Physics and Imaging in Radiation Oncology 23 (2022) 48–53

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

Clinical and radiomics prediction of complete response in rectal cancer pre-chemoradiotherapy

Peter Mbanu [a,*], Mark P. Saunders [a], Hitesh Mistry [b,f], Joe Mercer [c], Lee Malcomson [b,e], Saif Yousif [d], Gareth Price [b], Rohit Kochhar [c], Andrew G. Renehan [b,e], Marcel van Herk [b], Eliana Vasquez Osorio [b]

[a] Department of Clinical Oncology, Christie Hospital NHS Foundation Trust, Manchester, United Kingdom
[b] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom
[c] Department of Radiological Oncology, Christie Hospital NHS Foundation Trust, Manchester, United Kingdom
[d] Department of Clinical Oncology, Lancashire Teaching Hospital, Preston, United Kingdom
[e] Colorectal and Peritoneal Oncology Centre, Christie Hospital NHS Foundation Trust, Manchester, United Kingdom
[f] Division of Pharmacy, University of Manchester, Manchester, United Kingdom

ARTICLE INFO

Keywords:
Rectal cancer
Chemoradiotherapy
Magnetic resonance imaging
Radiomics
Treatment response

ABSTRACT

Background and purpose: Patients with rectal cancer could avoid major surgery if they achieve clinical complete response (cCR) post neoadjuvant treatment. Therefore, prediction of treatment outcomes before treatment has become necessary to select the best neo-adjuvant treatment option. This study investigates clinical and radiomics variables' ability to predict cCR in patients pre chemoradiotherapy.

Materials and methods: Using the OnCoRe database, we recruited a matched cohort of 304 patients (152 with cCR; 152 without cCR) deriving training (N = 200) and validation (N = 104) sets. We collected pre-treatment MR (magnetic resonance) images, demographics and blood parameters (haemoglobin, neutrophil, lymphocyte, alkaline phosphate and albumin). We segmented the gross tumour volume on T2 Weighted MR Images and extracted 1430 stable radiomics features per patient. We used principal component analysis (PCA) and receiver operating characteristic area under the curve (ROC AUC) to reduce dimensionality and evaluate the models produced.

Results: Using Logistic regression analysis, PCA-derived combined model (radiomics plus clinical variables) gave a ROC AUC of 0.76 (95% CI: 0.69–0.83) in the training set and 0.68 (95% CI 0.57–0.79) in the validation set. The clinical only model achieved an AUC of 0.73 (95% CI 0.66–0.80) and 0.62 (95% CI 0.51–0.74) in the training and validation set, respectively. The radiomics model had an AUC of 0.68 (95% CI 0.61–0.75) and 0.66 (95% CI 0.56–0.77) in the training and validation sets.

Conclusion: The predictive characteristics of both clinical and radiomics variables for clinical complete response remain modest but radiomics predictability is improved with addition of clinical variables.

1. Introduction

For patients with locally advanced rectal cancer, the standard of care treatment is neoadjuvant chemoradiotherapy followed by resective surgery either as a total mesorectal excision surgery or abdominoperineal resection [1]. Surgical resection is associated with considerable short and long-term morbidity; up to 3% risk of perioperative mortality and 40% risk of requiring a permanent stoma [2]. For two decades, pathological complete response (pCR) (the absence of microscopic disease post-resection) has been recognised in 15%–27% of patients who had resective surgery post-chemoradiotherapy [3]. Clinical complete response (cCR) is the absence of clinically and radiologically detectable disease post neoadjuvant chemoradiotherapy and pre resective surgery. The identification of cCR followed by a decision between the patient and oncologist to actively monitor or 'watch and wait (W&W)', pioneered by Habr-Gama et al. [4], has become a novel management strategy to

* Corresponding author at: Department of Clinical Oncology, Christie NHS Foundation Trust, 450 Wilmslow Road, Manchester M20 4BX, United Kingdom.
E-mail address: Peter.mbanu1@nhs.net (P. Mbanu).

https://doi.org/10.1016/j.phro.2022.06.010
Received 29 September 2021; Received in revised form 18 June 2022; Accepted 20 June 2022
Available online 28 June 2022

reduce surgery-related morbidity, mortality and permanent stoma. Patients with cCR on watch and wait are carefully monitored on a surveillance follow-up plan, and surgery is only needed in the event of a disease re-growth. About 25–30% of patients on this surveillance treatment plan will require surgery within the first three years after neoadjuvant treatment, while the rest will be managed without the need for surgery [5,6]. The growing evidence of organ preservation in rectal cancer has shown this management plan to be valid and safe without any decline in clinical outcome [5,7–9]. Pooled analysis of trials showed that pCR is associated with a good prognosis and an indicator of a biologically favourable tumour [3]. Patients with cCR also have a comparable excellent long term outcome similar to those with pCR [5]. Clinical complete response has become a surrogate for the pathological complete response for selecting patients who may not require surgery. Therefore it remains essential to predict patients likely to follow this treatment plan from the onset of their diagnosis in the era of personised medicine.

However, there is no robust predictor of either pCR or cCR before neoadjuvant chemoradiotherapy. Ryan et al. reported a systematic review, including 85 studies, evaluating predictors (including biochemical, gene expression, mutational, and protein expression analyses) for pCR but concluded that there were 'no robust markers' [10]. Since the mid-2000s, magnetic resonance imaging (MRI) has been the standard of care for pre-treatment staging in rectal cancer patients [11]. MRI has shown good accuracy in determining the size and stage of the rectal tumours. It is also crucial in determining invasion into the mesorectal fascia, which is an essential factor in deciding if neo-adjuvant treatment is required [12]. However, the review by Ryan et al. concluded that volumetric measurement on standard pre-treatment MRI had not been shown accurately to predict the response [10].

MRI radiomics is an alternative dimension beyond standard clinical MR imaging, which might predict complete response. Radiomics is the mining and analysis of large amounts of advanced quantitative imaging features from routinely performed radiological investigations [13,14], from which statistical modelling can be used to predict treatment outcomes. Thus, radiomics analysis has the potential utility as a biomarker in treatment selection. Many studies [15,16–18] have evaluated radiomics features' predictive abilities for pCR using pre-treatment MRI scans alone. Notably, the end-points of these studies were either pCR or pathological tumour regression, which requires surgical resection to assess. In addition, the number of complete response cases was relatively small, which may lead to overfitting radiomics features' contributions in the prediction models.

Currently, different treatment options are available in neo-adjuvant rectal cancer treatment with differing toxicities. Predicting treatment response before treatment is essential in selecting the best treatment plan for a patient. Patients predicted to have cCR with chemoradiotherapy could have this treatment. In contrast, others could have an alternative neoadjuvant plan such as total neoadjuvant therapy or intensive doublet chemotherapy [19,20–21] in other to improve their overall treatment outcome. To produce a robust predictive model, we performed a matched cohort study enhancing the number of cCR cases to 152 to create a predictive model for cCR combining radiomics features and routinely collected clinical parameters.

2. Methods

2.1. Study population

All appropriate research governance and ethics approval was obtained before starting this study (IRAS 265989). All patients recruited received their treatment either at the Christie NHS Foundation Trust or Lancashire University Teaching Hospital, both cancer centres in the north of England, UK. We recruited patients primarily from the OnCoRe (The Rectal Cancer Oncological Complete Response Database) database, The OnCoRe is a research database of patients who achieved clinical complete response.

All patients had locally advanced rectal adenocarcinoma and received neo-adjuvant long course chemoradiotherapy between 2008 and 2019. 395 patients were selected consecutively from the database – 165 patients with cCR and 230 patients without cCR, non-clinical complete response (NcCR). From these, MR images for four patients with cCR were not available. Propensity score matching of 0.1 callipers based on T-stage, age, N-Stage and performance status was used to select 161 patients out of 230 patients without complete response. Propensity score matching was used in this study to ensure that patients in both cohorts have equal numbers of similar baseline characteristics. A propensity matching of 0.1 resulted in the lowest bias in a study comparing different propensity widths [22]. After segmentation, 9 and 4 patients out of 161 patients belonging to the cCR and NcCR groups respectively were removed due to either low-quality MR images or incomplete tumour coverage in the required MR sequence. A re-run of the propensity matching was done after segmentation to select an equal number of patients in both cohorts. Finally, 152 patients from both groups (304 patients) were enrolled in this study. (See supplementary A, Fig. S1).

Clinical variables of each patient: demographics (age, gender, T-stage, N-stage tumour diameter) and blood parameters (haemoglobin, neutrophil, lymphocyte, alkaline phosphate and albumin) were obtained from the clinical records held at the treating institution. All the clinical parameters including the blood parameters were taken pre-treatment**.** Patients were then split into two groups; a training group and a validation group at the ratio of 2:1. In line with prospective studies to limit the selection bias of retrospective studies, allocation to the training and validation cohort was done using the patient's date of diagnosis rather than random assignment. Thus, the first 100 patients of the cCR and NcCR group were placed in the training cohort and the last 52 patients of the cCR and NcCR group in the validation cohort.

2.2. Neo-adjuvant chemo-radiotherapy

All patients were aged 18 and over and underwent conformal planned pelvic radiotherapy, concurrent with capecitabine 825 mg/m^2 twice daily during treatment. They all received a prescribed dose of 45 Gy in 25 fractions of pelvic radiotherapy. Post radiotherapy, they were all restaged with a pelvic MRI and CT imaging at 8–10 weeks. Patients that did not have a viable radiological tumour on this imaging (MRTRG 1 and 2) were further investigated with a digital rectal examination (DRE) and colonoscopy. The multi-disciplinary team meeting independently verified the investigations. The absence of residue disease in all three investigatory modalities is defined as 'clinical complete response'. Patients with cCR were offered 'watch and wait' surveillance. The patient's population characteristics are summarized in Table 1.

2.3. MRI and segmentation

Retrospective pre-treatment MR pelvis sequences of recruited patients were acquired. All images were scanned on a 1.5 T diagnostic MR with a 24 cm field of view, 3-mm slide thickess and no intersection gap. Transverse T2-weighted (T2W) high-resolution axial MR images was the selected sequence. The T2W fast spin-echo sequence images were acquired in a plane orthogonal to the tumour longitudinal axis. No contrast was given during image acquisition. T2WI sequence is chosen to reflect the most commonly used sequence in previous published MR radiomics work in rectal cancer [23]. The images were segmented in the contouring software RayStation v6.99.

A clinical oncologist and a radiologist, both with expertise in lower GI malignancies, performed image segmentation. The region of interest (ROI) in this study is the segmented tumour volume. (A representation of a segmented slide is seen in the supplementary A, Fig. S2).

Twenty-one patients were randomly selected and independently segmented by both the clinical oncologist and the radiologist to investigate inter-observer variations. The two volumes were assessed for consistency, using volumetric differences, dice coefficient, distance to

Table 1
Demographic table. Table showing the baseline characteristics of the two groups.

Characteristics	cCR group (n = 152)	NcCR group (n = 152)
Age in years (Mean and range)	66.3 (41–86)	66.8 (31–89)
Gender (Male: Female)	111(73%) Male 41 (27%) Female	99(65%) Male 53 (35%) Female
T staging T2	31 (20%)	10 (7%)
T3	108 (71%)	125 (82%)
T4	13 (9%)	17 (11%)
N staging N0	39 (26%)	35 (23%)
N1	65 (43%)	66 (43%)
N2	48 (31%)	47 (31%)
N3	0	4 (3%)
Tumour diameter* (cm) (Mean / range)	4.8 (2–10)	5.5 (2–10)
Height above anal margin (cm) (Mean/range)**	5.9 (0–15)	6.2 (0–18)

*Tumour diameter is the maximum craino-caudal length of the tumour measured on the sagittal MRI planes.
**Height above the anal margin is the length from the most distal part of the tumour to the anal verge measured on a sagittal MR image plane.

agreement, Hausdorff distance and intra-class correlation (ICC). (See supplementary B, table S1).

2.4. Feature extraction and image normalisation

DICOM files containing the MR image segmentation were exported from RayStation. Nifti files for the MR and the rasterised delineations (masks) were then created from the Dicom files using in-house software. The images were then normalised before extraction of features using histogram intensity normalisation. Histogram normalisation has shown to increase radiomics features reproducibility in a recent work looking at normalisation effects on the reproducibility of radiomics features relating to T2WI of the pelvis [24]. MR images intensity were normalised in this study by applying histogram intensity matching using an arbitrary MR image as the reference (first image in the folder) [25].

As the images were acquired with different angles, we followed the recommendation of IBSI [26] and resample the images to a 3 mm isotropic resolution. In addition, we used Fixed Bin Size (FBS) as recommended in several reports [24,27] on MR feature reproducibility. All features available in pyradiomics v 3.0 were calculated (except for 2D specific features) on the original image and on the following filtered images: Laplacian of Gaussian (LOG, for edge detection, using sigma 3 and 5), Wavelet, Square, SquareRoot, Logarithm, Exponential, Gradient and local binary pattern (LBP) [28] (See supplementary C). A total of 1781 radiomics features were extracted per patient. The quantitative values of these features were tableted for feature selection and statistical analysis.

2.5. Feature selection

The features extracted from the twenty-one patients contoured by both clinicians were assessed to determine their stability. Features were extracted from both sets of images independently segmented by two observers. Using the intra-class correlation coefficient (ICC), the features with excellent correlations in the two cohorts were selected as stable. An ICC greater than 0.90 suggests excellent reliability [29]. We accepted features with an ICC of more than 0.9 to be stable features.

2.6. Principal component analysis (PCA)

Feature reduction is achieved through two forms of dimensionality reduction process; supervised or unsupervised. We choose to use unsupervised feature reduction in this study due to its beneficial characteristics over supervised feature reduction which are prone to overfitting [30]. Unsupervised dimensional reduction is robust against overfitting and, therefore, more suitable [31]. The most commonly used unsupervised approaches in radiomics work are cluster analysis and PCA (28). PCA creates new variables from the existing ones which are uncorrelated and which maximise the variance captured in the data-set. PCA has returned the highest predictive performance in radiomics studies [32,33]. We performed principal components analysis on the dataset and clustered the observations using hierarchical clustering on the factor map. Using the Euclidean distance metric, the optimal number of clusters was determined by assessing the loss in entropy.

2.7. Multivariable analysis

First, PCA was applied to the radiomics features as above mentioned. Then the PCA generated variables were clustered using hierarchical clustering. The clusters were assessed for variation in tumour diameter and volume to evaluate whether the variations captured by PCA is only representing differences in tumour size or diameter. Next, we pooled the first two principal components, which accounts for most of the variation in the data, PC1 and PC2, as explanatory variables to construct three logistic regression models; combined radiomics and clinical model, clinical only model, and a radiomics model. These models were built on the training cohort, assessed with the validation cohort using ROC AUC.

3. Results

3.1. Interobserver analysis and feature selection

Contours between the clinical oncologist and the radiologist were consistent. The average dice coefficient was 0.85 (range 0.78 to 0.92), average mean distance-to-agreement (DTA) was 0.08 cm (range 0.05 to 0.15). The average Hausdorff distance was 0.55 cm (range 0.3 to 1.3 cm). The ICC of the volumes generated was 0.998 (CI 0.995–0.999), showing excellent consistency (See supplementary B, table S1). 1430 out of 1781 features were selected for analysis as stable features.

3.2. Principal component analysis/ hierarchical clustering

Four clusters were identified by applying PCA hierarchical clustering to the selected 1430 radiomics features. We found that the probability of cCR correlates with the cluster groups (likelihood ratio-test p-value = 0.007). The odds ratio for cluster 4 vs 1 is 3.14 (95% CI: 1.56–6.46), 3 vs 1 is 2.11 (95% CI: 1.05–4.29).

The distribution of voxel volume (tumour volume) and diameter within the PCA derived clusters showed no significant differences (see Fig. 1). Moreover, we found that PC1 and PC2 remained correlated to cCR after adjusting for tumour diameter within a logistic regression analysis. Multivariable analysis of the two principal components and tumour diameter against cCR showed that PC1 has an odds ratio of 1.26 (95% CI 1.12–1.42) and a p-value of < 0.001. PC2 had an odds ratio of 0.92 (95% CI 0.84–1.00), p-value of 0.062 and the tumour diameter 0.85 (95% CI 0.71–1.03) and a p-value of 0.094. Indicating that variations represented by PC1 and PC2 are independent of volumetric tumour measurements.

3.3. Multivariable logistic regression models

The multivariable logistic regression generated three models; clinical, radiomics and combined clinical and radiomics model. Comparing model likelihoods, we found that the inclusion of the radiomics variable improved the model fit of the combined model, $p = 0.006$ (see Tables 2 and 3). The accumulative weights of the radiomics feature group that form part of the model is seen on supplementary D, tables S2 and S3.

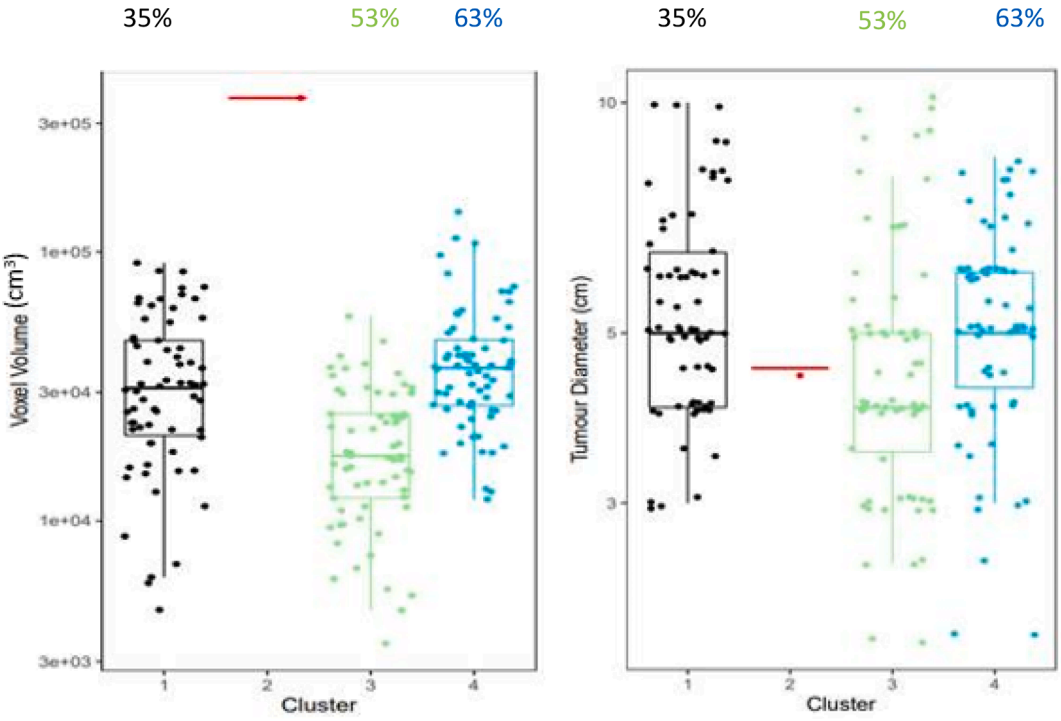

Fig. 1. Hierarchical clustering using the leading principal components is plotted against tumour volume (Voxel volume in cm^3) and Diameter in cm. The distribution across the cluster groups shows that the clusters are independent of volume and diameter. Cluster 2 is an outlier in our database.

Table 2
Multivariable clinical and radiomics logistic regression analysis – training set.

	ROCAUC-0.76 (95% CI: 0.69–0.83)	
	OR (95% CI)	p-value
PC1/10	**1.23 (1.07–1.41)**	**0.004**
PC2/10	0.90 (0.80–1.01)	0.061
Diameter (cm)	0.89 (0.72–1.11)	0.309
Age/10 (years)	0.86 (0.62–1.20)	0.375
Sex		
Female v Male	0.86 (0.40–1.84)	0.691
T-Stage		
3 v 2	0.41 (0.14–1.24)	0.115
4 v 2	**0.21 (0.05–0.96)**	**0.044**
N-Stage		
1 v 0	0.93 (0.40–2.16)	0.869
2/3 v 0	0.75 (0.30–1.89)	0.545
Hb/10 (g/L)	**1.27 (1.00–1.60)**	**0.047**
Neutrophils ($x10^9$/L)	1.01 (0.83–1.22)	0.945
Lymphocytes ($x10^9$/L)	1.27 (0.86–1.88)	0.232
log(Alkalinephosphatase(log iu/L)	**0.23 (0.06–0.83)**	**0.024**
Albumin (g/L)	0.92 (0.82–1.04)	0.196

Hb- Haemoglobin, g/L- grams per litre, iu/L- units per litre, cm-centimetre. Hightlighted variables have a p value < 0.05.

Table 3
Evaluation of the models. Table is comparing the AUC value between the training and validation cohort of each model. The models with clinical variables have notable differences in AUC.

	ROC AUC (95% CI)	
	Training	Validation
Clinical alone	0.73 (0.66–0.80)	0.62 (0.51–0.74)
Radiomics alone	0.68 (0.61–0.75)	0.66 (0.56–0.77)
Clinical and Radiomics	0.76 (0.69–0.83)	0.68 (0.57–0.79)

3.4. Evaluation of the model

Table 3 gives the evaluation of the three models in both cohorts. In the models containing clinical variables, the AUC values have dropped significantly. The radiomics only model has a similar AUC between the training and validation model.

The drop in ROC AUC between the validation and training cohorts was not due to differences in patient demographics (see Table 4). To assess why such a significant drop was seen when using clinical variables, we performed a multivariable logistic regression analysis on the validation cohort and compared the results to the training cohort (see Table 5).

Looking at the clinical variables and comparing the odd ratios of individual variables in the training and validation cohorts, it is clear that the main clinical drivers of cCR have changed significantly, with shifts in some of the clinical variables which may have affected the validation of the models containing the clinical variables (highlighted variables in Table 5).

4. Discussion

Our study compared the discriminative characteristics performance of a combined radiomics and clinical model with a clinical or radiomics only model. We found that the clinical variables on their own (based on ROC AUC) are potentially a better predictor of cCR than radiomics variables alone. However, the models containing the clinical variables failed to validate successfully. Even though the study was designed to minimise its risk, overfitting could cause this discrepancy. Another possible reason is calibration drift. The predictive model reduces performance with calibration drift [34,35] as the outcomes are reported differently over time. A recent paper [36] recommended that clinical prediction models be continuously updated and monitored to remain relevant over time. A dynamic prediction model approach [37], whereby a model consecutively adjusts to changes in population demographics, disease incidence, and clinical practice over time, has been proposed as a potential solution to this problem. A notable example of a

Table 4
Patient's characteristics between the two cohorts. The training and validation cohort show similar baseline characteristics.

	Training (N = 200)	Validation (N = 104)
PC1		
median (range)	−6.8 (−53.2–95.0)	−2.1 (−48.1–86.6)
PC2		
median (range)	−7.9 (−65.6–513.2)	−8.6 (−60.1–144.7)
Diameter (cm)		
median (range)	5 (2, 10)	5 (2, 9)
Age (years)		
median (range)	66 (31–89)	68 (36–90)
Sex – N (%)		
Female	62 (31)	31 (30)
Male	138 (69)	73 (70)
T-Stage – N (%)		
2	24 (12)	16 (15)
3	155 (78)	79 (76)
4	21 (10)	9 (9)
N-Stage (%)		
0	49 (25)	24 (23)
1	86 (43)	45 (43)
2	61 (31)	35 (34)
3	4 (2)	0 (0)
Hb (g/L)		
median (range)	13.4 (7.7–16.6)	13.5 (8.7–16.9)
Neutrophils ($x10^9$/L)		
median (range)	4.7 (1.7–12.4)	5.0 (1.9–11.4)
Lymphocytes ($x10^9$/L)		
median (range)	1.7 (0.3–6.1)	1.7 (0.4–4.7)
Alkaline Phosphatase(iu/L)		
median (range)	80 (40–155)	83 (42–158)
Albumin (g/L)		
median (range)	44 (24–51)	44 (31–49)

Hb- Haemoglobin, g/L- grams per litre, iu/L- units per litre, cm-centimetre.

Table 5
Multivariable logistic regression analysis in the training and validation cohort. This table shows the odd-ratios and p values of the variables. Highlighted variables show a major difference in odd ratios between the two cohorts.

	Training (N = 200)		Validation (N = 104)	
	OR (95% CI)	p-value	OR (95% CI)	p-value
PC1/10	1.23 (1.07–1.41)	0.004	1.23 (0.98–1.54)	0.078
PC2/10	0.90 (0.80–1.01)	0.061	1.02 (0.86–1.20)	0.853
Diameter (cm)	**0.89 (0.72–1.11)**	**0.309**	**0.58 (0.38–0.88)**	**0.012**
Age/10 (years)	0.86 (0.62–1.20)	0.375	1.30 (0.73–2.31)	0.377
Sex				
Female v Male	0.86 (0.40–1.84)	0.691	1.13 (0.29–4.42)	0.856
T-Stage				
3 v 2	0.41 (0.14–1.24)	0.115	0.07 (0.01–0.48)	0.007
4 v 2	0.21 (0.05–0.96)	0.044	0.35 (0.02–0.52)	0.447
N-Stage				
1 v 0	0.93 (0.40–2.16)	0.869	0.95 (0.22–4.20)	0.947
2 v 0	**0.75 (0.30–1.89)**	**0.545**	**5.86 (1.16–29.7)**	**0.033**
Hb/10 (g/L)	1.27 (1.00–1.60)	0.047	1.14 (0.76–1.69)	0.531
Neutrophils ($x10^9$/L)	1.01 (0.83–1.22)	0.945	0.77 (0.54–1.09)	0.144
Lymphocytes ($x10^9$/L)	**1.27 (0.86–1.88)**	**0.232**	**0.56 (0.21–1.50)**	**0.250**
log(Alkaline Phosphatase) (log iu/L)	0.23 (0.06–0.83)	0.024	0.85 (0.09–7.74)	0.887
Albumin (g/L)	0.92 (0.82–1.04)	0.196	0.91 (0.74–1.12)	0.381

Hb-Haemoglobin, g/L-grams per litre, iu/L-units per litre, cm-centimetre.

clinical prediction model updated yearly and revised to include additional predictors is the QRISK [38]. Our results showed that even though the predictability of the radiomics only model is lower at ROC AUC of 0.68, it is unaffected by calibration drift and was validated successfully. This study represents the most extensive MR radiomics work with patients who had chemoradiotherapy in rectal cancer to the best of our knowledge. It also recruited patients who had a clinical complete response (cCR), the target group for organ preservation treatment pathway, unlike other studies that predominantly used patients with pCR. We used a large 1:1 matched cohort of patients (Ccr; N = 152) and those without; N = 152), representing the largest proportion of patients with complete response in any rectal cancer radiomics work.

Very few studies have used mono sequence pre-treatment images in the prediction of complete response. A similar study [39] used an intensity histogram to predict pCR with external validation. Their results showed an AUC of 0.73 and 0.75 on external validation. Although this study demonstrated good predictability of the radiomics variable, the proportion of patients with pCR in the whole database is less than 30% which could skew the results. Patients in this study were also not matched to ensure a reduction in selection bias.

This study comes with some limitations; firstly, all patients received their radiotherapy treatment in either two hospitals in the same region. A more diverse database would have been preferable to further reduced selection bias, which can be inherent in a few centre studies. This bias could be said to have been reduced by the use of propensity matching. The validation cohort in this study was chosen internally even though the recruited patients were treated in two institutions, an external validation cohort from a different regional hospital may have provided extra validity. The analysis of this study assumed cCR to be a binary response. The reality is that patients without cCR have a wide variety of responses; near-complete, partial, stable, and no response, so a future radiomics study should look at predictors of good response to neoadjuvant treatment by combining preferred clinical outcomes in one group. Inter-observer variation in the segmentation of radiomics work has been a source of bias. This bias was reduced in this study by ensuring consistency between the two clinicians involved in the segmentation. For sizeable radiomics work in the future, there is a need to develop automatic contouring software to allow radiomics in day-to-day clinical practice. In the future, it will be expected that contouring of ROI in radiomics will be done by automatic delineation tools [40]. The most frequently used MR radiomics sequence is the T2WI; this was used as the protocol sequence in this work. It could be that combining different sequences might improve the predictability of radiomics features. A study [17] using MR radiomics in rectal cancer to predict pCR, showed that combining different image sequences performed better than using one sequence. Although this study showed an improved radiomics model with multiple sequences, it had only 31 patients with pCR out of 186 patients recruited in the study, which is a significant drawback.

The predictive abilities of our clinical variables, with or without radiomics, are modest, as demonstrated in this study. The predictive capabilities of the radiomics variables for cCR are improved by adding the clinical variables, but the absolute gains remain low. New approaches are essential to improve the predictability of cCR for neoadjuvant treatment selection in rectal cancer. Future approaches could investigate the addition of radiotherapy biomarkers such as hypoxia, gene expression signatures and deep learning techniques. Molecular markers could also be a valuable addition to a clinical model to improve the model's predictability.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2022.06.010.

References

[1] Glynne-Jones R, Wyrwicz L, Tiret E, Brown G, Rödel C, Cervantes A, et al. ESMO Guidelines Committee. *Ann Oncol* 2018;28:22–40. https://doi.org/10.1093/annonc/mdx224.

[2] National Bowel Cancer Audit Annual Report 2020 – HQIP. Available from: https://www.hqip.org.uk/resource/national-bowel-cancer-audit-annual-report-2020.

[3] Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo L-J, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. Lancet Oncol 2010;11:835–44.

[4] Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro U, Silva e Sousa AH, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg* 2004; 240: 711–717.

[5] van der Valk MJM, Hilling DE, Bastiaannet E, Meershoek-Klein Kranenbarg E, Beets GL, Figueiredo NL, et al. Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. Lancet 2018;391:2537–45.

[6] Fernandez LM, Julião GPS, Figueiredo NL, Beets GL, van der Valk MJM, Bahadoer RR, et al. Conditional recurrence-free survival of clinical complete responders managed by watch and wait after neoadjuvant chemoradiotherapy for rectal cancer in the International Watch & Wait Database: a retrospective, international, multicentre registry study. Lancet Oncol 2021;22:43–50.

[7] Hupkens BJP, Maas M, Martens MH, van der Sande ME, Lambregts DMJ, Breukink SO, et al. Organ preservation in rectal cancer after chemoradiation: should we extend the observation period in patients with a clinical near-complete response? Ann Surg Oncol 2018;25:197–203.

[8] Renehan AG, Malcomson L, Emsley R, Gollins S, Maw A, Myint AS, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): A propensity-score matched cohort analysis. Lancet Oncol 2016;17:174–83.

[9] Sammour T, Price BA, Krause KJ, Chang GJ. Nonoperative management or "watch and wait" for rectal cancer with complete clinical response after neoadjuvant chemoradiotherapy: a critical appraisal. Ann Surg Oncol 2017;24:1904–15.

[10] Ryan JE, Warrier SK, Lynch AC, Ramsay RG, Phillips WA, Heriot AG. Predicting pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. Color Dis 2016;18:234–46.

[11] Van CE, Verheul HMW, Flamen P, Rougier P, Beets-Tan R, Glynne-Jones R, et al. Imaging in colorectal cancer: progress and challenges for the clinicians. Cancers (Basel) 2016;8(9):81. https://doi.org/10.3390/cancers8090081.

[12] Beets-Tan R, Beets G, Vliegen R, Kessels A, Van BH, De BA, et al. Accuracy of magnetic resonance imaging in prediction of tumour-free resection margin in rectal cancer surgery. Lancet 2001;357:497–504.

[13] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234–48.

[14] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6.

[15] Meng Y, Zhang C, Zou S, Zhao X, Xu K, Zhang H, et al. MRI texture analysis in predicting treatment response to neoadjuvant chemoradiotherapy in rectal cancer. Oncotarget 2018;9:11999–2008.

[16] Cusumano D, Dinapoli N, Boldrini L, Chiloiro G, Gatta R, Masciocchi C, et al. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. Radiol Med 2018;123:286–95.

[17] Cui Y, Yang X, Shi Z, Yang Z, Du X, Zhao Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Eur Radiol 2019;29:1211–20.

[18] Zhou X, Yi Y, Liu Z, Cao W, Lai B, Sun K, et al. Radiomics-based pretherapeutic prediction of non-response to neoadjuvant therapy in locally advanced rectal cancer. Ann Surg Oncol 2019;26:1676–84.

[19] Conroy T, Bosset J-F, Etienne P-L, Rio E, FRANCOIS E, Mesgouez-Nebout N, et al. Neoadjuvant chemotherapy with FOLFIRINOX and preoperative chemoradiotherapy for patients with locally advanced rectal cancer (UNICANCER-PRODIGE 23): a multicentre, randomised, open-label, phase 3 trial trial'. Lancet Oncol. 2021; 22: 702–715.

[20] Bahadoer RR, Dijkstra EA, van Etten B, Marijnen C, Putter H, et al. Short-course radiotherapy followed by chemotherapy before TME in locally advanced rectal cancer: the randomized RAPIDO trial. Lancet Oncol 2021;22:29–42.

[21] Garcia-Aguilar J, Patil S, Kim JK, Yuval JB, Thompson H, Verheij F, et al. Preliminary results of the organ preservation of rectal adenocarcinoma (OPRA) trial. J Clin Oncol 2020;38:4008.

[22] Wang Y, Cai H, Li C, Jiang Z, Wang L, Song J, et al. 'Optimal caliper width for propensity score matching of three treatment groups: a Monte Carlo study. *PLoS One*; 8(12): e81045. Doi: 10.1371/journal.pone.0081045.

[23] Horvat N, Bates DDB, Petkovska I. Novel imaging techniques of rectal cancer: what do radiomics and radiogenomics have to offer? A literature review. *Abdom Radiol* 2019;44(11):3764–74. https://doi.org/10.1007/s00261-019-02042-y.

[24] Scalco E, Belfatto A, Mastropietro A, Rancati T, Avuzzi B, Messina A, et al. T2w-MRI signal normalization affects radiomics features reproducibility. Med Phys 2020;47:1680–91.

[25] Nyul LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 2000;19:143–50.

[26] Zwanenburg A, Vallières M, Abdalah M, Aerts H, Andrearczyk V, Aditya A, et al. Image biomarker standardisation initiative. Radiology 2020;295:328–38.

[27] Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. PLoS ONE 2019;14:e0213459.

[28] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339.

[29] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63.

[30] Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507–17.

[31] Dy J, Brodley C. Feature selection for unsupervised learning. J Machine Learn Res 2004;5:845–89.

[32] Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-based prognosis analysis for non-small cell lung cancer. Sci Rep 2017;7:46349. https://doi.org/10.1038/srep46349.

[33] Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018;2:36.

[34] Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017;356:i6460.

[35] Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny M E. Detection of calibration drift in clinical prediction models to inform model updating. *J Biomed Inform* 2020;112: 103611.

[36] Jenkins DA, Martin GP, Sperrin M, Riley RD, Debray TPA, Collins GS, et al. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? Diagnostic Progn Res 2021;5:1.

[37] Jenkins DA, Sperrin M, Martin GP, Peek N. Dynamic models to predict health outcomes: current status and methodological challenges. Diagnostic Progn Res 2018;2:23.

[38] Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017;357:j2099.

[39] Dinapoli N, Barbaro B, Gatta R, Chiloiro G, Casà C, Masciocchi C, et al. 'Magnetic resonance, vendor-independent, intensity histogram analysis predicting pathologic complete response after radiochemotherapy of rectal cancer. *Int J Radiat Oncol* Biol Phys 2018;102:765–74.

[40] Wang S, Zhou M, Liu Z, Liu Z, Gu D, Zang Y, et al. Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med Image Anal 2017;40:172–83.