Repeatability and Reproducibility of MRI-based Radiomic Features in Rectal Cancer

Robba Rai (✉ robba.rai@health.nsw.gov.au)

University of New South Wales

Michael B. Barton

University of New South Wales

Phillip Chlap

University of New South Wales

Gary P. Liney

Ingham Institute for Applied Medical Research

Carsten Brink

University of Southern Denmark

Shalini Vinod

University of New South Wales

Monique Heinke

GenesisCare Sydney

Yuvnik Trada

Calvary Mater Newcastle Hospital

Lois C. Holloway

University of New South Wales

Research Article

Keywords: MRI, Radiomics, Rectal Cancer, Repeatability, Reproducibility, Stability

Posted Date: December 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1105008/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Radiomics of magnetic resonance images (MRI) in rectal cancer can non-invasively characterise tumour heterogeneity with potential to discover new imaging biomarkers. However, for radiomics to be reliable; the imaging features measured must be stable and reproducible. The aim of this study is to quantify the repeatability and reproducibility of MRI-based radiomic features in rectal cancer.

An MRI radiomics phantom was used to measure the longitudinal repeatability of radiomic features and the impact of post-processing changes related to image resolution and noise. Repeatability measurements in rectal cancers were also quantified in a cohort of ten patients with test-retest imaging amongst two observers.

We found that many radiomic features; particularly from texture classes, were highly sensitive to changes in image resolution and noise. 49% of features had coefficient of variations $\leq 10\%$ in longitudinal phantom measurements. 75% of radiomic features in *in vivo* test-retest measurements had an intraclass correlation coefficient of ≥ 0.8 . We saw excellent interobserver agreement with mean dice similarity coefficient of 0.95 ± 0.04 for test and retest scans.

The results of this study show that even when using a consistent imaging protocol many radiomic features were unstable. Therefore, caution must be taken when selecting features for potential imaging biomarkers.

Introduction

Radiomics has gained popularity during the last decade and is proving to be a promising method to non-invasively characterise tumour heterogeneity, treatment response, radiosensitivity and outcome prediction based on imaging features^{1–5}. With radiomics, medical images of tumours can be segmented using medical images and information on the first and second order features from the histogram, pixel intensity relationships and shape can be extracted using complex algorithms. The advantage of radiomics is the methodology can be applied to a number of tumour site groups² and is a non-invasive method of characterising tumours compared to traditional methods such as biopsy. With biopsies, a single tumour sample is often analysed which may not be representative of the tumour genomic landscape due to complex intratumour heterogeneity⁶. Radiomics can address the limitation of biopsy by using tumour segmentations from 3D volumetric medical imaging such as computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). Imaging can be repeated over time to measure the temporal heterogeneity of tumours.

For radiomics to be useful, the predictive models need to have high accuracy, reliability and efficiency⁷. Many radiomic features are highly correlated with one another and are therefore redundant and need to be removed prior to radiomics analysis to reduce the risk of overfitting the predictive models^{8,9}. This will also ensure that only the most informative features are included in the radiomics signature for a

predictive model. For this to occur, many studies in radiomics use a variety of datasets in their analysis, for training and validation^{2,10-12}.

Further investigations into the repeatability and reproducibility of features used, need to be performed¹³. Repeatability is defined as radiomic features that do not change when the object of interest is imaged multiple times¹³. Reproducibility can be defined as features that remain the same when imaged using different imaging apparatus such as different scanners, software or sequence parameters¹³.

Repeatability and reproducibility of radiomic features is imperative if predictive radiomic models are to be used in multicentre trials, where image quality can vary greatly for a number of reasons including different protocol acquisition parameters and machine vendors and fluctuations in image quality¹⁴. Quantifying the repeatability and reproducibility of radiomic features can be achieved with a rigorous assessment of feature stability including quality assurance (QA) of imaging protocols as well as measuring feature stability using in vivo test-retest datasets.

There has been recent interest in rectal cancer radiomics to predict response to treatments using MRI, because MRI provides superior soft tissue contrast compared to CT and PET. MRI is often used to diagnose, stage and restage rectal cancer¹⁵ and to determine the status of the circumferential resection margin which is associated with the risk of local recurrence¹⁶.

Recent studies into MRI in rectal cancer have explored creating radiomics models that can be used to predict pathological complete response (pCR) to neoadjuvant chemoradiotherapy (CRT). The studies often use a combination of anatomical and functional imaging such as DWI and DCE^{17,18} or anatomical T2-weighted (T2-w) imaging alone^{19,20}. Although the findings from this growing body of evidence is yielding promising results, there is still a need to explore the repeatability and reproducibility of radiomics features. This is particularly important for the building of predictive radiomics models which require reproducible and stable imaging features⁷. Traverso et al. (2018) in a systematic review, found that there was limited investigation into the most stable features for MRI-based radiomics analysis and recommended standardisation of radiomics analysis¹³.

The purpose of this study is to investigate the repeatability and reproducibility of MRI-based rectal cancer radiomic features using both phantom and clinical test-retest datasets.

Results

Patient Demographics

Table 1 outlines patient demographics. 10 patients were analysed in this study with each patient having two MRI scans. The average time $\pm$ SD between the test and retest scans was 11 ± 2 minutes.

Table 1
Patient characteristics

Characteristics	Patient Cohort
Number of Patients	10
Age (Years)	
Median	60
Q1-Q3	55 - 70
Gender	
Male	7
Female	3
TNM Staging	
IIA	0
IIIA	1 (1%)
IIIB	7 (70%)
IIIC	0
IVA	1 (1%)
IVB	1 (1%)
Time between Scans, mean $\pm$ SD	11 $\pm$ 2 (min)

Repeatability of radiomic features in the patient test-retest cohort

Table 2 summarises the ICC values for all radiomic feature classes. The mean ICC for observer 1 and 2 for each feature was used to categorise the features into three levels of agreement strength. Overall, 74.6% of radiomic features returned a high ICC > 0.8 with all shape features having a high strength of agreement. A complete list of ICC values for each radiomic feature for each observer can be found in Table 1 of supplementary material.

Table 2
Feature stability as measured by the mean ICC of observer 1 and 2 in the tumour

	Low (<0.5)	Moderate (0.5-0.8)	High (>0.8)
First Order Statistics	3	2	10
GLCM	2	1	20
GLRLM	1	5	7
GLSZM	1	4	7
NGTDM	-	1	4
Shape	-	-	14
Number of features per feature matrix			
First Order Statistics = 15, GLCM =23, GLRLM = 13, GLSZM = 13, NGTDM = 5, Shape = 14			

Interobserver Variability in the patient test-retest cohort

Table 2 of supplementary material shows the HD and DSC results from the tumour volume analysis between observers for the test and retest scans. There were no statistically significant differences between the tumour volumes of observer one and two for the test scans and retest scans using a paired sample T-test.

The mean $\pm$ SD volume for observer 1 and 2 for the test scans was 38.01 ± 18.89 ml and 36.83 ± 14.82 ml, respectively, with a mean difference of 1.18 ml ($p = 0.878$ [95% CI: -14.77-17.13]). The mean $\pm$ SD volume for observer 1 and 2 for the retest scan was 36.68 ± 17.26 ml and 39.19 ± 19.79 ml, respectively, with a mean difference of 2.52 ml ($p = 0.766$ [95% CI: -19.96-14.93]).

There was good interobserver agreement when considering the DSC with a mean $\pm$ SD DSC of 0.95 ± 0.04 for both the test and retest scans. The mean $\pm$ SD HD for the test and retest datasets was 0.061 ± 0.037 cm and 0.057 ± 0.042 cm

Feature stability in reference tissue

Table 3 lists the ICC values for the reference muscle across both texture and first order statistics. Overall, 75% of radiomic features returned a high ICC > 0.8 . Only one feature, variance from first order statistics, had a low ICC of 0.03.

Table 3
Feature stability as measured by the ICC in the reference muscle

	Low (<0.5)	Moderate (0.5-0.8)	High (>0.8)
First Order Statistics	1	4	10
GLCM	-	4	19
GLRLM	-	5	8
GLSZM		2	11
NGTDM		1	4
Number of features per feature matrix			
First Order Statistics = 15, GLCM = 23, GLRLM = 13, GLSZM = 13, NGTDM = 5			

Temporal stability of radiomic features in phantom

Table 4 shows a summary of the temporal stability of all features that had $\leq 10\%$ COV. 41/83 features had stable values measured as COV less than 10% on repeat imaging over a period of 4 months (Table 2 – Supplementary material).

Table 4
Radiomic features that had $\leq 10\%$ COV in the phantom using the MRI pulse sequence of the test-retest patient study

Feature Class	Feature	COV $\pm$ SD %
First Order	<i>Entropy</i>	3.9 $\pm$ 1.1
	<i>Kurtosis</i>	7.8 $\pm$ 1.9
	<i>Maximum</i>	6.4 $\pm$ 2.2
	<i>Mean</i>	6.1 $\pm$ 0.9
	<i>Mean Absolute Deviation</i>	8.0 $\pm$ 3.3
	<i>Median</i>	7.4 $\pm$ 3.1
	<i>Range</i>	6.6 $\pm$ 2.0
	<i>Root Mean Squared</i>	5.8 $\pm$ 0.7
	<i>Skewness</i>	9.5 $\pm$ 16.8
	<i>Uniformity</i>	8.5 $\pm$ 2.1
GLCM	<i>Difference Average</i>	9.2 $\pm$ 2.7
	<i>Difference Entropy</i>	4.3 $\pm$ 1.3
	<i>Inverse Difference</i>	3.1 $\pm$ 0.6
	<i>Inverse Difference Moment</i>	4.0 $\pm$ 0.9
	<i>Inverse Difference Moment Normalised</i>	0.4 $\pm$ 0.2
	<i>Inverse Difference Normalised</i>	0.8 $\pm$ 0.2
	<i>Information Measure of Correlation 1</i>	7.5 $\pm$ 3.1
	<i>Information Measure of Correlation 2</i>	5.7 $\pm$ 3.1
	<i>Inverse Variance</i>	4.5 $\pm$ 1.6
	<i>Joint Average</i>	5.1 $\pm$ 0.2
	<i>Joint Entropy</i>	4.3 $\pm$ 1.1
	<i>Maximum Probability</i>	6.5 $\pm$ 5.1
	<i>Sum Average</i>	5.1 $\pm$ 0.2
	<i>Sum Entropy</i>	3.5 $\pm$ 1.0
GLRLM	<i>Run Entropy</i>	1.7 $\pm$ 0.2
	<i>Run Length Non Uniformity Normalised</i>	3.4 $\pm$ 1.1

Feature Class	Feature	COV $\pm$ SD %
	<i>Run Percentage</i>	2.7 $\pm$ 0.7
	<i>Short Run Emphasis</i>	1.6 $\pm$ 0.6
GLSZM	<i>Zone Entropy</i>	4.6 $\pm$ 4.1
	<i>Elongation</i>	0.6 $\pm$ 0.3
	<i>Flatness</i>	6.1 $\pm$ 1.9
	<i>Least Axis Length</i>	6.9 $\pm$ 2.7
	<i>Major Axis Length</i>	1.4 $\pm$ 0.6
	<i>Maximum 2D Diameter Column</i>	2.0 $\pm$ 0.5
	<i>Maximum 2D Diameter Row</i>	1.4 $\pm$ 0.4
	<i>Maximum 2D Diameter Slice</i>	1.4 $\pm$ 0.7
	<i>Maximum 3D Diameter</i>	1.4 $\pm$ 0.4
	<i>Mesh Volume</i>	10.0 $\pm$ 3.4
	<i>Minor Axis Length</i>	1.6 $\pm$ 0.4
	<i>Sphericity</i>	2.5 $\pm$ 0.4
	<i>Surface Area</i>	4.5 $\pm$ 1.9
	<i>Surface Volume Ratio</i>	5.7 $\pm$ 0.6
	<i>Voxel Volume</i>	10.0 $\pm$ 3.4

Impact of post-processing on feature reproducibility

The original phantom image had an SNR of 73 (a.u.). The datasets with added Gaussian noise of 2, 5, 20 and 20 SD had a SNR of 57, 38, 18 and 10 (a.u.), respectively.

Figure 1-a is a heatmap demonstrating the ICC across the different noise iterations for each individual texture feature against each phantom test object. With an increase in noise the repeatability of some features from all feature classes decreases. Features from the GLSZM appear to be more sensitive to smaller changes in noise compared to features from GLCM, GLRLM and NGTDM.

In contrast, first order features (Figure 2 – a) appear to be more robust in the presence of added noise with the exception of maximum, uniformity and variance. Figure 2 shows the ICC across the different noise levels for each first order feature against each phantom test object. It appears that even with an added 20 SD of Gaussian noise, the majority of features have ICC $\geq$ 0.8.

Figure 1-b highlights the ICC across the different resolution settings for each individual texture feature against each phantom test object. Texture features appear to be more sensitive to changes in resolution particularly with GLSZM and NGTDM, with GLCM and GLRLM having more stable features out of the four feature classes.

The change in resolution also impacted the reproducibility of first order features particularly maximum, uniformity and variance, however the majority of features were highly reproducible (Figure 2 – b).

Discussion

This study has used phantom and test-retest imaging in a rectal cancer patient cohort to assess the stability of MRI-based radiomics features.

The results of the study show that shape features, in both the phantom and test-retest cohorts, was the most stable feature class which provided the most repeatable and reproducible results which is in agreement with previous repeatability studies in cervical²³ and prostate^{24,25} cancer.

In the phantom repeatability measurements, shape which was auto-defined, performed the best with all features having a $COV \leq 10\%$. The remaining features from first order and texture features classes had more varied results. One feature from GLSZM and no features from NGTDM class had $\leq 10\%$ COV for the test-retest imaging protocol. This is similar to the results of Gourtsoyianni et al.²⁶ who found COV ranging from 30 – 50% or greater for features from GLSZM and NGTDM, concluding that these features were not reliable for texture analysis in primary rectal cancer.

We found that there was a varied range of ICC in repeatability of texture and first order statistics features for the patient test-retest cohort. All shape features had an $ICC \geq 0.8$ with excellent interobserver agreement. This is not surprising as there was good interobserver agreement between the contoured volumes for both the test and retest scans as measured using DSC and HD. There was no statistically significant difference between the contoured volumes of both observers for the test and retest scans. The second set of tumour segmentations was completed by one observer using a copy and adjust method instead of an independent blinded contour. This may have contributed to the excellent interobserver agreement in DSC and mean HD values.

Similarly to this study, Timmeren et al. (2016)²⁷ found that shape features analysed with CT were the most stable in a cohort of rectal and lung cancer patients. Contrary to the findings in this study, Traverso et al.²⁸ found that first order statistics had higher repeatability compared to GLCM and GLRLM features, but found very poor inter observer repeatability of shape features in their rectal cancer dataset which they attributed to interobserver variability of the tumour delineation.

The investigations into the impact of added Gaussian noise and varied spatial resolution yielded noteworthy results. First order features were more robust in the presence of increased Gaussian noise and significant decreases of SNR compared to texture features. Overall, increasing or decreasing spatial

resolution using clinically relevant and realistic values severely impacted the reproducibility of most texture features. First order features and texture features from GLCM had more moderate ICC values with changes to resolution compared to GLRLM, GLSZM and NGTDM. These results are in agreement with previous studies where resolution was found to significantly impact the reliability of texture features in both phantom²⁹ and in vivo datasets^{28,30}. Mayerhoefer et al.²⁹ also found that texture features from the GLCM are superior to features from the GLRM with their ability to distinguish texture in the presence of spatial resolution perturbations.

There are several limitations to this study. The phantom repeatability measurements were only performed on three test objects. Further investigations into the impact of image quality changes should be performed on the remaining test objects as some texture measurements may not be suitable for analysis of tumours with more complex shapes¹⁴.

The patient numbers were also low, therefore the observed correlations should be validated and performed in a larger patient cohort. The radiomics analysis was only performed on T2-w images. Further investigations should look at assessing the repeatability and reproducibility for a range of sequences used in rectal cancer treatment and staging such as diffusion weighted imaging and dynamic contrast enhanced imaging.

Conclusion

This study has evaluated the performance of radiomics analysis using phantom and test-retest imaging of rectal cancers. Even using a consistent imaging protocol, many radiomic features are unstable in both in vivo and phantom experiments. Texture features from the GLRLM, GLSZM and NGTDM appear to be more sensitive to changes in image quality such as varied noise and resolution. Shape features performed the best of all features classes with consistently high reproducibility and repeatability in vivo and in the phantom. There is a need for controlled image quality and regular QA of imaging protocols if these features are to be used for radiomic models in multicentre trials

Materials & Methods

Radiomics Phantom

The radiomics phantom was developed by our group previously¹⁴ and was used to assess the repeatability of radiomic features for the imaging sequences used in clinical rectal cancer patients. The phantom was imaged twice during each imaging session over a four-month period to measure the temporal stability of radiomic features.

The phantom was positioned on two oil filled containers to ensure adequate signal to trigger a localiser scan (Figure 3-a). A non-slip mat was placed under the phantom base to minimise movement and vibrations caused by image acquisition.

The phantom is comprised of eight test objects of which three were used in this study for radiomics analysis including test objects M1, M2 and M8 (Figure 3-b). Only three test objects were selected for segmentation as these objects have a variation of textures and shapes, including a solid homogenous structure (M1), heterogeneous structure with fine texture details (M2) and a complex shape (M8). The test objects were segmented using a global threshold-based segmentation method described by Rai et al.¹⁴. The threshold values are based on the minimum and maximum pixel values. This was performed on one reference scan (test) and then propagated on the second scan (retest) to minimise contour variability. The upper and lower threshold pixel values were based on the histogram of the models.

Patient population

Retrospective analysis of a test-retest cohort of rectal cancer patients was performed. This study was approved by the South Western Sydney Human Research Ethics Committee (SWSLHD HREC) (HREC Number: HREC/15/LPOOL/555). All patients included in this study provided written informed consent. All experiments were performed in accordance with relevant guidelines and regulations.

Ten patients treated at Liverpool and Macarthur Cancer Therapy Centre between 2015 and 2017 with stage IIIA-IVB rectal adenocarcinoma were recruited. Inclusion criteria included any patient with confirmed rectal adenocarcinoma requiring a simulation MRI for their radiation therapy planning. Exclusion criteria were patients with a contraindication to MRI such as non-safe or non-compatible medical devices.

Image acquisition

All imaging in this study was conducted on a radiotherapy dedicated 3 Tesla MRI (MAGNETOM Skyra, Siemens Healthineers, Erlangen, Germany), using an 18-channel receiver only surface coil and 32-channel spine coil integrated into the MRI bed. A flat top radiation therapy table (CIVCO Medical Solutions, Coralville, USA) was used for both phantom and patient imaging to replicate patient treatment conditions (Figure 3-a).

The imaging protocol acquired for both phantom and patients was a T2-weighted (T2-w) turbo spin echo (TSE) sequence. The sequence was acquired with a TE/TR of 96/10000 ms, 3/0 mm slice thickness/spacing, 1 signal average, 320 x 224 matrix, 160° flip angle, 400Hz/Px receiver bandwidth, 220 mm field of view (FOV), and 0.7 x 0.7 mm² in-plane resolution.

Phantom Imaging

The phantom test objects were placed on a 3D printed horizontal plate in their corresponding position. The horizontal plate was aligned using the external laser bridge system (LAP Laser, Luneburg, Germany) at every scan session to ensure accurate reproducibility of position and to minimise setup errors. The phantom was imaged weekly for a period of 4 months.

Patient Imaging

All patients in the test-retest cohort were scanned in their radiation therapy (RT) treatment position including a flat table top and radiation therapy immobilisation equipment. The external laser system was used to align the patient to their planning tattoos (given in CT simulation) to assist with reproducing the patient position. After the first sequence was acquired (test) the patient left the MRI room and went for a short walk. They were then repositioned in the same RT planning positioning and the second T2-w TSE scan was acquired (retest). All ten patients received two scans for a total of 20 datasets.

Image post-processing

The impact of varied image quality on the reproducibility of radiomic features in the phantom was investigated to assess the impact it may have on radiomic feature stability.

To assess the impact on feature stability, the original phantom images were pre-processed with added noise and variable spatial resolution. To adjust the spatial resolution, the images were resampled to 0.5, 0.8, 1 and 1.2 mm². To add noise to the original images a Gaussian noise filter was added to the datasets. The standard deviation of the noise levels tested were set to 2, 5, 10 and 20 with the range of image intensities ranging from 0 to 302. For all images (both original and processed), the SNR within the individual images was assessed as the mean signal intensity within one phantom test object (M1) divided by the standard deviation of a nearby air signal. All post-processing was performed in 3D Slicer (Version 4.13.0, available at: <http://slicer.org/>).

Tumour and Tissue Segmentation

All segmentations were performed in imaging processing software MiM (MiM Maestro, Cleveland, OH, USA).

Tumour Segmentation

The rectal tumour volumes were contoured manually by two radiation oncologists (MH & YT) on both repeat T2-w sequences. Observer 2 (YT) performed a copy and adjust contour from the original contour of observer 1 (MH). The rectal tumour volume was defined as a high signal region on T2-w imaging by both observers. Care was taken to ensure only the tumour volumes excluded healthy rectal wall, lumen and faecal matter (Figure 4). This process was completed on both test and retest imaging by both observers.

Normal Tissue Segmentation

The right gluteus maximus was selected as a reference for normal tissue. Radiomic feature stability within a reference muscle was performed to quantify features that are stable in an otherwise normal human tissue.

The gluteus maximus was selected as it is covered in the field of view of all MRI scans across all datasets and was consistently outside the highest planning dose areas (< 20% of Max Dose). A spherical ROI was placed over the middle of the gluteus maximus on axial imaging. The size of the spherical ROI

was kept consistent across all scans with an average volume of 4 cc. The reference muscle was contoured manually by a single observer (RR) on T2-w imaging

Radiomics Analysis

Both in-vivo and phantom segmentations were analysed using PyRadiomics²¹ which is available in-house as a plug in within MiM software. 83 radiomic features were calculated including 15 first order statistics, 14 shape, and various textural features including 23 GLCM, 13 GLRLM, 13 GLSZM and 5 from NGTDM. All features were computed with a fixed bin width of 25. A list of all features examined can be found in supplementary material (Table 1).

Statistical Analysis & Feature Selection

For both the original and post-processed phantom imaging, the coefficient of variation (COV) was calculated to assess the degree of variability in features. The COV was calculated as:

$$\text{COV} = \frac{\text{standard deviation}}{\text{mean}} \times 100\%$$

where the standard deviation (SD) and the mean is derived from the radiomic features over all time points of imaging for each model.

The features were divided into four groups based on COV as very small ($\text{COV} \leq 5\%$), small ($5\% < \text{COV} \leq 10\%$), intermediate ($10\% < \text{COV} \leq 20\%$), and large ($\text{COV} > 20\%$) ranges of variation ²².

For the test-retest patient cohort, the intraclass correlation coefficient (ICC) test was performed to measure the reproducibility and repeatability of radiomic features between the test and retest scans for both observers for the rectal tumour volumes and normal muscles.

The impact of spatial resolution and Gaussian noise filtering on the reproducibility and repeatability of each radiomic feature were quantified using ICC. Features that returned an ICC of ≥ 0.8 were considered to have an almost perfect strength of agreement. ICC was calculated in R using the DescTools package (Version 3.5.1).

References

1. Leijenaar, R. T. H. *et al.* Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol. (Madr)*. **52**, 1391–7 (2013).
2. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**, 4006 (2014).
3. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–6 (2012).

4. Gillies, R. J., Anderson, A. R., Gatenby, R. A. & Morse, D. L. The Biology Underlying Molecular Imaging in Oncology: From Genome to Anatome and Back Again. *Clin. Radiol.* **65**, 517–521 (2010).
5. Kumar, V. *et al.* QIN “Radiomics: The Process and the Challenges”. *Magn. Reson. Imaging* **30**, 1234–1248 (2013).
6. Clark, W. M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **353**, 977–987 (2005).
7. Yip, S. S. F. & Aerts, H. J. W. L. Applications and limitations of radiomics. *Phys. Med. Biol.* **61**, R150–R166 (2016).
8. Papanikolaou, N., Matos, C. & Koh, D. M. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* **20**, 1–10 (2020).
9. Larue, R. T. H. M., Defraene, G., De Ruysscher, D., Lambin, P. & Van Elmpt, W. Quantitative radiomics studies for tissue characterization: A review of technology and methodologic procedures. *Br. J. Radiol.* **90**, (2017).
10. Yang, L. *et al.* Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur. Radiol.* **29**, 2196–2206 (2019).
11. Parmar, C. *et al.* Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Nat. Sci. reports* **5**, 1–10 (2015).
12. Shen, C. *et al.* 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. *Transl. Oncol.* **10**, 886–894 (2017).
13. Traverso, A., Wee, L., Dekker, A. & Gillies, R. Repeatability and reproducibility of radiomic features: A systematic review. *Int. J. Radiat. Oncol.* **102**, 1143–1158 (2018).
14. Rai, R. *et al.* Multicenter evaluation of MRI-based radiomic features: A phantom study. *Med. Phys.* **47**, 3054–3063 (2020).
15. Beets-Tan, R. G. H. *et al.* Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur. Radiol.* **28**, 1465–1475 (2018).
16. LeBlanc, J. K. Imaging and management of rectal cancer. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **4**, 665–676 (2007).
17. Nie, K. *et al.* Rectal cancer: Assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin. Cancer Res.* **22**, 5256–5264 (2016).
18. Liu, Z. *et al.* Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin. Cancer Res.* clincanres.1038.2017 (2017) doi:10.1158/1078-0432.CCR-17-1038.
19. Horvat, Nataly and Veeraraghavan, Harini and Khan, Monika and Blazic, Ivana and Zheng, Junting and Capanu, Marinela and Sala, Evis and Garcia-Aguilar, Julio and Gollub, Marc J and Petkovska, I. MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology* **287**, 833–843 (2018).

20. De Cecco, C. N. *et al.* Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. *Invest. Radiol.* **50**, 239–245 (2015).
21. Van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
22. Yan, J. *et al.* Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. *J. Nucl. Med.* **56**, 1667–1673 (2015).
23. Fiset, S. *et al.* Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother. Oncol.* **135**, 107–114 (2019).
24. Fedorov, A., Vangel, M. G., Tempany, C. M. & Fennessy, F. M. Multiparametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. *Invest. Radiol.* **52**, 538–546 (2017).
25. Schwier, M. *et al.* Repeatability of Multiparametric Prostate MRI Radiomics Features. *Sci. Rep.* **9**, 1–16 (2019).
26. Gourtsoyianni, S. T. D. R. of M. I. T. F. in P. R. C. G. *et al et al.* Primary rectal cancer: Repeatability of Global and Local- Regional MR Imaging Texture Features. *Radiol. n Radiol. Radiol.* **284**, (2017).
27. Timmeren, J. E. Van *et al.* Test–Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? *Tomography* **2**, 361–365 (2016).
28. Traverso, A. *et al.* Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing. *Phys. Medica* **61**, 44–51 (2019).
29. Mayerhoefer, M. E., Szomolanyi, P., Jirak, D., Materka, A. & Trattnig, S. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. *Med. Phys.* **36**, 1236–1243 (2009).
30. Brynolfsson, P. *et al.* Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters. *Sci. Rep.* **7**, 1–11 (2017).

Figures

Figure 1

Heatmap of all intraclass correlation coefficients (ICC) for each variation in image quality across the three test objects M1, M2 and M8 for texture features GLCM, GLRLM, GLSZM and NGTDM. (a) shows the ICC for all noise levels (2, 5, 10 and 20 SD) and (b) shows the ICC for all resampled resolution iterations introduced, except for 0.8 mm which is the original resolution. The columns correspond to each feature and the rows correspond to the test object.

Figure 2

Heatmap of all intraclass correlation coefficients (ICC) for each variation in image quality across the three test objects M1, M2 and M8 for first order features. (a) shows the ICC for all noise levels (2, 5, 10 and 20 SD) and (b) shows the ICC for all resampled resolution iterations. The columns correspond to each feature and the rows correspond to the test object.

Figure 3

(a) Phantom setup on radiation therapy table, and (b) individual test objects part of the radiomics phantom

Figure 4

Example of contours of a rectal tumour on axial (a), coronal (b) and sagittal (c) reconstructions. Care was taken to ensure rectal wall, lumen and faecal matter was not included in the tumour volumes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)