

Magnetic Resonance, Vendor-independent, Intensity Histogram Analysis Predicting Pathologic Complete Response After Radiochemotherapy of Rectal Cancer

Citation for published version (APA):

Dinapoli, N., Barbaro, B., Gatta, R., Chiloiro, G., Casa, C., Masciocchi, C., Damiani, A., Boldrini, L., Gambacorta, M. A., Dezio, M., Mattiucci, G. C., Balducci, M., van Soest, J., Dekker, A., Lambin, P., Fiorino, C., Sini, C., De Cobelli, F., Di Muzio, N., ... Valentini, V. (2018). Magnetic Resonance, Vendor-independent, Intensity Histogram Analysis Predicting Pathologic Complete Response After Radiochemotherapy of Rectal Cancer. *International Journal of Radiation Oncology Biology Physics*, 102(4), 765-774. <https://doi.org/10.1016/j.ijrobp.2018.04.065>

Document status and date:

Published: 15/11/2018

DOI:

[10.1016/j.ijrobp.2018.04.065](https://doi.org/10.1016/j.ijrobp.2018.04.065)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 07 Mar. 2021

Prognostication

Magnetic Resonance, Vendor-independent, Intensity Histogram Analysis Predicting Pathologic Complete Response After Radiochemotherapy of Rectal Cancer



Nicola Dinapoli, MD, PhD,* Brunella Barbaro, MD,[†]
Roberto Gatta, PhD,* Giuditta Chiloire, MD,[†] Calogero Casà, MD,[†]
Carlotta Masciocchi, MSc,[†] Andrea Damiani,* Luca Boldrini, MD,[†]
Maria Antonietta Gambacorta, MD, PhD,[†] Michele Dezio, MD,[†]
Gian Carlo Mattiucci, MD, PhD,[†] Mario Balducci, MD,[†]
Johan van Soest, MSc,[‡] Andre Dekker, PhD,[‡]
Philippe Lambin, MD, PhD,[‡] Claudio Fiorino, PhD,[§]
Carla Sini, PhD,[§] Francesco De Cobelli, MD,^{||} Nadia Di Muzio, MD,[¶]
Calogero Gumina, MD,[¶] Paolo Passoni, MD,[¶] Riccardo Manfredi, MD,[†]
and Vincenzo Valentini, MD[†]

*Fondazione Policlinico Universitario A. Gemelli IRCCS, Dipartimento Scienze Radiologiche, Radioterapiche ed Ematologiche, Roma, Italia; [†]Fondazione Policlinico A. Gemelli IRCCS - Università Cattolica Sacro Cuore, Dipartimento Scienze Radiologiche, Radioterapiche ed Ematologiche, Istituto di Radiologia, Roma, Italia; [‡]Department of Radiation Oncology MAASTRO Clinic GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, The Netherlands; [§]Medical Physics, San Raffaele Scientific Institute, Milan, Italy; ^{||}Radiology, San Raffaele Scientific Institute, Milan, Italy; and [¶]Radiotherapy, San Raffaele Scientific Institute, Milan, Italy

Received Dec 29, 2017, and in revised form Apr 13, 2018. Accepted for publication Apr 23, 2018.

Summary

This magnetic resonance (MR)-based, vendor-independent intensity based histogram (IBH) model can

Purpose: The objective of this study is finding an intensity based histogram (IBH) signature to predict pathologic complete response (pCR) probability using only pre-treatment magnetic resonance (MR) and validate it externally in order to create a workflow for the external validation of an MR IBH signature and to apply the model out of the environment where it has been tuned. The impact of pCR and the final predictors on the survival outcome were also evaluated.

Reprint requests to: Carlotta Masciocchi, MSc, Fondazione Policlinico A. Gemelli IRCCS - Università Cattolica Sacro Cuore, Dipartimento Scienze Radiologiche, Radioterapiche ed Ematologiche, Istituto di Radiologia, Roma, Italia. Tel: +39 393 9913958; E-mail: carlotta.masciocchi@unicatt.it

Conflict of interest: none.

Supplementary material for this article can be found at <https://doi.org/10.1016/j.ijrobp.2018.04.065>.

Disclosure: All authors individually contributed and approved the final article.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

predict pathologic complete response (pCR) probability in locally advanced rectal cancer patients using only pre-treatment imaging. The use of the radiomic features, combined with the clinical information, can add information in predicting pCR. Furthermore, it is an externally validated model with 2 data sets coming from 2 different institutions.

Methods and Materials: Three centers using different MR scanners were involved in this retrospective study. The first center recruited 162 patients for model training, and the second and third centers provided 34 plus 25 patients for external validation. Patients provided written consent. Accrual period was from May 2008 to December 2014. After surgery pathologic response was defined. T2-weighted MR scans acquired before chemoradiation therapy (CRT) were used for analysis addressed on primary lesions. Images were pre-processed using Laplacian of Gaussian (LoG) filter with multiple σ , and first order intensity histogram-based features (kurtosis, skewness, and entropy) were extracted. Features selection was performed using Mann-Whitney test. Tumor staging (cT, cN) was added to build a logistic regression model and predict pCR. Model performance was evaluated with internal and external validation using area under the curve (AUC) of the receiver operator characteristic (ROC) and calibration with Hosmer-Lemeshow test. The linear cross-correlation matrix (Pearson's coefficient) and the variance inflation factor (VIF) were used to check the correlation and the co-linearity among the final predictors. The amount of the information added through the radiomics features was estimated by using the DeLong's test, and the impact of pCR and the final predictors on survival outcomes were evaluated through the Kaplan-Meier curves by using the log-rank test and the multivariate Cox model.

Results: Candidate-to-analysis features were skewness ($\sigma = 0.485$, P value = .01) and entropy ($\sigma = 0.344$, P value < .05). Logistic regression analysis showed as significant covariates cT (P value < .01), skewness- $\sigma = 0.485$ (P value = .01), and entropy- $\sigma = 0.344$ (P value < .05). Model AUCs were 0.73 (internal) and 0.75 (external).

Conclusions: This MR-based, vendor-independent model can be helpful for predicting pCR probability in locally advanced rectal cancer (LARC) patients only using pre-treatment imaging. © 2018 Elsevier Inc. All rights reserved.

Introduction

The staging process in oncology is based on radiological images analysis and reporting. The TNM staging process itself is widely based on radiological definition of boundaries of primary lesion, lymph-nodal, and distant metastases. Nowadays radiological images can represent a source of data that can be analyzed by using automated computer-based techniques, working on numerical informations coded within the Digital Imaging and Communications in Medicine (DICOM) files (1): this images numerical analysis has been named "radiomics" (2). It uses a high number of numerical features that can be extracted automatically by images and analyzed for decoding the tumor "phenotype" as a set of significant numerical values. An accurate process of feature selection always underlies the final radiomics model extraction. In the radiomics workflow the use of computed tomography (CT) images has shown a meaningful correlation with clinical outcome in many anatomic sites, such as lung tumors (2, 3). Considering the locally advanced rectal cancer (LARC), first reports showed the possibility to achieve similar findings using the CT scan in mixed population patients (rectum and colon) (4) and MR (5, 6), but neither an external validation (for the CT) nor a wide number of patients (for the MR) were provided to strengthen the value of the radiomics features. Magnetic

resonance imaging (MRI) has good accuracy for both circumferential resection margin involvement and T category definition and should be considered for preoperative rectal cancer staging (7) as a routine procedure beside CT scan. MR has proven to be effective for LARC staging and prognosis evaluation in several publications (8-10). Another important aspect in LARC treatment is the use of preoperative CRT, nowadays considered the standard for reducing the probability to address patients to abdominal-perineal resection and for increasing local control (11, 12). The pCR after CRT has been shown to be significant for predicting sphincter preservation and survival outcomes (13, 14), and its correlation with MR imaging findings has been already proven (10). Starting from this background, this work primarily intended to define an MR IBH signature able to distinguish patients that will show, or not, a pCR after CRT for LARC cases. The main goal of this work was trying to create an IBH signature-based predictive model, able to classify patients *before* starting the CRT treatment, adding prediction to the traditional clinical information and supporting the possibility, in the future, to prescribe treatments that could be really *tailored* to tumor phenotype as coded by the radiomics purpose. A secondary objective was creating a workflow for the external validation of an MR IBH signature in order to apply the model out of the environment where it has been tuned. Another important

issue in radiomics is the reproducibility of the signature. This fact led us in trying to use only first order statistical features that are a statistical evaluation of intensity histograms, choosing features that are not influenced by inter-patients variability of the signals when average shift about intensity values is linear (see [Appendix E1](#); available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>).

Methods and Materials

Patients selection

Patients enrolled in this study have been collected from three different centers. Patients coming from center #1 were used for model training (training set-TS) and patients coming from the centers #2 and #3 for model external validation (verification set-VS1 and VS2). Patients enrollment criteria were: (1) pathologically proven LARC, without distant metastases at diagnosis; (2) a pelvic MR staging study with good image quality (absence of macroscopic artifacts due, for example, to hip prostheses); (3) age at diagnosis time greater than 18 years; (4) availability of tumor regression grade (TRG) (13) classification in the pathologic report; (5) signed informed consent for retrospective data collection. The same criteria were used for collecting patients in the second and third centers and building the VSs. For VS1 patients were recruited in retrospective fashion, for VS2 patients belong to THUNDER trial (15) (NCT 00969657) for which patients had provided consent for MR images analysis. Data in TS were collected retrospectively from May 2008 to December 2014.

Staging, MR imaging technique, and delineation of gross tumor volume (GTV)

Before CRT treatment, patients recruited in the TS were staged by using an MR 1.5 T unit (GE Signa Excite, Medical Systems), according to a protocol (10) that comprises T2-weighted fast spin-echo 2D oblique images ([Table E1](#); available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>) acquired in a transverse plane orthogonal to the tumor longitudinal axis. No intravenous contrast medium was administered. Images were loaded in a radiation therapy delineation console (Eclipse, Varian Medical System) for the definition of lesion outline (“GTV” as defined in International Commission on Radiation Units and Measurements [ICRU] n. 83 [16]). Patients in VS1 were recruited after local ethical committee approval of data sharing for external validation process. Patients in the VS2 were extracted from the list of THUNDER trial (15) patients who had available MR diagnostic imaging. The selection of the MR studies has been conducted by selecting those who had acquisition parameters close to the range of TS pre-defined protocol and a pixel spacing value not higher than 0.76 mm regardless of the scanner brand. The only criterion for the scanner hardware

selection was the magnetic field power (1.5 T). Finally, all selected patients resulted to be scanned by a Philips Achieva Nova in center 1 and Siemens Magnetom AVANTO in center 2 (17, 18). Patients population characteristics are summarized in [Table 1](#). The GTV delineation was assessed by cooperation of 2 radiation oncologists (ND, CC) and 2 radiologists (BB, MD) experienced in rectal cancer diagnosis and treatment for TS and VS2, in VS1 GTV delineation was locally defined by different experienced radiation oncologist and radiologist (NDM, FDC). According to the TRG score (10, 13) all patients were divided in 2 classes: complete responders (corresponding to TRG = 1 or pCR) or not (corresponding to TRG >1).

Imaging radiomics analysis

Images were analyzed using moddicom, an in-house developed R statistical software package (19). Moddicom has in charge the complete workflow of features extraction and images analysis: it creates a data set composed by records containing arrays of voxel data (one for each patient) within the borders of delineated GTVs. Images were filtered using Laplacian of Gaussian (LoG) convolution kernel filter, a tool chosen in order to decrease the MRI high frequency signal noise and, at the same time, reduce the impact of large variations of signal that can be detected within a single image slice (eg, variation due to magnetic field irregularities): numerical data were pre-processed by using the convolution kernel filter implemented in moddicom and tuned according to the size of standard deviation (σ) in the LoG equation (1); value of σ was set in mm; x and y are the coordinates of pixels surrounding the central one (on which equation is used to calculate the convolution):

$$(1) \text{LoG}(x,y) = -1/(\pi\sigma^4)\{1 - [(x^2 + y^2)/(2\sigma^4)]\} \\ * \exp[-(x^2 + y^2)/(2\sigma^2)]$$

Features selection process

After image data set completion moddicom has been used to extract image features. Two types of features were extracted: (1) geometrical (GTV volume, GTV surface) and (2) histogram-based features (skewness, kurtosis, and Shannon entropy), these latter widely used in radiomics publications in the past (2, 4, 20-22). In moddicom, a function for scanning every possible σ value from 0.2 mm to 1.4 mm (step 0.01 mm) was created extracting 1200 features analyzed versus final outcome (pCR and not-pCR patients) by univariate Mann-Whitney test and the area under the receiver operating characteristic curve (AUC). P values under .05 were considered significant.

Statistical analysis

R statistical software version 3.3.1 was used for statistical analysis. Heterogeneity between patient groups (TS vs VS) was evaluated by Pearson’s χ^2 test with Yates’ continuity

Table 1 Patients characteristics and descriptive statistics of variables in Training Set (TS), Verification Set 1 (VS1), and Verification Set 2 (VS2)

	TS (162)		VS1 (39)		P values differences TS vs VS1		VS 2 (25)		P values differences TS vs VS2	
	Patient characteristics									
	Age				Statistic test		Age		Statistic test	
	Median	Range	Median	Range	χ^2	MW	Median	Range	χ^2	MW
Sex										
Male	66.0	28-83	62.0	43-77	-	0.18	66	52-80	-	0.661
Female	61.0	43-80	56.0	35-79	-	0.25	66.4	47-73	-	0.588
Age (overall)	65.0	28-83	60.5	35-79	-	0.14	66	47-80	-	0.963
Interval between MRI-start chemoradiation	1.4	0-10	0.5	0-2	-	<0.01	1.8	0-11	-	0.945
Interval between Surgery-CRT [mo]	2.6	1-10	2.7	1-5	-	0.9	3	1-9	-	0.942
	Clinical features									
	Number	%	Number	%	χ^2	MW	Number	%	χ^2	MW
Stage										
cT					0.14	-			0.12	-
2	15	9	2	6			1	4		
3	95	59	26	76			20	80		
4	52	32	6	17			4	16		
cN					0.17	-			0.41	-
0	9	5.5	3	9			3	12		
1	58	36	17	50			7	28		
2	95	59	14	41			15	60		
Response					0.47	-			1.0	-
TRG = 1	46	28	7	21			7	28		
TRG > 1	116	72	27	79			18	72		
RT dose*										
43.8	-	-	1	3.0	-	-	-	-	-	-
45	-	-	8	23.5	-	-	-	-	-	-
46.2	-	-	25	73.5	-	-	-	-	-	-
50.4	143	88	-	-	-	-	25	100	-	-
55	19	12	-	-	-	-	-	-	-	-
Concomitant CT type					-	-			-	-
CAP	45	28	-	-			25	100		
CAP+OX	114	70	-	-			-	-		
FU	3	2	-	-			-	-		
OX+FU	-	-	34	100			-	-		

Abbreviations: χ^2 = Pearson's χ^2 ; CAP = capecitabine only; CAP+OX = capecitabine plus oxaliplatin; cN = clinical N stage; CRT = chemoradiation therapy; CT = chemotherapy; cT = clinical T stage; MW = Mann-Whitney; OX+FU = oxaliplatin + fluorouracil; RT = radiation therapy; TRG = 1 = pathologic complete response; TRG > 1 = pathologic not complete response; TS = Training Set; VS1 = Verification Set 1; VS2 = Verification Set 2.

* Dose prescription details, doses are referred on GTV.

TS: 50.4 Gy = 1.8 Gy \times 25 fr + 1.8 \times 3 fr; 55 Gy = 1.8 Gy \times 15 fr + 2.8 \times 10 fr;

VS1: 43.8 Gy = 2.3 Gy \times 12 fr + 2.7 Gy \times 6 fr; 45 Gy = 2.3 Gy \times 12 fr + 2.9 Gy \times 6 fr, 46.2 = 2.3 Gy \times 12 fr + 3.1 Gy \times 6 fr;

VS2: 50.4 Gy = 1.8 \times 25 fr + 1.8 \times 3 fr.

correction for categorical variables and Mann-Whitney test for numerical ones. A logistic regression model was computed for modeling the binary outcome (pCR vs not-pCR) by backward elimination, using the best univariate analysis performing features, clinical T stage (cT) and clinical N stage (cN) on TS: 2-tailed *P* value

corresponding to Z-ratio based on normal distribution was used for evaluating statistical significance. The correlation and the co-linearity of the final predictor features were evaluated by using the linear cross-correlation matrix (Pearson's coefficient) and the Variance Inflation Factor (VIF), respectively. Discrimination power was calculated

using AUC of ROC of the model, and external validation was performed on VS. Calibration of the model was calculated for both TS and VS using Hosmer and Lemeshow goodness-of-fit (GOF) test: P values $< .05$ indicate lack of fit of the model. In order to evaluate how much information the 2 radiomic features added to the cT in predicting the pCR, the AUC values of the cT model (trained in the univariate way) and the final model were compared through the DeLong's test. Survival analysis was made by using Kaplan-Meier log-rank test in univariate analysis and Cox model in multivariate analysis in order to evaluate the impact of the final predictors on the overall survival (OS), loco-regional control (LRC), metastases-free survival (MFS) and disease-free survival (DFS). Moreover, the impact of TRG on survival outcomes has been evaluated to assess actual impact of this parameter on patients' survival. OS was defined as time from diagnosis (biopsy) and death or last follow-up (censor), LRC was defined as time from surgery and loco-regional recurrence or last follow-up (censor), MFS was defined as time from surgery and metastasis finding or last follow-up, and DFS survival was defined as the minor value reported in MFS or LRC.

Results

A total number of 162 patients were selected in the TS; 34 patients were included from center 1 in VS1 and 25 patients were included from center 2 in the VS2 after the analysis of MR DICOM header. All patients have been treated by using 3D conformal radiation therapy and concurrent chemotherapy. Only patients with complete treatment course have been selected. All histological types were adenocarcinoma and the median time between CRT, and surgery was 2.6 months (range 1-10 months). Details of TS and VSs patients and their treatment features are summarized in Table 1: no significant differences are observed in the distribution of clinical characteristics of the patients and in the proportion of pCR. The overall response rate (TRG = 1) was 28% (46/162) in TS, 21% (7/34) in VS1, and 28% (7/25) in VS2 (not significantly different among different centers: P value TS-VS1 = 0.47, TS-VS2 = 1.00). The application of LoG filter on raw MR images produces images that are different from the original ones, enhancing the short-range differences in pixel intensities and revealing the texture appearance according to the σ value. An example of application is given in Figure 1, where it is possible to see how, with increasing the σ value of LoG, the coarseness of MR GTV texture appears greater. This process produces different values of features calculated over the total amount of analyzed voxels. And this is the reason why a features selection process is needed for browsing through a continuous range of σ values and finding where the Mann-Whitney test returns the most significant (lowest) P values: these σ values correspond to highest AUC and have been used for extracting the covariates to be put in

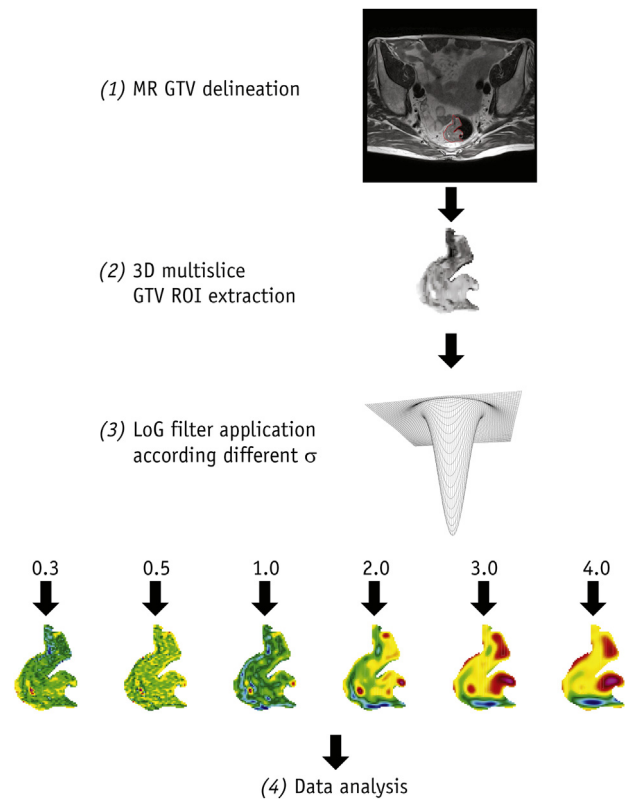


Fig. 1. Radiomics workflow: after image acquisition (1) the gross tumor volume (GTV) is delineated. Voxel belonging to GTV are extracted in 3D fashion building a 3D array (2) and afterwards processed by applying Laplacian of Gaussian filter (LoG - 3). The σ parameter of LoG has to be tuned in order to identify the values that better enhance the features of the GTV related to outcome prediction.

the logistic regression model and predicting the final pCR probability. The most significant features were skewness with $\sigma = 0.485$ mm (SKE0485, P value = .010) and entropy with $\sigma = 0.344$ mm (ENT0344, P value $< .031$); kurtosis didn't show significant values (Fig. 2). The multivariate analysis performed by logistic regression model, putting inside geometrical and IBH features besides the cT and cN for clinical classification, showed that only SKE0485, ENT0344, and cT were statistically significant (see Table 2). The Pearson's coefficient and the VIF were < 0.4 and < 1.2 , respectively, showing no linear correlation and co-linearity among the final predictors (23, 24) (for additional information see Fig. E2 and Table E2; available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>). The logistic model has been represented by a nomogram that can be used for calculating the prediction value starting from known clinical and IBH features achieved for a single case (Fig. E3; available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>). The discrimination performance of the model, evaluated by AUC of the ROC, showed an AUC = 0.73 (95% CI 0.65-0.82) (Fig. 3A); the internal calibration of the model didn't show significant lack of fit (GOF test P value = .77, Fig. 3B). After model tuning the

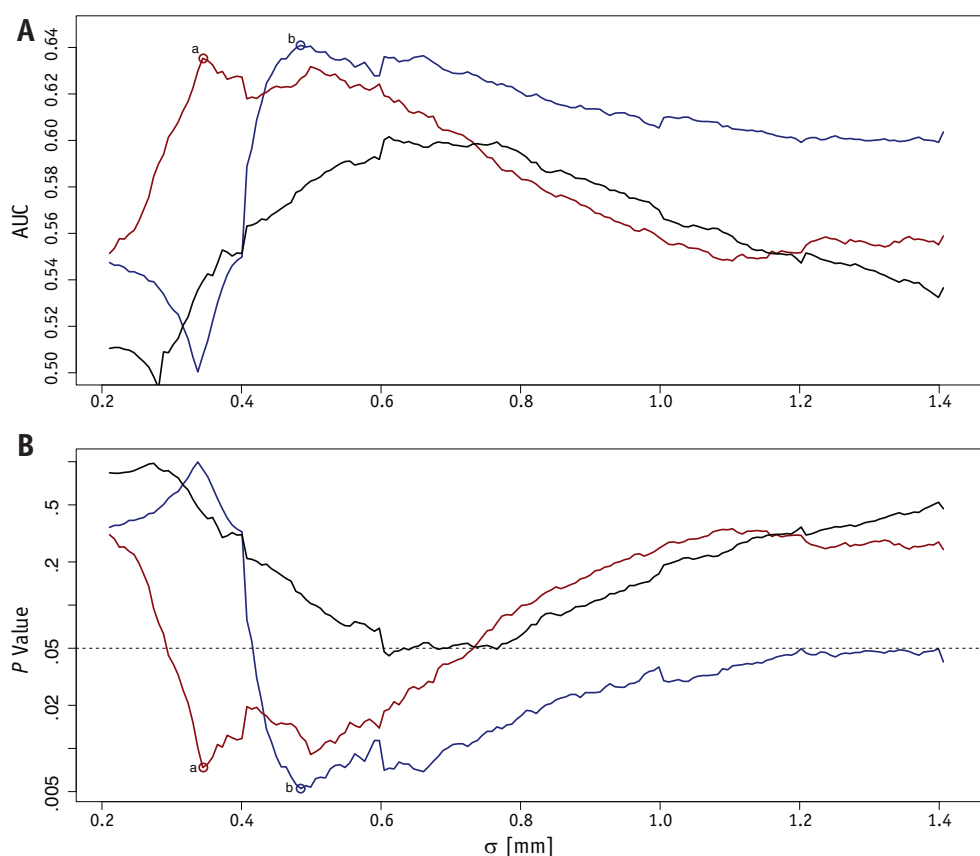


Fig. 2. Features selection: the two plots show the process of features selection for radiomics modeling. In the plot A the univariate AUC is provided, and it can be easily related to the P value of not parametric univariate Mann-Whitney test for TRG = 1 versus TRG > 1 discrimination power (B). Only the most significant features (points *a* and *b*) in the set have been chosen for the definition of multivariate logistic model. P Values = .05 in univariate test (horizontal solid reference line in plot B) are the threshold level chosen for excluding features from selection. Features: kurtosis (black line), skewness (blue line), entropy (red line). (A color version of this figure is available at <https://doi.org/10.1016/j.ijrobp.2018.04.065>.)

external verification was conducted on VS1 and VS2 separately and by merging the 2 data sets: both sets showed similar AUCs (VS1 AUC = 0.79, 95% CI 0.55-0.94, VS2 AUC = 0.75, 95% CI 0.59-0.98, Fig. 3C) and

not significant P value in the GOF test (VS1 P value = .94, VS2 P value = .94, Fig. 3D). The combination of VS1 and VS2 showed combined AUC = 0.75 (95% CI 0.61-0.88) and again not significant P value in GOF test (P value = .98). The difference between the 2 AUCs (only cT vs final model) was recorded as statistically significant difference ($P < .01$) (for additional information see Fig. E4; available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>). The number of events in the TS for survival outcomes were: death = 17, local recurrence = 7, distant metastases = 25, overall disease relapse = 15. Median follow-up time calculated by Kaplan-Meier estimate is 40.9 months (range 4-83 months). The Kaplan-Meier log-rank tests between pCR and not-pCR calculated on survival outcomes give back significant P value for DFS (P value < .05, hazard-ratio (HR) = 0.24, 95% CI 0.07-0.79, Fig. 4A) and MFS (P value = .05, HR = 0.28, 95% CI 0.08-0.93 Fig. 4B), an almost not significant P value for LRC (P value = .10, HR = 3.4e-09, 95% CI 0-Inf) (for additional information see Fig. E5; available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>), and not

Table 2 Result of logistic regression model tuned after backward elimination on clinical and intensity based histogram features

Radiomics modeling results	Estimate	Std. error	z value	P value ($> z $) [*]
cT	-0.947	0.3604	-2.628	.0086
cN	0.531	0.3446	1.542	.1232
SKE0485	-3.013	1.17	-2.571	.010
ENT0344	3.610	1.676	2.154	.031
(Intercept)	-6.1850	3.003	-1.545	.1224

Abbreviations: cN = clinical N stage; cT = clinical T stage; ENT0344 = entropy with Laplacian of Gaussian filter $\sigma = 0.344$ mm; SKE0485 = skewness with Laplacian of Gaussian filter $\sigma = 0.485$ mm.

^{*} Bold P values are considered significant and enter in the final pathologic complete response predictive model.

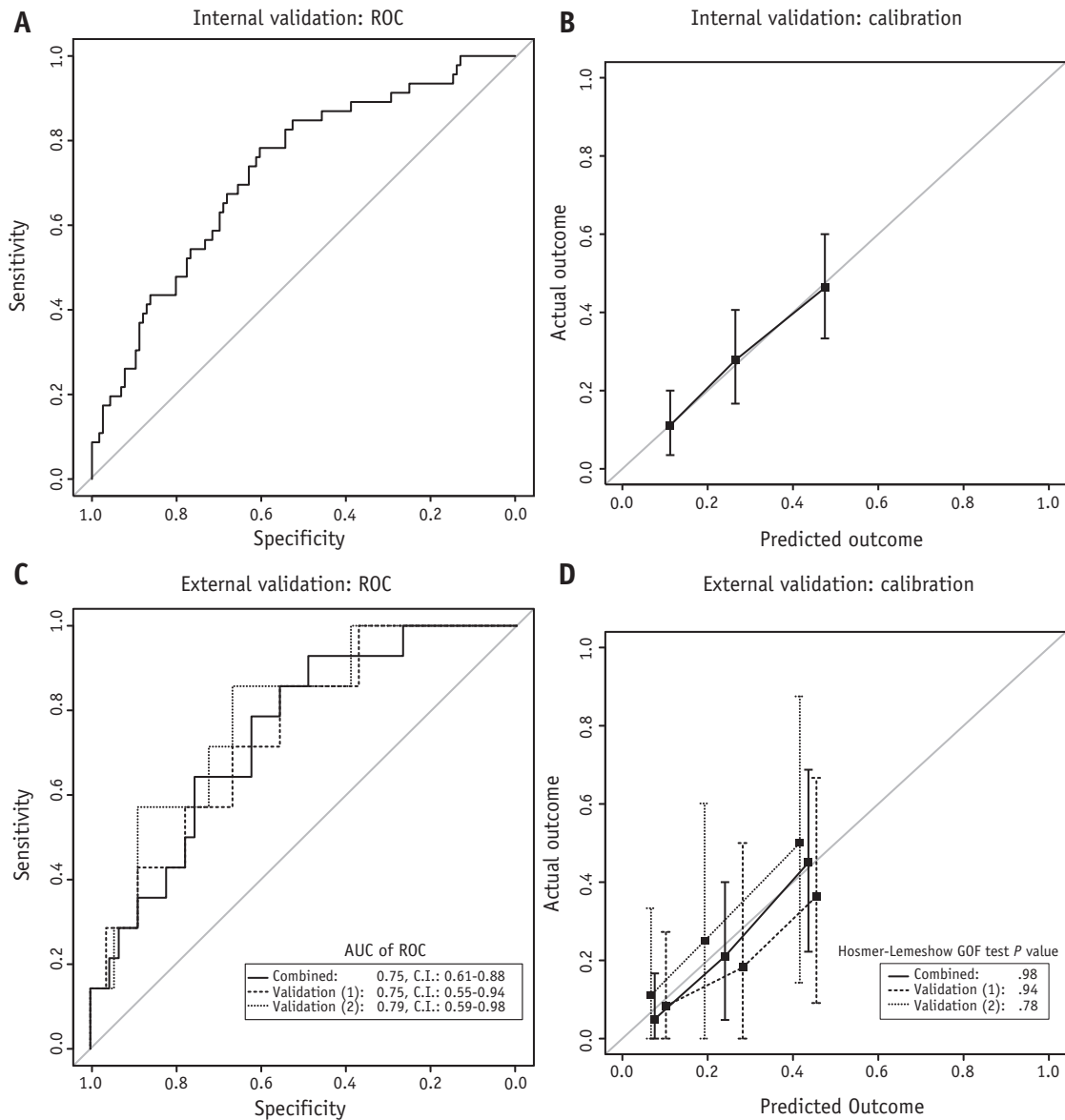


Fig. 3. Model validation. (A) Internal validation: receiver operating characteristic (ROC) curve of multivariate logistic model calculated on the original data set. Area under the curve (AUC) is 0.73 (95% CI 0.65-0.82). (B) Internal validation: calibration plot showing discrepancy between predicted probabilities of TRG = 1 (x axis) and actual outcome (y axis). The brackets show 95% CI of prediction; square dots show grouped tertiles of training set population. (C) External validation: receiver operating characteristic (ROC) curve of multivariate logistic model calculated on the validation data sets. Area under the curve (AUC) is 0.75 for VS1 (95% CI 0.55-0.94), 0.79 for VS2 (95% CI 0.59-0.98), combined AUC is 0.75 (95% CI 0.61-0.88). (D) External validation: calibration plot showing discrepancy between predicted probabilities of TRG = 1 (x axis) and actual outcome (y axis). The brackets show 95% CI of prediction, square dots show grouped tertiles of validation set population, continuous line shows the combined VS1 and VS2, dashed line shows VS1, and dot line shows VS2.

significant impact in OS (P value = .07, HR = 0.18, 95% CI 0.02-1.38, Fig. 4C). Other evaluated covariates (sex, cT, and cN) didn't return significant P values in univariate analysis. Multivariate survival analysis (Cox model) didn't give significant values in this cohort of patients for the chosen covariates (Sex, Age, cT, cN, SKE0485, ENT0344, chemotherapy regimen, radiation therapy dose, time gap between end of radiation therapy treatment and surgery).

Discussion

Imaging-based predictive models for rectal cancer have been published in the past, mostly using PET (25, 26). Radiomics in MR can return issues that depend from typical absolute value variations of MR signal recorded in numerical format inside DICOM files (27) changing according patients, sequences, acquisition parameters, or

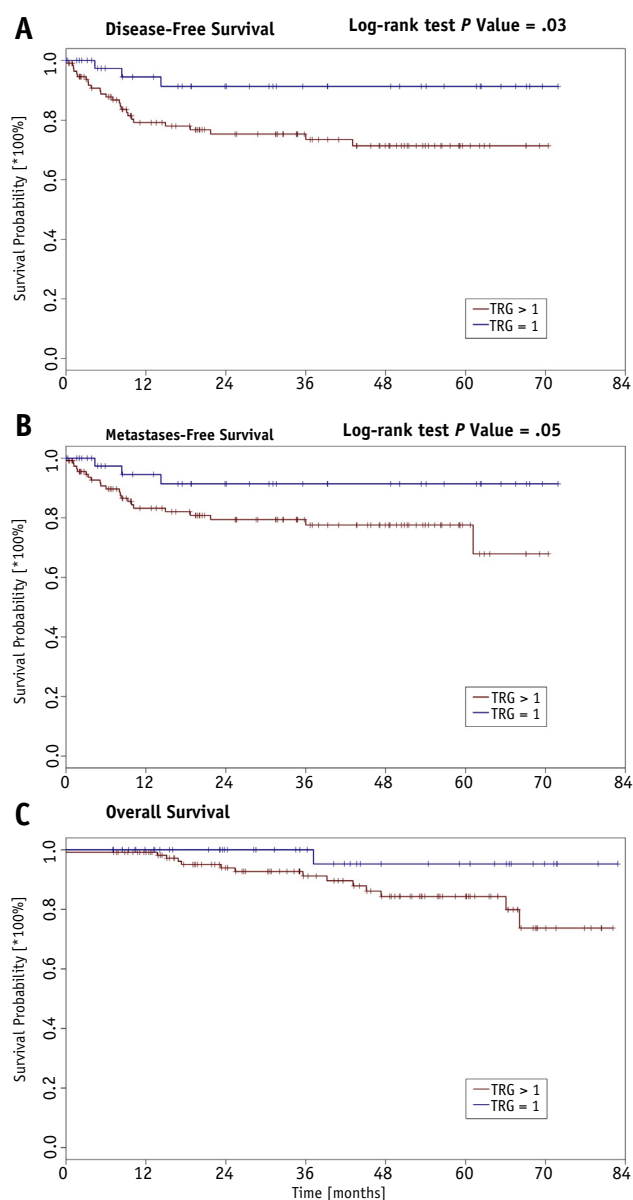


Fig. 4. (A) Disease-free survival curve in the training set population: patients are divided in cases showing TRG = 1 (blue line) and cases showing TRG > 1 (red line). The log-rank test shows significant result (P value = .03); (B) Metastases-free survival curve in the training set population: patients are divided in cases showing TRG = 1 (blue line) and cases showing TRG > 1 (red line). The log-rank test shows significant result (P value = .05); (C) Overall survival curve in the training set population: patients are divided in cases showing TRG = 1 (blue line) and cases showing TRG > 1 (red line). The log-rank test doesn't show significant result (P value > .05). (A color version of this figure is available at <https://doi.org/10.1016/j.ijrobp.2018.04.065>.)

simply time. Furthermore, MR can be affected by several patterns of noise that can interfere with quantitative analysis of signal values (28), and heterogeneity of MR

numerical data is a well-known issue for the analysis (29). The LoG filter has been extensively used in radiomics in literature, mostly in lung, and using CT images (30-32). LoG filter has also been used in the analysis of colon-rectum imaging by CT (4) and in 1 early rectal cancer radiomics experience based on a 3T MR imaging (22). LoG filter enhances textural details that characterize the tumor lesions as shown in Figure 1. Furthermore, it is straightforward that the use of first order features invariant for linear transformations overcomes the issue of inter-patients signal variations that can affect MR radiomics analysis (see Appendix E1; available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>). The extraction of features strongly depends on the size of the σ value. In this work the IBH is able to predict the probability of pCR in patients undergoing to pre-operative CRT treatment *without* the need to compare the results of this analysis with other (posttreatment) imaging data, while previous imaging based predictive models (even when not strictly *radiomics* oriented) were mainly based on pre- and posttreatment imaging comparison (9, 26, 33). A previous publication has shown a good predictive power using MR in pathologic response prediction, using T1/T2, diffusion-weighted-MRI (DWI), and dynamic-contrast-enhanced (DCE)-MRI (5), but it used a small patient population (48 patients), and the model was not externally validated. As in our work, this paper classified patients using the surrogate outcome of the pCR, but in our work we furthermore applied *an independent external validation process*. This validation uses 2 different brands of MR scanners (but using similar protocol and pixel spacing resolution and equal magnetic field power 1.5T), and the only prerequisite for determining images comparability was the analysis of data coming from DICOM tags and related to scanning protocol details. As a consequence, the level of reliability of the model is granted of Type 3 (on 4 levels) according to the TRIPOD classification (34). The reason of this result is to ascribe to LoG filter use, coupled with features that describe the *shape* of the histograms (skewness, entropy), without being influenced by the absolute values in the histograms (and MR images) themselves. This process is able to overcome the inter-patients variations of signals that are common findings in the numerical analysis of MR images (35). The pCR is entitled to be a good surrogate endpoint widely used in the past in order to predict which patients show a higher risk of recurrence (13, 36, 37). Our work confirms this finding as shown by significantly lower HR for local or distant recurrence in patients showing pCR although not showing direct significant correlation between IBH parameters and survival outcome and furthermore demonstrates that the addition of radiomics features added information to the cT alone in predicting the pCR. This model has further limitations described hereafter: (1) the validation sets have been obtained in order to admit as many as possible patients coming from case series in which MR protocols for patients staging were different in detail. This fact leads to obtain small number of patients despite

requiring to 2 different institutions series: 34 patients from the first institution and 25 patients from the second one. So, the first issue of such a model is its applicability outside of the range of the acquisition protocols described in section 3 of Appendix E1 (available online at <https://doi.org/10.1016/j.ijrobp.2018.04.065>). In our opinion this is a common weak point of all MR-based radiomics models as previously described in literature (38). (2) Another issue in the modeling process is given by the GTV delineation process: in literature, when feasible, examples of automated delineation of regions of interests (ROI) have been proposed (20), but they are used above all when voxel intensity-based delineation algorithms allow to shape them (eg, lung), while, in our knowledge, they lack for anatomic sites and imaging type like the ones we used in this work. At the moment TS and VS2 were delineated by the center that provided the patients in the TS, VS1 was delineated in another center. There could be a residual operator-based bias, but looking just at VS1 testing results the predictive power of this IBH signature seems to overcome such an issue. (3) The final issue about this model is given by the lack of significance in survival outcomes for the same IBH features found significant in predicting pCR, as stated before: not all patients not showing pCR showed events in survival outcome despite the consistence of median follow-up time, so, coupling this factor with the different mathematical approach in computing survival statistics, these features didn't show a significant level differently from the logistic model predicting pCR. In order to encompass this issue, a further validation approach could be realized by increasing patient population and observing the impact of IBH model prediction on survival outcomes themselves.

Conclusion

Radiomics in MR is going to be used day after day for helping physicians in characterizing prognosis of the patients. This evidence achieved in LARC confirms previous smaller case series. This MR-based, vendor-independent IBH model can be considered the first example of a model able to predict pCR probability in LARC patients *only using pre-treatment imaging*. It is also an example of a consistent externally validated model, a basic prerequisite for starting radiomics analysis and driving similar studies as stated recently by Lambin et al (39).

References

- Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441-446.
- Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
- Ganeshan B, Goh V, Mandeville HC, et al. Non-small cell lung cancer: Histopathologic correlates for texture. *Radiology* 2013;266:326-336.
- Ng F, Ganeshan B, Kozarski R, et al. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: Contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology* 2013;266:177-184.
- Nie K, Shi L, Chen Q, et al. Rectal cancer: Assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res* 2016;21:5256-5264.
- De Cecco CN, Ciolina M, Caruso D, et al. Performance of diffusion-weighted imaging, perfusion imaging, and texture analysis in predicting tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3T MR: Initial experience. *Abdom Radiol* 2016;9:1728-1735.
- Al-Sukhni E, Milot L, Fruitman M, et al. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: A systematic review and meta-analysis. *Ann Surg Oncol* 2012;19:2212-2223.
- Vliegen RFA, Beets GL, Lammering G, et al. Mesorectal fascia invasion after neoadjuvant chemotherapy and radiation therapy for locally advanced rectal cancer: Accuracy of MR imaging for prediction. *Radiology* 2008;246:454-462.
- Lambrechts DMJ, Vandecaveye V, Barbaro B, et al. Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: A multicenter study. *Ann Surg Oncol* 2011;18:2224-2231.
- Barbaro B, Fiorucci C, Tebala C, et al. Locally advanced rectal cancer: MR imaging in prediction of response after preoperative chemotherapy and radiation therapy. *Radiology* 2009;250:730-739.
- Sauer R, Liersch T, Merkel S, et al. 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* 2012;30:1926-1933.
- Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004;351:1731-1740.
- Vecchio FM, Valentini V, Minsky BD, et al. The relationship of pathologic tumor regression grade (TRG) and outcomes after preoperative therapy in rectal cancer. *Int J Radiat Oncol Biol Phys* 2005;62:752-760.
- Valentini V, van Stiphout RGPM, Lammering G, et al. Nomograms for predicting local recurrence, distant metastases, and overall survival for patients with locally advanced rectal cancer on the basis of European randomized clinical trials. *J Clin Oncol* 2011;29:3163-3172.
- Validation of a Predictive Model After Complete Response in Rectal Cancer (Thunder). 2011. <https://clinicaltrials.gov/ct2/show/NCT00969657>.
- ICRU. ICRU 83: 4. Definition of volumes. *J ICRU* 2010;10:41-53.
- Palmisano A, Esposito A, Di Chiara A, et al. Could early tumour volume changes assessed on morphological MRI predict the response to chemoradiation therapy in locally-advanced rectal cancer? *Clin Radiol* 2018;73:555-563.
- Passoni P, Fiorino C, Slim N, et al. Feasibility of an adaptive strategy in preoperative radiochemotherapy for rectal cancer with image-guided tomotherapy: Boosting the dose to the shrinking tumor. *Int J Radiat Oncol Biol Phys* 2013;87:67-72.
- Dinapoli N, Alitto AR, Vallati M, et al. Moddicom: a complete and easily accessible library for prognostic evaluations relying on image features. *Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf* 2015;2015:771-774.
- Parmar C, Velazquez ER, Leijenaar R, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* 2014;9:1-8.

21. Miles KA, Ganeshan B, Hayball MP, et al. CT texture analysis using the filtration-histogram method: What do the measurements mean? *Cancer Imaging* 2013;13:400-406.
22. De Cecco CN, Rengo M, Meinel FG, 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* 2015;50:239-245.
23. Tabachnick B, Fidell LS. *Using Multivariate Statistics*. 5th ed. Boston, PA: Allyn & Bacon/Pearson Education; 2007.
24. Akinwande MO, Dikko HG, Samson A, et al. Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open J Stat* 2015;5:754.
25. Janssen MHM, Aerts HJWL, Buijsen J, et al. Repeated positron emission tomography-computed tomography and perfusion-computed tomography imaging in rectal cancer: Fluorodeoxyglucose uptake corresponds with tumor perfusion. *Int J Radiat Oncol Biol Phys* 2012; 82:849-855.
26. Van Den Bogaard J, Janssen MHM, Janssens G, et al. Residual metabolic tumor activity after chemo-radiotherapy is mainly located in initially high FDG uptake areas in rectal cancer. *Radiother Oncol* 2011;99:137-141.
27. Tofts P. *Concepts: Measurement and MR*. In: Tofts P, editor. *Quant MRI Brain*. Chichester, UK: John Wiley & Sons, Ltd; 2003.
28. Parker DL, Gullberg GT. Signal-to-noise efficiency in magnetic resonance imaging. *Med Phys* 1990;17:250-257.
29. Jäger F. Normalization of magnetic resonance images and its application to the diagnosis of the scoliotic spine. *Universität Erlangen-Nürnberg* 2010.
30. Thompson M, Gonzalez RC, Wintz P, et al. *Digital Image Processing*. Leonardo: Pearson Prentis Hall; 2008.
31. Davnall F, Yip CSP, Ljungqvist G, et al. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? *Insights Imaging* 2012;3:573-589.
32. Ng F, Kozarski R, Ganeshan B, et al. Assessment of tumor heterogeneity by CT texture analysis: Can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol* 2013;82:342-348.
33. Liu Z, Zhang X-Y, Shi Y-J, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res* 2017;23:7253-7262.
34. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med* 2015;162:55-63.
35. Jäger F, Balda M, Hornegger J. Correction of intensity inhomogeneities utilizing histogram-based regularization. 4th Russ Conf Biomed Eng. 2008:23-27.
36. vanStiphout RGPM, Valentini V, Buijsen J, et al. Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: A multicentric prospective study with external validation. *Radiother Oncol* 2014;113:215-222.
37. Capirci C, Valentini V, Cionini L, et al. Prognostic value of pathologic complete response after neoadjuvant therapy in locally advanced rectal cancer: Long-term analysis of 566 ypCR patients. *Int J Radiat Oncol Biol Phys* 2008;72:99-107.
38. Larue RT, Defraene G, De Ruyscher D, et al. Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures. *Br J Radiol* 2017;90:20160665.
39. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;12:749-762.