cancers

MDPI

Article

Inter- and Intra-Patient Repeatability of Radiomic Features from Multiparametric Whole-Body MRI in Patients with Metastatic Prostate Cancer

Ricardo Donners [1,*], Antonio Candito [2], Mihaela Rata [2], Adam Sharp [2], Christina Messiou [2], Dow-Mu Koh [2], Nina Tunariu [2] and Matthew D. Blackledge [2,*]

[1] University Hospital Basel, Petersgraben 4, 4031 Basel, Switzerland
[2] The Institute of Cancer Research, 15 Cotswold Road, Sutton SM2 5NG, UK; antonio.candito@icr.ac.uk (A.C.); mihaela.rata@icr.ac.uk (M.R.); adam.sharp@icr.ac.uk (A.S.); christina.messiou@icr.ac.uk (C.M.); dow-mu.koh@icr.ac.uk (D.-M.K.); nina.tunariu@icr.ac.uk (N.T.)
* Correspondence: ricardo.donners@usb.ch (R.D.); matthew.blackledge@icr.ac.uk (M.D.B.)

Simple Summary: Prostate cancer bone metastases are a heterogonous disease with heterogeneous therapy-response, not adequately captured by one-dimensional imaging biomarker measurements. DWI and Dixon MRI radiomics analysis may tackle this shortcoming, but technical assessment of repeatability is an essential prerequisite before implementation. In this manuscript we identified whole-body MRI radiomics features in prostate cancer bone disease with good inter- and intra-patient repeatability. These features may be further explored to improve outcome predictions and therapy response assessment in prostate cancer patients.

Abstract: (1) Background: We assessed the test–re-test repeatability of radiomics in metastatic castration-resistant prostate cancer (mCPRC) bone disease on whole-body diffusion-weighted (DWI) and T1-weighted Dixon MRI. (2) Methods: In 10 mCRPC patients, 1.5 T MRI, including DWI and T1-weighted gradient-echo Dixon sequences, was performed twice on the same day. Apparent diffusion coefficient (ADC) and relative fat-fraction-percentage (rFF%) maps were calculated. Per study, up to 10 target bone metastases were manually delineated on DWI and Dixon images. All 106 radiomic features included in the Pyradiomics toolbox were derived for each target volume from the ADC and rFF% maps. To account for inter- and intra-patient measurement repeatability, the log-transformed individual target measurements were fitted to a hierarchical model, represented as a Bayesian network. Repeatability measurements, including the intraclass correlation coefficient (ICC), were derived. Feature ICCs were compared with mean ADC and rFF ICCs. (3) Results: A total of 65 DWI and 47 rFF% targets were analysed. There was no significant bias for any features. Pairwise correlation revealed fifteen ADC and fourteen rFF% feature sub-groups, without specific patterns between feature classes. The median intra-patient ICC was generally higher than the inter-patient ICC. Features that describe extremes in voxel values (minimum, maximum, range, skewness, and kurtosis) showed generally lower ICCs. Several mostly shape-based texture features were identified, which showed high inter- and intra-patient ICCs when compared with the mean ADC or mean rFF%, respectively. (4) Conclusions: Pyradiomics texture features of mCRPC bone metastases varied greatly in inter- and intra-patient repeatability. Several features demonstrated good repeatability, allowing for further exploration as diagnostic parameters in mCRPC bone disease.

Keywords: radiomics; diffusion magnetic resonance imaging; neoplasm metastases

Citation: Donners, R.; Candito, A.; Rata, M.; Sharp, A.; Messiou, C.; Koh, D.-M.; Tunariu, N.; Blackledge, M.D. Inter- and Intra-Patient Repeatability of Radiomic Features from Multiparametric Whole-Body MRI in Patients with Metastatic Prostate Cancer. *Cancers* **2024**, *16*, 1647. https://doi.org/10.3390/cancers16091647

Academic Editor: Dania Cioni

Received: 12 February 2024
Revised: 13 April 2024
Accepted: 22 April 2024
Published: 25 April 2024

1. Introduction

Metastatic castration-resistant prostate cancer (mCRPC) is a lethal disease. Bone metastases develop in 90% of mCRPC patients and are a major cause of morbidity and mortality [1]. However, conventional anatomic imaging techniques, including CT, bone

scans, and standard MRI are inadequate for the response assessment of malignant bone disease [2,3].

By contrast, whole-body MRI (WB-MRI), including diffusion-weighted imaging (DWI) and T1-weighted fat/water (Dixon) sequences, can assess the treatment response of bone disease [4]. DWI informs on tumour cellularity, whilst Dixon acquisition assesses the relative tissue fat content. Both techniques facilitate the identification, staging, and response assessment of bone metastases and may provide quantitative response biomarkers [4–11]. The DWI-derived apparent diffusion coefficient (ADC) and Dixon MRI-derived relative fat-fraction percentage (rFF%) correlate with tumour cellularity in prostate cancer bone metastases and show increases with therapy response [5,6,12–18]. However, the simple averaging of imaging biomarker values within delineated regions of interest (ROIs) fails to capture the commonly heterogeneous appearance of mCRPC disease. Studies have suggested that more complex evaluations of tumour texture features can improve therapy outcome predictions [19–21].

The computerised extraction of quantitative features from medical images to describe different cancer phenotypes is called "radiomics". Many radiomic features have been described, but there is still no routine implementation in clinical practise for mCRPC. Several factors contribute to this disparity between research and clinical application, which include the lack of an integral clinical pipeline for data curation, a lack of capacity for tumour annotation, and no clinical processing tools for the disease. It is not established which features provide consistent, repeatable results, which can be harnessed in a test–retest setting, to inform on relevant changes between baseline and follow-up imaging in cancer patients undergoing therapy [22]. Knowing a feature's repeatability is important—if a parameter shows poor repeatability, its predictive power is low; thus, excellent repeatability can be considered a prerequisite for meaningful parameter selection among the large number of radiomic features [23]. The study of MRI radiomics repeatability is challenging, because in contrast to CT, there is no inherent normalisation of signal intensities, making test–re-test comparisons between examinations difficult. DWI-derived ADC and Dixon-derived rFF% maps may tackle this shortcoming, as both parametric maps offer inherent normalisation, enabling inter-study comparisons [24].

To date, no radiomics repeatability study has been published in conjunction with the MRI assessment of metastatic bone disease. Without technical validation, no meaningful radiomics features can be identified, as was highlighted by expert consensus statements [25,26]. Consequently, the purpose of this study is to contribute to the work on this knowledge gap by assessing the test–re-test repeatability of radiomic features in mCPRC bone disease assessed on WB-MRI DWI and T1-weighted Dixon sequences. We considered all texture features included in the open-source Pyradiomics package, which are implemented according to consensus definitions of the Imaging Biomarker Standardisation Initiative (IBSI) [26,27].

2. Materials and Methods

2.1. Study Design

This prospective repeatability study was approved by the local research and ethics committee. Prostate cancer patients were recruited and consented in one institution. The study inclusion criteria were the histopathology diagnosis of prostate cancer, history of bone metastases, castration-resistant disease, and no contraindication for MRI acquisition. The exclusion criterion was contraindications for MRI acquisition. In total, eleven mCRPC patients were recruited.

2.2. Imaging Acquisition

Initial and repeat WB-MRI acquired on a Siemens MAGNETOM Aera 1.5T MRI system (Siemens Healthineers, Erlangen, Germany) were evaluated. Patients were scanned twice in one setting, with repositioning between the examinations. The median time interval between the initial and re-test imaging sequences was 54 min. The imaging protocol

included DWI and CAIPIRINHA (Controlled Aliasing in Parallel Imaging Results in Higher Acceleration)-accelerated T1-weighted Dixon MRI, as shown in Table 1. ADC and rFF% maps were calculated using in-house routines.

Table 1. MRI acquisition parameters, DWI—diffusion-weighted imaging, GE—gradient-echo.

Parameter	DWI	rFF
b-values	B50, b600 b900	
TE	69	2.39
TR	11,300	7.63
Slice	6 mm	5 mm
Inversion	STIR 180	
Averages	3-5-5	1
Slice spacing	6	6
Px bandwidth	1955	400
Aqu Matrix	128 × 104	256 × 156
Image matrix	256 × 208	256 × 208
Flip angle	90	10

2.3. Disease Delineation

Disease delineation was performed on commercially available post-processing software (OsiriX, version 56, PixmeoSARL Bernex, Switzerland) by a dedicated radiology fellow with four years of experience in the functional imaging of malignant bone disease. In each patient, up to 10 bone metastases were chosen as target lesions, facilitating the identification of inter-lesion heterogeneity. Lesions were selected across the body (where present) in the cervical spine, thoracic spine, lumbar spine, sacrum, pelvis, ribs, shoulders, and long bones. Target bone metastases were defined as a focal lesion with a low signal on rFF% images (<20%) compared with adjacent bone marrow, together with an unsuppressed high signal on b50 and b900 DWI and a mean ADC value of $<1400 \times 10^{-6}$ mm^2/s. Lesions with mean ADC $> 1400 \times 10^{-6}$ mm^2/s were evaluated together with previous imaging and were suitable for inclusion when they showed unequivocal increases or decreases in size and/or in ADC $\geq$ 30% (4). We did not analyse lesions < 1 mL in volume. (Figure 1).

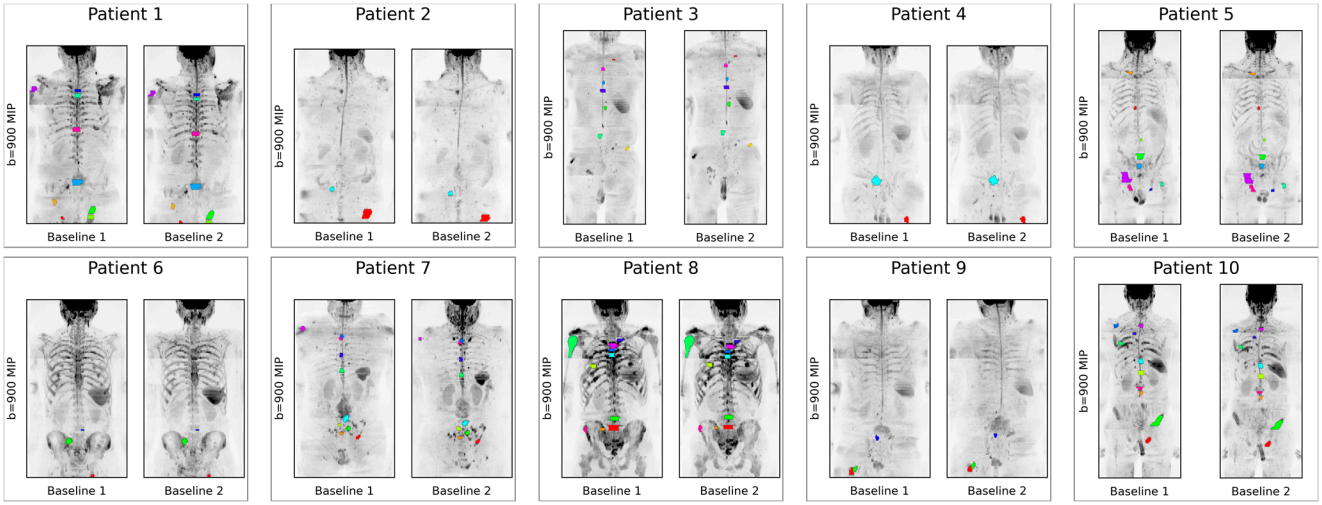

Figure 1. Maximum intensity projections (MIPs) of the high-b-value images for all 10 repeatability patients available to the study, including regions of interest displayed as colour overlays (each individual lesion displayed as a different colour).

The whole target lesion was segmented in consecutive slices in the b900 and rFF% images. The b900 segmentation masks were copied onto the corresponding ADC maps. As there was no ground truth, the absolute accuracy of segmentation was not evaluated as part of the study.

2.4. Extraction of Radiomics Features

Radiomics features were derived using the Pyradiomics toolbox [27], including the first order, shape, grey-level co-occurrence matrix (GLCM), grey-level run length matrix (GLRLM), grey-level size zone matrix (GLSZM), grey-level dependence matrix (GLDM), and neighbouring grey tone difference matrix (NGTDM) features (total of 106 features). Feature extraction definition files are presented in the Supplementary Information S1 and S2. As many derived features demonstrate heteroscedastic repeatability, all features were subsequently log-transformed, except for skewness (first order), minimum (first order), correlation (GLCM), and cluster shade (GLCM), which were observed to have both positive and negative and/or zero values [28]. Furthermore, the inverse difference normalised (GLCM) and inverse difference moment normalised (GLCM) were transformed according to $y = 1 - x$, and informational measure of correlation (GLCM) was normalised by $y = 1 - x$ prior to log-transformation. No wavelet-filtered features were investigated. We used a fixed bin size of 100×10^{-6} mm^2/s and 3.333% for ADC and rFF% maps, respectively, such that approximately 30 bins were applied in each case.

2.5. Repeatability Model

We considered the (log-transformed) measurements derived from each lesion within each patient to be derived from a hierarchical model, graphically represented as a Bayesian network in Figure 2.

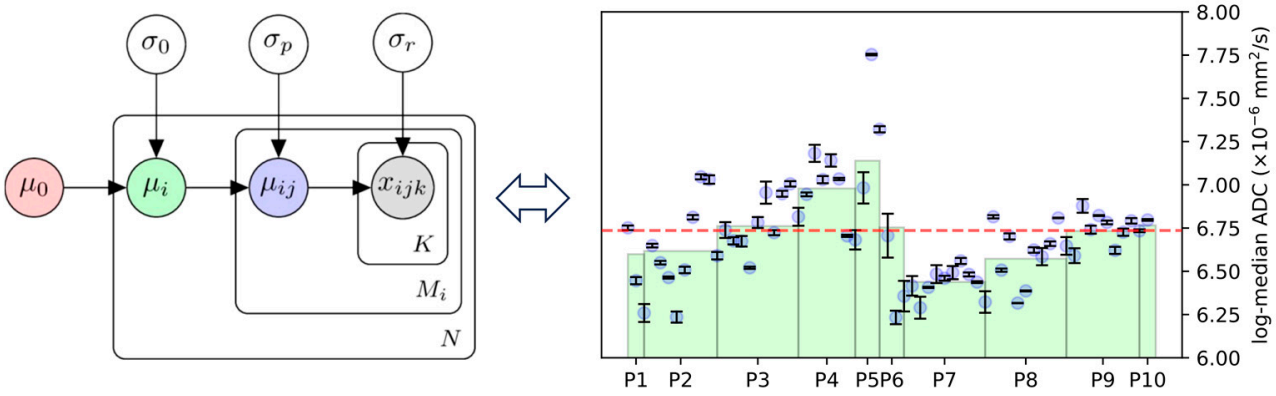

Figure 2. A Bayesian network of per-lesion measurements, x, within a whole-body MRI experiment. The model consists of three hierarchical normal distributions for the population, (μ_0, σ_0), the ith patient (μ_i, σ_p), and the jth lesion within the patient $\left(\mu_{ij}, \sigma_r\right)$. Note that whilst separate mean values μ are determined for each patient/lesion, global values for standard deviation σ are defined.

We denote the kth repeat measurement made in the jth lesion of patient i as x_{ijk}, where $k \in \{1,2\}$, $j \in \{1,\ldots,M_i\}$, and $i \in \{1,\ldots,N\}$. Each measurement is assumed to be normally distributed about the true lesion value, μ_{ij}, with inter-measurement error, σ_r, each lesion value being normally distributed about the true patient value μ_i with intra-patient variation, σ_p, and each patient value being normally distributed about the population average μ_0 with inter-patient variation, σ_0: $x_{ijk} \sim \mathcal{N}\left(\mu_{ij}, \sigma_r\right)$, $\mu_{ij} \sim \mathcal{N}(\mu_i, \sigma_p)$, and $\mu_i \sim \mathcal{N}(\mu_0, \sigma_0)$. A key advantage to this model is that it allows us to disentangle variation amongst lesions within an individual patient from variation between different patients and thus understand whether a particular measurement is sensitive to changes occurring on a per-lesion and/or per-patient level. Once estimates of the parameters from

this model were determined, we extracted several summary statistics, in line with those from the traditional repeatability literature.

Equation	Description
$RC = 1.96\sqrt{2}\sigma_r$	Repeatability coefficient. Useful in the context of assessing response after treatment. Any change above +RC or below −RC is considered to be statistically significant and thus might be a direct result of treatment rather than due to measurement error (assuming a p-value of 0.05).
$ICC_\delta = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_r^2}$ $(0 \leq ICC_\delta \leq 1)$	Intra-patient intraclass correlation. Compares the magnitude of the inter-measurement error with intra-patient variation in lesion values. A value closer to 1 indicates better measurement repeatability in the context of measuring changes to individual lesions.
$ICC_\Delta = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_r^2}$ $(0 \leq ICC_\Delta \leq 1)$	Inter-patient intraclass correlation. Compares the magnitude of the inter-measurement error with inter-patient variation in lesion values. A value closer to 1 indicates better measurement repeatability in the context of measuring changes with groups of lesions within each patient.
$wlCV = \sqrt{\frac{1}{\sum_i M_i}\sum_{i,j}\frac{\sigma_r^2}{\mu_{ij}^2}}$ $(wlCV \geq 0)$	Average within-lesion coefficient of variation. Describes the magnitude of inter-measurement error in the context of true lesion values. A large value represents potentially poor repeatability compared with the expected values for each lesion.
$blCV = \sqrt{\frac{1}{N}\sum_i \frac{\sigma_p^2}{\mu_i^2}}$ $(blCV \geq 0)$	Average between-lesion coefficient of variation. Describes the magnitude of inter-lesion variation in the context of average patient values. A large value represents higher intra-patient heterogeneity.
$bCV = \frac{\sigma_0}{\mu_0}$ $(bCV \geq 0)$	Between-patient coefficient of variation. Describes the magnitude of inter-patient variation in the context of the average population value. A large value represents higher inter-patient heterogeneity.
$LoA = \left(e^{\pm 1.96\sqrt{2}\sigma_r'} - 1\right) \times 100\%$	Limits of agreement (log-transformed features only). Defines the percentage difference after treatment needed to deem that change significantly different.

It is important to note that in this setting, we defined σ_p as a population-wide parameter, where, in theory, it might be estimated for each patient. However, given the numbers of lesions we encounter in certain patients, it can become very difficult to meaningfully estimate this parameter on a per-patient basis. Furthermore, by considering it as a population-wide estimate, it is much simpler to compare it with the measurement error σ_r.

We also defined a bias parameter ε in the model that represents the average difference between both baseline measurements: $x_{ij1} \sim \mathcal{N}\left(\mu_{ij}, \sigma_r\right)$ and $x_{ij2} \sim \mathcal{N}\left(\mu_{ij} + \varepsilon, \sigma_r\right)$. This allowed us to confirm that no systematic bias occurred between both baseline measurements. The total number of parameters in this model was $8 + N + \sum_i M_i$.

2.6. Model Fitting

We used the Markov Chain Monte Carlo (MCMC) optimisation with Stan for hierarchical modelling [29]. This technique draws samples from the posterior probability distribution of model parameters given the available data, thereby fully characterising uncertainty in parameter estimation (and subsequently generated statistics). The Stan code for our implementation is presented in Supplementary S3.

Before sampling, we standardised our data using the convention $x'_{ij1} = \left(x_{ij1} - \overline{x_1}\right) / \sqrt{Var(x_1)}$ and $x'_{ij2} = \left(x_{ij2} - \overline{x_1}\right) / \sqrt{Var(x_1)}$, where $\overline{x_1}$ and $Var(x_1)$ represent the mean and variance

in the data from baseline measurement 1, respectively. Furthermore, we initialise the parameter values for sampling using the following:

$\hat{\mu}_{ij}^0 = \frac{1}{2}\left(x_{ij1} + x_{ij2}\right)$	$\hat{\sigma}_r^0 = \sqrt{\frac{1}{2\sum_i M_i}\sum_{i,j}\left(x_{ij1} - x_{ij2}\right)^2}$
$\hat{\mu}_i^0 = \frac{1}{M_i}\sum_j \hat{\mu}_{ij}^0$	$\hat{\sigma}_p^0 = \sqrt{\frac{1}{\sum_i M_i - N}\sum_{i,j}\left(\hat{\mu}_{ij}^0 - \hat{\mu}_i^0\right)^2}$
$\hat{\mu}_0^0 = \frac{1}{N}\sum_i \hat{\mu}_i^0$	$\hat{\sigma}_0^0 = \sqrt{\frac{1}{N-1}\sum_i \left(\hat{\mu}_i^0 - \hat{\mu}_0^0\right)^2}$
$\hat{\varepsilon}^0 = \frac{1}{\sum_i M_i}\sum_{i,j}\left(x_{ij2} - x_{ij1}\right)$	

Prior distributions for parameters σ_0, σ_p, and σ_r were set to be half-Cauchy distributions with location 0 and scale 5, whilst priors for μ_0 and ε were zero-mean normal distributions with a standard deviation of 10. Checks were made that the range of these distributions covered the range of initial model parameters $\hat{\sigma}_r^0$, $\hat{\sigma}_p^0$, $\hat{\sigma}_0^0$, $\hat{\varepsilon}^0$, and $\hat{\mu}_r^0$ for all radiomic features investigated in this study. Sampling parameters included the following: number of chains = 3, number of samples = 2000, number of warmup samples = 500, no thinning, and fixed random seed initialisation.

To assess the independence of successive samples and good mixing of multiple sampling chains, we used the Gelman–Rubin R-hat ($\hat{R}$) convergence diagnostic: Calculated as the ratio of the pooled variance of parameters across multiple chains to the average variance within each individual chain, good mixing was observed as $\hat{R} \to 1$. Our sampling regime consisted of checking that 99% of all parameters had $\hat{R} \leq 1.02$; otherwise, samples were rejected and repeated up to 10 times. In our study, this schema needed, at most, two retries until adequate convergence was found over all radiomics features considered.

Fixed thresholds for repeatability interpretation can be problematic [30]. Consequently, we compared the ICCs of the extracted texture features to the mean ADC and mean rFF% ICCs to allow for some classification. Features with equal or greater ICCs than these reference metrics were considered to offer good repeatability.

3. Results

As one patient did not show bone metastases on WB-MRI, a total of 10 mCRPC patients with a median age of 67.5 years were included for analysis. All patients had undergone all lines of standard-of-care treatment and were undergoing systemic therapy at the time of study inclusion. In one patient, Dixon imaging had been performed with erroneous pixel spacing during one of the repeat measurements, and so this patient was removed from rFF% analysis. In total, 65 delineated target lesions were used for ADC analysis and 47 for rFF% analysis.

Median repeatability model parameters for all radiomics features are presented in Supplementary S4 and S5. For all features, for both ADC and rFF%, there was no evidence of significant bias.

Heatmaps of pairwise distance correlation for radiomic features are presented in Figure 3 (and Supplementary S6). Fifteen and fourteen sub-groups were identified for ADC-derived and rFF%-derived features, respectively. No specific patterns were identified in terms of the radiomic feature classes that were identified within these groups, other than typical parameters such as “10th percentile”, “90th percentile, “median”, “mean”, and “root-mean-squared” being grouped together in both cases. Although the first-order “minimum” was identified to exist in its own subgroup for rFF%, we note that this feature was zero for many lesions and thus likely not a reliable biomarker.

Bland–Altman plots for a single representative radiomics feature from each correlated sub-group are presented in Figure 4, where in each sub-group, the feature with the maximum intra-patient ICC is presented. Repeatability limits are adjusted to account for the observed linearity using the method of Euser et al. [28].

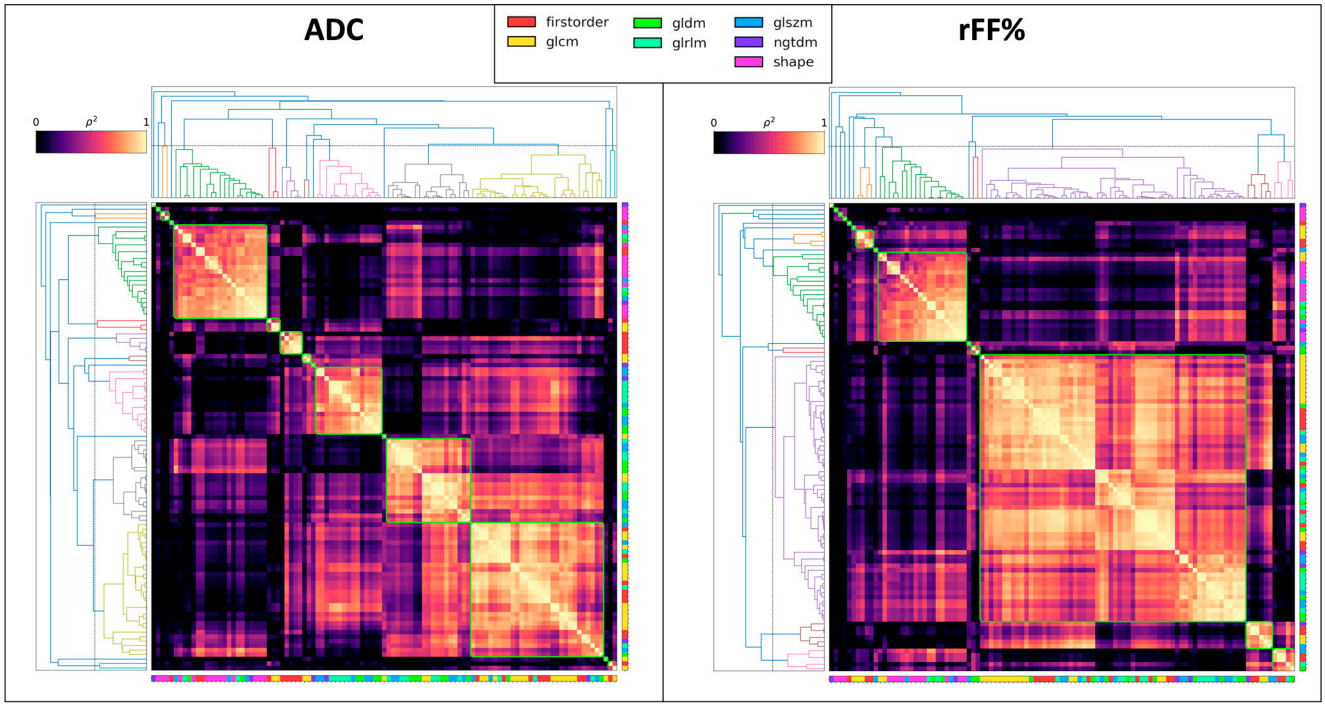

Figure 3. Correlation heatmaps of radiomics features for both ADC and fat fraction values. It is noted that a large proportion of the radiomics features are highly correlated; 15 sub-groups are identified as ADC-derived features, whilst 14 are identified for fat-fraction-derived features, using hierarchical clustering with a pairwise Spearman correlation threshold of $1 - \rho^2 = 0.51$. A high-resolution copy of this figure is provided in Supplementary Information S6, which depicts the name of all features.

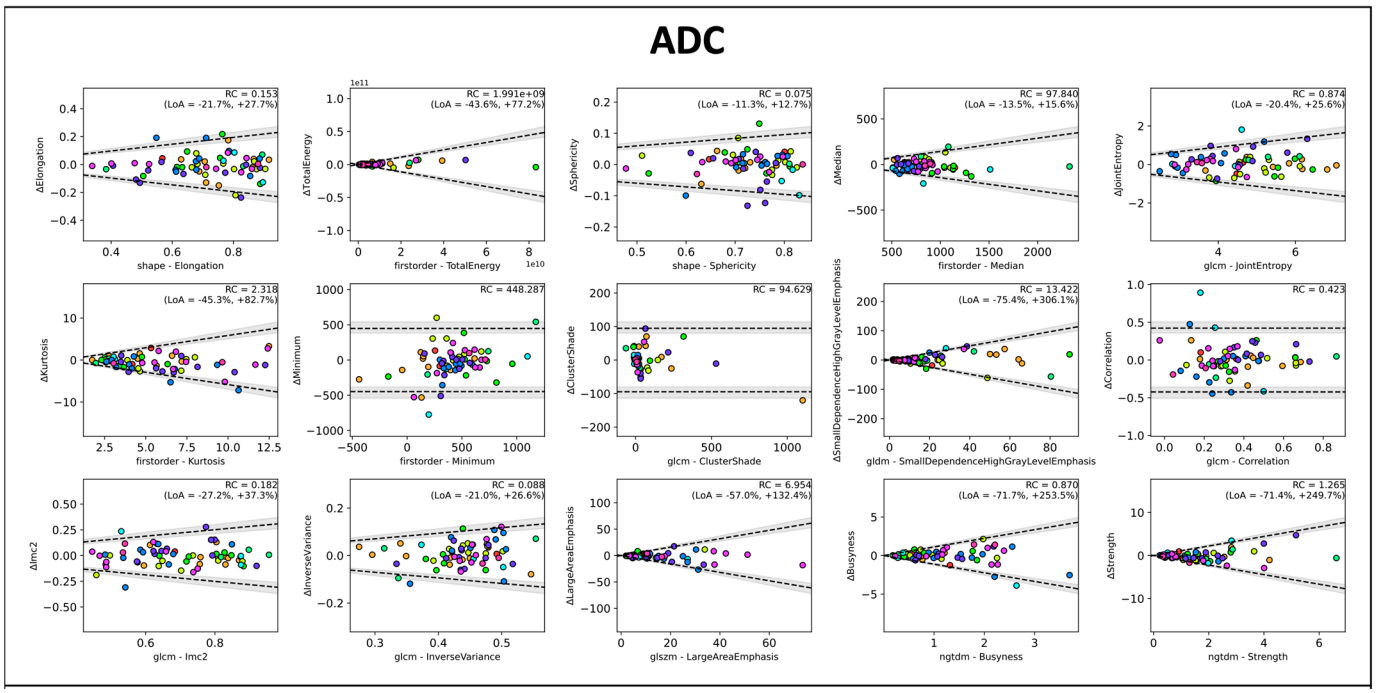

Figure 4. *Cont.*

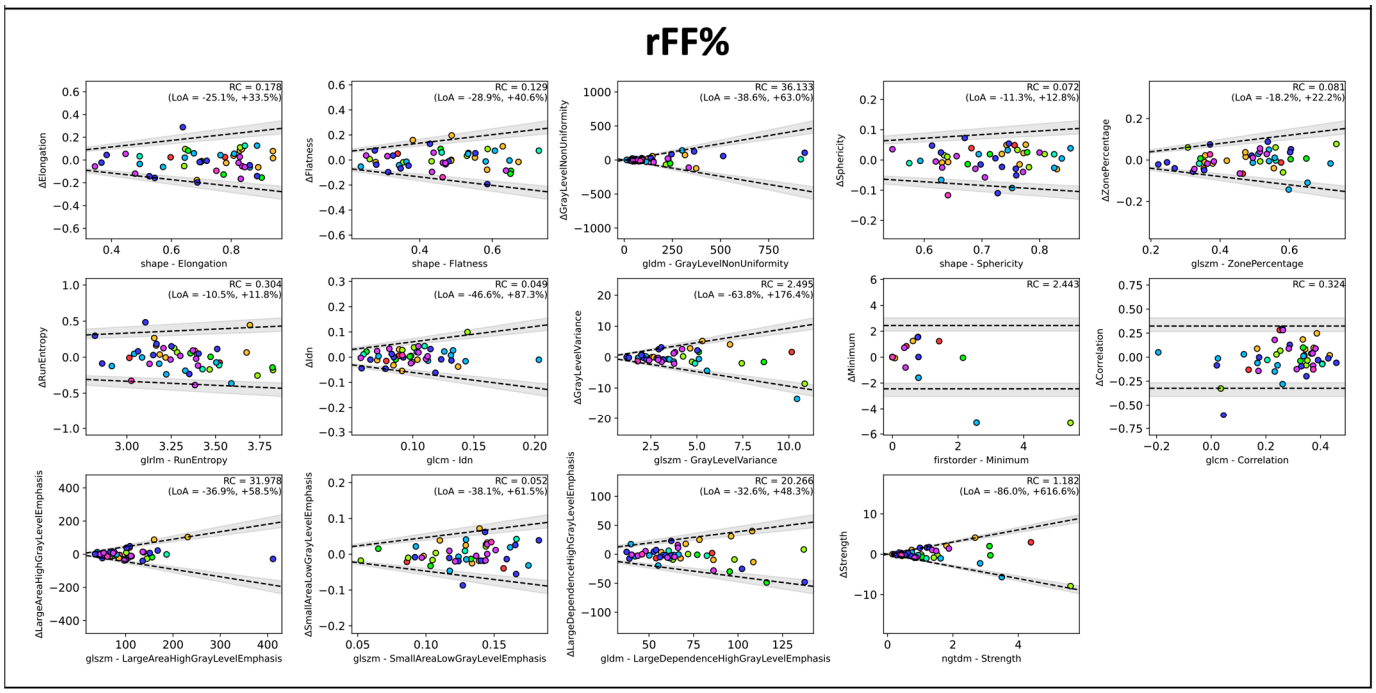

Figure 4. Bland–Altman plots for radiomics features from our study. Presented are those features that demonstrated the best median ICC_{δ} within each correlation group. The scatter plots are color-coded according to patient to demonstrate the inter-patient variability. Black dashed lines represent limits of repeatability (95% confidence interval illustrated as grey areas), and the values of the repeatability coefficient (RC) and limits of agreement (LoA) are provided.

Waterfall plots for the median intra- and inter-patient ICCs are presented in Figure 5 (and Supplementary S7) for all radiomics features investigated, along with inter-quartile ranges (IQRs). The median intra-patient ICC, ICC_{δ}, was generally higher than the median inter-patient ICC, ICC_{Δ}, though the uncertainty in the estimated ICC_{Δ} was also generally much larger (due to a smaller effective sample size).

This is echoed in Figure 6, which presents scatter plots of ICC_{Δ} and ICC_{δ} for both ADC and rFF%. Features that describe extremes in voxel values (minimum, maximum, range, skewness, and kurtosis) appear to have much lower ICC_{δ} values for rFF% than for ADC. Conversely, many texture features appear to have higher ICC_{δ} values for rFF% than for ADC. It was difficult to interpret ICC_{Δ} differences between both quantitative metrics, as error bars were much larger, and therefore, any trends needed to be interpreted with caution.

Comparison of Features with the Reference Metrics

The mean ADC and mean rFF% were chosen as reference metrics to allow for the assessment and comparison of radiomics features' inter- and intra-patient repeatability. Supplementary S4 and S5 show the repeatability measurements for all analysed features.

The inter-patient ICC, ICC_{Δ}, for the mean ADC was 0.93. Among the other 17 first-order features only the median and root-mean-squared showed equivalent ICC_{Δ} values. None of the 22 glcm, 14 gldm, 16 glrlm, 16 glszm, 5 ngtdm, or 14 shape features showed comparable ICC_{Δ} values (group maximum 0.74–0.83). The mean ADC intra-patient ICC, ICC_{δ}, was 0.95. Three first-order, one glcm, one gldm, two glrlm, one glszm, one ngtdm, and five shape features yielded equivalent or higher ICC_{δ} values (0.95–0.97).

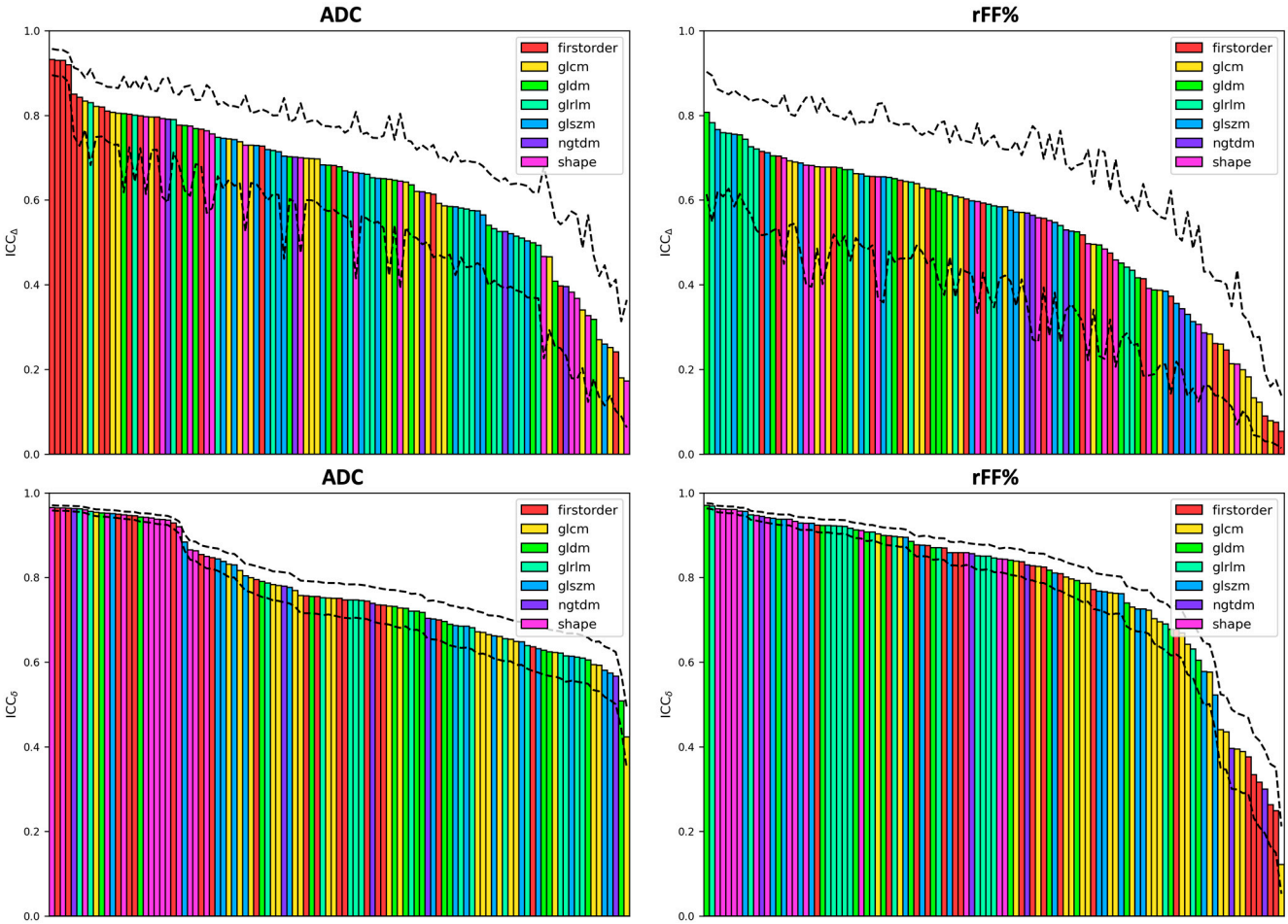

Figure 5. Waterfall plots of the inter- and intra-patient intraclass correlation coefficients for all radiomics features (ICC_Δ top row, and ICC_δ bottom row, respectively). Values for ADC measurements are presented in the left column, whilst values for fat fraction measurements are shown in the right column. Bars are colour-coded according to the radiomic feature type, and dashed lines represent the interquartile range of ICC values. A high-resolution copy of this figure is provided in Supplementary Information S7, which depicts the name of all features.

The mean rFF% inter-patient ICC, ICC_Δ, was 0.70. The median rFF% ICC_Δ was 0.71. The gldm parameters GrayLevelNonUniformity and SmallDependenceEmphasis, 6/16 glrlm, 3/16 glszm, and 1/14 shape features showed equivalent or higher ICC_Δ values (group maximum 0.70–0.81). All features in the glcm and ngtdm groups had lower ICC_Δ values (maximum 0.69 and 0.57, respectively). However, Bayesian sampling of ICC_Δ revealed relatively large confidence intervals in parameter estimation for all features, and thus, overlap between feature precision was apparent in all groups. The mean rFF% intra-patient ICC_δ was 0.90. The median rFF% intra-patient ICC_δ was 0.92. Two glcm, six gldm, nine glrlm, five glszm, one ngtdm, and eleven shape features showed equivalent or higher ICC_δ values (group maximum 0.90–0.97).

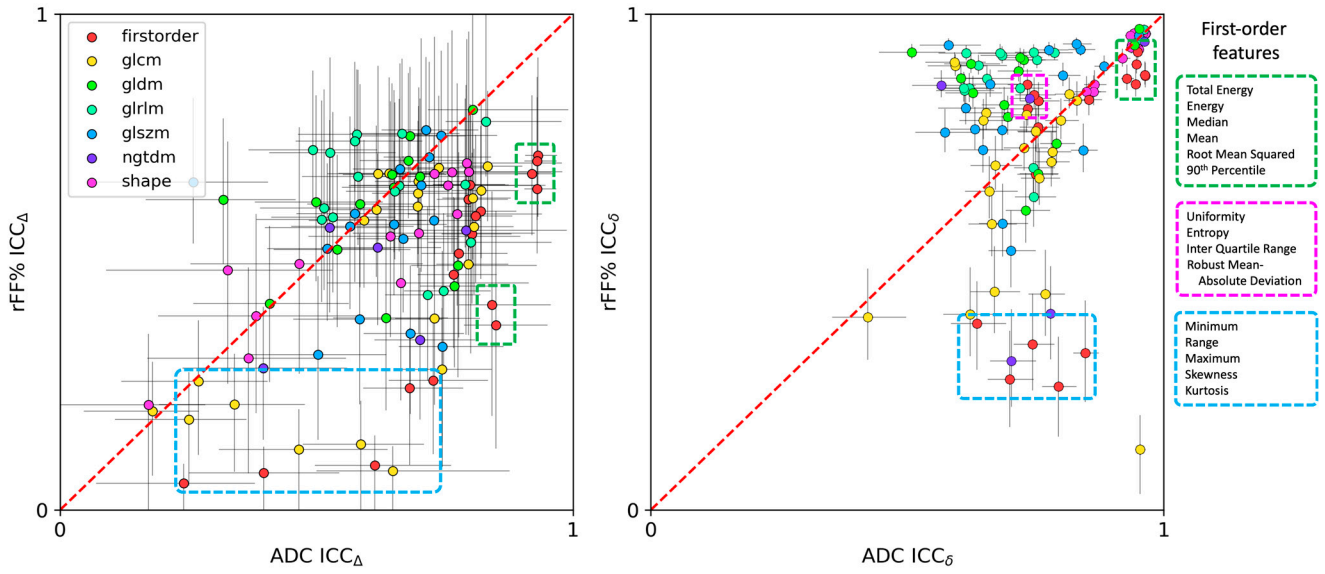

Figure 6. Scatter plots for inter- and intra-patient intraclass correlation coefficients for all radiomics features (ICC_Δ left, and ICC_δ right, respectively), comparing metrics between ADC and fat fraction. The red dashed line represents the line of equality, and interquartile ranges for the measured ICC are displayed as error bars around each scatter point. Our data seem to indicate that the inter-lesion ICC for fat fraction is significantly lower than for ADC for features that capture the extremes in data and may not be robust to outliers (blue dashed box). However, as might be expected, features that capture data averages demonstrate high ICC_δ in both cases (green dashed box).

4. Discussion

In this study of the intra- and interpatient repeatability of radiomic features in mCRPC bone disease, we found that the intra-patient ICC, ICC_δ, was generally higher than inter-patient ICC, ICC_Δ. This suggests that Pyradiomics features are more stable and thus might be more sensitive to changes occurring for individual lesions rather than total-body measurements.

Regarding ADC map radiomic analyses, the most repeatable features were shape-based or first-order features, demonstrating excellent repeatability (ICC_Δ and $ICC_\delta > 0.8$). We note that many first-order features are highly correlated (mean, median, and root-mean-squared, for example), as shown by the fact that only 15 uncorrelated sub-groups were found from our correlation analysis of ADC features. The mean ADC is a commonly used biomarker in cancer imaging, and part of contemporary imaging and interpretation guidelines for prostate cancer bone disease [4], and has been shown to correlate negatively with tumour cellularity [5,31]. It is considered to offer good measurement repeatability and is therefore commonly employed for malignant bone marrow lesion comparison in a test–re-test setting [6,32]. Nonetheless, we identified fourteen texture features with equivalent ICC_δ values, with at least one being from each feature class. These features likely infer information about the heterogeneity of tumour cellularity, which may be compared between imaging time points, allowing for the monitoring of cancer evolution in patients undergoing oncology therapy.

Among texture classes for rFF% repeatability, the best performance was found for grey-level non-uniformity (GLDM and GLRLM), though only the GLDM version demonstrated $ICC_\Delta > 0.8$. Multiple texture features from various feature groups demonstrated equivalent or higher ICCs when compared to the mean rFF% ($ICC_\Delta \geq 0.7$ and $ICC_\delta > 0.9$). Similar to the ADC, rFF% can also provide information on bone metastases and their evolution under cancer therapy—while a vital metastasis is assumed to contain no fat, a return of fatty bone marrow may suggest favourable response to therapy. The latter may be detected by the comparison of rFF% features between baseline and follow-up MRI.

When comparing the repeatability of radiomics texture features, GLRLM, GLDM, and GLSZM features generally had higher ICC_{δ} values for rFF% maps than for equivalent features derived from ADC maps. Many first-order feature ICC_{δ} values were equivalent between rFF% and ADC, with the exception of those that describe extremes in the voxel values including minimum, maximum, range, skewness, and kurtosis, which had significantly lower ICC_{δ} values than those computed using ADC. Shape-based features demonstrated similar ICC values between ADC and rFF%, which is likely because they should be independent of the imaging modality from which they were derived.

Recently, researchers applied radiomics to detect visually non-perceivable prostate cancer bone metastases on CT [33]. Hounsfield units are inherently normalised, which facilitates inter-study comparisons. By contrast, MRI signal intensity values are relative; however, ADC and rFF% maps provide inherent normalisation, enabling inter-study comparison.

In primary cancer of the prostate gland, research into MRI-derived radiomics is ahead of the current body of literature on radiomics in bone metastases. Two recent studies identified 12 and 15 features, respectively, extracted from pre-treatment T2- and dynamic contrast-enhanced T1-weighted MRI of the gland, which were significantly associated with the presence of bone metastases [34,35]. One may hypothesise that mCRPC bone metastases' radiomics may likewise be used as predictors of response or even overall patient survival in future scenarios. Published research does not yet provide evidence for this hypothesis. Nonetheless, researchers have found that DWI-derived radiomics may be used to classify spinal tumours [36], dynamic contrast-enhanced spine MRI-derived radiomics can discriminate between lung cancer and non-lung cancer spine metastases [37], and MRI radiomics may be able to differentiate between malignant and benign spinal lesions [38], with more evidence to be expected in the near future.

Our study results suggest that several Pyradiomics features derived from ADC and rFF% maps have sufficiently high levels of repeatability to be utilised for predictive, diagnostic models. We identified ADC and rFF% radiomics with good repeatability in a test–re-test setting, which may contribute to a better understanding of the heterogenous responses seen in metastatic bone disease in mCRPC based on DWI/rFF alone. Our results are unique to this dataset. Consequently, general recommendations on which features may yield the highest diagnostic value in a follow-up setting in a patient undergoing oncology treatment cannot be made. For the common scenario of mCRPC patients undergoing repeat examinations for the surveillance of malignant bone disease, we have identified several parameters with a sufficient level of repeatability to be tested in future studies. Moreover, repeatability studies on conventional, single imaging biomarkers usually aim to determine limits of agreement to allow for the identification of meaningful parameter change thresholds in a test–re-test scenario—"meaningful change" is the measured difference in the quantitative imaging biomarker between two time points, which represents a true biological effect, such as response to therapy, rather than measurement variability. A quantitative parameter change between two time points which is larger than the determined LoA or repeatability coefficient can be considered meaningful. In this study, we clearly demonstrated linear agreement between the tested radiomics features. However, with the current level of evidence, we do not consider it sensible to conclude fixed parameter thresholds for the tested features.

This study has several limitations. First, only 10 patients were included. For the requirement of repeat scans, repeatability studies are time and labour intensive and usually include few patients. This is a key motivator for the development of our novel Bayesian pipeline for analysing repeatability data in whole-body MRI studies. Second, the diagnostic performance of the analysed texture features was not evaluated, as this is beyond the scope of the manuscript. Third, multiple factors affect feature repeatability, including the consistency of lesion segmentation between test and re-test examination. Although an experienced radiologist performed these measurements, a degree of variation must be expected. This is representative of clinical practise, where lesion measurements and comparisons will not be perfectly matched. A baseline ex vivo phantom study was beyond

the scope of this article, but previous authors noted the good test–retest repeatability of DWI and ADC radiomics, acknowledging the limited implications for in vivo analyses [39].

Finally, the use of ICCs to compare repeatability across different biomarkers can be problematic [40]. For any response biomarker to be effective, two conditions should be met: Firstly, it must be precise enough to reliably detect genuine changes in the tumour property of interest, meaning the repeatability coefficient should be considerably smaller than the change anticipated after treatment. Secondly, the biomarker should exhibit a significant change in response to successful treatment, implying that the expected effect size is substantially greater than the measurement error, as determined using a repeatability assessment. While an ICC does not directly measure these conditions, it can still offer some insights into the relative repeatability of different biomarkers against their expected range of variation within a given patient cohort. However, without post-treatment data, an ICC can only suggest, not confirm, goodness of repeatability, and interpretations based on an ICC should be approached with caution. In determining the clinical utility of a biomarker, it is advisable to consider additional indices of repeatability, beyond the ICC (including the repeatability coefficient and coefficient of variation, also presented in this article), that better capture the precise nature required for clinical application.

5. Conclusions

In conclusion, mCRPC bone metastases Pyradiomics texture features vary greatly in inter- and intra-patient repeatability. In the presented dataset, we were not able to determine several universally stable features; however, we found several features with good repeatability, allowing for their further exploration as diagnostic parameters in mCRPC bone disease.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers16091647/s1, S1–S4: Tables of ADC repeatability statistics for each radiomic features tested. Median values are given with 95% confidence interval provided in brackets (other than for R-hat statistics, where the minimum and maximum over all parameters is demonstrated); S5: Tables of Fat Fraction repeatability statistics for each radiomic features tested. Median values are given with 95% confidence interval provided in brackets (other than for R-hat statistics, where the minimum and maximum over all parameters is demonstrated); S6: High resolution version of Figure 3 that includes features names; S7: High resolution version of Figure 5 that includes features names.

Author Contributions: Conceptualisation, R.D., N.T. and M.D.B.; methodology, R.D., N.T. and M.D.B.; software, R.D., N.T. and M.D.B.; validation, A.C., M.R., A.S., C.M., D.-M.K., N.T. and M.D.B.; formal analysis, R.D. and M.D.B.; investigation, R.D., N.T. and M.D.B.; resources, D.-M.K., N.T. and M.D.B.; data curation, R.D., N.T. and M.D.B.; writing—original draft preparation, R.D.; writing—review and editing, A.C., M.R., A.S., C.M., D.-M.K., N.T. and M.D.B.; visualisation, R.D. and M.D.B.; supervision, A.S., C.M., D.-M.K., N.T. and M.D.B.; project administration, R.D., N.T., D.-M.K. and M.D.B.; funding acquisition, D.-M.K., N.T. and M.D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Prostate Cancer UK, the Movember Foundation through the London Movember Centre of Excellence (CEO13_2-002), the Prostrate Cancer Foundation, Cancer Research UK (Centre Programme grant), and Experimental Cancer Medicine Centre grant funding from Cancer Research UK.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board Trial registration: Committee for Clinical Research of the Royal Marsden Hospital, registration number CCR1406, Sutton, Surrey, UK.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: Data are not publicly available but can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bubendorf, L.; Schöpfer, A.; Wagner, U.; Sauter, G.; Moch, H.; Willi, N.; Gasser, T.C.; Mihatsch, M.J. Metastatic patterns of prostate cancer: An autopsy study of 1589 patients. *Hum. Pathol.* **2000**, *31*, 578–583. [CrossRef] [PubMed]
2. Schwartz, L.H.; Seymour, L.; Litière, S.; Ford, R.; Gwyther, S.; Mandrekar, S.; Shankar, L.; Bogaerts, J.; Chen, A.; Dancey, J.; et al. RECIST 1.1—Standardisation and disease-specific adaptations: Perspectives from the RECIST Working Group. *Eur. J. Cancer* **2016**, *62*, 138–145. [CrossRef] [PubMed]
3. Scher, H.I.; Morris, M.J.; Stadler, W.M.; Higano, C.; Basch, E.; Fizazi, K.; Antonarakis, E.S.; Beer, T.M.; Carducci, M.A.; Chi, K.N.; et al. Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. *J. Clin. Oncol.* **2016**, *34*, 1402–1418. [CrossRef] [PubMed]
4. Padhani, A.R.; Lecouvet, F.E.; Tunariu, N.; Koh, D.M.; De Keyzer, F.; Collins, D.J.; Sala, E.; Schlemmer, H.P.; Petralia, G.; Vargas, H.A.; 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. *Eur. Urol.* **2017**, *71*, 81–92. [CrossRef]
5. Perez-Lopez, R.; Nava Rodrigues, D.; Figueiredo, I.; Mateo, J.; Collins, D.J.; Koh, D.M.; de Bono, J.S.; Tunariu, N. Multiparametric Magnetic Resonance Imaging of Prostate Cancer Bone Disease: Correlation With Bone Biopsy Histological and Molecular Features. *Invest. Radiol.* **2018**, *53*, 96–102. [CrossRef]
6. Perez-Lopez, R.; Mateo, J.; Mossop, H.; Blackledge, M.D.; Collins, D.J.; Rata, M.; Morgan, V.A.; Macdonald, A.; Sandhu, S.; Lorente, D.; et al. Diffusion-weighted Imaging as a Treatment Response Biomarker for Evaluating Bone Metastases in Prostate Cancer: A Pilot Study. *Radiology* **2017**, *283*, 168–177. [CrossRef] [PubMed]
7. Perez-Lopez, R.; Lorente, D.; Blackledge, M.D.; Collins, D.J.; Mateo, J.; Bianchini, D.; Omlin, A.; Zivi, A.; Leach, M.O.; de Bono, J.S.; et al. Volume of Bone Metastasis Assessed with Whole-Body Diffusion-weighted Imaging Is Associated with Overall Survival in Metastatic Castration-resistant Prostate Cancer. *Radiology* **2016**, *280*, 151–160. [CrossRef]
8. Padhani, A.; Gogbashian, A. Bony metastases: Assessing response to therapy with whole-body diffusion MRI. *Cancer Imaging* **2011**, *11*, S129–S145. [CrossRef] [PubMed]
9. Costelloe, C.M.; Madewell, J.E.; Kundra, V.; Harrell, R.K.; Bassett, R.L.; Ma, J. Conspicuity of bone metastases on fast Dixon-based multisequence whole-body MRI: Clinical utility per sequence. *Magn. Reson. Imaging* **2013**, *31*, 669–675. [CrossRef]
10. Padhani, A.R.; Makris, A.; Gall, P.; Collins, D.J.; Tunariu, N.; de Bono, J.S. Therapy monitoring of skeletal metastases with whole-body diffusion MRI. *J. Magn. Reson. Imaging* **2014**, *39*, 1049–1078. [CrossRef]
11. Donners, R.; Blackledge, M.; Tunariu, N.; Messiou, C.; Merkle, E.M.; Koh, D.M. Quantitative Whole-Body Diffusion-Weighted MR Imaging. *Magn. Reson. Imaging Clin. N. Am.* **2018**, *26*, 479–494. [CrossRef] [PubMed]
12. Woo, S.; Suh, C.H.; Kim, S.Y.; Cho, J.Y.; Kim, S.H. Diagnostic Performance of DWI for Differentiating High- From Low-Grade Clear Cell Renal Cell Carcinoma: A Systematic Review and Meta-Analysis. *AJR Am. J. Roentgenol.* **2017**, *209*, W374–W381. [CrossRef] [PubMed]
13. Sun, Y.; Tong, T.; Cai, S.; Bi, R.; Xin, C.; Gu, Y. Apparent Diffusion Coefficient (ADC) value: A potential imaging biomarker that reflects the biological features of rectal cancer. *PLoS ONE* **2014**, *9*, e109371. [CrossRef] [PubMed]
14. Choi, S.Y.; Chang, Y.W.; Park, H.J.; Kim, H.J.; Hong, S.S.; Seo, D.Y. Correlation of the apparent diffusion coefficiency values on diffusion-weighted imaging with prognostic factors for breast cancer. *Br. J. Radiol.* **2012**, *85*, e474–e479. [CrossRef] [PubMed]
15. Costantini, M.; Belli, P.; Rinaldi, P.; Bufi, E.; Giardina, G.; Franceschini, G.; Petrone, G.; Bonomo, L. Diffusion-weighted imaging in breast cancer: Relationship between apparent diffusion coefficient and tumour aggressiveness. *Clin. Radiol.* **2010**, *65*, 1005–1012. [CrossRef]
16. Wang, Y.; Chen, Z.E.; Yaghmai, V.; Nikolaidis, P.; McCarthy, R.J.; Merrick, L.; Miller, F.H. Diffusion-weighted MR imaging in pancreatic endocrine tumors correlated with histopathologic characteristics. *J. Magn. Reson. Imaging* **2011**, *33*, 1071–1079. [CrossRef]
17. Donners, R.; Hirschmann, A.; Gutzeit, A.; Harder, D. T2-weighted Dixon MRI of the spine: A feasibility study of quantitative vertebral bone marrow analysis. *Diagn. Interv. Imaging* **2021**. [CrossRef] [PubMed]
18. Donners, R.; Obmann, M.M.; Boll, D.; Gutzeit, A.; Harder, D. Dixon or DWI—Comparing the utility of fat fraction and apparent diffusion coefficient to distinguish between malignant and acute osteoporotic vertebral fractures. *Eur. J. Radiol.* **2020**, *132*, 109342. [CrossRef]
19. Scalco, E.; Rizzo, G. Texture analysis of medical images for radiotherapy applications. *Br. J. Radiol.* **2017**, *90*, 20160642. [CrossRef]
20. Alobaidli, S.; McQuaid, S.; South, C.; Prakash, V.; Evans, P.; Nisbet, A. The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning. *Br. J. Radiol.* **2014**, *87*, 20140369. [CrossRef]
21. Miles, K.A.; Ganeshan, B.; Hayball, M.P. CT texture analysis using the filtration-histogram method: What do the measurements mean? *Cancer Imaging* **2013**, *13*, 400–406. [CrossRef] [PubMed]
22. Traverso, A.; Wee, L.; Dekker, A.; Gillies, R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int. J. Radiat. Oncol. Biol. Phys.* **2018**, *102*, 1143–1158. [CrossRef] [PubMed]
23. Gudmundsson, S.; Runarsson, T.P.; Sigurdsson, S. Test-retest reliability and feature selection in physiological time series classification. *Comput. Methods Programs Biomed.* **2012**, *105*, 50–60. [CrossRef]

24. Donners, R.; Candito, A.; Blackledge, M.; Rata, M.; Messiou, C.; Koh, D.-M.; Tunariu, N. Repeatability of quantitative individual lesion and total disease multiparametric whole-body MRI measurements in prostate cancer bone metastases. *Br. J. Radiol.* **2023**, *96*, 20230378. [CrossRef] [PubMed]
25. Fournier, L.; Costaridou, L.; Bidaut, L.; Michoux, N.; Lecouvet, F.E.; de Geus-Oei, L.F.; Boellaard, R.; Oprea-Lager, D.E.; Obuchowski, N.A.; Caroli, A.; et al. Incorporating radiomics into clinical trials: Expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. *Eur. Radiol.* **2021**, *31*, 6001–6012. [CrossRef] [PubMed]
26. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [CrossRef] [PubMed]
27. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef] [PubMed]
28. Euser, A.M.; Dekker, F.W.; le Cessie, S. A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *J. Clin. Epidemiol.* **2008**, *61*, 978–982. [CrossRef] [PubMed]
29. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.33. Available online: https://mc-stan.org/docs/2_33/stan-users-guide-2_33.pdf (accessed on 21 April 2024).
30. Raunig, D.L.; McShane, L.M.; Pennello, G.; Gatsonis, C.; Carson, P.L.; Voyvodic, J.T.; Wahl, R.L.; Kurland, B.F.; Schwarz, A.J.; Gönen, M.; et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat. Methods Med. Res.* **2015**, *24*, 27–67. [CrossRef]
31. Chen, L.; Liu, M.; Bao, J.; Xia, Y.; Zhang, J.; Zhang, L.; Huang, X.; Wang, J. The correlation between apparent diffusion coefficient and tumor cellularity in patients: A meta-analysis. *PLoS ONE* **2013**, *8*, e79008. [CrossRef]
32. ElGendy, K.; Barwick, T.D.; Auner, H.W.; Chaidos, A.; Wallitt, K.; Sergot, A.; Rockall, A. Repeatability and test-retest reproducibility of mean apparent diffusion coefficient measurements of focal and diffuse disease in relapsed multiple myeloma at 3T whole body diffusion-weighted MRI (WB-DW-MRI). *Br. J. Radiol.* **2022**, *95*, 20220418. [CrossRef] [PubMed]
33. Hinzpeter, R.; Baumann, L.; Guggenberger, R.; Huellner, M.; Alkadhi, H.; Baessler, B. Radiomics for detecting prostate cancer bone metastases invisible in CT: A proof-of-concept study. *Eur. Radiol.* **2022**, *32*, 1823–1832. [CrossRef]
34. Wang, Y.; Yu, B.; Zhong, F.; Guo, Q.; Li, K.; Hou, Y.; Lin, N. MRI-based texture analysis of the primary tumor for pre-treatment prediction of bone metastases in prostate cancer. *Magn. Reson. Imaging* **2019**, *60*, 76–84. [CrossRef] [PubMed]
35. Zhang, W.; Mao, N.; Wang, Y.; Xie, H.; Duan, S.; Zhang, X.; Wang, B. A Radiomics nomogram for predicting bone metastasis in newly diagnosed prostate cancer patients. *Eur. J. Radiol.* **2020**, *128*, 109020. [CrossRef] [PubMed]
36. Gitto, S.; Bologna, M.; Corino, V.D.A.; Emili, I.; Albano, D.; Messina, C.; Armiraglio, E.; Parafioriti, A.; Luzzati, A.; Mainardi, L.; et al. Diffusion-weighted MRI radiomics of spine bone tumors: Feature stability and machine learning-based classification performance. *Radiol. Med.* **2022**, *127*, 518–525. [CrossRef] [PubMed]
37. Lang, N.; Zhang, Y.; Zhang, E.; Zhang, J.; Chow, D.; Chang, P.; Yu, H.J.; Yuan, H.; Su, M.Y. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magn. Reson. Imaging* **2019**, *64*, 4–12. [CrossRef] [PubMed]
38. Chianca, V.; Cuocolo, R.; Gitto, S.; Albano, D.; Merli, I.; Badalyan, J.; Cortese, M.C.; Messina, C.; Luzzati, A.; Parafioriti, A.; et al. Radiomic Machine Learning Classifiers in Spine Bone Tumors: A Multi-Software, Multi-Scanner Study. *Eur. J. Radiol.* **2021**, *137*, 109586. [CrossRef] [PubMed]
39. Dreher, C.; Kuder, T.A.; König, F.; Mlynarska-Bujny, A.; Tenconi, C.; Paech, D.; Schlemmer, H.P.; Ladd, M.E.; Bickelhaupt, S. Radiomics in diffusion data: A test-retest, inter- and intra-reader DWI phantom study. *Clin. Radiol.* **2020**, *75*, 798.e13–798.e22. [CrossRef]
40. Prescott, R.J. Editorial: Avoid being tripped up by statistics: Statistical guidance for a successful research paper. *Gait Posture* **2019**, *72*, 240–249. [CrossRef]