ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF PHYSICS AND ASTRONOMY ”A. RIGHI”

SECOND CYCLE DEGREE

PHYSICS

QUANTITATIVE VALIDATION AND APPLICATION OF CLINICAL BIOMARKERS IN 3T WB-MRI

Supervisor:
Prof. Mauro Villa

Co-supervisors:
Dr. Roberto Spighi
Dr. Anna Sarnelli
Dr. Giacomo Feliciani

Defended by:
Alessia Pushi

Graduation Session 25-26 September 2025

Academic Year 2024/2025

To all the times I believed I could not make it

Abstract

This work investigates the quantitative validation and clinical application of two MRI-derived biomarkers in 3T Whole-Body MRI (WB-MRI): the Apparent Diffusion Coefficient (ADC), reflecting water mobility and microstructural density, and the Fat Fraction (FF), representing tissue fat content. The objectives were to assess their robustness, define reference values in healthy tissues and evaluate their diagnostic potential in Multiple Myeloma (MM). Technical validation was performed using a water phantom for ADC accuracy and a custom-built fat–water phantom for FF quantification. In vivo characterisation in 20 healthy females across seven anatomical sites — parenchymal organs (liver, spleen, pancreas, kidneys) and skeletal regions (vertebrae, pelvic bone, femoral head) — examined repeatability, variability and radiomic feature stability. ADC showed greater consistency in parenchymal organs, whereas FF proved more stable in adipose-rich skeletal tissues. The clinical study analysed 107 MM patients (38 Smoldering, the early asymptomatic stage; 46 active; 23 Relapsed/Refractory) against the 20 healthy controls. In the pelvic bone the replacement of normal marrow by malignant plasma cells was associated with increased ADC and decreased FF. FF consistently showed higher discriminative power in distinguishing healthy from infiltrated marrow even at early stages, while ADC reached high significance mainly in advanced disease. Radiomic descriptors further enhanced classification performance, with FF-based models achieving Area Under the Curve (AUC) values above 0.95 and ADC providing complementary information in multiparametric models. The results demonstrate that ADC and FF can be robustly quantified in WB-MRI. FF emerges as the more stable and diagnostically sensitive biomarker for bone marrow infiltration in MM. The proposed framework supports the integration of quantitative biomarkers into non-invasive tools for diagnosis, staging and longitudinal monitoring of MM patients.

Sommario

Questo lavoro tratta la validazione quantitativa e l'applicazione clinica di due biomarcatori derivati dalla risonanza magnetica Whole-Body MRI (WB-MRI) a 3T: il coefficiente di diffusione apparente (ADC), che riflette la mobilità delle molecole d'acqua e quindi la densità microstrutturale, e la percentuale di grasso (FF), che rappresenta il contenuto lipidico dei tessuti. Scopo del lavoro è stato quello di valutare la robustezza quantitativa di tali biomarcatori, definire i loro valori di riferimento nei tessuti sani e analizzarne il potenziale diagnostico nel mieloma multiplo (MM). La validazione tecnica è stata condotta utilizzando un fantoccio di acqua per verificare l'accuratezza delle misure dell'ADC e un fantoccio a emulsione grasso–acqua appositamente realizzato per valutare la quantificazione della FF. La caratterizzazione in vivo, svolta su 20 soggetti sani di sesso femminile e comprendente sette distretti anatomici - organi parenchimali (fegato, milza, pancreas, reni) e regioni scheletriche (vertebre, osso pelvico e testa femorale) - ha esaminato ripetibilità, variabilità e stabilità delle caratteristiche radiomiche. L'ADC ha mostrato maggiore affidabilità negli organi parenchimali, mentre la FF si è dimostrata più stabile nei tessuti scheletrici ricchi di adiposità. Lo studio clinico ha coinvolto 107 pazienti con MM (38 Smoldering MM, fase precoce e asintomatica; 46 nella fase attiva MM; 23 Recidivi/Refrattari MM), confrontati con i 20 soggetti sani. A livello dell'osso pelvico la sostituzione del midollo sano da parte di plasmacellule neoplastiche è risultata associata a un incremento dell'ADC e a una riduzione della FF. Quest'ultima ha dimostrato una superiore capacità discriminativa nel distinguere in modo significativo il midollo sano da quello infiltrato già negli stadi precoci, mentre l'ADC ha raggiunto alta significatività soprattutto nelle fasi avanzate. I descrittori radiomici (caratteristiche radiomiche) hanno ulteriormente potenziato le performance di classificazione, con modelli basati sulle feature delle mappe FF che hanno raggiunto valori di Area Under the Curve (AUC, area sotto alla curva) superiori a 0.95, mentre le caratteristiche derivate dalle mappe ADC hanno fornito un contributo complementare nei modelli multiparametrici. I risultati dimostrano che ADC e FF possono essere quantificati in modo robusto nella WB-MRI. La FF emerge come biomarcatore più stabile e sensibile per l'infiltrazione midollare nel MM. Il framework proposto rappresenta un quadro metodologico riproducibile per l'integrazione dei biomarcatori di imaging quantitativo in strumenti non invasivi per la diagnosi, la stadiazione e il monitoraggio longitudinale dei pazienti affetti da MM.

Contents

List of Figures

List of Tables

List of Acronyms

ADC	Apparent Diffusion Coefficient
AUC	Area Under the Curve
DWI	Diffusion Weighted Imaging
FF	Fat Fraction
IP	In-Phase
MM	Multiple Myeloma
MRI	Magnetic Resonance Imaging
MY-RADS	MYeloma Response Assessment and Diagnosis System
NMR	Nuclear Magnetic Resonance
OP	Opposed-Phase
QIB	Quantitative Imaging Biomarker
QIBA	Quantitative Imaging Biomarkers Alliance
qMRI	Quantitative MRI
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RRMM	Relapsed/Refractory Multiple Myeloma
SIBEX	Standardised Imaging Biomarker EXtraction
SMM	Smoldering Multiple Myeloma
WB-MRI	Whole-Body Magnetic Resonance Imaging

Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique widely used in clinical diagnostics due to its ability to provide high-resolution images of soft tissues without the use of ionising radiation. Unlike nuclear imaging techniques such as Positron Emission Tomography (PET) or radiological modalities like Computed Tomography (CT) — which rely on the administration of radioactive tracers or ionising radiation — MRI is based on the interaction between magnetic fields and hydrogen nuclei naturally present in biological tissues. This allows for safe and repeatable imaging across multiple anatomical regions. Moreover, MRI combines anatomical and functional information within a single acquisition, enabling both morphological assessment and tissue characterisation.

Recent developments in acquisition techniques and signal modelling have progressively extended the capabilities of MRI from purely morphological imaging to the quantitative characterisation of tissue properties. This evolution has led to quantitative MRI (qMRI), in which specific acquisition sequences and reconstruction strategies are used to derive physical and physiological parameters directly from the MR signal. When combined with standardised protocols and appropriate post-processing, qMRI enables the derivation of Quantitative Imaging Biomarkers (QIBs) — scalar values that reflect underlying microstructural or compositional properties. These biomarkers have potential diagnostic, prognostic and predictive value. As reproducible and non-invasive metrics, QIBs enable longitudinal assessment of tissue alterations, supporting both clinical interpretation and computational modelling. They also enhance diagnostic precision by providing objective measurements that complement visual analysis and can be used to inform personalised treatment planning through quantitative insight into disease burden and tissue composition.

Two quantitative biomarkers of particular interest in current clinical research are the Apparent Diffusion Coefficient (ADC) and the Fat Fraction (FF), derived respectively from diffusion-weighted imaging and fat quantification sequences.
ADC quantifies the mobility of water molecules within biological tissues and is sensitive to microstructural changes such as increased cellularity, fibrosis or necrosis.
FF, on the other hand, is an emerging clinical biomarker that measures the proportion of fat within a tissue and provides insight into its lipid content. It is particularly useful in the analysis of organs where fat content varies with metabolic, inflammatory or infiltrative conditions.
Although ADC and FF are derived from different physical principles — water diffusion and chemical shift, respectively — they provide complementary information about tissue composition and structure. Both biomarkers hold clinical relevance, as they are

expected to follow physiologically consistent patterns across different biological tissues and deviations from these patterns may serve as indicators of pathological change.
These concepts, along with the basic principles of MRI and qMRI, are further detailed in Chapter 1.

This study focuses on the evaluation of both ADC and FF within the context of 3T Whole-Body MRI (WB-MRI), a non-invasive imaging modality that enables comprehensive coverage of the entire body in a single examination. WB-MRI combines high anatomical resolution with advanced functional imaging, allowing for the extraction of quantitative biomarkers across multiple organs and tissue types. This makes it particularly suitable for the assessment of systemic diseases, in which it is essential to evaluate the extent of involvement beyond a single anatomical region.
All MRI acquisitions and data analyses presented in this work were conducted at the Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori" in Meldola (FC, Italy), as part of ongoing clinical and research protocols carried out by the Medical Physics Unit, with a specific focus on the development and validation of quantitative imaging techniques for oncological applications.

To ensure the robustness of ADC and FF biomarkers before their clinical application, this study begins by verifying that both parameters satisfy the technical requirements expected for any quantitative imaging biomarker: accuracy, repeatability, reproducibility and — ideally — linearity across the physiologically relevant range. These properties were evaluated through phantom studies, which provide a stable ground truth against which the performance of a sequence can be evaluated.
The ADC protocol was tested using a water phantom, confirming whether the measured diffusion values were consistent with known physical references at the corresponding acquisition temperature, as detailed in Chapter 2.
For FF, validation was carried out using a custom-built fat–water emulsion phantom specifically designed to simulate a wide range of fat concentrations. This enabled the assessment of signal stability, linearity and quantitative reliability across the full spectrum of expected values. As described in Chapter 3, this experimental setup also served to compare different fat quantification sequences available on the scanner, providing a practical framework for selecting the most accurate and robust acquisition protocol for clinical and research applications.

Following phantom validation, the in vivo behaviour of each biomarker was assessed in physiologically normal tissues. Establishing reference values in healthy subjects is a fundamental step for clinical interpretation, as it provides a baseline against which pathological deviations can be measured. In addition to absolute values, the evaluation of intra-subject repeatability and inter-subject variability is essential to understand the

consistency and robustness of the biomarkers across different anatomical regions and individuals.
Chapters 2 and 4 are dedicated to the systematic characterisation of ADC and FF, respectively, using data from a cohort of 20 healthy adult female subjects undergoing multiple WB-MRI acquisitions. The analysis was conducted across seven anatomical regions, including both parenchymal organs (liver, spleen, pancreas, kidneys) and bone tissues (vertebrae, pelvic bone and femoral head). For each biomarker, the statistical analysis includes assessments of repeatability across the study cohort, longitudinal stability in repeated scans from a single subject and intra-organ variability. Furthermore, the structural complexity and feature stability of each biomarker beyond its average value was explored by performing radiomic analysis, which consists in the extraction of mathematically defined features — numerical descriptors of image properties such as intensity, texture and spatial organisation — that quantitatively characterise tissue morphology and heterogeneity.

Once validated and characterised in healthy tissue, the clinical applicability of these biomarkers was evaluated. In this work, the selected clinical context is Multiple Myeloma (MM), a plasma cell malignancy characterised by bone marrow infiltration and progressive alteration of marrow composition. Infiltrated regions are expected to show increased ADC values — due to greater water mobility — and reduced FF values — reflecting the loss of adipose tissue. These theoretical expectations were examined through quantitative analysis of a cohort including 20 healthy controls and 107 patients at different stages of myeloma: 38 smoldering (SMM), 46 active (MM) and 23 relapsed/refractory (RRMM). Both mean signal values and radiomic features were extracted from the pelvic bone ADC and FF maps and used in predictive models to evaluate their discriminative power. These analyses are presented in detail in Chapter 5, which focuses on the clinical evaluation of ADC and FF in multiple myeloma, exploring their individual and combined potential for the non-invasive assessment of bone marrow involvement in the pelvic bone of affected patients.

The thesis reflects the logical progression of the study and is structured as follows:

- Chapter 1 provides the theoretical background of MRI, with a focus on whole-body and quantitative imaging techniques. It explains the physical principles and clinical relevance of ADC and FF and describes the acquisition protocols adopted throughout the study.

- Chapter 2 presents the quantitative analysis of ADC in healthy tissues. It includes phantom validation using a water-based reference and describes the dataset, the anatomical ROI placement strategy and the feature extraction protocol. The analysis evaluates intra-subject repeatability, longitudinal stability and intra-organ variability across seven anatomical regions, as well as radiomic descriptors to assess feature stability beyond mean intensity.

- Chapter 3 focuses on the validation of fat quantification sequences. A custom-built fat–water emulsion phantom is used to compare different acquisition methods and assess the linearity and accuracy of FF estimation. The analysis highlights the impact of different fat modelling strategies on signal behaviour and quantification accuracy and informed the selection and interpretation of FF maps in the subsequent in vivo analyses.

- Chapter 4 reports the characterisation of FF in healthy tissues, following the same dataset and methodological framework adopted for ADC. It includes the evaluation of repeatability, variability and longitudinal consistency across parenchymal and skeletal regions and a radiomic analysis to investigate structural and compositional stability in the evaluated anatomical regions.

- Chapter 5 explores the clinical application of ADC and FF in multiple myeloma. Building on the characterisation of both biomarkers in healthy tissues, this chapter focuses on their use for the non-invasive assessment of bone marrow involvement in the pelvic bone. It describes the behaviour of ADC and FF across disease stages — from smouldering to multiple myeloma and relapsed/refractory cases — and evaluates their diagnostic performance through groupwise comparisons and predictive models based on intensity and radiomic features.

1 Principles and Techniques of Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique used for clinical purposes to provide information about the human body without using ionising radiation. It is based on the physical phenomenon of Nuclear Magnetic Resonance, which involves the interaction between atomic nuclei and external magnetic fields.

This chapter presents the fundamental principles of MRI and their applications in quantitative assessment of tissue properties through Apparent Diffusion Coefficient (ADC) and Fat Fraction (FF) imaging biomarkers. Furthermore, it introduces the concept of Whole-Body MRI (WB-MRI) and gives an overview of the acquisition systems and sequences employed in this work as a foundation for the subsequent experimental and clinical analyses.

1.1 Fundamentals of Magnetic Resonance Imaging

MRI is a high-resolution, non-invasive imaging technique that allows the acquisition of detailed three-dimensional anatomical images and is currently used in clinical practice for disease detection, diagnosis and treatment monitoring. It is able to provide excellent soft tissue contrast in anatomical images and also to exploit different physical properties of tissues — such as proton density, relaxation times, functional activity, diffusion and fat content — to generate various types of contrast and quantitative maps. Since the working principles involve radiofrequency pulses in the low-energy range of the electromagnetic spectrum, MRI is considered safer than other radiographic modalities such as Computed Tomography (CT) and Positron Emission Tomography (PET) that use ionising radiation, respectively X-rays or γ-rays, to produce diagnostic images of the body.

The physical mechanism of MRI is Nuclear Magnetic Resonance (NMR), a phenomenon discovered and explained simultaneously by F. Bloch[1] at Stanford University and E. M. Purcell[2] at Harvard University in 1946. Their theoretical and experimental observations lead to their Nobel Prize in 1952 and are essentials to understand how atomic nuclei respond to external magnetic fields and radiofrequency excitation[3].

In order to interact with an external static magnetic field $\vec{B}_0$ nuclei must have a nuclear magnetic moment $\vec{\mu}$ which arises from a non-zero nuclear spin $\vec{I}$ in nuclei with an odd number of protons or neutrons. The magnetic moment is proportional to the nuclear angular momentum according to the relation

$$\vec{\mu} = \gamma \left(\frac{h}{2\pi} \right) \vec{I} = \gamma \hbar \vec{I} \tag{1.1}$$

where γ (or γ_n) is the gyromagnetic ratio, a constant characteristic of each nucleus. The value of $\frac{\gamma}{2\pi}$ is particularly high for the ^{1}H atom and this is one of the main reasons why hydrogen nuclei are the most commonly used in MRI: their high gyromagnetic ratio and large abundance in biological tissues lead to a stronger detectable signal.

Atomic nuclei are characterised by states of quantum mechanical nature. The projection of the nuclear magnetic moment along the direction of the external magnetic field $\vec{B}_0$ is denoted as the z-component and is given by

$$\mu_z = \gamma \hbar m \tag{1.2}$$

where m is the magnetic quantum number that for spin-$\frac{1}{2}$ nuclei can take values equal to $\frac{1}{2}$ for spin-up state, aligned parallel to the external field, or $-\frac{1}{2}$ for the spin-down state, in the antiparallel configuration. Therefore, for the ^{1}H atom that has nuclear spin $I = \frac{1}{2}$, only two possible values exist for μ_z.
When there is no external magnetic field, the two energy levels of spin-$\frac{1}{2}$ nuclei are degenerate and equally populated at thermal equilibrium. When an external magnetic field $\vec{B}_0$ is applied, Zeeman interaction between the spins and the magnetic field takes place and cause the splitting of the degenerate energy level into two distinct states with different energies, corresponding to spin orientations either aligned or anti-aligned with the external field.

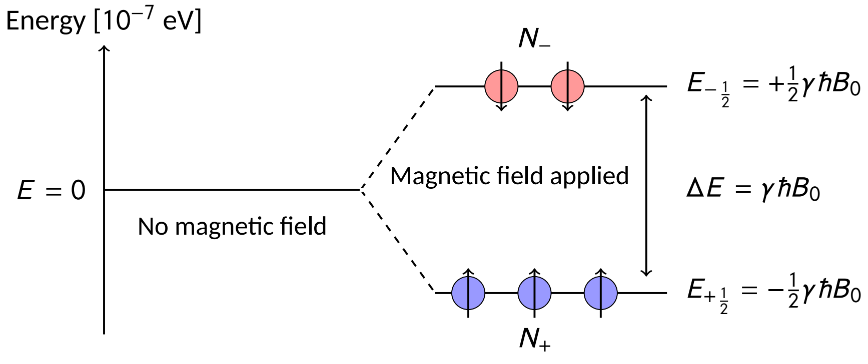


Figure 1.1: Schematic representation of the Zeeman splitting in a population of spin-$\frac{1}{2}$ nuclei.

Considering $\vec{B}_0$ uniform along the z-axis ($\vec{B}_0 = B_0\hat{z}$), the Zeeman Hamiltonian is

$$\hat{H}_Z = -\gamma \hbar I_z B_0 \tag{1.3}$$

and, as illustrated in Figure 1.1, the energy values are defined by the eigenvalues

$$E_m = -\gamma \hbar m B_0 \tag{1.4}$$

so that the splitting of the energy levels generated by the magnetic field can be quantified by the energy gap between them

$$\Delta E = \gamma \hbar B_0 \,. \tag{1.5}$$

Transitions between energy levels can occur when the system absorbs energy through electromagnetic radiation with a frequency equal to

$$\omega_0 = \gamma B_0 \tag{1.6}$$

which is known as Larmor frequency and determines the resonance condition required to excite nuclear spins.

At thermal equilibrium, nuclei distribute among the $2I + 1$ possible energy levels according to Boltzmann distribution

$$\frac{N_+}{N_-} = \exp\left(\frac{\Delta E}{kT}\right) = \exp\left(\frac{\hbar\gamma B_0}{kT}\right) \tag{1.7}$$

where N_+ and N_- are the number of nuclei per unit volume in the lowest and highest energy configuration (respectively the ones with up and down spins), k is the Boltzmann constant and T is the absolute temperature measured in kelvin.

Since the lowest energy state will be the most populated, in the total amount of nuclei per unit volume ($N = N_+ + N_-$) there will be an excess of up spins over the down spins, as schematically represented in Figure 1.1, and this quantity will give rise to a net nuclear magnetization, known as nuclear magnetization at equilibrium

$$M_0 = \mu(N_+ - N_-) = N\mu \tanh \frac{\gamma \hbar B_0}{2kT} \,. \tag{1.8}$$

More generally, the magnetization at equilibrium per unit volume is a vector $\vec{M}_0$ that has the same direction of $\vec{B}_0$ and a magnitude given by Curie Law

$$M_0 = N \frac{\gamma^2 \hbar^2 I(I+1)}{3kT} B_0 \,. \tag{1.9}$$

In NMR, an additional oscillating magnetic field $\vec{B}_1$ is generated by a radiofrequency (r.f.) pulse delivered through a transmit coil, tuned to the resonance frequency and applied perpendicularly to $\vec{B}_0$ (i.e. in the xy-plane). This excitation tilts the bulk magnetization vector $\vec{M}$ from its equilibrium alignment along $\vec{B}_0$ by an angle which is called nutation angle

$$\alpha = \gamma B_1 \tau \tag{1.10}$$

with τ equal to the duration of the r.f. pulse application. This rotation causes the magnetization to begin precessing around $\vec{B_0}$ at the Larmor angular frequency ω_0.
The time evolution of $\vec{M}$ in the Cartesian coordinate system is described by the Bloch equation:

$$\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} = \gamma \vec{M(t)} \times \vec{B(t)} \tag{1.11}$$

where $\vec{B(t)} = \vec{B_0} + \vec{B_1}$ is the total magnetic field applied to the system.
The time-varying magnetization induces an oscillating voltage and current in the receiver coil, according to Faraday's law of electromagnetic induction. The resulting signal is known as Free Induction Decay (FID) and constitutes the observable quantity used in NMR experiments to reconstruct information about the motion of the total amount of spins in the real system and describe the tissue properties and characteristics.

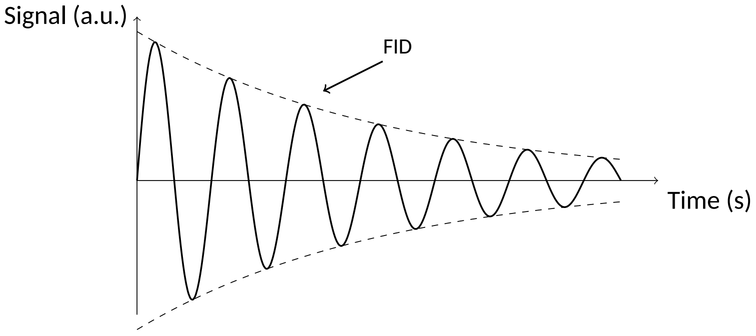


Figure 1.2: Free Induction Decay (FID) signal following radiofrequency excitation in NMR.

As shown in Figure 1.2, the FID is an exponentially damped oscillatory signal that decays towards zero. In particular, it is obtained by following the temporal evolution of the transverse magnetization M_{xy} that, after the NMR excitation phase, returns back to its equilibrium state, where the transverse component is zero. Once equilibrium is reached on the transverse plane, also the longitudinal component M_z is restored to its initial value M_0. Starting from the acquired signal, it is therefore possible to reconstruct the two different relaxation processes through which $\vec{M(t)}$ evolves back to equilibrium.

The transverse relaxation is an entropic process characterised by the loss of phase coherence within the spin system. Each spin experiences a local magnetic field resulting from the combination of the externally applied field and the magnetic influence of surrounding spins and, as a result, magnetic moments spread over the *xy*-plane and individual spins precess at slightly different rates, gradually losing phase coherence.
Transverse relaxation is therefore mainly due to spin-spin interactions and it is described by an exponential decay of the measurable transverse component M_{xy} as modelled by

the following relation

$$M_{xy}(t) = M_{xy}(0)\, e^{-\frac{t}{T_2}} \tag{1.12}$$

in which the time constant T_2 is the transverse relaxation time. In order to take into account also the additional dephasing induced by magnetic field inhomogeneities, common in experiments, T_2 will be substituted by an effective transverse relaxation time calculated as

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{\text{inhomo}}} \; . \tag{1.13}$$

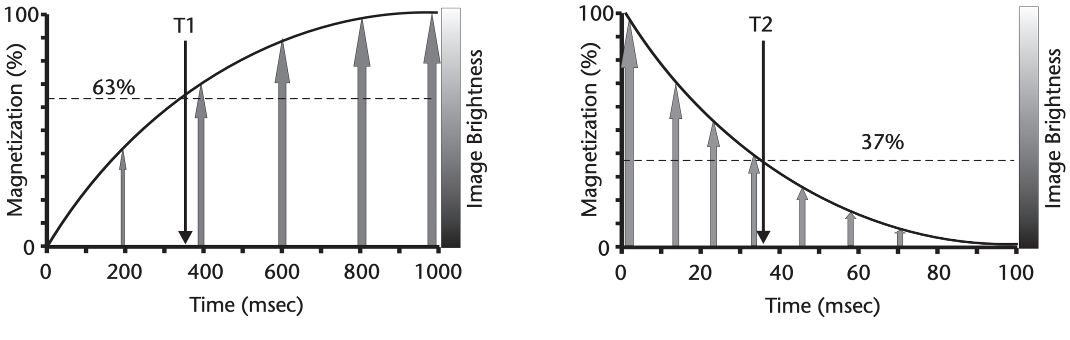
In practice, the observed decay of transverse magnetization towards equilibrium is governed by this shorter time constant T_2^* that accounts for both intrinsic and extrinsic dephasing mechanisms.

Longitudinal relaxation is instead an energetic process related to spin-lattice interactions that allow a transfer of energy from the nuclear spins to the surrounding lattice, allowing the spins to return to their lower energy state aligned with the static magnetic field.

The recovery of the longitudinal component of the magnetization $M_z(t)$ of ^{1}H nuclei occurs gradually over time and is described by an exponential relation

$$M_z(t) = M_0 \left(1 - e^{-t/T_1}\right) \; . \tag{1.14}$$

In Equation 1.14 M_0 represents the equilibrium magnetization and T_1 is the longitudinal relaxation time, which characterises the rate at which the system returns to thermal equilibrium along the z-axis. The value of T_1 depends on the difference in transition probabilities between the lower and higher energy states: even in the absence of the r.f. pulse, transitions between the two levels are possible due to fluctuations around the Larmour frequency but transitions to the lower energy state are more probable.



[A] [B]

Figure 1.3: Diagrams of the longitudinal [A] and transverse [B] relaxation processes[4].

Figure 1.3 illustrates the exponential recovery of the longitudinal magnetization M_z with the time constant T_1 and the exponential decay of transverse magnetization M_{xy} with time constant T_2. It is important to note that, since transverse relaxation leads the system towards internal equilibrium without any energy exchange, T_2 is always shorter than T_1 and in biological tissues $T_2 \ll T_1$.

T_1 and T_2 relaxation times are different for each specific tissue analysed with NMR because variations in molecular composition, water content, cellular architecture and microenvironmental interactions affect the efficiency of energy exchange and dephasing processes. This has a direct impact on contrast and visibility of different tissue types in MRI and allows to obtain images which are T_1-weighted or T_2-weighted, where images voxel brightness predominantly reflects longitudinal and transverse relaxation time.
An example of images acquired by T_1-weighting or T_2-weighting is shown in Figure 1.4, where the spine is imaged using both contrasts. In the T_1-weighted image fluids appear dark, fat is bright and muscle and bone tissue show an intermediate grey intensity. The T_2-weighted image is instead generally brighter and particularly useful to highlight fluids or to detect pathological conditions such as tumour, inflammation or trauma.

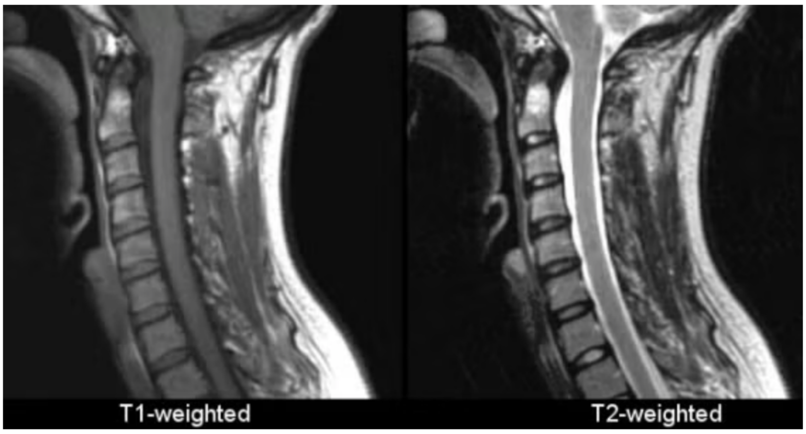


Figure 1.4: Comparison of T_1 and T_2-weighted images for the spine[5].

The relaxation processes described above determine the measurable image contrast, through which anatomical information can be represented. As in other diagnostic modalities, MRI images are displayed by slicing the reconstructed volume along three orthogonal reference planes: the axial plane, oriented perpendicular to the longitudinal body axis, the coronal plane, oriented parallel to the frontal surface of the body, and the sagittal plane, oriented parallel to the median division between left and right sides.
A schematic representation of the three anatomical planes is reported in Figure 1.5.

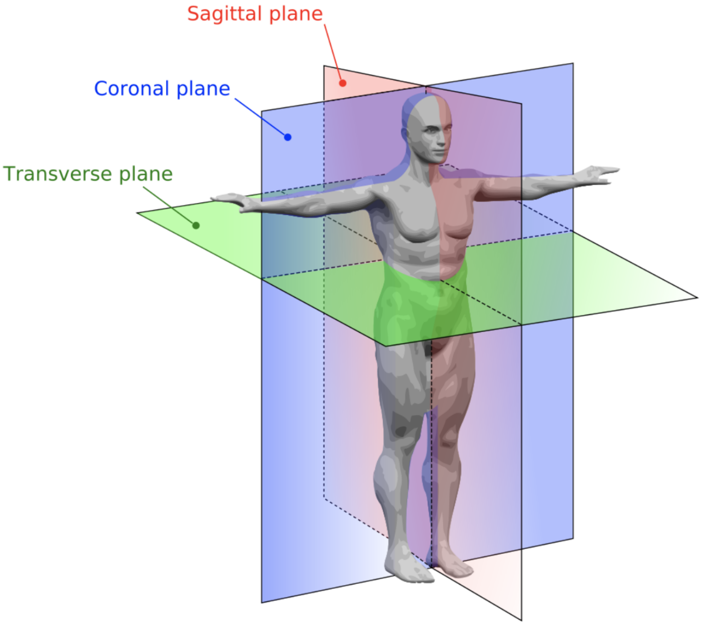


Figure 1.5: Schematic representation of the three anatomical planes conventionally used in MRI[6].

1.2 Quantitative MRI

In conventional MRI, contrast-weighted images exploit differences in tissue relaxation properties to produce high-quality anatomical images. While the resulting soft tissue contrast is excellent, it remains limited to qualitative analysis and is subject to human interpretation.

Quantitative MRI (qMRI) was developed to overcome these limitations by measuring intrinsic physical properties of tissues and expressing them in values associated to meaningful physical units. Its implementation has been enabled by technological progress, including improved scanner hardware, faster data acquisition and new advanced software algorithms. The rationale behind qMRI is therefore to provide objective, reproducible and standardised parameters in the form of quantitative maps that, unlike conventional MRI, offer voxel-wise measurements of specific physical or physiological properties rather than qualitative signal intensities.

qMRI refers to a range of techniques that enables the quantification of several tissue characteristics such as relaxation times, diffusion, perfusion, iron fraction, elastic properties, temperature, chemical composition and exchange. All of these measurable parameters can serve as quantifiable indicators of biological processes, pathological changes or treatment response and they are defined as Quantitative Imaging Biomark-

ers or QIBs[7].
To explain the physical principles behind qMRI, this section will concentrate on the two QIBs relevant to the experimental validation of this work: the Apparent Diffusion Coefficient (ADC) and the Fat Fraction (FF).

1.2.1 Diffusion and Apparent Diffusion Coefficient

Diffusion is the movement of particles down their concentration gradients, from regions of higher concentration to regions of lower concentration. In a free medium, the flux of particles due to diffusion ($\vec{J}$) is described by Fick's First Law

$$\vec{J} = -D\vec{\nabla}c \tag{1.15}$$

where c is the concentration and D the diffusion coefficient.
This movement is a consequence of the microscopic Brownian motion of particles: due to thermal energy, molecules follow a random translational motion and for the ideal case of free diffusion, there will be a net mean square displacement equal to

$$\langle x^2(t) \rangle = 2Dt \tag{1.16}$$

in each spatial direction in a given time t, corresponding to a Gaussian probability distribution of displacements. The entity of this phenomenon is related to the molecular dimension and shape since the diffusion coefficient, according to Stokes-Einstein equation[8], is inversely proportional to the radius of the spherical particle and the viscosity of the fluid, besides being directly proportional to the temperature.

Diffusion Weighted Imaging (DWI) is an MRI technique that makes the signal sensitive to diffusion, allowing the detection of tissue microstructural changes based on the mobility of water molecules. In MRI, diffusion is encoded by including additional magnetic field gradients into the pulse sequence, translating the entity of water molecules displacement into an attenuation of the measurable signal.

The most common method for image acquisition is the Spin Echo Diffusion Weighted Imaging (SE-DWI), schematically represented in Figure 1.6.
As in a standard SE sequence, there is a first 90° r.f. pulse that tips the longitudinal magnetization into the transverse plane, starting the spins dephasing, and then a 180° refocusing pulse that reverses the dephasing and leads to the formation of an echo at time TE. Immediately before and after the 180° pulse there are two diffusion sensitising gradients, applied on the three axes for a time duration δ, both with the same amplitude (gradient strength G) but with inverse polarities and separated by a time gap equal to Δ. In this way, the first gradient after excitation generates a position-dependent phase shift among the spins and during the Δ period spins that diffuse acquire an additional

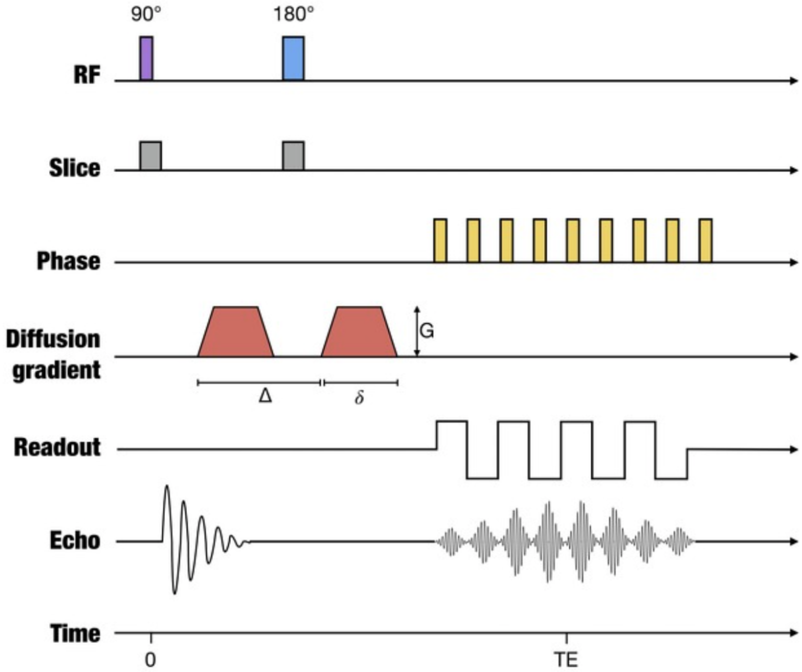


Figure 1.6: Schematic diagram of the Spin Echo sequence for Diffusion Weighted Imaging (SE-DWI)[9].

phase shift. The second gradient exactly realigns only the spins that remained stationary in their original position: only the spins that did not diffuse will be rephased and will contribute to the echo signal, leading to a diffusion-dependent signal loss[10].

The signal attenuation depends on water molecule displacement and can be modelled using a mono-exponential relation

$$S = S_0\, e^{-b\cdot D} \tag{1.17}$$

where S is the signal measured with diffusion weighting, S_0 is the signal acquired without diffusion gradients and D is the diffusion coefficient, usually expressed in mm^2/s. The signal sensitivity to diffusion depends on the physical characteristics of the diffusion gradients and is quantified by the *b-value*

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right) \tag{1.18}$$

that combines the parameters of the gradients (gradient amplitude G, time duration δ and time gap between the two gradients Δ) with the gyromagnetic ratio of the proton γ. An higher *b-value* increases the sensitivity to diffusion but also leads to greater signal attenuation and therefore to a lower signal-to-noise ratio.

From Equation 1.17 it is possible to understand that to measure the diffusion coefficient

$$D = -\frac{1}{b}\ln\left(\frac{S}{S_0}\right) \tag{1.19}$$

it is necessary to acquire at least two measurements: one for $b = 0$ to determine S_0 and one at $b > 0$ for D. In real acquisitions, the images acquired at different *b-values* are fitted to the monoexponential model of Equation 1.17 that in its logarithmic form returns the value of D as the negative slope of the fitted line and the value of S_0 as the intercept.

Deviations from the mono-exponential model can occur, particularly at low *b-values* due to molecular perfusion, as described by the IntraVoxel Incoherent Motion (IVIM Model[11]), or at high *b-values* where diffusion becomes non-Gaussian due to the tissue structure, requiring more sophisticated models such as Diffusion Kurtosis Imaging (DKI[12]). To avoid these deviations and more advanced models, the mono-exponential signal decay is still assumed over an intermediate range of *b-values* (from 0 to 800 s/mm^2), where perfusion and non-Gaussian effects are minimised[7].

In biological tissues, diffusion is not completely free but restricted by structural boundaries such as cell membranes, organelles and extracellular matrix components. Moreover, the intrinsic anisotropy of many tissues influences water mobility since physical properties can vary with the direction along which the analysis is run. The diffusion measured by MRI is therefore affected by restrictions and anisotropies and, for this reason, the coefficient D in the ideal signal relation (Equation 1.17) is replaced by the Apparent Diffusion Coefficient (ADC), which is operationally defined through the same expression — explicitly formulated in Equation 1.19 — but with D substituted by ADC to account for non-ideal effects. In practice, ADC values are consistently lower than the free diffusion coefficient D[13].

The ADC is the QIB that reflects the diffusion of water molecules within tissues, including the effects of both true molecular diffusion and microstructural constraints. In ADC maps, high values of ADC (i.e. bright regions) indicate high water mobility, typical in tissues with low cellularity or disrupted structural integrity such as in necrotic lesions, edema or bone marrow infiltrations. Low ADC values (i.e. dark regions) are instead typical in tissues with restricted diffusion due to dense cellular structures.

An example of an ADC Map MRI of the upper abdomen acquired in the axial plane is illustrated in Figure 1.7. The liver is shown on the left side of the image (patients' anatomical right), the stomach on the upper right, the spleen on the lower right and the vertebral column centrally located in the lower region.

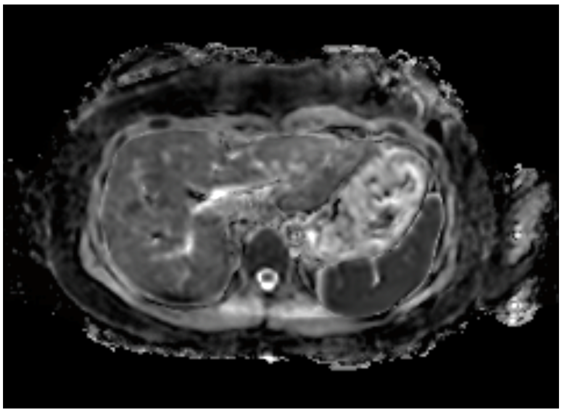


Figure 1.7: ADC Map of the upper abdomen acquired through DWI in the axial plane.

1.2.2 Fat Quantification and Fat Fraction

The quantification of fat in MRI starts from the physical phenomenon of chemical shift[14]: fat and water protons are surrounded by different molecular environments which modify the effective local magnetic field experienced by each nucleus. Due to differences in molecular size and electronic configuration, the shielding from the external magnetic field varies between the two, leading to slightly different Larmor frequencies. This frequency offset is referred to as chemical shift and is defined as the difference in resonance frequency between fat and water, normalised to the resonance frequency of water, which is used as a reference.
Fat is composed of multiple spectral components, but in a first approximation, a single peak centred on the methylene group is considered. The resonance frequency of fat is slightly lower than that of water and the chemical shift between the two is approximately 3.5 parts per million (ppm)[15].

The Dixon method, originally proposed by W. T. Dixon in 1984[16], enables the separation of water and fat signals in MRI by exploiting their chemical shift and the resulting phase evolution over time. Precessing at different rates, water and fat protons gradually accumulate a phase difference and at specific echo times their signals may be either in-phase (IP) or opposed-phase (OP). In the first case, the magnetization vectors of fat and water are aligned and add constructively in the total signal while in the second one they are oriented in opposite directions and there is signal cancellation.
Under the assumption that water and fat are the only chemical species contributing to the signal, the resulting complex image obtained after applying the Fourier transform to the acquired data can be written as

$$S(x,y) = [W(x,y) + F(x,y) \cdot e^{i\alpha'}] \cdot e^{i\phi(x,y)} \cdot e^{i\phi_0(x,y)} \tag{1.20}$$

where (x, y) are the pixel coordinates, W and F are the magnitudes of the magnetizations, α' is the phase angle of fat relative to water due to the chemical shift, ϕ is the error phase due to magnetic field inhomogeneity and ϕ_0 is the error phase due to the acquisition system[15].
The vector representation of such complex signal is illustrated in Figure 1.8.

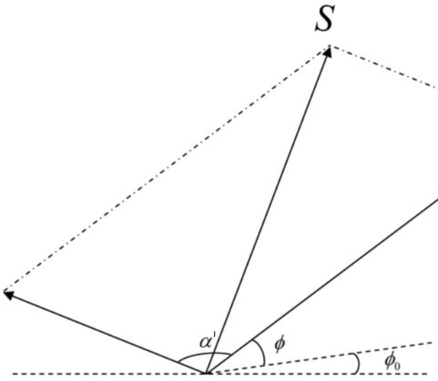


Figure 1.8: Graphical representation of the vector quantities involved in the acquisition of the complex signal through the water (W) and fat (F) spectral components[15].

The original two-point Dixon technique involves the acquisition and elaboration of two images, one in the in-phase condition and one in the opposite-phase one. In their complex form, they are respectively indicated as

$$S_0 = (W + F) \cdot e^{i\phi_0} \tag{1.21}$$

$$S_1 = (W - F) \cdot e^{i\phi} \cdot e^{i\phi_0} \,. \tag{1.22}$$

When the error phase due to the magnetic field inhomogeneity $\phi = 0$ it is possible to obtain the water-only and fat-only images as an algebraic combination of Equation 1.21 and Equation 1.22 as follows

$$W = \frac{1}{2} \cdot |S_0 + S_1| \tag{1.23}$$

$$F = \frac{1}{2} \cdot |S_0 - S_1| \,. \tag{1.24}$$

Due to the absolute value, the error phase ϕ_0 does not affect the calculations. For this reason, the complex form is usually avoided and Equation 1.23 and Equation 1.24 are written using only the magnitudes

$$W = \frac{IP + OP}{2} \quad , \quad F = \frac{IP - OP}{2} \,. \tag{1.25}$$

Finally, for each pixel the Fat Fraction (FF) is given by the ratio between the fat signal and the total signal as

$$\mathrm{FF} = \frac{F}{F+W} = \frac{IP - OP}{2IP} \,. \tag{1.26}$$

The fat quantification from two-point Dixon technique is a good approximation and is a simple method, both from a theoretical and a computational point of view. However, it suffers from several important limitations[15, 17, 18, 19]:

- T_1 and T_2^* bias: there is no compensation for the relaxation differences between fat and water. Short repetition times (TR) introduce T_1-weighting that can bias the signal intensity, while the T_2^* decay during the echo time (TE) leads to signal loss and an underestimation of fat fraction.
- B_0 inhomogeneity: the magnetic field inhomogeneities affect the value of the error phase ϕ. Spatial inhomogeneities alter the phase evolution of fat and water, distorting the relationships that assume to know exactly how the two magnetization are aligned at each TE.
- Incomplete modelling of the fat spectrum: modelling the fat spectrum as a single peak centred on the methylene group resonance frequency is not enough in clinical applications. In reality, fat spectroscopy shows a multi-peak structure with different amplitudes and chemical shifts and, especially at higher field strengths, this is an important source of errors in fat quantification.

In order to overcome these limitations, more advanced Dixon-based techniques have been proposed and validated, including explicit phase corrections, compensations for T_1 and T_2^* decays and multi-peak spectral model of fat. Several methods for phase unwrapping have been studied[15] and the Dixon method was generalised by acquiring multiple echoes to better estimate the field map, water and fat images.

An efficient solution is the semi-flexible two-point method that is still based on the acquisition of two echoes but allows more flexibility in the choice of echo times[20, 21].
The acquired signal at each echo time TE_n (with $n = 1, .., N_E$) is modelled as

$$S_n = \left(W + F \cdot e^{i\theta_n}\right) \cdot e^{i\phi_n} \tag{1.27}$$

with W and F the water and fat signals in image space. θ is the dephasing angle

$$\theta_n = 2\pi \Delta f \, t_n \tag{1.28}$$

directly proportional to the resonance frequency offset of the central resonance peak of fat with respect to water and the echo time at which the sample is taken. $\Phi = e^{i\phi}$ is the phasor corresponding to the phase error, dependent on ΔB_0 and TE. The important difference with the original two-point method is that here Φ is not linearly related to TE.

In the reconstruction following the acquisition, this method analytically computes the phasor Φ and solves the complex equations to obtain the F and W images from the in-phase and out-of-phase echoes, including signal dephasing and decay and also eddy current corrections[21].

In this protocol, the imaging time is kept short and the spatial resolution is optimised. Additional improvements can be obtained by including the complete 7-peaks spectrum of fat-water, particularly relevant for the B_0 correction in the case of large fields of view. This spectrum, illustrated in Figure 1.9, shows the dominant water resonance peak alongside six smaller peaks corresponding to the different spectral components of fat.

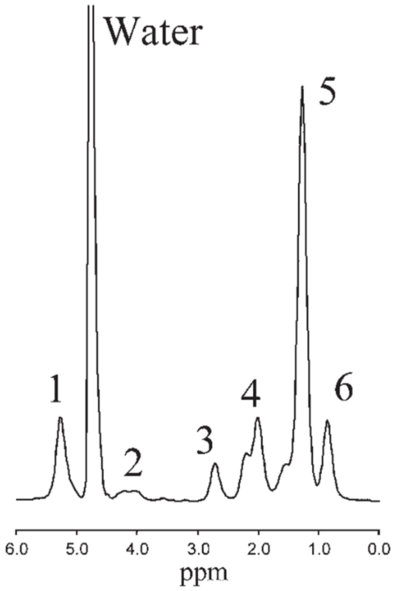


Figure 1.9: Fat-water emulsion MR spectrum acquired at 3T[22].

In qualitative fat imaging, SE sequences can be exploited if adapted to generate multiple echoes in each repetition time. Turbo Spin Echo sequences (TSE) repeat the refocusing pulse multiple times in each TR, obtaining the generation of a train of echoes that allows to collect multiple *k-lines* after each 90° excitation pulse, reducing the time needed to acquire the whole *k-space* and generate the image.

However, for an accurate fat reconstruction and quantification, the literature supports the use of multi-echo Gradient Echo sequences (GRE)[17, 20]. As shown in Figure 1.10 these sequences are composed by an initial r.f. pulse of a flip angle α, tipically low and less than 90°. The echo is generated by applying two bipolar gradients. The first one is negative and dephases the transverse magnetization, introducing spatially dependent

phase shifts due to variations in precessional frequencies. This dephasing is not caused by intrinsic field inhomogeneities alone but is induced by the gradient itself. The second gradient is positive and used as a readout gradient, effectively rephasing the spins. By reversing the polarity and duration of the initial dephasing gradient, it realigns the spins and leads to the formation of a gradient echo at the time of acquisition.
With respect to SE sequences, the echo time at which the echo is generated is significantly lower and this enables rapid readouts, allowing the sampling of the phase evolution between water and fat signals over time. Beyond being sensitive to phase variations caused by chemical shift, GRE is also compatible with multi-echo acquisition schemes and flexible echo times but is much more sensible to the T_2^* decay because the B_0 inhomogeneities are not canceled out when the echo is generated by gradients.

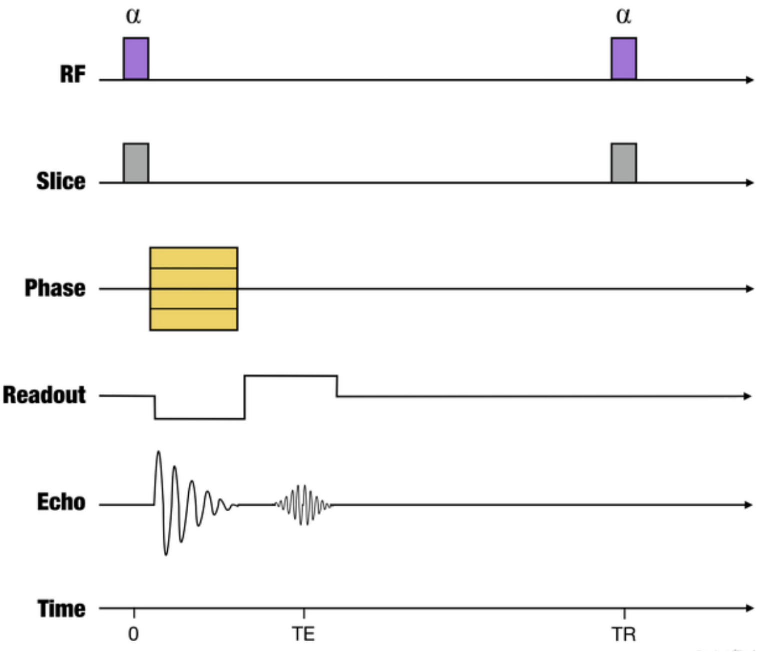


Figure 1.10: Schematic diagram of the Gradient Echo sequence (GRE)[23].

The fat fraction has gained clinical importance for the non-invasive assessment of tissue lipid content and now represents a valid quantitative imaging biomarker in several diagnostic contexts.
In living organisms, each anatomical location has its own physiological fat percentage according to the tissue function, structural characteristics and metabolic demand. The possibility to accurately measure the actual fat fraction with MRI allows to detect non-physiological changes for early diagnosis, disease monitoring and treatment response evaluation of pathological processes that involve fat accumulation, infiltration or fat

depletion. These alterations can be associated with a wide range of diseases such as obesity, hepatic steatosis, metabolic dysfunctions, neuromuscular disorders and bone marrow pathologies[20, 17].
An example of axial fat fraction imaging and its corresponding FF map acquired in a healthy subject at the upper abdomen level is shown in Figure 1.11. In FF maps, brighter regions indicate a higher fat content, whereas darker regions correspond to tissues with a lower fat fraction. The anatomical structures visible in the reported slice are the same as those described in the ADC map of Figure 1.7 but here the skin contour is more clearly delineated due to its high and homogeneous fat content.

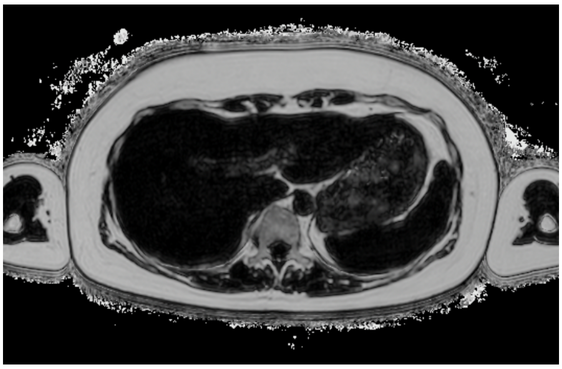


Figure 1.11: FF Map of the upper abdomen acquired in the axial plane.

1.3 Whole-Body MRI: Imaging Techniques and Applications

Recent advancements in MRI techniques have led to the development of extremely efficient sequences with high spatial resolution, reduced acquisition times and specific post-processing techniques that enable excellent tissue contrast.
These improvements have provided the foundation for the development of Whole-Body MRI (WB-MRI), a technique that allows complete imaging of the entire body within a single examination. WB-MRI combines anatomical and quantitative sequences and is particularly useful for systemic disease evaluation in oncologic staging and skeletal screening.

The next sections of this chapter will focus on the techniques and applications of WB-MRI. Beginning with its general principles and implementation in clinical practice, the specific role of WB-MRI in the diagnosis and treatment monitoring of multiple myeloma will be discussed in accordance with international guidelines.

1.3.1 WB-MRI Techniques and Clinical Applications

Whole-body MRI (WB-MRI) is a multi-station, multi-sequence imaging technique that enables the evaluation of the entire body in a single examination without the use of ionising radiation. Anatomical, functional and quantitative sequences are acquired across multiple contiguous anatomical regions in a multi-station scanning approach, in which the patient table moves to image successive body sections. The full coverage typically spans from vertex (i.e. head) to mid-thigh but can be adjusted to reach knees or feet, according to specific clinical needs.
The complete procedure generally lasts from 25 to 90 minutes, depending on the number of stations, the sequences included and scanner capabilities[24, 25].

The typical WB-MRI protocol includes anatomical, functional and quantitative sequences. Morphological T_1 and T_2-weighted images are acquired through GRE or TSE sequences. T_1-weighted images provide anatomical details and highlight bone metastases, while T_2-weighted ones are implemented with fat-suppression techniques such as STIR and can provide information about other organs and liquid components.
DWI is typically acquired during free-breathing using 2 *b-values* to reconstruct the ADC parameter: the lowest *b-value* is at least 50 s/mm^2 in order to minimise the perfusion effects and the highest is recommended to be between 800 and 1000 s/mm^2. To achieve a high signal-to-noise ratio while maintaining short acquisition times and enabling the coverage of large anatomical regions, a single-shot echo planar imaging (EPI) readout is employed in combination with a single diffusion encoding direction. This also reduces the image blurring and artefacts caused by eddy currents, without introducing biases in the tissues characterised by isotropic diffusion as the ones commonly investigated by WB-MRI[24].
In a multiparametric approach, WB-MRI protocols finally integrate Dixon-based gradient echo acquisitions to reconstruct the fat fraction (FF) maps. This is essential to evaluate bone marrow involvement and fat distribution across different tissues[26, 17].

WB-MRI is particularly useful to detect systemic, multifocal or diffuse disease patterns in oncologic, haematologic, metabolic and musculoskeletal contexts[24, 25].
Its clinical applications include the assessment of multiple myeloma, skeletal metastases from prostate or breast cancer, lymphoma, melanoma and gynecologic malignancies, as well as the evaluation of inflammatory lesions and treatment response in arthritides and myopathies . In addition, since WB-MRI does not involve ionising radiation and the associated risks, this technique is increasingly used for pediatric imaging and cancer screening in individuals with hereditary cancer predisposition syndromes such as Li-Fraumeni or Neurofibromatosis Type 1 (NF1).

In clinical practice, WB-MRI is typically acquired either in the coronal or axial plane.

Coronal acquisitions may reduce the overall scan time by requiring fewer stations (i.e. coils) but they are more subject to distortions in DWI compared to axial scans with equivalent field of view and slice coverage. Axial imaging is instead also compatible with cross-sectional anatomy of other modalities.
To ensure consistent interpretation, slice thickness must be kept constant across all sequences and its typical numerical value in protocols ranges from 5 to 7 mm[24].
WB-MRI can be performed on both 1.5 T and 3 T scanners. In particular, 3 T scanners are preferred because they offer high signal-to-noise ratio (SNR) which is an essential characteristics for quantitative techniques. However, it is important to note that at higher field strength the impact of field inhomogeneities is much more pronounced and distortions could be observed, especially near air-tissue interfaces and in Dixon scans[27].
Among the other technical difficulties, WB-MRI requires the coordinated use of multiple surface coils — including head-neck, spine and body arrays — to ensure consistent image quality across the extended field of view. The acquisitions performed in separate stations must be integrated in a post-processing step to obtain a single composite volume from spatial alignment and merge.

In order to overcome the main limitations of WB-MRI, dedicated guidelines such as the Metastasis Reporting and Data System for Prostate Cancer (MET-RADS) and the Myeloma Response Assessment and and Diagnosis System (MY-RADS) have been developed to promote protocol standardisation and improve the reproducibility of WB-MRI examinations, especially in the context of multicentre studies. These international recommendations promote the standardisation of acquisition protocols in terms of sequence selection and parameter optimisation and also provide structured criteria for image interpretation and results reporting, enhancing consistency across institutions and over time[28, 25, 29].

1.3.2 WB-MRI in Multiple Myeloma

Multiple Myeloma (MM) is a malignant plasma cell disorder characterised by the clonal proliferation of plasma cells in the bone marrow. It is the second most common haematological cancer and accounts for 1 % of all cancers, with an increasing worldwide incidence[30, 31, 32]. In Europe, approximately 50000 new cases are diagnosed annually, while in Italy the incidence is estimated at 5500–6000 new cases each year, with nearly 30000 patients currently living with the disease and a median age at diagnosis of 72 years[33].

Myeloma arises from genetic damage occurring during the development of plasma cells in the bone marrow, leading to their uncontrolled proliferation. These malignant myeloma cells produce large quantities of a single, non-functional antibody known as M-protein, monoclonal immunoglobulin or paraprotein. This accumulation process, illustrated in Figure 1.12, interferes with the normal haematopoietic function of bone marrow, preventing normal blood cells - red and white blood cells and platelets - from being produced.

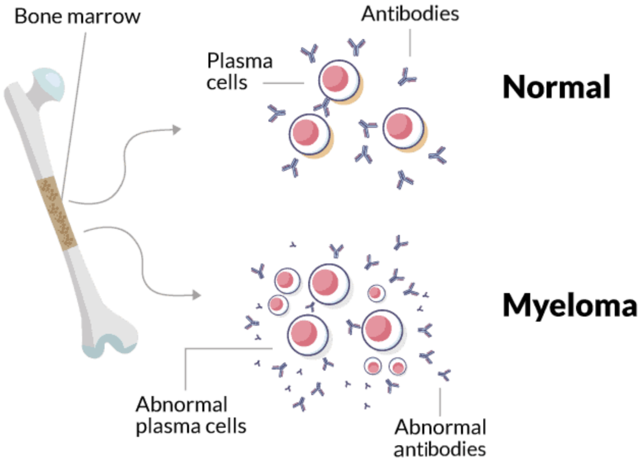


Figure 1.12: Monoclonal plasma cell proliferation and accumulation in the bone marrow in Multiple Myeloma[34].

Multiple Myeloma condition is preceded by asymptomatic precursor conditions, biologically related but clinically different.
The earliest stage is Monoclonal Gammopathy of Undetermined Significance (MGUS), defined by the presence of a low concentration of M-protein in the serum, less than 10% clonal plasma cells in the bone marrow and no evidence of end-organ damage. MGUS is a benign condition, present in 5% of the population above 50 years old, but can progress to malignant Smoldering Multiple Myeloma (SMM) with a rate of 1% per year.
SMM is still an asymptomatic condition but M-protein concentration in serum exceeds 3 g/dL and bone marrow infiltration of plasma cells ranges between 10% and 60%. In the 5 years following SMM diagnosis, 10% of patients evolve to the MM condition.

The transition from the illustrated precursor stages to active disease (MM) is defined by the detection of at least one Myeloma-Defining Event (MDE), in combination with either $\geq$10% clonal plasma cells in the bone marrow or a biopsy-confirmed plasmacytoma. MDEs include clonal plasma cell infiltration $\geq$60% in the bone marrow, a serum free light chain (FLC) ratio $\geq$100 (with the involved FLC concentration $\geq$100 mg/L) and the presence of more than one focal lesion on MRI. In addition, the proliferative activity of malignant plasma cells leads to various organ dysfunctions that serve as further evidence of active disease. These clinical complications include hypercalcemia, renal insufficiency, anemia and lytic bone lesions and are known as the CRAB criteria[35].

MM is a systemic disease since plasma cell infiltrations in the bone marrow are distributed across the entire skeleton in a diffuse pattern. Whole-body imaging is necessary to evaluate the bone marrow involvement in the axial and appendicular skeleton, both for diagnosis and monitoring of MM.
Compared to other imaging modalities such as CT or PET, WB-MRI has shown greater sensitivity in detecting both focal and diffuse pattern of bone marrow infiltration, particularly in early-stage disease. While CT primarily detects cortical bone destruction and PET results may be biased by low metabolic activities, WB-MRI is the preferred imaging technique to assess marrow composition and to detect early bone involvement. With the advantage of avoiding radiometabolites and ionising radiation exposure, WB-MRI is the first choice also in patients follow-ups and treatment monitoring[36].

As cited in WB-MRI Techniques and Clinical Applications, the Myeloma Response Assessment and Diagnosis System (MY-RADS) was developed to promote standardised use of Whole-Body MRI in patients with multiple myeloma. MY-RADS recommendations focus on acquisition protocols, image interpretation and reporting and provide a structured framework for staging, monitoring and treatment response evaluation, based on a combination of morphological, diffusion-weighted and quantitative sequences.
The core clinical WB-MRI protocol recommended for multiple myeloma is summarised in Table 1.1.

Sequence	**Detected anatomical region**	**Acquisition plane**	**Slice thickness**	**Derived calculations**
T_1-weighted FSE	Whole spine	Sagittal	4-5 mm	
STIR or fat-suppressed T_2	Whole spine	Sagittal	4-5 mm	
DWI (b=50-100 and 800-900 s/mm^2) with STIR fat suppression	Whole body (vertex to knees)	Axial	5 mm	ADC map from monoexponential data fitting
T_1-weighted GRE Dixon technique	Whole body (vertex to knees)	Axial (or coronal)	5 mm	FF map from fat and water image reconstructions

Table 1.1: Core clinical WB-MRI protocol for multiple myeloma according to MY-RADS guidelines[28]

The detection of focal lesions is based on the identification of hyperintense regions on high *b-value* DWI images, coregistered with Dixon-derived fat fraction maps[28]. In cases of pathological marrow infiltration, the replacement of adipocytes and haematopoietic cells by myeloma cells alters the tissue microenvironment, resulting in increased water diffusivity and a corresponding decrease in fat fraction relative to normal marrow.

Beyond standardised image interpretation and reporting of focal lesions, recent developments have enabled the use of radiomics and deep learning algorithms in WB-MRI image analysis.
Radiomics refers to the extraction of several mathematically defined features from medical images. The so obtained radiomic features quantitatively describe structure, morphology, texture and heterogeneity of tissues, providing a deeper characterisation of the investigated regions. Radiomic features serve as QIBs and can be integrated into machine learning models to predict plasma cell infiltration, cytogenetic aberrations and risk status directly from MRI images of the pelvic bone, in a non-invasive, repeatable and multi-focal assessment of the bone marrow[37].
In multiple myeloma, a further application is the evaluation of QIBs in regions near bone marrow biopsy sites, restricting the image field in order to target the disease diffusion patterns in the pelvic bone. This investigation is denoted as radiopsy and aims to create predictive models for clinical diagnosis based on relevant radiological features that are correlated with histological findings. The quantification and modelling of image characteristics near the biopsy site enable a virtual, non-invasive biopsy approach to monitor patients during treatment and to estimate tumor burden by assessing disease infiltration in areas distant from the biopsy site[38].

1.4 MRI Acquisition System and Protocols

All the MRI sequences analysed in this study were acquired using a Philips Ingenia 3T scanner (Philips Healthcare, Best, The Netherlands), installed at the Istituto Romagnolo per lo Studio dei Tumori (IRST) “Dino Amadori” in Meldola, Italy and shown in Figure 1.13.
This system is equipped with a 3.0 Tesla superconducting magnet and the dStream digital broadband architecture, which allows for direct digital signal transmission from the coil, improving signal-to-noise ratio (SNR) and minimising analog signal degradation.
Whole-body acquisitions were performed using a combination of coils, including Q-Body, dS Anterior, dS Base, dS HeadNeck and dS Posterior coil. The continuous coverage from head to knees is obtained with automatic table movement across stations and the final datasets are reconstructed into a single seamless image using the vendor’s MobiView algorithm for multistation image composition.
The scanner supports parallel imaging acceleration (SENSE) and is compatible with

Figure 1.13: Philips Ingenia 3T scanner[39].

Compressed-SENSE which reduces scan time while preserving image resolution and contrast.
Signal transmission is achieved using the integrated body transmit coil, while reception is performed with multi-array surface coils, automatically selected for each anatomical region. This setup provides high image homogeneity and consistent spatial resolution across the full field of view[39, 40, 41].

The scan parameters of both the imaging protocols used in this work — WB-MRI for Healthy Tissue Characterisation and WB-MRI for Multiple Myeloma Imaging — are reported in WB-MRI Protocols and Acquisition Parameters for Quantitative Sequences, with a special focus on the sequences relevant to the quantitative analysis of healthy tissues and disease-related alterations. Although both protocols included conventional anatomical sequences (such as T_1 and T_2-weighted images) for diagnostic and structural assessment, the description is limited to the sequences considered in the present work for their ability to provide reproducible and spatially resolved maps of tissue diffusivity and lipid content: diffusion-weighted imaging (DWI) and Dixon-based fat quantification techniques.

1.4.1 WB-MRI Protocols and Acquisition Parameters for Quantitative Sequences

The two WB-MRI protocols used in this work, both implemented and acquired at IRST in Meldola, are the one dedicated to breast cancer patients, employed for healthy tissue characterisation, and the one used for multiple myeloma imaging.
Both protocols include T_1-weighted TSE and STIR TSE as conventional anatomical sequences for diagnostic and structural evaluation.

The two protocols slightly differs in the anatomical coverage: the multiple myeloma protocol extends from the vertex to the knees, while the breast cancer protocol covers from the vertex to approximately mid-thigh or just below the pelvic region. This difference in coverage is reflected in the acquisition durations, which tends to increase when additional anatomical stations are included, although sequence-specific parameters may also contribute to variability between protocols.

DWI is included in both protocols and can be classified as spin-echo echo-planar imaging (SE-EPI) scheme with inversion recovery fat suppression, offering rapid acquisition with robust sensitivity to microscopic water motion. This allows for voxel-wise ADC computation through mono-exponential fitting of the signal attenuation between two diffusion weightings (i.e. two images acquired at different *b-values*).
The acquisition parameters for both protocols are summarised in Table 1.2.

Acquisition Parameter	**WB-MRI for Healthy Tissue Characterisation**	**WB-MRI for Multiple Myeloma Imaging**
Acquisition Duration [s]	204.62	172.84
Pulse Sequence Name	DwiIR	DwiIR
Echo Pulse Sequence	SE	SE
Acquisition Plane	Axial	Axial
Slices Thickness [mm]	6	6
FOV [mm]/Acquisition Matrix	450/128 × 128	450/128 × 106
Pixel Spacing [mm]	1.8 × 1.8	1.8 × 1.8
Phase Encoding Direction	AP	AP
SENSE Acceleration	Yes	Yes
TR/TE [ms]	4946.64/62.80	8642.15/53.02
Echo Train Length	47	31
Echo Numbers	1	1
Flip Angle [°]	90	90
Pixel Bandwidth [Hz/pixel]	3130.00	3734.53
Fat Suppression Technique	SPIR	SPIR
b-values [s/mm^2]	50, 800	50, 800
Diffusion Gradient Directions	3	3

Table 1.2: Acquisition parameters for DWI sequences in both WB-MRI protocols.

Despite minor variations in acquisition parameters between the DWI sequences used in the two protocols such as repetition time, echo time, acquisition matrix size, echo train length and pixel bandwidth, these differences are not expected to significantly affect the comparability of ADC values. Such parameters primarily influence image quality factors like signal-to-noise ratio, resolution and scan time. However, both sequences share the same pulse sequence design (DwiIR), acquisition plane (axial), *b-values* (50 and 800 s/mm^2), number of diffusion gradient directions (3) and fat suppression technique, ensuring consistent diffusion weighting and anatomical coverage. Therefore, the two protocols can be considered technically comparable for the voxel-wise quantification of tissue diffusivity and the results obtained for patients scanned with either protocols can be considered equivalent and suitable for direct comparison.

The fat fraction (FF) maps analysed in this study were generated using a dual-echo gradient-echo (GRE) sequence, implemented on the Philips Ingenia 3T scanner as part of the vendor-specific mDixon family (Pulse Sequence Name: T1FFE). This implementation, often referred to as mDixonAll, acquires two echoes with different echo times and applies a two-point Dixon reconstruction algorithm to separate fat and water components. The resulting parametric maps provide voxel-wise estimates of proton-density fat fraction expressed as a percentage. In both protocols, the short echo train length (Echo Train Length = 2) is consistent with the dual-echo design and enables fat-water decomposition with minimised artefacts.

Acquisition parameters used in the two WB-MRI protocols are summarised in Table 1.3. The two FF sequences are technically equivalent and, as stated for DWI sequences, the differences between the two protocols are not expected to significantly alter the quality or quantitative consistency of the resulting maps.

Acquisition Parameter	**WB-MRI for Healthy Tissue Characterisation**	**WB-MRI for Multiple Myeloma Imaging**
Acquisition Duration [s]	13.67	16.05
Pulse Sequence Name	T1FFE	T1FFE
Echo Pulse Sequence	GRE	GRE
Acquisition Plane	Axial	Axial
Slice Thickness [mm]	3	3
FOV [mm]/Acquisition Matrix	430/268 × 217	480/300 × 237
Pixel Spacing [mm]	0.8 × 0.8	0.8 × 0.8
Phase Encoding Direction	AP	AP
SENSE Acceleration	Yes	Yes
TR/TE [ms]	3.79/shortest	3.64/shortest
Echo Train Length	2	2
Echo Numbers	1	1
Flip Angle [°]	10	10
Pixel Bandwidth [Hz/pixel]	1110	1344.09
Fat Quantification Model	2-point Dixon	2-point Dixon

Table 1.3: Acquisition parameters for Fat Fraction sequences (mDixonAll-based) in both WB-MRI protocols.

In addition to the conventional dual-echo mDixon-based sequences used for FF map-

ping, the more recent mDixonQuant implementation has been lately integrated into the WB-MRI protocols for both breast cancer and multiple myeloma imaging.
This sequence is based on a low-flip-angle, multi-echo gradient-echo acquisition (Pulse Sequence Name: T1FFE), specifically designed to support advanced fat quantification through a multi-peak fat spectral model. The mDixonQuant framework acquires six echoes with different echo times (Echo Train Length = 6), enabling simultaneous estimation of fat, water and T_2^* maps while correcting for confounding effects such as eddy currents and field inhomogeneities. This sequence was initially developed for the accurate quantification of hepatic fat content but was subsequently optimised for whole-body imaging applications with high spatial resolution and short acquisition times over a wide range of echo times and fields of view[20].

Acquisition parameters used for the mDixonQuant sequences in both WB-MRI protocols are summarised in Table 1.4.
The sequence design and acquisition parameters are identical in both the protocols. The only different value is the acquisition duration and it is due to the anatomical coverage of the protocols.

Acquisition Parameter	**WB-MRI for Healthy Tissue Characterisation**	**WB-MRI for Multiple Myeloma Imaging**
Acquisition Duration [s]	12.61	15.20
Pulse Sequence Name	T1FFE	T1FFE
Echo Pulse Sequence	GR	GR
Acquisition Plane	Axial	Axial
Slice Thickness [mm]	6	6
FOV [mm]/Acquisition Matrix	400/160 × 160	400/160 × 140
Pixel Spacing [mm]	2.1 × 2.1	2.1 × 2.1
Phase Encoding Direction	AP	AP
SENSE Acceleration	Yes	Yes
TR/TE [ms]	5.61/shortest	5.61/shortest
Echo Train Length	6	6
Echo Numbers	1	1
Flip Angle [°]	3	3
Pixel Bandwidth [Hz/pixel]	2367	2367
Fat Quantification Model	Multi-peak Dixon	Multi-peak Dixon

Table 1.4: Acquisition parameters for Fat Quant sequences (mDixonQuant-based) in both WB-MRI protocols.

2 Quantitative ADC Analysis in Healthy Tissues

In order to apply the Apparent Diffusion Coefficient (ADC) as a reliable clinical biomarker, it is essential to define its typical range and variability in normal tissues. Deriving reference distributions and statistical measures from a well-controlled healthy cohort is necessary to distinguish physiological variability from pathological alterations in clinical practice and research applications.

This chapter presents the quantitative assessment of ADC in healthy tissues, focusing on seven anatomically distinct organs.
Regions of interest (ROIs) were manually placed on WB-MRI images acquired with a standardised protocol from a large cohort of healthy subjects and later processed for feature extraction. Statistical analysis was performed to describe the behaviour of ADC values and the results are presented in terms of sample repeatability, organ-specific characterisation and intra-subject repeatability for a single patient.

2.1 Patients

This study was based on a dataset of 20 adult female patients who underwent WB-MRI at the IRST ”Dino Amadori” (Meldola, Italy) between 2023 and 2025, as part of clinical staging or follow-up for oligometastatic breast cancer. All imaging was acquired using a standardised WB-MRI protocol[29], described in detail in WB-MRI Protocols and Acquisition Parameters for Quantitative Sequences. Diffusion-Weighted Imaging (DWI) was performed with two *b-values* (0 and 800 s/mm^2) and Apparent Diffusion Coefficient (ADC) maps were generated through mono-exponential fitting.

Patients with invasive ductal or lobular carcinoma were excluded to minimise the likelihood of undetected lesions in the tissues of interest. Only scans showing no visible abnormalities in the target organs were considered for analysis. Furthermore, images were reviewed to ensure high diagnostic quality: acquisitions containing severe motion artefacts or scanner-induced distortions were excluded so that only anatomically intact regions were included in the ROI analysis.

The dataset includes 20 patients (mean age 57 ± 13 years, range: 35–79 years) for a total of 41 acquisitions. For most subjects, multiple time points were selected, as patients typically underwent repeated WB-MRI examinations for clinical staging and monitoring. Among the available scans, a subset was selected for each patient based on image quality and study requirements. Specifically, two acquisitions were selected for 14 patients,

one acquisition for four patients, three acquisitions for one patient and six longitudinal scans for one patient, with inter-scan intervals ranging from three to nine months.
This structure enables both inter-subject analysis across the whole healthy cohort and intra-subject repeatability assessment based on longitudinal data from a single case.

2.2 Data Collection and Processing

Following image acquisition, all data were anonymised and processed to extract quantitative diffusion metrics from anatomically healthy tissues. All 41 acquisitions underwent the same standardised analysis workflow to maximise reproducibility and comparability across subjects. The following sections describe in detail the methodology used for ROIs selection and placement and radiomic feature extraction using the SIBEX software.

To verify the accuracy of ADC measurements, a preliminary test acquisition was performed using a spherical water phantom at room temperature. The mean ADC value measured in the phantom (2047 ± 61 mm^2/s) was consistent with expected reference values, supporting the reliability of the imaging protocol[42].

2.2.1 ROIs Placement

WB-MRI visualization and ROIs placement were performed using MIM® (MIM Software Inc., Cleveland, OH, USA[43]), a commercial platform widely used in clinical imaging and research applications. The software allowed for interactive manipulation of the multi-planar image data and facilitated consistent ROIs placement across patients and anatomical planes.

To evaluate diffusion properties across multiple tissue types, seven anatomically distinct regions were selected: liver, spleen, pancreas, kidneys, vertebrae, pelvic bone and femoral head. These organs were chosen for their clinical relevance, their tissue heterogeneity and their inclusion in Whole-Body MRI assessments. Importantly, they represent a mix of parenchymal and osseous tissues with varying microstructural properties and diffusion behaviours, providing a comprehensive sampling of physiological ADC values.

For each organ, three regions of interest (ROIs) were manually placed to improve statistical robustness. ROIs were drawn with a cylindrical shape and a target volume of approximately 2.6 cm^3. Based on voxel dimensions ($1.8 \times 1.8 \times 6$ mm^3), this corresponds to an average of $\sim$140 voxels per ROI, although minor variations occurred due to pixel interpolation introduced by the MIM® software during placement.
Due to its smaller anatomical size and more irregular shape, the pancreas required a reduced ROI volume of 0.5 cm^3, in order to ensure precise placement and avoid partial

volume effects.
The ROIs were initially positioned on axial slices and subsequently refined using coronal and sagittal views to maximise anatomical accuracy across all three planes. An example of ROI placement across axial, sagittal and coronal planes is shown in Figure 2.1, illustrating the multi-organ sampling strategy adopted in this study. Note that not all 21 ROIs are simultaneously visible across the three views, as only those intersecting the slice position defined by the reference cursor (green crosshair) in the MIM® viewer are displayed in each plane.

Special attention was given to anatomical consistency across patients: ROIs were placed in visually homogeneous areas of tissue, avoiding visible vessels, lesion-like structures or partial volume effects. This approach was adopted to minimise bias and ensure internal tissue uniformity in the sampled regions.

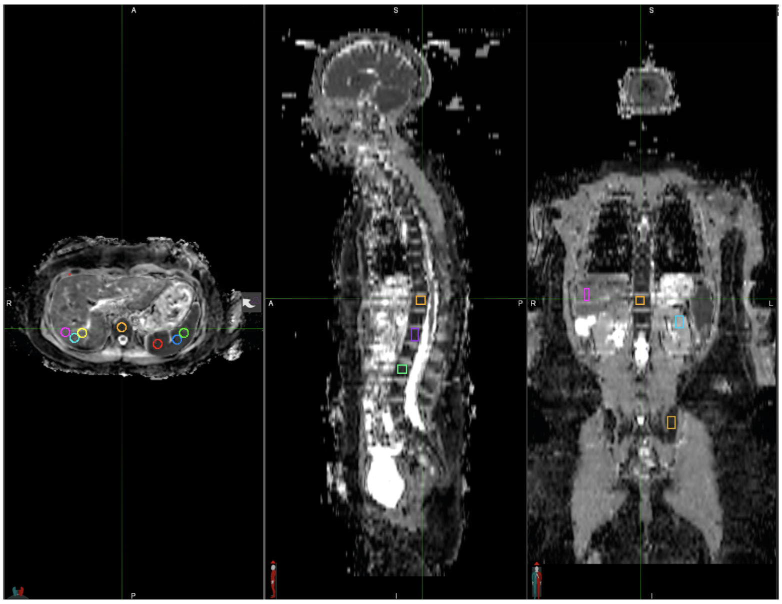


Figure 2.1: ROIs placement for ADC analysis.

2.2.2 Features Extraction

Quantitative feature extraction was carried out using SIBEX (Standardised Imaging Biomarker EXtraction), a MATLAB®-based radiomic platform developed for reproducible analysis in accordance with IBSI guidelines for feature computation[44, 45]. Prior to feature extraction, image volumes and corresponding ROI structures (RTSTRUCT) were exported from MIM® in order to avoid potential inconsistencies related to the MobiView multistation reconstruction, particularly at station (i.e. coil) junctions.

SIBEX was configured to process each ROI individually, applying a fixed binning strategy (32 bins, fixed bin number rescaling) and a standard outlier exclusion preprocessing step based on a 3σ rule. This means that for each ROI, voxels with intensities more than three standard deviations from the mean (3σ) were excluded to reduce noise and the influence of extreme values and all the remaining voxel intensities were rescaled to fit 32 equally populated bins, ensuring feature comparability.
The software computed a total of 144 radiomic features across several IBSI-defined categories. The complete set of extracted features is reported in Appendix A with selected categories and representative features described below:

- First-order intensity features: mean, median, variance, skewness, kurtosis, percentiles, energy and entropy;
- Histogram-based features: interquartile range, uniformity, mode and histogram gradient metrics;
- Grey Level Size Zone Matrix (GLSZM) and Grey Level Distance Zone Matrix (GLDZM): zone emphasis, non-uniformity and zone entropy;
- Neighbourhood Intensity Difference (NID): busyness, contrast, complexity and texture strength;
- Neighbourhood Grey Level Dependence (NGLD): dependence emphasis, grey level count metrics, non-uniformity and entropy.

Additional categories included Local Intensity Features and Intensity-Volume Histogram (IVH) descriptors, providing a comprehensive characterisation of signal distribution and spatial heterogeneity within each ROI.
All the extracted features were stored in structured matrices for subsequent statistical analysis.

2.3 ADC Statistical Analysis

The structure of the dataset enabled three distinct but complementary types of statistical analysis, each contributing to the assessment of ADC measurement reliability from a different perspective. The analyses are presented in an order that reflects both methodological coherence and interpretative progression. First, a repeatability analysis was performed across the entire sample (ADC Repeatability Across the Sample), introducing the metrics recommended by the Quantitative Imaging Biomarkers Alliance

(QIBA) and establishing a baseline estimation of intra- and inter-subject variability in realistic clinical conditions. A single-subject longitudinal analysis (Intra-Subject ADC Repeatability) was then conducted to isolate intra-subject variability without the effect of inter-subject biological heterogeneity. Finally, an intra-organ analysis (Organ ADC Variability and Feature Stability) explored the spatial consistency of ADC values within anatomical regions and extended the investigation to additional radiomic features beyond the mean ADC value.

The conceptual and methodological framework for these analyses is based on QIBA guidelines which define standardised metrics and procedures to ensure the reliability and reproducibility of quantitative imaging biomarkers such as ADC across time, subjects and imaging platforms[46].

2.3.1 ADC Repeatability Across the Sample

To assess the repeatability of ADC measurements in healthy tissues, a statistical analysis was conducted on the subset of subjects who underwent multiple Whole-Body MRI scans. The aim was to quantify the intra-subject variability which is the expected variation in ADC values for a given individual across different acquisitions.
The analysis is in accordance to the statistical methodology and reporting metrics recommended by QIBA[47, 48] adapted to a real-world clinical setting where follow-up acquisitions are typically taken some months apart, instead of applying a test-retest protocol.

To reduce measurement noise, the ADC value was first averaged across all ROIs belonging to the same subject, organ and acquisition date. This aggregated mean was considered to be the representative ADC value for that subject, organ and scan combination.
In order to estimate the intra-subject repeatability, only the 16 patients with at least two acquisitions were retained for this part of the analysis.

Subsequently, the inter-subject averages across sessions were calculated for each organ to serve as the reference values for the repeatability analysis. These averages form the baseline for estimating the within- and between-subject standard deviations.
The resulting distributions are graphically represented in the boxplot in Figure 2.2 and the corresponding numerical results will be reported in Table 2.1 at the end of this section. In the boxplot, the maximum variation is represented by the whiskers, which extend to the most extreme data points not considered outliers, while the coloured boxes indicate the interquartile range, representing the middle 50% of the data. The horizontal line inside each box represents the median value, whereas the upper and lower edges of the box correspond to the third and first quartiles, respectively. Circles mark

outliers, defined as values lying more than 1.5 times the IQR above the third quartile or below the first quartile.

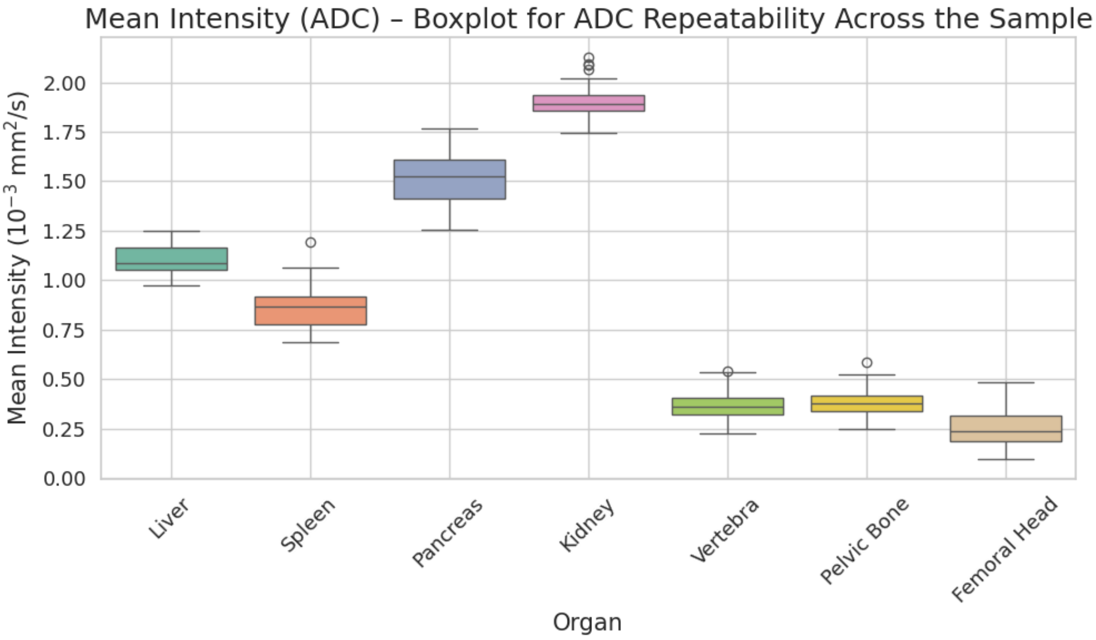


Figure 2.2: Boxplot of the mean intensity (ADC value) across the different organs for the ADC Repeatability Analysis Across the Sample.

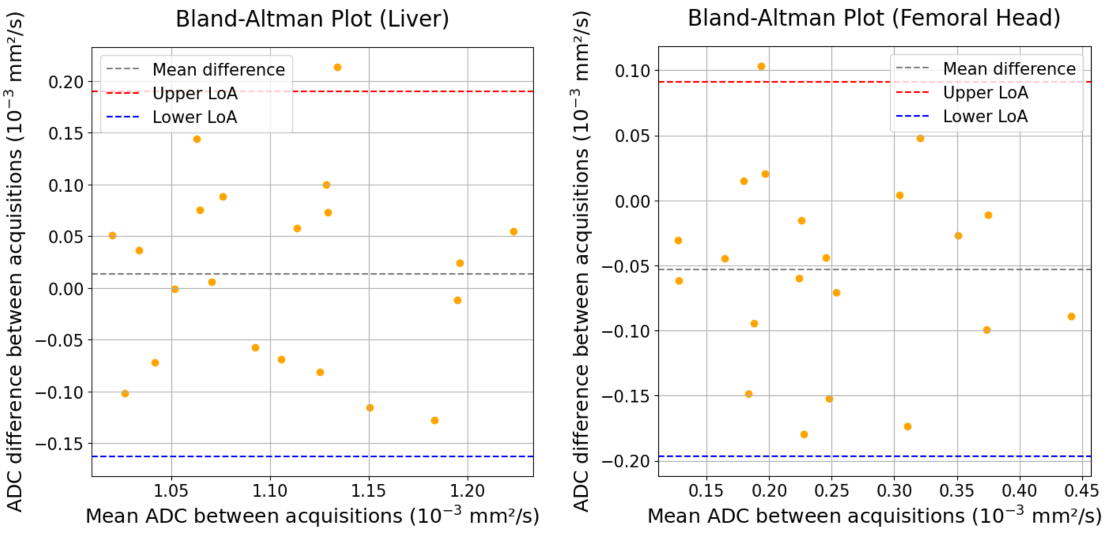
For each of the seven organs, a linear mixed-effects model (MixedLM) was fitted to the ADC values. The model included a random intercept for each subject and allowed the decomposition of the total variance into σ_b^2, a between-subject component representing the differences in mean ADC among the different patients of the sample, and σ_w^2, the within-subject residual component that represents the variability across repeated measurements within the same individual. The total variance σ^2 is given by the sum of these components.

The primary metric of interest was the within-subject standard deviation, from which the Repeatability Coefficient (RC) was computed following the QIBA definition. This coefficient estimates the maximum expected difference between two repeated measurements on the same subject in 95% of cases. In addition, the within-subject Coefficient of Variation (wCV) was calculated to provide a normalised measure of dispersion.
The definitions and mathematical formulas of the repeatability metrics used in this analysis are reported in Appendix B.

The results of the statistical analysis, computed across all organs across all the available repeated scans, are summarised in Table 2.1.

Organ	Mean ADC	σ	σ_w	σ_b	RC	wCV (%)	LOA$_\text{L}$	LOA$_\text{U}$
Liver	1.10	0.08	0.06	0.05	0.17	7	-0.16	0.19
Spleen	0.86	0.10	0.09	0.06	0.24	12	-0.21	0.26
Pancreas	1.50	0.15	0.13	0.08	0.35	10	-0.34	0.33
Kidney	1.92	0.09	0.05	0.08	0.14	5	-0.14	0.14
Vertebra	0.37	0.08	0.07	0.02	0.20	21	-0.18	0.12
Pelvic bone	0.38	0.07	0.05	0.05	0.15	19	-0.16	0.13
Femoral head	0.25	0.10	0.09	0.04	0.24	38	-0.20	0.09

Table 2.1: Results for the ADC Repeatability Analysis Across the Sample. All values are expressed in 10^{-3} *mm*^{2}*/s, except the ones explicitly expressed as percentages.*



[A] *[B]*

Figure 2.3: Bland-Altman plots for the ADC Repeatability Analysis Across the Sample for the liver[A] and the femoral head[B].

Along with the quantitative analysis, Bland–Altman plots were generated for each organ to visualise the distribution of differences between repeated measurements. For every patient-organ pair with multiple acquisitions, only successive pairs were considered, ordered chronologically. Each plot displays the individual differences as a function of the average ADC, along with the mean difference (bias) and Limits of Agreement (LoAs). LoAs were calculated as defined by the general form of the Bland-Altman method[49], deviating from the QIBA profile that defines LoAs using a symmetric formulation around

zero. The general form explicitly accounts for any systematic bias between paired measurements, instead of assuming that no bias exists between repeated scans. In this clinical setting, where repeated scans are separated by months, the general form was considered to be more appropriate since it allows for the detection of systematic differences between acquisitions without imposing the assumptions of perfect agreement and zero bias.
Two representative plots are shown in Figure 2.3 for the liver and the femoral head, highlighting variability patterns in soft tissue and osseous structures with different diffusion characteristics.

2.3.2 Intra-Subject ADC Repeatability

Following the repeatability analysis conducted on the whole sample, a subject-specific investigation was conducted to assess the temporal consistency of ADC measurements in the absence of inter-individual biological variability. For this purpose, the single patient with six longitudinal scans (Patient 20) was considered.

Unlike standard test–retest protocols, which typically involve scan–rescan procedures with patient repositioning and short intervals between acquisitions, the imaging sessions considered for this analysis were acquired over a period of 24 months, with inter-

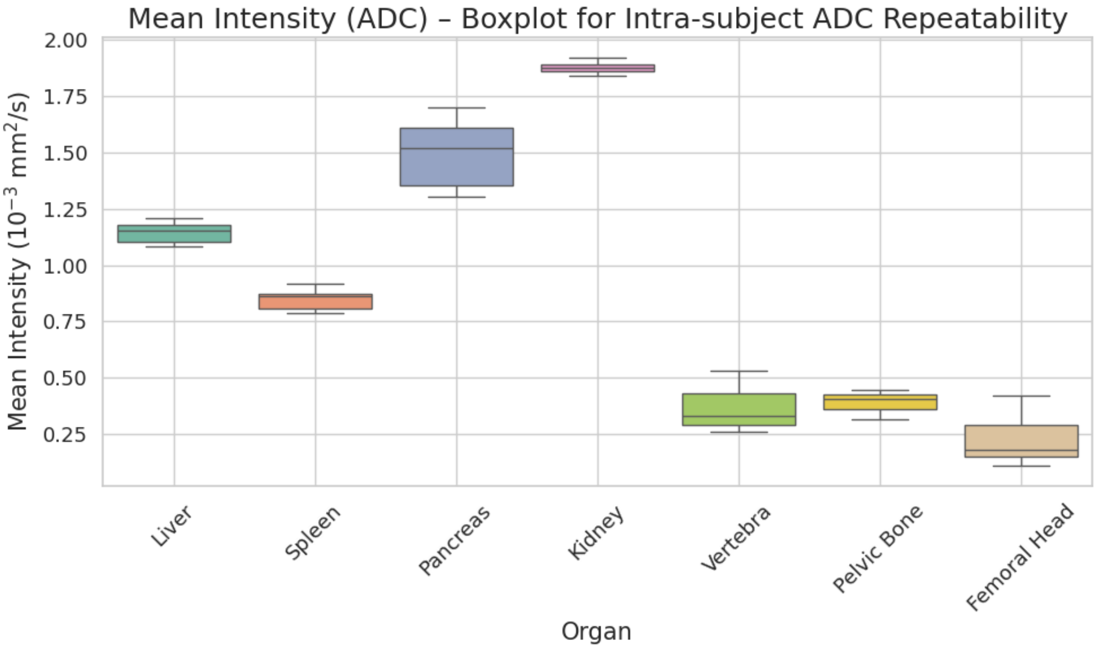


Figure 2.4: Boxplot of the mean intensity (ADC value) across the different organs for the Intra-Subject ADC Repeatability Analysis.

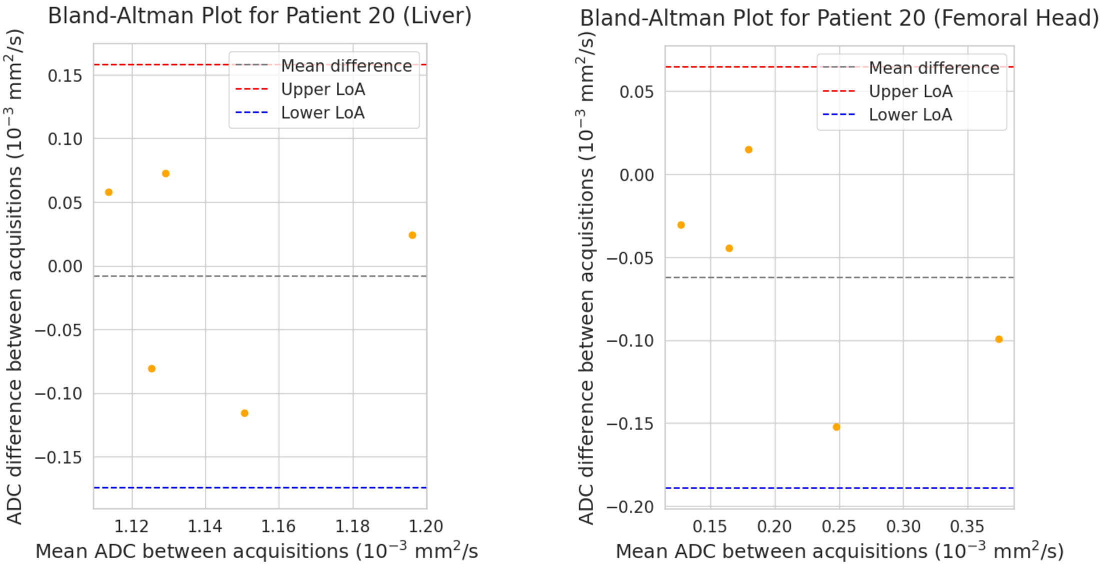
scan intervals ranging from three to nine months.
This approach allows for the evaluation of ADC stability under realistic clinical follow-up conditions, where measurements are expected to remain consistent over time in the absence of pathological changes or scanner-related inconsistencies.

For each of the seven anatomical regions considered, the three ROI values per organ and acquisition were averaged to obtain a single ADC value per time point. For the patient with six longitudinal acquisitions, the mean ADC value was computed for each organ across all time points. These values are expected to be consistent with those derived from the group-level analysis presented earlier.
The graphical results are shown in Figure 2.4.

The time series for the patient were then processed to compute the within-subject variance (σ_w^2), the within-subject coefficient of variation (wCV) and the repeatability coefficient (RC). The lower and upper LoAs were computed and, for each organ, a Bland–Altman analysis was performed by computing the pairwise differences between successive acquisitions.
The results of this repeatability analysis for each organ are reported in Table 2.2 while the Bland-Altman plots for the liver and the femoral head are shown in Figure 2.5.

Organ	**Mean ADC**	σ_w	**RC**	**wCV (%)**	**LoA$_L$**	**LoA$_U$**
Liver	1.15	0.05	0.14	4	-0.17	0.16
Spleen	0.85	0.05	0.14	6	-0.15	0.10
Pancreas	1.50	0.16	0.46	11	-0.46	0.34
Kidney	1.88	0.03	0.08	1	-0.10	0.08
Vertebra	0.37	0.11	0.30	29	-0.22	0.14
Pelvic Bone	0.39	0.05	0.14	13	-0.14	0.14
Femoral Head	0.23	0.12	0.34	53	-0.19	0.06

Table 2.2: Results for the Intra-Subject ADC Repeatability Analysis for a single patient with six longitudinal acquisitions. All values are expressed in 10^{-3} mm^2/s, except the ones explicitly expressed as percentages.



[A] *[B]*

Figure 2.5: Bland-Altman plots for Intra-Subject ADC Repeatability Analysis for a single subject for the liver[A] and the femoral head[B].

2.3.3 Organ ADC Variability and Feature Stability

In order to quantify the statistical variability of the mean ADC value within each anatomical region, all 41 acquisitions of the dataset were considered, using the three ROIs defined for each acquisition, and an exploratory statistical analysis was performed organ by organ.
This methodology was extended to the full set of radiomic features extracted by SIBEX. The evaluation of the intra-organ stability of all the features allowed the identification of those with consistent behavior within each anatomical region. These stable descriptors are the ones suitable for tissue characterisation and can be considered as candidate Quantitative Imaging Biomarkers (QIBs), potentially integrable into predictive models of clinical outcomes.

Initially, boxplots, histograms and violin plots were generated to visualise the distribution of mean ADC values within each organ. These plots allowed for the preliminary identification of organs exhibiting greater internal heterogeneity or the presence of potential outliers.
Histograms were chosen to show the distribution shape and value range for ADC in each tissue and the plots are shown in Figure 2.6 for all the seven organs.

Subsequently, descriptive statistics were computed for each organ, including mean,

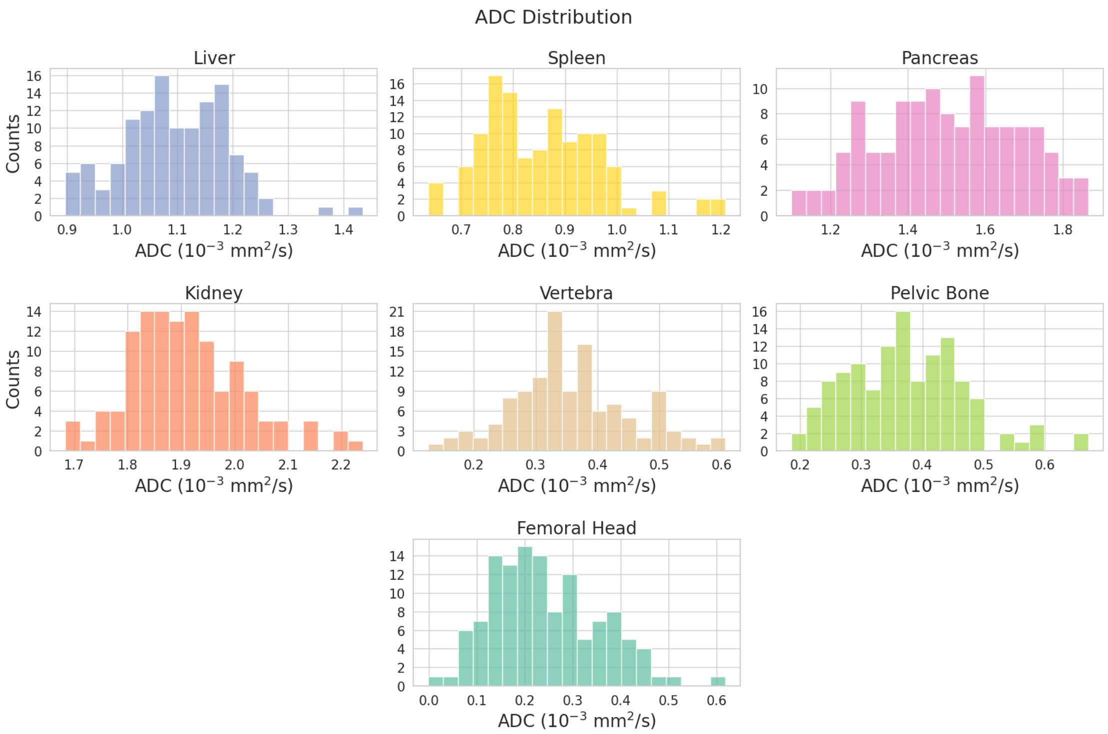


Figure 2.6: Histograms of the distributions of the mean ADC value in the different organs.

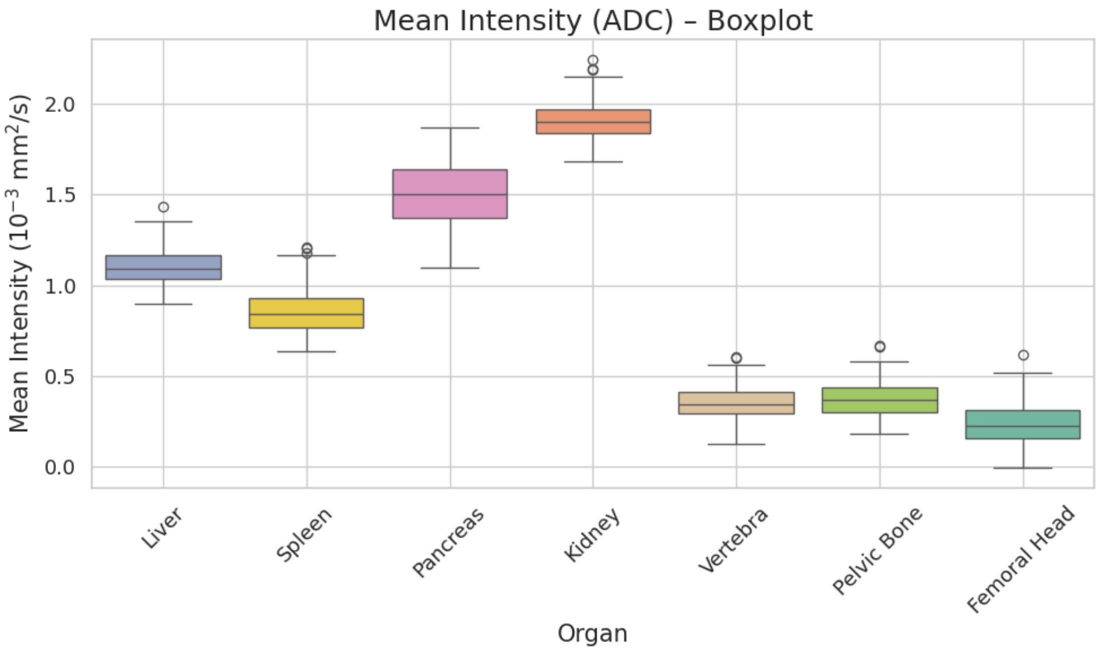


Figure 2.7: Boxplot of the mean intensity (ADC value) across the different organs.

standard deviation, median, minimum and maximum value, interquartile range (IQR) and Coefficient of Variation (CoV). The CoV, defined as the ratio between the standard deviation and the mean, provides a normalised measure of relative dispersion of data around the mean, usually expressed as a percentage, and is commonly used in radiomic and quantitative imaging to assess feature repeatability and robustness.
The boxplot for the ADC values across the different organs is shown in Figure 2.7. This visualisation provides a clear graphical summary of the central tendency, spread and presence of statistical outliers in each anatomical region, allowing for direct comparison of variability levels across organs and highlighting tissues with more homogeneous or more dispersed ADC distributions.

Outliers were removed using the standard IQR method, as suggested by QIBA guidelines and reported in Appendix B, applied independently for each organ. This step is especially important in this setting, as unreliable values may result from segmentation inaccuracies, image noise or atypical local tissue variability.
Overall, the outlier removal step led to minimal data loss, with less than 2.5% of ROIs excluded per organ: three outliers (2.4%) were removed for the kidney and spleen, two (1.6%) for the pelvic bone and vertebrae and only one (0.8%) for the liver and femoral head. No outliers were detected in the pancreas and therefore the analysis proceeded with the full set of 123 available ROIs for this organ.
After outlier removal, descriptive metrics were recomputed to provide more robust estimates. The results are reported in Table 2.3.

Organ	Mean ADC	σ	Median	Min	Max	IQR	CoV (%)
Liver	1.09	0.09	1.09	0.90	1.35	0.13	8
Spleen	0.85	0.10	0.84	0.64	1.17	0.16	12
Pancreas	1.50	0.18	1.50	1.10	1.87	0.27	12
Kidney	1.90	0.10	1.90	1.68	2.15	0.13	5
Vertebra	0.35	0.09	0.34	0.13	0.56	0.11	26
Pelvic bone	0.37	0.09	0.37	0.19	0.58	0.14	24
Femoral head	0.24	0.11	0.22	0.00	0.52	0.15	46

Table 2.3: Results for Organ ADC Variability Analysis after outlier removal. All values are expressed in 10^{-3} *mm*^{2}*/s, except the ones explicitly expressed as percentages.*

Up to this point, the statistical analysis focused on the mean ADC value, corresponding to the Mean Intensity of Grey Level feature. However, the same methodology can be extended to the entire set of radiomic features extracted via SIBEX. For each feature, distribution plots were generated separately for each organ, using both boxplots and histograms for a visual assessment.
To quantitatively evaluate the intra-organ variability of each radiomic feature, the Co-

efficient of Variation (CoV) was computed across all ROIs for each organ-feature pair, after outlier removal.
Since CoV expresses relative variability as a percentage of the mean, lower values indicate higher stability. Based on this principle, features were categorised into three stability levels: high stability (i.e. low variability, CoV $\leq$ 15%), moderate stability (15% $<$ CoV $\leq$ 30%) and low stability (30% $<$ CoV $\leq$ 45%). Features with a CoV exceeding 45% were considered too unstable and excluded from further analysis. Features that are by definition constant or degenerate such as Discretised Intensity Range, Minimum or Maximum Discretised Intensity and Dependence Count Percentage were also excluded from the start.

To assess redundancy and reduce the dimensionality of the feature space, Pearson correlation matrices were computed for each organ within each CoV range. An illustrative

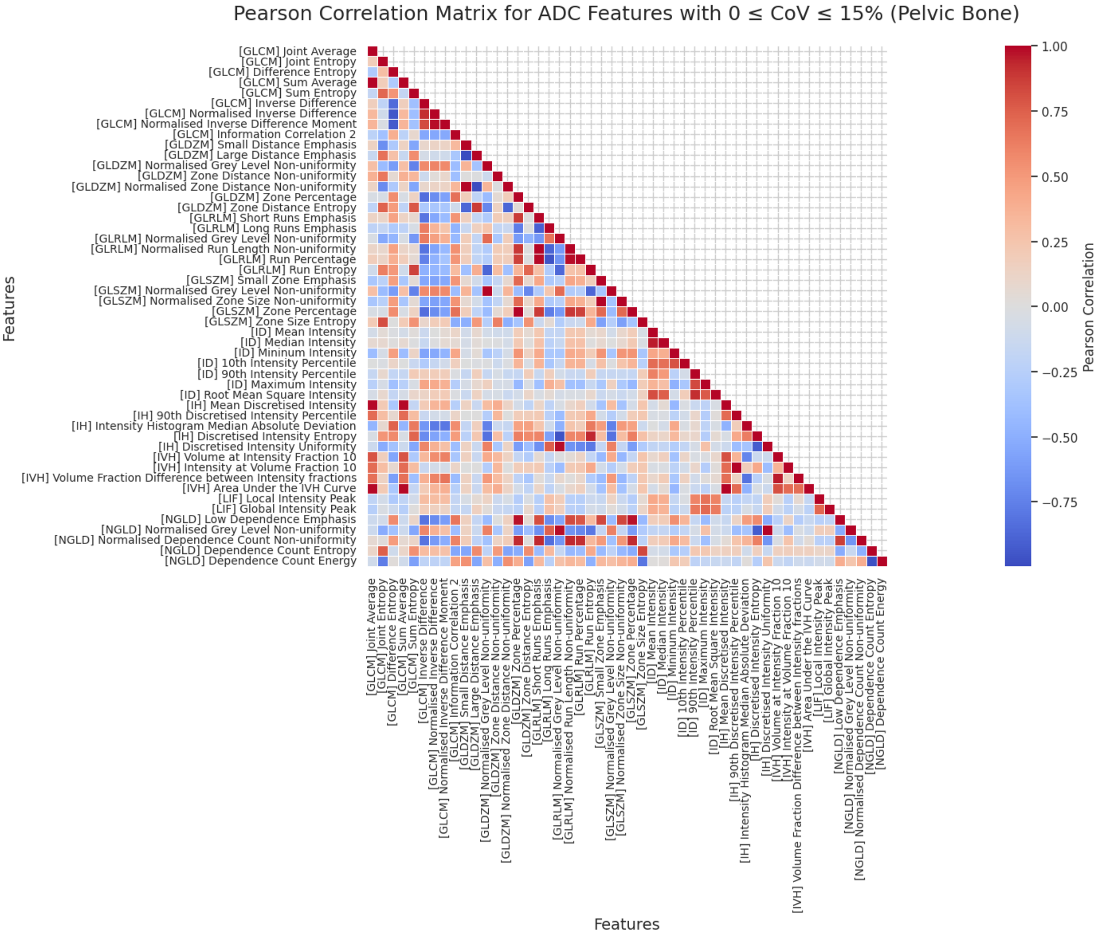


Figure 2.8: Pearson correlation matrix of ADC features with CoV ≤ 15% in the pelvic bone.

example of the obtained correlation matrices is shown in Figure 2.8, for the case of the pelvic bone and radiomic features with CoV $\leq$ 15%.
Regions of the matrix exhibiting strong linear correlation ($r > 0.85$) were interpreted as clusters of redundant features. From each group of highly correlated features, one representative was selected, preferring lowest-order descriptors or, when equivalent, most stable ones, defined as those with the lowest CoV.

In parallel, feature stability was also assessed using the Intraclass Correlation Coefficient (ICC), through a linear mixed-effects model accounting for inter-subject variability via a random effect. In this model, all ROIs belonging to the same subject were grouped together, without distinguishing between separate acquisitions or timepoints. This reflects the assumption that the three ROIs drawn for each organ within the same subject are independent samples from the same underlying anatomical structure. As a result, the ICC quantifies the proportion of variance attributable to differences between subjects, relative to the total variance observed across all ROIs.
The ICC and CoV are widely used and complementary metrics for quantifying feature reproducibility and variability in radiomics[50]. However, ICC is particularly suited to repeated measurements on identical anatomical locations, such as test-retest acquisitions of the same ROI[47]. In this context, ROIs were distinct and anatomically non-overlapping, though located within the same organ. As a result, the conditions required for a proper application of ICC are not fully satisfied and its use must be interpreted with caution.

Empirically, the ICC was computed for each feature and each organ, following the same structure adopted for the CoV analysis.
A threshold of ICC $\geq$ 0.5 was applied to identify features with at least moderate reliability.
Compared to CoV-based selection, this resulted in a more restrictive filtering, with a smaller number of features kept and generally sparser Pearson correlation matrices.

2.4 Discussions

The results presented in this chapter have been interpreted with reference to the existing scientific literature, in order to assess their methodological and clinical consistency. The discussion is structured into two main sections. The first one addresses the repeatability of ADC measurements, based on two complementary test–retest analyses, both involving temporally separated acquisitions: one conducted on a cohort of healthy subjects and the other on a single subject. The second one instead focuses on the characterisation of physiological ADC values across healthy tissues, with particular attention to their stability and intra-organ variability.

It is important to highlight that all analyses were performed on WB-MRI acquisitions. Compared to region-specific MRI, WB-MRI faces additional challenges due to its broader anatomical coverage, increased variability across tissue types and higher potential for acquisition-related artefacts. This context is essential when evaluating the reproducibility and clinical applicability of the results.

2.4.1 Repeatability of ADC Measurements

Repeatability of ADC measurements was assessed using two extended test–retest designs, both involving temporally separated acquisitions in a real-world clinical setting. The first analysis, described in ADC Repeatability Across the Sample, was conducted on the subset of subjects who underwent multiple Whole-Body MRI scans (at least two). The second, presented in Intra-Subject ADC Repeatability, was performed on a single subject who underwent six separate acquisitions over a total period of 24 months.
The aim was to evaluate both the impact of biological variability and the technical repeatability of ADC measurements under routine clinical conditions.

In parenchymal organs such as the liver, spleen, kidneys and pancreas, repeatability was generally high across both analyses. In the liver, for example, the repeatability coefficient (RC) and within-subject coefficient of variation (wCV) improved from 0.17 and 7% across the sample to 0.14 and 4% in the intra-subject setting. In both cases, the LoAs were narrow and essentially symmetrically distributed around zero, as observed in the corresponding Bland–Altman plots (Figure 2.3[A] and Figure 2.5[A]), suggesting robust repeatability. Similar trends were found in the kidneys, where wCV decreased from 5% to 1% and in the spleen, where RC improved modestly (0.24 vs. 0.14).
In the case of the liver, the repeatability coefficients also fell below the threshold of 0.20×10^{-3} mm^2/s defined in the QIBA Profile for Diffusion-Weighted Imaging[46], supporting the reliability of the measurements according to established quantitative imaging standards.

In the pancreas, repeatability was comparatively lower: although the mean ADC remained stable across both analyses at 1.50±0.11 and 1.50±0.16 $\times 10^{-3}$ mm^2/s, RC increased from 0.35 to 0.46 and wCV from 10% to 11.0%. This reduced stability is likely attributable to the anatomical and functional heterogeneity of the pancreas, which is known to be affected by various physiological and technical factors, including Body Mass Index (BMI) and acquisition parameters[51].
In this study, a single ROI encompassing the entire pancreatic parenchyma was used, without distinguishing between head, body and tail. This approach may have introduced additional anatomical heterogeneity compared to studies that analyse pancreatic subregions separately[51] and could partly account for the slightly lower repeatability observed. However, the measured mean is consistent with previously reported

reference values in healthy subjects[52], supporting the plausibility of the findings despite the observed intra-subject variability.
Although no established reference values for RC and wCV in the pancreas are available for direct comparison, assuming a wCV below 15% as an acceptable threshold for stability, the repeatability observed in this study can still be considered satisfactory.

In osseous tissues, repeatability was significantly lower, especially in the femoral head, where RC and wCV respectively reached 0.34 and 53% in the intra-subject design. Although the limits of agreement were not particularly wide, they appeared asymmetric (Figure 2.3[B] and Figure 2.5[B]), with a more pronounced lower bound. This is likely the result of low signal-to-noise ratio in the femoral head region, which increases sensitivity to outlier effects and amplifies the impact of measurement noise on individual acquisitions.
In the pelvic bone and vertebral bodies — analysed in this study without distinction between thoracic or lumbar levels - repeatability was also limited. RC ranged from 0.14 to 0.30×10^{-3} mm^2/s and wCV from 13% to 29%, depending on the region and analysis. These findings are consistent with those reported by Michoux et al.[53], who observed RC values ranging from 0.23-0.44 $\times10^{-3}$ mm^2/s for the L5 vertebra, 22-46 $\times10^{-3}$ mm^2/s for the iliac crest and 22-55 $\times10^{-3}$ mm^2/s for the femur in two separate test-retest settings.
The close agreement in absolute values supports the robustness of the present data, despite the inherently reduced reproducibility observed in bone, which is documented in the context of bone diffusion imaging[53, 46].

	Across the Sample		**Intra-Subject**		**Variation**	
Organ	**RC**	**wCV (%)**	**RC**	**wCV (%)**	**ΔRC**	**ΔwCV (%)**
Liver	0.17	7	0.14	4	-0.03	-3
Spleen	0.24	12	0.14	6	-0.10	-6
Pancreas	0.35	10	0.46	11	+0.11	+1
Kidney	0.14	5	0.08	1	-0.06	-4
Vertebra	0.20	21	0.30	29	+0.10	+8
Pelvic bone	0.15	19	0.14	13	-0.01	-6
Femoral head	0.24	38	0.34	53	+0.10	+15

Table 2.4: Comparison of RC (expressed in 10^{-3} mm^2/s) and wCV between ADC Repeatability Across the Sample and Intra-Subject ADC Repeatability, with the corresponding variation between analyses.

The relevant results for the two repeatability analyses are reported in Table 2.4, directly comparing the ADC Repeatability Analysis Across the Sample and the Intra-Subject one. This table highlights how the restriction of the sample — corresponding to a reduced biological heterogeneity — is associated with improved repeatability metrics, as reflected by lower RC and wCV values in several anatomical regions. This behavior confirms the expected trend that measurement stability increases when variability due to inter-subject differences is minimised.

In summary, repeatability of ADC measurements was satisfactory in parenchymal tissues and improved when biological variability was controlled. In osseous tissues, the observed variability was larger but consistent with existing literature. These results confirm that ADC can serve as a repeatable quantitative imaging biomarker under clinical conditions, particularly in soft-tissue organs where technical and biological variability are more easily controlled.
Since measurement precision is a prerequisite for the interpretation of quantitative values, repeatability was assessed first to ensure the stability of the method. The analysis of average ADC values and their physiological interpretation across organs is presented as part of the organ-level analysis that follows, in a more general characterisation of healthy tissues.

2.4.2 Physiological Characterisation of ADC Values in Healthy Tissues

The intra-organ analysis presented in Organ ADC Variability and Feature Stability explored the physiological distribution of ADC values across different anatomical regions in the healthy population. Unlike the repeatability-focused analyses, this section included all 41 available WB-MRI acquisitions in the dataset, including multiple sessions and subjects. The objective was to characterise the quantitative profile of normal tissues using a more general and exploratory approach, not constrained by test–retest design, but rather intended to reflect the variability and distribution of ADC values under routine clinical conditions.

To evaluate the plausibility of the measured values, average ADCs for each organ were compared with those reported by Michoux et al.[53], who provided reference values with associated 95% confidence intervals in a smaller cohort of eight healthy volunteers. While their protocol was optimised for reproducibility — using three b-values (0, 150, 1000 s/mm^2), radiologist-defined 2D ROIs on identical slices and repeated acquisitions under highly standardised conditions — the present study relied on a more diverse sample acquired in a clinical setting, using two b-values (0 and 800 s/mm^2) and broader ROI placement across entire organs. These differences are expected to influence absolute ADC values, particularly in perfused tissues, and justify small deviations from the literature benchmarks.

Average ADCs and 95% confidence intervals (CI_{95}) were compared with those reported by Michoux et al.[53]. In the parenchymal organs, the measured values were generally compatible with published references. In the liver, the mean ADC was $(1.09 \pm 0.09) \times 10^{-3}$ mm^2/s, corresponding to a CI$_{95}$ of [1.06; 1.12], which lies below the reference interval of 1.20 [1.16; 1.23]. In the pancreas, the measured value was $(1.50 \pm 0.18) \times 10^{-3}$ mm^2/s, with a CI$_{95}$ of [1.44; 1.56], near to the reference of $(1.40 \pm 0.19) \times 10^{-3}$ mm^2/s reported in [52].

The spleen and kidneys presented borderline compatibility. The measured ADC in the spleen was $(0.85 \pm 0.10) \times 10^{-3}$ mm^2/s, CI$_{95}$ [0.82; 0.88], compared with a reference of 0.89 [0.86; 0.92]; for the kidney, the observed value was $(1.90 \pm 0.10) \times 10^{-3}$ mm^2/s, CI$_{95}$ [1.87; 1.93], compared with 1.95 [1.93; 1.97] for the renal cortex and 1.77 [1.74; 1.80] for the renal medulla.

In osseous tissues, the measured ADC values showed greater deviation from published references. For the vertebrae the mean ADC was $(0.35 \pm 0.09) \times 10^{-3}$ mm^2/s, with a 95% confidence interval of [0.32; 0.38], compared to the reference value of 0.29 [0.27; 0.31] reported for the L5 vertebra. In the pelvic bone (ilium), the measured value was $(0.37 \pm 0.09) \times 10^{-3}$ mm^2/s, CI$_{95}$ [0.34; 0.40], versus a reference of 0.30 [0.29; 0.32]. For the femoral head, the mean ADC was $(0.24 \pm 0.11) \times 10^{-3}$ mm^2/s , CI$_{95}$ [0.21; 0.27], compared with the reference 0.29 [0.28; 0.30].

These small discrepancies were expected given the differences in acquisition parameters and ROI placement strategies.

Welch's t-test was performed to assess the statistical significance of the observed differences between the present measurements and the reference values from the literature. This test evaluates the null hypothesis that the means of two independent samples are equal, without assuming equal variances between groups. A small p-value (typically below 0.05) indicates that the difference in means is unlikely to have occurred by chance, suggesting a statistically significant deviation from the reference values.

The resulting p-values (Table 2.5) were as follows: 1.41×10^{-5} for the liver, 7.42×10^{-2} for the spleen, 1.01×10^{-2} for the kidney, 2.57×10^{-2} for the pancreas, 1.09×10^{-3} for the vertebra, 6.08×10^{-5} for the pelvic bone and 8.34×10^{-3} for the femoral head.

All comparisons yielded statistically significant differences ($p < 0.05$) except for the spleen that exhibited statistical and physiological compatibility with the reference and the pancreas and the kidneys that showed moderate but statistically significant deviations. The spleen's compatibility is similarly supported by its well-known physiological stability and low ADC variability, which contribute to the consistent values observed across studies[54].

Organ	**Measured ADC [CI_{95}]**	**Reference Value [CI_{95}]**	**Compatibility**
Liver	1.09 [1.06; 1.12]	1.20 [1.16; 1.23]	✗
Spleen	0.85 [0.82; 0.88]	0.89 [0.86; 0.92]	✓
Pancreas	1.50 [1.44; 1.56]	1.40 [1.33; 1.47]	✗
Kidney	1.90 [1.87; 1.93]	1.95 [1.93; 1.97]	✗
Vertebra	0.35 [0.32; 0.38]	0.29 [0.27; 0.31]	✗
Pelvic bone	0.37 [0.34; 0.40]	0.30 [0.29; 0.32]	✗
Femoral head	0.24 [0.21; 0.27]	0.29 [0.28; 0.30]	✗

Table 2.5: Comparison between measured ADC values, literature reference values[53, 52] and assessment of their compatibility. All values are expressed in 10^{-3} *mm*^{2}*/s.*

Although the comparisons for the remaining tissues yielded statistically significant differences, the absolute deviations between the measured and reference ADC values were consistently small, typically in the range of $10^{-2}\times10^{-3}$ mm^2/s. As discussed above in this section, the slight discrepancies with respect to the reference values can be reasonably attributed to protocol differences and the limited sample size of the study by Michoux et al.

The standard deviations associated with the mean values reported in this study were generally on the order of 10^{-4} mm^2/s, i.e., one order of magnitude lower in most of the cases. This relative precision suggests that the observed discrepancies are minor in physiological terms and likely to be clinically acceptable. Overall, these results reinforce the plausibility of the measured ADC values and their reliability for tissue characterisation, particularly when considering the expected inter-subject variability and differences in acquisition protocols across studies, considering also the limitations in bone tissues.

Analysing the CoV and the IQR statistical measures across all tissues, as summarised in Table 2.6, clear differences emerged between parenchymal and osseous regions.

Parenchymal organs showed lower CoV values — ranging from 5% in the kidney to 12% in the spleen — while osseous tissues exhibited substantially higher CoVs, exceeding 24% in the vertebra and pelvic bone and reaching 46% in the femoral head. The IQR values were more homogeneous across tissues, ranging from 0.11 to 0.27 $\times10^{-3}$ mm^2/s. The pancreas exhibited the highest IQR (0.27), consistent with its greater anatomical and functional variability, as shown by its broader distribution in Figure 2.6. These trends indicate limited relative variability in ADC values for homogeneous tissues and reflect, in the case of bone, the increased dispersion likely caused by the inherent heterogeneity of bone structure, lower signal-to-noise ratio and susceptibility to imaging artefacts.

Organ	IQR	CoV (%)
Liver	0.13	8
Spleen	0.16	12
Pancreas	0.27	12
Kidney	0.13	5
Vertebra	0.11	26
Pelvic bone	0.14	24
Femoral head	0.15	46

Table 2.6: Summary of Organ ADC Variability metrics: IQR (expressed in 10^{-3} *mm*^{2}*/s) and CoV.*

Although ADC-based variability was more pronounced in osseous tissues, extending the analysis to a broader set of radiomic features revealed several parameters with low dispersion even in these anatomically heterogeneous regions. This was observed consistently across all organs. The pelvic bone is shown in Figure 2.8 as a representative example, given its relevance for the predictive modelling of pathological cases discussed in Clinical Application of ADC and FF in Multiple Myeloma. The matrix highlights a number of features with limited variability (i.e. $0 \leq \text{Cov} \leq 15\,\%$) in healthy subjects, including metrics from texture-based families such as GLCM, GLDZM, GLRLM and GLSZM, as well as intensity-based descriptors (for example Root Mean Square Intensity, Intensity Histogram Entropy) and IVH-derived parameters.

While the primary quantitative biomarker remains the mean ADC value ([ID] Mean Intensity of ADC maps), the observed stability of several additional radiomic features suggests that they may serve as complementary biomarkers in classification tasks, even within structurally complex regions such as the pelvic skeleton.

3 Phantom Validation of Fat Fraction Sequences

This chapter presents a quantitative validation of the sequences dedicated to Fat Fraction (FF) mapping, implemented on the Philips Ingenia 3T MRI scanner - Fat Fraction from mDixonAll and FatQuant from mDixonQuant.

An in-house built phantom, containing defined mixtures of water and lipid emulsions, was specifically constructed and subsequently imaged using the two available Dixon-based sequences.
In order to establish FF as a reliable quantitative imaging biomarker, the linearity, accuracy and calibration behaviour of each sequence were evaluated across a wide range of fat concentrations. By characterising the performance of the two sequences, this analysis aims to understand the limitations and diagnostic capabilities of each acquisition strategy within clinical and research contexts.

3.1 Phantom Preparation

The preparation of the fat–water phantom was based on the protocol developed by Bush et al. for the construction of MRI-calibrated fat fraction phantoms using water–lipid emulsions[55]. This method provides a flexible and cost-effective approach to create a stable phantom using standard laboratory equipment and easily accessible reagents. It enables the preparation of emulsions spanning a wide range of fat percentages — from 0% to 100% — facilitating direct comparison between measured and known fat content values for sequence validation purposes.

Following the published protocol, two separate base solutions were first prepared: a water-based solution and a fat-based solution.

The water solution consisted of 300mL of distilled water, 9.0 g of agar, 0.3 g of sodium benzoate and 0.6 mL of a water-soluble surfactant (Tween 20). In order to modulate the relaxation properties of the acqueous solution, improving its similarity to biological tissues, 0.24 mL of gadolinium-DTPA contrast agent were added.
The mixture was heated to 90°C under constant stirring (100 rpm) and then brought to 350°C until the agar was completely melted, forming a uniform amber solution.

The fat solution was prepared by heating 300 mL of peanut oil to 90°C under constant stirring. Peanut oil was chosen as the lipid component due to its nuclear magnetic resonance spectrum, which exhibits spectral characteristics comparable to those of human

adipose tissue triglycerides.
To facilitate emulsification with the aqueous phase, the oil-soluble surfactant (soy lecithin) was added at a concentration of 1.0% of the oil volume, corresponding to 3 mL.
The mixture was stirred at 1100 rpm for approximately 5 minutes to ensure complete mixing, achieving a soluble and homogeneous solution.

Figure 3.1: Preparation of the fat–water emulsion.

Emulsion preparation was carried out by combining the pre-heated water and oil solutions in the appropriate proportions using calibrated volumetric pipettes. The volumes were selected based on the target fat fraction of each sample.
Throughout the process, the water solution was maintained on a stirring hotplate to prevent premature gelation and ensure homogeneous mixing. The oil solution was slowly added and, as the emulsification progressed, the initially visible separation between the two phases gradually disappeared. Once the mixture became fully opaque and exhibited a creamy white consistency, the emulsion was immediately transferred into a 50 mL Falcon tube and left to cool and solidify at room temperature.
A representative image of the emulsification process is shown in Figure 3.1, illustrating the early phase of mixing before the formation of a homogeneous, stable emulsion.

This process was repeated for each target fat concentration, adjusting the amount of soy lecithin in the oil phase as needed. In general, higher fat content required increased surfactant concentrations, due to the greater difficulty in achieving a stable dispersion of lipid droplets within the aqueous matrix.
The optimal lecithin concentration for each sample was determined empirically to ensure that, once solidified, the emulsion exhibited no significant local inhomogeneities, visible bubbles, phase stratification or fat–water separation.

Each prepared tube, corresponding to a specific target fat fraction, was labelled and securely placed inside a custom-built phantom designed to mimic realistic anatomical conditions. The phantom consisted of a central water-filled cavity in which the sample tubes were arranged in fixed positions, two lateral compartments filled with polystyrene beads to simulate lung tissue and a separate compartment filled with pure peanut oil to serve as a 100% fat reference. The water compartment also served as a 0% fat reference and helped reduce susceptibility artefacts at air–interface boundaries during MRI acquisition.

The fat-water phantom is shown in Figure 3.2 in its final assembled configuration.
To ensure thermal equilibrium and minimise temperature-induced variability in MR signal, the phantom was placed in the MRI suite one day prior to image acquisition.

Figure 3.2: Final configuration of the fat-water phantom.

The composition of all phantom regions is summarised in Table 3.1, which reports, for each sample, the name, the nominal percentage of oil used in the mixture, the concentration of soy lecithin added to the oil solution and the resulting estimated fat fraction (FF_{true}).
The FF_{true} values were calculated as the proportion of fat within the total emulsion, taking into account both the oil and the surfactant volumes in the lipid phase. In other words, the true fat content corresponds to the fraction of oil relative to the combined volume of oil and lecithin, as this mixture constitutes the effective fat-containing component introduced into the emulsion.

Sample Name	Oil Solution (%)	Lecithin in Oil (%)	FF_{true} (%)
Water	0.0	0.0	0.0
1	15.0	8.0	13.8
2	15.0	5.0	14.3
3	25.0	1.0	25.0
4	44.0	1.0	44.0
5	50.0	1.0	50.0
6	55.0	8.0	50.6
7	55.0	5.0	52.3
8	60.0	8.0	55.2
9	65.0	1.0	65.0
Oil	100.0	0.0	100.0

Table 3.1: Composition of the fat–water phantom regions.

3.2 MRI Acquisition

The phantom was positioned on the scanner table, covered with the Q-Body coil and carefully centred within the magnet using alignment lasers and the table positioning system.

Once the correct placement was ensured, both fat fraction quantification sequences — Fat Fraction from mDixonAll and FatQuant from mDixonQuant — were acquired five times under identical conditions to support subsequent validation.

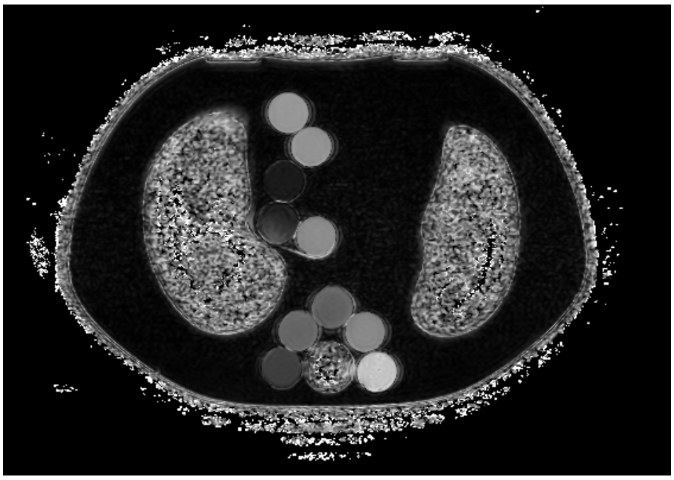


Figure 3.3: Axial image of the fat–water phantom acquired with the mDixonAll-based Fat Fraction (FF) sequence.

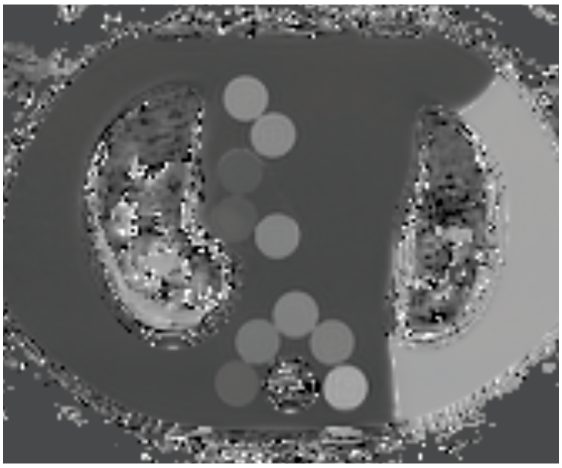


Figure 3.4: Axial image of the fat–water phantom acquired with the mDixonQuant-based FatQuant sequence.

Figure 3.3 and Figure 3.4 show representative images of the phantom acquired with the mDixonAll-based FF sequence and the mDixonQuant-based FatQuant sequence, respectively.

3.2.1 Comparison of Acquisition Parameters

The main acquisition parameters for both the acquired sequences - Fat Fraction (FF) from mDixonAll and FatQuant from mDixonQuant - are summarised in Table 3.2.

Acquisition Parameter	**FF (mDixonAll)**	**FatQuant (mDixonQuant)**
Acquisition Duration [s]	13.81	14.09
Pulse Sequence Name	T1FFE	T1FFE
Echo Pulse Sequence	GR	GR
Acquisition Plane	Axial	Axial
Slice Thickness [mm]	3	6
FOV [mm]/Acquisition Matrix	480/236 × 300	400/140 × 160
Pixel Spacing [mm]	0.8 × 0.8	2.1 × 2.1
Phase Encoding Direction	RL	RL
SENSE Acceleration	Yes	Yes
TR/TE [ms]	3.66/shortest	6.27/shortest
Echo Train Length	2	6
Echo Numbers	1	1
Flip Angle [°]	10	3
Pixel Bandwidth [Hz/pixel]	1344	2298
Fat Quantification Model	2-point Dixon	Multi-peak Dixon

Table 3.2: Acquisition parameters for the two sequences dedicated to FF mapping.

The differences in acquisition parameters between the two sequences are clearly reflected in the visual appearance of the resulting images.
In particular, the larger acquisition matrix used in the FF (mDixonAll) sequence (236 × 300 vs 140 × 160 in FatQuant) leads to a higher in-plane spatial resolution and a wider field of view, as already evident from the phantom images shown in Figure 3.3 and Figure 3.4. The reduced slice thickness in FF (3 mm vs 6 mm) also enhance the anatomical definition in coronal and sagittal reconstructions, owing to the increased through-plane resolution.

Beyond spatial resolution, several additional parameters are expected to influence the sequences' behaviour in fat quantification, potentially leading to differences in the reliability and precision of the resulting fat fraction estimates.
The echo train length is considerably shorter in FF (2 vs 6), reflecting the use of a two-point Dixon model as opposed to the multi-echo, multi-peak spectral model employed by FatQuant. The latter enables more comprehensive modelling of fat and more robust correction for confounding factors such as B_0 inhomogeneities and T_2^* decay.
Furthermore, the flip angle differs substantially (10° for FF vs 3° for FatQuant), which may lead to greater T1-related bias in FF, particularly in voxels containing both fat and water. FatQuant, by contrast, is designed to minimise such effects and is therefore expected to provide more accurate and reproducible fat fraction estimates, despite its lower spatial resolution.

In summary, while the FF (mDixonAll) sequence offers superior image sharpness, FatQuant (mDixonQuant) is expected to provide improved quantitative reliability, particularly in challenging regions or at intermediate fat concentrations.

3.3 Quantitative Results and Calibration

In order to quantitatively assess the performance of the two fat fraction sequences, numerical measurements were obtained from the phantom acquisitions and compared against the known fat concentrations used during preparation. This evaluation involved extracting fat fraction values from the parametric maps generated by each sequence, constructing calibration curves and examining their agreement with the reference values. The analysis aimed to assess the accuracy, linearity and consistency of both methods across a wide range of fat content.

3.3.1 Measured Fat Fraction Values

For each fat concentration present in the phantom, a cylindrical region of interest (ROI) with a target volume of 2.6 cm^3 was manually placed in the centre of the corresponding sample region, in accordance with the nominal compositions reported in Table 3.1. The

ROI volume was selected to ensure adequate voxel sampling while avoiding boundary effects or partial volume contamination.

Although 10 tubes are visible in the phantom images, the analysis was limited to 11 well-defined fat–water regions, including the pure water and pure oil references.
One additional tube — corresponding to a nominal fat content of 85.0% with 8.0% lecithin — was included in the phantom but excluded from the quantitative analysis. Due to its high fat concentration, the sample proved difficult to prepare with sufficient stability and the emulsion appeared visibly inhomogeneous upon inspection of the acquired images. For this reason, it was excluded from the analysis and was not included in the summary table of phantom composition and in the measured fat fraction results.

For each of the five repeated acquisitions performed per sequence, ROIs were positioned consistently using the same spatial coordinates and slice index, in order to minimise variability due to repositioning or reader-dependent factors.
Fat fraction values were extracted from each ROI through the MIM® software and then averaged across the five repetitions.
The resulting mean values, along with the standard deviation of the mean, are reported in Table 3.3 for both the FF (mDixonAll) and FatQuant (mDixonQuant) sequences. All values are expressed as absolute fat fraction percentages (%) and have been rounded to one decimal place in order to preserve measurable differences between closely similar values.

Sample Name	**FF_{true} (%)**	**FF (%)**	**σ_{FF} (%)**	**FatQuant (%)**	**$\sigma_{FatQuant}$ (%)**
Water	0.0	3.6	2.6	0.7	0.5
1	13.8	15.0	0.0	16.1	1.8
2	14.3	10.8	0.8	15.6	0.6
3	25.0	26.6	1.3	22.9	1.4
4	44.0	43.6	1.5	44.4	0.7
5	50.0	60.8	1.1	52.5	0.6
6	50.6	64.0	1.0	58.8	0.4
7	52.3	42.0	1.2	57.5	1.0
8	55.2	66.4	0.6	64.6	0.4
9	65.0	64.4	4.2	65.3	1.4
Oil	100.0	96.6	0.6	98.4	0.6

Table 3.3: True fat fraction values (FF_{true}) and mean fat fraction measurements with standard deviations from five repeated acquisitions using FF (mDixonAll) and FatQuant (mDixonQuant).

3.3.2 Calibration Curve

To quantitatively assess the calibration performance of the two sequences, a linear regression analysis was conducted by fitting the measured fat fraction values against the known reference concentrations ($\mathrm{FF}_{\mathrm{true}}$).
The analysis was based on the mean values obtained from the five repeated acquisitions for both FF (mDixonAll) and FatQuant (mDixonQuant), as reported in Table 3.3. The corresponding standard deviations were used to define horizontal error bars in the plot, reflecting the variability in the measured values.

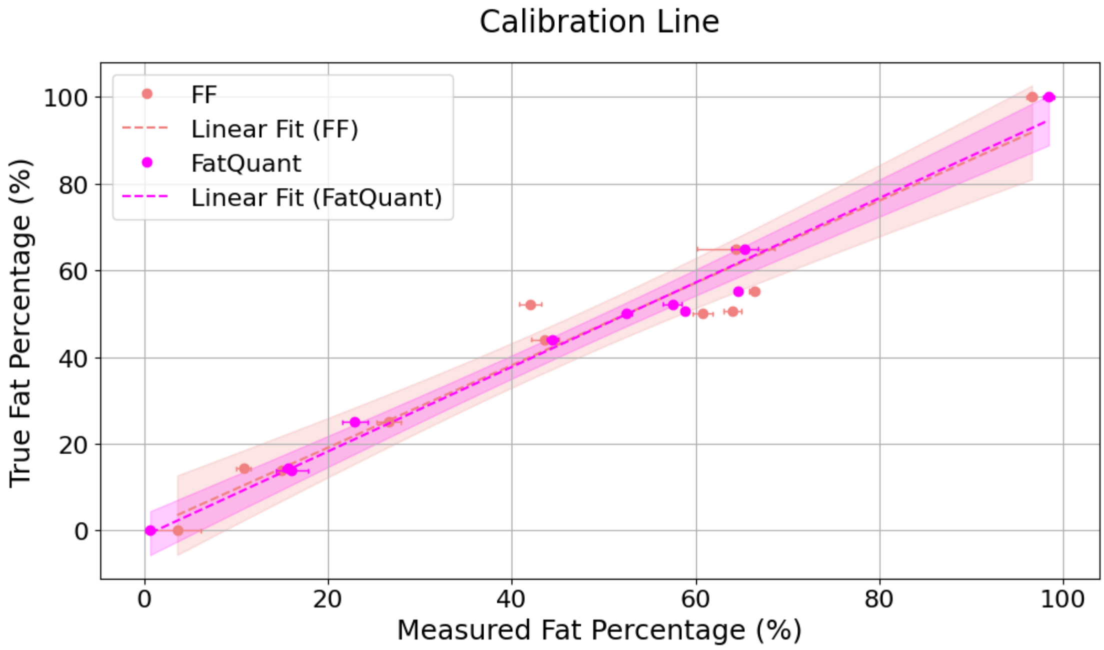


Figure 3.5: Calibration curves for FF (mDixonAll) and FatQuant (mDixonQuant) with linear regression lines, 95% confidence bands and standard deviation error bars.

In the resulting calibration curves, shown in Figure 3.5, the reference values ($\mathrm{FF}_{\mathrm{true}}$) are represented on the y-axis, while the measured fat fractions are plotted on the x-axis. Each point thus indicates how well a given measurement aligns with its true target value.
The fitted regression lines for both sequences are shown, together with 95% confidence bands that represent the uncertainty associated with the estimated regression function. These confidence bands were calculated analytically using the standard error of the fit and the Student's t-distribution. They define the interval within which the true regression line is expected to lie with 95% probability, taking into account both the residual variance and the spread of the data around the predictor mean.

The regression lines were estimated using Ordinary Least Squares (OLS), according to the general model

$$y_i = \beta_0 + \beta_1 \cdot x_i + \epsilon_i \tag{3.1}$$

where y_i is the reference value ($\mathrm{FF}_{\mathrm{true}}$), x_i is the measurand value from the sequence under analysis, β_0 is the intercept, β_1 is the slope and ϵ_i are the residual errors.
For FF, the fitted model was

$$\mathrm{FF}_{\mathrm{true}} = 0.11 + 0.95 \cdot \mathrm{FF}_{\mathrm{measured}} \quad (R^2 = 0.94) \tag{3.2}$$

while for FatQuant the best-fit equation was

$$\mathrm{FF}_{\mathrm{true}} = -1.27 + 0.97 \cdot \mathrm{FatQuant}_{\mathrm{measured}} \quad (R^2 = 0.98). \tag{3.3}$$

Statistical tests confirmed the significance of both models and the strength of the linear relationships, already supported by the slope estimates in Equation 3.2 and Equation 3.3 (0.95 for FF and 0.97 for FatQuant). The corresponding coefficients of determination (R^2) were 0.94 for FF and 0.98 for FatQuant, indicating that 94% and 98% of the variance in the reference values is explained by the respective regression models. Such high R^2 values reflect an excellent fit of the linear calibration within the tested range of fat fractions.
The significance of the slope coefficients was evaluated using a two-sided t-test for the null hypothesis $H_0 : \beta_1 = 0$, which assesses whether the slope is significantly different from zero. Both slopes were found to be highly significant ($p = 1.02 \times 10^{-6}$ for FF and $p = 2.78 \times 10^{-9}$ for FatQuant), confirming a strong association between measured and reference values. The corresponding model F-statistics (134.93 for FF and 522.82 for FatQuant) assess the overall regression significance compared to an intercept-only model. In simple linear regression with one predictor, the F-test is mathematically equivalent to the t-test of the slope ($F = t^2$), which explains why the F-tests yield identical p-values to those of the slope coefficients.
Model residuals, defined as

$$\epsilon_i = \mathrm{FF}_{\mathrm{true},i} - \widehat{\mathrm{FF}}_{\mathrm{true},i} \tag{3.4}$$

were tested for normality using the Shapiro-Wilk test, which evaluates the null hypothesis that the residuals follow a normal distribution, a necessary assumption for the validity of the statistical tests reported above. A p-value greater than the conventional 0.05 threshold indicates no evidence to reject normality. In this analysis, the p-values ($p = 0.83$ for FF and $p = 0.71$ for FatQuant) confirm that the residuals are consistent with a normal distribution, supporting the validity of standard inference for the regression coefficients.

To formally assess whether the calibration behaviour differed between the two sequences, a multiple linear regression was fitted including the measured fat fraction (x), the sequence type (FF = 0, FatQuant = 1) and their interaction term ($x \times group$). This specification allows both the intercept and the slope to vary depending on the sequence.
This model achieved a coefficient of determination $R^2 = 0.96$ and an overall F-statistic of 145.00 ($p = 8.53 \times 10^{-13}$), confirming that the measured fat fractions strongly explain the reference values. However, neither the interaction term ($p = 0.79$) nor the group-specific intercept difference ($p = 0.78$) reached statistical significance, indicating that neither the slope nor the intercept differ between the two sequences. These results suggest that the linear calibration trends of FF and FatQuant are statistically comparable.

3.4 Evaluation of Sequences Performance

This section provides a critical evaluation of the performance of the two fat fraction sequences implemented in the MRI scanner used in this study, in order to clarify their differences and justify their potential use in fat quantification studies.
The phantom experiment was specifically designed to assess whether both sequences — FF (mDixonAll) and FatQuant (mDixonQuant) — can be considered suitable for quantitative fat fraction estimation. This issue is particularly relevant in the context of retrospective studies, where FatQuant may not be available due to its introduction in more recent imaging protocols. Establishing whether these sequences provide comparable results is therefore essential to ensure methodological consistency and interpretability when both are used across different datasets.

From a methodological point of view, the two sequences are based on different Dixon implementations. FF uses a two-point Dixon model, which estimates fat and water content from two echo times, without explicit correction for confounding factors such as B_0 inhomogeneities or T_2^* decay. In contrast, FatQuant employs a multi-echo acquisition and incorporates a multi-peak spectral model of fat, along with corrections for B_0 field variation and T_2^* relaxation. These features are designed to improve the accuracy and reliability of fat quantification[19].
On a qualitative level, as observed in the acquired phantom images (Figure 3.3 and Figure 3.4), FF generally produces sharper and more stable fat–water separation, making it more suitable for anatomical visualisation, whereas FatQuant is better suited for quantitative applications due to its advanced signal modelling.

It is important to note that FatQuant images exhibited occasional fat-water swaps, which are signal inversions where voxels expected to be water-dominant are instead reconstructed as fat-dominant and vice versa. These artefacts were not observed in the FF sequence and are typically caused by ambiguities in the phase estimation used to solve the multi-echo Dixon model. The risk of swaps increases in regions affected by low signal-to-noise ratio (SNR), susceptibility effects or steep B_0 field gradients, all of which can destabilise the field map estimation and phase unwrapping process used in multi-echo reconstructions.

In the phantom used in this study, swaps were primarily observed in the most water-rich region (i.e. the water cavity where $\text{FF}_{\text{true}} = 0\%$), where the fat signal is virtually absent. In such conditions, even small phase inconsistencies can dominate the decomposition process and lead to misclassification.

Several strategies have been proposed in the literature to address these artefacts[15, 56, 57, 58] but in the context of this study they were not implemented since fat-water swaps did not affect the validity of the quantitative results. In the phantom, where ground-truth fat concentrations were known, artifactual regions were visually identifiable and were excluded from the analysis. Similarly, in the patient data which will be presented in Quantitative FF Analysis in Healthy Tissues, all images were manually inspected and voxels affected by swaps were discarded prior to quantification, ensuring the integrity of the quantitative analysis by retaining only high-confidence values.

Both sequences demonstrated an instability region around $\text{FF}_{\text{true}} \approx 50 - 55\%$, where fat and water signals are approximately equal in amplitude. In this regimen, the Dixon model becomes more sensitive to noise and small phase inconsistencies, as the phase difference between fat and water approaches critical values that can make the system ill-conditioned[59]. This leads to greater variability and reduced precision in fat fraction estimation in regions where fat and water coexist in similar proportions. This effect is intrinsic to the Dixon reconstruction process and has been attributed to a reduction in contrast and sensitivity when fat and water signals are nearly equivalent, as observed in clinical studies[60].

This behaviour is further illustrated in Figure 3.6. The plot shows the residuals of FF and FatQuant measurements relative to the reference values, displayed as a function of the true fat concentration. Each point corresponds to a phantom ROI, with the dashed line indicating zero residual. The shaded band highlights the 50–55% range, where an increased spread of residuals is observed, visually supporting the reduced precision of both sequences in this instability region. It should be noted that, unlike the regression residuals defined above, here residuals are simply computed as the difference between the measured and the true fat fraction values, in order to directly visualise estimation errors across the full range of fat concentrations.

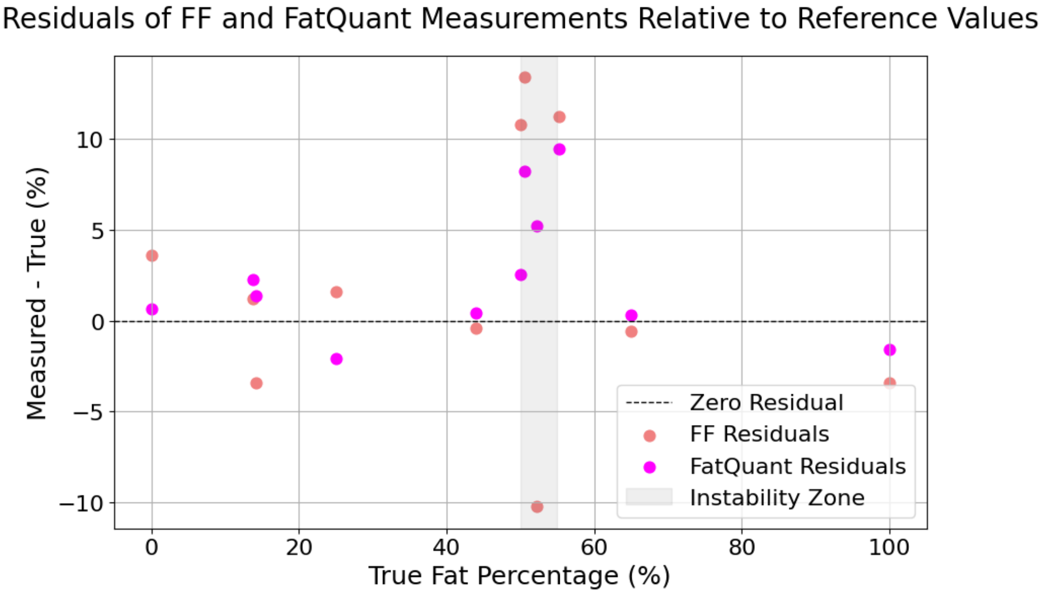


Figure 3.6: Residuals of FF and FatQuant relative to reference values, with the shaded area marking the 50–55% instability region.

Despite these limitations, the calibration results confirm that both sequences provide excellent linear agreement with the known fat concentrations. FatQuant yielded a regression slope of 0.97 with $R^2 = 0.98$, while FF produced a slope of 0.95 with
$R^2 = 0.94$.
The confidence bands associated with the FatQuant regression line were visibly narrower. Since both models were based on the same number of data points, the difference in band width is primarily attributable to the residual variance: for the FatQuant sequence, the estimation of the regression line was more precise as data deviated less from the fitted line. The smaller dispersion observed for FatQuant suggests that its fat fraction estimates were more consistent and less affected by random noise, resulting in a tighter and more stable linear relationship across the full range of fat concentrations.

However, a direct statistical comparison between the two calibration lines using a linear model with interaction showed no significant difference in either slope or intercept ($p = 0.79$ and $p = 0.78$, respectively), indicating that the quantitative behaviour of the two sequences is statistically comparable.
In other words, the data do not provide sufficient evidence to support a difference in the linear calibration trends between FF and FatQuant. Despite this does not imply that the two models are identical, it suggests that any observed differences may be attributed to random variability rather than systematic bias between the sequences.

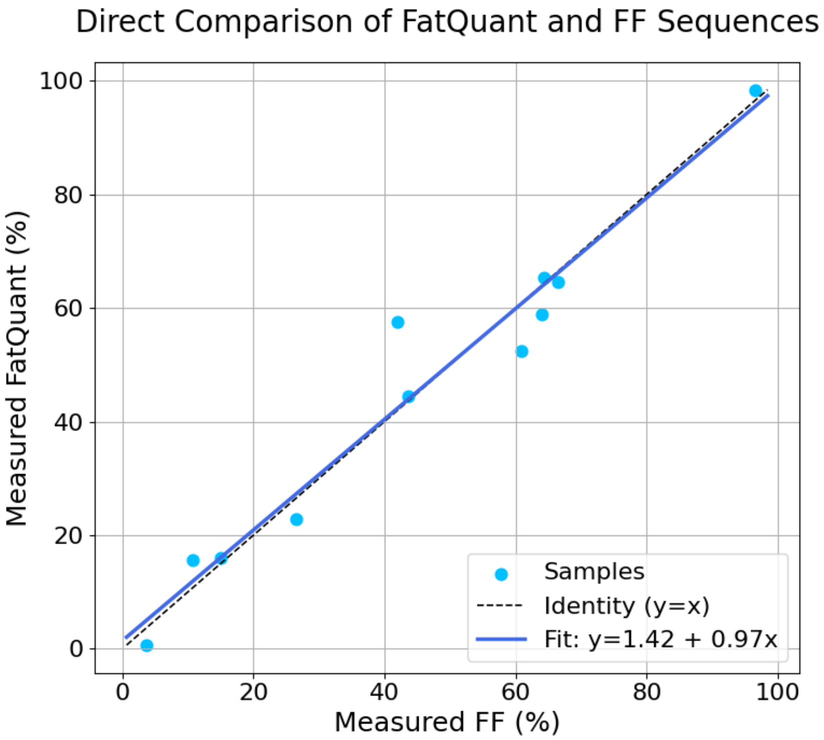


Figure 3.7: Direct comparison between FatQuant (y-axis) and FF (x-axis) measurements in the phantom to assess the consistency between the two sequences.

Figure 3.7 further illustrates this comparability through a direct method-to-method analysis. The scatter plot shows FatQuant measurements plotted against the corresponding FF measurements for each phantom ROI. The dashed line represents the identity ($y = x$), which corresponds to the condition of perfect agreement between the two methods, while the solid line indicates the fitted regression line. Points clustering close to the identity denote strong concordance between the sequences, whereas deviations from the line would reflect systematic or random differences. In the presented graph, the regression line is nearly overlapping with the identity, with a slope of 0.97 which is close to the expected value of 1. The dispersion of points is limited, confirming the absence of systematic bias and the overall consistency of the two sequences.

These findings support the use of both FF and FatQuant for fat quantification, provided that their respective limitations are understood and appropriately accounted for.
FatQuant should be preferred when available, due to its more advanced multi-echo and multi-peak spectral signal modelling, reduced susceptibility to B_0 and T_2^* bias and tighter confidence intervals.
FF can nevertheless be considered a valid and reliable alternative. Its simpler recon-

struction algorithm and better visual contrast make it particularly suitable for anatomical interpretation and for retrospective analyses, where FatQuant may not have been included in earlier imaging protocols.
The statistical comparability demonstrated in the phantom analysis provides empirical supports for the interchangeable use of FF and FatQuant sequences in studies constrained by protocol variations or limited data availability, contributing to greater methodological standardisation across heterogeneous datasets.

Finally, it is important to acknowledge the inherent limitations of the presented phantom-based experiment. The phantom was specifically designed to assess the comparability of FF and FatQuant sequences under controlled conditions, where the fat–water emulsions exhibit stable and chemically homogeneous properties.
Peanut oil was used as the fat component due to its spectral composition, which closely resembles that of human adipose tissue, improving the physiological relevance of the model. While this setup is well suited to test linearity and reproducibility, it does not replicate the biological complexity of living tissues, such as spatially heterogeneous fat distribution, microscopic susceptibility gradients and variability in relaxation parameters (T_1, T_2, and T_2^*). Potential non-linear effects arising from differences in tissue composition — such as those related to temperature variation, iron accumulation or pathological infiltration — cannot be reproduced in a synthetic phantom. Future validation studies involving in vivo acquisitions will therefore be essential to confirm the generalisability of these results and to investigate how biological factors may affect the quantitative performance of each sequence.

4 Quantitative FF Analysis in Healthy Tissues

The Fat Fraction (FF) is an emerging quantitative imaging biomarker used to characterise tissue composition and detect metabolic or infiltrative abnormalities[17, 61, 62, 63, 64]. In this study, FF values were exclusively derived from maps generated using the FatQuant sequence, whose accuracy and reliability were previously validated through phantom experiments (see Phantom Validation of Fat Fraction Sequences).

This chapter presents a quantitative analysis of FF in healthy tissues, following the same methodology and anatomical sampling strategy used for ADC analysis in Quantitative ADC Analysis in Healthy Tissues. A consistent workflow was adopted for image selection, ROIs placement and radiomic feature extraction. The statistical analysis followed the same structure and includes intra-subject repeatability across the sample, longitudinal intra-subject repeatability for a single patient and intra-organ variability and feature stability.
This framework provides a comprehensive characterisation of FF in normal tissues and establishes reference distributions to support future clinical applications and modelling based on FF-derived biomarkers.

4.1 Patients

The dataset analysed in this chapter corresponds to the same cohort of 20 healthy adult female subjects described in Quantitative ADC Analysis in Healthy Tissues, with a mean age of 57 ± 13 years and a broad age range (35–79 years). All subjects underwent WB-MRI acquisitions at IRST ”Dino Amadori” (Meldola, Italy), using a standardised WB-MRI protocol that includes both diffusion-weighted and fat quantification sequences. As a result, the FatQuant maps analysed in this chapter were acquired during the exact same imaging sessions as the ADC maps previously used in this work, ensuring full spatial and temporal correspondence between the two quantitative biomarkers.

The 20 subjects collectively contributed to a total of 41 WB-MRI examinations, including 14 patients with two scans each (at different time points), four patients with a single scan, one patient with three scans and one patient with six longitudinal acquisitions, with inter-scan intervals ranging from three to nine months.
All the acquisitions were characterised by high diagnostic quality and anatomical integrity, with complete absence of motion-related, acquisition or reconstruction artefacts. Each dataset was reviewed to confirm uniform signal behavior, correct fat–water decomposition and consistency across anatomical stations, ensuring full suitability for

quantitative analysis.

The WB-MRI protocol applied to the 20 subjects included both the standard fat fraction (FF) sequence — based on a two-point Dixon reconstruction — and the more advanced FatQuant sequence (Philips mDixonQuant), designed for accurate fat quantification using a multi-echo and multi-peak spectral model.
For the statistical analysis presented in this chapter, FF values were exclusively derived from FatQuant acquisitions, based on the findings reported in Phantom Validation of Fat Fraction Sequences. The phantom validation demonstrated that FatQuant provides superior accuracy, linearity and stability across a wide range of fat concentrations, while reducing susceptibility to signal bias and modelling limitations observed in simpler Dixon implementations.
Although some fat–water swaps were noted in phantom data, they occurred only in regions with extremely low fat content and were easily identifiable. In all patient scans, each FF map was individually reviewed and any visibly affected regions were systematically excluded from the analysis, ensuring artefact-free ROI placement.

Throughout this chapter, the acronym FF refers to the fat fraction biomarker as estimated from FatQuant images only. This choice reflects the quantitative nature of the sequence and aligns with standard nomenclature in the literature, where FF is used to indicate the biomarker regardless of the specific acquisition protocol employed.

4.2 Data Collection and Processing

The entire dataset underwent a standardised post-processing workflow identical to the one described for ADC analysis in Quantitative ADC Analysis in Healthy Tissues. Each FF map was used to define organ-specific regions of interest (ROIs), from which quantitative radiomic features were extracted using a consistent and reproducible protocol.

Except for the early outlier removal strategy — discussed in Outlier Removal — all steps were carried out following the same procedures applied to ADC maps, including ROI definition criteria, dimensional constraints and feature computation settings.

4.2.1 ROIs placement and Features Extraction

Organ-specific ROIs were manually placed on each FatQuant-derived FF map, following the same anatomical strategy described for the ADC analysis. ROI definition and visualisation were performed using the MIM® software[43] which enabled multi-planar inspection and accurate positioning across axial, sagittal and coronal views.

Three cylindrical ROIs were drawn for each of the seven target organs - liver, spleen,

pancreas, kidneys, vertebrae, pelvic bone and femoral head - in order to ensure consistent sampling and improve statistical robustness across repeated acquisitions.
The ROI standard volume was fixed at 2.6 cm^3 for all the organs except for the pancreas that required a smaller volume (0.5 cm^3) due to its limited anatomical size and complex geometry.

ROIs initially defined on the ADC maps were transferred to the corresponding FF maps by anatomical matching. Since the FF and DWI sequences were acquired in separate blocks of the whole-body protocol, slight spatial misalignments were observed — particularly in abdominal organs affected by respiratory motion. All ROIs were therefore manually realigned on the FF maps to ensure correct placement and tissue consistency. Each ROI was carefully positioned to sample homogeneous signal regions, free from fat-water separation artefacts, visible vessels and lesions that could bias the subsequent analysis.
The overall placement strategy is illustrated in Figure 4.1, which shows representative examples of ROI locations on FatQuant-derived FF maps across axial, sagittal and coronal planes. Only the ROIs intersecting the slice position defined by the reference cursor are visible in each view.

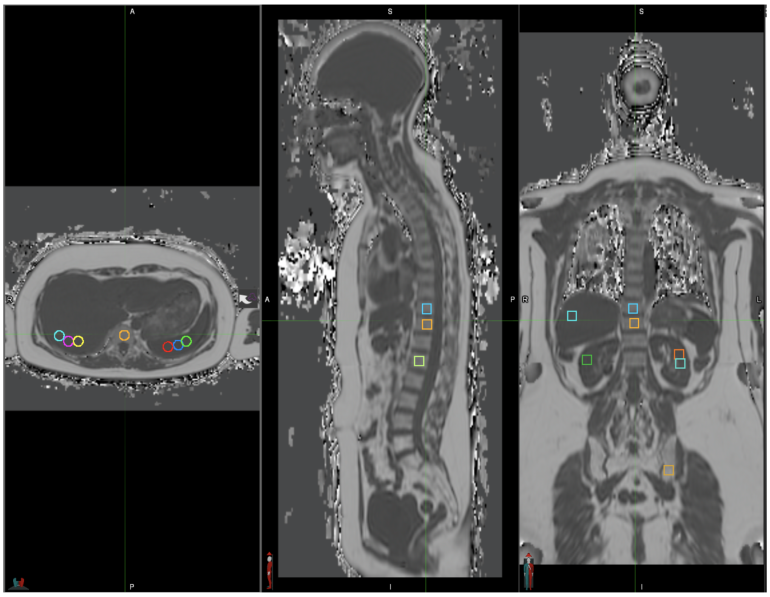


Figure 4.1: ROIs placement for FF analysis.

Image volumes and associated ROI structures (RTSTRUCT) were exported from MIM® and processed using the SIBEX (Standardised Imaging Biomarker Extraction) software, implemented in MATLAB® [44, 45]. Each ROI was analysed individually to extract a set of

144 quantitative radiomic features covering intensity, histogram, texture and structural categories as described in Appendix A.

4.2.2 Outlier Removal

Differently from the ADC analysis described in Quantitative ADC Analysis in Healthy Tissues, an early outlier removal step was necessary prior to any statistical evaluation of FF values.

The patient cohort used in this study included women with a history of breast cancer, spanning a wide age range and displaying high biological heterogeneity. While all ROIs were placed in homogeneous regions with no visible lesions or structural abnormalities, some subjects may have been affected by systemic or metabolic conditions — such as obesity, hepatic steatosis, thyroid dysfunctions or sarcopenia — which can alter tissue fat content without producing clear imaging findings.
These factors have little to no effect on ADC values, but may introduce large inter-subject variability in fat fraction measurements, especially in parenchymal organs such as the liver, spleen, pancreas or kidneys. For this reason, a pre-filtering step was applied to remove non-physiological values and ensure the robustness of subsequent statistical analyses.

Outliers were identified and excluded based on the interquartile range (IQR) method, as recommended by the Quantitative Imaging Biomarkers Alliance (QIBA) for statistical outlier detection in biomarker analysis[47]. Specifically, for each organ, ROI-level FF values falling below Q1 - 1.5 × IQR or above Q3 + 1.5 × IQR were flagged as outliers. A detailed description of the method is provided in Appendix B.
To better understand the origin of extreme FF values, clinical records were reviewed when available. Relevant information was retrieved from the institutional oncology information system MOSAIQ® (Elekta AB, Stockholm, Sweden) which includes patient histories, diagnostic reports and treatment documentation[65].

It is important to note that all patients included in this study were originally classified as “healthy” based on the normal appearance of their Whole-Body MRI scans. However, the outlier removal step was necessary to isolate cases in which clinical conditions — although not morphologically evident — might have influenced FF values at a biochemical or structural level.
This approach ensured that the statistical analysis was based on physiologically consistent data, minimising the risk of hidden biases related to subclinical alterations in fat metabolism, hormonal imbalances or localised anomalies such as cystic lesions.
Table 4.1 reports the cases in which ROIs were excluded, grouped by patient and organ, with the number of removed ROIs and clinical rationale when available. The table shows

that several outliers, although identified statistically, were also supported by relevant clinical context.

Patient	Organ	Removed ROIs	Clinical Justification
Patient 1	Femoral Head	1	Bone lesion (right pelvic bone)
Patient 3	Pancreas	1	Diabetes and hypothyroidism
Patient 3	Kidneys	4	Renal insufficiency
Patient 4	Liver	6	Hepatic cysts
Patient 6	Liver	3	Hepatic hemangioma
Patient 7	Spleen	4	Hypercholesterolemia
Patient 12	Spleen	3	Unknown
Patient 12	Pancreas	2	Unknown
Patient 17	Kidneys	3	Hypothyroidism
Patient 18	Kidneys	1	Total thyroidectomy

Table 4.1: Summary of removed outlier ROIs in FF analysis.

In addition to statistical outliers, another correction was applied prior to feature analysis: all negative FF values were replaced with zero.
Although fat fraction is expressed as a percentage and is by definition a non-negative quantity, small negative values can arise in magnitude-based Dixon reconstructions due to low signal-to-noise ratio (SNR), imperfect fat–water separation or model fitting limitations, especially in tissues with very low fat content such as spleen, pancreas or kidneys[66]. As these values are not physiologically meaningful, they were replaced with zero to ensure interpretability, consistently with the methods previously adopted in literature[67].

4.3 FF Statistical Analysis

The overall approach for the statistical analysis performed on FF values replicated the methodology previously adopted for ADC in Quantitative ADC Analysis in Healthy Tissues, following QIBA recommendations to evaluate repeatability and precision in quantitative imaging biomarkers[46, 47, 48]. These guidelines were used as a reference to choose the statistical metrics and the structure of the analyses.

As done for ADC, three complementary analyses were conducted. First, a repeatability analysis was performed across the whole cohort and, focusing on the mean intensity feature (i.e. the average FF in each ROI), it was used to evaluate intra-subject and inter-subject variabilities. An intra-subject longitudinal analysis was then performed to assess the temporal stability of FF measurements, considering a single subject who underwent six acquisitions. Finally, intra-organ variability and feature stability was evaluated using the full cleaned dataset.

4.3.1 FF Repeatability Across the Sample

The statistical analysis of repeatability for Fat Fraction (FF) values was conducted following the same methodology used for ADC analysis (see ADC Repeatability Across the Sample). This analysis estimates both intra-subject and inter-subject variability of the average FF across different acquisitions, considering only patients with at least two scans. In this way, it quantifies both the temporal repeatability of measurements and the biological heterogeneity within the healthy cohort.

For each subject, organ, and acquisition time point, FF values were averaged across the three corresponding ROIs — when available after outlier removal — to obtain a single representative value per observation. In cases where fewer than three ROIs remained, the available ones were used.

The reference mean FF values for each organ were then computed across the entire sample of subjects and acquisitions, including the subset of 16 patients with at least two acquisitions and using only cleaned data after the outlier removal procedure described in Outlier Removal. The graphical result is shown in Figure 4.2.

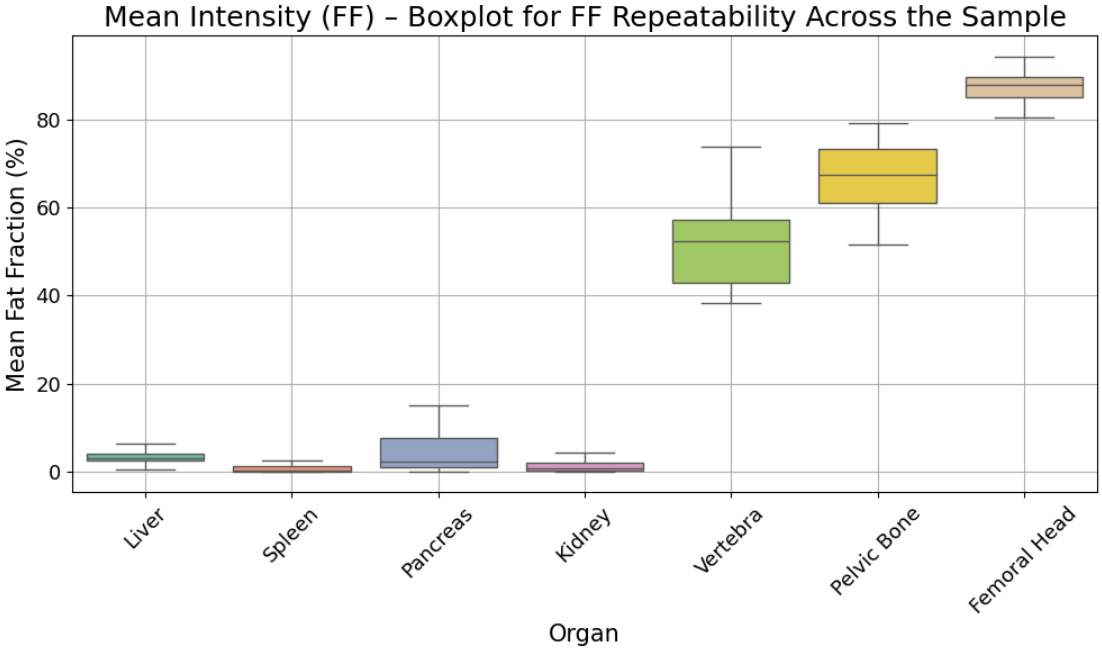


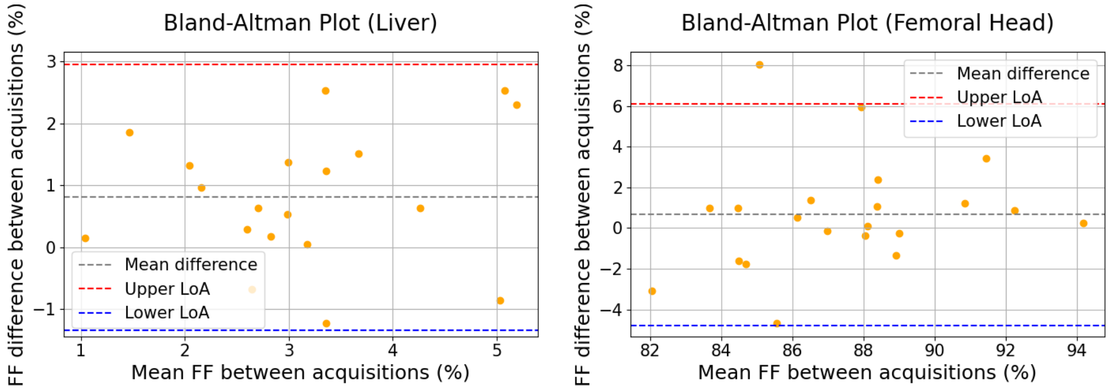
Figure 4.2: Boxplot of the mean intensity (FF value) across the different organs for the FF Repeatability Analysis Across the Sample.

The aggregated values were then analysed using a linear mixed-effects model (MixedLM), which allowed the variance to be split into intra-subject (σ_w^2) and inter-subject (σ_b^2) components. From this model, the key repeatability metrics indicated by QIBA guidelines[47, 48] were derived, providing a quantitative assessment of FF consistency over time and across subjects.

Organ	**Mean FF**	σ	σ_w	σ_b	**RC**	**wCV (%)**	**LOA$_L$**	**LOA$_U$**
Liver	3.19	1.46	0.96	1.11	2.65	44	-1.33	2.95
Spleen	0.67	0.80	0.60	0.52	1.67	117	-1.93	1.37
Pancreas	4.12	4.17	2.09	3.61	5.78	98	-5.17	6.47
Kidney	1.14	1.31	0.84	1.00	2.33	111	-2.21	1.92
Vertebra	51.77	9.23	4.02	8.31	11.14	18	-11.46	11.29
Pelvic bone	66.76	8.05	3.69	7.15	10.22	11	-10.12	10.18
Femoral head	87.46	3.67	2.14	2.99	5.93	4	-4.79	6.11

Table 4.2: Results for the FF Repeatability Analysis Across the Sample. All values are expressed as absolute percentages of fat fraction, except the ones explicitly indicated as percentages which are expressed as relative percentages.

Table 4.2 shows the obtained results and reports the mean FF value, the overall standard deviation (σ), the within-subject and between-subject standard deviations (σ_w and σ_b), the repeatability coefficient (RC), the within-subject coefficient of variation (wCV) and the lower and upper limits of agreement (LoA$_L$ and LoA$_U$).



[A] [B]

Figure 4.3: Bland-Altman plots for the FF Repeatability Analysis Across the Sample for the liver[A] and the femoral head[B].

To visually assess the agreement between repeated acquisitions, Bland–Altman plots

were generated for each organ. For every patient-organ combination, only chronologically successive acquisition pairs were considered. The plots were computed using the general form of the Bland–Altman method[49], which does not assume zero bias and allows for the detection of potential systematic differences between scans.
Two representative plots are shown in Figure 4.3: one for the liver, where FF values are typically low and subject to biological variability, and one for the femoral head, where FF is high.

4.3.2 Intra-Subject FF Repeatability

The intra-subject repeatability of Fat Fraction (FF) measurements was further evaluated through a longitudinal analysis conducted on a single subject who underwent six WB-MRI scans over a period of 24 months. The methodology applied in this section mirrors the one used for ADC analysis in Intra-Subject ADC Repeatability, with the goal of assessing the temporal consistency of FF measurements within the same individual, in the absence of inter-subject biological variability.

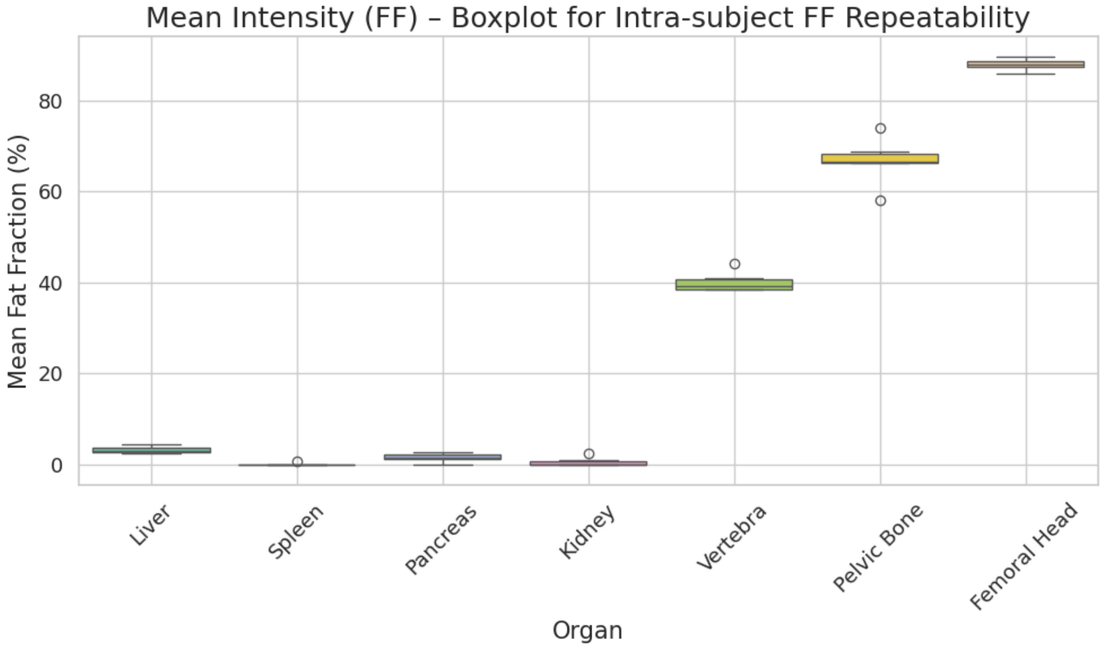


Figure 4.4: Boxplot of the mean intensity (FF value) across the different organs for the Intra-Subject FF Repeatability Analysis.

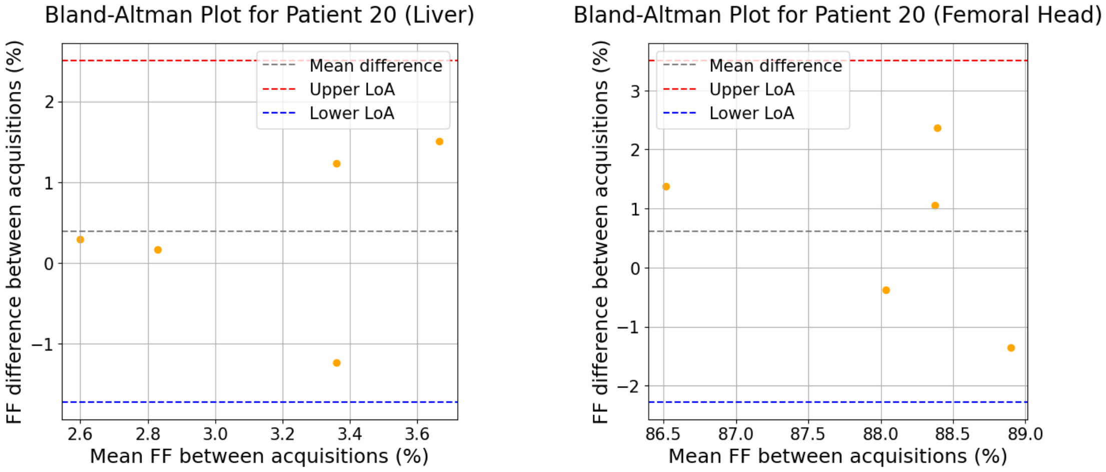
For each time point and organ, FF values were averaged across the three corresponding ROIs to obtain a single representative value per acquisition. In this subject, no outliers were identified or removed during preprocessing, so all three ROIs were available and

included in the analysis for each organ and time point.
The mean FF value for the single patient considered was calculated for each organ across all time points and the graphical results are in Figure 4.4.

For each organ and each of the six longitudinal acquisition time points, the average FF across the three ROIs was used to calculate the within-subject standard deviation (σ_w), the repeatability coefficient (RC), the within-subject coefficient of variation (wCV) and the limits of agreement ($\mathrm{LOA_L}$ and $\mathrm{LOA_U}$).
Results of all metrics for the seven analysed organs are reported in Table 4.3.

Organ	**Mean FF**	σ_w	**RC**	**wCV (%)**	**$\mathrm{LoA_L}$**	**$\mathrm{LoA_U}$**
Liver	3.21	0.79	2.20	25	-1.72	2.51
Spleen	0.13	0.29	0.79	224	-0.77	0.48
Pancreas	1.49	0.93	2.59	63	-3.93	3.64
Kidney	0.58	1.02	2.82	174	-1.98	1.00
Vertebra	40.09	2.30	6.37	6	-7.37	6.83
Pelvic Bone	66.71	5.13	14.23	8	-16.11	12.84
Femoral Head	87.93	1.32	3.66	2	-2.28	3.51

Table 4.3: Results for the Intra-Subject FF Repeatability Analysis for a single patient with six longitudinal acquisitions. All values are expressed as absolute percentages of fat fraction, except the ones explicitly indicated as percentages which are expressed as relative percentages.



[A] *[B]*

Figure 4.5: Bland-Altman plots for Intra-Subject FF Repeatability Analysis for a single subject for the liver[A] and the femoral head[B].

For each organ, Bland–Altman analysis was performed by computing the pairwise differences between successive acquisitions, using the general form of the Bland–Altman method[49].
The liver and femoral head were again selected as representative examples and are reported in Figure 4.5.

4.3.3 Organ FF Variability and Feature Stability

In analogy with the ADC analysis presented in Organ ADC Variability and Feature Stability, intra-organ variability of the Fat Fraction (FF) was evaluated across the full dataset, including all ROIs from the 41 WB-MRI acquisitions. The objective was to quantify the statistical dispersion of FF values within each anatomical region, as measured across different patients and time points, and to identify stable patterns in physiological fat content.

The same methodology was also applied to the entire set of radiomic features extracted via the SIBEX platform, with the goal of identifying the most stable descriptors for tissue characterisation beyond mean FF value alone (i.e the Mean Intensity of Grey Level feature of FF maps).

All analyses were performed on the cleaned dataset, where outlier ROIs were excluded according to the IQR-based method described in Outlier Removal. In the cleaned dataset, all negative FF values were also replaced with zero to ensure physiological consistency and prevent bias in the statistical summaries — a correction particularly relevant in organs with low fat content such as the spleen, pancreas and kidneys.

For each organ, FF values were averaged across ROIs per acquisition and their distributions were initially explored through boxplots, violin plots and histograms. Among these, the histograms were selected for visual presentation and are shown in Figure 4.6, representing the shape and range of FF distributions across the seven analysed tissues.

In addition to histograms, a single boxplot was generated to summarise the distribution of mean FF values across all seven organs in a compact and comparative format. The resulting plot is shown in Figure 4.7.
This representation complements the histograms by highlighting the differences in central tendency, variability and the presence of outliers between regions. It is important to note that the outliers shown in the boxplot are not residual artefacts from the original dataset but rather statistical extremes recomputed locally by the plotting algorithm. These values are flagged as outliers based on the standard boxplot convention within the context of each organ's distribution and are fully consistent with the cleaned dataset used in this analysis.

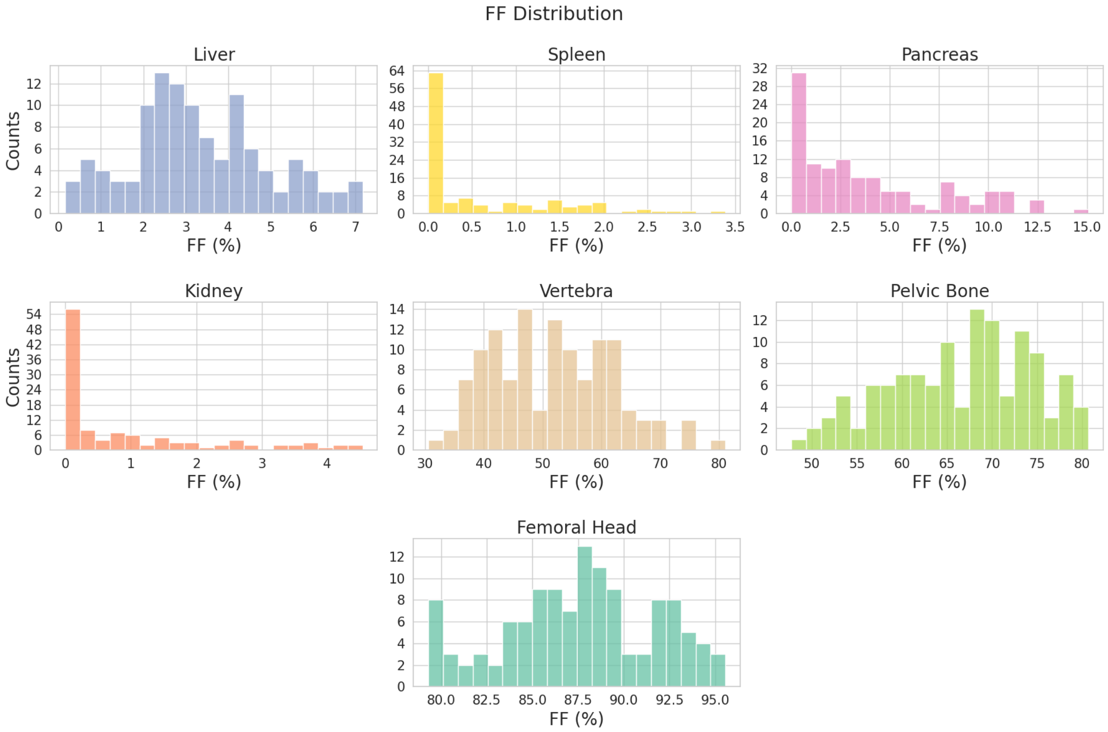


Figure 4.6: Histograms of the distributions of the mean FF value in the different organs.

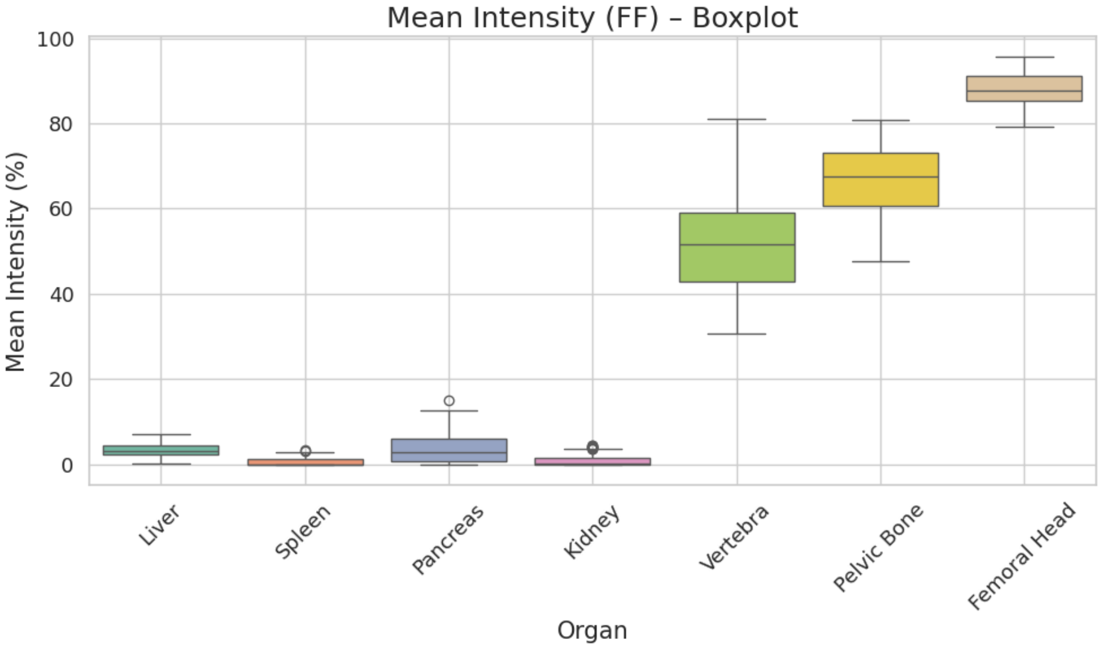


Figure 4.7: Boxplot of the mean intensity (FF value) across the different organs.

The descriptive metrics for intra-organ FF variability are summarised in Table 4.4. For each of the seven analysed organs, the table reports the mean, standard deviation (σ), median, minimum and maximum value, interquartile range (IQR) and coefficient of variation (CoV).
As done for all the statistical analyses presented throughout this chapter, all values were computed on the cleaned dataset obtained after outlier removal and setting all negative FF values to zero.

Organ	**Mean FF**	σ	**Median**	**Min**	**Max**	**IQR**	**CoV (%)**
Liver	3.35	1.64	3.14	0.16	7.16	2.07	49
Spleen	0.59	0.83	0.01	0.00	3.39	1.14	140
Pancreas	3.90	3.74	2.82	0.00	15.07	5.33	96
Kidney	0.93	1.28	0.26	0.00	4.54	1.43	138
Vertebra	51.84	10.22	51.52	30.57	81.09	16.06	20
Pelvic bone	66.49	7.91	67.66	47.71	80.70	12.34	12
Femoral head	87.65	4.14	87.83	79.29	95.56	5.82	5

Table 4.4: Results for Organ FF Variability Analysis. All values are expressed as absolute percentages of fat fractions, except the ones explicitly indicated as percentages which are expressed as relative percentages.

As in the ADC analysis (Organ ADC Variability and Feature Stability), the same methodology was extended to the entire set of radiomic features extracted via SIBEX, in order to evaluate their intra-organ variability and to identify stable descriptors for tissue characterisation. For each organ-feature pair, the coefficient of variation (CoV) was computed across all available ROIs and features were classified into three stability ranges: CoV $\leq$ 15% (high stability), 15% $<$ CoV $\leq$ 30% (moderate) and 30% $<$ CoV $\leq$ 45% (low). Features with CoV $>$ 45% or constant values were excluded from the analysis.

To assess redundancy, Pearson correlation matrices were generated within each stability group and representative, non-redundant features were selected from clusters of highly correlated descriptors ($r > 0.85$). As done before, only the pelvic bone is shown here as an example (Figure 4.8), with the matrix corresponding to the most stable group (CoV $\leq$ 15%).

Feature reliability was also assessed using the Intraclass Correlation Coefficient (ICC), as described in Organ ADC Variability and Feature Stability. Although ICC captures the proportion of variance due to inter-subject differences, its assumptions are only partially satisfied in this context, where ROIs are anatomically distinct. As a result, ICC-based selection produced much sparser correlation matrices and in this work only CoV was adopted as the main stability criterion for future feature selection and modelling.

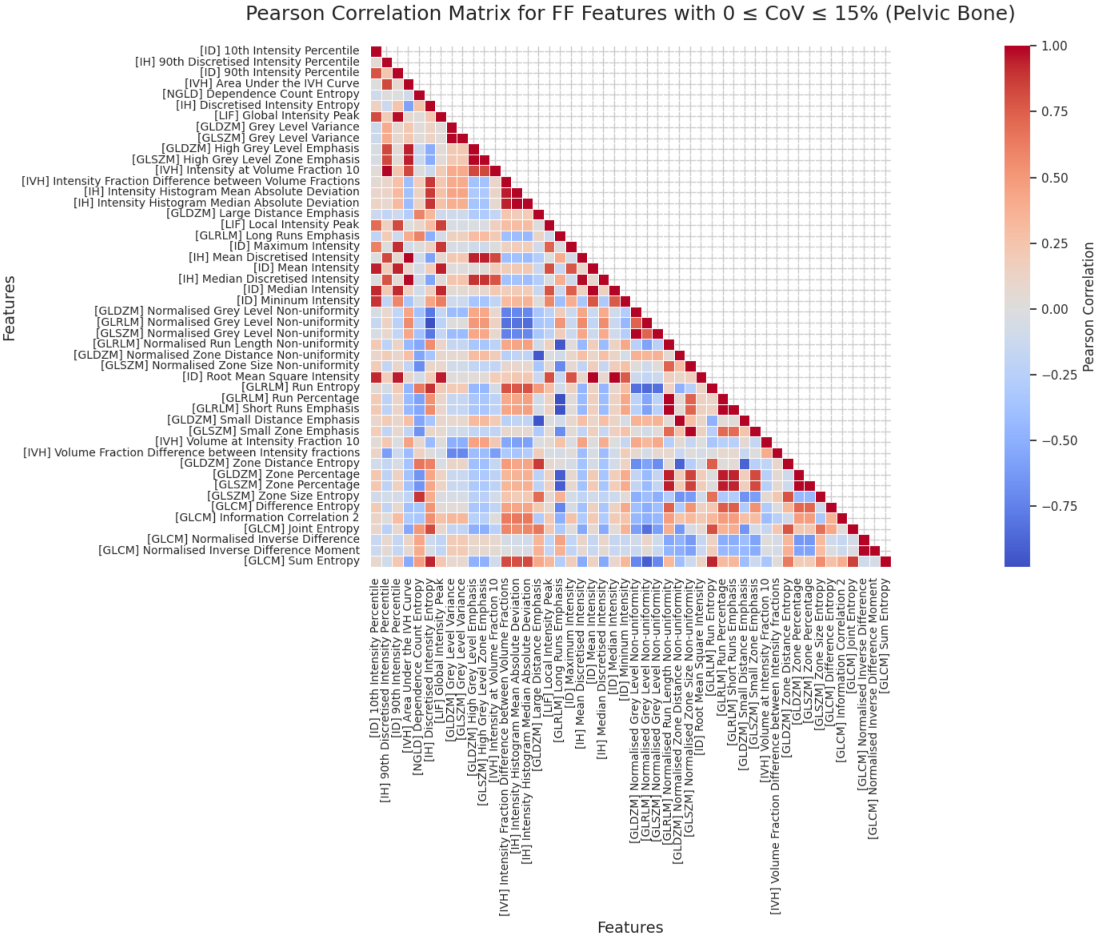


Figure 4.8: Pearson correlation matrix of FF features with CoV $\leq$ 15% in the pelvic bone.

4.4 Discussions

As done for the ADC in Quantitative ADC Analysis in Healthy Tissues, the results obtained in this chapter have been interpreted with the aim of first assessing the stability of the FF biomarker across repeated acquisitions, using the two complementary repeatability approaches illustrated in FF Repeatability Across the Sample and Intra-Subject FF Repeatability. Subsequently, the lipid content of different healthy tissues has been characterised in terms of fat fraction percentage, examining the distributions observed through descriptive statistical metrics and, where available, comparing them with reference values reported in the literature.

As previously noted for the ADC analysis discussion, it is important to emphasise that all evaluations were performed on WB-MRI data. Compared to region-specific imaging, WB-MRI is characterised by additional complexity due to its extensive anatomical coverage, heterogeneous tissue composition and greater susceptibility to acquisition-related artefacts. These aspects must be considered when interpreting the stability and reliability of FF measurements.

4.4.1 Repeatability of FF Measurements

Repeatability of Fat Fraction (FF) measurements was assessed to evaluate the temporal stability of the biomarker across repeated MRI acquisitions, following an approach consistent with the one employed for ADC. The first analysis was conducted on patients of the dataset with at least two scans, providing a broad characterisation of intra-subject repeatability. A second longitudinal assessment was then performed on a single subject who underwent six WB-MRI scans over 24 months, enabling a controlled evaluation of intra-individual variability over time.

The sample analysis revealed two distinct patterns depending on the tissue type. In adipose-rich regions such as the femoral head, pelvic bone and vertebrae, repeatability was high across all metrics. In these regions, the within-subject standard deviation (σ_w) was consistently smaller than the between-subject standard deviation (σ_b), indicating that most of the variability originated from biological differences between subjects rather than measurement noise. For instance, in the femoral head, σ_w was approximately 2.14% compared to σ_b of 2.99%, resulting in a low wCV of 4% and an RC of 5.93%. Similar ratios were observed in the pelvic bone and vertebrae, where σ_w remained in the 3.69–4.02% range, consistently lower than σ_b (7.15–8.31%). These findings suggest that FF can be considered a stable, with changes over time that are small relative to inter-individual variation.

In contrast, parenchymal tissues such as the liver, spleen, pancreas and kidneys exhibited a different behaviour. Here, σ_w and σ_b were often of comparable magnitude, suggesting that temporal variability within subjects was not negligible. In the liver, for example, σ_w was 0.96% versus σ_b of 1.11%; in the pancreas, the gap widened to σ_w = 2.09% and σ_b = 3.61%, indicating greater measurement noise relative to inter-subject variation. In the spleen and kidneys, σ_w and σ_b were very close (0.60% vs 0.52% and 0.84% vs 1.00%, respectively), making it difficult to distinguish between biological and technical sources of variability.

A complete summary of mean FF values together with σ_w and σ_b from the repeatability analysis across the sample is reported in Table 4.5. This table, extracted from the broader results presented in Table 4.2, is included here for clarity, summarising the behaviours observed across parenchymal and skeletal tissues.

Organ	**Mean FF**	σ_w	σ_b
Liver	3.19	0.96	1.11
Spleen	0.67	0.60	0.52
Pancreas	4.12	2.09	3.61
Kidney	1.14	0.84	1.00
Vertebra	51.77	4.02	8.31
Pelvic bone	66.76	3.69	7.15
Femoral head	87.46	2.14	2.99

Table 4.5: Mean values and corresponding σ_w and σ_b from the FF Repeatability Analysis Across the Sample. All values are expressed as absolute percentages of fat fractions.

Beyond the relative size of σ_w and σ_b, it is also important to consider the practical impact of measurement variability. The repeatability coefficient (RC), which estimates the maximum expected difference between two measurements acquired under identical conditions, was lower in absolute terms for parenchymal tissues than for osseous regions, but often comparable to — or even greater than — the mean FF values themselves. This limits the interpretability of small temporal changes in lean organs.
The within-subject coefficient of variation (wCV), a normalised index of dispersion that expresses variability relative to the mean of the measurements, was consistently higher in parenchymal tissues, further highlighting the reduced stability of FF in low-fat regions.

The longitudinal analysis conducted on a single subject provided additional insight into the repeatability of FF measurements over time. By eliminating inter-subject variability, it allowed for a clearer assessment of intra-subject temporal stability across organs.
In adipose-rich tissues, the within-subject standard deviation remained small relative to the mean FF, resulting in low RC values and confirming the high reliability of the biomarker even over a 24-month period (Table 4.3). In contrast, parenchymal tissues displayed similar or greater levels of dispersion despite biological constancy, with σ_w often approaching or exceeding the mean, leading to RC values that were disproportionately large in relative terms. This confirmed that technical variability dominates in low-fat regions. Additionally, the analysis highlighted a known limitation of relative metrics such as wCV: when the mean FF is close to zero, even small absolute deviations generate inflated coefficients of variation, reducing the interpretability of the measure. A quantitative summary of RC and wCV from both the repeatability analyses is reported in Table 4.6. Overall, RC and wCV values tended to be lower in the longitudinal setting, confirming greater stability of FF within the same subject over time. Exceptions were observed in organs with very low mean FF, such as the spleen and kidneys, where the relative coefficients were inflated by the near-zero signal, and in the pelvic bone, where RC increased despite a reduction in wCV.

	Across the Sample		**Intra-Subject**		**Variation**	
Organ	**RC**	**wCV (%)**	**RC**	**wCV (%)**	**ΔRC**	**ΔwCV (%)**
Liver	2.65	44	2.20	25	-0.45	-19
Spleen	1.67	117	0.79	224	-0.88	+107
Pancreas	5.78	98	2.59	63	-3.19	-35
Kidney	2.33	111	2.82	174	+0.49	+63
Vertebra	11.14	18	6.37	6	-4.77	-12
Pelvic bone	10.22	11	14.23	8	+4.01	-3
Femoral head	5.93	4	3.66	2	-2.27	-2

Table 4.6: Comparison of RC (in absolute percentage) and wCV between FF Repeatability Across the Sample and Intra-Subject FF Repeatability, with the corresponding variation between analyses.

The Limits of Agreement (LoAs) further confirmed distinct behaviours between tissue types. In osseous structures, LoAs were narrow relative to the mean FF and generally symmetric, indicating strong agreement across acquisitions while parenchymal tissues showed broader ranges, often approaching or exceeding the mean FF itself, reflecting limited measurement stability. In most cases, the absence of inter-subject variability in the single-subject analysis lead to slightly narrower LoAs compared to the cohort ones, with the only exception of the pelvic bone that can be attributed to single anomalous values, clearly observable in Figure 4.4.
The Bland-Altman plots reported as examples in Figure 4.5 were representative of the broader patterns observed across all tissues. In the liver, LoAs spanned approximately –1.7% to +2.5% around a mean FF of 3.21%, suggesting that measurement fluctuations could mask small physiological changes, while for the femoral head LoAs remained within -2.28% to 3.51% around a mean of 87.93%, confirming excellent repeatability.

These considerations demonstrate that FF is a repeatable and robust biomarker in adipose-rich tissues, where intra-subject variability remains low even across extended time intervals. In lean parenchymal organs, both absolute and relative dispersion limit the interpretability of repeated FF measurements, particularly when the expected physiological values are close to the detection threshold.

Although these results demonstrate strong internal consistency, the lack of published test–retest studies in healthy subjects prevents a systematic external validation. However, several works indirectly support the reliability of FF in bone tissue. Chemical shift encoding-based water–fat imaging is routinely employed to assess bone marrow composition and detect physiological or pathological changes, implying that FF is sufficiently stable to reflect meaningful biological variations[68]. The robustness of FF measurements in osseous regions is further supported by the consistent spatial gradients observed across anatomical subregions of the proximal femur in large healthy cohorts[69] and by quantitative assessments of repeatability coefficients across acquisition protocols using CSE-MRI[18].

4.4.2 Physiological Characterisation of FF Values in Healthy Tissues

The physiological distribution of FF values across the seven anatomical regions in the healthy cohort was derived from the analysis presented in Organ FF Variability and Feature Stability, which included all 41 WB-MRI acquisitions in the dataset.
As a methodological note, due to the exclusion of outlier ROIs based on the IQR method (Outlier Removal), not all acquisitions contributed uniformly to each organ and the number of retained acquisitions per organ ranged approximately between 38 and 41.

For the parenchymal organs considered in this study the mean fat fraction (FF) values reported in Table 4.4 were compared against reference values available in the literature for healthy tissues. Differently from skeletal structures, the fat content in parenchymal organs is more heterogeneous and subject to inter-individual variability since FF in these tissues is influenced not only by structural anatomy but also by personal factors such as age, metabolic status and body composition. As a result, the literature typically does not define strict physiological ranges for these organs but rather reports threshold values below which the tissue is generally considered healthy in comparison to pathological states.

In the liver, the average FF value observed in this cohort was (3.35 ± 1.64)%, which is considered normal and is below the commonly used diagnostic cut-off of 5.00% for hepatic steatosis (i.e. the excessive deposition of triglycerides in the liver)[70].
For the spleen, the mean FF was (0.59% ± 0.83)%, in line with its known lack of physiologic fat: even using advanced CSE-MRI techniques, measured FF values in the spleen of healthy subjects remain close to zero and are often considered artefactual or noise-driven[71].
The pancreas showed a mean FF of (3.90 ± 3.74)%, which aligns with values reported for metabolically healthy adults individuals such as a median pancreatic FF of 4.9% (IQR 3.1–7.4%)[72]. The variability observed in the dataset is in line with the known heterogeneity of pancreatic fat distribution, which is sensitive to both anatomical and metabolic factors and higly influenced by various metabolic, inflammatory and neoplastic diseases[64].
For the kidneys, the mean FF was (0.93 ± 1.28)% which is compatible with a healthy renal profile, as it remains well below the 4% threshold typically used to define ”fatty kidney” (i.e. the pathological fat accumulation in the kidneys).

The vertebral marrow in our cohort exhibited a mean fat fraction of (51.84 ± 10.22)%, with a 95% confidence interval of [48.71; 54.97]. The discrepancy with reference values reported in the literature — such as a mean of 64.43 [64.09; 64.77]% for lumbar vertebrae in healthy subjects[73] — was found to be statistically significant (Welch's t-test resulted in $p = 1.11 \times 10^{-9}$ but can be justified by several physiological and methodological considerations. First, the vertebral analysis performed in this study included all available vertebrae across the spine but ROI placement was limited to morphologically healthy vertebrae, based on visual inspection, in order to avoid regions affected by metastases or structural abnormalities. This selection was necessary given that the dataset consisted of oncological patients and only non-pathological tissues were included in the analysis. No distinction was made between vertebral anatomical levels, while literature has shown that vertebral fat content increases caudally along the spine. In addition, vertebral marrow fat fraction exhibits a wide physiological range and its distribution can vary considerably even within healthy populations, being influenced by factors such as sex and age[68]. For a more accurate comparison, a closer demographic matching between cohorts would be necessary.

The femoral head in our cohort exhibited a mean fat fraction of (87.65 ± 4.14)%, with a 95% confidence interval of [86.37; 88.93] and an interquartile range (IQR) of 5.82%. This value is in close agreement with literature data for the proximal femoral epiphysis in healthy adults, where Teixeira et al. reported a median FF of 89% (IQR: 84–92%) in a cohort of 131 subjects[69]. Although a formal Welch's t-test could not be performed due to the lack of mean and standard deviation in the reference study, the substantial overlap between our confidence interval and the interquartile range reported in the literature supports the consistency of our results with established physiological values. The narrow confidence interval and limited IQR observed in the data further support the low inter-subject variability of FF in this region, likely due to the homogeneous composition of yellow marrow typically found in the femoral epiphysis, where the adipose content in healthy adults is uniform.

No direct reference values were available in the literature for the iliac bone and no previous studies were found to apply an analysis comparable to the one presented in this work. However, the consistency of our findings in the other skeletal districts - the strong agreement observed in the femoral head and the physiologically interpretable deviation in the vertebral marrow - supports the robustness of FF as a quantitative biomarker in bone tissue. The pelvic bone, like the femur and vertebrae, contains bone marrow with a predominantly adipose composition in healthy adults. In our dataset, the fat fraction measured in this region was (66.49 ± 7.91)%, with a 95% confidence interval of [64.07; 68.91]. This value lies between the vertebral marrow and femoral head fat fraction and is physiologically plausible, reflecting the expected gradient in adipose content across skeletal regions, which generally increases from central axial regions towards peripheral appendicular sites like the femur[74].

A summary of the measured FF values together with their comparison against available literature evidence and physiological expectations is provided in Table 4.7. The table offers a structured summary of the considerations discussed for each of the seven analysed organs, while emphasising that the present analysis is not directly comparable with literature data, whose heterogeneous reporting formats often preclude a direct one-to-one comparison.

Organ	**Measured FF [CI_{95}]**	**Comparison with literature and physiological expectations**
Liver	3.35 [2.75; 3.95]	Below the 5% cut-off for hepatic steatosis.
Spleen	0.59 [0.33; 0.85]	Consistent with the well-established absence of physiological fat content.
Pancreas	3.90 [2.74; 5.06]	Aligned with values reported in metabolically healthy adults, reflecting the known heterogeneity of pancreatic fat distribution.
Kidney	0.93 [0.53; 1.33]	Below the 4% threshold for fatty kidney.
Vertebra	51.84 [48.71; 54.97]	Within the wide physiological range reported in the literature; methodological differences justify potential discrepancies.
Pelvic bone	66.49 [64.07; 68.91]%	Intermediate between vertebral and femoral head FF values, as expected for skeletal bone marrow composition.
Femoral head	87.65 [86.37; 88.93]	The confidence interval showed substantial overlap with literature IQR ranges.

Table 4.7: Comparison between measured FF values and literature or physiological expectations. All values are expressed as absolute percentages.

The consistency observed across the three bone structures - vertebra, pelvic bone and femoral head - supports the reliability of our measurements and confirms the suitability of FF as a robust imaging biomarker for the quantitative characterisation of healthy bone marrow. This expectation is supported by the statistical indicators reported in Table 4.8. The interquartile range (IQR), although not normalised, appears well-contained when considered relative to the mean, indicating a limited spread of values in absolute terms, in particular in the femoral head where the IQR was below 6%.

In addition, the coefficient of variation (CoV), which expresses dispersion relative to the mean, further highlights the contrast between skeletal and parenchymal tissues. In the latter, CoV values were generally above 50% and exceeded 100% in the spleen and kidneys due to their very low mean FF. In contrast, the skeletal regions demonstrated substantially lower variability and high stability: the vertebrae had a CoV of 20%, the pelvic bone of 12% and the femoral head of only 5%.

Organ	**IQR**	**CoV (%)**
Liver	2.07	49
Spleen	1.14	140
Pancreas	5.33	96
Kidney	1.43	138
Vertebra	16.06	20
Pelvic bone	12.34	12
Femoral head	5.82	5

Table 4.8: Summary of Organ FF Variability metrics: IQR (in absolute percentage) and CoV.

In contrast to what was observed for ADC in Quantitative ADC Analysis in Healthy Tissues, where osseous tissues exhibited greater variability due to lower diffusivity and reduced signal-to-noise ratio, fat fraction showed an inverted trend: variability was markedly lower in skeletal regions compared to parenchymal organs.
This shift in behaviour was also reflected in the radiomic domain. As shown in Figure 4.8 for the pelvic bone, the number and distribution of stable features (CoV $\leq$ 15%) were highly comparable between the two biomarkers (49 for ADC and 52 for FF) but their anatomical interpretation differed. For ADC, the identification of low-variability features in bone was necessary to mitigate the inherent instability of the primary metric, affected by low signal and noise. In contrast, FF was already physiologically homogeneous in bone marrow and the feature matrix reflected this consistency across multiple families, including GLCM, GLDZM, GLRLM and first-order descriptors.
In both cases, extending the analysis to the full radiomic feature set yielded a broad selection of stable descriptors in the pelvic region. However, since these features act as complementary elements with divergent physiological implications, such biomarker-specific behaviours should be carefully considered when developing radiomic-based predictive models.

5 Clinical Application of ADC and FF in Multiple Myeloma

The characterisation of healthy tissues in Quantitative ADC Analysis in Healthy Tissues and Quantitative FF Analysis in Healthy Tissues highlighted distinct behaviours for the imaging biomarkers ADC and FF in different anatomical regions, especially in terms of repeatability and stability.
In particular, the findings observed in bone tissues such as the pelvic bone, vertebrae and femoral head provide the foundation for their clinical application in diseases that alter the composition of bone marrow. Among these, a particularly relevant and clinically significant target is Multiple Myeloma (MM) - a malignant plasma cell disorder characterised by the clonal proliferation of plasma cells in the bone marrow.

In this chapter, the expected behaviour of ADC and FF is first examined in relation to myeloma disease progression, as pathological infiltration alters the marrow microenvironment. This is followed by an exploratory analysis of a new dataset including patients with multiple myeloma at different clinical stages. Predictive modelling approaches are then applied, using both mean values and additional radiomic features extracted from ADC and FF maps. Finally, the clinical relevance of these findings is discussed, with particular focus on the potential integration of imaging biomarkers with genetic and laboratory data to enable more comprehensive patient stratification.

5.1 Biomarker Expectations in Disease Progression

Quantitative Imaging Biomarkers (QIBs) such as the Apparent Diffusion Coefficient (ADC) and Fat Fraction (FF) are central in the evaluation of bone marrow involvement in Multiple Myeloma (MM). These biomarkers reflect microstructural changes occurring in the bone marrow and can be used for the non-invasive assessment and monitoring of tumour burden. Due to their distinct signal properties and physiological sensitivity, their combined interpretation is essential to characterise marrow composition, allowing effective discrimination between healthy and pathological conditions.

In the context of Whole-Body MRI (WB-MRI), the two biomarkers exhibit markedly different behavior in healthy bone marrow, as demonstrated in Quantitative ADC Analysis in Healthy Tissues and Quantitative FF Analysis in Healthy Tissues.
ADC is a well-established biomarker that reflects water diffusivity within tissues. In healthy bone marrow, ADC values tend to be low due to the limited water content and compact microarchitecture, resulting in low signal intensity and a poor signal-to-noise ratio (SNR) in diffusion-weighted sequences. This combination leads to higher measure-

ment variability and makes ADC less sensitive to subtle changes in normal bone tissue, although its responsiveness increases significantly in the presence of pathological infiltrations.
FF, on the other hand, is an emerging imaging biomarker that quantifies the proportion of fat in the bone marrow[64]. In healthy conditions, FF values are typically high, owing to the physiological abundance of adipocytes, which serve as an energy reservoir for the skeletal system and play a regulatory role in haematopoiesis. This high fat content generates a strong and homogeneous FF signal, which translates into excellent stability and repeatability in quantitative measurements. The FF signal remains robust even in repeated acquisitions and its intrinsic stability makes FF particularly effective in distinguishing healthy bone marrow from pathological infiltration.

Given their distinct biophysical sensitivities — ADC to water mobility and FF to fat content — the combined use of both biomarkers is essential to obtain a complete and reliable characterisation of bone marrow tissue. While ADC is more responsive to changes in cellularity and tissue microstructure, FF reflects variations in fat composition. Together, they allow for the detection of alterations in marrow architecture and fat content that occur during the progression of infiltrative bone marrow diseases, enabling a more accurate distinction between physiological and pathological states and potentially supporting longitudinal assessment of disease burden.

Multiple Myeloma (MM) is an haematological cancer characterised by the clonal proliferation of plasma cells within the bone marrow.
As outlined in WB-MRI in Multiple Myeloma, this malignant plasma cell disorder evolves through a continuum of clinical stages: Monoclonal Gammopathy of Undetermined Significance (MGUS), Smoldering Multiple Myeloma (SMM) and finally Multiple Myeloma (MM), which represents the symptomatic and clinically active phase of the disease. This biological progression reflects increasing tumour burden and marrow infiltration, transitioning from indolent to active disease. The shift from SMM to MM is clinically defined by the appearance of Myeloma-Defining Events (MDEs), such as 60% bone marrow plasma cells, elevated serum free light chain ratios or the presence of more than one focal lesion on MRI[28, 30, 31].

While these clinical stages reflect systemic tumour evolution, WB-MRI allows for spatial characterisation of marrow involvement, capturing not only the extent of disease but also its distribution pattern. The MY-RADS guidelines[28] describe how infiltration patterns can be assessed through diffusion-weighted imaging, where healthy bone marrow typically appears with a low, homogeneous signal on high *b-value* DWI, consistent with normal fat-rich composition and absence of disease. In contrast, tumour-infiltrated marrow presents as hyperintense areas on DWI and shows increased ADC values due to higher cellularity and water content.

Marrow involvement in MM can be radiologically classified into several progression-linked patterns. The focal pattern is characterised by one or more discrete lesions within the marrow and reflects localised infiltration, marking the transition from smoldering to active disease. In the diffuse pattern, the entire marrow appears uniformly infiltrated, with globally increased signal on DWI that indicates widespread replacement of fat by tumour cells and is typically associated with advanced disease and high tumour burden. Focal on diffuse is considered an even more severe pattern, where focal lesions emerge on a background of diffuse infiltration, suggesting both structural disorganisation and active proliferative disease. At more advanced stages, tumour growth can extend beyond the marrow cavity: in the paramedullary pattern, the tumour mass lies adjacent to but external to the bone marrow and in the extramedullary pattern infiltration progresses further into surrounding soft tissues, indicating highly aggressive disease with extensive systemic dissemination[75].

These radiological patterns not only correspond to distinct morphological stages of infiltration but also correlate with quantitative changes in bone marrow composition, particularly in ADC and FF values. As the disease progresses, their interpretation provides both visual and quantitative insight into the extent and aggressiveness of marrow involvement.
From a tissue composition standpoint, the infiltration of bone marrow by malignant plasma cells leads to significant microstructural alterations. The progressive replacement of adipocytes and haematopoietic elements by dense, lipid-poor tumour tissue results in

- an increase in ADC, due to greater extracellular water content and reduced microstructural restriction;
- a decrease in FF, reflecting the depletion of marrow fat.

This transformation reflects the remodelling of the marrow environment during disease progression, as bone marrow adipose cells are replaced by focal or diffuse malignant infiltration[76]. This trend is also supported by quantitative MRI studies, which have shown consistent increases in ADC and decreases in FF across infiltration patterns, in correlation with clinical and prognostic markers[77], and form the basis to use Quantitative Imaging Biomarkers (QIBs) in the non-invasive assessment of disease progression, from diagnosis to longitudinal follow-up.
A visual summary of these expected microstructural and quantitative changes between healthy and myeloma-infiltrated bone marrow is presented in the following schematic overview (Figure 5.1).

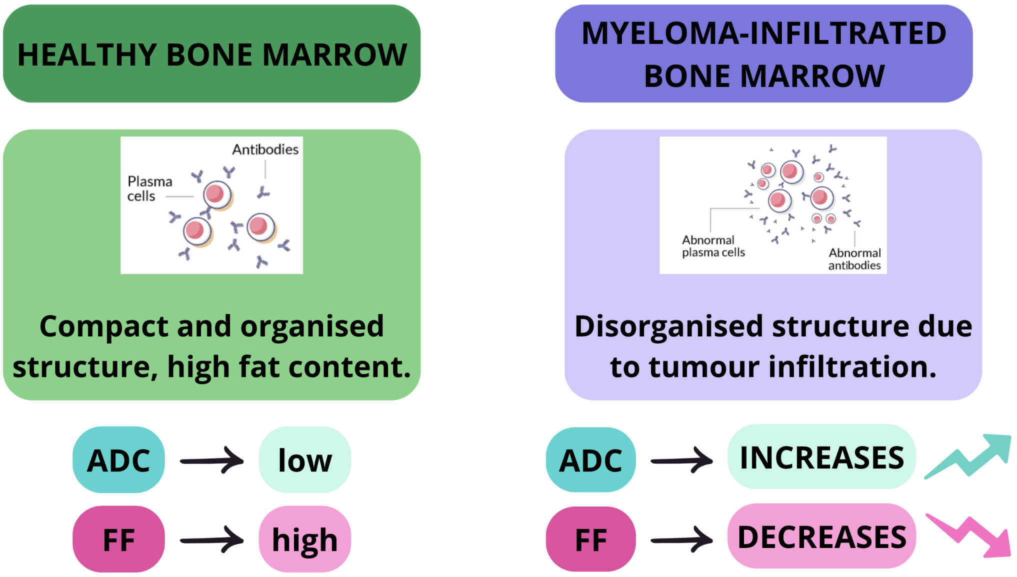


Figure 5.1: Schematic comparison of marrow structure and corresponding ADC and FF behaviour in healthy and myeloma-infiltrated bone.

The conceptual trends illustrated above are corroborated by quantitative imaging studies that validate the expected behavior of ADC and FF in the clinical setting. These biomarkers have been shown to correlate with bone marrow plasma cell percentage (BMPC), serum monoclonal protein levels and other prognostic indicators such as β_2-microglobulin, hemoglobin concentration and flow cytometry findings[78, 77]. From a therapeutic perspective, their quantitative dynamics are also useful for treatment monitoring, with FF emerging as the more stable and reproducible parameter to assess early response to therapy[79].

To support the theoretical trends described above, a preliminary visual inspection can provide qualitative confirmation of the expected behaviours of ADC and FF in healthy and myeloma-infiltrated bone marrow. This comparison can be performed by analysing axial images extracted from WB-MRI-derived ADC and FF maps.
Figure 5.2 displays axial slices from WB-MRI acquisitions, comparing healthy and myeloma-infiltrated bone marrow in both ADC and FF maps. The pelvic bone has been selected due to its high clinical relevance, frequent involvement in MM and the availability of histological confirmation from diagnostic bone marrow biopsies routinely performed in this location. In Figure 5.2, the right and left pelvic bones have been manually contoured in red to clearly identify the regions of interest. It should be noted that the images [A] and [C] belong to the same healthy subject, a female patient, while the pathological cases [B] and [D] come from two different male patients. Minor anatomical differences

between scans are thus attributable to physiological inter-subject variability.

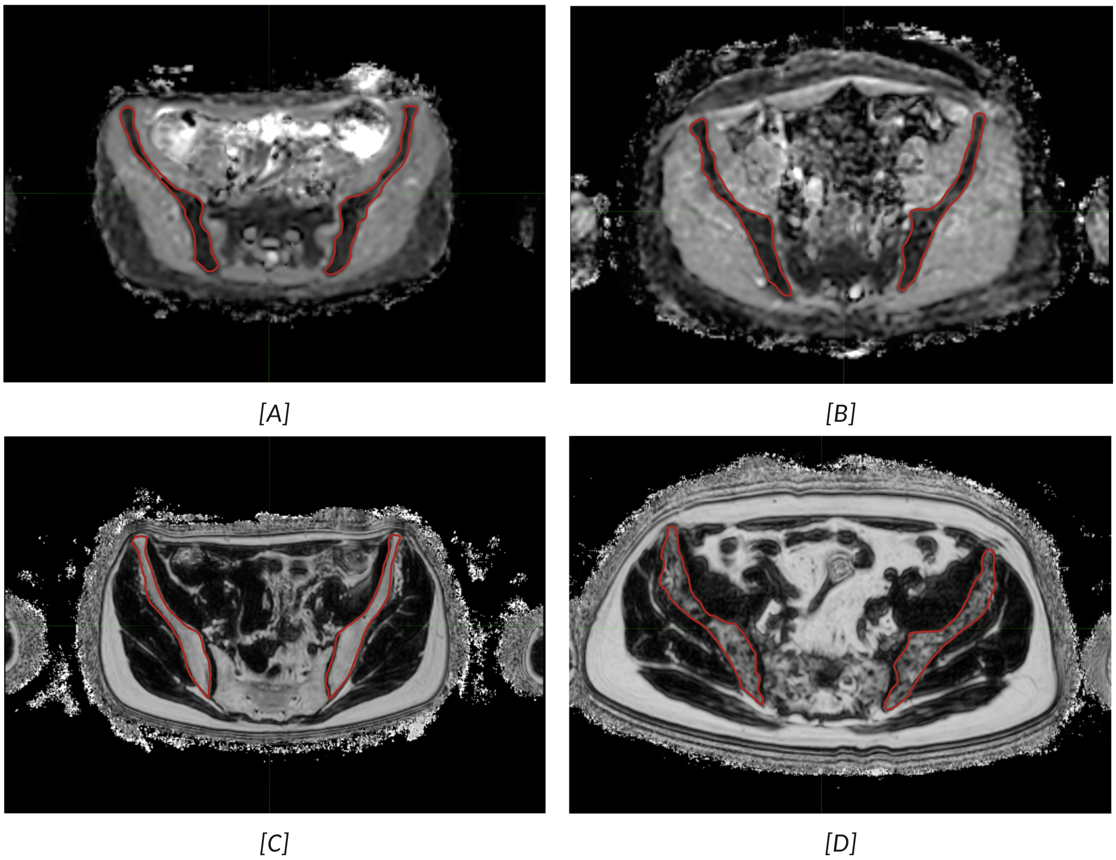


Figure 5.2: Qualitative comparison between healthy and myeloma-infiltrated pelvic bone marrow (outlined in red) in WB-MRI. ADC maps are shown in [A] for a healthy subject and in [B] for a patient with myeloma. FF maps are displayed in [C] and [D], respectively for a healthy subject and a myeloma patient.

In Figure 5.2, images [A] and [B] show the ADC maps: in the healthy subject ([A]), the bone marrow of the pelvic region appears dark, reflecting the typically low ADC signal due to compact structure and limited water diffusivity. In the pathological case ([B]), the marrow shows slightly hyperintense areas, indicating increased diffusivity. The distinction remains visually subtle and affected by noise, which is a known limitation of ADC imaging in osseous tissue due to inherently low signal-to-noise ratio.
In contrast, the FF maps shown in [C] and [D] reveal a more apparent qualitative difference. In the healthy subject ([C]), the pelvic bone displays uniformly high signal, consistent with the expected high fat content of normal haematopoietic marrow. In the pathological case ([D]), several darker areas are clearly visible within the red contours

of the pelvic bone, indicating reduced fat fraction in the infiltrated marrow.

5.2 Patient Cohort and Exploratory Analysis

The theoretical expectations regarding the behaviour of ADC and FF in healthy versus myeloma-infiltrated bone marrow were further examined through the analysis of a comprehensive patient cohort, including subjects across different clinical stages. A visual exploration of biomarker variations in relation to disease progression is presented to qualitatively assess emerging trends and provide context for the predictive modelling presented in the subsequent sections.

5.2.1 Patient Groups and Image Data

The dataset used for this analysis includes both healthy and diseased subjects across different clinical stages of multiple myeloma. A total of 148 regions of interest (ROIs) were analysed, each positioned within the right iliac bone of the pelvis and with a fixed volume of 2.6 cm^3. The pelvic region was selected due to its clinical relevance in detecting marrow infiltration. To minimise variability in signal intensity, the right side was consistently chosen for all subjects as the pelvis lies near the interface between the body and lower-limb coil arrays in WB-MRI and this transition zone could introduce subtle signal discontinuities between the left and right sides.

The healthy group of the cohort consists of 20 patients from a breast cancer imaging protocol, previously analysed in Quantitative ADC Analysis in Healthy Tissues and Quantitative FF Analysis in Healthy Tissues. Although these individuals were oncological patients, the tissues selected for analysis were free from pathological infiltration and considered physiologically normal. Several patients underwent multiple MRI acquisitions and one ROI was selected from the right pelvic bone in each acquisition. This resulted in a total of 41 non-pathological ROIs.
The pathological portion of the dataset includes three distinct clinical groups:

- SMM ($n = 38$): patients in an early, asymptomatic stage with detectable plasma cell proliferation in the marrow but no clinical end-organ damage.
- MM ($n = 46$): patients in the active, symptomatic form of the disease, characterised by marrow infiltration and CRAB criteria.
- RRMM ($n = 23$): relapsed/refractory patients who did not achieve remission or showed progression after treatment.

In all diseased patients, one ROI was positioned in the right pelvic bone, in proximity to the biopsy site used for diagnosis.

The composition of the entire dataset is summarised in Table 5.1.

Group	Description	Number of Patients	Total ROIs
Healthy	Control	20	41
SMM	Early, asymptomatic stage	38	38
MM	Active disease	46	46
RRMM	Relapsed or refractory disease	23	23
Total		127	**148**

Table 5.1: Composition of the multiple myeloma dataset.

Each ROI was placed on both the ADC and FF maps obtained from the patients' WB-MRI acquisitions, carefully adjusting for potential spatial misalignments between the sequences to ensure anatomically consistent placement within the pelvic bone tissue.
The acquisition protocols followed are described in detail in WB-MRI Protocols and Acquisition Parameters for Quantitative Sequences. All myeloma patients underwent imaging according to the MY-RADS-based WB-MRI protocol, while healthy controls were scanned using the breast cancer protocol. As discussed in the corresponding section, although the acquisition protocols slightly differ, their technical characteristics are sufficiently comparable to ensure that FF and ADC values can be consistently interpreted across the two protocols.
Since this dataset includes patients who underwent WB-MRI in years prior to the introduction of the FatQuant sequence, fat fraction estimation was based on the FF map derived from the mDixonAll acquisition for all patients. As highlighted in Phantom Validation of Fat Fraction Sequences, this sequence has been validated as suitable for quantitative fat assessment in bone marrow.

Each ROI was processed using the SIBEX radiomics platform[44, 45] to extract not only the mean signal intensity values of ADC and FF but also a comprehensive set of radiomic features. This approach allowed for an extended analysis in the radiomic domain, consistent with the methodology adopted in Quantitative ADC Analysis in Healthy Tissues and Quantitative FF Analysis in Healthy Tissues.

5.2.2 Distribution of ADC and FF Across Clinical Groups

To investigate the behaviour of quantitative imaging biomarkers across clinical stages, an initial exploratory analysis was conducted on the dataset comprising both healthy controls and multiple myeloma patients. Specifically, the cohort was subdivided into four clinical categories: Healthy, SMM, MM and RRMM. Additionally, an aggregated category labelled ”All Diseased” was included to represent the entire spectrum of pathological cases. This group combines all stages of myeloma and provides a summarised view of disease-related alterations, which is particularly useful in the context of binary classification tasks presented later in this chapter.

The analysis focused on the distribution of the mean values of ADC and FF within each group. For each biomarker, a boxplot was generated to provide a visual representation of its variation across categories and to identify potential trends related to disease progression. To support these qualitative observations, non-parametric statistical tests were also performed to assess whether the observed differences between groups were statistically significant.

The first biomarker analysed was the Apparent Diffusion Coefficient (ADC). Figure 5.3 shows the distribution of mean ADC values across all clinical groups. Each box represents the interquartile range of the ADC values in the corresponding group, with the median indicated by the horizontal line.

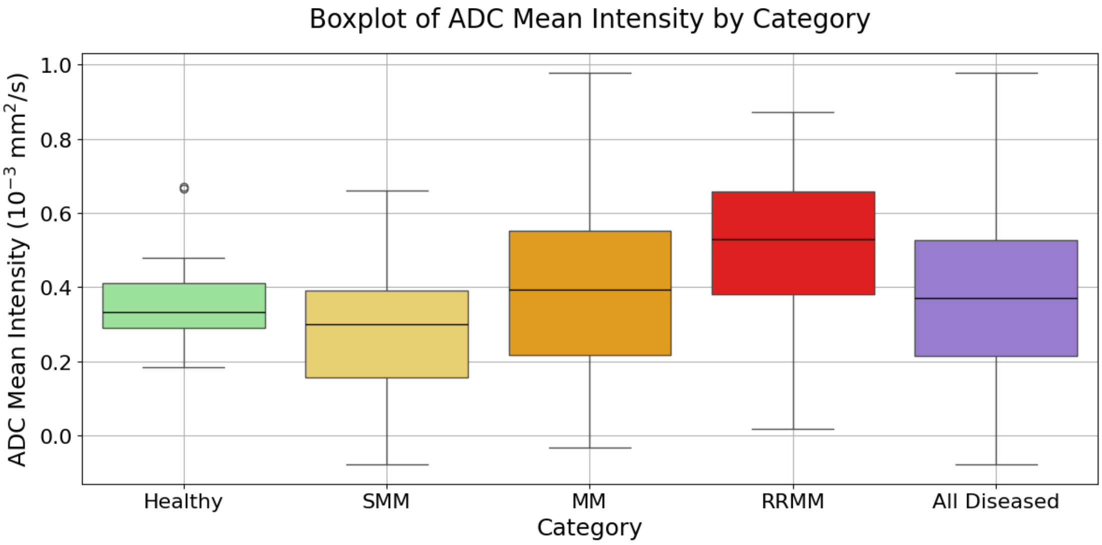


Figure 5.3: Boxplot of the mean intensity (ADC value) across the different categories of the multiple myeloma dataset.

As illustrated in the figure, ADC values tend to increase with disease severity. The Healthy group exhibits low median values, consistent with the limited water diffusivity of normal, fat-rich bone marrow. A progressive rise is observed in the MM and RRMM groups, reflecting the increased cellularity and extracellular water content associated with pathological infiltration.

Interestingly, the SMM group shows slightly lower ADC values than the Healthy group, a finding that diverges from the expected monotonic trend. This inversion may be par-

tially explained by technical factors: all WB-MRI scans of myeloma patients were acquired in earlier years compared to those of healthy controls, possibly before full optimisation of DWI sequences. Additionally, SMM is a biologically heterogeneous stage where some patients may exhibit minimal or indolent marrow involvement, contributing to this variability.
The variance in ADC values is notably higher in the diseased groups compared to the healthy cohort. This observation is consistent with the diversity of infiltration patterns in multiple myeloma, which range from focal to diffuse and lead to broader heterogeneity in tissue composition and water diffusivity.

To determine whether the visual trends observed in ADC values were statistically significant across clinical categories, non-parametric tests were employed. Given the non-Gaussian distribution of the data and the unequal group sizes, the Kruskal–Wallis test was applied to compare ADC distributions across the four clinical stages: Healthy, SMM, MM, and RRMM. This test evaluates the null hypothesis that all groups originate from the same population distribution and is a rank-based method that does not assume normality[80]. The result revealed a statistically significant difference ($H = 20.79$, $p = 1.16 \times 10^{-4}$), confirming that at least one group differs significantly from the others in terms of marrow diffusivity.

In order to identify which specific groups were responsible for this difference, Dunn's post-hoc test with Bonferroni correction was performed. This non-parametric post-hoc procedure evaluates all possible pairwise comparisons between groups while controlling the family-wise error rate; the Bonferroni correction adjusts the significance threshold to account for multiple testing, thereby reducing the risk of false positives[80]. The results for ADC are reported in Table 5.2, showing adjusted *p*-values for all pairwise comparisons in matrix form.

	Healthy	**SMM**	**MM**	**RRMM**
Healthy	1.00	0.64	1.00	**0.01**
SMM	0.64	1.00	0.05	**5.20×10^{-5}**
MM	1.00	0.05	1.00	0.11
RRMM	**0.01**	**5.20×10^{-5}**	0.11	1.00

Table 5.2: Pairwise group comparisons for ADC mean values using Dunn's test with Bonferroni correction (adjusted p-values). Statistically significant results are highlighted in bold.

These results indicate that statistically significant differences in ADC are primarily associated with the RRMM group, which showed elevated values compared to both Healthy and SMM subjects. No significant differences were detected between Healthy, SMM and MM, suggesting that ADC alone may have limited discriminative power in distinguishing early or intermediate stages of myeloma.

To further assess the potential of ADC as a binary classifier between healthy and pathological marrow, an additional Kruskal–Wallis test was performed comparing Healthy subjects with the aggregated category All Diseased (comprising SMM, MM and RRMM). Interestingly, this test did not yield statistical significance ($H = 0.55$ and $p = 0.46$), highlighting that while ADC values tend to increase in advanced disease, their variability and overlap across early and intermediate stages limit their ability to clearly separate healthy from diseased marrow in a dichotomous setting. This finding reinforces the importance of incorporating disease staging into any ADC-based stratification and suggests that additional biomarkers may be necessary to enhance diagnostic accuracy in early-phase multiple myeloma.

The same exploratory analysis was applied to the Fat Fraction (FF) biomarker. Figure 5.4 displays the distribution of mean FF values across the clinical groups, including the aggregated All Diseased category.

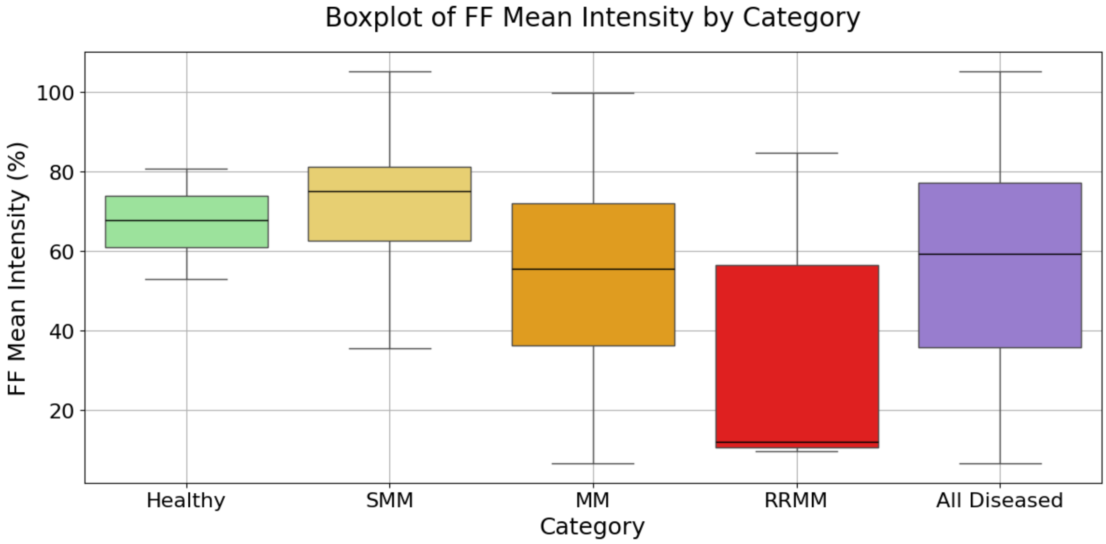


Figure 5.4: Boxplot of the mean intensity (FF value) across the different categories of the multiple myeloma dataset.

A general decreasing trend in FF values is observable along the clinical continuum, consistent with the expected reduction in marrow adiposity due to tumour infiltration. RRMM patients, in particular, exhibit markedly low FF values, in some cases approaching ~10%, which reflects extensive degradation of fat content in the bone marrow.
Conversely, the SMM group displays higher median FF values than the healthy controls.

This unexpected result may be attributed to the technical differences between the optimisation of acquisition protocols or to biological heterogeneity within the SMM cohort, as previously discussed.
As observed for ADC, the variance in FF is visibly greater among diseased patients, supporting the notion that myeloma infiltration alters marrow composition in a non-uniform manner across individuals and disease stages.

To assess statistical significance, a Kruskal–Wallis test confirmed a significant difference in FF values among the four clinical groups ($H = 37.53$ and $p = 3.56 \times 10^{-8}$). In contrast to ADC, the FF biomarker also demonstrated a significant difference when comparing Healthy subjects against the aggregated All Diseased group ($H = 4.24$ and $p = 0.04$), suggesting a more robust ability to distinguish healthy from infiltrated marrow even when early-stage patients are included.

Pairwise comparisons using Dunn's test with Bonferroni correction are reported in Table 5.3. Statistically significant differences were found between Healthy and RRMM ($p = 6.63 \times 10^{-5}$), SMM and MM ($p = 3.69 \times 10^{-4}$) and SMM and RRMM ($p = 2.05 \times 10^{-7}$).

	Healthy	**SMM**	**MM**	**RRMM**
Healthy	1.00	1.00	0.05	**6.63×10^{-5}**
SMM	1.00	1.00	**3.69×10^{-4}**	**2.05×10^{-7}**
MM	0.05	**3.69×10^{-4}**	1.00	0.14
RRMM	**6.60×10^{-5}**	**2.05×10^{-7}**	0.14	1.00

Table 5.3: Pairwise group comparisons for FF mean values using Dunn's test with Bonferroni correction (adjusted p-values). Statistically significant results are highlighted in bold.

These results highlight the superior discriminative power of FF in differentiating both between and within disease stages. In particular, the sharp reduction in fat content observed in RRMM patients reinforces the role of FF as a sensitive biomarker for advanced marrow infiltration. The presence of significant differences across multiple pairwise comparisons also suggests that FF may have stronger potential than ADC for early disease detection and stratification.

5.3 Predictive Modelling

The exploratory analyses presented in Distribution of ADC and FF Across Clinical Groups have demonstrated that both Apparent Diffusion Coefficient (ADC) and Fat Fraction (FF) exhibit systematic variations across different clinical stages of multiple myeloma. These findings suggest that Quantitative Imaging Biomarkers (QIBs) may have potential not

only for descriptive analysis but also for classification and disease stratification.
The aim of this section is to explore whether such biomarkers can support automated disease classification and contribute to clinically meaningful patient stratification.

Predictive modelling approaches were applied to investigate the diagnostic performance of ADC and FF in discriminating healthy subjects from patients at different disease stages. Logistic regression models were implemented to evaluate the ability of these biomarkers to distinguish between classes based on imaging features extracted from the right pelvic bone. Receiver Operating Characteristic (ROC) curves were generated to quantify classification performance, using the Area Under the Curve (AUC) as the primary evaluation metric. In the initial phase, models were tested using only the mean intensity values of ADC and FF, both individually and in combination, as outlined in Preliminary Results. This was followed by the development of extended models incorporating radiomic features, selected through a two-step process based on statistical variability and regularised regression, as described in Methods and Feature Selection. Final performance metrics and classification outcomes are reported in Multiparametric Classification Modelling with ADC, FF and Radiomic Features.

5.3.1 Preliminary Results

In the preliminary phase of modelling, the diagnostic potential of ADC and FF was assessed using only their respective mean intensity values, computed from the right pelvic bone regions of interest (ROIs) previously defined in the study cohort described in Patient Groups and Image Data. The objective was to evaluate whether these individual imaging biomarkers are sufficient to effectively discriminate between healthy subjects and patients at different clinical stages of multiple myeloma.

For each comparison, a binary classification problem was formulated: one class consisted of healthy controls, while the second class included patients in a specific disease stage (SMM, MM, RRMM) or all stages combined (All Diseased). Classification was performed using logistic regression, a statistical model commonly used for binary outcomes. It estimates the probability that a sample belongs to a given class as a function of one or more input features — in this case, either ADC or FF mean intensity values. The model outputs a probability score for each sample, which is then compared to a threshold to assign a predicted class[81].

To ensure robust evaluation, a Repeated Stratified K-Fold cross-validation strategy was employed, with three splits and 30 repeats. This method divides the dataset into stratified folds (preserving class proportions), repeatedly training and testing the model on different partitions. This approach mitigates overfitting and provides a more reliable estimate of the model's generalisability[82].

Model performance was assessed using Receiver Operating Characteristic (ROC) curves. These curves plot the true positive rate (i.e. sensitivity) against the false positive rate (i.e. 1 – specificity) across all classification thresholds. The Area Under the Curve (AUC) provides a scalar summary of performance: a value of 1.0 indicates perfect classification, while a value of 0.5 indicates random guessing. For each model, the mean AUC across cross-validation folds was computed along with its 95% confidence interval (CI_{95}), reflecting the uncertainty around the estimate.

The results for ADC and FF, tested independently, are summarised in Table 5.4.

	ADC Mean Intensity	**FF Mean Intensity**
Comparison	**AUC [CI_{95}]**	**AUC [CI_{95}]**
Healthy vs SMM	0.63 [0.62; 0.65]	0.63 [0.61; 0.65]
Healthy vs MM	0.52 [0.50; 0.55]	0.70 [0.68; 0.72]
Healthy vs RRMM	0.78 [0.76; 0.80]	0.82 [0.80; 0.85]
Healthy vs All Diseased	0.52 [0.50; 0.53]	0.61 [0.59; 0.62]

Table 5.4: AUC values and 95% confidence intervals (CI_{95}) for logistic regression models using ADC or FF mean intensity as single predictors.

Both biomarkers showed progressively higher AUC values with increasing disease severity, particularly in the RRMM group, where ADC and FF reached values of 0.78 and 0.82, respectively.

In early-stage comparisons, however, classification performance was limited. For instance, in the Healthy vs SMM setting, both ADC and FF achieved modest AUCs of 0.63, suggesting weak but symmetrical predictive ability. A marked discrepancy was observed in the Healthy vs MM comparison: while FF achieved an AUC of 0.70, ADC yielded a substantially lower value of 0.52, suggesting no classification ability. This trend persisted in the aggregated comparison between Healthy and All Diseased subjects, where ADC again resulted in an AUC of 0.52 — a value approaching random classification — while FF reached 0.61.

These results are consistent with the findings of the exploratory statistical analysis presented in Distribution of ADC and FF Across Clinical Groups, where no significant difference was found for ADC between Healthy and All Diseased groups, while FF exhibited a significant distinction.

Overall, the results indicate that neither ADC nor FF mean intensity values, when used individually, are sufficient to robustly classify disease status across the full clinical spectrum.

To assess whether combining information from both biomarkers improves classification performance, a second set of models was tested using the joint mean intensity values of ADC and FF as input features.

The same logistic regression framework and validation strategy were applied. The resulting AUC values are reported in Table 5.5 and the ROC curves are shown in Figure 5.5.

	ADC and FF Mean Intensities
Comparison	**AUC [CI_{95}]**
Healthy vs SMM	0.66 [0.64; 0.68]
Healthy vs MM	0.70 [0.68; 0.71]
Healthy vs RRMM	0.83 [0.81; 0.85]
Healthy vs All Diseased	0.61 [0.59; 0.62]

Table 5.5: AUC values and 95% confidence intervals (CI_{95}) for logistic regression models using combined ADC and FF mean intensities as predictors.

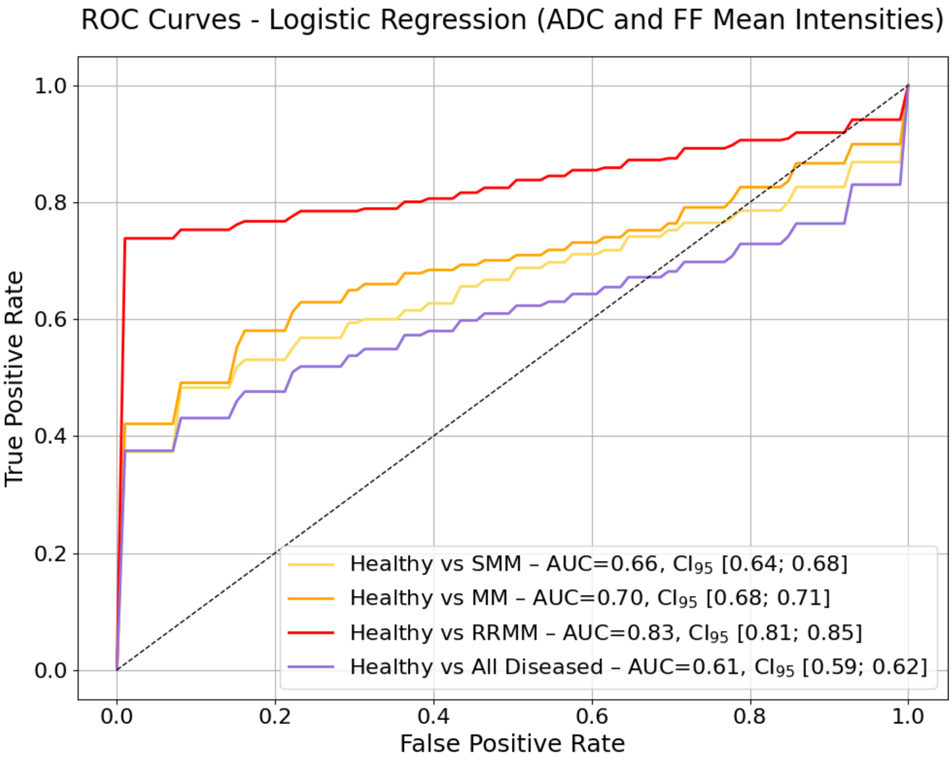


Figure 5.5: ROC curves for logistic regression models using combined ADC and FF mean intensities.

While the inclusion of both biomarkers led to slightly higher AUC values in most comparisons, the improvement was modest relative to using FF alone. In particular, the combined model achieved AUC values of 0.66 [0.64; 0.68] for Healthy vs SMM, 0.70

[0.68; 0.71] for Healthy vs MM, 0.83 [0.81; 0.85] for Healthy vs RRMM and 0.61 [0.59; 0.62] for Healthy vs All Diseased — values that closely reflect those obtained using FF mean intensity as a single predictor.
These findings reinforce the superior predictive performance of FF over ADC when using mean intensity values alone, while also supporting the potential of multiparametric approaches to improve classification when extended to more complex feature sets.

5.3.2 Methods and Feature Selection

As observed in the previous section, both ADC and FF mean intensity values showed a clear association with disease stage, reflecting biologically meaningful changes in bone marrow composition. However, their standalone predictive power remained limited, particularly for intermediate stages such as smoldering myeloma. While the combination of ADC and FF mean intensities provided a slight improvement, the overall performance suggested that additional information would be needed to capture more subtle or spatially heterogeneous disease patterns. These observations motivated a shift towards a multiparametric radiomic approach, intended to exploit the rich textural and structural information encoded in ADC and FF maps beyond simple average signal levels.

In selecting candidate features for classification, a natural strategy was to prioritise those with low variability in healthy tissues, particularly in the pelvic bone, as previously analysed in Organ ADC Variability and Feature Stability and Organ FF Variability and Feature Stability. In those sections, several features were identified as stable based on a Coefficient of Variation (CoV) below 15%, with a total of 49 ADC features and 52 FF features falling within this low-variability range, as illustrated in the corresponding correlation matrices in Figure 2.8 and Figure 4.8.
Directly using this extended set of stable features was not appropriate given the relatively small size of the multiple myeloma dataset (148 ROIs in total as described in Patient Groups and Image Data): including such a large number of predictors in a model with limited observations would have led to poor generalisability and a high risk of overfitting, violating standard principles of statistical learning.

For this reason, a dedicated selection strategy was developed to identify a compact, robust and non-redundant subset of radiomic features specific to the classification task. Among the features previously extracted using the SIBEX platform[44, 45], only those exhibiting a CoV between 0% and 5% within the healthy control group were selected, carrying out the procedure independently for ADC and FF maps.
Redundancy among the retained features was addressed by computing pairwise Pearson correlation coefficients. When two features exhibited high linear correlation (absolute value above 0.85), only the more stable of the pair — that is, the one with the

lower CoV — was kept. This ensured that the selected feature set was not only stable but also minimally redundant, providing complementary information and reducing the risk of multicollinearity in the model.
These steps reduced the initial high-dimensional feature set to a smaller, more stable subset for each modality: 9 features were retained for ADC and 7 for FF.

Following this, an embedded selection method was applied to identify the most informative predictors (features) for classification. Specifically, a multiclass logistic regression model with L1 regularisation (LASSO) was trained to distinguish between the four diagnostic groups: Healthy, SMM, MM and RRMM. In this framework, the feature selection occurs automatically during model fitting: the L1 penalty encourages sparsity in the model coefficients by driving many of them exactly to zero. Features with non-zero coefficients in at least one class are retained, as these contribute to the discrimination between groups. This regularised approach yields compact models with only a few active predictors, improving generalisability and reducing the risk of overfitting while concentrating on the most relevant descriptors[83].

In addition to the selected radiomic features, the mean intensity of each map was also included in the analysis. This decision was justified by the biological interpretation of these metrics: the mean ADC reflects water diffusivity, which tends to increase with marrow infiltration, while the mean FF represents fat content, which typically decreases as disease progresses. Including these values thus added a complementary summary of tissue composition with known clinical relevance, as already demonstrated in Distribution of ADC and FF Across Clinical Groups.

To facilitate a structured comparison between different modelling strategies, three distinct datasets were prepared: one containing only features derived from ADC maps and their corresponding mean intensity, one including only FF-based features and mean intensity and a third one combining all features from both modalities. This design enabled a systematic evaluation of the predictive contribution of each biomarker, both individually and in combination.

The final feature sets consisted of the following descriptors:

- ADC [GLDZM] Zone Distance Entropy;
- ADC [NGLD] Dependence Count Entropy;
- ADC [GLSZM] Zone Size Entropy;
- ADC [ID] Mean Intensity;
- FF [IH] Discretised Intensity Entropy;

- FF [NGLD] Dependence Count Entropy;
- FF [ID] Mean Intensity.

Since these features were selected through the combined application of variability filtering, redundancy reduction and embedded modelling, they were both stable in healthy tissue and informative for classification. This dual property suggests that the selected descriptors capture relevant alterations in marrow architecture and composition associated with disease. Notably, many of these features are entropy-based — quantifying the degree of spatial or distributional disorder within the ROI — and belong to distinct radiomic families: GLDZM (Grey Level Distance Zone Matrix), GLSZM (Grey Level Size Zone Matrix), NGLD (Neighbourhood Grey Level Dependence) and IH (Intensity Histogram). Their selection indicates that pathological marrow infiltration introduces structural disorganisation and heterogeneity, which manifests as increased entropy in affected regions.

5.3.3 Multiparametric Classification Modelling with ADC, FF and Radiomic Features

The final set of radiomic features described in the previous section was used to develop predictive models for classifying healthy and diseased subjects. The goal was to assess whether a multiparametric approach — combining mean intensity values and radiomic descriptors from both ADC and FF maps — could improve the classification of bone marrow infiltration in multiple myeloma compared to the performance achieved by using only ADC and FF mean intensities as done in Preliminary Results.

Classification was performed by comparing healthy subjects against the pathological classes of the dataset, using the three different feature configurations: ADC-derived features (radiomic descriptors and mean intensity), FF-derived features and a combined multiparametric set including both modalities.
To evaluate the predictive performance of each configuration, a Repeated Stratified K-Fold cross-validation strategy was applied (3 folds, 30 repetitions), ensuring balanced class distributions and robust estimation. A logistic regression model without regularisation was used, and performance was summarised by the average AUC across repetitions, with 95% confidence intervals computed via percentile bootstrapping.

The ROC curves for the classification task ”Healthy vs All Diseased” are shown in Figure 5.6, comparing the three tested feature configurations. Each curve represents the mean true positive rate across 90 validation folds and the shaded areas indicate the standard deviation. The FF-based model achieved the highest classification performance (AUC = 0.97), closely followed by the combined ADC+FF model (AUC = 0.96). The ADC-based configuration alone showed reduced discriminative power (AUC = 0.60) but still contributed complementary information in the multiparametric setting.

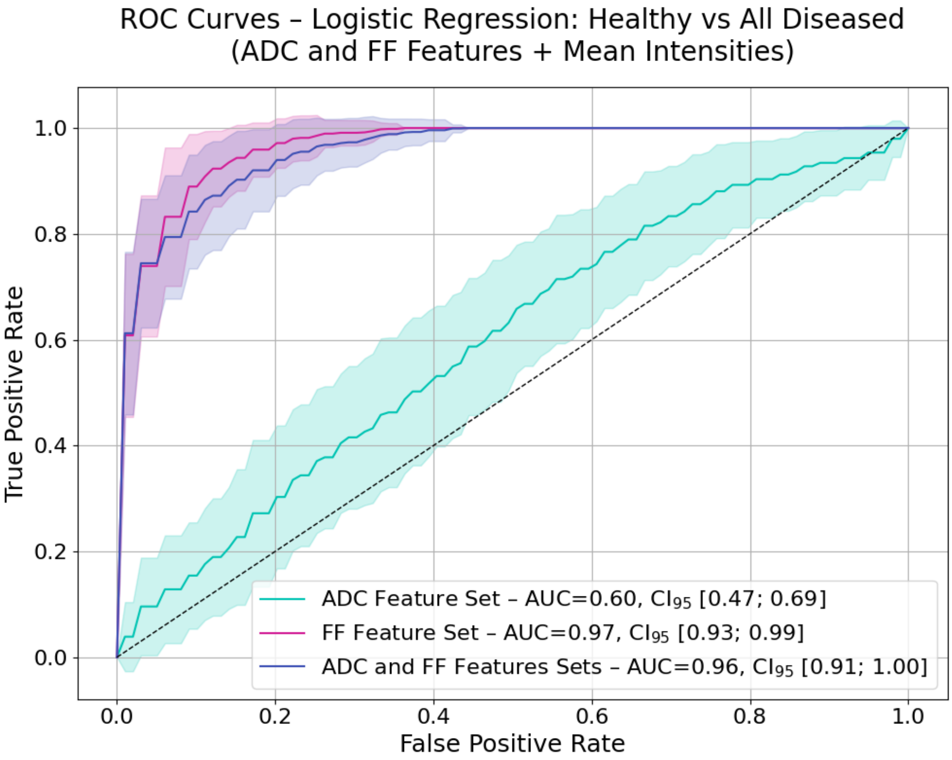


Figure 5.6: ROC curves for logistic regression models using multiparametric features from ADC and FF maps, including both mean intensities and radiomic descriptors.

Only the ”Healthy vs All Diseased” task is shown in the ROC plot for clarity and brevity: this specific comparison was selected for visualisation as it represents the most clinically relevant scenario, aiming to distinguish healthy bone marrow from any degree of pathological infiltration, without stratifying disease severity. Despite being the most challenging task — as also reflected in the lower AUC values reported in Preliminary Results — it provides a comprehensive summary of the model’s overall discriminative ability.
AUC values with 95% confidence intervals were computed for all binary classification tasks, comparing healthy subjects against a specific disease stage. The complete results are summarised in Table 5.6.

The results reported in Table 5.6 confirm the trends already observed in the analysis of Preliminary Results, with FF-derived features systematically outperforming ADC-based ones across all classification tasks. The FF feature set alone achieved remarkably high AUC values — exceeding 0.95 in all comparisons — and showed excellent discriminative power even in the most challenging cases such as “Healthy vs All Diseased”. In contrast, the ADC feature set resulted in moderate classification performance, with AUC values

	ADC Feature Set	**FF Feature Set**	**ADC and FF Feature Sets**
Comparison	**AUC [CI_{95}]**	**AUC [CI_{95}]**	**AUC [CI_{95}]**
Healthy vs SMM	0.64 [0.48; 0.82]	0.98 [0.90; 1.00]	0.91 [0.73; 1.00]
Healthy vs MM	0.68 [0.53; 0.81]	0.95 [0.89; 1.00]	0.93 [0.76; 0.99]
Healthy vs RRMM	0.73 [0.53; 0.92]	0.99 [0.91; 1.00]	0.93 [0.77; 1.00]
Healthy vs All Diseased	0.60 [0.47; 0.69]	0.97 [0.93; 0.99]	0.96 [0.91; 1.00]

Table 5.6: AUC values and 95% confidence intervals (CI_{95}) for logistic regression models using radiomic and intensity features from ADC, FF and both modalities combined.

ranging from 0.61 to 0.73 depending on the comparison. When the two modalities were combined in a multiparametric model, the performance remained comparable to that of FF alone, without significant improvement. This indicates that, although ADC descriptors may offer some complementary insight, their contribution does not appear to substantially enhance the classification performance beyond what is already achieved using FF-derived features.

The fact that the dominant predictive value is associated with the fat fraction signal was further confirmed by an additional validation based on a stratified 70/30 train/test split, ensuring that the relative proportions of the four diagnostic groups were preserved in both sets. This evaluation assesses the model's ability to generalise to previously unseen data, thereby providing an indirect check for overfitting. Even under this independent partition, models based on FF - as well as the combined configuration - mantained high classification performance, whereas ADC-based classifiers remained less reliable. The close agreement between training and test set accuracies across all binary classification tasks indicates that the models did not overfit the training data.
These results, although obtained with a simplified evaluation protocol, further support the consistency and robustness of the FF modality in detecting pathological infiltration across the different stages of multiple myeloma.

Overall, the proposed multiparametric classification approach — integrating radiomic descriptors and quantitative intensities from ADC and FF maps — demonstrated high accuracy in distinguishing healthy from diseased bone marrow. The combination of quantitative MRI and radiomic analysis provided an effective framework to characterise bone marrow alterations, supporting the potential of imaging biomarkers in non-invasive disease assessment and patient stratification.

5.4 Future Directions

The findings presented in this chapter confirm the potential of ADC and FF as Quantitative Imaging Biomarkers (QIBs) for the assessment of bone marrow involvement in multiple myeloma (MM). Both biomarkers demonstrated systematic variations across clinical stages and enabled highly accurate classification performance, particularly when combined and when radiomic descriptors were integrated. These results not only support the relevance of QIBs in clinical imaging but also reveal methodological and practical limitations, which define clear and actionable directions for future research.

A primary limitation of this study was the relatively small sample size of the myeloma patient dataset, particularly in the relapsed/refractory subgroup. Although the radiomic analysis was grounded in a solid foundation of feature stability derived from healthy pelvic bone tissue, the limited cohort size imposed constraints on the number of features that could be reliably employed. Many radiomic descriptors that demonstrated low variability in healthy tissues (CoV $\leq$ 15%) had to be excluded due to insufficient statistical power. Expanding the dataset would allow for the inclusion of a broader set of stable features, potentially enhancing predictive performance — particularly for ADC. To improve statistical robustness and model generalisability, future studies should consider multicentric data collection. The inclusion of a larger and more diverse patient population, scanned across multiple centers with standardised acquisition protocols, would be essential to assess reproducibility and to develop clinically applicable models.

Another distinguishing aspect of this work lies in its methodological design. Unlike many studies that begin directly with pathological analysis, this study originated from a systematic characterisation of healthy bone marrow. This approach allowed for the identification of intrinsically stable features, providing a biologically and statistically grounded basis for radiomic selection. Such a bottom-up strategy improves interpretability and improves the robustness and transparency of subsequent modelling.

The proposed workflow, based on reproducible radiomic extraction from fixed-volume ROIs, could be integrated with automated tools to facilitate clinical translation and scalability. In this study, segmentation was performed manually, introducing inter-operator variability and limiting overall efficiency. Future implementations could incorporate Artificial Intelligence (AI)-based solutions for automated ROI definition, feature extraction and classification. Such automation would improve reproducibility, reduce processing time and support integration into routine radiological practice.

Although this work focused on imaging-derived features, the integration of QIBs with complementary clinical, laboratory and genetic data represents a natural and necessary evolution. This type of integrative approach — known as multi-omics — combines heterogeneous data sources (such as imaging, genomics and blood biomarkers) to enable more comprehensive disease characterisation and personalised risk stratification. In future studies, the inclusion of bone marrow plasma cell percentage (BMPC), serum markers such as the free light chain (FLC) ratio and β_2-microglobulin, as well as gene expression profiles like SKY92, could significantly enhance the predictive power of radiomic models. While such parameters were not available in the current dataset, the proposed workflow is readily adaptable to multi-omic integration and could support the development of more robust and personalised decision-support tools, enabling a comprehensive characterisation of disease progression from multiple biological perspectives.

Finally, while this analysis was cross-sectional, longitudinal applications represent a particularly promising area for future development. Quantitative biomarkers such as ADC and FF could serve as dynamic indicators of therapeutic response, detecting marrow changes before conventional laboratory markers become abnormal. Incorporating QIBs into longitudinal follow-up may support early treatment adaptation and allow for more precise monitoring of minimal residual disease. Moreover, by providing image-based assessments of marrow composition and disease dynamics, these biomarkers could reduce the clinical need for repeated bone marrow biopsies, offering a less invasive alternative for disease monitoring.

In conclusion, this study provides a solid foundation for the use of QIBs in multiple myeloma imaging, demonstrating the applicability and value of radiomic modelling based on ADC and FF. Future work should prioritise dataset expansion, workflow automation, integration with multi-omic data and longitudinal validation in order to fully realise the clinical potential of quantitative imaging in this context.

Conclusions

This thesis has focused on the quantitative validation and clinical application of imaging biomarkers in 3T Whole-Body Magnetic Resonance Imaging (WB-MRI), with specific attention to the Apparent Diffusion Coefficient (ADC) and the Fat Fraction (FF).
The research followed a consistent and methodically connected framework, beginning with the technical validation of acquisition protocols on dedicated phantoms to ensure measurement accuracy and — for FF — linearity. Afterwards, the in vivo characterisation of both biomarkers in healthy tissues was carried out to establish physiological reference values in seven distinct organs, assessing repeatability across subjects and over time and comparing results with existing literature. Building on this foundation, the study then examined the behaviour of ADC and FF in pathological conditions, specifically multiple myeloma, visualising their variation across distinct clinical stages and quantifying disease-related changes. Finally, predictive models were developed to explore the potential of these biomarkers — both through mean values and through stable radiomic features — to distinguish healthy from infiltrated bone marrow directly from WB-MRI data, thereby enabling a fully non-invasive imaging-based assessment capable of supporting diagnosis, patient stratification and potentially treatment monitoring.

The technical validation phase ensured that both biomarkers met the fundamental requirements expected of quantitative imaging biomarkers, including accuracy, repeatability and, where applicable, linearity across the physiologically relevant range.
For ADC, validation on a temperature-controlled water phantom produced measurements in close agreement with known physical reference standards (ADC = 2047 ± 61 mm^2/s), confirming the accuracy of the well-established diffusion-weighted imaging protocol and requiring no further optimisation beyond this accuracy check.
For FF, a more in-depth investigation was necessary, as the quantification process is influenced by the modelling approach used in the available sequences. A custom-built fat–water emulsion phantom was therefore employed to simulate a broad range of lipid concentrations, allowing a detailed evaluation of stability, accuracy and linearity and enabling direct comparison between the simplified fat fraction estimation from the mDixonAll sequence and the more advanced multi-peak fat spectral modelling implemented in mDixonQuant. Both sequences performed well and produced comparable quantitative estimates across the tested range: linear regression against reference fat fractions yielded $R^2 = 0.94$ for mDixonAll and $R^2 = 0.98$ for mDixonQuant. Multiple linear regression further confirmed that slope and intercept were not significantly different from the identity line ($p > 0.05$), supporting quantitative agreement between the two sequences. The mDixonQuant sequence showed slightly superior linearity and narrower confidence intervals, although occasional fat–water swaps were clearly ob-

servable at very low fat fractions. Based on these results, mDixonQuant was selected for the subsequent in vivo analyses but, given the close agreement between the two methods, it was concluded that FF values from mDixonAll can be confidently used when mDixonQuant data are not available.

The in vivo characterisation was performed on a cohort of 20 female patients undergoing WB-MRI for breast cancer imaging, from whom only morphologically healthy tissues were selected to represent the healthy reference sample, yielding a total of 41 acquisitions across the cohort.

Both biomarkers were assessed across seven anatomical regions - liver, spleen, pancreas, kidney, vertebra, pelvic bone, femoral head - to establish physiological reference values and to quantify stability and repeatability across subjects, time points and regions. Repeatability was evaluated both at the cohort level and through a longitudinal single-subject analysis.

For ADC, the across-sample results showed relatively low within-subject variability in parenchymal tissues (wCV $<$ 12%) and greater variability in skeletal regions (19% $<$ wCV $<$ 38%), largely attributable to the lower signal-to-noise ratio and restricted diffusivity typical of bone marrow.

FF exhibited the opposite behaviour, with excellent repeatability in adipose-rich bone marrow regions such as the femoral head (wCV = 4 %), pelvic bone and vertebrae and lower stability in lean parenchymal organs with inherently low fat content.

The single-subject longitudinal analysis, removing the component of inter-subject biological heterogeneity, confirmed these patterns: FF remained consistently stable in skeletal regions, whereas low-fat parenchymal organs produced less constant measurements, reflecting the combined influence of technical noise and the proximity of true values to the detection threshold. Conversely, ADC values were more stable in parenchymal tissues and less in skeletal regions, consistently with the trends observed in the across-sample analysis.

The intra-organ variability analysis extended the assessment beyond repeatability, providing a detailed characterisation of both biomarkers across the seven distinct tissue types. This evaluation confirmed the opposing patterns of variability observed in the repeatability analysis, with ADC showing lower variability (i.e. lower coefficients of variation) in parenchymal organs and higher variability in skeletal regions and FF displaying the inverse trend.

Where literature reference values were available, the measured biomarker distributions demonstrated good agreement.

For ADC, all organs showed values that, although statistically different from literature references, deviated only marginally in absolute terms, within ranges attributable to protocol differences and inter-study variability. In skeletal tissues, these differences were accompanied by the expected lower diffusivity and greater heterogeneity of bone marrow, further supporting the physiological plausibility of the measurements. All the observed discrepancies remained minor in clinical terms.

For FF, this study provides the first comprehensive reports of WB-MRI measurements, as reference data are limited and not systematically published in the literature. Accordingly, comparisons with previous studies were more qualitative and focused on physiological plausibility rather than formal statistical testing. Liver values matched non-steatotic thresholds, pancreatic FF fell within ranges reported for metabolically healthy adults and spleen and kidney values were consistent with the expectation of negligible physiological fat. Vertebral marrow FF was significantly different from some published data, a discrepancy attributable to the inclusion of vertebrae from multiple spinal levels, the exclusion of morphologically abnormal vertebrae and demographic differences between cohorts. Femoral head FF closely matched reported values in healthy adults, while pelvic bone FF laid between vertebral and femoral measurements, reflecting known gradients of marrow adiposity from axial to appendicular sites.

The overall consistency of these results, together with physiologically interpretable deviations, confirmed the robustness of the acquisition and analysis workflow, while the radiomic stability analysis identified a substantial set of features with a coefficient of variation $\leq$15 % across all analysed organs for both biomarkers, providing a reliable basis for subsequent clinical modelling.

The clinical application of this work focused on Multiple Myeloma (MM), a malignant plasma cell disorder characterised by progressive bone marrow infiltration and structural disruption of the medullary environment that accounts for 1 % of all cancers, with nearly 50000 new cases each year in Europe and about 5500-6000 in Italy. In MM, as adipocyte-rich marrow is progressively replaced by densely cellular, lipid-poor tumour tissue, its microstructural and compositional properties change in a predictable way: water diffusivity increases, leading to elevated ADC values, while fat content decreases, resulting in reduced FF.

To investigate these effects, the pelvic bone was selected as the site of analysis, owing to its high clinical relevance, frequent involvement in MM and the availability of histological confirmation from diagnostic bone marrow biopsies routinely performed in this location. The study cohort comprised 20 healthy controls and 107 patients across three clinical stages — 38 smouldering myeloma (SMM), 46 active myeloma (MM) and 23 relapsed/refractory myeloma (RRMM) — enabling assessment of biomarker behaviour along the disease continuum.

Qualitative inspection of ADC and FF maps, supported by quantitative analysis, confirmed the expected trends of increasing ADC and decreasing FF with disease progression, although FF consistently demonstrated greater discriminative power. Statistical comparisons revealed that FF could significantly distinguish healthy from diseased mar-

row even when early-stage patients were included, whereas ADC achieved statistical significance primarily at advanced stages.
Predictive models based solely on mean biomarker values reflected these findings, with FF outperforming ADC across all binary classifications and combined models offering only marginal improvements. The inclusion of stable radiomic features markedly enhanced classification performance, particularly for FF-derived models, which consistently achieved AUC values greater than 0.95 in all pairwise group comparisons, including the most challenging Healthy versus All Diseased task. ADC-based models showed moderate performance (AUC between 0.60 and 0.73) but retained complementary value in multiparametric configurations. Model robustness was further confirmed on independent test sets, indicating the absence of overfitting and reinforcing the potential of this framework as a non-invasive, image-based approach able to reliably distinguish healthy from diseased bone marrow.

A notable methodological strength of this work lies in its bottom-up design: by starting with a rigorous technical and in vivo characterisation of healthy tissue, feature selection for clinical modelling was grounded in demonstrated stability, improving interpretability and reproducibility. The standardised ROI-based pipeline further supported the validity of cross-cohort comparisons. Limitations included the relatively small patient cohort, especially in the relapsed/refractory subgroup, manual ROI definition introducing potential inter-operator variability and minor differences in acquisition protocols between controls and patients, although both were technically comparable and clinically validated.

Future research should prioritise expanding the dataset through multicentre collaborations, incorporating automated ROI definition and feature extraction and integrating imaging biomarkers with clinical, laboratory and genomic data to enable multi-omic characterisation of disease. Longitudinal studies will be essential to evaluate ADC and FF as dynamic indicators of therapeutic response and minimal residual disease, potentially reducing the need for invasive bone marrow biopsies.

Overall, this work has demonstrated that ADC and FF can be robustly quantified in WB-MRI following careful technical validation and in vivo characterisation. FF emerges as the more stable and discriminative biomarker for detecting bone marrow infiltration in MM, while ADC retains complementary value in multiparametric frameworks. By combining rigorous methodological groundwork with clinical application, the study provides a reproducible framework for the development of QIB-based radiomic models, supporting the translation of these biomarkers into non-invasive tools for diagnosis, staging and potentially treatment monitoring in multiple myeloma.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Mauro Villa, for his interest in this work and for the valuable guidance provided during the writing of this thesis. I am also thankful to my co-supervisor, Dr. Roberto Spighi, for making this experience possible and for his continuous support and encouragement throughout my internship. Reaching the end of my academic journey and seeing the work accomplished finally brought together in this thesis make me feel profoundly grateful and satisfied with the choices I made, even when they were not always the easiest path to follow.

I am deeply grateful to the IRST “Dino Amadori” of Meldola for hosting my research, providing me with all the necessary support and making me feel part of its ongoing scientific projects. No words can express how grateful I am to the Medical Physics Unit, which welcomed me warmly and was always available to guide me, answer my questions and clarify any doubts. Being so encouraged and valued from the very first day made me feel truly in the right place.

I wish to thank Dr. Anna Sarnelli for giving me the opportunity to carry out this work, for recognising its value and for providing valuable scientific insights that contributed to the quality and growth of this research.

My heartfelt thanks go to Dr. Giacomo Feliciani (Jack), to whom I owe my deepest thanks. Your incredible support was essential not only for following every step of my research, but also for believing in me and providing endless guidance and encouragement. I truly saw in you the figure of a teacher: someone who generously shared his knowledge, valued my ideas and contributions and engaged with me in genuine scientific dialogues. I will never be able to express enough gratitude for such a constructive and enriching experience, both personally and academically.

There have been many times, both while writing this thesis and beyond, I believed I could not make it. I have been immensely fortunate, however, to have people beside me who reminded me otherwise. Even while holding on to my own belief, I hope that those around me could proudly see that they were right all along. A special thank you to all the people in my life who have been there, even when they had no obligation to be, and who chose to stay.

Appendix A

Radiomic Features Extracted with SIBEX

This appendix provides a comprehensive list of the radiomic features extracted from the WB-MRI acquisitions analysed in this study. Feature extraction was performed using SIBEX (Standardised Imaging Biomarker Extraction), a MATLAB®-based radiomics platform developed in accordance with the Image Biomarker Standardisation Initiative (IBSI) guidelines[44, 45].

A feature is a numerical descriptor extracted from an image that summarises a property of the signal within a given region such as its intensity, texture or spatial organisation. These values are computed algorithmically from the voxel data and are designed to capture underlying biological or structural characteristics in a reproducible way.
Each feature in this study quantifies a specific aspect of signal distribution or spatial heterogeneity within the manually defined regions of interest (ROIs), enabling the quantitative characterisation of anatomical structures and supporting comparisons between healthy and pathological tissues.

The extracted features are grouped into ten major IBSI-defined categories, each based on a different mathematical model or statistical formulation. These include first-order intensity statistics, texture matrices (e.g. GLCM, GLRLM, GLSZM), histogram-based measures, intensity-volume histograms and local or global intensity peaks.

The extraction process, implemented using the SIBEX platform, followed the IBSI guidelines for feature computation and terminology. For intensity- and texture-based features requiring discretisation, a fixed bin number of 32 was used with full-image rescaling. Other categories, such as Local Intensity Features and Intensity Direct statistics, were computed directly from the continuous image data without discretisation. Outlier exclusion was applied using a 3σ threshold on voxel intensities within each ROI.

The SIBEX software provides a graphical user interface in MATLAB®, through which the user first loads the imaging datasets and the corresponding ROI definitions, typically in DICOM RTSTRUCT format. Once the images and structures are registered, the user saves them as an internal dataset, which is later processed using the predefined feature set. In this study, this pipeline was applied consistently to all ADC and FF maps across all subjects and timepoints.

These extraction steps yielded a resulting dataset comprising exactly 144 radiomic features per ROI for each image modality (ADC and FF). The extracted features were subsequently used in the statistical analyses presented in Quantitative ADC Analysis in Healthy Tissues, Quantitative FF Analysis in Healthy Tissues and Clinical Application of ADC and FF in Multiple Myeloma to assess repeatability, intra-organ variability and feature stability in healthy tissues and pathological alterations in multiple myeloma patients.

A detailed listing of all extracted features, organised by category and accompanied by a description of the mathematical basis behind each group, is presented below[84, 85].

Grey Level Co-occurrence Matrix (GLCM) The grey level co-occurrence matrix (GLCM) describes the frequency with which combinations of discretised grey levels occur between neighbouring voxels, measured along a specific spatial direction. A separate GLCM is computed for each direction vector and the matrix elements define a joint probability distribution that serves as the basis for texture feature calculation.

> Features (n = 24): Joint Maximum, Joint Average, Joint Variance, Joint Entropy, Difference Average, Difference Variance, Difference Entropy, Sum Average, Sum Variance, Sum Entropy, Angular Second Moment, Contrast, Dissimilarity, Inverse Difference, Normalised Inverse Difference, Inverse Difference Moment, Normalised Inverse Difference Moment, Inverse Variance, Correlation, Autocorrelation, Cluster Tendency, Cluster Shade, Cluster Prominence, Information Correlation 1, Information Correlation 2.

Grey Level Distance Zone Matrix (GLDZM) The grey level distance zone matrix (GLDZM) counts the number of groups (or zones) of connected voxels that share a specific discretised grey level and are located at the same distance from the ROI boundary. This matrix captures the relationship between spatial location and grey level, highlighting structural homogeneity with respect to distance from the edge.

> Features (n = 16): Small Distance Emphasis, Large Distance Emphasis, Low Grey Level Zone Emphasis, High Grey Level Zone Emphasis, Small Distance Low Grey Level Emphasis, Small Distance High Grey Level Emphasis, Large Distance Low Grey Level Emphasis, Large Distance High Grey Level Emphasis, Grey Level Non-uniformity, Normalised Grey Level Non-uniformity, Zone Distance Non-uniformity, Normalised Zone Distance Non-uniformity, Zone Percentage, Grey Level Variance, Zone Distance Variance, Zone Distance Entropy.

Grey Level Run Length Matrix (GLRLM) The grey level run length matrix (GLRLM) is a statistical representation that quantifies the length of consecutive sequences — known as runs — of voxels sharing the same discretised grey level along a given direction. Unlike GLCM, which measures co-occurrence between neighbouring voxels, GLRLM captures the distribution and frequency of uniform grey-level runs, providing information on texture smoothness and directional patterns.

> Features (n = 16): Short Runs Emphasis, Long Runs Emphasis, 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, Long Run High Grey Level Emphasis, Grey Level Non-uniformity, Normalised Grey Level Non-uniformity, Run Length Non-uniformity, Normalised Run Length Non-uniformity, Run Percentage, Grey Level Variance, Run Length Variance, Run Entropy.

Grey Level Size Zone Matrix (GLSZM) The grey level size zone matrix (GLSZM) quantifies the number of homogeneous zones within a region, where each zone is defined as a group of connected voxels sharing the same discretised grey level. In 3D images, voxel connectivity is typically assessed using a 26-neighbourhood criterion. GLSZM features reflect the size and distribution of these zones, capturing information about texture granularity and spatial uniformity.

> Features (n = 16): Small Zone Emphasis, Large Zone Emphasis, Low Grey Level Zone Emphasis, High Grey Level Zone Emphasis, Small Zone Low Grey Level Emphasis, Small Zone High Grey Level Emphasis, Large Zone Low Grey Level Emphasis, Large Zone High Grey Level Emphasis, Grey Level Non-uniformity, Normalised Grey Level Non-uniformity, Zone Size Non-uniformity, Normalised Zone Size Non-uniformity, Zone Percentage, Grey Level Variance, Zone Size Variance, Zone Size Entropy.

Intensity Direct (ID) These first-order statistical features describe the distribution of voxel intensities within the region of interest (ROI), without requiring discretisation. They characterise properties such as central tendency, dispersion and shape of the intensity histogram. This family of features is particularly relevant when the intensity scale carries meaningful physical information, as in quantitative parametric maps like ADC or FF.

> Features (n = 18): Mean Intensity, Intensity Variance, Intensity Skewness, (Excess) Intensity Kurtosis, Median Intensity, Minimum Intensity, 10th Intensity Percentile, 90th Intensity Percentile, Maximum Intensity, Intensity Interquartile Range, Intensity Range, Intensity-based Mean Absolute Deviation, Intensity-based Robust Mean Absolute Deviation, Intensity-based Median Absolute Deviation, Intensity-based Coefficient of Variation, Intensity-based Quartile Coefficient of Dispersion, Intensity-based Energy, Root Mean Square Intensity.

Intensity Histogram (IH) This family of features is derived from the intensity histogram of the ROI, which is generated by discretising the original continuous intensity values into a fixed number of bins. These features provide statistical measures that describe the shape, spread and uniformity of the discretised intensity distribution, offering insights into tissue composition and heterogeneity.

Features (n = 23): Mean Discretised Intensity, Discretised Intensity Variance, Discretised Intensity Skewness, (Excess) Discretised Intensity Kurtosis, Median Discretised Intensity, Minimum Discretised Intensity, 10th Discretised Intensity Percentile, 90th Discretised Intensity Percentile, Maximum Discretised Intensity, Discretised Intensity Interquartile Range, Discretised Intensity Range, Intensity Histogram Mean Absolute Deviation, Intensity Histogram Robust Mean Absolute Deviation, Intensity Histogram Median Absolute Deviation, Intensity Histogram Coefficient of Variation, Intensity Histogram Quartile Coefficient of Dispersion, Discretised Intensity Entropy, Discretised Intensity Uniformity, Maximum Histogram Gradient, Maximum Histogram Gradient Intensity, Minimum Histogram Gradient, Minimum Histogram Gradient Intensity.

Intensity Volume Histogram (IVH) The cumulative intensity-volume histogram (IVH) describes the relationship between discretised intensity levels and the proportion of the ROI volume that contains voxel values greater than or equal to each intensity threshold. IVH features provide information about signal distribution in terms of both intensity and spatial extent and are particularly relevant when evaluating heterogeneity in parametric maps.

Features (n = 7): Volume at Intensity Fraction 10, Volume at Intensity Fraction 90, Intensity at Volume Fraction 10, Intensity at Volume Fraction 90, Volume Fraction Difference between Intensity Fractions, Intensity Fraction Difference between Volume Fractions, Area Under the IVH Curve.

Local Intensity Features (LIF) Local intensity features are computed by analysing voxel intensities within a defined neighbourhood surrounding each centre voxel. Unlike most other feature families, these features are not limited to values strictly within the ROI: although only voxels inside the ROI can serve as centre voxels, their neighbourhoods may include voxels outside the ROI boundary. This approach allows for the assessment of local contrast and peak intensities across boundaries.

Features (n = 2): Local Intensity Peak, Global Intensity Peak.

Neighbourhood Grey Level Dependence (NGLD) The neighbourhood grey level dependence matrix (NGLDM) provides an alternative approach to texture analysis compared to the co-occurrence matrix. It quantifies the number of neighbouring voxels that depend on a central voxel, based on grey level similarity within a defined neighbourhood. NGLDM features are rotationally invariant and aim to characterise texture coarseness. All voxels within the ROI are evaluated as centre voxels, with their neighbourhoods contributing to the overall matrix.

Features (n = 17): Low Dependence Emphasis, High Dependence Emphasis, Low Grey Level Count Emphasis, High Grey Level Count Emphasis, Low Dependence Low Grey Level Emphasis, Low Dependence High Grey Level Emphasis, High Dependence Low Grey Level Emphasis, High Dependence High Grey Level Emphasis, Grey Level Non-uniformity, Normalised Grey Level Non-uniformity, Grey Level Variance, Normalised Dependence Count Non-uniformity, Dependence Count Non-uniformity, Dependence Count Percentage, Dependence Count Variance, Dependence Count Entropy, Dependence Count Energy.

Neighbourhood Intensity Difference (NID) Neighbour intensity difference (NID) features are derived from the neighbourhood grey tone difference matrix (NGTDM), which provides an alternative to the co-occurrence matrix for describing texture. The matrix records the sum of absolute differences between the intensity of a voxel with a certain grey level and the average grey level of its neighbours, within a specified distance. A neighbourhood is considered valid if at least one neighbouring voxel lies within the ROI mask. These features capture local contrast, complexity and coarseness.

Features (n = 5): Coarseness, Contrast, Busyness, Complexity, Strength.

Appendix B

Statistical Measures

In the evaluation of Quantitative Imaging Biomarkers (QIBs) such as the Apparent Diffusion Coefficient (ADC) and Fat Fraction (FF), it is essential to employ statistical metrics that ensure the reliability, comparability and interpretability of measurements across repeated acquisitions, sessions and subjects.
The Quantitative Imaging Biomarkers Alliance (QIBA), promoted by the Radiological Society of North America (RSNA), provides technical recommendations and methodological standards to evaluate and validate such biomarkers. These recommendations are intended to minimise variability in quantitative measurements and to enhance the reliability of results in both clinical practice and research applications[46, 47, 48].

A Quantitative Imaging Biomarker (QIB) is defined as “an objective characteristic derived from an in vivo image measured on a ratio or interval scale as an indicator of normal biological processes, pathogenic processes, or a response to a therapeutic intervention”[86]. In order to meet the criteria of a QIB, a biomarker is required to generate quantitative data that are robust in terms of both technical performance and clinical applicability.

To be technically valid and clinically useful, a QIB must demonstrate key measurement properties including accuracy, precision, linearity and reliability[48].
Accuracy refers to the degree of closeness between a measured value and the true or accepted reference value, reflecting the level of systematic bias in the imaging process. In this study, accuracy was evaluated through phantom-based measurements: a temperature-controlled water phantom was used to verify the ADC values and a fat–water phantom provided reference standards for FF quantification.
Precision, in contrast, reflects the variability of repeated measurements under the same conditions and is concerned with random error rather than systematic bias.
Linearity refers to the ability of the imaging method to produce results that are directly proportional to the true values across a range of measurements.
Reliability is a more general concept that integrates these properties to express the confidence with which a QIB can be interpreted.

Within the framework of precision and technical validation of QIBs, two important components are repeatability and reproducibility.
Repeatability describes the degree of agreement among repeated measurements taken under identical conditions (same measurement procedure, same subject, same scanner and same operator) within short time intervals.

Reproducibility, on the other hand, accounts for the variability introduced when one or more of these conditions change, such as the imaging system, site or operator[48].

The present work specifically focuses on repeatability, as all measurements were acquired under tightly controlled and standardised conditions. This includes the use of consistent imaging protocols, scanner hardware and operator procedures. Although the analysis is centred on repeatability, it should be noted that the imaging protocol did not adhere to a strict test–retest design as defined by QIBA: rather than performing repeated acquisitions within the same imaging session, the measurements were obtained across separate time points. This approach was chosen to reflect a more realistic clinical scenario, allowing for a broader assessment of short-term variability while maintaining stable acquisition conditions.
The analyses presented in Quantitative ADC Analysis in Healthy Tissues and Quantitative FF Analysis in Healthy Tissues were carried out to quantify measurement variability in the absence of true biological change, finally characterising physiological ADC and FF values in healthy tissues.

The following notations and statistical measures were used in this work to quantify repeatability, variability and measurement precision of imaging biomarkers, in agreement with QIBA protocols and international standards in quantitative imaging metrology[47].

Within-Subject Standard Deviation (σ_w) The within-subject (or intra-subject) standard deviation, denoted as σ_w, represents the core measure of repeatability in quantitative imaging. It reflects the variability observed when the same subject is measured multiple times under identical conditions. It is used to characterise the random error inherent to the measurement process and serves as the basis for calculating other precision-related metrics. σ_w is typically estimated from repeated measures designs or analysis of variance models by isolating the within-subject component of variance and it is expressed in the same unit as the analysed biomarker. In general, when σ_w is small relatively to the measured value or the expected biomarker differences between normal and abnormal tissues, it is reasonable to assume reliability with confidence.

Between-Subject Standard Deviation (σ_b) The between-subject standard deviation (σ_b) quantifies the natural variability observed across different subjects in a given population. It reflects biological differences as well as potential sources of technical variation that persist despite standardisation. This measure captures the inter-subject component of variance and its accurate estimation enables meaningful interpretation of how well a biomarker discriminates between individuals beyond random noise.

Repeatability Coefficient (RC) The Repeatability Coefficient (RC) quantifies the maximum expected difference between two repeated measurements on the same subject under identical conditions, with a 95% confidence level. It reflects the combined influence of random noise and intra-subject variability in a single number and is particularly useful to assess whether a change in biomarker value exceeds the expected measurement error, providing a statistical threshold: any difference between two measurements exceeding the RC is unlikely to be attributable to measurement variability alone.
The RC is calculated from the within-subject standard deviation, assuming normally distributed measurement differences and is given by

$$\mathrm{RC} = 1.96 \cdot \sqrt{2\sigma_w^2} = 2.77 \cdot \sigma_w \ . \tag{B.1}$$

Within-subject Coefficient of Variation (wCV) The within-subject coefficient of variation (wCV) is a dimensionless metric that expresses intra-subject variability relative to the mean value of the measurement. It is particularly useful when comparing biomarkers with different units or magnitudes, as it normalises variability across scales. The wCV is calculated as the ratio of the within-subject standard deviation to the mean of the measurements (μ), according to

$$\mathrm{wCV} = \frac{\sigma_w}{\mu} \ . \tag{B.2}$$

This allows for a percentage-based expression of precision that is independent of the absolute values being measured. In this study, wCV values are reported as percentages and rounded to the nearest integer for clarity and consistency across all biomarkers and anatomical locations.

Interquartile Range (IQR) The interquartile range (IQR) is a non-parametric measure of statistical dispersion based on the division of a dataset into quartiles. The first quartile (Q1) corresponds to the 25th percentile and marks the value below which 25% of the observations fall, while the third quartile (Q3) corresponds to the 75th percentile, indicating the point below which 75% of the data lie. The IQR represents the range within which the central 50% of the data are distributed and is defined as

$$\mathrm{IQR} = \mathrm{Q3} - \mathrm{Q1} \ . \tag{B.3}$$

Unlike standard deviation, the IQR is robust to outliers and does not assume any specific distribution of the data. In this work, the IQR was used as a descriptive measure of variability in the intra-organ analysis, by summarising the spread of biomarker values across all acquisitions and subjects without making distributional assumptions.

Coefficient of Variation (CoV) The coefficient of variation (CoV) is a relative measure of dispersion defined as the ratio between the standard deviation (σ) and the mean value (μ) of a dataset:

$$\mathrm{CoV} = \frac{\sigma}{\mu}\,. \tag{B.4}$$

The higher the coefficient of variation, the greater the level of dispersion around the mean. It is generally expressed as a percentage and, being dimensionless, it enables comparison between distributions of values that have different units or scales. When used to characterise estimated quantities, a lower CoV indicates greater precision of the estimate[87].

In this study, the CoV was not used in the repeatability analysis but instead to characterise the general behaviour of the biomarkers within each organ by aggregating all acquisitions from all subjects as independent observations. This allowed for a descriptive assessment of biomarker stability across the population, providing insight into the variability of measurements in a broader, clinically relevant setting.

For consistency with wCV, in all the analyses CoV was expressed as a percentage, rounded to the nearest integer.

Intraclass Correlation Coefficient (ICC) The intraclass correlation coefficient (ICC) is a statistical measure of repeatability that quantifies the consistency of repeated measurements relative to the total variability in the population, commonly used in quantitative imaging to evaluate the stability of biomarker values under repeated acquisitions. It is defined as the proportion of total variance that is not attributable to measurement error and provides an aggregate assessment of measurement reliability across subjects. Mathematically, it is expressed as

$$\mathrm{ICC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}\,. \tag{B.5}$$

Although the ICC is often interpreted as a measure of relative variance, high values may not always reflect high measurement precision. When the between-subject variance (σ_b^2) greatly exceeds the within-subject variance (σ_w^2), the ICC can approach 1 even in the presence of considerable intra-subject noise. Therefore, ICC values should always be interpreted cautiously, evaluating intra and inter-subject variances first.

The following sections describe a set of additional methods used to support the interpretation and quality of the statistical analyses. Although most of these tools are referenced within the QIBA framework[47], they are not considered statistical metrics in the strict sense, but rather graphical methods, confidence intervals or technical procedures that complement the core measures of repeatability and variability.

95% Confidence Interval (CI_{95}) The 95% confidence interval (CI_{95}) provides a range within which the true value of a population parameter is expected to lie with 95% probability, based on sample data. It reflects the uncertainty associated with the estimation of a mean or other summary statistic.
In this work, CI_{95} was not part of the QIBA-defined precision metrics, but was calculated to support the comparison between the values obtained in this study and those reported in the literature.
Given the relatively large sample sizes and the use of sample standard deviations, the confidence intervals were estimated approximating the t-distribution using to the normal approximation, as justified by the Central Limit Theorem. .
Specifically, the 95% confidence interval for the sample mean was computed as

$$\mathrm{CI}_{95} = \mu \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} \, . \tag{B.6}$$

where μ is the sample mean, σ is the sample standard deviation and n is the number of independent observations[88].

Limits of Agreement (LoAs) According to QIBA[47], the limits of agreement (LoAs) represent the interval within which the difference between two repeated measurements under repeatability conditions for a randomly selected subject is expected to lie 95% of the time. In this framework, the LoA are assumed to be symmetric and centred around zero and are expressed as

$$\mathrm{LoA} = [\text{-RC}; \text{+RC}] \, . \tag{B.7}$$

This formulation implicitly assumes that there is no systematic bias between repeated measurements, and that the expected difference between them is zero.

In this study, the LoA were instead estimated using the standard Bland–Altman method, which explicitly accounts for any systematic bias between paired measurements. This approach allows for asymmetry around the origin and directly incorporates the empirical mean and standard deviation of the observed differences. The general equation is defined as

$$\mathrm{LoA} = \bar{d} \pm 1.96 \cdot \sigma_d \tag{B.8}$$

where $\bar{d}$ is the mean difference and σ_d is the standard deviation of the differences between the two measurements[49].
This formulation provides a more realistic characterisation of measurement agreement, particularly when small biases are present and was therefore adopted throughout the analysis presented in this work.

Bland–Altman Plots Bland–Altman plots are graphical tools used to assess the agreement between two repeated measurements of the same variable. Each point on the plot represents a subject, with the x-axis showing the mean of the two measurements and the y-axis showing their difference. This allows for the visual inspection of potential systematic bias, magnitude-dependent variability and outliers.

In the context of quantitative imaging, Bland–Altman plots are particularly valuable for identifying trends in measurement differences across the range of biomarker values. The central horizontal line corresponds to the mean difference ($\bar{d}$) and the lower and upper lines represent the 95% LoAs ($\mathrm{LoA}_{\mathbf{L}}$ and $\mathrm{LoA}_{\mathbf{U}}$).

This graphical representation complements the numerical LoA values and helps reveal whether measurement differences are consistent or vary depending on biomarker magnitude. In an ideal scenario, the differences should be randomly scattered around the mean difference, without any discernible trend and the majority of points should lie within the limits of agreement. A narrow spread of differences and a clustering of points close to the zero line typically indicate good measurement agreement and low variability across the range of values.

In this work, Bland–Altman plots were used alongside the numerical LoAs to visually evaluate the agreement between repeated acquisitions and to support the detection of systematic trends or inconsistencies in the biomarker measurements.

Outlier Detection and Removal Outliers are data points that deviate substantially from the rest of the dataset and can distort statistical estimates, particularly in small or moderately sized samples. In quantitative imaging studies, such extreme values may arise from technical artifacts, segmentation errors or unmodelled physiological variability. For this reason, outlier removal should be applied prior to statistical analysis to ensure the robustness and interpretability of the results.

In QIBA guidelines, the identification of outliers is based on the interquartile range (IQR). Specifically, a data point is considered an outlier if it satisfies the condition

$$\mathrm{QIB}_{\mathbf{outlier}} < \mathrm{Q1} - 1.5 \cdot \mathrm{IQR} \quad \text{or} \quad \mathrm{QIB}_{\mathbf{outlier}} > \mathrm{Q3} + 1.5 \cdot \mathrm{IQR}\,. \tag{B.9}$$

This criterion is adopted in exploratory data analysis after visual inspection of the data distributions, in order to ensure that the exclusion of values is justified and not the result of natural variation or expected heterogeneity within the dataset.

Outlier filtering was applied independently to each biomarker and at different stages of the analysis for ADC and FF, depending on the context. In all cases, the exclusion of outliers was explicitly reported in the corresponding sections of this work to ensure transparency and reproducibility.

References

[1] F. Bloch et al. *Nuclear Induction*. In: *Physical Review* 69.127 (1946).

[2] E. M. Purcell et al. *Resonance Absorption by Nuclear Magnetic Moments in a Solid*. In: *Physical Review* 69.37 (1946).

[3] R. W. Brown. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley & Sons, Ltd, 1999.

[4] P. Sprawls. *Magnetic Resonance Imaging: Principles, Methods, and Techniques*. Medical Physics Publishing, 2000, pp. 39–42.

[5] Case Western Reserve University. *Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics*. URL: `https://case.edu/med/neurology/NR/MRI%20Basics.htm` (accessed on 06/23/2025).

[6] Lumen Learning. *Body Plan and Organization: Anatomical Location*. URL: `https://courses.lumenlearning.com/suny-dutchess-anatomy-physiology/chapter/practice-anatomical-location/` (accessed on 08/19/2025).

[7] V. Gulani and N. Seiberlich et al. *Advances in Magnetic Resonance Technology and Applications*. Vol. 1. Academic Press, 2020, pp. xxxvii–lv.

[8] M. A. Islam. *Einstein–Smoluchowski Diffusion Equation: A Discussion*. In: *Physica Scripta* 70.2-3 (2004), p. 120.

[9] U. Bashir, Y. Yap, and J. Feeney et al. *Diffusion-weighted imaging*. 2012. URL: `https://radiopaedia.org/articles/diffusion-weighted-imaging-2` (accessed on 06/22/2025).

[10] E. O. Stejskal and J. E. Tanner. *Spin Diffusion Measurements: Spin Echoes in the Presence of a Time Dependent Field Gradient*. In: *The Journal of Chemical Physics* 42.1 (1965), pp. 288–292.

[11] D Le Bihan D and E. Breton et al. *Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging*. In: *Radiology* 168.2 (1988), pp. 497–505.

[12] J. H. Jensen and J. A. Helpern et al. *Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging*. In: *Magn Reson Med.* 53.6 (2005), pp. 1432–40.

[13] J. Cohen-Adad and C. A. M. Wheeler-Kingshott. *Quantitative MRI of the Spinal Cord*. Academic Press, 2014. DOI: `https://doi.org/10.1016/C2011-0-07594-4`.

[14] P. Pezeshk, A. Alian, and A. Chhabra. *Role of chemical shift and Dixon based techniques in musculoskeletal MR imaging*. In: *European Journal of Radiology* 94 (2017), pp. 93–100.

[15] J. Ma. *Dixon Techniques for Water and Fat Imaging*. In: *J Magn Reson Imaging* 28 (2008), pp. 543–558.

[16] W. T. Dixon. *Simple proton spectroscopic imaging*. In: *Radiology* 53.1 (1984), pp. 189–94.

[17] H. H. Hu and H. E. Kan. *Quantitative proton MR techniques for measuring fat*. In: *NMR Biomed* 26.12 (1984), pp. 1609–29.

[18] Y. N. Azma, N.Boci, and K. Abramowicz et al. *Influence of imaging method on fat fraction estimation for assessing bone marrow in metastatic prostate cancer*. In: *Eur Radiol* (2025). DOI: `10.1007/s00330-025-11564-7`.

[19] D. Hernando, Z. P. Liang, and P. Kellman. *Chemical shift-based water/fat separation: a comparison of signal models*. In: *Magn Reson Med* 64.3 (2010), pp. 811–22.

[20] T.G. Perkins et al. *Philips Whitepaper: The next generation fat-free imaging*. URL: `https://www.documents.philips.com/assets/20180126/1e42bcf9b1934fc597fea87400c078a8.pdf` (accessed on 06/23/2025).

[21] H. Heggers and B. Brendel et al. *Dual-echo Dixon imaging with flexible choice of echo times*. In: *Magn Reson Med* 65.1 (2011), pp. 96–107.

[22] G. Hamilton and T. Yokooa et al. *In vivo characterization of the liver fat 1H MR spectrum*. In: *NMR Biomed* 24.7 (2011), pp. 784–90.

[23] U. Bashir, A. Murphy, and Y. Glick et al. *Gradient echo sequences*. 2012. URL: `https://radiopaedia.org/articles/gradient-echo-sequences-1` (accessed on 06/25/2025).

[24] P. Summers, G. Saia, and A. Colombo et al. *Whole-body magnetic resonance imaging: technique, guidelines and key applications*. In: *Ecancermedicalscience* 15 (2021), p. 1164.

[25] S. Ahlawat, P. Debs, and B. Amini et al. *Clinical Applications and Controversies of Whole-Body MRI: AJR Expert Panel Narrative Review*. In: *AJR Am J Roentgenol* 220.4 (2023), pp. 463–475.

[26] F. E. Lecouvet, C. Chabot, and L. Taihi et al. *Present and future of whole-body MRI in metastatic disease and myeloma: how and why you will do it*. In: *Skeletal Radiol* 53.9 (2024), pp. 1815–1831.

[27] A. Cattabriga, B. Renzetti, and F. Galuppi et al. *Multiparametric Whole-Body MRI: A Game Changer in Metastatic Prostate Cancer*. In: *Cancers (Basel)* 16.14 (2024), p. 2531.

[28] C. Messiou, J. Hillengass, and S. Delorme et al. *Guidelines for Acquisition, Interpretation, and Reporting of Whole-Body MRI in Myeloma: Myeloma Response Assessment and Diagnosis System (MY-RADS)*. In: *Radiology* 291.1 (2019), pp. 5–13.

[29] A. R. Padhani, F. E. Lecouvet, and N. Tunariu et al. *METastasis Reporting and Data System for Prostate Cancer: Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-body Magnetic Resonance Imaging-based Evaluations of Multiorgan Involvement in Advanced Prostate Cancer*. In: *European Urology* 71.1 (2017), pp. 81–92.

[30] N. W. C. J. Van Der Donk, C Pawlyn, and K. L. Yong. *Multiple myeloma*. In: *Lancet* 397.10272 (2021), pp. 410–427.

[31] R. A. Kyle and S. V. Rajkumar. *Multiple myeloma*. In: *Lancet* 111.6 (2008), pp. 2962–72.

[32] A. Mafra, M. Laversanne, and R. Marcos-Gragera et al. *The global multiple myeloma incidence and mortality burden in 2022 and predictions for 2045*. In: *J Natl Cancer Inst* 117.5 (2025), pp. 907–914.

[33] Osservatorio Malattie Rare. *Mieloma multiplo, in Italia si stimano circa 6.000 nuovi casi all’anno*. 2020. URL: `https://www.osservatoriomalattierare.it/i-tumori-rari/mieloma-multiplo/16849-mieloma-multiplo-in-italia-si-stimano-circa-6-000-nuovi-casi-all-anno#:~:text=UNA%20MALATTIA%20SEMPRE%20PIU’%20DIFFUSA,all’armamentario%20terapeutico%20oggi%20disponibile.` (accessed on 08/26/2025).

[34] Myeloma Patients Europe. *Myeloma and AL Amyloidosis - What is myeloma*. 2022. URL: `https://www.mpeurope.org/about-myeloma/what-is-myeloma/?_gl=1*1ng3q48*_up*MQ..*_gs*MQ..&gclid=Cj0KCQjwpf7CBhCfARIsANIETVogd5lbz0JCrYj3AFYf66CsgFdHKiBevJ4sG3CAtDcKoUllKK9HlLcaAq9BEALw_wcB&gbraid=0AAAAA-PX-QFgfaue5RLpVDu-zR72e5nro` (accessed on 06/29/2025).

[35] S. V. Rajkumar. *Multiple myeloma: 2024 update on diagnosis, risk-stratification, and management*. In: *Am J Hematol* 99.9 (2024), pp. 1802–1824.

[36] S. V. Rajkumar, M. A. Dimopoulos, and A. Palumbo et al. *International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma*. In: *Lancet Oncol* 15.12 (2014), pp. 538–548.

[37] M. Wennmann, W. Ming, and F. Bauer et al. *Prediction of Bone Marrow Biopsy Results From MRI in Multiple Myeloma Patients Using Deep Learning and Radiomics*. In: *Invest Radiol* 58.10 (2023), pp. 754–765.

[38] G. Feliciani, E. Loi, and E. Amadori et al. *Radiopsy: a prospective observational study on quantitative multiparametric whole-body magnetic resonance imaging to discriminate between smoldering and multiple myeloma*. In: *Haematologica* (2025). DOI: `10.3324/haematol.2025.287563`.

[39] Philips. *Ingenia 3.0T*. Best, The Netherlands. URL: `https://www.philips.it/healthcare/product/HC781342/sistema-rm-ingenia-30t#specifications` (accessed on 07/01/2025).

[40] Philips. *Smartpath MR system upgrade to dStream*. Best, The Netherlands. URL: `https://www.documents.philips.com/assets/20170523/28a6689771014ce1b5f8a77c014ac83c.pdf` (accessed on 07/01/2025).

[41] Philips. *MR MobiView: Automatic review of total body MR data*. Best, The Netherlands. URL: `https://www.usa.philips.com/healthcare/product/NICA703/mr-mobiview-automatic-review-of-total-body-mr-data` (accessed on 07/01/2025).

[42] L. Fedeli, G. Belli, and A. Ciccarone. *Dependence of apparent diffusion coefficient measurement on diffusion gradient direction and spatial position – A quality assurance intercomparison study of forty-four scanners for quantitative diffusion-weighted imaging*. In: *Physica Medica* 55 (2018), pp. 135–141.

[43] MIM Software Inc. *Mim Software Platform*. Cleveland, OH, USA. 2003. URL: `https://www.mimsoftware.com` (accessed on 07/03/2025).

[44] A. Zwanenburg A, M. Vallières, and A. Abdalah et al. *The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping*. In: *Radiology* 295.2 (2020), pp. 328–338.

[45] A. Bettinelli, M. Branchini, and F. De Monte et al. *Technical Note: An IBEX adaption toward image biomarker standardization*. In: *Med Phys* 47.3 (2020), pp. 1167–1173.

[46] M. A. Boss, D. Malyarenko, and S. Partridge et al. *The QIBA Profile for Diffusion-Weighted MRI: Apparent Diffusion Coefficient as a Quantitative Imaging Biomarker*. In: *Radiology* 313.1 (2024), e233055.

[47] QIBA Technical Performance Working Group. *Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment*. In: *Stat Methods Med Res* 24.1 (2015), pp. 27–67.

[48] A. Shukla-Dave, N. A. Obuchowski, and T. L. Chenevert et al. *Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials*. In: *J Magn Reson Imaging* 49.7 (2019), e101–e121.

[49] J. M. Bland and D. G. Altman. *Statistical methods for assessing agreement between two methods of clinical measurement*. In: *Lancet* 1.8476 (1986), pp. 307–310.

[50] H. Yu, B. Tang, and Y. Fu et al. *Quantifying the reproducibility and longitudinal repeatability of radiomics features in magnetic resonance Image-Guide accelerator Imaging: A phantom study*. In: *European Journal of Radiology* 181 (2024), p. 111735.

[51] A. K. Nazarlou, F. Faeghi, and M. H. Abdkarimi et al. *ADC values in diffusion-weighted MRI and their relationship with age, gender and BMI in healthy people's pancreases*. In: *Br J Radiol* 88.1047 (2015), p. 20140449.

[52] M. A Jacobs, K. J. Macura, and A. Zaheer et al. *Multiparametric Whole-body MRI with Diffusion-weighted Imaging and ADC Mapping for the Identification of Visceral and Osseous Metastases From Solid Tumors*. In: *Acad Radiol* 25.11 (2018), pp. 1405–1414.

[53] N. F. Michoux, J. W. Ceranka, and J. Vandemeulebroucke et al. *Repeatability and reproducibility of ADC measurements: a prospective multicenter whole-body-MRI study*. In: *Eur Radiol* 31.7 (2021), pp. 4514–4527.

[54] N. Papanikolaou, S. Gourtsoyianni, and S. Yarmenitis et al. *Comparison between two-point and four-point methods for quantification of apparent diffusion coefficient of normal liver parenchyma and focal lesions. Value of normalization with spleen.* In: *Eur J Radiol* 73.2 (2010), pp. 305–309.

[55] E. C. Bush, A. Gifford, and C. L. Coolbaugh et al. *Fat-Water Phantoms for Magnetic Resonance Imaging Validation: A Flexible and Scalable Protocol*. In: *J Vis Exp* 139 (2018), p. 57704.

[56] S. B. Reeder, A. R. Pineda, and Z. Wen et al. *Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): Application with fast spin-echo imaging*. In: *Magn Reson Med* 54.3 (2005), pp. 636–644.

[57] T. A. Bley, O. Wieben, and C. J. François et al. *Fat and water magnetic resonance imaging*. In: *J Magn Reson Imaging* 31.1 (2010), pp. 4–18.

[58] B. Glocker, E. Konukoglu, and I. Lavdas et al. *Correction of Fat-Water Swaps in Dixon MRI*. In: *MICCAI 2016* Lecture Notes in Computer Science 9902 (2016).

[59] C. Y. Liu, C. A. McKenzie, and H. Yu et al. *Fat quantification with IDEAL gradient echo imaging: correction of bias from T(1) and noise*. In: *Magn Reson Med* 58.2 (2007), pp. 354–364.

[60] T. Yookoo, M. Bidder, and G. Hamilton et al. *Hepatic fat quantification by low flip-angle multi-echo gradient-echo MR imaging: a clinical study with validation with MR spectroscopy*. In: *Proc 16th Annual Meeting ISMRM, Toronto* Lecture Notes in Computer Science 9902.Abstract 706 (2008).

[61] J. G. Stine and R. Loomba. *Magnetic resonance imaging proton density fat fraction as an imaging-based biomarker of treatment response in patients with non-alcoholic steatohepatitis.* In: *Clin Liver Dis (Hoboken)* 20.6 (2022).

[62] M. Mazzola, R. Cannella, and G. Pilato et al. *Quantitative liver MRI biomarkers: what the radiologist should know*. In: *J Med Imaging Intervent Radiol* 11.13 (2024). DOI: `https://doi.org/10.1007/s44326-024-00015-0`.

[63] T. J. Bray, M. D. Chouhan, and S. Punwani et al. *Fat fraction mapping using magnetic resonance imaging: insight into pathophysiology.* In: *Br J Radiol* 91.1089 (2018), p. 20170344.

[64] I. S. Idilman, A. E. Yildiz, and A. D. Karaosmanoglu et al. *Proton density fat fraction: magnetic resonance imaging applications beyond the liver*. In: *Diagn Interv Radiol* 28.1 (2022), pp. 83–91.

[65] Elekta. *MOSAIQ ® Plaza for Medical Oncology*. Elekta AB, Stockholm, Sweden. 2025. URL: `https://www.elekta.com/products/oncology-informatics/elekta-one/oncology-care/medical-oncology/` (accessed on 07/24/2025).

[66] H. H. Hu, T. Yookoo, and M. R. Bashir et al. *Linearity and Bias of Proton Density Fat Fraction as a Quantitative Imaging Biomarker: A Multicenter, Multiplatform, Multivendor Phantom Study*. In: *Radiology* 298.3 (2021), pp. 640–651.

[67] M. S. Burhans, N. Balu, and K. A. Schmidt et al. *Impact of the Analytical Approach on the Reliability of MRI-Based Assessment of Hepatic Fat Content*. In: *Curr Dev Nutr* 4.12 (2020), nzaa171.

[68] D. C. Karampinos, S. Ruschke, and M. Dieckmeyer et al. *Quantitative MRI and spectroscopy of bone marrow*. In: *J Magn Reson Imaging* 47.2 (2018), pp. 332–353.

[69] P. A. G. Teixeira, T. Cherubin, and S. Badr et al. *Proximal femur fat fraction variation in healthy subjects using chemical shift-encoding based MRI*. In: *Sci Rep* 9.1 (2019), p. 20212.

[70] A. Gupta, R. Dixit, and A. Prakash et al. *Non-invasive hepatic fat quantification: Can multi-echo Dixon help?* In: *Radiol Bras* 7.57 (2024), e20230125.

[71] C. W. Hong, G. Hamilton, and C. Hooker et al. *Measurement of spleen fat on MRI-proton density fat fraction arises from reconstruction of noise*. In: *Abdom Radiol (NY)* 44.10 (2019), pp. 3295–3303.

[72] S. D. Heber, H. Hetterich, and R. Lorbeer et al. *Pancreatic fat content by magnetic resonance imaging in subjects with prediabetes, diabetes, and controls from a general population without cardiovascular disease*. In: *PLoS One* 12.5 (2017), e0177154.

[73] X. J. Pei et al. *Fat fraction quantification of lumbar spine: comparison of T1-weighted two-point Dixon and single-voxel magnetic resonance spectroscopy in diagnosis of multiple myeloma*. In: *Diagn Interv Radiol* 26.5 (2020), pp. 492–497.

[74] C. Le Ster, J. Lasbleiz, and S. Kannengiesser et al. *A fast method for the quantification of fat fraction and relaxation times: Comparison of five sites of bone marrow*. In: *Magn Reson Imaging* 39 (2017), pp. 157–161.

[75] A. Baur-Melnyk, S. Buhmann, and H. R. Dürr et al. *Role of MRI for the diagnosis and prognosis of multiple myeloma*. In: *Eur J Radiol* 55.1 (2005), pp. 56–63.

[76] V. Koutoulidis, E. Terpos, and N. Papanikolaou et al. *Comparison of MRI Features of Fat Fraction and ADC for Early Treatment Response Assessment in Participants with Multiple Myeloma*. In: *Radiology* 304.1 (2022), pp. 137–144.

[77] M. Sun, J. Cheng, and C. Ren et al. *Quantitative whole-body MR imaging for assessment of tumor burden in patients with multiple myeloma: correlation with prognostic biomarkers*. In: *Quant Imaging Med Surg* 11.8 (2021), pp. 3767–3780.

[78] X. Ji, W. Huang, and H. Dong et al. *Evaluation of bone marrow infiltration in multiple myeloma using whole-body diffusion-weighted imaging and T1-weighted water-fat separation Dixon*. In: *Quant Imaging Med Surg* 11.2 (2021), pp. 641–651.

[79] A. Latifoltojar, M. Hall-Craggs, and A. Bainbridge A et al. *Whole-body MRI quantitative biomarkers are associated significantly with treatment response in patients with newly diagnosed symptomatic multiple myeloma following bortezomib induction*. In: *Eur Radiol* 27.12 (2017), pp. 5325–5336.

[80] A. Dinno. *Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test*. In: *The Stata Journal* 55.1 (2015), pp. 292–300.

[81] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant. *Applied Logistic Regression*. John Wiley & Sons, Ltd, 2013. DOI: `10.1002/9781118548387`.

[82] R. Kohavi. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In: *Proc 14th IJCAI, San Francisco* 2 (1995), pp. 1137–1143.

[83] R. Tibshirani. *Regression Shrinkage and Selection via the Lasso*. In: *J. R. Statis. Soc. B* 58.1 (1996), pp. 267–288.

[84] Image Biomarker Standardisation Initiative. *IBSI Documentation*. URL: `https://ibsi.readthedocs.io/en/latest/index.html` (accessed on 07/26/2025).

[85] R. M. Haralick, K. Shanmugam, and I. Dinstein. *Textural Features for Image Classification*. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (1973), pp. 610–621.

[86] L. G. Kessler, H. X. Barnhart, and A. J. Buckler et al. *The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions*. In: *Stat Methods Med Res* 24.1 (2015), pp. 9–26.

[87] Institut national de la statistique et des études économiques. *Coefficient of variation / CV*. 2016. URL: `https://www.insee.fr/en/metadonnees/definition/c1366#:~:text=The%20coefficient%20of%20variation%20(CV,the%20more%20precise%20the%20estimate.` (accessed on 07/30/2025).

[88] A. Hazra. *Using the confidence interval confidently*. In: *J Thorac Dis* 9.10 (2017), pp. 4125–4130.