



OPEN

## Preoperative prediction of perineural invasion with multi-modality radiomics in rectal cancer

Yu Guo<sup>1</sup>, Quan Wang<sup>2</sup>, Yan Guo<sup>3</sup>, Yiyang Zhang<sup>1</sup>, Yu Fu<sup>1✉</sup> & Huimao Zhang<sup>1✉</sup>

Perineural invasion (PNI) as a grossly underreported independent risk predictor in rectal cancer is hard to identify preoperatively. We aim to predict PNI status in rectal cancer using multi-modality radiomics. In total, 396 radiomics features were extracted from T2-weighted images (T2WIs), diffusion-weighted images (DWIs), and portal venous phase of contrast-enhanced CT (CE-CT) respectively of 94 consecutive patients with histologically confirmed rectal cancer. T2WI score, DWI score, and CT score were calculated via the radiomics features selection and optimization. Discrimination, calibration, and clinical benefit ability were used to evaluate the performance of the radiomics scores in both training and testing datasets. CT score and T2WI score were independent risk predictors [CT score, OR (95% CI) = 4.218 (1.070–16.620); T2WI score, OR (95% CI) = 105.721 (3.091–3615.790)]. The concise score which combined CT score and T2WI score, showed the best performance [training dataset, AUC (95% CI) = 0.906 (0.833–0.979); testing dataset, AUC (95% CI) = 0.884 (0.761–1.000)] and good calibration ( $P > 0.05$  in the Hosmer–Lemeshow test for the training and testing datasets). Decision curve analysis showed that the multi-modality radiomics nomogram had a higher clinical net benefit. The multi-modality radiomics score could be used to preoperatively assess PNI status in rectal cancer.

Rectal cancer is one of the most common cancers and the leading causes of death around the world<sup>1,2</sup>. Perineural invasion (PNI), malignant invasion of nervous structures and nerve sheaths that could be a source of metastatic spread beyond the extend of any local invasion, indicating an increased risk for poor prognosis and decreased survival in rectal cancer<sup>3–6</sup>.

Although the molecular mechanisms behind PNI are obscure, nerves are also regarded as another route of malignant cells dissociate from the local primary tumor to deposit metastases except for vascular and lymphatic channels<sup>3</sup>. PNI is a grossly underreported predictor in rectal cancer, the National Comprehensive Cancer Network describes PNI as an important parameter to be reported in the standard pathological reports<sup>7</sup>. Total mesorectum excision (TME) is the main therapy choice for rectal cancer, and recent studies proved that PNI-positive rectal cancer proves to experience incomplete tumor resection and local tumor recurrence<sup>5,8</sup>. Liebig et al. reported that about 30–40% of rectal cancer are PNI-positive rectal cancer who with a 5-years overall survival of less than 16% and median survival time (MST) of 25 months after surgical resection<sup>8</sup>. Some studies proved that PNI-positive patients could get potential benefits from adjuvant therapy and PNI status should be considered when stratifying rectal cancer patients for adjuvant treatment. Hence, there is a significant clinical implication to detect PNI status preoperatively<sup>9,10</sup>.

The identification of the PNI status preoperatively is based on biopsy specimens which were reviewed by pathologists in photomicrographs. Efficiency and timeliness of detecting PNI status can limit the application to assist cancer treatment decisions<sup>11–13</sup>. Magnetic resonance imaging (MRI) and computer tomography (CT) play a vital role in the rectal tumor detecting, location and the extent of disease defining, treatment planning, and longitudinal response monitoring, which are noninvasively compared with other procedures such as preoperative biopsy and serum test<sup>14–16</sup>. Nevertheless, the traditional radiological image fails to detect the PNI status preoperatively.

<sup>1</sup>Department of Radiology, The First Hospital of Jilin University, Jilin Provincial Key Laboratory of Medical Imaging and Big Data, Changchun, China. <sup>2</sup>Department of Gastric and Colorectal Surgery, The First Hospital of Jilin University, Changchun, China. <sup>3</sup>GE Healthcare, Shanghai, China. ✉email: fuyujdy@163.com; huimaozhanglinda@163.com

Combined with artificial intelligence (AI), MRI and CT would be used to assess tumor phenotype and its local microenvironment which could provide additional information to determining potential treatments in the advanced image analysis algorithms<sup>17</sup>. Currently, this kind of AI-based process to analyze medical imaging and provide more quantitative image features is referred to as "radiomics"<sup>18–20</sup>. In recent years, radiomics have shown effective diagnostic, prognostic, and predictive ability in the clinical decision support systems of oncology<sup>14,21–26</sup>. As important components of the tumor microenvironment, Huang et al. reported that radiomics-based CT model status could predict PNI status in colorectal cancer<sup>27</sup>. However, most of radiomics analyses in oncology usually depend on the unimodality equipment such as CT, MRI, or ultrasound, et cetera<sup>19,20</sup>. Multi-modal machine learning (MMML) is a process that could be used to combine information from multiple modalities and perform a target prediction<sup>28,29</sup>. However, there is still a rare of multi-modality radiomics-based methods to predict PNI status for rectal cancer.

Thus, we aim to investigate the potential value of multi-modality radiomics derived from MRI and CE-CT for the preoperative prediction of PNI status in rectal cancer patients.

## Results

**Patient characteristics.** A total of 94 patients were included. For stage I-II rectal cancer, 40 (88.9%) patients performed radical surgical treatment. 5 (11.1%) patients performed surgery followed by chemotherapy; For stage III rectal cancer, 22 (44.9%) patients performed radical surgical treatment. 24 (49.0%) patients performed surgery followed by chemotherapy. 2 (4.1%) patients performed surgery followed by radiotherapy. 1 (2.0%) patient performed surgery followed by chemoradiotherapy. Included patients were divided into the training dataset (21 PNI-positive rectal cancer and 44 PNI-negative rectal cancer) and testing dataset (10 PNI-positive rectal cancer and 19 PNI-negative rectal cancer) randomly. There was no significant difference between the difference of clinical characteristics between the training dataset and testing dataset ( $P = 0.099–1.000$ ) (Table 1). In the univariable logistic regression analysis, the initially selected clinical risk predictors included CA19-9, T staging, and N staging diagnosed by pathology (all  $P < 0.05$ ). After the multivariable logistic regression analysis, N staging was regarded as independent risk predictors to PNI status ( $P < 0.05$ ) (Table 2). There was more T3-4 rectal cancer (97.09% vs 55.55%) and N1-2 rectal cancer (80.65% vs 38.09%) in the PNI-positive group than in the PNI-negative group (both  $P < 0.05$ ). However, the radiomics model was established preoperatively, N staging wouldn't be contained in the radiomics model.

**Radiomics scores.** 128 radiomics features in the modality of T2WI, 181 radiomics features in the modality of DWI, 65 radiomics features in the modality of portal venous phase CE-CT, which showed stability with good inter-/intra observer agreement (ICCs > 0.70) had been remained. The radiomics features extracted from reader 1's segmentation with stable inter-/intra-observer reproducibility went through the selection process.

In the process of building radiomics score in the modality of T2WI, DWI, and portal venous phase CE-CT, the minimum redundancy maximum correlation (mRMR) selection was used to select the most relevant and the least redundant 20 radiomics features to PNI status respectively. Then the least absolute shrinkage and selection operator (LASSO) method was used. The lambdas were chosen as 0.051, 0.095, and 0.004 for the T2WI radiomics features, DWI radiomics features, and portal venous phase CE-CT radiomics features, respectively and which gave the minimum binomial deviance. 5 T2WI radiomics features (Supplementary Information 1: Figure S1), 1 DWI radiomics feature (Supplementary Information 1: Figure S2), and 12 portal venous phase CE-CT radiomics features (Supplementary Information 1: Figure S3) remained. The radiomics scores were calculated by the selected features multiplied the corresponding coefficients for each modality as follows:

$$\begin{aligned} T2WI \text{ score} = & -0.761 + (-0.258 \times \text{LongRunEmphasis\_angle0\_offset4} - 0.131 \times \text{DifferenceVariance} \\ & - 0.102 \times \text{LongRunEmphasis\_angle45\_offset4} - 0.087 \\ & \times \text{RunLengthNonuniformity\_AllDirection\_offset4\_SD} + 0.408 \times \text{MaxIntensity}) \end{aligned}$$

$$DWI \text{ score} = -0.719 - 0.144 \times \text{InverseDifferenceMoment\_angle0\_offset1}$$

$$\begin{aligned} CT \text{ score} = & -0.988 + (-0.436 \times \text{Sphericity} + 2.838 \times \text{Quantile 0.975} + 0.857 \\ & \times \text{LongRunLowGreyLevelEmphasis\_angle45\_offset4} + 0.874 \times \text{Percentile15} \\ & - 0.665 \times \text{ShortRunLowGreyLevelEmphasis\_AllDirection\_offset7\_SD} + 1.25 \\ & \times \text{HaralickCorrelation\_AllDirection\_offset4\_SD} - 0.511 \times \text{GLCMEnergy\_angle135\_offset7} \\ & + 0.828 \times \text{ClusterShade\_AllDirection\_offset1\_SD} + 2.131 \\ & \times \text{LowGreyLevelRunEmphasis\_angle90\_offset1} \\ & - 3.153 \times \text{Percentile90} - 2.377 \times \text{GLCMEnergy\_angle45\_offset7} \\ & - 2.067 \times \text{ShortRunLowGreyLevelEmphasis\_angle0\_offset1}) \end{aligned}$$

T2WI score and CT score showed a significant difference between the PNI-positive group and the PNI-negative group in the training and testing datasets (all  $P < 0.05$ ) (Table 1). In the univariable logistic regression analysis, the initially selected radiomics risk predictors included T2WI score and CT score (all  $P < 0.05$ ). After the multivariable logistic regression analysis, the T2WI score and CT score were regarded as independent risk predictors to PNI status (all  $P < 0.05$ ) (Table 2). T2WI score and DWI score were combined as MR score. T2WI score and CT score were combined as concise score. T2WI score, DWI score, and CT score were combined as

Variables	All patients (N = 94)	Training dataset (N = 65)	Testing dataset (N = 29)	P value
<b>PNI status</b>				1
Negative	63 (67.02)	44 (67.69)	19 (65.52)	
Positive	31 (32.98)	21 (32.31)	10 (34.48)	
Age	59.20 ± 12.43	59.37 ± 12.94	58.83 ± 11.41	0.846
<b>Gender</b>				0.707
Male	72 (76.60)	51 (78.46)	21 (72.41)	
Female	22 (23.40)	14 (21.54)	8 (27.59)	
<b>CEA level</b>				0.757
≤ 3.4 ng/ml	46 (48.94)	33 (50.77)	13 (44.83)	
> 3.4 ng/ml	48 (51.06)	32 (49.23)	16 (55.17)	
<b>CA19-9 level</b>				1
≤ 22 ng/ml	78 (82.98)	54 (83.08)	24 (82.76)	
> 22 ng/ml	16 (17.02)	11 (16.92)	5 (17.24)	
<b>Tumor location</b>				0.916
Middle and high	51 (54.26)	36 (55.38)	15 (51.72)	
Low	43 (45.74)	29 (44.62)	14 (48.28)	
Tumor length (cm)	5.25 (4.10, 6.78)	5.60 (4.60, 6.70)	4.70 (3.40, 7.10)	0.271
Tumor thickness(cm)	1.40 (1.10, 1.60)	1.40 (1.10, 1.60)	1.40 (1.10, 1.70)	0.908
<b>R_T staging</b>				0.466
T 1–2	23 (24.47)	14 (21.54)	9 (31.03)	
T 3–4	71 (75.53)	51 (78.46)	20 (68.97)	
<b>R_N staging</b>				1
N0	25 (26.60)	17 (26.15)	8 (27.59)	
N1–2	69 (73.40)	48 (73.85)	21 (72.41)	
<b>P_T staging</b>				0.443
T 1–2	32 (34.04)	20 (30.77)	12 (41.38)	
T 3–4	62 (65.96)	45 (69.23)	17 (58.62)	
<b>R_N staging</b>				0.864
N0	45 (47.87)	32 (49.23)	13 (44.83)	
N1–2	49 (52.13)	33 (50.77)	16 (55.17)	
T2WI score	− 0.79 (− 1.10, − 0.45)	− 0.77 (− 1.09, − 0.39)	− 0.80 (− 1.18, − 0.53)	0.697
DWI score	− 0.69 (− 0.78, − 0.59)	− 0.69 (− 0.78, − 0.60)	− 0.69 (− 0.76, − 0.59)	0.990
CT score	− 0.90 (− 1.57, − 0.41)	− 0.79 (− 1.52, − 0.40)	− 1.14 (− 1.93, − 0.76)	0.099

**Table 1.** Clinical characteristics and radiomics scores for the perineural invasion in the training and testing datasets. Data are presented as mean ± SD, No. (%) or median (25%, 75%). *PNI* perineural invasion, *CEA* carcinoembryonic antigen, *CA19-9* cancer antigen 19–9, *R\_T staging* T staging was diagnosed based on rectal MR and CT images, *R\_N staging* N staging was diagnosed based on rectal MR and CT images, *P\_T staging* T staging was diagnosed by pathology, *P\_N staging* N staging was diagnosed by pathology.

integrated score. The multi-modality radiomics scores were combined by the single modality radiomics score multiplied the corresponding coefficients as follows:

$$\text{MR score} = 2.921 + 2.551 * \text{T2WI score} + 2.756 * \text{DWI score}$$

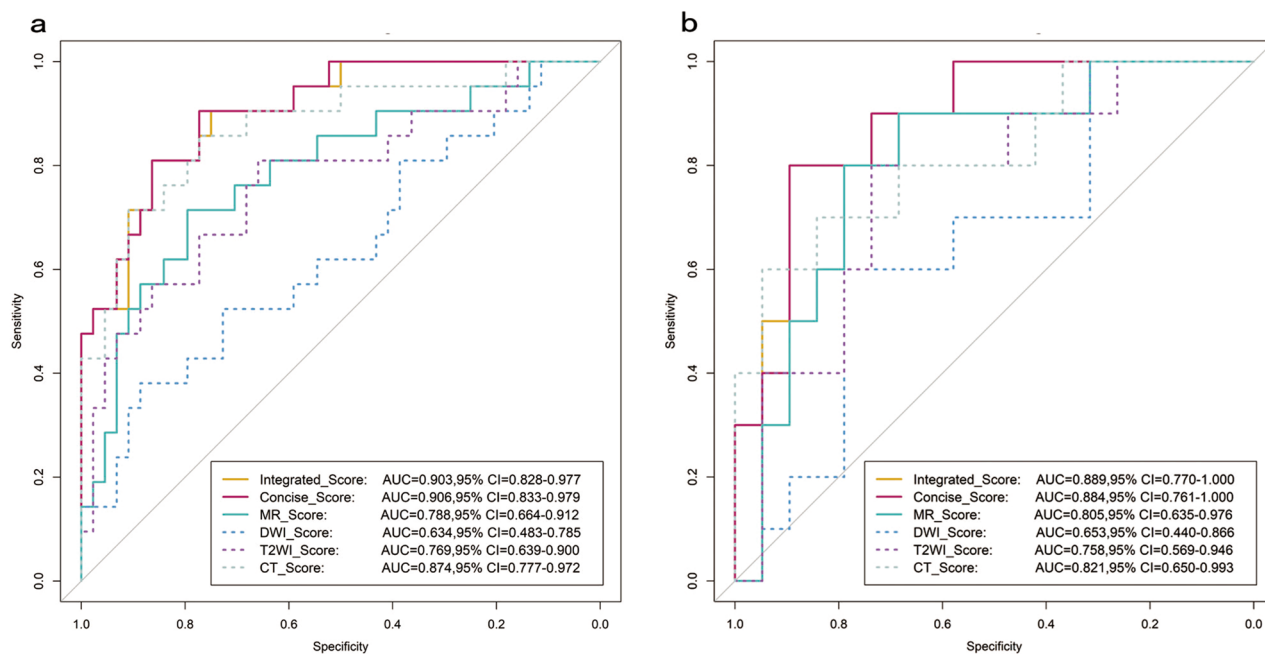
$$\text{Concise score} = 2.078 + 2.400 * \text{T2WI score} + 1.842 * \text{CT score}$$

$$\text{Integrated score} = 1.898 + 2.422 * \text{T2WI score} - 0.298 * \text{DWI score} + 1.856 * \text{CT score}$$

**Performance of the radiomics scores.** Among the single modality models, the CT score showed the best discrimination. The area under the curve (AUC) of CT score was 0.874 (95% CI 0.777–0.972) in the training dataset and 0.821 (95% CI 0.650–0.993) in the testing dataset. The discrimination performance was better than the DWI score in the training and testing datasets (both  $P < 0.05$ ) (Fig. 1). The best discrimination performance of derived multi-modality radiomics models was the concise score which combined the T2WI score and the CT score. The AUC of the concise score was 0.906 (95% CI 0.833–0.979) in the training dataset and 0.884 (95% CI 0.761–1.000) in the testing dataset. The concise score was higher in the PNI-positive group than the PNI-negative group in both training and testing datasets (Supplementary Information 1: Figure S4). AUC, specificity, sensitivity, and accuracy of every single modality radiomics score and derived radiomics scores were showed in

Characteristics	Univariate analysis			Multivariate analysis		
	$\beta$	Odds ratio (95% CI lower-upper)	P value	$\beta$	Odds ratio (95% CI lower-upper)	P value
Age	-0.019	0.982 (0.943-1.022)	0.369	-	-	-
Gender	0.693	2.000 (0.493-8.109)	0.332	-	-	-
CEA (> 3.4 ng/ml)	0.968	2.632 (0.888-7.794)	0.081	-	-	-
CA19-9 (> 22 ng/ml)	-1.609	0.200 (0.051-0.788)	0.021	-2.133	0.118 (0.009-1.514)	0.101
Tumor location	0.395	1.484 (0.514-4.287)	0.466	-	-	-
Tumor length	0.062	1.064 (0.817-1.386)	0.643	-	-	-
Tumor thickness	0.607	1.834 (0.768-4.382)	0.172	-	-	-
R_T staging	1.270	3.562 (0.719-17.662)	0.120	-	-	-
R_N staging	1.030	2.800 (0.706-11.097)	0.143	-	-	-
P_T staging	-1.884	0.152 (0.031-0.735)	0.019	-0.151	0.860 (0.086-8.606)	0.898
P_N staging	-2.007	0.134 (0.039-0.469)	0.002	-2.404	0.09 (0.009-0.950)	0.045
T2WI score	2.662	14.331 (2.955-69.488)	0.001	4.661	105.721 (3.091-3615.790)	0.010
DWI score	4.294	73.234 (0.841-6374.624)	0.060	0.454	1.575 (0.001-2063.812)	0.901
CT score	1.889	6.615 (2.182-20.055)	0.001	1.439	4.218 (1.070-16.620)	0.040

**Table 2.** Univariate and multivariable logistic regression analysis for the prediction of PNI status.  $\beta$  regression coefficient, 95% CI lower-upper, the lower value to the upper value of 95% confidence interval, CEA carcinoembryonic antigen, CA19-9 cancer antigen 19-9, R\_T staging T staging was diagnosed based on rectal MR and CT images, R\_N staging N staging was diagnosed based on rectal MR and CT images, P\_T staging T staging was diagnosed by pathology, P\_N staging N staging was diagnosed by pathology.



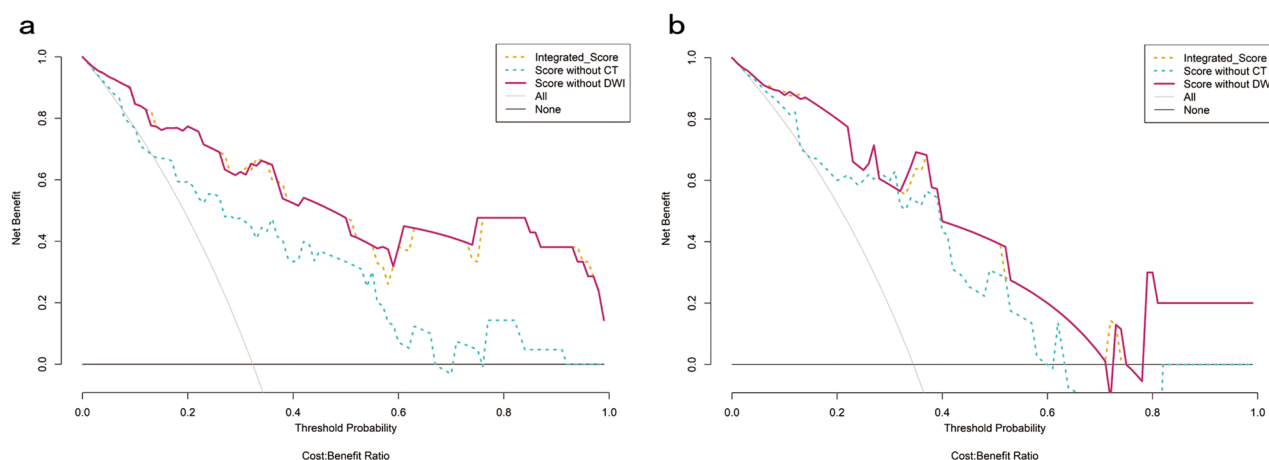
**Figure 1.** Receiver operating characteristic curves analysis in training and testing datasets. Receiver operating characteristic curves analysis in training (a) and testing datasets (b) for each radiomics models to predict perineural invasion status; AUC area under the curve, 95% CI 95% confidence interval.

Table 3, in which the Delong test was used to verify the radiomics scores were robust both in the training and testing datasets.

In most sections of the DCA curves, the concise score is at the top right comparing with the MR score. The concise score and the integrated score have a comparable net benefit, which show several overlaps within the full range of threshold probabilities. The concise score and the integrated score have similar clinical application value, both of which are higher than MR score (Fig. 2). The concise score was appointed as the final model, which had a better discrimination ability and clinical benefit. The calibration curve of the concise score showed good agreement between predicted probability value and the real value in both training and testing datasets ( $P > 0.05$ ) (Fig. 3). In order to ensure the model to be used easily, the concise score consisting of two radiomics scores was presented as the nomogram (Fig. 4).

Model	Cutoff	Training dataset				Testing dataset				P value
		AUC (95% CI lower–upper)	SPE	SEN	ACC	AUC (95% CI lower–upper)	SPE	SEN	ACC	
T2WI score	−0.768	0.769 (0.639–0.900)	0.659	0.810	0.708	0.758 (0.569–0.946)	0.759	0.800	0.759	0.922
DWI score	−0.578	0.634 (0.483–0.785)	0.886	0.789	0.723	0.653 (0.440–0.866)	0.789	0.200	0.586	0.890
CT score	−0.713	0.874 (0.777–0.972)	0.773	0.857	0.800	0.821 (0.650–0.993)	0.947	0.600	0.828	0.598
MR score	−0.555	0.788 (0.664–0.912)	0.795	0.714	0.769	0.805 (0.635–0.976)	0.789	0.600	0.724	0.872
Integrated score	−0.656	0.903 (0.828–0.977)	0.864	0.810	0.846	0.889 (0.770–1.000)	0.895	0.500	0.759	0.856
Concise score	−1.254	0.906 (0.833–0.979)	0.773	0.905	0.815	0.884 (0.761–1.000)	0.895	0.600	0.793	0.768

**Table 3.** The discrimination performance of models in the training and testing datasets. When the model value lesser than the corresponding cut-off value means perineural invasion status is negative; *AUC* the area under receiver operating characteristic (ROC) curves, *95% CI lower–upper* the lower value to the upper value of 95% confidence interval, *SPE* specificity, *SEN* sensitivity, *ACC* accuracy; P value was derived from Delong test between the training and testing datasets.

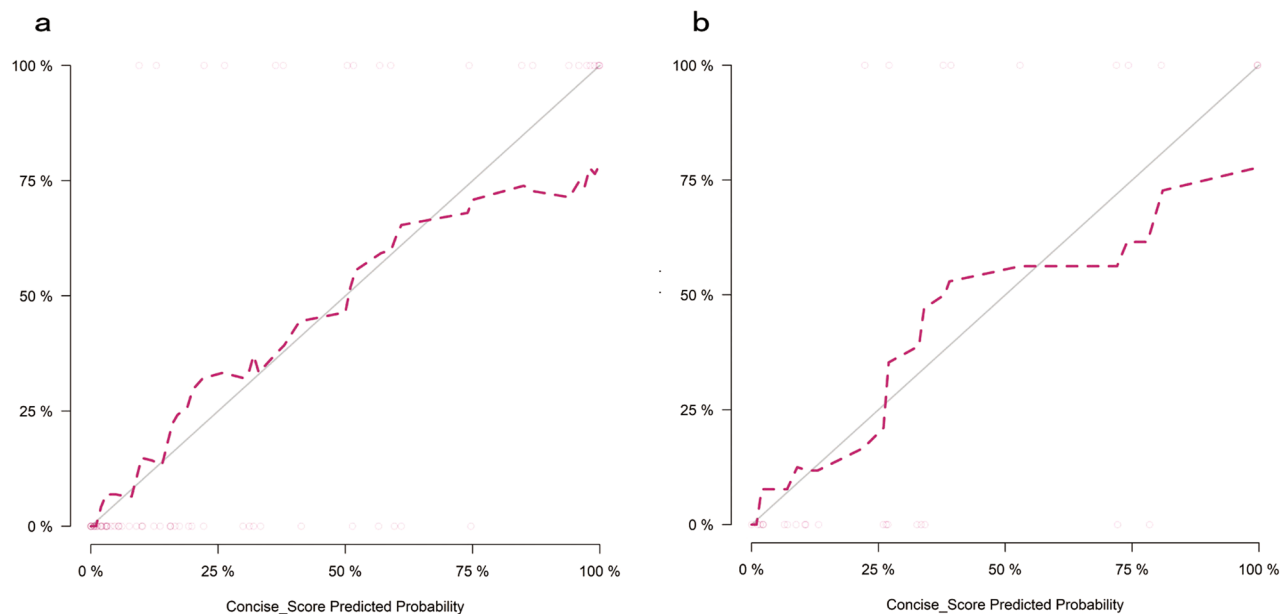


**Figure 2.** Decision curve analysis of radiomics models for predicting perineural invasion status. (a) Training dataset; (b) Testing dataset. The y-axis measures the probability of net benefit. The x-axis measures the threshold probability of perineural invasion positive. The red solid curve represents concise score. The blue and yellow dotted curve represents MR score and integrated score, respectively. The grey curve represents the assumption that all perineural invasion status is positive. The black line represents the assumption that all perineural invasion status is negative.

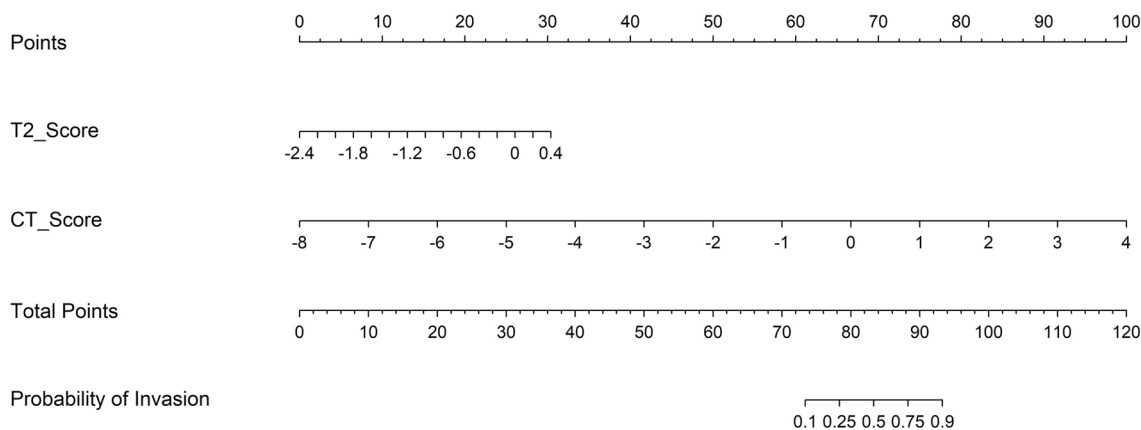
## Discussion

An initial objective of our study was to develop and validate a multi-modality radiomics model for preoperative prediction of PNI status in rectal cancer and the multi-modality radiomics model was converted into a nomogram which is easy to be used in the individual cancer treatment decision system. In the results, the concise score (multi-modality radiomics model integrating T2WI score and CT score) showed excellent discrimination and calibration ability. Nevertheless, the DWI score failed to provide additional predictive value to assess PNI status in rectal cancer.

Among the clinical characteristics, there was no statistically significant difference between the training dataset and the testing dataset. As for the clinical candidate predictor, CA19-9 showed a statistical difference between abnormal tumor marker-level and PNI status in the univariate logistic regression analysis, which is consistent with the published study<sup>30</sup>. Some studies showed the statistical difference between normal level and abnormal levels of CA19-9 and CEA in the PNI-positive and PNI-negative rectal cancer<sup>6,31</sup>. Whereas, there is no sufficient evidence to prove tumor markers could be regarded as an independent risk predictor for PNI status in rectal cancer. It is argued that nonsignificant statistical association risk factor should be kept in the model development. However, with a small sample size, caution must be applied in the candidate factor selection avoiding overfitting. Finally, it is failed to be absorbed in the multi-modality radiomics model. As for the histopathological predictors, N staging diagnosed by pathology is an independent risk predictor for PNI status in rectal cancer in the multivariate logistic regression analysis. Following the present results, previous studies have demonstrated a correlation between PNI status and advanced tumor stage<sup>5,6</sup>. However, the integration of histopathological predictors from surgical tissue may cause sampling bias and tend to high false-negative rate. On the other hand, the multi-modality radiomics model was established before the surgery, the relevant postoperative pathological characteristics were not involved in our model. Our results confirmed that radiomics can use noninvasive CT and MR scans performed in the clinical diagnosis procedure to predict the PNI status in rectal cancer<sup>20,32</sup>, which



**Figure 3.** Calibration curves of the radiomics nomogram in the training and testing datasets. **(a)** Training dataset; **(b)** Testing dataset.  $P > 0.05$  in the Hosmer–Lemeshow test for the training and testing datasets, which indicating an appropriate agreement between the predicted perineural invasion status and actual observed perineural invasion status. The y-axis represents the actual perineural invasion probability. The x-axis represents the predicted perineural invasion probability. The grey solid line represents the best model with the perfect agreement. The red dotted curves have a closer fit to the grey solid line represents the radiomics nomogram has a better calibration.



**Figure 4.** Radiomic nomogram was presented to predict Perineural Invasion status in rectal cancer. The nomogram was built in the training dataset with the T2 score and CT score using a multivariate logistic regression model. The range of T2 score and CT score are respectively  $-2.4$  to  $0.4$  and  $-8$  to  $4$ . The proportional regression coefficients of T2 score and CT score were scaled in the points. Sum the points of T2 score and CT score to get the total points in points scale. The probability of perineural invasion in a rectal cancer patient is the corresponding numbers on the lower probability of Invasion scale.

is corroborate the previous work that taking different aspects marker is the most promising way to facilitates the individualized clinical management<sup>33</sup>. This main finding is based on the central concept that biomedical images reflect underlying pathophysiology via mineable high-dimensional and quantitative image analysis, such as tumor microenvironment. PNI status is regarding as an important component of the tumor microenvironment<sup>17–19</sup>. To construct the radiomics score for each imaging modality of VOIs, an integrated radiomics selection was performed. The inter- and intra-class ICCs were used to select the stable radiomics features. This method could avoid the feature fluctuations caused by segmentation and accomplish the robustness of model development. The mRMR selection and LASSO algorithm were applied to reduce superfluous redundancy radiomics features. This combined method could choose the most relevant radiomics features and eliminate the most superfluous radiomics features to the PNI status. Completing the process of fivefold cross-validation to ensure the stability of the selected radiomics features and summing a panel of selected radiomics features weighted by the regression

coefficients to construct the radiomics score. In the formula of our radiomics scores, the Energy of the Grey level Co-occurrence Matrix (GLCM) were selected regarding as a negative relationship with PNI-positive rectal cancer. It not only supports the published research<sup>27</sup> but also verifies the theory that higher the heterogeneity of tumor represents stronger the invasiveness<sup>34,35</sup>. Comparing with the previous research, besides first-order statistics features and high order statistics features, morphological features were also be extracted from the VOIs of the rectal tumor. Sphericity was chosen as a negative element and a possible explanation for this might be that rectal cancer with the shape of the round may tend to be less aggressive. Most of the selected radiomics features in the modality of T2WI belonged to Gray Level Run-Length Matrix (GLRM), including Long Run Emphasis (LRE) and Run Length Nonuniformity (RLN) which also could be used to assess the homogeneity of the tumor<sup>36</sup>.

In our results, the radiomics scores of T2WI and CT are independent risk predictors to PNI-positive rectal cancer, which demonstrated the significant discrimination in the training dataset, and were approved in the testing dataset. CT score showed the best discrimination and calibration performance of the unimodality radiomics scores (training dataset, AUC = 0.874; testing dataset, AUC = 0.821). Our results confirmed that the published research revealed that integrated CT score and CEA level showed good calibration and discrimination performance (training dataset, AUC = 0.817; testing dataset, AUC = 0.803)<sup>27</sup>. Our research also highlights the role of T2WI score for the prediction of PNI status in rectal cancer. This result is in keeping with the published paper that the imaging biomarkers from MRI can be used to predict the biologic aggressiveness<sup>37</sup>. Following the present results, there was no statistical difference in the discrimination ability between the T2WI score and CT score (Supplementary Information 1: Table S2 and Table S3). Our research displayed a particular application of multi-modality radiomics in the interpretation of rectal cancer and gives a hypothesis that every modality radiomics feature provides different and independent information<sup>28,38</sup>. The concise score showed the best discrimination and calibration performance among the multi-modality radiomics scores that included T2WI score and CT score which were independent risk variables in the multivariate analysis. The discrimination ability of concise score is comparable in the training and testing datasets implies that radiomics score was robust for prediction and could be applied in the testing dataset. This is consistent with the result of clinical decision curve which showed the concise score and the integrated score have similar clinical application value, both of which are higher than MR score. Basing on the need for individual medical treatment the concise score was converted to the nomogram. It could generate an individual predictive probability of PNI-positive for each patient preoperatively. This easy-to-use system could help physicians and patients perform a preoperative prediction of PNI status and individualize the clinical management.

Our study found DWI score didn't have additional predictive value to estimate PNI status in rectal cancer, while DWI is an important preoperative examination in rectal cancer<sup>39,40</sup>. We guessed several possible explanations for this phenomenon. First, lower spatial contrast resolution of DWI contrasting with CT images and T2WI affects the accuracy and robustness in the tumor segmentation. Our results revealed that 54% DWI radiomics features were eliminated because of low inter-/intra-observer reproducibility which is consistent with the published researches, only a small part of radiomics features from DWI remained independent and reproducible<sup>41-43</sup>. Second, portal venous phase CE-CT represents the enhancement of the tumor and therefore extracted quantitative features relating macroscopic blood flow<sup>44</sup>. In our results, twelve CT radiomics features were retained in the feature selection, even though more CT radiomics features were eliminated by ICCs than DWI radiomics features. This reason may be that PNI-positive rectal cancer is more likely to develop perivascular infiltration<sup>3</sup>. Furthermore, artifact-inducing influence can't be neglected, and motion artifact is one of the most important causes limit DWI image quality<sup>45,46</sup>.

Our findings may be somewhat interpreted by some limitations. First, this is a retrospective study that excluded a lot of patients who didn't perform preoperative rectal MR and CE-CT within 2 weeks, and thus there may exist potential selection bias in the study dataset. Second, Patients included in this study from a single-center in China without prognostic information, which restricts the predictive value and generalizability of these findings. In further research, a multi-center validation dataset should be included to evaluate the real benefit in this stratification and verify the robustness of the multi-modality Radiomics model. Third, there is no image pre-processing before the segmentation and radiomics feature extraction. Recent studies only indicated that signal normalization and voxel size resampling could decrease radiomics features' variability. In the future, we will discuss how to improve the robustness of the model through image normalization of different devices and parameters<sup>47,48</sup>.

In conclusion, this research revealed the strong correlation of multi-modality radiomics features to the PNI status in rectal cancer. The multi-modality radiomics score may provide a valuable reference for clinicians in the individual clinical decision system.

## Methods

**Patients.** This retrospective study was approved by the ethics committee of The First Hospital of Jilin University. The informed consent requirement was waived (IRB No. 19K060-001) and all procedures carried out were consistent with the ethical principles such as: International Health Organization good clinical practice (ICH-GCP) (shown in Supplementary Information 2 and 3). We reviewed the medical records of patients who underwent surgical resection for rectal cancer from June 2016 to October 2018. The inclusion criteria were as follow: (I) patients who underwent surgery were diagnosed rectal cancer pathologically; (II) both rectal MRI and contrast-enhanced CT (CE-CT) were performed within 2 weeks before surgery; (III) PNI status was assessed by histopathological examination. The exclusion criteria were as follows: (I) lack of thick slices in the venous phase CT (n = 47); (II) patient underwent anti-tumor treatments before image examination (n = 33); (III) poor image quality (n = 7). The flowchart of the patient enrollment process is shown in Supplementary Information 1: Figure S5. Finally, we enrolled 94 consecutive patients, 70 (74.46%) men, and 24 (25.54%) women, with a median

age of 60.5 years (interquartile range 50–68 years). These patients were randomly divided into training and testing datasets at a ratio of 7:3. Clinicopathologic data of eligible patients, including age, gender, carcinoembryonic antigen (CEA) and cancer antigen 19-9 (CA19-9), were derived from the medical records.

**Histopathology evaluation and perineural invasion status analysis.** Histopathology results were obtained from the pathological information system (PIS). The rectal cancer staging and PNI status were evaluated based on the hematoxylin and eosin slides of the resected specimens according to the 8th AJCC staging system<sup>49</sup>. When the nervous structures and nerve sheaths were involved or more than 1/3 perineural space occupied by rectal tumor cells were defined as PNI-positive rectal cancer, otherwise were defined as PNI-negative rectal cancer.

**MR and CT imaging acquisition.** All rectal MR and CT examinations were performed at our institution. MR imaging was acquired by a 3.0-T scanner (Ingenia, Philips Medical Systems, Netherlands) by using a phase-array coil. The analytic protocols and parameters were as follows: (I) high-resolution axial T2-weighted image (T2WI) was performed using fast recovery fast spin echo, repetition time (TR)=3500 ms, echo time (TE)=100 ms, slice thickness=3.0 mm, gap=0.3 mm, matrix=288×256, echo train length=24, and field of view (FOV)=18 cm×18 cm; (II) Diffusion-weighted imaging (DWI) was performed on b value=1000 s/mm<sup>2</sup>, TR=2800 ms, TE=70 ms, slice thickness=4.0 mm, gap=1.0 mm, matrix=256×256, and FOV=34 cm×34 cm. CT imaging was acquired by one of these two CT scanners (256-detector row, Brilliance, Philips Medical Systems, Netherlands; 64 slices dual source, Definition, Siemens Medical Systems). The protocol parameters of these two scanners were as follows: tube current=250 mA; tube voltage=120 kV; slice thickness=5 mm; rotation time=0.5- or 0.4-s; detector collimation=8×2.5 mm or 64×0.625 mm; matrix=512×512; FOV=350×350 mm. The iodinated contrast (Ultravist 370, Bayer Schering Pharma, Berlin, Germany) was injected at a rate of 2.0–3.0 mL/s after a plain scan. The arterial and portal venous phase imaging were obtained at 25 s and 60 s after injecting. The final dose of iodinated contrast is 1.5 ml/kg.

**Imaging analysis and segmentation.** MR and CT imaging were derived from the picture archiving and communication system (PACS, Agfa, Belgium) in the format of digital imaging and communications in medicine (DICOM). Two radiologists (Reader 1 and reader 2, respectively with 4 years' and 10 years' experience in rectal imaging) reviewed the MR and CT imaging of all included patients. Blinded to the histopathologic results, the radiologists had the primary tumor location and TNM staging according to the 8th AJCC staging system<sup>50</sup>. The thickness and length were respectively measured in the largest axial and sagittal plane. After daily readout sessions, consensus results were reached for all the patients.

Two radiologists (Reader 1 and reader 3, respectively with 3 years' and 10 years' experience in rectal imaging) manually segmented primary tumors using an open-source software package (ITK-SNAP version 3.4.0, <http://www.itksnap.org>) for each tumor slice in the modality of T2WI, DWI and portal venous phase of CE-CT. Either radiologist was blinded to the clinicopathological results. The volumes of interest (VOIs) segmented were defined as follows: (I) high-resolution T2WI images: manually drawn along the contour of the primary tumor lesion, which is slightly high signal; (II) DWI (b=1000): manually drawn along the contour of the primary tumor lesion, which is high signal; (III) Portal venous phase of CE-CT: manually drawn along the primary tumor region enhanced heterogeneously which is could be differentiated from normal bowel structure. All VOIs were segmented on each primary tumor slice containing the circumambient chords and burrs, excluding the fluid in the intestinal lumen.

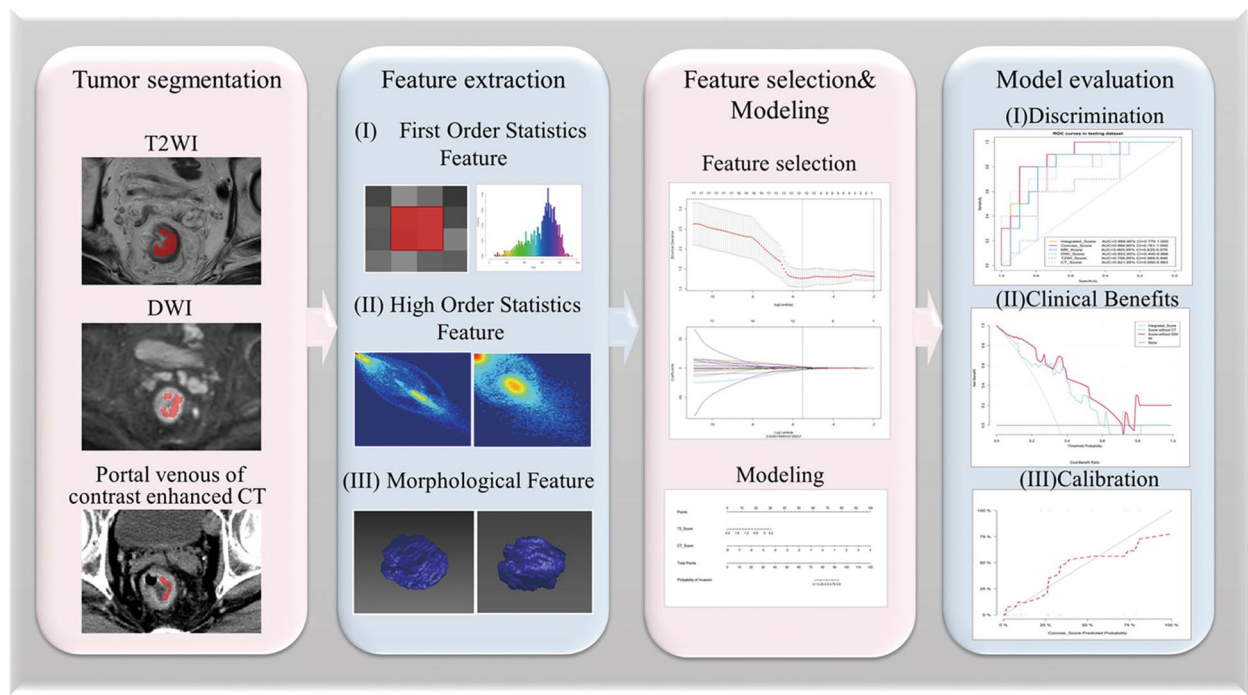
**Radiomics features extraction and selection.** All VOIs in the modality of T2WI, DWI, portal venous phase of CE-CT were loaded to the A.K. software (Artificial Intelligence Kit, AK, version V3.0.0.R, GE Healthcare, China). 396 radiomics features (including 42 first-order statistics features, 345 high order statistics features, 9 Morphological Features) were extracted from every primary lesion VOI. Extraction and standardization of radiomics features was provided in the Supplementary Information 1: Figure S6.

30 VOIs were randomly selected to assess the inter-/intra-observer reproducibility of the Radiomics features extraction from each modality. Reader 1 is one of the two radiologists who performed tumor segmentation repeated this procedure after 2 weeks. Inter- and intra-class correlation coefficients (ICCs) were used to assess the reproducibility between the features which were extracted by reader 1 twice, as well as the features extracted by the two radiologists in the first-time. To ensure the robustness of the radiomics model, the features with ICCs more than 0.70 for both inter-observer and intra-observer were selected for subsequent analysis<sup>50</sup>. The mRMR algorithm and the LASSO algorithm were conducted to select the features with the best correlation to PNI status in the training dataset.

**Radiomics scores building and evaluation.** The radiomics score was built via a linear combination of the selected features weighted by their respective coefficients based on the single modality alone. The radiomics score of T2WI, DWI, portal venous phase of CE-CT respectively present as T2WI score, DWI score, and CT score. Univariate logistic regression analysis was conducted to determine the Odd Ratio of the radiomics scores. Then multivariate logistic regression analysis with a backward stepwise elimination method was used to combine the radiomics scores using the Akaike information criterion (AIC).

Accuracy, specificity, sensitivity, positive predictive value, negative predictive value, and the area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the prediction performance of the models in both training and testing datasets. Decision curve analysis (DCA) was used to estimate the clinical value of the prediction models by contrasting the standardized net benefit at different risk threshold probabilities.





**Figure 5.** Flowchart describing the workflow of this study to construct and validate the radiomics models.

The final model was selected which had a better discrimination ability and clinical benefit. The calibration curve depicted the consistency between the predicted probability value and the real value in the final model. A radiomics nomogram from the final model to ensure it is easy to use. All procedures building and validating radiomics models were shown in Fig. 5.

**Statistical analysis.** All the statistical analyses were performed with R software (version 3.5.1, <http://www.r-project.org>). Age regarding as a continuous variable was expressed as mean  $\pm$  standard deviation (SD). Sex was considered as a binary variable. CEA and CA19-9 were handled as binary variables cause the thresholds were set at 3.4 ng/ml and 22 ng/ml according to the clinical consensuses. The radiomics scores were expressed as a continuous variable and expressed as median (25% quantile, 75% quantile). The *t* test or Mann–Whitney *U* tests was performed to compare the continuous variables as appropriate. Fisher’s exact test or chi-squared test was applied to assess the distribution differences between the training and testing datasets in binary variables. Univariate and multivariate logistic regression analyses were respectively used to select the significant predictors from relevant clinicopathologic characteristics and radiomics scores to build prediction models. Delong tests were performed to assess the differences in AUC values between different radiomics models. Hosmer–Lemeshow goodness-of-fit test was applied in the calibration plot via bootstrapping resamples. In the statistical analysis  $P < 0.05$  presented significant statistically. All packages used were listed in the supplementary Information 1: Table S1.

**Ethical statement.** All the authors are responsible for guaranteeing the integrity and accuracy of the data collection and analysis process. This retrospective study was approved by the first hospital institutional ethics committee of Jilin University (IRB No. 19K060-001).

### Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Received: 18 July 2020; Accepted: 14 April 2021

Published online: 03 May 2021

### References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA-Cancer J. Clin.* **70**, 7–30. <https://doi.org/10.3322/caac.21590> (2020).
2. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* <https://doi.org/10.3322/caac.21601> (2020).
3. Liebig, C., Ayala, G., Wilks, J. A., Berger, D. H. & Albo, D. Perineural invasion in cancer: A review of the literature. *Cancer* **115**, 3379–3391. <https://doi.org/10.1002/cncr.24396> (2009).
4. Knijn, N., Mogk, S. C., Teerenstra, S., Simmer, F. & Nagtegaal, I. D. Perineural invasion is a strong prognostic factor in colorectal cancer: A systematic review. *Am. J. Surg. Pathol.* **40**, 103–112. <https://doi.org/10.1097/PAS.0000000000000518> (2016).

5. Poeschl, E. M. *et al.* Perineural invasion: Correlation with aggressive phenotype and independent prognostic variable in both colon and rectum cancer. *J. Clin. Oncol.* **28**, e358–e362. <https://doi.org/10.1200/JCO.2009.27.3581> (2010).
6. Al-Sukhni, E. *et al.* Lymphovascular and perineural invasion are associated with poor prognostic features and outcomes in colorectal cancer: A retrospective cohort study. *Int. J. Surg.* **37**, 42–49. <https://doi.org/10.1016/j.ijsu.2016.08.528> (2017).
7. Benson, A. B. *et al.* Rectal cancer, version 2.2018, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Cancer Netw.* **16**, 874–901. <https://doi.org/10.6004/jnccn.2018.0061> (2018).
8. Liebig, C. *et al.* Perineural invasion is an independent predictor of outcome in colorectal cancer. *J. Clin. Oncol.* **27**, 5131–5137. <https://doi.org/10.1200/JCO.2009.22.4949> (2009).
9. Cienfuegos, J. A. *et al.* Impact of perineural and lymphovascular invasion on oncological outcomes in rectal cancer treated with neoadjuvant chemoradiotherapy and surgery. *Ann. Surg. Oncol.* **22**, 916–923. <https://doi.org/10.1245/s10434-014-4051-5> (2014).
10. Song, J. H. *et al.* Significance of perineural and lymphovascular invasion in locally advanced rectal cancer treated by preoperative chemoradiotherapy and radical surgery: Can perineural invasion be an indication of adjuvant chemotherapy?. *Radiother. Oncol.* **133**, 125–131. <https://doi.org/10.1016/j.radonc.2019.01.002> (2019).
11. Nikberg, M. *et al.* Lymphovascular and perineural invasion in stage II rectal cancer: A report from the Swedish colorectal cancer registry. *Acta Oncol.* **55**, 1418–1424. <https://doi.org/10.1080/0284186X.2016.1230274> (2016).
12. Chablani, P. *et al.* Perineural invasion predicts for distant metastasis in locally advanced rectal cancer treated with neoadjuvant chemoradiation and surgery. *Am. J. Clin. Oncol.* **40**, 561–568. <https://doi.org/10.1097/COC.0000000000000214> (2017).
13. Giger, O. T., Comtesse, S. C. M., Lugli, A., Zlobec, I. & Kurrer, M. O. Intra-tumoral budding in preoperative biopsy specimens predicts lymph node and distant metastasis in patients with colorectal cancer. *Mod. Pathol.* **25**, 1048–1053. <https://doi.org/10.1038/modpathol.2012.56> (2012).
14. Garcia-Figueiras, R. *et al.* Advanced imaging techniques in evaluation of colorectal cancer. *Radiographics* **38**, 740–765. <https://doi.org/10.1148/rg.2018170044> (2018).
15. Glynne-Jones, R. *et al.* Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **28**, iv22–iv40. <https://doi.org/10.1093/annonc/mdx224> (2017).
16. Beets-Tan, R. G. H. *et al.* Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur. Radiol.* **28**, 1465–1475. <https://doi.org/10.1007/s00330-017-5026-2> (2018).
17. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **278**, 563–577. <https://doi.org/10.1148/radiol.2015151169> (2016).
18. Lambin, P. *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446. <https://doi.org/10.1016/j.ejca.2011.11.036> (2012).
19. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006. <https://doi.org/10.1038/ncomms5006> (2014).
20. Lambin, P. *et al.* Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762. <https://doi.org/10.1038/nrclinonc.2017.141> (2017).
21. Limkin, E. J. *et al.* Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann. Oncol.* **28**, 1191–1206. <https://doi.org/10.1093/annonc/mdx034> (2017).
22. Liu, Z. *et al.* Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin. Cancer Res.* **23**, 7253–7262. <https://doi.org/10.1158/1078-0432.CCR-17-1038> (2017).
23. Chen, L.-D. *et al.* Multiparametric radiomics improve prediction of lymph node metastasis of rectal cancer compared with conventional radiomics. *Life Sci.* **208**, 55–63. <https://doi.org/10.1016/j.lfs.2018.07.007> (2018).
24. Golia Pernicka, J. S. *et al.* Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom. Radiol. (New York)* **44**, 3755–3763. <https://doi.org/10.1007/s00261-019-02117-w> (2019).
25. Horvat, N. *et al.* MR imaging of rectal cancer: Radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology* **287**, 833–843. <https://doi.org/10.1148/radiol.2018172300> (2018).
26. Li, Y. *et al.* Radiomics signature for the preoperative assessment of stage in advanced colon cancer. *Am. J. Cancer Res.* **9**, 1429–1438 (2019).
27. Huang, Y. *et al.* Individualized prediction of perineural invasion in colorectal cancer: Development and validation of a radiomics prediction model. *Chin. J. Cancer Res.* **30**, 40–50. <https://doi.org/10.21147/j.issn.1000-9604.2018.01.05> (2018).
28. Baltrusaitis, T., Ahuja, C. & Morency, L. P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 423–443. <https://doi.org/10.1109/tpami.2018.2798607> (2019).
29. Li, Y. F., Wu, F. X. & Ngom, A. A review on machine learning principles for multi-view biological data integration. *Brief. Bioinform.* **19**, 325–340. <https://doi.org/10.1093/bib/bbw113> (2018).
30. Huh, J. W., Lee, J. H., Kim, H. R. & Kim, Y. J. Prognostic significance of lymphovascular or perineural invasion in patients with locally advanced colorectal cancer. *Am. J. Surg.* **206**, 758–763. <https://doi.org/10.1016/j.amjsurg.2013.02.010> (2013).
31. Park, I. J., Choi, G. S., Lim, K. H., Kang, B. M. & Jun, S. H. Serum carcinoembryonic antigen monitoring after curative resection for colorectal cancer: Clinical significance of the preoperative level. *Ann. Surg. Oncol.* **16**, 3087–3093. <https://doi.org/10.1245/s10434-009-0625-z> (2009).
32. Bi, W. L. *et al.* Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J. Clin.* **69**, 127–157. <https://doi.org/10.3322/caac.21552> (2019).
33. Birkhahn, M., Mitra, A. P. & Cote, R. J. Molecular markers for bladder cancer: The road to a multimarker approach. *Expert Rev. Anticancer Ther.* **7**, 1717–1727. <https://doi.org/10.1586/14737140.7.12.1717> (2007).
34. Hardiman, K. M. *et al.* Intra-tumor genetic heterogeneity in rectal cancer. *Lab. Invest.* **96**, 4–15. <https://doi.org/10.1038/labinvest.2015.131> (2016).
35. Greenbaum, A. *et al.* Tumor heterogeneity as a predictor of response to neoadjuvant chemotherapy in locally advanced rectal cancer. *Clin. Colorectal Cancer* **18**, 102–109. <https://doi.org/10.1016/j.clcc.2019.02.003> (2019).
36. Sagaert, X., Vanstapel, A. & Verbeek, S. Tumor heterogeneity in colorectal cancer: What do we know so far?. *Pathobiology* **85**, 72–84. <https://doi.org/10.1159/000486721> (2018).
37. Yeo, D. M. *et al.* Correlation of dynamic contrast-enhanced MRI perfusion parameters with angiogenesis and biologic aggressiveness of rectal cancer: Preliminary results. *J. Magn. Reson. Imaging* **41**, 474–480. <https://doi.org/10.1002/jmri.24541> (2015).
38. Zhang, Y. *et al.* A novel multimodal radiomics model for preoperative prediction of lymphovascular invasion in rectal cancer. *Front. Oncol.* **10**, 457. <https://doi.org/10.3389/fonc.2020.00457> (2020).
39. Gollub, M. J. *et al.* Value of adding dynamic contrast-enhanced MRI visual assessment to conventional MRI and clinical assessment in the diagnosis of complete tumour response to chemoradiotherapy for rectal cancer. *Eur. Radiol.* **29**, 1104–1113. <https://doi.org/10.1007/s00330-018-5719-1> (2019).
40. Tang, X., Luo, Y., Zhang, S. P., Xia, L. G. & Gong, J. S. Local staging of rectal cancer using fused high resolution diffusion weighted imaging and modified MR rectography. *Quant. Imaging Med. Surg.* **9**, 1592–1596. <https://doi.org/10.21037/qims.2019.08.02> (2019).
41. Schwier, M. *et al.* Repeatability of multiparametric prostate MRI radiomics features. *Sci. Rep.* **9**, 16. <https://doi.org/10.1038/s41598-019-45766-z> (2019).
42. Mackin, D. *et al.* Effect of tube current on computed tomography radiomic features. *Sci. Rep.* **8**, 10. <https://doi.org/10.1038/s41598-018-20713-6> (2018).

43. Scalco, E. *et al.* T2w-MRI signal normalization affects radiomics features reproducibility. *Med. Phys.* **47**, 1680–1691. <https://doi.org/10.1002/mp.14038> (2020).
44. Hayano, K. *et al.* Quantitative measurement of blood flow using perfusion CT for assessing clinicopathologic features and prognosis in patients with rectal cancer. *Dis. Colon Rectum* **52**, 1624–1629. <https://doi.org/10.1007/DCR.0b013e3181afbd79> (2009).
45. Xia, C. C. *et al.* Readout-segmented echo-planar imaging improves the image quality of diffusion-weighted MR imaging in rectal cancer: Comparison with single-shot echo-planar diffusion-weighted sequences. *Eur. J. Radiol.* **85**, 1818–1823. <https://doi.org/10.1016/j.ejrad.2016.08.008> (2016).
46. Le Bihan, D., Poupon, C., Amadon, A. & Lethimonnier, F. Artifacts and pitfalls in diffusion MRI. *J. Magn. Reson. Imaging* **24**, 478–488. <https://doi.org/10.1002/jmri.20683> (2006).
47. Shafiq-ul-Hassan, M. *et al.* Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* **44**, 1050–1062. <https://doi.org/10.1002/mp.12123> (2017).
48. Traverso, A. *et al.* Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. *Radiother. Oncol.* **143**, 88–94. <https://doi.org/10.1016/j.radonc.2019.08.008> (2020).
49. Weiser, M. R. AJCC 8th edition: Colorectal cancer. *Ann. Surg. Oncol.* **25**, 1454–1455. <https://doi.org/10.1245/s10434-018-6462-1> (2018).
50. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012> (2016).

## Acknowledgements

This research was supported by the Foundation of Department of Finance, Jilin Province, China (Grant number:2018SCZWSZX-026, JLSCZD2019-062); Science and Technology Innovation Talents Cultivation Program of the Science and Technology Department, Jilin Province, China (Grant number: 20180519008JH); Key Laboratory of Jilin Province, China (Grant number: 20200601003JC); and the Foundation of Department of Finance, Jilin Provincial, China (project name: Establishment of a standardized database for colorectal cancer and exploration of new diagnosis and treatment model based on big data analysis).

## Author contributions

Conception and design: H.Z., Y.F., Y.G.<sup>1</sup>; Administrative support: H.Z., Y.F.; Provision of study materials or patients: H.Z., Y.F., Y.G.<sup>1</sup>, Q.W.; Collection and assembly of data: Y.G.<sup>1</sup>, Y.F., Y.Z.; Data analysis and interpretation: Y.G.<sup>1</sup>, Y.G.<sup>2</sup>, Y.Z.; Manuscript writing: Y.G.<sup>1</sup>. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-88831-2>.

**Correspondence** and requests for materials should be addressed to Y.F. or H.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021