**Name of Journal:** *World Journal of Gastroenterology*

**Manuscript NO:** 56842

**Manuscript Type:** ORIGINAL ARTICLE

***Retrospective Study***

**Radiomics of rectal cancer for predicting distant metastasis and overall survival**

Li M *et al.* Radiomics for prognosis evaluation

Mou Li, Yu-Zhou Zhu, Yong-Chang Zhang, Yu-Feng Yue, Hao-Peng Yu, Bin Song

**Mou Li, Yu-Feng Yue, Hao-Peng Yu, Bin Song,** Department of Radiology, West China Hospital of Sichuan University, Chengdu 610041, Sichuan Province, China

**Yu-Zhou Zhu,** Department of Gastrointestinal Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan Province, China

**Yong-Chang Zhang,** Department of Radiology, Chengdu Seventh People’s Hospital, Chengdu 610213, Sichuan Province, China

**Author contributions:** All authors helped to perform the research; Song B and Zhu YZ designed the research; Li M and Zhang YC performed the research; Li M, Yue YF, Song B, and Yu HP analyzed the data; Li M and Song B wrote the paper.

**Corresponding author: Bin Song, MD, PhD, Chief Doctor, Professor,** Department of Radiology, West China Hospital of Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, Sichuan Province, China. songlab\_radiology@163.com

**Received:** May 27, 2020

**Revised:** June 16, 2020

**Accepted:** August 13, 2020

**Published online:** September 7, 2020

**Abstract**

BACKGROUND

Rectal cancer (RC) patient stratification by different factors may yield variable results. Therefore, more efficient prognostic biomarkers are needed for improved risk stratification, personalized treatment, and prognostication of RC patients.

AIM

To build a novel model for predicting the presence of distant metastases and 3-year overall survival (OS) in RC patients.

METHODS

This was a retrospective analysis of 148 patients (76 males and 72 females) with RC treated with curative resection, without neoadjuvant or postoperative chemoradiotherapy, between October 2012 and December 2015. These patients were allocated to a training or validation set, with a ratio of 7:3. Radiomic features were extracted from portal venous phase computed tomography (CT) images of RC. The least absolute shrinkage and selection operator regression analysis was used for feature selection. Multivariate logistic regression analysis was used to develop the radiomics signature (Rad-score) and the clinicoradiologic risk model (the combined model). Receiver operating characteristic curves were constructed to evaluate the diagnostic performance of the models for predicting distant metastasis of RC. The association of the combined model with 3-year OS was investigated by Kaplan-Meier survival analysis.

RESULTS

A total of 51 (34.5%) patients had distant metastases, while 26 (17.6%) patients died, and 122 (82.4%) patients lived at least 3 years post-surgery. The values of both the Rad-score (consisted of three selected features) and the combined model were significantly different between the distant metastasis group and the non-metastasis group (0.46 ± 0.21 *vs* 0.32 ± 0.24 for the Rad-score, and 0.60 ± 0.23 *vs* 0.28 ± 0.26 for the combined model; *P* < 0.001 for both models). Predictors contained in the combined model included the Rad-score, pathological N-stage, and T-stage. The addition of histologic grade to the model failed to show incremental prognostic value. The combined model showed good discrimination, with areas under the curve of 0.842 and 0.802 for the training set and validation set, respectively. For the survival analysis, the combined model was associated with an improved OS in the whole cohort and the respective subgroups.

CONCLUSION

This study presents a clinicoradiologic risk model, visualized in a nomogram, that can be used to facilitate individualized prediction of distant metastasis and 3-year OS in patients with RC.

**Key words:** Radiomics; Rectal cancer; Overall survival; Distant metastasis; Computed tomography

**Citation:** Li M, Zhu YZ, Zhang YC, Yue YF, Yu HP, Song B. Radiomics of rectal cancer for predicting distant metastasis and overall survival. *World J Gastroenterol* 2020; 26(33): 5008-5021

**URL:** https://www.wjgnet.com/1007-9327/full/v26/i33/5008.htm

**DOI:** https://dx.doi.org/10.3748/wjg.v26.i33.5008

**Core tip:** We developed and validated a combined model that incorporated radiomic features and clinical factors. This model showed excellent potential for predicting distant metastasis of rectal cancer (RC) within 3 years after surgery. We used this model to stratify the patients with RC into low-risk and high-risk groups for the survival analysis. Overall survival rates between the low-risk and high-risk groups were significantly different. This model may aid in individualized prediction of distant metastasis and 3-year overall survival in patients with RC.

**INTRODUCTION**

More than 1.8 million cases of colorectal cancer were diagnosed in 2018 worldwide, making it the third most prevalent malignancy around the world. In terms of morbidity, colorectal cancer accounts for approximately 9.2% of cancer-related deaths worldwide, similar to that of stomach and liver cancer[1]. In up to 70% of patients with rectal cancer (RC), surgical removal of the primary tumor is successful. However, local recurrence and distant metastases are commonly detected in approximately 30% of RC patients, often within 3 years after surgery[2-4]. The overall prognosis becomes poor once distant metastases have developed, thus demonstrating the importance of prompt diagnosis and treatment of RC[5]. Some patients may be at a higher risk of developing adverse outcomes post-surgery. In these patients, alternative and adjunctive therapies, such as chemotherapy, radiotherapy, or other targeted therapies, may be needed to minimize the risk of developing distant metastases[6]. Hence, improved patient outcomes may be feasible by identifying unfavorable prognostic characteristics that could hinder the overall survival of patients. In return, personalized treatment strategies could be implemented to achieved improved outcomes in patients with RC[7]. Recently, the American Society of Clinical Oncology guidelines outlined the need for newer surveillance technologies to better characterize and detect rectal tumors during the early stages of the disease, which could aid in risk stratification and subsequent surveillance[8].

Currently, there is no universal list of clinicoradiologic prognostic factors used for the detection of RC patients more likely to benefit from treatment. While the TNM classification system is widely utilized for staging cancer in the clinic, it has certain pitfalls that limit its clinical utility in RC[9]. While TNM considers the degree of cancer invasion, the involvement of surrounding lymph nodes, and metastatic spread of the tumor[10], it does not consider the importance of spatial heterogeneity. Spatial heterogeneity is an important characteristic of RC, as it indicates high cell density, hemorrhage, and necrosis[11].

Radiomics can provide a comprehensive overview of intratumor heterogeneity by extracting multiple quantitative features from medical images. A wide array of parameters can be utilized to assess and quantify the degree of heterogeneity at relevant scales, such as skewness, entropy, kurtosis, and uniformity. Radiomic analysis has been successfully utilized to predict survival outcomes in other diseases[12-14]. When integrated with the clinicopathologic features, the radiomics signature is superior to that of a single biomarker, in terms of prognostic value[15-17]. Hence, in the current study, the aim was to compare the predictive abilities, in terms of distant metastases and 3-year OS, of a radiomics signature and clinicoradiologic risk model in patients with RC.

**MATERIALS AND METHODS**

***Patients***

This study was approved by the medical ethics committee of our institution (No. 2019-1159; date: December 26, 2019). Patient approval or informed consent for the review of medical images was waived by the committee due to the retrospective nature of the study.

The primary cohort of this study comprised an evaluation of the institutional database for medical records from October 2012 to December 2015 to identify patients with histologically-confirmed RC, who underwent curative surgery alone. A total of 148 patients (76 males and 72 females; mean age: 59.7 ± 11.7 years; age range: 27-94 years) were enrolled in our study according to the following inclusion criteria: (1) Patients with histologically-confirmed rectal adenocarcinoma who were treated with curative resection alone; (2) preoperative enhanced computed tomography (CT) was performed within 1 mo before resection; (3) follow-up was conducted for at least 3 years; and (4) T-stage, N-stage, and tumor grade were confirmed by pathology. The exclusion criteria were as follows: (1) Patients who received neoadjuvant or postoperative chemoradiotherapy; (2) patients who presented distant metastasis at the time of diagnosis; (3) patients who failed to undergo follow-up; (4) lack of preoperative CT images; and (5) poor CT image quality. The patients were allocated to a training or validation set at a ratio of 7:3, based on the scanning date. The early data before the 70th percentile scanning date were allocated to the training set, while the other data were allocated to the validation set. The patient recruitment pathway is shown in Figure 1.

The baseline clinical characteristics and pathological data of each patient, including age, sex, tumor size on CT, pathological TNM stage, histological grade, antigen Ki-67, circumferential resection margin, number of distant metastasis, and follow-up time were all derived from medical records, as shown in Table 1.

***CT examinations***

Enhanced CT was performed using a 128-channel multidetector CT scanner (Simens, SOMATOM Definition AS+) with the following scanning parameters: Tube voltage of 120 kV, tube current of 200-210 mA, and slice thickness of 2.0 mm. With the mass injection (dose: 1.2 mL/kg; injection rate: 3 mL/s) of iodine contrast agent (300 mg/mL), the portal venous phase images (30 s after the trigger) were obtained.

***Reference standard for pathology***

All 148 patients had histopathologically-verified RC, which was diagnosed with resected surgical specimens. The pathological confirmatory report was acquired from electronic medical records. Samples were processed using standard procedures, fixed in formalin, embedded in paraffin, and stained with hematoxylin and eosin (HE).

***Follow-up***

All patients were followed for at least 3 years after surgery (3-mo intervals in the first year; every 6 mo in the following 2 years). The minimum postoperative follow-up period was 1 mo for metastasis (three cases had liver, lung, and bone metastases, respectively), and 11 mo for OS.

The endpoint of this study was 3-year OS, defined as the time between the surgery and the date of the patient’s death. Local recurrence was defined as recurrence in the pelvis, and distant metastasis as recurrence at sites other than the pelvis. All distant metastatic cases were diagnosed by a multidisciplinary team based on clinical examinations, serum carcinoembryonic antigen levels, chest and abdominopelvic CT, abdominopelvic magnetic resonance imaging, endoscopy, and biopsy. Follow-up information was recorded in the database.

***Texture feature extraction***

A radiologist with 7 years of experience in RC imaging used the ITK-SNAP software (open source, www.itk-snap.org) for the three-dimensional (3D) manual segmentation of the primary tumor, as shown in Figure 2. The region of interst was carefully placed in an area that avoided the intestinal lumen, calcification, blood vessels, and necrosis. In all, 396 quantitative radiomic features of three types (first-order, second-order, and higher-order) were calculated and extracted automatically using the in-house Artificial Intelligence Kit software (GE Healthcare). First-order features are calculated from the pixel intensity histogram to quantify tumor intensity characteristics. Second-order features based on the co-occurrence matrix or run-length matrix account for the location of the pixels and analyze texture in a specific direction. Higher-order features, such as contrast, compare differences and relationships between multiple pixels.

***Model building and evaluation***

Before selecting features, the redundant features needed to be eliminated. When the Pearson correlation coefficient of any two features was higher than 0.6, one of them was selected at random, and the rest could be treated in the same manner. Then, the least absolute shrinkage and selection operator (LASSO) regression analysis was performed to identify the most useful prognostic features. The LASSO regression is a variable selection method for linear regression models. The LASSO does this by imposing a constraint on the model parameter (λ) that causes regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero are excluded from the model. Variables with non-zero regression coefficients are selected by the LASSO. After that, z-score transformation was used to standardize all selected features. The CT images of 30 patients were segmented twice by the same radiologist to calculate intraclass correlation coefficients for assessing the stability of the features. Radiomic features with an intraclass correlation coefficient > 0.75 were included in the analysis. Each radiomic feature was assessed by univariate logistic regression, and statistical significance was assumed at a confidence level of 0.2 to avoid missing important features. Features with *P* < 0.2 in univariate logistic regression analysis were entered into the multivariate analysis. A nomogram was generated for model visualization. Receiver operating characteristic curves for each model were constructed to evaluate the performance for predicting distant metastasis of RC. The association of the predicting models with 3-year OS was investigated by Kaplan-Meier survival analysis.

***Statistical analysis***

All statistical analyses were performed using R software (version 3.6.1, R Foundation for Statistical Computing, Vienna, Austria), Statistic Package for Social Science version 21 (Statistic Package for Social Science Inc., Chicago, IL, United States), Stata 15.0 (Stata Corp LLC, Lakeway Drive College Station, United States), and Medcalc 15.2.2 (Medcalc Software, Ostend, Belgium). Differences between the training and validation sets were assessed by the chi-square test, Mann-Whitney test, and *t*-test. Survival curves were compared by the log-rank test. *P* < 0.05 was considered statistically significant.

**RESULTS**

A total of 148 patients were enrolled in our study, of whom 51 had distant metastases, 26 died, and 122 survived to the date of the last follow-up visit at 3 years post-surgery (Table 2).

***Feature selection and model building***

A total of 17 radiomic features remained after removing the redundant features. Two features, “MaxIntensity” and “RelativeDeviation”, were selected by the LASSO method from the remaining features (Figure 3). To avoid missing important features, another feature “Inertia\_AllDirection\_offset7\_SD” was selected by the univariate logistic regression (*P* = 0.192). These three features were presented in this calculation formula: Rad-score = 1/{1 + EXP [-(-0.649 - 0.786 × MaxIntensity - 0.868 × Inertia\_AllDirection\_offset7\_SD - 2.743 × RelativeDeviation)]}. The radiomics signature (Rad-score), which consisted of three selected features, was statistically different between the metastasis and non-metastasis groups (Rad-score = 0.46 and 0.32, respectively; *P* < 0.001), as shown in Table 2.

For clinical features, pathological T-stage and N-stage were selected by the univariate logistic regression analysis (*P* < 0.001 and *P* = 0.002, respectively). The clinicoradiologic risk model was built by the multivariate logistic regression analysis, using the formula 1/{1 + EXP [-(-6.199 + 1.368 × T-stage + 0.738 × N-stage + 4.673 × Rad-score)]}. The results of the combined model were significantly different between the metastasis and non-metastasis groups by 3 years after surgery (*P* < 0.001), as shown in Table 2. The Hosmer-Lemeshow test yielded no statistical significance (*P* = 0.280 and 0.151 for the Rad-score and combined model, respectively), which suggested that there was no evidence of lack of fit.

A nomogram to predict the distant metastasis of RC was generated for the combined model visualization (Figure 4). The nomogram could be used by locating the score for each variable on the corresponding axis, adding the points together for all of the variables, and drawing a line from the total number of points to the risk axis, which allows for the determination of distant metastasis risk. Higher total scores were associated with a greater risk of distant metastasis. The contributions of these variables to the Rad-score and the combined model were measured by the value of the standardized logistic regression coefficient (Figure 5). The contribution of “MaxIntensity” to the Rad-score, and that of “T-stage” to the combined model were greater than those of the others.

***Classification results***

In the case of only considering the Rad-score, the resulting area under the curve (AUC) was 0.709 (95%CI: 0.612-0.795). Improved prediction of distant metastases could be achieved by combining the radiomic features from preoperative CT images with the clinical features. The AUCs of the clinical model (T-stage combined with N-stage) and the combined model were 0.782 (95%CI: 0.689-0.857) and 0.842 (95%CI: 0.757-0.906), respectively (Table 3 and Figure 6). There was a significant difference in AUCs between the Rad-score and combined model in the training cohort (AUC = 0.709 *vs* 0.842, *P* = 0.005), which was confirmed using the validation cohort (AUC = 0.687 *vs* 0.802, *P* = 0.020). We found that the clinical variables had higher classification contributions than Rad-score to build the combined model. This finding was consistent with the higher standardized logistic regression coefficient of T-stage than that of the Rad-score (Figure 5). Moreover, the AUC value of the clinical model was also higher than that of the Rad-score. Even so, the Rad-score greatly helped to increase the AUC value from 0.782 to 0.842 (Table 3 and Figure 6). In the subgroups of the receiver operating characteristic analysis, the AUC of the combined model decreased in the 1-year (distant metastasis occurred in the first year after surgery) and 2-year (distant metastasis occurred within 2 years after surgery) subgroups (Table 3).

***Assessment of the combined model for OS***

Using the clinicoradiologic risk model, stratified analyses were performed for the whole set, training set, and validation set to evaluate the association with OS. As shown in Figure 7, the grouping results of the combined model were significantly associated with OS in the whole, training, and validation groups (*P* < 0.0001, *P* = 0.0001, and *P* = 0.0137, respectively). Although there was a lack of statistical significance for stage Ⅲ in the stratified subgroup analysis, according to the overall pathological stage, the low-risk group displayed a longer OS than the high-risk group (*P* < 0.001, *P* = 0.0194, and *P* = 0.1401 for stages Ⅰ, Ⅱ, and Ⅲ, respectively), which was significant in terms of individualized treatment (Figure 8)[18].

**DISCUSSION**

In this study, we developed and validated a clinicoradiologic risk model that showed potential for predicting distant metastasis of RC within 3 years after surgery, which had a better prognostic performance than Rad-score (*P* = 0.005 and 0.020 for the training cohort and validation cohort, respectively). The combined model was used to stratify patients into low-risk and high-risk groups for the analysis of 3-year OS. The results showed that OS rates between low- and high-risk groups were significantly different in the training cohort, which was verified in the validation cohort.

Compared with traditional image explanation, which is qualitative or subjective, radiomic analyses permit high-throughput extraction of radiomic features that can quantify differences between tissues invisible to human eyes. Recently, the use of radiomics has appeared as a potential technique for constructing decision-support models based on high-throughput quantificational characters extracted from medical images. Radiomics-based prognosis prediction models have been reported for advanced nasopharyngeal carcinoma[19], early-stage non-small cell lung cancer[20], and RC[12-14]. In their pioneering retrospective studies about the prognosis of RC[12-14], robust models and strong independent prognostic factors have been developed for the prediction of OS in patients with RC. Our results were consistent with previous studies, suggesting that radiomics could help predict the prognosis of patients with RC. The building methods between our study and those of previous studies were similar (machine learning), and all these studies lacked external validation. However, there were still some differences that need to be explained. For the follow-up time, Wang *et al*[13] followed the patients for 5 years, and Lovinfosse *et al*[12] for 4 years, which were longer than the time in the current study. These authors[12-14] focused on predicting the prognosis in locally advanced RC patients treated with neoadjuvant chemoradiation followed by surgery, which is different from our current study. Different chemoradiotherapy regimens might influence the prognostic evaluation of patients with RC. Moreover, the sample sizes of these two studies[12,14] were less than ours.

In terms of feature selection, several relationships were uncovered between the specific features and their ability to predict distant metastasis. For the radiomic features included in this study, the first-order features show an excellent auxiliary classification effect, accounting for two of the three specified radiomic features. In the clinical features, we found that T-stage and N-stage had high standardized logistic regression coefficients, indicating the contribution to the combined model. This finding is consistent with the consensus that pathological T-stage and N-stage are very important for predicting the distant metastasis of RC. The histological grade, circumferential resection margin, Ki-67 score, and tumor diameter did not present enough predictive power for distant metastasis and prognosis. Subsequently, we integrated the Rad-score into a nomogram with clinical risk factors, and constructed a useful tool for individualized evaluation of distant metastasis and OS in patients with RC.

In our study, only patients who did not receive chemoradiotherapy were selected for the analysis of prognosis. On the one hand, the variables caused by the different intensive chemoradiotherapy methods can be controlled. On the other hand, we could build a novel model to stratify the high-risk patients. This was in step with the current trends toward personalized medicine[21]. Considering individualized evaluation of patients with RC of different stages, the subgroup survival analysis was performed. The low-risk groups had longer OS than the high-risk groups, which was statically significant between stage Ⅰ and Ⅱ (*P* < 0.001 and *P* = 0.0194, respectively). However, there was no statistically significant difference for stage Ⅲ (*P* = 0.1401). We speculate that this might be related to the small sample size of this study.

There are several limitations to this study. The first limitation is the relatively small sample size. Second, even if we controlled therapeutic methods to surgery alone, inevitable bias may exist due to the retrospective design of this study. Therefore, prospective and external validation studies are required in future studies. Third, all of the CT images were obtained from a single institution. In the future, multicenter verification is necessary to extend the versatility of the experimental results.

**Conclusion**

In conclusion, this study describes a combined model that incorporates a radiomics signature and clinical risk factors. The model can aid in the individualized prediction of distant metastasis and prognosis in patients with RC.

**ARTICLE HIGHLIGHTS**

***Research background***

Rectal cancer (RC) patient stratification by different factors may yield variable results. Therefore, more efficient prognostic biomarkers are needed for improved risk stratification, personalized treatment, and prognostication of RC patients.

***Research motivation***

In up to 70% of patients with RC, surgical removal of the primary tumor is successful. However, local recurrence and distant metastases are commonly detected in approximately 30% of RC patients, often within 3 years after surgery. Therefore, it is necessary to identify patients who may be at a higher risk of developing adverse outcomes post-surgery. In these patients, alternative and adjunctive therapies, such as chemotherapy, radiotherapy, or other targeted therapies, may be needed to minimize the risk of developing distant metastases.

***Research objectives***

To build a novel clinicoradiologic model for predicting the presence of distant metastases and 3-year overall survival (OS) rates in RC patients.

***Research methods***

This was a retrospective analysis of 148 patients (76 males and 72 females) with RC treated with curative resection, without neoadjuvant or postoperative chemoradiotherapy, between October 2012 and December 2015. These patients were allocated to a training or validation set, with a ratio of 7:3. Radiomic features were extracted from portal venous phase computed tomography images of RC. The least absolute shrinkage and selection operator regression analysis was used for feature selection. Multivariate logistic regression analysis was used to develop the Rad-score and the combined model. Receiver operating characteristic curves were constructed to evaluate the diagnostic performance of the models for predicting distant metastasis of RC. The association of the combined model with 3-year OS was investigated by Kaplan-Meier survival analysis.

***Research results***

A total of 51 (34.5%) patients had distant metastases, while 26 (17.6%) patients died, and 122 (82.4%) patients lived at least 3 years post-surgery. The values of both the Rad-score and the combined model were significantly different between the distant metastasis group and the non-metastasis group (0.46 ± 0.21 *vs* 0.32 ± 0.24 for the Rad-score, 0.60 ± 0.23 *vs* 0.28 ± 0.26 for the combined model; *P* < 0.001 for both models). Predictors contained in the combined model included the Rad-score, pathological N-stage, and T-stage. The combined model showed good discrimination, with an area under the curve of 0.842 and 0.802 for the training set and validation set, respectively. For the survival analysis, the combined model was associated with an improved OS in the whole cohort and the respective subgroups.

***Research conclusions***

This study presents a novel model, visualized in a nomogram, that can be used to facilitate individualized prediction of distant metastasis and 3-year OS in patients with RC.

***Research perspectives***

Radiomics may change the practice of medicine, particularly for patients with RC. However, there are challenges to be overcome before its routine implementation including challenges related to sample size, model design, and the lack of robust multicenter validation set. Therefore, prospective multicenter studies of a larger size are needed to externally validate our proposed model in the future.

**REFERENCES**

1 **Deng Y**. Rectal Cancer in Asian *vs* Western Countries: Why the Variation in Incidence? *Curr Treat Options Oncol* 2017; **18**: 64 [PMID: 28948490 DOI: 10.1007/s11864-017-0500-2]

2 **McArdle CS**, Hole D, Hansell D, Blumgart LH, Wood CB. Prospective study of colorectal cancer in the west of Scotland: 10-year follow-up. *Br J Surg* 1990; **77**: 280-282 [PMID: 1691033 DOI: 10.1002/bjs.1800770314]

3 **Sargent DJ**, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, Labianca R, Seitz JF, O'Callaghan CJ, Francini G, Grothey A, O'Connell M, Catalano PJ, Blanke CD, Kerr D, Green E, Wolmark N, Andre T, Goldberg RM, De Gramont A. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 2005; **23**: 8664-8670 [PMID: 16260700 DOI: 10.1200/jco.2005.01.6071]

4 **Yang ZF**, Wu DQ, Wang JJ, Lv ZJ, Li Y. Short- and long-term outcomes following laparoscopic *vs* open surgery for pathological T4 colorectal cancer: 10 years of experience in a single center. *World J Gastroenterol* 2018; **24**: 76-86 [PMID: 29358884 DOI: 10.3748/wjg.v24.i1.76]

5 **Sun Y**, Lin H, Lu X, Huang Y, Xu Z, Huang S, Wang X, Chi P. A nomogram to predict distant metastasis after neoadjuvant chemoradiotherapy and radical surgery in patients with locally advanced rectal cancer. *J Surg Oncol* 2017; **115**: 462-469 [PMID: 28105657 DOI: 10.1002/jso.24522]

6 **Sargent DJ**, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J, Buyse M, Labianca R, Seitz JF, O'Callaghan CJ, Francini G, Grothey A, O'Connell M, Catalano PJ, Kerr D, Green E, Wieand HS, Goldberg RM, de Gramont A; ACCENT Group. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol* 2007; **25**: 4569-4574 [PMID: 17876008 DOI: 10.1200/jco.2006.10.4323]

7 **Bown EJ**, Lloyd GM, Boyle KM, Miller AS. Rectal cancer: prognostic indicators of long-term outcome in patients considered for surgery. *Int J Colorectal Dis* 2014; **29**: 147-155 [PMID: 24051904 DOI: 10.1007/s00384-013-1772-z]

8 **Desch CE**, Benson AB 3rd, Somerfield MR, Flynn PJ, Krause C, Loprinzi CL, Minsky BD, Pfister DG, Virgo KS, Petrelli NJ; American Society of Clinical Oncology. Colorectal cancer surveillance: 2005 update of an American Society of Clinical Oncology practice guideline. *J Clin Oncol* 2005; **23**: 8512-8519 [PMID: 16260687 DOI: 10.1200/jco.2005.04.0063]

9 **Horvat N**, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ. MRI of Rectal Cancer: Tumor Staging, Imaging Techniques, and Management. *Radiographics* 2019; **39**: 367-387 [PMID: 30768361 DOI: 10.1148/rg.2019180114]

10 **Reggiani Bonetti L**, Lionti S, Domati F, Barresi V. Do pathological variables have prognostic significance in rectal adenocarcinoma treated with neoadjuvant chemoradiotherapy and surgery? *World J Gastroenterol* 2017; **23**: 1412-1423 [PMID: 28293088 DOI: 10.3748/wjg.v23.i8.1412]

11 **Nelson DA**, Tan TT, Rabson AB, Anderson D, Degenhardt K, White E. Hypoxia and defective apoptosis drive genomic instability and tumorigenesis. *Genes Dev* 2004; **18**: 2095-2107 [PMID: 15314031 DOI: 10.1101/gad.1204904]

12 **Lovinfosse P**, Polus M, Van Daele D, Martinive P, Daenen F, Hatt M, Visvikis D, Koopmansch B, Lambert F, Coimbra C, Seidel L, Albert A, Delvenne P, Hustinx R. FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. *Eur J Nucl Med Mol Imaging* 2018; **45**: 365-375 [PMID: 29046927 DOI: 10.1007/s00259-017-3855-5]

13 **Wang J**, Shen L, Zhong H, Zhou Z, Hu P, Gan J, Luo R, Hu W, Zhang Z. Radiomics features on radiotherapy treatment planning CT can predict patient survival in locally advanced rectal cancer patients. *Sci Rep* 2019; **9**: 15346 [PMID: 31653909 DOI: 10.1038/s41598-019-51629-4]

14 **Meng Y**, Zhang Y, Dong D, Li C, Liang X, Zhang C, Wan L, Zhao X, Xu K, Zhou C, Tian J, Zhang H. Novel radiomic signature as a prognostic biomarker for locally advanced rectal cancer. *J Magn Reson Imaging* 2018 [PMID: 29437271 DOI: 10.1002/jmri.25968]

15 **Ma X**, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features. *BMC Med Imaging* 2019; **19**: 86 [PMID: 31747902 DOI: 10.1186/s12880-019-0392-7]

16 **Shu Z**, Fang S, Ding Z, Mao D, Cai R, Chen Y, Pang P, Gong X. MRI-based Radiomics nomogram to detect primary rectal cancer with synchronous liver metastases. *Sci Rep* 2019; **9**: 3374 [PMID: 30833648 DOI: 10.1038/s41598-019-39651-y]

17 **Liu H**, Zhang C, Wang L, Luo R, Li J, Zheng H, Yin Q, Zhang Z, Duan S, Li X, Wang D. MRI radiomics analysis for predicting preoperative synchronous distant metastasis in patients with rectal cancer. *Eur Radiol* 2019; **29**: 4418-4426 [PMID: 30413955 DOI: 10.1007/s00330-018-5802-7]

18 **Lambin P**, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749-762 [PMID: 28975929 DOI: 10.1038/nrclinonc.2017.141]

19 **Ming X**, Oei RW, Zhai R, Kong F, Du C, Hu C, Hu W, Zhang Z, Ying H, Wang J. MRI-based radiomics signature is a quantitative prognostic biomarker for nasopharyngeal carcinoma. *Sci Rep* 2019; **9**: 10412 [PMID: 31320729 DOI: 10.1038/s41598-019-46985-0]

20 **Tang C**, Hobbs B, Amer A, Li X, Behrens C, Canales JR, Cuentas EP, Villalobos P, Fried D, Chang JY, Hong DS, Welsh JW, Sepesi B, Court L, Wistuba II, Koay EJ. Development of an Immune-Pathology Informed Radiomics Model for Non-Small Cell Lung Cancer. *Sci Rep* 2018; **8**: 1922 [PMID: 29386574 DOI: 10.1038/s41598-018-20471-5]

21 **Sheikh AR**, Yameen H, Hartshorn K. Treatment of Rectal Cancer in Older Adults. *Curr Oncol Rep* 2018; **20**: 102 [PMID: 30456634 DOI: 10.1007/s11912-018-0746-2]

**Footnotes**

**Institutional review board statement:** This study was reviewed and approved by the Ethics Committee of West China Hospital of Sichuan University.

**Informed consent statement:** Patients were not required to give informed consent to the study because the analysis used anonymous clinical data that were obtained after each patient agreed to treatment by written consent.

**Conflict-of-interest statement:** All authors declare no conflicts of interest related to this article.

**Data sharing statement:** No additional data are available.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

**Manuscript source:** Invited manuscript

**Peer-review started:** May 27, 2020

**First decision:** June 4, 2020

**Article in press:** August 13, 2020

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): B

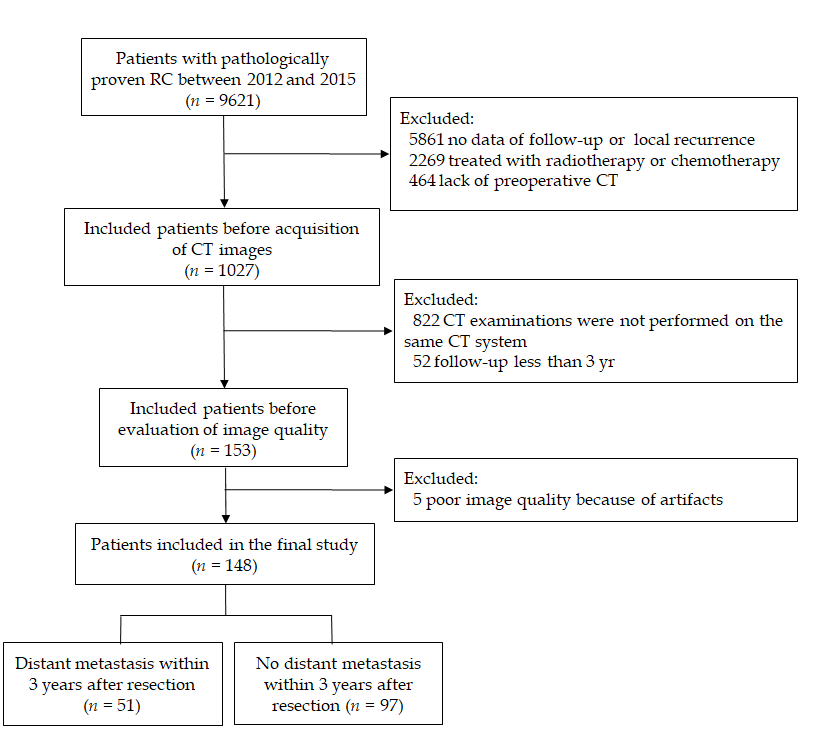
Grade C (Good): 0

Grade D (Fair): 0

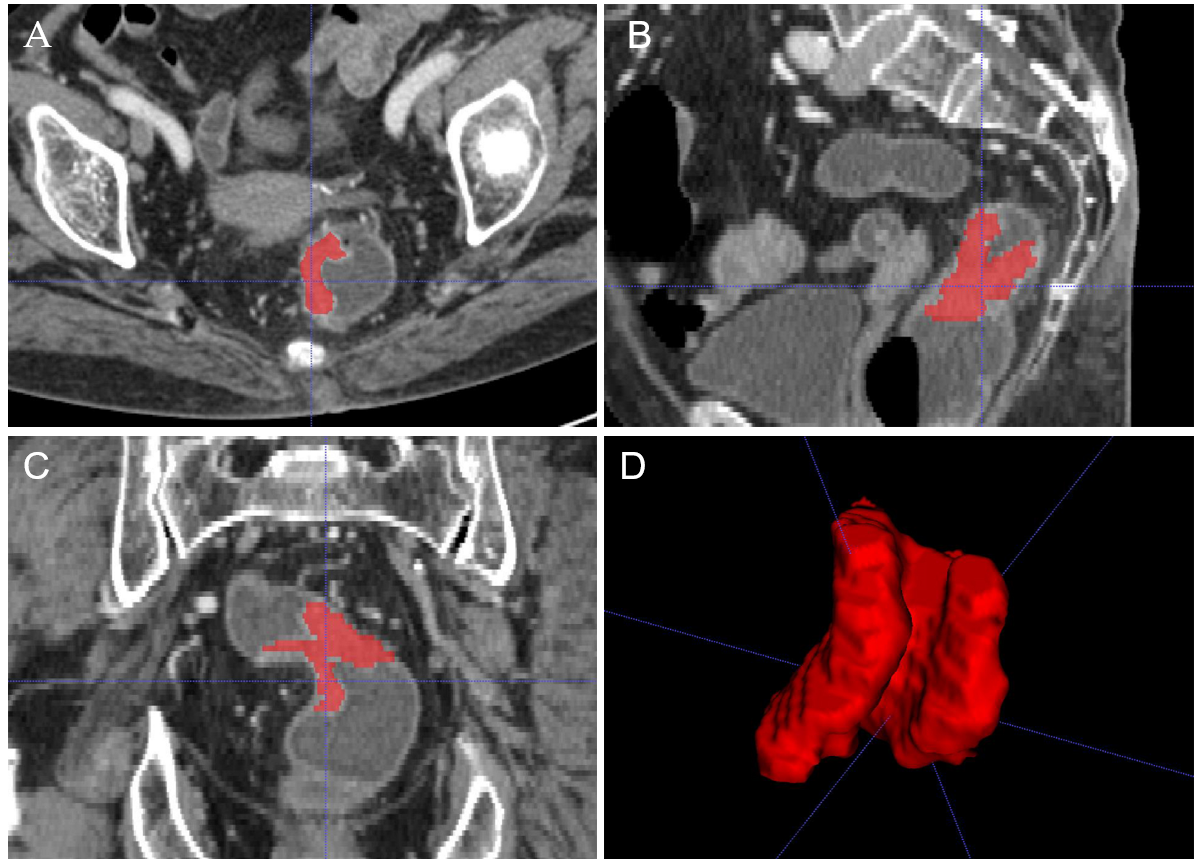
Grade E (Poor): 0

**P-Reviewer:** Rim CH **S-Editor:** Zhang L **L-Editor:** Wang TQ **P-Editor:** Ma YJ

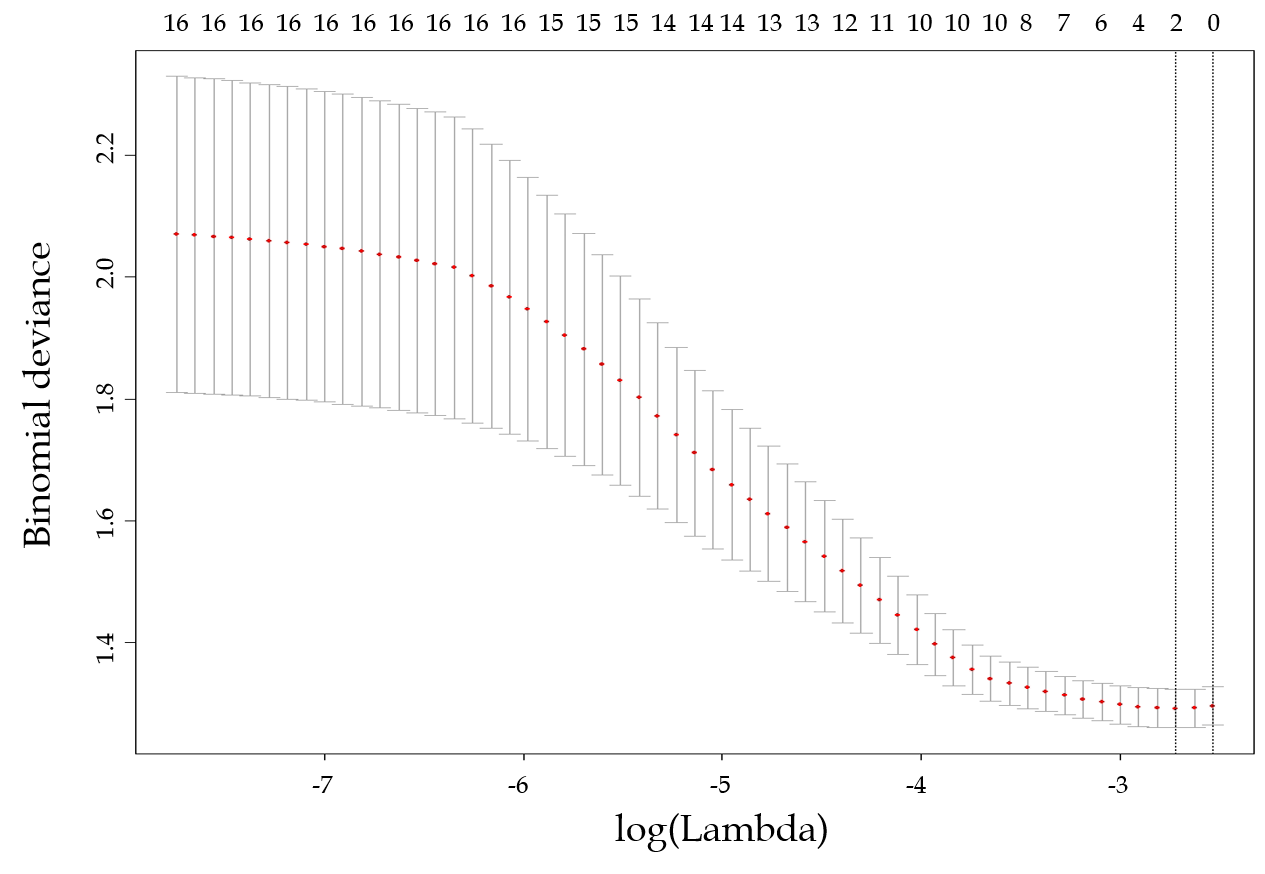
**Figure Legends**



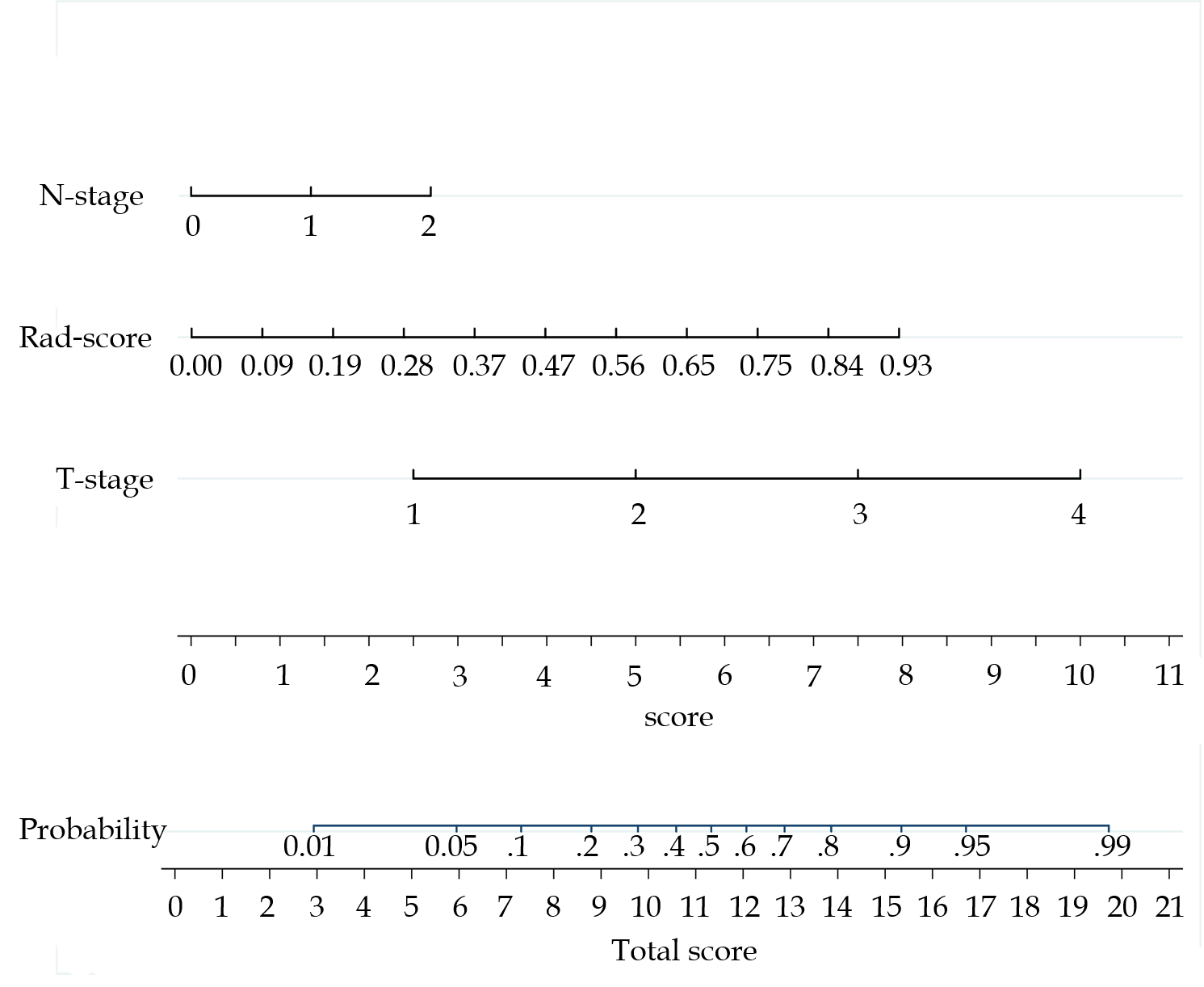
**Figure 1 Flow chart of the recruitment pathway for patients included in the study.** CT: Computed tomography.



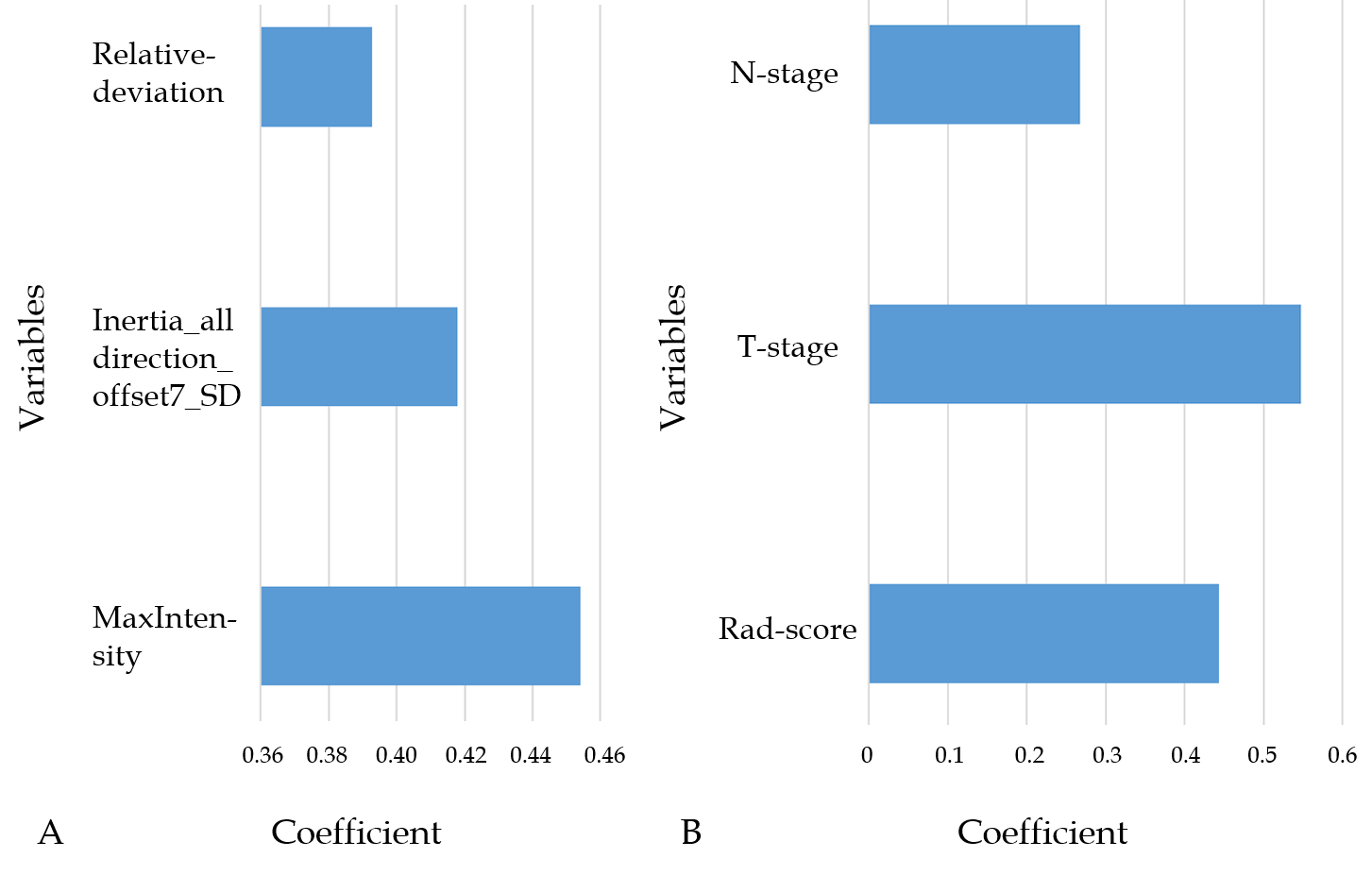
**Figure 2 A 58-year-old woman with rectal cancer.** A-C: Representative manual segmentation of the whole lesion in the axial, sagittal, and coronal enhanced computed tomography images. Red lines represent the delineations of the regions of interest used to derive the radiomic features; D: Three-dimensional volumetric reconstruction of the segmented lesion.



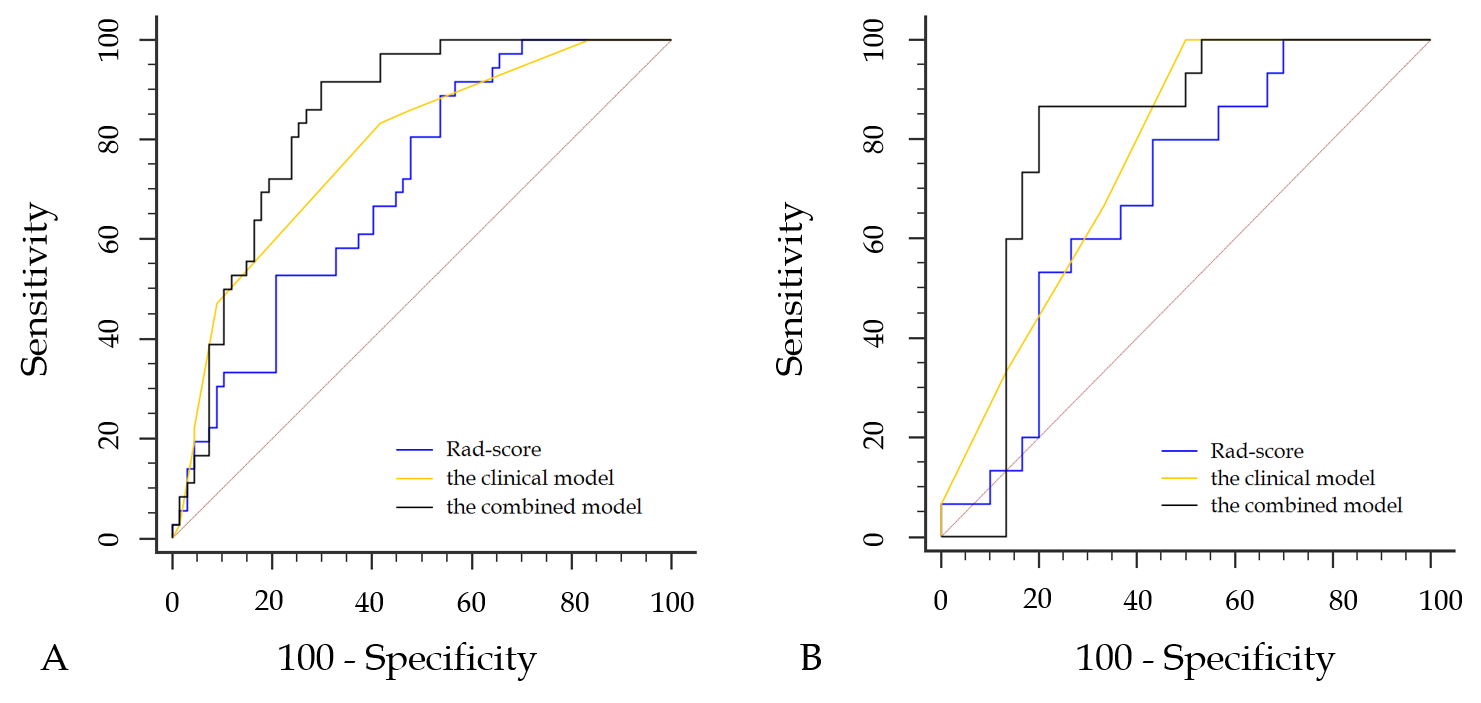
**Figure 3 Feature selection using the least absolute shrinkage and selection operator regression model.** Dotted vertical lines were drawn at the optimal values (two features were selected) by using the minimum criteria and one standard error of the minimum criteria (*i.e.*, the 1-SE criteria).



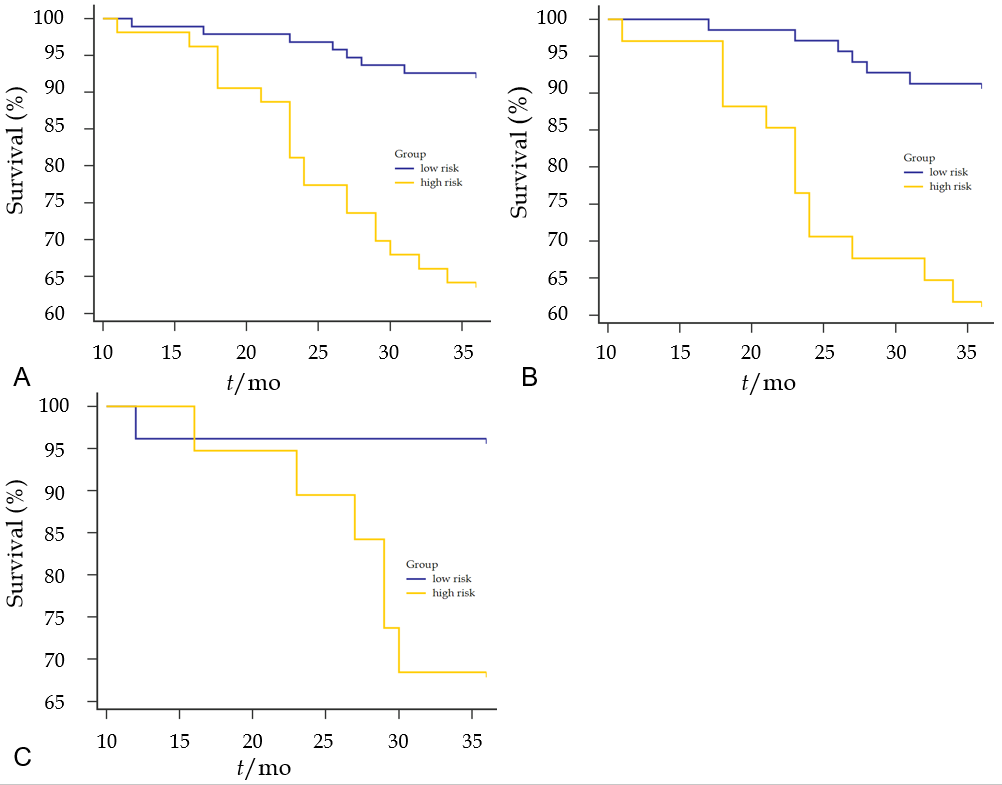
**Figure 4 Developed radiomics nomogram.** The radiomics nomogram was developed in the training cohort, with the Rad-score, T-stage, and N-stage incorporated.



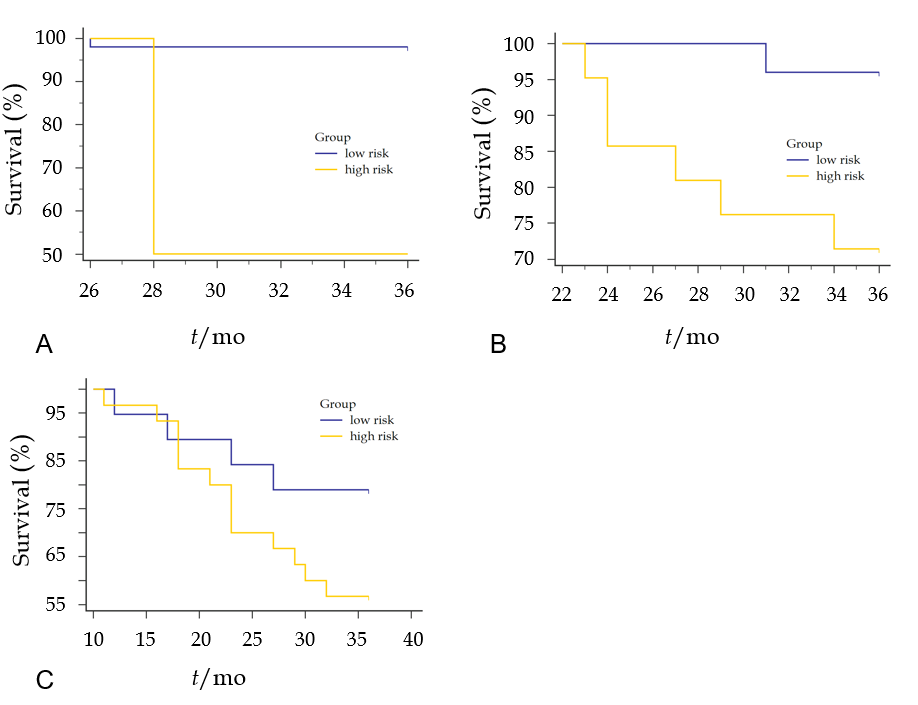
**Figure 5 Histogram showing the contribution of each variable to the models.** A: the Rad-score; B: the combined model. The contributions of the variables were measured by the values of the standardized logistic regression coefficients.



**Figure 6 Receiver operating characteristic curves of the models in the training and validation cohorts.** A: In the training set, the combined model [area under the curve under the curve (AUC) = 0.842] achieved a better performance than the Rad-score (AUC = 0.709) and the clinical model containing T-stage and N-stage (AUC = 0.782); B: In the validation set, the AUCs of the Rad-score, the clinical model, and the combined model were 0.687, 0.766, and 0.802, respectively.



**Figure 7 Kaplan-Meier curves for overall survival stratified by risk group, as identified by the combined model.** A: The whole set. *P* < 0.0001; B: The training set. *P* = 0.0001; C: The validation set. *P* = 0.0137.



**Figure 8 Overall survival curves for the low- and high-risk groups classified according to the combined model in the subgroups of the overall pathological stage.** A: Stage Ⅰ. *P* < 0.001; B: Stage Ⅱ. *P* = 0.0194; C: Stage Ⅲ. *P* = 0.1401.

**Table 1 Patient and tumor characteristics in the training and validation sets**

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Training (*n* = 103)** | **Validation (*n* = 45)** | ***P* value** |
| Age (yr) | 60 ± 12 | 59 ± 11 | 0.513 |
| Gender |  |  | 0.156 |
| Male | 57 | 19 |  |
| Female | 46 | 26 |  |
| Tumor size (long axis, mm) | 49.4 ± 11.6 | 54.1 ± 14.6 | 0.040 |
| Pathological T-stage |  |  | 0.218 |
| T1 | 11 | 2 |  |
| T2 | 34 | 13 |  |
| T3 | 55 | 29 |  |
| T4 | 3 | 1 |  |
| Pathological N-stage |  |  | 0.018 |
| N0 | 75 | 24 |  |
| N1 | 18 | 12 |  |
| N2 | 10 | 9 |  |
| M-stage | 0 | 0 |  |
| Overall Stage (pTNM) |  |  | 0.081 |
| Ⅰ | 39 | 14 |  |
| Ⅱ | 36 | 10 |  |
| Ⅲ | 28 | 21 |  |
| Pathological grade |  |  | 0.222 |
| Well | 3 | 0 |  |
| Moderate | 78 | 32 |  |
| Poor | 22 | 13 |  |
| CRM |  |  | 0.788 |
| Negativity | 91 | 39 |  |
| Positivity | 12 | 6 |  |
| Ki-67 (%) | 52 ± 20 | 47 ± 21 | 0.261 |
| Distant metastasis (+/-) | 36/67 | 15/30 | 0.849 |
| Distant metastasis1 |  |  | 0.394 |
| Lung | 16 | 13 |  |
| Liver | 19 | 8 |  |
| Other (bone, brain, *etc.*) | 6 | 6 |  |
| Follow-up time (months) |  |  |  |
| DFS | 29 ± 12 | 30 ± 12 | 0.638 |
| OS | 34 ± 5 | 34 ± 5 | 0.691 |

The data are presented as the mean ± the standard deviation. 1A total of 51 patients had distant metastases within 3 years after surgery, and some of these patients had multiple organic metastases. CRM: Circumferential resection margin; DFS: Disease-free survival; OS: Overall survival.

**Table 2 Comparison of the models between the metastasis group and the non-metastasis group, and between the death group and the survival group**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Metastasis** | **Non-Metastasis** | ***P* value** | **Death** | **Survival** | ***P* value** |
| 3-yr1 |  |  |  |  |  |  |
| *n* (%) | 51 (34.5) | 97 (65.5) |  | 26 (17.6) | 122 (82.4) |  |
| Rad-score | 0.46 ± 0.21 | 0.32 ± 0.24 | < 0.001 | 0.45 ± 0.20 | 0.35 ± 0.24 | 0.023 |
| Clinical | 0.52 ± 0.20 | 0.30 ± 0.21 | < 0.001 | 0.55 ± 0.20 | 0.34 ± 0.22 | < 0.001 |
| Combined | 0.60 ± 0.23 | 0.28 ± 0.26 | < 0.001 | 0.62 ± 0.22 | 0.34 ± 0.28 | < 0.001 |
| 2-yr1 |  |  |  |  |  |  |
| *n* (%) | 41 (27.7) | 107 (72.3) |  | 15 (10.1) | 133 (89.9) |  |
| Rad-score | 0.44 ± 0.20 | 0.34 ± 0.25 | 0.009 | 0.40 ± 0.17 | 0.35 ± 0.25 | 0.368 |
| Clinical | 0.51 ± 0.19 | 0.32 ± 0.23 | < 0.001 | 0.63 ± 0.15 | 0.35 ± 0.22 | < 0.001 |
| Combined | 0.56 ± 0.23 | 0.31 ± 0.28 | < 0.001 | 0.65 ± 0.19 | 0.36 ± 0.29 | < 0.001 |
| 1-yr1 |  |  |  |  |  |  |
| *n* (%) | 25 (16.9) | 123 (83.1) |  | 2 (1.4) | 146 (98.6) |  |
| Rad-score | 0.39 ± 0.19 | 0.36 ± 0.25 | 0.453 | 0.29 ± 0.36 | 0.37 ± 0.24 | 0.676 |
| Clinical | 0.55 ± 0.19 | 0.34 ± 0.22 | < 0.001 | 0.75 ± 0.24 | 0.37 ± 0.23 | 0.258 |
| Combined | 0.56 ± 0.23 | 0.35 ± 0.29 | 0.001 | 0.60 ± 0.52 | 0.38 ± 0.29 | 0.302 |

The data are presented as the mean ± standard deviation. 1During the time from surgery to the presence of distant metastasis or death.

**Table 3 Diagnostic performance of the three models in both training and validation cohorts**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Training set** | | | **Validation set** | | |
| 1Time |  | Rad-score | Clinical | Combined | Rad-score | Clinical | Combined |
| 3-yr | AUC | 0.709 | 0.782 | 0.842 | 0.687 | 0.766 | 0.802 |
|  | SEN | 88.89% | 88.33% | 91.67% | 80% | 100% | 86.67% |
|  | SPE | 42.67% | 58.21% | 70.15% | 56.67% | 50% | 80% |
| 2-yr | AUC | 0.700 | 0.760 | 0.816 | 0.609 | 0.720 | 0.691 |
| 1-yr | AUC | 0.614 | 0.771 | 0.769 | 0.545 | 0.742 | 0.632 |

1Time: The time from surgery to the presence of distant metastases. AUC: Area under the summary receiver operating characteristic curve; SEN: Sensitivity; SPE: Specificity.