" of the distribution. A higher value of *kurtosis* indicates that the contribution of the tails is greater with respect to the mean value. On the other hand, a lower *kurtosis* describes a distribution concentrated around the mean value.

$$Kurtosis = \frac{\frac{1}{N_v} \sum_{k=1}^{N_v} (X_{d,k} - \mu)^4}{\left(\frac{1}{N_v} \sum_{k=1}^{N_v} (X_{d,k} - \mu)^2\right)^2} - 3 \tag{A.2.2}$$

A.2.2 Discretised intensity skewness

Skewness quantifies the asymmetry of the intensity distribution inside the VOI. A distribution can be symmetrical or either have a positive or negative *skewness*, depending on the weight of the right or left tail, respectively.

$$Skewness = \frac{\frac{1}{N_v} \sum_{k=1}^{N_v} (X_{d,k} - \mu)^3}{\left(\frac{1}{N_v} \sum_{k=1}^{N_v} (X_{d,k} - \mu)^2\right)^{3/2}} \tag{A.2.3}$$

A.3 Grey level co-occurence based features

The Grey Level Co-occurrence Matrix (GLCM) has been defined in Chapter 2 and quantifies how many times a pair of voxels with a certain grey level and at fixed distance occurs in the VOI. Let N_g be the total number of grey levels in which the VOI intensity mask has been discretised, $i = 1, ..., N_g$ and $j = 1, ..., N_g$. Let p_{ij} be the normalised co-occurrence matrix of the VOI, $p_{i.} = \sum_{j=1}^{N_g} p_{ij}$ the row marginal probability and $p_{.j} = \sum_{i=1}^{N_g} p_{ij}$ the column marginal probability. Let $p_{i-j,l} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_{ij}$ be the diagonal probabilities, with $l = |i - j|$ and $l = 0, 1, ..., N_g - 1$. The following GLCM-based features can be extracted.

A.3.1 Inverse difference

Inverse difference quantifies the local homogeneity in the VOI intensities. The higher the feature value the higher the homogeneity (i.e. locally same grey levels).

$$Inv.\ Difference = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p_{ij}}{1+|i-j|} \tag{A.3.1}$$

A.3.2 Angular second moment

The *Angular second moment*, called also *Energy*, is another measure of homogeneity in the VOI. The higher the energy the higher the occurrences of pairs (i, j) in the VOI.

$$Angular\ second\ moment = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(p_{ij})^2 \tag{A.3.2}$$

A.3.3 Contrast

Contrast quantifies the local variations of intensity levels within the VOI, with a higher value associated to greater variations of grey levels in neighbouring voxels.

$$Contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-j)^2\,p_{ij} \tag{A.3.3}$$

A.3.4 Correlation

Correlation is a measure of the linear dependency of grey levels within the VOI.

$$Correlation = \frac{1}{\sigma_{i.}\sigma_{.j}}\left(-\mu_{i.}\mu_{.j} + \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijp_{ij}\right) \tag{A.3.4}$$

where $\mu_{i.} = \sum_{i=1}^{N_g} ip_{i.}$ and $\sigma_{i.} = \left(\sum_{i=1}^{N_g}(i-\mu_{i.})^2 p_{i.}\right)^{1/2}$ are the mean and standard deviation of $p_{i.}$. Similarly, $\mu_{.j}$ and $\sigma_{.j}$ are the mean and standard deviation of $p_{.j}$.

A.3.5 Joint entropy

Joint entropy gives an indication on the randomness of grey levels in neighbouring voxels.

$$Joint\ entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_{ij}\ log_2\ p_{ij} \tag{A.3.5}$$

A.3.6 Difference average

Difference average quantifies the variation in grey levels, with a higher value indicating a greater disparity in neighbouring voxels.

$$Difference\ average = \sum_{l=0}^{N_g-1} l\, p_{i-j,l} \tag{A.3.6}$$

A.4 Grey level run length based features

The Grey Level Run Length Matrix (GLRLM) has been defined in Chapter 2 and computes the run lengths in the VOI. Let N_r be the maximum length of a run observed in the image, $i = 1, ..., N_g$ and $j = 1, ..., N_r$. Let r_{ij} be the element of the GLRLM of the VOI and $N_s = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} r_{ij}$. Let $r_{i.}$ and $r_{.j}$ be defined as: $r_{i.} = \sum_{j=1}^{N_r} r_{ij}$ and $r_{.j} = \sum_{i=1}^{N_g} r_{ij}$. The following GLRLM-based features can be extracted.

A.4.1 Short run emphasis

Short run emphasis quantifies the short run lengths in the VOI. A higher value is associated with a fine-grained texture.

$$Short\ run\ emphasis = \frac{1}{N_s} \sum_{j=1}^{N_r} \frac{r_{.j}}{j^2} \tag{A.4.1}$$

A.4.2 Long run emphasis

Long run emphasis quantifies the long run lengths in the VOI. A higher value is associated with a coarse texture.

$$Long\ run\ emphasis = \frac{1}{N_s} \sum_{j=1}^{N_r} j^2 r_{.j} \tag{A.4.2}$$

A.4.3 Low grey level run emphasis

Low grey level run emphasis emphasises the low intensity levels in the VOI. A higher value indicates a higher occurrence of low grey levels.

$$Low\ grey\ level\ run\ emphasis = \frac{1}{N_s} \sum_{i=1}^{N_g} \frac{r_{i.}}{i^2} \tag{A.4.3}$$

A.4.4 High grey level run emphasis

High grey level run emphasis emphasises the high intensity levels in the VOI. A higher value indicates a greater occurrence of high grey levels in the VOI.

$$High\ grey\ level\ run\ emphasis = \frac{1}{N_s}\sum_{i=1}^{N_g} i^2 r_{i.} \tag{A.4.4}$$

A.4.5 Short run low grey level emphasis

Short run low grey level emphasis emphasis the GLRLM cells characterised with both short run and low grey levels.

$$Short\ run\ low\ grey\ level\ emphasis = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{r_{ij}}{i^2 j^2} \tag{A.4.5}$$

A.4.6 Short run high grey level emphasis

Short run high grey level emphasis emphasis the GLRLM cells characterised with both short run and high grey levels.

$$Short\ run\ high\ grey\ level\ emphasis = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{i^2 r_{ij}}{j^2} \tag{A.4.6}$$

A.4.7 Long run low grey level emphasis

Long run low grey level emphasis emphasis the GLRLM cells characterised with both long run and low grey levels.

$$Long\ run\ low\ grey\ level\ emphasis = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{j^2 r_{ij}}{i^2} \tag{A.4.7}$$

A.4.8 Long run high grey level emphasis

Long run high grey level emphasis emphasis the GLRLM cells characterised with both long run and high grey levels.

$$Long\ run\ high\ grey\ level\ emphasis = \frac{1}{N_s}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} i^2 j^2 r_{ij} \tag{A.4.8}$$

A.4.9 Grey level non-uniformity

Grey level non-uniformity quantifies the distribution of runs over the grey levels. A lower value indicates that the runs are evenly distributed over the grey levels.

$$Grey\ level\ non\ uniformity = \frac{1}{N_s}\sum_{i=1}^{N_g} r_{i.}^2 \tag{A.4.9}$$

A.4.10 Run length non-uniformity

Run length non-uniformity quantifies the distribution of runs over the run lengths. A lower value indicates that the runs are evenly distributed over the run lengths.

$$Run\ length\ non\ uniformity = \frac{1}{N_s}\sum_{j=1}^{N_r} r_{.j}^2 \tag{A.4.10}$$

A.4.11 Run percentage

Run percentage quantifies the coarseness of a texture, being defined as the ratio of number of effective runs and total number of possible runs in the VOI.

$$Run\ percentage = \frac{N_s}{N_v} \tag{A.4.11}$$

A.5 Neighbourhood grey tone difference based features

The Neighbourhood Grey Tone Difference Matrix (NGTDM) has been defined in Chapter 2 and is based on the difference of grey level between one voxel and its surrounding neighbours. Let s_i be a matrix element of the NGTDM, N_n the number of voxels with a valid neighbourhood, n_i the number of voxels of grey level i and with a valid neighbourhood, and $p_i = n_i/N_n$ the grey level probability. Let $N_{g,p} \leq N_g$ be the number of grey levels with $p_i > 0$ and $N_{v,c} = \sum n_i$ the number of voxels with at least one neighbour. The following NGTDM-based features can be extracted.

A.5.1 Coarseness

Coarseness quantifies the spatial rate of the change in intensity values within the VOI. A higher value of this feature is associated with a more uniform texture.

$$Coarseness = \frac{1}{\sum_{i=1}^{N_g} p_i s_i} \tag{A.5.1}$$

A.5.2 Contrast

A high *Contrast* is typical of a VOI showing large differences between the grey level of a voxel and the levels of its neighbours.

$$Contrast = \left(\frac{1}{N_{g,p}\left(N_{g,p}-1\right)} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_i p_j (i-j)^2 \right) \left(\frac{1}{N_{v,c}} \sum_{i=1}^{N_g} s_i \right) \tag{A.5.2}$$

with $p_i \neq 0$ and $p_j \neq 0$.

A.5.3 Busyness

Busyness indicates how much the texture in the VOI is *busy*. A texture is busy when there are large differences between the grey level of a voxel and the levels of its neighbours.

$$Busyness = \frac{\sum_{i=1}^{N_g} p_i s_i}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |ip_i - jp_j|} \tag{A.5.3}$$

with $p_i \neq 0$ and $p_j \neq 0$.

APPENDIX B

Selection of the material for the artificial breast tumour

This Appendix describes the tests performed on samples of different materials, to identify the one suitable for the realisation of the porous matrix inside the artificial breast tumour. The desired material should be non-homogeneous and should create a pattern of grey levels on the MR images acquired with a selected sequence.

B.1 Preliminary test

The materials tested were provided by the Interdisciplinary Centre for Nanostructured Materials and Interfaces (CIMaINa, Milan) and are listed in Table B.1. The samples consisted in cylindrical objects with a $\sim$2 cm diameter and $\sim$1 cm height. They were fixed on two bottle-shaped phantoms usually exploited for the quality controls on the considered MR scanner (1.5 T Optima MR450W, General Electric Healthcare), available at the European Institute on Oncology, Milan. The setup is shown in Figure B.1. The bottle-shaped phantoms contained 3.75 g $NiSO_4$ x 6 H_2O and 5 g NaCl per 1000 g distilled H_2O. Four samples were attached to each bottle (two for opposite sides of the bottle) and the procedure repeated for all the available test objects. To easily identify the objects on the MR image, a silicone marker was added as a reference on one side of the bottle. The two bottles were then positioned inside the breast coil normally used in the setup for breast imaging on the MR scanner. The images were acquired with a T_1-w sequence, with imaging parameters TR 560 ms, TE 20 ms, slice thickness 5.5 mm, slice spacing 5.9 mm, FoV 400x400 mm, and with a T_2-w sequence with TR 3126 ms, TE 93 ms, slice thickness 5.5 mm, slice spacing 5.9 mm, FoV 400x400 mm. Examples of the obtained MR T_1-w and T_2-w images are shown in Figure B.2. From this test, only 6 materials gave a visible signal on the obtained MR images: the expanded epoxy resin (19), the silicon rubber (20) and some of the polyurethane foams and elastomers (26, 27, 28, 29). The remaining materials were transparent on the images.

B.2 Targeted test

A second experiment was performed on materials 19, 20 and 28 (selected as representative of the polyurethane foams) in a geometry more similar to the breast phantom that had to be realised (closer to a real patient). The samples were immersed in two cups (20-28 in the same cup, material 19 in the other) filled with a 0.1 mM $MnCl_2$ solution,

Table B.1: List of materials tested on the MR scanner. Materials from 23 to 28 are commercial polyurethane foams and elastomers.

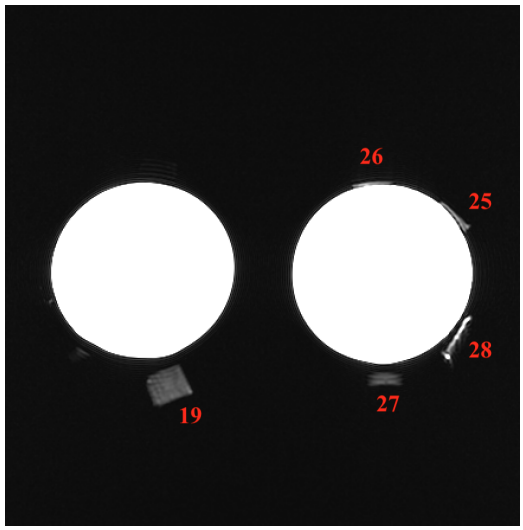
Sample number	Material
1	Polymethylmethacrylate (PMMA)
2	High molecular weight polyethylene (HMW-PE)
3	Glass-filled nylon
4	Polyvinyl chloride (PVC 1)
5	Polyvinyl chloride (PVC 3)
6	Polyvinyl chloride (PVC 4)
7	ABS natural 250 °C 100 % infill
8	ABS natural 250 °C 25 % infill
9	ABS clear 250 °C 100 % infill
10	ABS clear 250 °C 25 % infill
11	PET clear 230 °C 100 % infill
12	PET clear 230 °C 25 % infill
13	TPU clear 230 °C 100 % infill
14	TPU clear 230 °C 25 % infill
15	Form2 STD GREY 100% infill
16	Form2 Tough 100% infill
17	Polyurethane foam
18	Expanded silicon rubber
19	Expanded epoxy resin
20	Silicon rubber
21	Poli(HEMA-co-AN) - 1
22	Poli(HEMA-co-AN) - 2
23	Smooth-On pmc 780 dry
24	Smooth-On ecoflex 00-20
25	Smooth-On vytaflex 30
26	Smooth-On ecoflex 00-10
27	Smooth-On moldstar 15
28	Smooth-On ecoflex 00-30

as shown in B.3 (a). The cups were positioned inside the breast coil, as in the previous setup, and scanned with the T_1-w dynamic sequence included in the usual protocol for breast imaging. This sequence was chosen as it was the one of interest for acquiring the final breast phantom. The imaging parameters were TR 7.2 ms, TE 3.44 ms, slice thickness 1.4 mm, slice spacing 0.7 mm, FoV 350x350 mm. One of corresponding Maximum Intensity Projection (MIP) image is shown in B.3 (b). From this experiment, the expanded epoxy resin (19) showed an heterogeneous signal on the MR images (B.3b), also favored by its intrinsic porosity, whilst the other materials gave a homogeneous response. Thus, the resin resulted to be the most suitable material for the realisation of the artificial breast tumour.

(a)

(b)

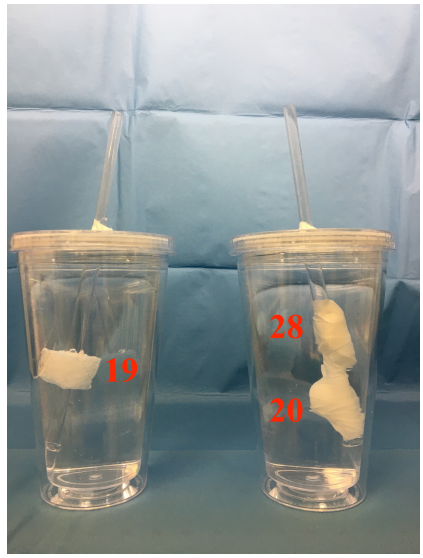
Figure B.1: Experimental setup. (a) The samples of different materials were fixed on the surface of preexisting phantoms. A silicone marker (attached between the two samples) was added to easily locate and distinguish the test objects on the MR images. (b) Each phantom was positioned inside the breast coil in the usual setup for diagnostic imaging.



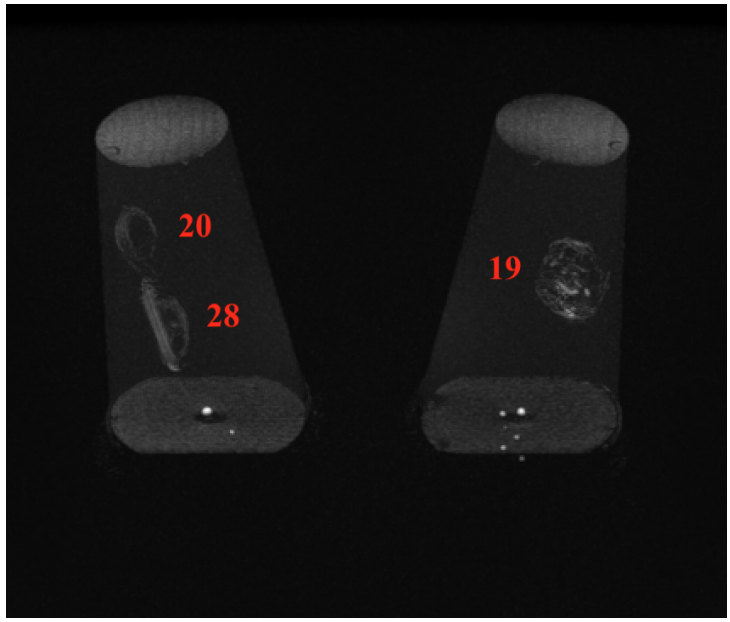
(a)

(b)

Figure B.2: Coronal (a) T_1-w and (b) T_2-w image of the tested materials attached on the bottle-shaped phantoms. The sample number is indicated in red.



(a)



(b)

Figure B.3: (a) Experimental setup for the targeted test. (b) MIP image of the phantom in the new configuration. The sample number is indicated in red.

Bibliography

[1] E. M. Purcell, H. C. Torrey, and R. V. Pound, "Resonance absorption by nuclear magnetic moments in a solid," *Phys. Rev*, vol. 69, no. 1-2, p. 37, 1946.

[2] F. Bloch, "Nuclear induction," *Physical Review*, vol. 70, no. 7-8, pp. 460–474, 1946.

[3] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, "Magnetic Resonance Imaging - Physical Principles and Sequence Design," 1999.

[4] R. W. Brown, Y.-c. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design Second Edition*. 2014.

[5] R. A. De Graaf, *In vivo NMR spectroscopy: principles and techniques*. 2019.

[6] E. L. Hahn, "Spin echoes," *Physical Review*, vol. 80, no. 4, p. 580, 1950.

[7] E. M. Carr, Herman Y. and Purcell, "Effects of diffusion on free precession in nuclear magnetic resonance experiments," *Physical Review*, vol. 94, no. 3, p. 630, 1954.

[8] S. Meiboom and D. Gill, "Modified spin-echo method for measuring nuclear relaxation times," *Review of Scientific Instruments*, vol. 29, no. 8, pp. 688–691, 1958.

[9] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.

[10] R. J. Gillies, A. R. Anderson, R. A. Gatenby, and D. L. Morse, "The biology underlying molecular imaging in oncology: from genome to anatome and back again," *Clinical Radiology*, vol. 65, no. 7, pp. 517–521, 2010.

[11] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[12] T. Cheng and X. Zhan, "Pattern recognition for predictive, preventive, and personalized medicine in cancer," *EPMA Journal*, vol. 8, no. 1, pp. 51–60, 2017.

[13] M. Aiello, C. Cavaliere, A. D'Albore, and M. Salvatore, "The Challenges of Diagnostic Imaging in the Era of Big Data," *Journal of Clinical Medicine*, vol. 8, no. 3, p. 316, 2019.

[14] P. Lambin, R. T. Leijenaar, T. M. Deist, J. Peerlings, E. E. De Jong, J. Van Timmeren, S. Sanduleanu, R. T. Larue, A. J. Even, A. Jochems, Y. Van Wijk, H. Woodruff, J. Van Soest, T. Lustberg, E. Roelofs, W. Van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger, and S. Walsh, "Radiomics: The bridge between medical imaging and personalized medicine," *Nature Reviews Clinical Oncology*, vol. 14, no. 12, pp. 749–762, 2017.

[15] R. T. Larue, G. Defraene, D. De Ruysscher, P. Lambin, and W. Van Elmpt, "Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures," *British Journal of Radiology*, vol. 90, no. 1070, 2017.

[16] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bellomi, "Radiomics: the facts and the challenges of image analysis," *European Radiology Experimental*, vol. 2, no. 1, p. 36, 2018.

[17] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, "Radiomics: The process and the challenges," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.

[18] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," *arXiv preprint arXiv:1612.07003*.

[19] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, 2014.

[20] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present... any future?," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 1, pp. 151–165, 2017.

[21] V. Parekh and M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Review of Precision Medicine and Drug Development*, vol. 1, no. 2, pp. 207–226, 2016.

[22] M. Amadasun and R. King, "Texural Features Corresponding to Texural Properties," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 5, pp. 1264–1274, 1989.

[23] A. Larroza, V. Bodí, and D. Moratal, "Texture analysis in magnetic resonance imaging: review and considerations for future applications," *Assessment of cellular and organ function and dysfunction using direct and derived MRI methodologies*, pp. 75–106,

2016.

[24] L. Zhang, D. V. Fried, X. J. Fave, L. A. Hunter, J. Yang, and L. E. Court, "IBEX: An open infrastructure software platform to facilitate collaborative work in radiomics," *Medical Physics*, vol. 42, no. 3, pp. 1341–1353, 2015.

[25] C. Nioche, F. Orlhac, S. Boughdad, S. Reuze, J. Goya-Outi, C. Robert, C. Pellot-Barakat, M. Soussan, F. erique Frouin, and I. Buvat, "LIFEx: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity," *Cancer Research*, vol. 78, no. 16, pp. 4786–4789, 2018.

[26] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[27] J. C. Peeken, M. Bernhofer, B. Wiestler, T. Goldberg, D. Cremers, B. Rost, J. J. Wilkens, S. E. Combs, and F. Nüsslin, "Radiomics in radiooncology – Challenging the medical physicist," *Physica Medica*, vol. 48, no. March, pp. 27–36, 2018.

[28] N. Papanikolaou and J. Santinha, "An Introduction to Radiomics : Capturing Tumour Biology in Space and Time," *Oncologic Imaging*, vol. 3, no. 1, pp. 61–71, 2018.

[29] U. Schick, F. Lucia, G. Dissaux, D. Visvikis, B. Badic, I. Masson, O. Pradier, V. Bourbonne, and M. Hatt, "MRI-derived radiomics: Methodology and clinical applications in the field of pelvic oncology," *British Journal of Radiology*, vol. 92, no. 1104, 2019.

[30] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," *BMJ (Online)*, vol. 350, no. January, pp. 1–9, 2015.

[31] S. Nougaret, H. Tibermacine, M. Tardieu, and E. Sala, "Radiomics: an Introductory Guide to What It May Foretell," *Current Oncology Reports*, vol. 21, no. 8, 2019.

[32] I. Castiglioni and M. C. Gilardi, "Radiomics: is it time to compose the puzzle?," *Clinical and Translational Imaging*, vol. 6, no. 5, pp. 411–413, 2018.

[33] M. Avanzo, J. Stancanello, and I. El Naqa, "Beyond imaging: The promise of radiomics," *Physica Medica*, vol. 38, pp. 122–139, 2017.

[34] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Materka, and S. Trattnig, "Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: An application-oriented study," *Medical Physics*, vol. 36, no. 4, pp. 1236–1243, 2009.

[35] D. L. Raunig, L. M. McShane, G. Pennello, C. Gatsonis, P. L. Carson, J. T. Voyvodic, R. L. Wahl, B. F. Kurland, A. J. Schwarz, M. G{\"o}nen, Z. Gudrun, M. Kondratovich, K. O'Donnell, N. Petrick, P. E. Cole, B. Garra, and D. C. Sullivan, "Quanti-

tative Imaging Biomarkers: A Review of Statistical Methods for Technical Performance Assessment," *Statistical methods in medical research*, vol. 24, no. 1, pp. 26–27, 2015.

[36] D. C. Sullivan, N. A. Obuchowski, L. G. Kessler, D. L. Raunig, C. Gatsonis, E. P. Huang, M. Kondratovich, L. M. McShane, A. P. Reeves, D. P. Barboriak, A. R. Guimaraes, and R. L. Wahl, "Metrology standards for quantitative imaging biomarkers," *Radiology*, vol. 277, no. 3, pp. 813–825, 2015.

[37] Y. Balagurunathan, Y. Gu, H. Wang, V. Kumar, O. Grove, S. Hawkins, J. Kim, D. B. Goldgof, L. O. Hall, R. A. Gatenby, and R. J. Gillies, "Reproducibility and prognosis of quantitative features extracted from CT images," *Translational Oncology*, vol. 7, no. 1, pp. 72–87, 2014.

[38] E. Scalco, A. Belfatto, A. Mastropietro, T. Rancati, B. Avuzzi, A. Messina, R. Valdagni, and G. Rizzo, "T2w-MRI signal normalization affects radiomics features reproducibility," *Medical Physics*, 2020.

[39] L. Duron, D. Balvay, S. V. Perre, A. Bouchouicha, J. Savatovsky, J. C. Sadik, I. Thomassin-Naggara, L. Fournier, and A. Lecler, "Gray-level discretization impacts reproducible MRI radiomics texture features," *PLoS ONE*, vol. 14, no. 3, pp. 1–14, 2019.

[40] E. J. Limkin, R. Sun, L. Dercle, E. I. Zacharaki, C. Robert, S. Reuzé, A. Schernberg, N. Paragios, E. Deutsch, and C. Ferté, "Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology," *Annals of Oncology*, vol. 28, no. 6, pp. 1191–1206, 2017.

[41] W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, T. Allison, O. Arnaout, C. Abbosh, I. F. Dunn, R. H. Mak, R. M. Tamimi, C. M. Tempany, C. Swanton, U. Hoffmann, L. H. Schwartz, R. J. Gillies, R. Y. Huang, and H. J. W. L. Aerts, "Artificial intelligence in cancer imaging: Clinical challenges and applications," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 2, pp. 127–157, 2019.

[42] O. Bane, S. J. Hectors, M. Wagner, L. L. Arlinghaus, M. P. Aryal, Y. Cao, T. L. Chenevert, F. Fennessy, W. Huang, N. M. Hylton, J. Kalpathy-Cramer, K. E. Keenan, D. I. Malyarenko, R. V. Mulkern, D. C. Newitt, S. E. Russek, K. F. Stupic, A. Tudorica, L. J. Wilmes, T. E. Yankeelov, Y. F. Yen, M. A. Boss, and B. Taouli, "Accuracy, repeatability, and interplatform reproducibility of T 1 quantification methods used for DCE-MRI: Results from a multicenter phantom study," *Magnetic Resonance in Medicine*, vol. 79, pp. 2564–2575, may 2018.

[43] A. Hellerbach, V. Schuster, A. Jansen, and J. Sommer, "MRI Phantoms - Are There Alternatives to Agar?," *PLoS ONE*, vol. 8, aug 2013.

[44] B. Baessler, K. Weiss, and D. Pinto Dos Santos, "Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study," *Investigative Radiology*, vol. 54, no. 4, pp. 221–228, 2019.

[45] N. I. Niebuhr, W. Johnen, G. Echner, A. Runz, M. Bach, M. Stoll, K. Giske, S. Greilich, and A. Pfaffenberger, "The ADAM-pelvis phantom - An anthropomorphic, deformable and multimodal phantom for MRgRT," *Physics in Medicine and Biology*, vol. 64, feb 2019.

[46] F. Yang, N. Dogan, R. Stoyanova, and J. C. Ford, "Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth," *Physica Medica*, vol. 50, pp. 26–26, 2018.

[47] J. Ford, N. Dogan, L. Young, and F. Yang, "Quantitative Radiomics: Impact of Pulse Sequence Parameter Selection on MRI-Based Textural Features of the Brain," *Contrast Media and Molecular Imaging*, vol. 2018, 2018.

[48] D. Mackin, X. Fave, L. Zhang, D. Fried, J. Yang, Brian Taylor, E. Rodriguez-Rivera, C. Dodge, A. K. Jones, and L. Court, "Measuring computed tomography scanner variability of radiomics features," *Investigative Radiology*, vol. 50, no. 11, pp. 757–765, 2015.

[49] A. Forgacs, H. Pall Jonsson, M. Dahlbom, F. Daver, M. D. Difranco, G. Opposits, A. K. Krizsan, I. Garai, J. Czernin, J. Varga, L. Tron, and L. Balkay, "A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images," *PLoS ONE*, vol. 11, no. 10, pp. 1–14, 2016.

[50] A. Valladares, T. Beyer, and I. Rausch, "Physical imaging phantoms for simulation of tumor heterogeneity in PET, CT, and MRI: An overview of existing designs," *Medical Physics*, pp. 1–15, 2020.

[51] N. Patel, A. Henry, and A. Scarsbrook, "The value of MR textural analysis in prostate cancer," *Clinical Radiology*, vol. 74, no. 11, pp. 876–885, 2019.

[52] Y. Sun, H. M. Reynolds, B. Parameswaran, D. Wraith, M. E. Finnegan, S. Williams, and A. Haworth, "Multiparametric MRI and radiomics in prostate cancer: a review," *Australasian Physical and Engineering Sciences in Medicine*, vol. 42, no. 1, pp. 3–25, 2019.

[53] Z. Li, H. Li, S. Wang, D. Dong, F. Yin, A. Chen, S. Wang, G. Zhao, M. Fang, J. Tian, S. Wu, and H. Wang, "MR-Based Radiomics Nomogram of Cervical Cancer in Prediction of the Lymph-Vascular Space Invasion preoperatively," *Journal of Magnetic Resonance Imaging*, pp. 1–7, 2018.

[54] Y. Kan, D. Dong, Y. Zhang, W. Jiang, N. Zhao, L. Han, M. Fang, Y. Zang, C. Hu, J. Tian, C. Li, and Y. Luo, "Radiomic Signature as a Predictive Factor for Lymph Node Metastasis in Early-Stage Cervical Cancer," *Journal of Magnetic Resonance Imaging*, pp. 1–7, 2018.

[55] J. Meng, S. Liu, L. Zhu, L. Zhu, H. Wang, L. Xie, Y. Guan, J. He, X. Yang, and Z. Zhou, "Texture Analysis as Imaging Biomarker for recurrence in advanced cervical cancer treated with CCRT," *Scientific Reports*, vol. 8, no. 1, pp. 1–9, 2018.

[56] X. Zhou, Y. Yi, Z. Liu, W. Cao, B. Lai, K. Sun, L. Li, Z. Zhou, Y. Feng, and J. Tian, "Radiomics-Based Pretherapeutic Prediction of Non-response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer," *Annals of Surgical Oncology*, 2019.

[57] S. H. Jeon, C. Song, E. K. Chie, B. Kim, Y. H. Kim, W. Chang, Y. J. Lee, J. H. Chung, J. B. Chung, K. W. Lee, S. B. Kang, and J. S. Kim, "Delta-radiomics signature predicts treatment outcomes after preoperative chemoradiotherapy and surgery in rectal cancer," *Radiation oncology (London, England)*, vol. 14, no. 1, p. 43, 2019.

[58] K. Nie, L. Shi, Q. Chen, X. Hu, S. K. Jabbour, N. Yue, T. Niu, and X. Sun, "Rectal cancer: Assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI," *Clinical Cancer Research*, vol. 22, no. 21, pp. 5256–5264, 2016.

[59] Z. Liu, X. Y. Zhang, Y. J. Shi, L. Wang, H. T. Zhu, Z. Tang, S. Wang, X. T. Li, J. Tian, and Y. S. Sun, "Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer," *Clinical Cancer Research*, vol. 23, no. 23, pp. 7253–7262, 2017.

[60] N. Horvat, H. Veeraraghavan, M. Khan, I. Blazic, J. Zheng, M. Capanu, E. Sala, J. Garcia-Aguilar, M. J. Gollub, and I. Petkovska, "Mr imaging of rectal cancer: Radiomics analysis to assess treatment response after neoadjuvant therapy," *Radiology*, vol. 287, no. 3, pp. 833–843, 2018.

[61] X. Meng, W. Xia, P. Xie, R. Zhang, W. Li, M. Wang, F. Xiong, Y. Liu, X. Fan, Y. Xie, X. Wan, K. Zhu, H. Shan, L. Wang, and X. Gao, "Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer," *European Radiology*, 2018.

[62] D. Cusumano, N. Dinapoli, L. Boldrini, G. Chiloiro, R. Gatta, C. Masciocchi, J. Lenkowicz, C. Casà, A. Damiani, L. Azario, J. Van Soest, A. Dekker, P. Lambin, M. De Spirito, and V. Valentini, "Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer," *Radiologia Medica*, vol. 123, no. 4, pp. 286–295, 2018.

[63] O. Jalil, A. Afaq, B. Ganeshan, U. B. Patel, D. Boone, R. Endozo, A. Groves, B. Sizer, and T. Arulampalam, "Magnetic resonance based texture parameters as potential imaging biomarkers for predicting long-term survival in locally advanced rectal cancer treated by chemoradiotherapy," *Colorectal Disease*, vol. 19, no. 4, pp. 349–362, 2017.

[64] L. Ge, Y. Chen, C. Yan, P. Zhao, P. Zhang, A. Runa, and J. Liu, "Study Progress of Radiomics With Machine Learning for Precision Medicine in Bladder Cancer Management," *Frontiers in Oncology*, vol. 9, no. November, 2019.

[65] C. B. van der Pol, A. Chung, C. Lim, N. Gandhi, W. Tu, M. D. McInnes, and N. Schieda, "Update on multiparametric MRI of urinary bladder cancer," *Journal*

of Magnetic Resonance Imaging, vol. 48, no. 4, pp. 882–896, 2018.

[66] S. Gourtsoyianni, G. Doumou, D. Prezzi, B. Taylor, J. J. Stirling, N. J. Taylor, R. Glynne-jones, and V. Goh, "Primary Rectal Cancer : Repeatability of Global and Local- Regional MR Imaging Texture," *Radiology*, vol. 000, no. 2, pp. 1–10, 2017.

[67] S. Fiset, M. L. Welch, J. Weiss, M. Pintilie, J. L. Conway, M. Milosevic, A. Fyles, A. Traverso, D. Jaffray, U. Metser, J. Xie, and K. Han, "Repeatability and reproducibility of MRI-based radiomic features in cervical cancer," *Radiotherapy and Oncology*, vol. 135, pp. 107–114, 2019.

[68] M. Schwier, J. van Griethuysen, M. G. Vangel, S. Pieper, S. Peled, C. Tempany, H. J. Aerts, R. Kikinis, F. M. Fennessy, and A. Fedorov, "Repeatability of Multiparametric Prostate MRI Radiomics Features," *Scientific Reports*, vol. 9, no. 1, pp. 1–16, 2019.

[69] M. E. Mayerhoefer, P. Szomolanyi, D. Jirak, A. Materka, and S. Trattnig, "Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: An application-oriented study," *Medical Physics*, vol. 36, no. 4, pp. 1236–1243, 2009.

[70] P. Chirra, N. B. Bloch, A. Rastinehead, A. Purysko, A. Madabhushi, S. E. Viswanath, P. Leo, M. Yim, and M. Rosen, "Empirical evaluation of cross-site reproducibility in radiomic features for characterizing prostate MRI," no. March, p. 10, 2018.

[71] F. Valdora, N. Houssami, F. Rossi, M. Calabrese, and A. S. Tagliafico, "Rapid review: radiomics and breast cancer," *Breast cancer research and treatment*, vol. 169, no. 2, pp. 217–229, 2018.

[72] M. Codari, S. Schiaffino, F. Sardanelli, and R. M. Trimboli, "Artificial intelligence for breast MRI in 2008-2018: A systematic mapping review," *American Journal of Roentgenology*, vol. 212, no. 2, pp. 280–292, 2019.

[73] D. Sheth and M. L. Giger, "Artificial intelligence in the interpretation of breast cancer on MRI," *Journal of Magnetic Resonance Imaging*, pp. 1–15, 2019.

[74] R. D. Chitalia and D. Kontos, "Role of texture analysis in breast MRI as a cancer biomarker: A review," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 927–938, 2019.

[75] B. Reig, L. Heacock, K. J. Geras, and L. Moy, "Machine learning in breast MRI," *Journal of Magnetic Resonance Imaging*, 2019.

[76] Q. Zhang, Y. Peng, W. Liu, J. Bai, J. Zheng, X. Yang, and L. Zhou, "Radiomics Based on Multimodal MRI for the Differential Diagnosis of Benign and Malignant Breast Lesions," *Journal of Magnetic Resonance Imaging*, 2020.

[77] S. Bickelhaupt, D. Paech, P. Kickingereder, F. Steudle, W. Lederer, H. Daniel, M. Götz, N. Gählert, D. Tichy, M. Wiesenfarth, F. B. Laun, K. H. Maier-Hein, H. P.

Schlemmer, and D. Bonekamp, "Prediction of malignancy by a radiomic signature from contrast agent-free diffusion MRI in suspicious breast lesions found on screening mammography.," *Journal of Magnetic Resonance Imaging*, vol. 46, no. 2, pp. 604–616, 2017.

[78] V. S. Parekh and M. A. Jacobs, "Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric MRI," *npj Breast Cancer*, vol. 3, no. 1, pp. 1–8, 2017.

[79] C. Liu, J. Ding, K. Spuhler, Y. Gao, M. Serrano Sosa, M. Moriarty, S. Hussain, X. He, C. Liang, and C. Huang, "Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 1, pp. 131–140, 2019.

[80] Y. Dong, Q. Feng, W. Yang, Z. Lu, C. Deng, L. Zhang, Z. Lian, J. Liu, X. Luo, S. Pei, X. Mo, W. Huang, C. Liang, B. Zhang, and S. Zhang, "Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI," *European Radiology*, vol. 28, no. 2, pp. 582–591, 2018.

[81] M. W. Juan, J. Yu, G. X. Peng, L. J. Jun, S. P. Feng, and L. P. Fang, "Correlation between DCE-MRI radiomics features and Ki-67 expression in invasive breast cancer," *Oncology Letters*, vol. 16, no. 4, pp. 5084–5090, 2018.

[82] H. Li, Y. Zhu, E. S. Burnside, K. Drukker, K. A. Hoadley, C. Fan, S. D. Conzen, G. J. Whitman, E. J. Sutton, J. M. Net, M. Ganott, E. Huang, E. A. Morris, C. M. Perou, Y. Ji, and M. L. Giger, "MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, oncotype DX, and PAM50 gene assays," *Radiology*, vol. 281, no. 2, pp. 382–391, 2016.

[83] M. Dietzel, R. Schulz-Wendtland, S. Ellmann, R. Zoubi, E. Wenkel, M. Hammon, P. Clauser, M. Uder, I. B. Runnebaum, and P. A. Baltzer, "Automated volumetric radiomic analysis of breast cancer vascularization improves survival prediction in primary breast cancer," *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[84] J. Koh, E. Lee, K. Han, S. Kim, D. kyu Kim, J. Y. Kwak, J. H. Yoon, and H. J. Moon, "Three-dimensional radiomics of triple-negative breast cancer: Prediction of systemic recurrence," *Scientific Reports*, vol. 10, no. 1, pp. 1–9, 2020.

[85] H. Li, Y. Zhu, E. S. Burnside, E. Huang, K. Drukker, K. A. Hoadley, C. Fan, S. D. Conzen, M. Zuley, J. M. Net, E. Sutton, G. J. Whitman, E. Morris, C. M. Perou, Y. Ji, and M. L. Giger, "Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set," *npj Breast Cancer*, vol. 2, no. 1, 2016.

[86] T. Xie, Z. Wang, Q. Zhao, Q. Bai, X. Zhou, Y. Gu, W. Peng, and H. Wang, "Machine learning-based analysis of MR multiparametric radiomics for the subtype classification of breast cancer," *Frontiers in Oncology*, vol. 9, no. JUN, pp. 1–10, 2019.

[87] J. Wang, F. Kato, N. Oyama-Manabe, R. Li, Y. Cui, K. K. Tha, H. Yamashita, K. Kudo, and H. Shirato, "Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced MRI: A pilot radiomics study," *PLoS ONE*, vol. 10, no. 11, pp. 1–17, 2015.

[88] W. Guo, H. Li, Y. Zhu, L. Lan, S. Yang, K. Drukker, E. A. Morris, E. S. Burnside, G. J. Whitman, M. Giger, and Y. Ji, "Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data," *Journal of Medical Imaging*, vol. 2, no. 4, p. 041007, 2015.

[89] Y. Zhu, H. Li, W. Guo, K. Drukker, L. Lan, M. L. Giger, and Y. Ji, "Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma," *Scientific Reports*, vol. 5, no. July, pp. 1–10, 2015.

[90] A. C. Yeh, H. Li, Y. Zhu, J. Zhang, G. Khramtsova, K. Drukker, A. Edwards, S. McGregor, T. Yoshimatsu, Y. Zheng, Q. Niu, H. Abe, J. Mueller, S. Conzen, Y. Ji, M. L. Giger, and O. I. Olopade, "Radiogenomics of breast cancer using dynamic contrast enhanced MRI and gene expression profiling," *Cancer Imaging*, vol. 19, no. 1, pp. 1–11, 2019.

[91] B. Reig, L. Heacock, A. Lewin, N. Cho, and L. Moy, "Role of MRI to Assess Response to Neoadjuvant Therapy for Breast Cancer," *Journal of Magnetic Resonance Imaging*, pp. 1–20, 2020.

[92] N. M. Braman, M. Etesami, P. Prasanna, C. Dubchuk, H. Gilmore, P. Tiwari, D. Pletcha, and A. Madabhushi, "Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI," *Breast Cancer Research*, vol. 19, no. 1, pp. 1–14, 2017.

[93] K. Robinson, H. Li, L. Lan, D. Schacht, and M. Giger, "Radiomics robustness assessment and classification evaluation: A two-stage method demonstrated on multivendor FFDM," *Medical Physics*, vol. 46, no. 5, pp. 2145–2156, 2019.

[94] A. Saha, X. Yu, D. Sahoo, and M. A. Mazurowski, "Effects of MRI scanner parameters on breast cancer radiomics," *Expert Systems with Applications*, vol. 87, pp. 384–391, 2017.

[95] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[96] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research," *Journal of Digital Imaging*, vol. 31, no. 3, pp. 290–303, 2018.

[97] F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.-P. Metges, L. Corcos,

and D. Visvikis, "Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer.," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 52, no. 3, pp. 369–78, 2011.

[98] J. J. Foy, K. R. Robinson, H. Li, M. L. Giger, H. Al-Hallaq, and S. G. Armato, "Variation in algorithm implementation across radiomics software," *Journal of Medical Imaging*, vol. 5, no. 04, p. 1, 2018.

[99] A. S. S. Shapiro and M. B. Wilk, "Biometrika Trust An Analysis of Variance Test for Normality (Complete Samples) Published by : Oxford University Press on behalf of Biometrika Trust Stable," *Biometrika*, vol. 52, no. 3, pp. 591–611, 1965.

[100] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[101] L. I.-k. Lin, "A Concordance Correlation Coefficient to Evaluate Reproducibility," *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989.

[102] H. X. Barnhart, M. Haber, and J. Song, "Overall Concordance Correlation Coefficient for Evaluating Agreement among Multiple Observers," *Biometrics*, vol. 58, no. 4, pp. 1020–1027, 2002.

[103] L. Bianchini, F. Botta, D. Origgi, S. Rizzo, M. Mariani, P. Summers, P. García-polo, M. Cremonesi, and A. Lascialfari, "PETER PHAN : An MRI phantom for the optimisation of radiomic studies of the female pelvis," *Physica Medica*, vol. 71, no. C, pp. 71–81, 2020.

[104] D. R. Messroghli, A. Radjenovic, S. Kozerke, D. M. Higgins, M. U. Sivananthan, and J. P. Ridgway, "Modified look-locker inversion recovery (MOLLI) for high-resolution T 1 mapping of the heart," *Magnetic Resonance in Medicine*, vol. 52, no. 1, pp. 141–146, 2004.

[105] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, "Magnetic Resonance Fingerprinting," *Nature*, vol. 495, no. 7440, p. 187, 2013.

[106] S. C. Deoni, B. K. Rutt, and T. M. Peters, "Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state," *Magnetic Resonance in Medicine*, vol. 49, no. 3, pp. 515–526, 2003.

[107] G. Ramos-Llorden, G. Vegas-Sanchez-Ferrero, M. Bjork, F. Vanhevel, P. M. Parizel, R. San Jose Estepar, A. J. Den Dekker, and J. Sijbers, "NOVIFAST: A Fast Algorithm for Accurate and Precise VFA MRI T 1 Mapping," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2414–2427, 2018.

[108] K. Thangavel and E. Ü. Saritaş, "Aqueous paramagnetic solutions for MRI phantoms at 3 T: A detailed study on relaxivities," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 25, no. 3, pp. 2108–2121, 2017.

[109] D. Jirák, M. Dezortová, and M. Hájek, "Phantoms for texture analysis of MR images. Long-term and multi-center study," *Medical Physics*, vol. 31, no. 3, pp. 616–622, 2004.

[110] E. Samei, J. Hoye, Y. Zheng, J. B. Solomon, and D. Marin, "Design and fabrication of heterogeneous lung nodule phantoms for assessing the accuracy and variability of measured texture radiomics features in CT," *Journal of Medical Imaging*, vol. 2, no. 2, p. 021606, 2019.

[111] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[112] H. Kato, M. Kuroda, K. Yoshimura, A. Yoshida, K. Hanamoto, S. Kawasaki, K. Shibuya, and S. Kanazawa, "Composition of MRI phantom equivalent to human tissues," *Medical Physics*, vol. 32, no. 10, pp. 3199–3208, 2005.

[113] R. Rai, Y. F. Wang, D. Manton, B. Dong, S. Deshpande, and G. P. Liney, "Development of multi-purpose 3D printed phantoms for MRI," *Physics in medicine and biology*, vol. 64, p. 075010, mar 2019.

[114] L. Bianchini, J. Santinha, N. Loução, M. Figueiredo, F. Botta, D. Origgi, M. Cremonesi, E. Cassano, N. Papanikolaou, and A. Lascialfari, "A multicenter study on radiomic features from T2-weighted images of a customized MR pelvic phantom setting the basis for robust radiomic models in clinics," *submitted*, pp. 1–32, 2020.

[115] S. Yu, G. Dai, Z. Wang, L. Li, X. Wei, and Y. Xie, "A consistency evaluation of signal-to-noise ratio in the quality assessment of human brain magnetic resonance images," *BMC Medical Imaging*, vol. 18, no. 1, pp. 1–9, 2018.

[116] S. Foxley, X. Fan, S. A. Jansen, M. Zamora, E. Markiewicz, H. Al-Ahmadie, and G. S. Karczmar, "High spectral and spatial resolution MRI of age-related changes in murine prostate," *Magnetic Resonance in Medicine*, vol. 60, pp. 575–581, sep 2008.

[117] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.," *Journal of chiropractic medicine*, vol. 15, no. 2, pp. 155–63, 2016.

[118] F. J. Brooks and P. W. Grigsby, "The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake," *Journal of Nuclear Medicine*, vol. 55, no. 1, pp. 37–42, 2014.

[119] F. J. Brooks and P. W. Grigsby, "Low-order non-spatial effects dominate second-order spatial effects in the texture quantifier analysis of 18F-FDG-PET images," *PLoS ONE*, vol. 10, no. 2, pp. 1–17, 2015.

[120] X. Fave, L. Zhang, J. Yang, D. Mackin, P. Balter, D. Gomez, D. Followill, A. K. Jones, F. Stingo, and L. E. Court, "Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung

cancer," *Translational Cancer Research*, vol. 5, no. 4, pp. 349–363, 2016.

[121] M. Shafiq-Ul-Hassan, K. Latifi, G. Zhang, G. Ullah, R. Gillies, and E. Moros, "Voxel size and gray level normalization of CT radiomic features in lung cancer," *Scientific Reports*, vol. 8, no. 1, pp. 1–9, 2018.

[122] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *International Journal of Nursing Studies*, vol. 47, no. 8, pp. 931–936, 2010.

[123] M. L. Welch, C. McIntosh, B. Haibe-Kains, M. F. Milosevic, L. Wee, A. Dekker, S. H. Huang, T. G. Purdie, B. O'Sullivan, H. J. Aerts, and D. A. Jaffray, "Vulnerabilities of radiomic signature development: The need for safeguards," *Radiotherapy and Oncology*, vol. 130, pp. 2–9, 2019.

[124] H. Fukuhara, H. Kume, M. Suzuki, T. Fujimura, Y. Enomoto, H. Nishimatsu, A. Ishikawa, and Y. Homma, "Maximum tumor diameter: A simple independent predictor for biochemical recurrence after radical prostatectomy," *Prostate Cancer and Prostatic Diseases*, vol. 13, no. 3, pp. 244–247, 2010.

[125] S. I. Lee, E. Oliva, P. F. Hahn, and A. H. Russell, "Malignant tumors of the female pelvic floor: Imaging features that determine therapy: Pictorial review," *American Journal of Roentgenology*, vol. 196, no. 3 SUPPL., pp. 15–23, 2011.

[126] A. Snoeckx, P. Reyntiens, D. Desbuquoit, M. J. Spinhoven, P. E. Van Schil, J. P. van Meerbeeck, and P. M. Parizel, "Evaluation of the solitary pulmonary nodule: size matters, but do not ignore the power of morphology," *Insights into Imaging*, vol. 9, no. 1, pp. 73–86, 2018.

[127] S. Liu, X.-D. Wu, W.-J. Xu, Q. Lin, X.-J. Liu, and Y. Li, "Is There a Correlation between the Presence of a Spiculated Mass on Mammogram and Luminal A Subtype Breast Cancer?," *Korean Journal of Radiology*, vol. 17, no. 6, p. 846, 2016.

[128] C. L. Schlett, T. Hendel, J. Hirsch, S. Weckbach, S. Caspers, J. Schulz-Menger, T. Ittermann, F. Von Knobelsdorff-Brenkenhoff, S. C. Ladd, S. Moebus, C. Stroszczynski, B. Fischer, M. Leitzmann, C. Kuhl, F. Pessler, D. Hartung, Y. Kemmling, H. Hetterich, K. Amunts, M. Günther, F. Wacker, E. Rummeny, H. U. Kauczor, M. Forsting, H. Völzke, N. Hosten, M. F. Reiser, and F. Bamberg, "Quantitative, Organ-Specific Interscanner and Intrascanner Variability for 3 T Whole-Body Magnetic Resonance Imaging in a Multicenter, Multivendor Study," *Investigative Radiology*, vol. 51, no. 4, pp. 255–265, 2016.

[129] J. K. Jang, S. S. Lee, B. Kim, E. S. Cho, Y. J. Kim, J. H. Byun, B. J. Park, S. Y. Kim, and J. H. Kim, "Agreement and Reproducibility of Proton Density Fat Fraction Measurements Using Commercial MR Sequences Across Different Platforms: A Multivendor, Multi-Institutional Phantom Experiment," *Investigative Radiology*, vol. 54, no. 8, pp. 517–523, 2019.

[130] N. Papanikolaou, C. Matos, and D. M. Koh, "How to develop a meaningful ra-

diomic signature for clinical use in oncologic patients," *Cancer Imaging*, pp. 1–10, 2020.

[131] P. Chirra, P. Leo, M. Yim, B. N. Bloch, A. R. Rastinehad, A. Purysko, M. Rosen, A. Madabhushi, and S. E. Viswanath, "Multisite evaluation of radiomic feature reproducibility and discriminability for identifying peripheral zone prostate tumors on MRI," *Journal of Medical Imaging*, vol. 6, no. 02, p. 1, 2019.

[132] K. E. Keenan, M. Ainslie, A. J. Barker, M. A. Boss, K. M. Cecil, C. Charles, T. L. Chenevert, L. Clarke, J. L. Evelhoch, P. Finn, D. Gembris, J. L. Gunter, D. L. Hill, C. R. Jack, E. F. Jackson, G. Liu, S. E. Russek, S. D. Sharma, M. Steckner, K. F. Stupic, J. D. Trzasko, C. Yuan, and J. Zheng, "Quantitative magnetic resonance imaging phantoms: A review and the need for a system phantom," *Magnetic Resonance in Medicine*, vol. 79, no. 1, pp. 48–61, 2018.

[133] D. Hernando, S. D. Sharma, M. Aliyari Ghasabeh, B. D. Alvis, S. S. Arora, G. Hamilton, L. Pan, J. M. Shaffer, K. Sofue, N. M. Szeverenyi, E. B. Welch, Q. Yuan, M. R. Bashir, I. R. Kamel, M. J. Rice, C. B. Sirlin, T. Yokoo, and S. B. Reeder, "Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat–water phantom," *Magnetic Resonance in Medicine*, vol. 77, no. 4, pp. 1516–1524, 2017.

[134] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[135] R. Cattell, S. Chen, and C. Huang, "Robustness of radiomic features in magnetic resonance imaging: review and a phantom study," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, 2019.

List of Publications

Refereed publications

- L. Bianchini, F. Botta, D. Origgi, S. Rizzo, M. Mariani, P. Summers, P. García-polo, M. Cremonesi, and A. Lascialfari, "PETER PHAN : An MRI phantom for the optimisation of radiomic studies of the female pelvis," *Physica Medica*, vol. 71, no. C, pp. 71-81, 2020.
- L. Bianchini, J. Santinha, N. Loução, M. Figueiredo, F. Botta, D. Origgi, M. Cremonesi, E. Cassano, N. Papanikolaou, and A. Lascialfari, "A multicenter study on radiomic features from T_2-weighted images of a customized MR pelvic phantom setting the basis for robust radiomic models in clinics," *Magnetic Resonance in Medicine*, pp. 1-14, 2020.

Presentations at conferences and schools

- L. Bianchini, F. Botta, D. Origgi, S. Rizzo, M. Mariani, P. Summers, P. García-Polo, M. Cremonesi and A. Lascialfari, "Development of an MRI phantom for the optimisation of radiomic studies on the female pelvis" *106 Congresso Nazionale SIF*, 2020 (oral presentation).
- L. Bianchini, J. Santinha, N. Loução, M. Figueiredo, F. Botta, D. Origgi, M. Cremonesi, N. Papanikolaou, and A. Lascialfari, "Blindly trusting MRI Radiomics? A study on radiomic features repeatability and reproducibility with a dedicated phantom" *ISMRM and SMRT Virtual Conference and Exhibition*, 2020 (poster).
- L. Bianchini, J. Santinha, N. Loução, F. Botta, M. Cremonesi, D. Origgi, N. Papanikolaou, and A. Lascialfari, "Investigation of radiomic features repeatability and reproducibility in Magnetic Resonance Imaging (MRI) with a dedicated pelvic phantom" *European Congress of Radiology*, 2020 (oral presentation).
- L. Bianchini, F. Botta, M. Cremonesi, D. Origgi, and A. Lascialfari, "Optimization of the radiomic analysis in MRI" *Biomedical Physics Workshop*, Pavia, 2020 (oral presentation).

- L. Bianchini, J. Santinha, N. Loução, F. Botta, M. Cremonesi, D. Origgi, P. Summers, N. Papanikolaou, and A. Lascialfari, "Investigation of radiomic features repeatability and reproducibility in Magnetic Resonance Imaging (MRI) with a dedicated pelvic phantom" *International Day of Medical Physics 2019*, Matera, 2019 (poster).

- L. Bianchini, F. Botta, D. Origgi , M. Cremonesi, P. Arosio and A. Lascialfari, "The impact of MR images acquisition process on radiomic features: phantom studies to support clinical research" *ISMRM 27^{th} annual meeting and exhibition*, Montreal, Canada, 2019 (poster).

- L. Bianchini, F. Botta, D. Origgi, , M. Cremonesi, M. Mariani, P. Arosio and A. Lascialfari, "PETER PHAN: An MRI phantom for radiomic studies on gynaecological cancers" *X Congresso AIRMM*, Milano, 2019 (poster).

- L. Bianchini, F. Botta, D. Origgi , M. Cremonesi, P. Arosio and A. Lascialfari, "Phantom studies for the investigation of the repeatability and robustness of radiomic features extracted from magnetic resonance images" *MAGNET2019: VI Italian Conference On Magnetism*, Messina, 2019 (oral presentation).

- L. Bianchini, F. Botta, D. Origgi , M. Cremonesi, P. Arosio and A. Lascialfari, "Repeatability and robustness of radiomic features extracted from magnetic resonance images of pelvic district: a phantom study" *XLVII National Congress on Magnetic Resonance*, Torino, 2018 (oral presentation).

Acknowledgements

I would like to thank Prof. Daniel Topgaard and Prof. José Javier Serrano Olmedo for being the external Referees of my thesis and for their thoughtful feedback.

I would like to express my own deepest gratitude to my supervisor, Prof. Alessandro Lascialfari, for the opportunity he gave me to enter the programme and for his constance presence - both scientific and human - in this experience.

I am extremely grateful to Prof. Ivan Veronese, who became my supervisor for this last year of Ph.D. and guided me until the end of my doctoral journey.

I wish to convey my sincere and full appreciation to Dr. Daniela Origgi and Dr. Marta Cremonesi for welcoming me to the European Institute of Oncology and for having encouraged my research spirit through these years. I can see your teachings reflected in the pages of this thesis and I am proud of being part of your group.

I am grateful to Prof. Nickolas Papanikolaou for having offered me the possibility of experiencing a fellowship at the Champalimaud Centre for the Unknown in Lisbon. During my stay I learnt a lot, in a stimulating and dynamic environment, with a strong sense of community.

To Dr. Francesca Botta: I do not have enough words to communicate in the proper way how much I am thankful to you. You have been my guide since the first day and a constant inspiration for me.

I am grateful to Prof. Paolo Milani, Dr. Tommaso Santaniello and all the CIMaINa staff for their work on the breast phantom inserts.

A very special thanks to João Santinha for being a smart and collaborative colleague and a precious friend.

Thanks to Dr. Paul Summers for sharing his knowledge on Magnetic Resonance and for always being ready to answer my questions.

To Nuno Loução: thank you for your support and help with my project. I will never forget arriving at Champalimaud with a beautiful sunrise on the Tejo for all the early acquisitions on the scanner!

To Francisco, Rui, Mauro, the hardware platform and all the staff at the Champalimaud that helped me during the experiments: thank you so much for your work, ideas and suggestions.

I wish to express my appreciation to Prof. Manuel Mariani, Dr. Paolo Arosio and Dr. Francesco Orsini for their contribution to the work and for what they taught me.

A special thanks to Dr. Mahila Ferrari for her support during all the project and for having the greatest smile everyday. You are always filling me with positivity and hope.

Thanks to Dr. Cristina Garibaldi, who has always been ready to help when I needed and for being part of this adventurous experience.

I wish to thank the physicians I worked with during these years for their availability and support: Dr. Stefania Rizzo, Dr. Anna Rotili and Dr. Valeria Dominelli. Thanks to Dr. Filippo Pesapane for being so collaborative and positive in our work together.

Thanks to Dr. Guido Pedroli for being a reference point in these years and thanks to Alessia for being the first smile and good morning every day.

Thank you Simone for your help and wise advice on the statistical analyses.

To the other members of the radiomic team, including Lisa, Noemi, Alberto and Johannes: thank you for your advice and useful discussion.

I would like to thank all the radiology technicians who helped me with patience and care to learn how to use the scanner.

A special thanks to the volunteers who decided to help my research by undertaking an MRI scan: Mahila, Eleonora, Chiara, Marta, Francesca, Sara, Marialetizia, Lia, Silvia, Aurora, Diana, Laura, Marzia, Sara, Delia, Paola, Angela, Giulia, Patrizia, Giulia, Teresa, Martina, Chiara, Vincenzo, Paul, Paolo, Stefano and Alfio.

Un ringraziamento davvero speciale è sicuramente per Alfio: non solo hai reso possibile questo lavoro, ma l'hai fatto con entusiasmo e reale interesse. Se mai dovessi aprire una fabbrica di fantocci, sicuramente saresti mio socio. Grazie anche a Stefano per avermi accolta in radiofarmacia.

Grazie Gennaro per i consigli e per avermi regalato i barattoli del frullatore, parte essenziale dell'esperimento!

Un ringraziamento speciale al comitato pavese della Croce Rossa Italiana, che mi ha accolta proprio negli anni del dottorato, ha nutrito la mia anima e ha dato valore al mio essere umana, attribuendo un significato più profondo alle mie azioni: *"Mais les femmes de Castiglione, voyant que je ne fais aucune distinction de nationalité, suivent mon exemple en témoignant la même bienveillance à tous ces hommes d'origines si diverses, et qui leur sont tous également étrangers. <<Tutti fratelli>>, répétaient-elles avec émotion."*(Henry Dunant, Un souvenir de Solférino, 1862).

Alla mia maestra preferita
Irene e alle mie meravigliose ballerine Michela,
Ilaria, Raffa, Cri, Elisa e Chiara:
i saggi, le chiacchiere in spogliatoio. grazie per tutte le nostre lezioni,
Mi avete regalato l'energia e la carica
necessaria ad affrontare questa avventura. A Sílvia, Cláudia, Mónica, Ana Sofia,
Sara, Juliana, Eunice, Maria Inês que fizeram a minha aventura em Portugal única e
memorável, a minha maior gratidão. A Mark: sei sempre stato al mio fianco in tutti
questi anni, hai seguito (anche se da lontano) tutti i miei spostamenti, le mie avventure, i miei traguardi. Grazie per esserci sempre stato e perché so che ci sarai sempre.
A Chiara, amica di una vita: ieri abbiamo condiviso i momenti più importanti della
nostra vita, oggi sei qui con me in questo viaggio che volge al termine e domani,
qualsiasi siano le nostre strade, saremo ancora insieme, a sostenerci sempre. Grazie
amica mia! Grazie Anna e Cri per avermi fatto sentire la vostra presenza, la vostra amicizia e il vostro affetto anche a centinaia (e migliaia) di chilometri di distanza. Grazie anche al Classe e al dott. Fabio per esserci sempre stati. Il 2020
ormai è passato, confido nel 2022?! Ai miei genitori un ringraziamento
gigante, per aver sempre creduto in me e supportato ogni mia
scelta, per avermi riempito il freezer di ragù e polpette e per
avermi tenuto compagnia nei lunghi viaggi casa-lavoro
quotidiani. A te, Mario, il mio grazie più dolce, per
avermi accompagnato lungo questo percorso
e per aver sempre tenuto la mia mano,
donandomi tutto ciò di cui
avevo bisogno e anche di più.
♡