European Journal of Radiology 118 (2019) 1–9

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

MR-based artificial intelligence model to assess response to therapy in locally advanced rectal cancer

R. Ferrari[a], C. Mancini-Terracciano[b], C. Voena[b,*], M. Rengo[c], M. Zerunian[c], A. Ciardiello[b,d], S. Grasso[d], V. Mare'[d,e], R. Paramatti[b,d], A. Russomando[f], R. Santacesaria[b], A. Satta[g], E. Solfaroli Camillocci[b,d,h], R. Faccini[b,d], A. Laghi[i]

[a] Az. Osp. San Camillo Forlanini, Department of Emergency Radiology, Viale Gianicolense 87, 00152, Rome, Italy
[b] Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Piazzale A. Moro 2, 00185, Rome, Italy
[c] "Sapienza", University of Rome, Department of Radiological Science, Oncology and Pathology, Polo Pontino, Icot Hospital, via Franco Faggiana 1680, 04100, Latina, Italy
[d] "Sapienza", University of Rome, Department of Physics, Piazzale A. Moro 2, 00185, Rome, Italy
[e] University "Cattolica del Sacro Cuore", Specialty School of Medical Physics, Largo Francesco Vito 1, 00198, Rome, Italy
[f] Centro Científico Tecnológico de Valparaíso-CCTVal, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaiso, Chile
[g] Istituto Nazionale di Fisica Nucleare, Sezione di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133, Rome, Italy
[h] "Sapienza", University of Rome, Specialty School of Medical Physics, Piazzale Aldo Moro 2, 00185, Rome, Italy
[i] "Sapienza", University of Rome, Department of Radiological Science, Oncology and Pathology, Sant'Andrea University hospital, via di Grottarossa 1035, 00189, Rome, Italy

ARTICLE INFO

Keywords:
Rectal cancer
Magnetic resonance imaging
Texture analysis
Artificial intelligence
Neoadjuvant chemoradiotherapy

ABSTRACT

Purpose: To develop and validate an Artificial Intelligence (AI) model based on texture analysis of high-resolution T2 weighted MR images able 1) to predict pathologic Complete Response (CR) and 2) to identify non-responders (NR) among patients with locally-advanced rectal cancer (LARC) after receiving neoadjuvant chemoradiotherapy (CRT).

Method: Fifty-five consecutive patients with LARC were retrospectively enrolled in this study. Patients underwent 3 T Magnetic Resonance Imaging (MRI) acquiring T2-weighted images before, during and after CRT. All patients underwent complete surgical resection and histopathology was the gold standard. Textural features were automatically extracted using an open-source software. A sub-set of statistically significant textural features was selected and two AI models were built by training a Random Forest (RF) classifier on 28 patients (training cohort). Model performances were estimated on 27 patients (validation cohort) using a ROC curve and a decision curve analysis.

Results: Sixteen of 55 patients achieved CR. The AI model for CR classification showed good discrimination power with mean area under the receiver operating curve (AUC) of 0.86 (95% CI: 0.70, 0.94) in the validation cohort. The discriminatory power for the NR classification showed a mean AUC of 0.83 (95% CI: 0.71,0.92). Decision curve analysis confirmed higher net patient benefit when using AI models compared to standard-of-care.

Conclusions: AI models based on textural features of MR images of patients with LARC may help to identify patients who will show CR at the end of treatment and those who will not respond to therapy (NR) at an early stage of the treatment.

1. Introduction

Locally advanced rectal cancer (LARC) is usually treated with neoadjuvant chemoradiotherapy (CRT) followed by radical surgery with total mesorectal excision (TME) [1,2]. This therapeutic strategy results in local pelvic recurrence rate lower than 10% [3]. However, after surgery, 20% to 25% of patients present pathologic complete response (CR) [4]. In those patients, TME can be considered a substantial

Abbreviations: T2w, T2-weighted; AI, artificial Intelligence; CRT, chemo-radiotherapy; RF, random forest; CR, complete responder; PR, partial responder; NR, non-responder; LARC, locally-advanced rectal cancer; TME, total mesorectal excision

* Corresponding author.
E-mail address: cecilia.voena@roma1.infn.it (C. Voena).

https://doi.org/10.1016/j.ejrad.2019.06.013
Received 17 January 2019; Received in revised form 31 May 2019; Accepted 11 June 2019
0720-048X/

overtreatment, since they might benefit from either less invasive surgical approach (i.e trans-anal endoscopic microsurgery) or wait-and-watch strategy [5–7].

The opposite scenario includes patients who do not respond to therapy, i.e., non-responders (NR). An early recognition of NR during treatment would be beneficial since patients might stop treatment, reducing potential adverse effects, and be referred to alternative treatments.

Although MRI is the most accurate imaging modality for primary staging of rectal cancer [8–10], this is not so for assessing response to therapy. In fact, on one side identification of CR on the basis of morphological imaging has been showed to be cumbersome because it is difficult to differentiate fibrosis from residual tumor [11–15]; on the other side, conventional morphological MR images are unable to select *a priori* those patients who will not respond to therapy.

Because of limitations of morphological imaging, quantitative approaches, based on non-morphological parameters, like diffusion-weighted imaging (DWI) and perfusion MRI, were explored with alternate results [11,16]. The development of radiomics, i.e. the extraction of quantitative data from digital images (in particular CT and MRI) [17] has led many researchers to explore if radiomics is able to overcome common limitations of morphological images. Preliminary experiences using texture analysis (TA) showed promising results [13,16,18,19]. A few studies have investigated the ability of MRI-based textural or radiomics analysis to predict response to CRT in LARC [20–22]. However, these studies were designed to identify only CR while none have investigated the accuracy for the detection of NR.

Thus, the aim of this study is to develop artificial intelligence (AI) models based on textural analysis of volumetric T2-weighted (T2w) MR imaging of rectal cancer to assess tumor response to therapy and, in particular, to categorize separately CR and NR, developing two different AI models: one model to discriminate CR from Partial Responders (PR) and NR and another model to discriminate NR from CR and PR.

2. Materials and methods

2.1. Study population

This study was designed as a retrospective data analysis of a prospectively acquired patient population. Good Clinical Practice (GCP)-International Conference on Harmonization (ICH) was applied [23]. All patients signed a written informed consent to be enrolled in the study. The protocol was approved by the Local Ethical Committee (Rif. 2737/28.03.2013).

Patient population was enrolled following specific inclusion and exclusion criteria (Table 1) from February 2014 to November 2016 as illustrated in the flow-chart diagram (Fig. 1). Ninety patients were enrolled in the study. Thirty-five patients were excluded because unable to complete either MRI or CRT protocols or because they refused surgery. Final population consisted of 55 patients (24 females, 31 males) with an average age of 64 ± 9 years (see Table 2).

Table 1
Inclusion and exclusion criteria of the study.

Inclusion Criteria	Exclusion Criteria
Histologically-confirmed rectal adenocarcinoma[a]	Contraindications to MR examinations[b] Incomplete MR acquisitions Lack of histopathological data Contraindications to neoadjuvant therapy[c] Concomitant or recent[d] participation to another clinical trial with experimental drug administration treatment

[a] Stage II and Stage III according to the International Union Against Cancer (IUCC) classification).
[b] MR incompatible implanted devices.
[c] drug hypersensitivity, severe comorbidities that contraindicate the use of the specific chemotherapy chosen, incomplete neoadjuvant therapy, prior pelvic radiation treatment for other tumors.
[d] 30 days before the enrollment.

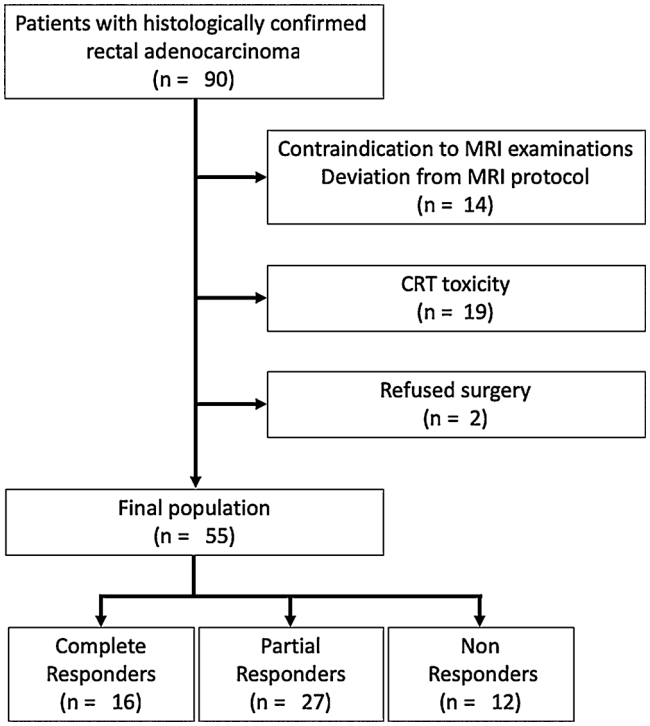

Fig. 1. Flow chart diagram of the enrollment in the study.

Table 2
Population group and pathologic staging at surgery.

Total No. of patients	**55**
Gender	
Female	24
Male	31
Mean age Female, yr ± SD (range)	64 ± 10 (47-81)
Mean age Male, yr ± SD (range)	65 ± 8 (47-83)
Pathological T Staging	
ypT0	11
ypTx	5
ypTis	1
ypT1	6
ypT2	12
ypT3	20
Pathological response to treatment (%)	
CR[a]	16 (29)
PR[b]	27 (49)
NR[c]	12 (22)

[a] CR: pathological complete responder.
[b] PR: pathological partial responder.
[c] NR: pathological non-responder.

All patients underwent optical colonoscopy with biopsy for immunohistochemical analysis and an MRI study for tumor staging. All patients then underwent neoadjuvant CRT protocol and total mesorectal excision (TME) 6–8 weeks after the end of CRT, according to the standard-of-care of our center [24]. One single pathologist, blinded to MR and biopsy, divided the patients in pathological CR, PR and NR by assessing the response to therapy according to the following procedure. All the gross specimens were analyzed in a random order. The rectal segment harboring the neoplasm was examined by sectioning orthogonal to the long axis, obtaining 2–3 mm thick macro section specimens. According to the Dowrak-Rodel technique tumor regression was semi-quantitatively assessed by the amount of viable tissue versus the amount of fibrosis, ranging from no evidence of fibrosis to a complete

response with no residual tumor identifiable.

Patients were followed up for 30 months at intervals of 3 months after surgery with physical examination, routine blood tests and yearly whole body computed tomography to assess local recurrences or distant metastases.

2.2. Imaging acquisition protocol

Patients underwent three MRI acquisitions, the first performed for tumor staging before CRT, the second after 21 days from the beginning of CRT and the third from 6 to 8 weeks after the end of it.

MRI examinations were performed using a 3 T scanner (Discovery MR750, General Electrics, Milwaukee, Wisconsin, USA), following a standard imaging protocol. The latter includes high-resolution T2w fast recovery fast-spin echo (2D FRFSE) sequence (TR, 2086-4172 ms; TE, 11.4–122.3 ms; Nex, 2; slice thickness, 4 mm; slice gap, 0.5 mm; matrix, 512 × 512) acquired on three conventional planes plus dedicated axial oblique and coronal oblique planes obtained orthogonal and parallel to the long axis of the rectal cancer, respectively. No intravenous contrast medium was administered.

2.3. Image analysis

2.3.1. Segmentation

An abdominal radiologist (MR, 10 years of experience), blind to the histopathological results, manually drew a ROI around each tumor area, avoiding rectal lumen, on the axial oblique T2w images, slice by slice through the whole longitudinal extension of the tumor using an open-source software (3D-slicer, Brigham and Women's Hospital, Boston, MA, USA). Additional ROIs (mean size: 10 mm^2) were placed on the right internal obturator muscle for signal intensity normalization.

2.3.2. Image pre-processing

Images were analyzed using a custom open-source software (Fig. 2) in python language (Dicom-tools, https://pypi.org/project/dicom-tools/), where segmentation, filtering, intensity normalization and texture analysis tools are implemented. We tested two gray-level intensity normalization schemes (a histogram matching algorithm [25] and a normalization to the muscle ROI mean grey-level intensity).

In order to study the texture of the tissue at different spatial scales, a Laplacian of Gaussian band pass filter at various scales (1 mm–4 mm) was applied.

We also tried to remove from the ROI the lower intensity values that are not expression of tumor in T2w MRI. We applied an intensity cut (the same on all the ROIs of all the patients) before computing the textural features and we repeated the analysis. We tested 3 thresholds (10-15-25% of the maximum intensity). Since we did not find any improvement in the performances of the classifiers after applying the grey-level intensity normalization, Laplacian filter or intensity cut, we did not use these algorithms in the results of this study.

2.3.3. Textural features

Textural features were extracted automatically using DICOM-tools for each image in the ROI slice with the maximum tumor area and in the whole ROI volume. The list of the features is in Table 3, and the feature extraction algorithms are presented in the Appendix A in Supplementary material.

Each feature was calculated at three different times: before CRT (pre-CRT), during CRT (mid-CRT) and after CRT (post-CRT) to include time evolution of the tumor in the analysis. We derived additional features to be used in the classifiers as the ratios and the differences of each textural feature at different times. Thus for each feature we have three ratios (post-CRT/pre-CRT, mid-CRT/pre-CRT, post-CRT/mid-CRT) and the three corresponding differences. In this way, instrumental or biological effects that affect the distributions of the original feature

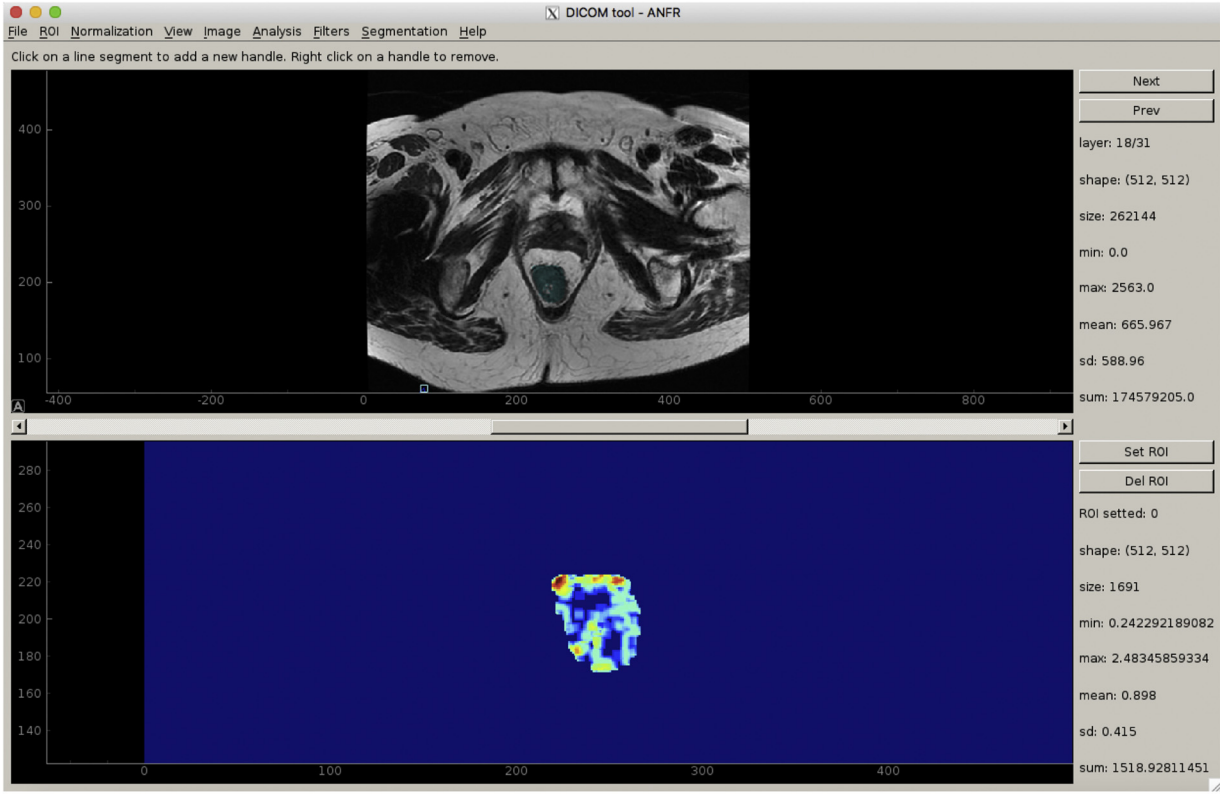

Fig. 2. Software user interface (https://pypi.org/project/dicom-tools/). In the upper panel a slice of an axial oblique T2w image is shown with the ROI highlighted in cyan. The lower panel shows the local entropy map of the ROI (radius = 5pixels).

Table 3
Textural features used in the analysis. First order features are the statistical parameters of the grey-level intensity histogram. See Appendix A in Supplementary material for details on the algorithms.

First order	Second order
Mean	Haralick Homogeneity
Standard Deviation	Haralick Correlation
Skewness	Haralick Contrast
Kurtosis	Haralick Energy
	Haralick Dissimilarity
	Minimum of the Shannon Entropy
	Maximum of the Shannon Entropy
	Mean of the Shannon Entropy
	Standard Deviations of the Shannon Entropy

may cancel out, possibly increasing the sensitivity to the tumor-non tumor separation. We performed both two-dimensional (in the slice where tumor has the maximum area) and three-dimensional analysis (all tumor volume), and we used 10 radii for local entropy computation. We also included in the list of features two morphological parameters, the volume in the ROI and the area in the ROI slice with maximum tumor area. A total of 855 features (taking into account the three times, ratios ad differences) were considered.

2.4. Statistical analysis, radiomics signature and AI models

Descriptive statistics are provided as averages, together with their standard deviation, of the features over the different patients' classes. To estimate the potential separation power of each feature, a statistical analysis with a 2-tail t-Student test was performed and the corresponding p-value computed.

We then combined a sub-set of the features (radiomics signature) in an AI approach to provide a model for the stratification of the patient response. In the radiomics signature we included the features with the p value of the above-mentioned t-Student test smaller than 0.01. In order to eliminate redundancy, we considered only the entropy related variables at the most significant pixel radius, since entropies at different radii are highly correlated. We used a random forest (RF) [26], a classifier of the family of the decision trees, which is considered robust in dealing with small data samples. The RF algorithm first calculates a probability-like continuous output and categories are determined by its value being above a given threshold or not. For each value of such threshold a pair of sensitivity and specificity values are obtained, thus allowing to draw the receiver operating characteristic (ROC) curve. The quality of the classifier can be expressed by the ROC area under the curve (AUC), which is related to the probability that the algorithm will rank a randomly chosen positive instance higher than a randomly chosen negative instance (assuming "positive" rank higher than "negative"). The patients were divided into a training cohort (28 patients), used to teach the RF to distinguish between the various classes and a validation cohort (27 patients), used to evaluate the classification performances. In order to evaluate the dependence of the classifier on the division of the patients into a training and a validation cohort we considered 50 different divisions of equal size. We quote the mean AUC with the 95% confidence interval from this distribution.

Two different AI models were built, one to discriminate CR from PR + NR and one to discriminate NR from CR + PR. We built the RF with 2000 trees and a bagged sampling fraction of 0.6. In the NR vs. CR + PR case, post-CRT features were not considered, since our purpose is to identify NR earlier before the end of CRT.

Finally, a decision curve analysis [27] was performed to calculate the net benefit from the use of our AI models using the output of the RF to compute the probability of either being a CR or a NR for each of the 50 validation cohorts. A comparison with a model using only the tumor size (morphological information) for the CR classifier and a model using only the standard deviation of the entropy in 2D for the NR classifier, was performed. The decision curves were calculated separately for the overall benefit, the benefit for the treated, and for the untreated.

3. Results

3.1. Patient cohort

A population of 55 patients was included for the analysis. At histology after surgery, 16 patients (29%) achieved pathologic complete response (CR), 27 (49%) pathologic partial response (PR) and 12 (22%) were pathologic non-responders (NR). No differences in terms of gender, age or tumor characteristics were observed between CR and PR or NR. Patients characteristics are summarized in Table 2.

3.2. Texture analysis

We studied each feature individually, to understand if a separation of the patients in two classes (CR vs PR + NR and NR vs CR + PR), according to the known response to CRT, was possible. A feature was considered as discriminant if the p-value was below 0.01. Descriptive statistics of selected features, together with the p-value of a 2-tail t-Student test, are reported in Table 4 choosing only one entropy pixel radius (R = 5 for CR and R = 3 for NR), the most significant one. This choice allows to reduce redundancy since entropies at different radii are highly correlated.

First, we considered the possibility to separate CR from PR + NR patients. As reported in Table 4, CR patients are characterized by higher pre-CRT energy, mid-CRT homogeneity, lower mid-CRT contrast and dissimilarity. The behavior of these Haralick parameters can be interpreted as a lower tissue heterogeneity for CR patients with respect to PR + NR. Moreover, there is a strong indication of a characteristic time evolution of the minimum of the local three-dimensional entropy for CRT patients. As an example, Fig. 3 shows the distribution of the ratio of the 3D local entropy minimum post-CRT/pre-CRT for the different patients' classes, showing that this feature tends to become lower after CRT in CR patients and higher in the other two categories of patients. Fig. 3 shows also that, although the average values are significantly different for the two patient classes, the distributions are too spread out around their mean and the two classes overlap, so that a separation patient by patient, using this single variable, would not be efficient. It has also been found that pre-CRT two-dimensional tumor area, although not a textural feature, is significantly lower in CR ($p < 0.001$).

Next, we searched for features able to discriminate NR versus CR + PR. The most significant features in this case turned out to be the standard deviation of the gray level intensity histogram in the two-dimensional analysis and the minimum, the maximum and the standard deviation of the local entropy pre-CRT, which were lower for NR than for CR + PR.

We post hoc performed a power calculation using the distributions of the most significant textural variables of Table 4 obtained on our 55 patients population. We find that the sample size required to separate, on average, the different classes of patients, choosing $\alpha = 0.05$ one tailed and $\beta = 0.10$ (power 90%) is always smaller than our sample size. Thus the study is not underpowered.

3.3. Decision model

Two different models were separately trained through RF classifiers, one to discriminate CR from PR + NR and one to discriminate NR from CR + PR, in the training cohort. The list of the features used in the CR and NR classifiers is in Table 4; we use 8 features for the CR RF and 4 features for the NR RF. ROC curves are shown in Figs. 4 and 5 for the 50 different divisions of the patients into training and validation cohorts. We obtained a mean ROC AUC of 0.86 (95% CI: 0.70,0.94) and 0.83 (95% CI: 0.71,0.92) for the CR vs PR + NR and for the NR vs CR + PR

Table 4
Average values and standard deviations of the most discriminant features or feature changes over time (p value < 0.01) computed separately for CR[a], PR[b] + NR[c] (CR vs PR + NR discrimination) patients and for NR, CR + PR patients (NR vs CR + PR discrimination). The values are also reported separately for the training and the validation set. Only the most discriminating scale of the entropy is considered. The p-value from the two-tails t-Student test is also shown.

CR[a] vs PR[b] + NR[c] discrimination (training + validation set)

Parameter	CR	PR + NR	p-value
Min Entropy[d]:post-CRT[e]/pre-CRT[f] (3D)	0.97 ± 0.13	2.49 ± 0.40	0.0008
Min Entropy: post-CRT-pre-CRT (3D)	−0.038 ± 0.030	0.14 ± 0.05	0.003
Min Entropy: post-CRT/mid-CRT[g] (3D)	1.07 ± 0.13	2.41 ± 0.42	0.004
ROI Area Pre-CRT (2D)	2537 ± 202	4016 ± 366	0.0009
Dissimilarity Mid-CRT (2D)	0.00297 ± 0.00025	0.00505 ± 0.00060	0.002
Contrast Mid-CRT (2D)	0.0050 ± 0.0006	0.0098 ± 0.0013	0.001
Homogeneity Mid-CRT (2D)	0.99870 ± 0.00010	0.99789 ± 0.00024	0.003
Energy Pre-CRT (2D)	0.99019 ± 0.00078	0.9859 ± 0.0012	0.004

CR vs PR + NR discrimination (training set)

Parameter	CR	PR + NR	p-value
Min Entropy:post-CRT/pre-CRT (3D)	1.18 ± 0.20	2.55 ± 0.67	0.06
Min Entropy:post-CRT-pre-CRT (3D)	0.011 ± 0.033	0.16 ± 0.07	0.06
Min Entropy: post-CRT/mid-CRT (3D)	1.10 ± 0.22	2.57 ± 0.69	0.05
ROI Area pre-CRT (2D)	2945 ± 146	3837 ± 370	0.03
Dissimilarity mid-CRT (2D)	0.00322 ± 0.00038	0.00438 ± 0.00059	0.11
Contrast mid-CRT (2D)	0.00563 ± 0.00088	0.00823 ± 0.0015	0.14
Homogeneity mid-CRT (2D)	0.99861 ± 0.00015	0.99815 ± 0.00023	0.11
Energy Pre-CRT (2D)	0.98859 ± 0.00058	0.9850 ± 0.0014	0.03

CR vs PR + NR discrimination (validation set)

Parameter	CR	PR + NR	p-value
Min Entropy:post-CRT/pre-CRT (3D)	0.76 ± 0.12	2.43 ± 0.44	0.0016
Min Entropy:post-CRT-pre-CRT (3D)	−0.088 ± 0.048	0.12 ± 0.07	0.02
Min Entropy: post-CRT/mid-CRT (3D)	1.036 ± 0.16	2.25 ± 0.50	0.03
ROI Area pre-CRT (2D)	2128 ± 324	4203 ± 653	0.009
Dissimilarity mid-CRT (2D)	0.0027 ± 0.0003	0.0058 ± 0.0011	0.012
Contrast mid-CRT (2D)	0.0043 ± 0.0008	0.0113 ± 0.0021	0.005
Homogeneity mid-CRT (2D)	0.99880 ± 0.00013	0.99763 ± 0.00043	0.017
Energy pre-CRT (2D)	0.99179 ± 0.0012	0.98679 ± 0.0021	0.05

NR vs CR + PR discrimination (training + validation set)

Parameter	NR	CR + PR	p-value
Standard Deviation pre-CRT (2D)	123 ± 11	172 ± 8	0.002
Mean Entropy[h] pre-CRT (2D)	0.618 ± 0.061	0.814 ± 0.028	0.009
Standard Deviation Entropy pre-CRT (2D)	0.4180 ± 0.0082	0.469 ± 0.011	0.0004
Max Entropy pre-CRT (2D)	1.870 ± 0.078	2.248 ± 0.056	0.0005

NR vs CR + PR discrimination (training set)

Parameter	NR	CR + PR	p-value
Standard Deviation pre-CRT (2D)	137 ± 20	175 ± 13	0.005
Mean Entropy pre-CRT (2D)	0.68 ± 0.10	0.800 ± 0.051	0.3
Standard Deviation Entropy pre-CRT (2D)	0.424 ± 0.012	0.464 ± 0.013	0.04
Max Entropy pre-CRT (2D)	2.00 ± 0.12	2.26 ± 0.10	0.13

NR vs CR + PR discrimination (validation set)

Parameter	NR	CR + PR	p-value
Standard Deviation pre-CRT (2D)	109 ± 9	169 ± 10	0.0003
Mean Entropy pre-CRT (2D)	0.558 ± 0.064	0.828 ± 0.029	0.0022
Standard Deviation Entropy pre-CRT (2D)	0.412 ± 0.011	0.475 ± 0.016	0.004
Max Entropy pre-CRT (2D)	1.727 ± 0.068	2.241 ± 0.056	0.0002

[a] Complete Responder patients.
[b] Partial Responder patients.
[c] Non Responder patients.
[d] Entropy is computed in a spherical volume with radius = 5 pixels.
[e] After chemo-radiotherapy.
[f] Before chemo-radiotherapy.
[g] During chemo-radiotherapy (21st day).
[h] Entropy is computed in a circle volume with radius = 3 pixels.

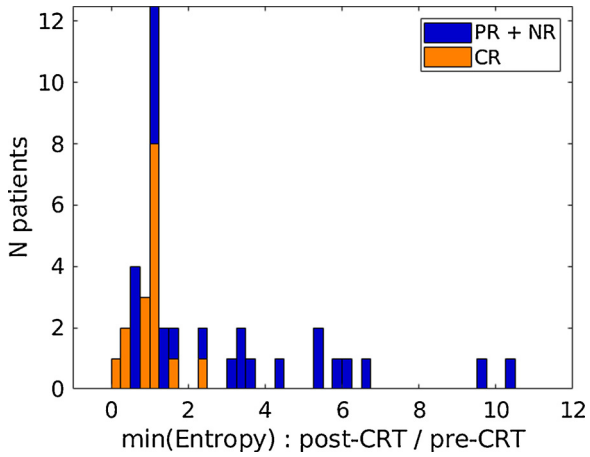

Fig. 3. Distribution over the patient population of one of the features used: the ratio of the minimum value (over the 3D ROI) of the entropy post-CRT and pre-CRT. The entropy is computed for each pixel in a spherical volume with radius = 5 pixels around it. CR patients (orange histogram) and PR + NR patients (blue histogram) are shown.

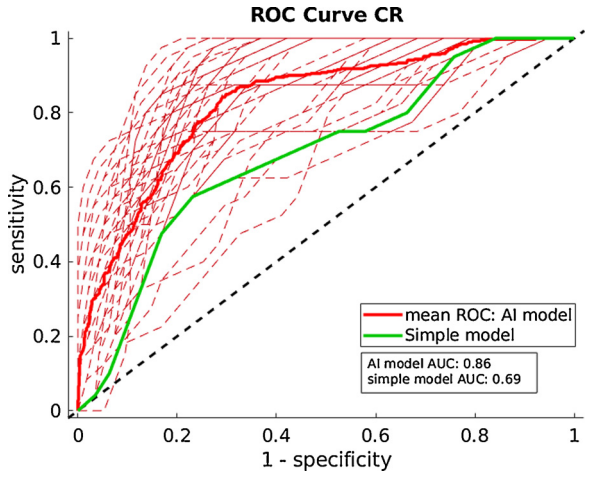

Fig. 4. ROC curves (sensitivity vs 1-specificity) for the RF classifier (dashed red lines) developed to discriminate CR vs PR + NR patients over the 50 different validation cohorts. The mean of the ROC curves is shown is the red curve. The mean AUC is 0.86. The ROC for a simple model using only the dimension of the tumor at the pre-CRT stage as discriminating variable (morphological parameter, green curve) is also shown.

case, respectively. Table 5 reports the obtained performances for the two classifiers.

Figs. 4 and 5 also show the ROC curves of two simple decision models, built using a single textural feature, the size of the tumor at the pre-CRT stage (morphological parameter) for the CR classifier and the standard deviation of the 2D local entropy (radius = 3 pixels) at the pre-CRT stage for the NR responder classifier. In both cases the AI models outperform the simple models in terms of AUC.

The mean decision curves over the 50 validation cohorts of the overall net benefits for the AI models are reported in Figs. 6 and 7. Fig. 6 also shows the decision curve for the simple morphological model. It can be concluded that the AI model built for CR discrimination outperforms the other models for threshold probabilities lower than 70%. In the case of the model for NR discrimination it gives better results than the case where no prediction model is assumed for threshold probabilities in the range 5%–50%. More details on the decision curve analysis are reported in Figs. 8 and 9 where the benefit for the treated and the untreated CR and NR, respectively, are separately shown.

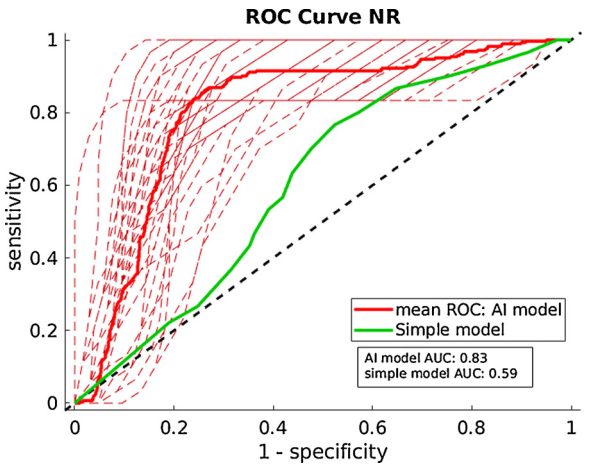

Fig. 5. ROC curves (sensitivity vs 1-specificity) for the RF classifier (dashed red lines) developed to discriminate NR vs CR + NR patients over the 50 different validation cohorts. The mean of the ROC curves is shown is the red curve. The mean AUC is 0.83. The ROC for a simple model using only the standard deviation of the local entropy in 2D (computed in a circle with radius = 3 pixels) at the pre-CRT stage as discriminating variable (green curve) is also shown.

Table 5
Performances of the CR and NR classifiers. The accuracy, the sensitivity and the specificity are computed at the maximum value of the Youden index for the mean ROC curves of Figs. 4 and 5.

	CR classifier	NR classifier
AUC	0.86	0.83
Accuracy	0.79	0.83
Sensitivity	0.85	0.86
Specificity	0.70	0.74

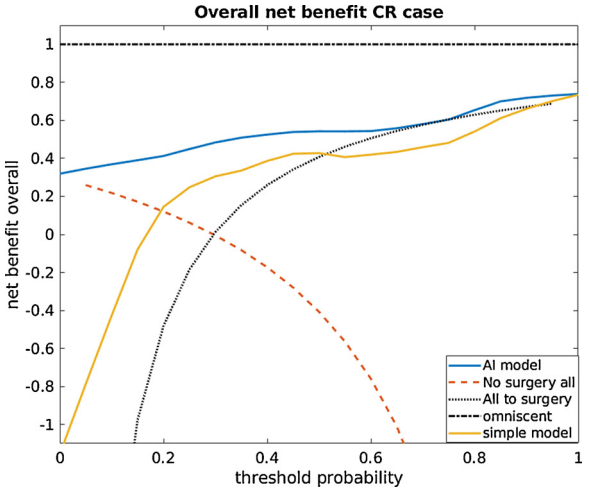

Fig. 6. Decision curve analysis for the AI model built for CR identification on the validation patient cohort; data shown are the mean over the 50 different compositions of the validation cohorts. The dotted line represents the overall net benefit in the case where all patients are sent to surgery. The dashed line represents the case where no patient is sent to surgery. The dashed-dotted line shows the benefit obtainable from an omniscient model. The blue line corresponds to the mean curve for the AI model over the 50 different validation cohorts and the yellow line to a model built using only the dimension of the tumor at the pre-CRT stage as discriminating variable (morphological parameter). It can be seen that the AI model outperforms the other cases for threshold probabilities lower than 70%.

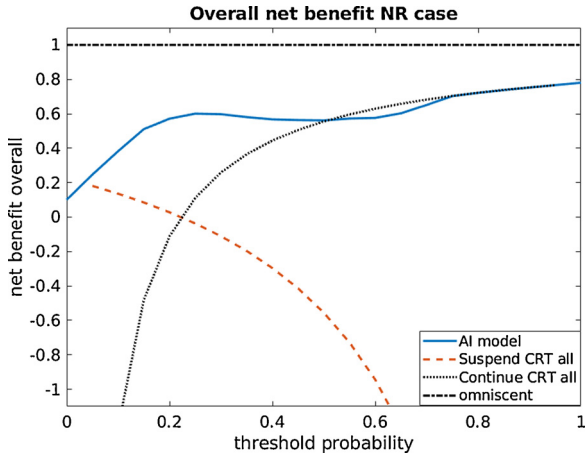

Fig. 7. Decision curve analysis for the AI model built for NR identification on the validation patient cohort; data shown are the mean over the 50 different compositions of the validation cohorts. The dotted line represents the overall net benefit in the case where all patients continue CRT. The dashed line represents the case where all patients suspend CRT. The dashed-dotted line shows the benefit obtainable from an omniscient model. The blue line corresponds to the mean curve for the AI model over the 50 different validation cohorts. It can be seen that the AI model outperforms the case where no prediction model is assumed for threshold probabilities in the range 5–50%.

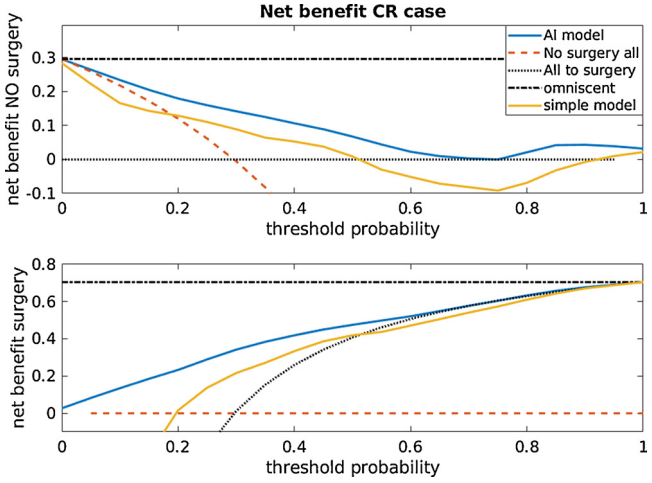

Fig. 8. Decision curve analysis for the AI model built for CR identification on the validation patient cohort; data shown are the mean over the 50 different compositions of the validation cohorts. The upper figure shows the net benefit for patients that are recognized as CR and thus benefit from not receiving surgery while the lower figure shows the net benefit for patients that are classified as PR or NR and thus should receive surgery. The dotted line represents the net benefit in the case where all patients are sent to surgery. The dashed line represents the case where no patient is sent to surgery. The dashed-dotted line shows the benefit obtainable from an omniscient model. The blue line corresponds to the mean curve for the AI model over the 50 different validation cohorts and the yellow line to a model built using only the dimension of the tumor at the pre-CRT stage as discriminating variable (morphological parameter).

4. Discussion

Our results show that an AI-based analysis of textural features extracted from high-resolution T2w MR images, has a good predictivity of the response to CRT in patients affected by LARC and therefore it can be used in the choice of the best therapeutic strategy. Our work simultaneously considers the three-dimensional tumor volume data for the textural analysis, the study of the time evolution of textural features (before, during, after CRT) and the use of AI models. Another novel

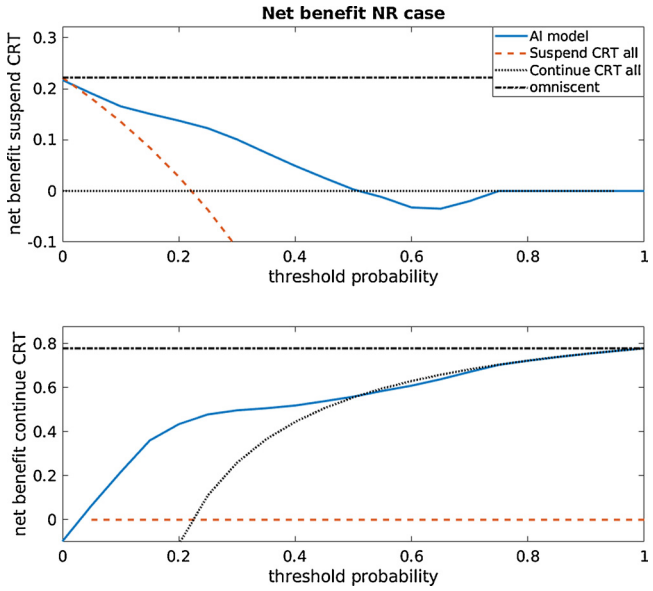

Fig. 9. Decision curve analysis for the AI model built for NR identification on the validation patient cohort; data shown are the mean over the 50 different compositions of the validation cohorts. The upper figure shows the net benefit for patients that are recognized as NR and thus benefit from suspending CRT while the lower figure shows the net benefit for patients that are classified as CR or PR and thus should continue CRT. The dotted line represents the net benefit in the case where all patients continue CRT. The dashed line represents the case where all patients suspend CRT. The dashed-dotted line shows the benefit obtainable from an omniscient model. The blue line corresponds to the mean curve for the AI model over the 50 different validation cohorts.

aspect of our work is the attempt to discriminate NR patients using only the information at pre- and mid-CRT stage. Indeed, we built two models, one to select patients that show a complete pathological response and another one to select patients classified as non-responders.

Statistical analysis of textural features indicates that both CR and NR tend to have more homogeneous tissues than PR in relation to the most significant features, mainly concerning the local entropy and the Haralick parameters, which are related to tissue dishomogeneity. Also, we found that a small tumor volume at pre-CRT stage highly correlates with a complete response to therapy. In the case of NR vs CR + PR separation the most significant textural features are the standard deviation of the gray level intensity histogram in the 2D analysis and minimum, maximum and standard deviation of local entropy, all of them computed pre-CRT.

The CR and NR groups are both characterized by features related to tissue homogeneity. However, features significant for classification purposes (see Table 4) are not the same in the two cases: this suggests they are different kind of homogeneous tissues.

The average of the above parameters is significantly different in the various classes, but patient-by-patient variability is quite large and does not allow discrimination on a single patient basis. This variability could be due to biological differences among patients or to instrumental effects; this problem is partially overcome by combining the most relevant features in an automatic AI approach. Automatic classification techniques can simultaneously analyze a large number of features exploiting possible correlations, thus performing significantly better than visual analysis by a clinician. Differently from previous studies [21,25] we built AI models using a RF classifier, an algorithm from the class of the decision trees, particularly robust against overtraining in small datasets [26,28–31].

We obtained a good discrimination power with a mean AUC of 0.86 (95% CI: 0.70,0.94) for the discrimination of CR against PR + NR. A comparable result was obtained by another study using a similar approach but based on radiomics features extracted from multiparametric

imaging (anatomical, diffusion and perfusion imaging) [25]. Another study [20], based on the same multiparametric model, provided a similar AUC of 0.84 for the identification of CR. In our opinion, the use of multiparametric models might introduce further uncontrolled variables in the acquisition parameters, possibly jeopardizing the entire model. Thus, in this preliminary phase, we preferred to concentrate our analysis on the most robust and reproducible images available for the radiologists, i.e. high-resolution T2 weighted images.

We also developed a dedicated algorithm to discriminate NR from the others and a mean AUC of 0.83 (95% CI: 0.71,0.92) is found. It should be considered that in this case we did not used post-CRT features since it is needed to identify NR patients before the end of CRT. It should be noted that in the case of the NR classifier we have only 6 patients in the training set. Such a small sample can hardly represent the variety of the biological anatomical variations thus our study should be considered a feasibilty study and an indication that this method may have a clinical application after the classifier is trained with a larger population.

A decision curve analysis confirms the potential usefulness of our AI models in clinical practice. All these figures of merit must be improved before a clinical use of the algorithm can be considered, but they are promising since these algorithms could learn and could work completely automatically giving simple and clear information to the radiologist once a ROI on the tumor is designed. The perspective of creating an automatic tool with quantitative data could be the easiest approach to perform this kind of analysis.

The main limit of this work is the size of the patient cohorts, both in training and validations sets. However we minimized this limit by using the RF classifier that can be considered the best method to classify small data samples. Thus, we are convinced that our data are quite robust.

Another limit of this study is the manual tumor delineation. This method has been used in most of the studies previously published. However, the delineation of tumors after CRT, especially for the CR, can be difficult since little tissue is usually visible. This limit may be overcome by the use of an automatic segmentation algorithm; such algorithms, to our knowledge, are not validated yet thus the human experience may be considered still superior. Finally, the design of this study should be suitable for a wait & watch prospective strategy. Thus its application in such setting, after the validation on a bigger population, may be interesting.

The real problem with MR data, as demonstrated by the high variability of the results published in previous studies [20–25], is the high variability of data acquisition, intrinsically related to the complexity of the technique and the high number of parameters that can be modified. This variability could affect reproducibility of the results: a multicenter trial using standardized acquisition parameters is highly advisable.

5. Conclusion

In conclusion, the huge amount of radiomics features extracted from MR images were combined by an AI algorithm which is intended to give the radiologists and clinicians solid answers to individuate rapidly NR or CR. The correct stratification of complete responders will help the affirmation of less invasive therapeutic strategies such as mucosectomy or “wait and watch”, while the identification of NR during the therapy will allow to address these patients promptly to more effective therapies.

Funding

This work was supported byAssociazione Italiana per la Ricerca sul Cancro (AIRC) [Investigator Grant 2013/14129]; Dr. Andrea Russomando was partially financed by CONICYT [Fondecyt postdoctorado n. 3180167].

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2019.06.013.

References

[1] R. Sauer, H. Becker, W. Hohenberger, C. Rodel, C. Wittekind, R. Fietkau, P. Martus, J. Tschmelitsch, E. Hager, C.F. Hess, J.H. Karstens, T. Liersch, H. Schmidberger, R. Raab, Preoperative versus postoperative chemoradiotherapy for rectal cancer, N. Engl. J. Med. 351 (17) (2004) 1731–1740, https://doi.org/10.1056/NEJMoa040694.

[2] J.E. Krook, C.G. Moertel, L.L. Gunderson, H.S. Wieand, R.T. Collins, R.W. Beart, T.P. Kubista, M.A. Poon, W.C. Meyers, J.A. Mailliard, et al., Effective surgical adjuvant therapy for high-risk rectal carcinoma, N. Engl. J. Med. 324 (11) (1991) 709–715, https://doi.org/10.1056/nejm199103143241101.

[3] R. Sauer, T. Liersch, S. Merkel, R. Fietkau, W. Hohenberger, C. Hess, H. Becker, H.R. Raab, M.T. Villanueva, H. Witzigmann, C. Wittekind, T. Beissbarth, C. Rodel, Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years, J. Clin. Oncol. 30 (16) (2012) 1926–1933, https://doi.org/10.1200/jco.2011.40.1836.

[4] A.B. Benson 3rd, A.P. Venook, M.M. Al-Hawary, L. Cederquist, Y.J. Chen, K.K. Ciombor, S. Cohen, H.S. Cooper, D. Deming, P.F. Engstrom, J.L. Grem, A. Grothey, H.S. Hochster, S. Hoffe, S. Hunt, A. Kamel, N. Kirilcuk, S. Krishnamurthi, W.A. Messersmith, J. Meyerhardt, M.F. Mulcahy, J.D. Murphy, S. Nurkin, L. Saltz, S. Sharma, D. Shibata, J.M. Skibber, C.T. Sofocleous, E.M. Stoffel, E. Stotsky-Himelfarb, C.G. Willett, E. Wuthrick, K.M. Gregory, L. Gurski, D.A. Freedman-Cass, Rectal cancer, version 2.2018, NCCN clinical practice guidelines in oncology, J. Compr. Canc. Netw. 16 (7) (2018) 874–901, https://doi.org/10.6004/jnccn.2018.0061.

[5] M. Maas, R.G. Beets-Tan, D.M. Lambregts, G. Lammering, P.J. Nelemans, S.M. Engelen, R.M. van Dam, R.L. Jansen, M. Sosef, J.W. Leijtens, K.W. Hulsewe, J. Buijsen, G.L. Beets, Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer, J. Clin. Oncol. 29 (35) (2011) 4633–4640, https://doi.org/10.1200/jco.2011.37.7176.

[6] A. Habr-Gama, G.P. Sao Juliao, R.O. Perez, Nonoperative management of rectal cancer: identifying the ideal patients, Hematol. Oncol. Clin. North Am. 29 (1) (2015) 135–151, https://doi.org/10.1016/j.hoc.2014.09.004.

[7] B.D. O'Neill, G. Brown, R.J. Heald, D. Cunningham, D.M. Tait, Non-operative treatment after neoadjuvant chemoradiotherapy for rectal cancer, Lancet Oncol. 8 (7) (2007) 625–633, https://doi.org/10.1016/s1470-2045(07)70202-4.

[8] V. Valentini, C. Aristei, B. Glimelius, B.D. Minsky, R. Beets-Tan, J.M. Borras, K. Haustermans, P. Maingon, J. Overgaard, L. Pahlman, P. Quirke, H.J. Schmoll, D. Sebag-Montefiore, I. Taylor, E. Van Cutsem, C. Van de Velde, N. Cellini, P. Latini, Multidisciplinary rectal cancer management: 2nd european rectal cancer consensus conference (EURECA-CC2), Radiother. Oncol. 92 (2) (2009) 148–163, https://doi.org/10.1016/j.radonc.2009.06.027.

[9] R. Glynne-Jones, L. Wyrwicz, E. Tiret, G. Brown, C. Rodel, A. Cervantes, D. Arnold, Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up, Ann. Oncol. 28 (Suppl_4) (2017) iv22–iv40, https://doi.org/10.1093/annonc/mdx224.

[10] R.G.H. Beets-Tan, D.M.J. Lambregts, M. Maas, S. Bipat, B. Barbaro, L. Curvo-Semedo, H.M. Fenlon, M.J. Gollub, S. Gourtsoyianni, S. Halligan, C. Hoeffel, S.H. Kim, A. Laghi, A. Maier, S.R. Rafaelsen, J. Stoker, S.A. Taylor, M.R. Torkzad, L. Blomqvist, 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 (4) (2018) 1465–1475, https://doi.org/10.1007/s00330-017-5026-2.

[11] D.M. Lambregts, V. Vandecaveye, B. Barbaro, F.C. Bakers, M. Lambrecht, M. Maas, K. Haustermans, V. Valentini, G.L. Beets, R.G. Beets-Tan, Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study, Ann. Surg. Oncol. 18 (8) (2011) 2224–2231, https://doi.org/10.1245/s10434-011-1607-5.

[12] U.B. Patel, F. Taylor, L. Blomqvist, C. George, H. Evans, P. Tekkis, P. Quirke, D. Sebag-Montefiore, B. Moran, R. Heald, A. Guthrie, N. Bees, I. Swift, K. Pennert, G. Brown, Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience, J. Clin. Oncol. 29 (28) (2011) 3753–3760, https://doi.org/10.1200/jco.2011.34.9068.

[13] M.J. Lahaye, S.M. Engelen, P.J. Nelemans, G.L. Beets, C.J. van de Velde, J.M. van Engelshoven, R.G. Beets-Tan, Imaging for predicting the risk factors–the circumferential resection margin and nodal disease–of local recurrence in rectal cancer: a meta-analysis, Semin. Ultrasound CT MR 26 (4) (2005) 259–268.

[14] M. Rengo, S. Picchia, S. Marzi, D. Bellini, D. Caruso, M. Caterino, M. Ciolina, D. De Santis, D. Musio, V. Tombolini, A. Laghi, Magnetic resonance tumor regression grade (MR-TRG) to assess pathological complete response following neoadjuvant radiochemotherapy in locally advanced rectal cancer, Oncotarget 8 (70) (2017) 114746–114755, https://doi.org/10.18632/oncotarget.21778.

[15] L. Curvo-Semedo, D.M. Lambregts, M. Maas, G.L. Beets, F. Caseiro-Alves, R.G. Beets-Tan, Diffusion-weighted MRI in rectal cancer: apparent diffusion coefficient as a potential noninvasive marker of tumor aggressiveness, J. Magn. Reson. Imaging 35 (6) (2012) 1365–1371, https://doi.org/10.1002/jmri.23589.

[16] M.J. Gollub, D.H. Gultekin, O. Akin, R.K. Do, J.L. Fuqua 3rd, M. Gonen, D. Kuk, M. Weiser, L. Saltz, D. Schrag, K. Goodman, P. Paty, J. Guillem, G.M. Nash,

L. Temple, J. Shia, L.H. Schwartz, Dynamic contrast enhanced-MRI for the detection of pathological complete response to neoadjuvant chemotherapy for locally advanced rectal cancer, Eur. Radiol. 22 (4) (2012) 821–831, https://doi.org/10.1007/s00330-011-2321-1.

[17] M. Avanzo, J. Stancanello, I. El Naqa, Beyond imaging: the promise of radiomics, Phys. Med. 38 (2017) 122–139, https://doi.org/10.1016/j.ejmp.2017.05.071.

[18] F. De Felice, A.L. Magnante, D. Musio, M. Ciolina, C.N. De Cecco, M. Rengo, A. Laghi, V. Tombolini, Diffusion-weighted magnetic resonance imaging in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy, Eur. J. Surg. Oncol. 43 (7) (2017) 1324–1329, https://doi.org/10.1016/j.ejso.2017.03.010.

[19] D. Caruso, M. Zerunian, M. Ciolina, D. de Santis, M. Rengo, M.H. Soomro, G. Giunta, S. Conforto, M. Schmid, E. Neri, A. Laghi, Haralick's texture features for the prediction of response to therapy in colorectal cancer: a preliminary study, Radiol. Med. 123 (3) (2018) 161–167, https://doi.org/10.1007/s11547-017-0833-8.

[20] K. Nie, L. Shi, Q. Chen, X. Hu, S.K. Jabbour, N. Yue, T. Niu, X. Sun, Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI, Clin. Cancer Res. 22 (21) (2016) 5256–5264, https://doi.org/10.1158/1078-0432.ccr-15-2997.

[21] Z. Liu, X.Y. Zhang, Y.J. Shi, L. Wang, H.T. Zhu, Z. Tang, S. Wang, X.T. Li, J. Tian, Y.S. Sun, Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer, Clin. Cancer Res. 23 (23) (2017) 7253–7262, https://doi.org/10.1158/1078-0432.ccr-17-1038.

[22] N. Dinapoli, B. Barbaro, R. Gatta, G. Chiloiro, C. Casa, C. Masciocchi, A. Damiani, L. Boldrini, M.A. Gambacorta, M. Dezio, G.C. Mattiucci, M. Balducci, J. van Soest, A. Dekker, P. Lambin, C. Fiorino, C. Sini, F. De Cobelli, N. Di Muzio, C. Gumina, P. Passoni, R. Manfredi, V. Valentini, Magnetic Resonance, Vendor-independent, intensity histogram analysis predicting pathologic complete response after radiochemotherapy of rectal cancer, Int. J. Radiat. Oncol. Biol. Phys. 102 (4) (2018) 765–774, https://doi.org/10.1016/j.ijrobp.2018.04.065.

[23] N. Horvat, H. Veeraraghavan, M. Khan, I. Blazic, J. Zheng, M. Capanu, E. Sala, J. Garcia-Aguilar, M.J. Gollub, I. Petkovska, MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy, Radiology 287 (3) (2018) 833–843, https://doi.org/10.1148/radiol.2018172300.

[24] Y. Meng, C. Zhang, S. Zou, X. Zhao, K. Xu, H. Zhang, C. Zhou, MRI texture analysis in predicting treatment response to neoadjuvant chemoradiotherapy in rectal cancer, Oncotarget 9 (15) (2018) 11999–12008, https://doi.org/10.18632/oncotarget.23813.

[25] Y. Cui, X. Yang, Z. Shi, Z. Yang, X. Du, Z. Zhao, X. Cheng, Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer, Eur. Radiol. 29 (3) (2018) 1211–1220, https://doi.org/10.1007/s00330-018-5683-9.

[26] R.C. Carlos, C.E. Kahn, S. Halabi, Data science: big data, machine learning, and artificial intelligence, J. Am. Coll. Radiol. 15 (3 Pt B) (2018) 497–498, https://doi.org/10.1016/j.jacr.2018.01.029.

[27] V. Rousson, T. Zumbrunn, Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies, BMC Med. Inform. Decis. Mak. (2011) 45, https://doi.org/10.1186/1472-6947-11-45.

[28] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A.E. Samir, O.S. Pianykh, J.R. Geis, P.V. Pandharipande, J.A. Brink, K.J. Dreyer, Current applications and future impact of machine learning in radiology, Radiology 288 (2) (2018) 318–328, https://doi.org/10.1148/radiol.2018171820.

[29] S.H. Park, K. Han, Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction, Radiology 286 (3) (2018) 800–809, https://doi.org/10.1148/radiol.2017171920.

[30] J.H. Thrall, X. Li, Q. Li, C. Cruz, S. Do, K. Dreyer, J. Brink, Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success, J. Am. Coll. Radiol. 15 (3 Pt B) (2018) 504–508, https://doi.org/10.1016/j.jacr.2017.12.026.

[31] T.K. Ho, Random decision forests, Proceedings of the Third International Conference on Document Analysis and Recognition, IEEE Computer Society Vol. 1 (1995) p. 278.