An endorectal ultrasound-based radiomics signature for preoperative prediction of lymphovascular invasion of rectal cancer

Yu-quan Wu1†, Rui-zhi Gao1†, Peng Lin1, Rong Wen1, Hai-yuan Li1, Mei-yan Mou1, Feng-huan Chen1, Fen Huang1, Wei-jie Zhou2, Hong Yang1, Yun He1* and Ji Wu1*

Abstract

Objective: To investigate whether radiomics based on ultrasound images can predict lymphovascular invasion (LVI) of rectal cancer (RC) before surgery.

Methods: A total of 203 patients with RC were enrolled retrospectively, and they were divided into a training set (143 patients) and a validation set (60 patients). We extracted the radiomic features from the largest gray ultrasound image of the RC lesion. The intraclass correlation coefficient (ICC) was applied to test the repeatability of the radiomic features. The least absolute shrinkage and selection operator (LASSO) was used to reduce the data dimension and select significant features. Logistic regression (LR) analysis was applied to establish the radiomics model. The receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA) were used to evaluate the comprehensive performance of the model.

Results: Among the 203 patients, 33 (16.7%) were LVI positive and 170 (83.7%) were LVI negative. A total of 5350 (90.1%) radiomic features with ICC values of ≥ 0.75 were reported, which were subsequently subjected to hypothesis testing and LASSO regression dimension reduction analysis. Finally, 15 selected features were used to construct the radiomics model. The area under the curve (AUC) of the training set was 0.849, and the AUC of the validation set was 0.781. The calibration curve indicated that the radiomics model had good calibration, and DCA demonstrated that the model had clinical benefits.

Conclusion: The proposed endorectal ultrasound-based radiomics model has the potential to predict LVI preoperatively in RC.

Keywords: Ultrasound, Radiomics, Model, Rectal cancer, Lymphovascular invasion

Background

According to global cancer statistics in 2020, rectal cancer (RC) is a major threat to humans. The RC has the third-highest recorded incidence of cancer, which is listed after lung cancer and breast cancer, and the cancer-related mortality rate of RC is ranked second after lung cancer [1]. Early detection and diagnosis is crucial to effectively treat RC. Unfortunately, most patients are diagnosed with locally advanced rectal cancer (LARC), which means RC with stage T3 or T4 [2]. The standard
treatment option for LARC is neoadjuvant chemoradiotherapy followed by total mesorectal excision (TME) [3, 4]. However, even though patients with LARC have already received neoadjuvant chemoradiotherapy, the 5-year overall survival rate is only 66.3% [5].

The poor prognosis of RC is related to various factors, including lesion size, stage, lymphovascular invasion (LVI) status, tumor marker status, immune indicators that are positive, RC’s sensitivity to radiotherapy and chemotherapy, and so on [6].

Previous studies have shown that vascular invasion is an independent risk factor for poor prognosis of RC [7, 8]. Now, the LVI status can be obtained by only postoperative pathology, which therefore cannot guide whether the patient should be treated more aggressively preoperatively. Therefore, preoperative prediction of LVI in RC is helpful for clinical decision-making.

There are numerous means of preoperative examination for RC. Colonoscopy can observe the general surface of the mass and perform the biopsy; however, the disadvantage is that the depth of infiltration and extraintestinal invasion could not be observed. Magnetic resonance imaging (MRI), especially high-resolution MRI, is the first choice for preoperative imaging diagnosis and postoperative follow-up review, which can accurately perform tumor node metastasis (TNM) staging, but its high cost makes it difficult to repeat multiple examinations, and it is always difficult to distinguish the T2 stage from the T3 stage [9, 10]. Endoscopic ultrasound (EUC) and transrectal ultrasound (TRUS) are cheap and repeatable and can visualize structures at multiple levels of the intestinal wall, but they still do not solve the problem of the preoperative prediction of LVI [11, 12]. To sum up, to optimize individualized clinical treatment, a method that can noninvasively and preoperatively predict the presence or absence of LVI is an urgent clinical need.

In recent years, radiomics has been introduced and has the potential to provide help for the above clinical challenges. Radiomics, using computers to excavate the data of medical images (ultrasound, CT, MR, etc.) to discover the intrinsic relationship between medical images and features, such as diagnostic localization, biological behavior, and treatment outcome of diseases, could provide a reference for clinical decision-making. Several recent reviews systematically reviewed artificial intelligence (AI) models and machine learning about radiomics based on high-resolution MRI which could be used to predict the response of chemotherapy and distant metastasis of RC, and the results showed that this novel imaging technology might be helpful to clinical decision-making [13–16]. Chinese scholars have reported that CT radiomics can predict whether lymph node metastasis exists preoperatively in RC [17], and CT radiomics and MR radiomics can predict colorectal cancer LVI [18, 19]. Also, high-resolution MR radiomics could predict peripheral nerve invasion [20] and disease-free survival [21] in RC, assessing response to treatment for RC [22, 23]. Ultrasound radiomics can predict subtypes of liver cancer [24] and determine whether lymph node metastasis exists in thyroid cancer [25]. However, to the best of our knowledge, studies of endorectal ultrasound-based radiomics in the LVI have not yet been reported. Our study aims to utilize endorectal ultrasound-based radiomics techniques to predict LVI preoperatively and noninvasively to aid in making clinical decisions.

**Materials and methods**

**Subjects**

The Ethics Committee of our hospital approved this retrospective study. Informed consent was waived. Patients with RC who underwent surgery in our hospital from January 2018 to February 2021, were included in this study. Inclusion criteria were the following: (1) RC that was confirmed by postoperative pathology, (2) TRUS examination that was conducted before surgery, (3) the biggest section of the RC lesion image that was clear and complete (in the case of multiple lesions, the largest lesion was to be selected), and (4) the required clinicopathological data that were complete. The exclusion criteria were as follows: (1) previous history of RC, (2) unclear diagnosis of LVI, and (3) neoadjuvant chemoradiotherapy if performed before ultrasound examination.

**Ultrasound examination methods**

First, the patient was asked to lie on the examination bed with a left lateral coxa and knee bending position, take off pants, and fully expose the anus. Before the TRUS examination, a digital rectal examination (DRE) was conducted to judge whether anal stenosis exists and the distance between the lower edge of the lesion and the anus, the lesion’s size, and its position. Then, a 50-ml mixture of water and starch was instilled into the rectal cavity through the anus, and the TRUS probe was wrapped in a condom for isolation when the probe was extended into the anal canal for examination. A convex or line array probe was selected according to the lesion’s size and distance between the lower edge of the lesion and the anus. An experienced sonographer performed a comprehensive scan of the lesion to assess the lesion’s size, shape, margin, and internal echogenicity, the distance between the lower edge of the lesion and the anus, and whether muscularis, serosal layer, internal anal sphincter, and external anal sphincter invasion existed. The images were stored for future use.
**Instruments and image acquisition**

Esaote MyLab Class C and General Electric Company (GE) LOGIQ E9 ultrasound diagnostic instruments were used to acquire ultrasound images of the lesions. The frequency of the probe was 6–15 MHz, and the images were recorded and saved in the Digital Imaging and Communications in Medicine (DICOM) format.

**Clinical and pathological data**

Clinical and pathological data, such as the patient’s gender, age, tumor size, stage, vascular invasion, and serum CEA, CA125, CA199, CA242, CA742, etc., were collected.

**Region of interest (ROI) segmentation, feature extraction, and consistency analysis**

In this study, the ROIs of the lesions were outlined using the image software application ITK-SNAP (version 3.8.0), which is an open source for medical image segmentation (Fig. 2) [24]. The ROIs were outlined manually by two independent radiologists (reader 1 with 5 years of experience in TRUS examination and reader 2 with 10 years). Neither radiologist knew the pathology results. Also, 50 images were randomly selected from all the images and drawn independently by two doctors to evaluate consistency between different observers. Subsequently, we performed feature extraction of the tumor ROI using Intelligent Foundry software (version 1.3, GE Company, Shanghai). This software extracted massive amounts of image features and represented them as quantitative data, which could be used to analyze the potential heterogeneity within the tumors [17]. Finally, we used consistency analysis to gain consistency between different observers, and we considered the features with consistency tests greater than 0.75 as those with high confidence and retained them for subsequent studies.

**Feature preprocessing and selection**

In this study, we used the Cart package in R language (version 4.3.0) to stratify all patients in a 7:3 ratio by the computer randomization method, with 70% of patients entering the training set for model construction and 30% of patients entering the validation set for model verification. Different evaluation metrics (radiomic features) often had different scales and scale units. To increase the comparability of quantitative radiomic features, we performed Z-score normalization for quantitative features in training and validation sets, respectively. Subsequently, to select the radiomic features associated with LVI in RC from the numerous image features, we performed a hypothesis test on all features in the training set and retained the features with statistically significant significance in the hypothesis test for the multifactor regression analysis. The massive number of radiomic features constructed in the model easily led to model overfitting, and to reduce the influence of high-dimensional features on the model, we used LASSO regression for feature downscaling and tenfold cross-validation to filter the optimal subset of features to predict LVI in RC from radiomic features. Meanwhile, we performed a correlation analysis of all features to optimize the covariance effect of the
subset of features. We considered a correlation coefficient of ≥ 0.8 to be a high correlation.

Construction, evaluation, and validation of the imaging radiomics model
To construct the radiomics model, we performed multifactor logistic regression modeling using the optimal subset of features in the training set. Meanwhile, we used the area under the curve (AUC) of the receiver operating characteristic (ROC) curve of both training and validation sets to evaluate the comprehensive discrimination ability of the radiomics model. We also used calibration curves to calculate the agreement between the predicted and observed probabilities. We used decision curve analysis (DCA) to estimate the outcome of the prediction model under different threshold probabilities.

Statistical analysis
SPSS 22.0 and R language (version 4.3.0) were used for the analysis. For statistical analysis of clinical parameters and hypothesis testing of radiomics characteristics, the chi-square test or Fisher’s exact probability method was used for statistical analysis of count data, and an independent samples t-test or Mann–Whitney U test was performed for measurement data. Means ± standard deviations were used for measures that conformed to a normal distribution and medians for those abnormal distributions. A P-value of < 0.05 was considered statistically significant.

Results
Clinicopathological data of the patients
According to the above-mentioned inclusion and exclusion criteria, 203 patients with RC were finally included (Fig. 1). The age range of the patients was 19–89 years, and the mean was 59.4 ± 12.4 years. There were 170 cases in the LVI-negative group and 33 cases in the LVI-positive group. The clinicopathological information of the patients is shown in Table 1. The differences in age, gender, maximum tumor diameter, thick diameter, ultrasound T-stage, N-stage, CEA, CA199, CA125, CA242, and CA724 between the two sets have no statistical significance, suggesting that the random grouping of these two groups is reasonable.

Extraction and selection of endorectal ultrasound-based radiomic features
The Intelligence Foundry software was used to extract the features of the outlined ROIs (Fig. 2), and a total of 5936 image features were obtained. The radiomic features were listed as follows: (1) 122 original features (first-order statistics, texture classes, shape descriptors, etc.); (2) 48 intra-perinodular textural transitions (ipris); (3) 432 wavelets + local binary patterns (LBPs); (4) 1170 co-occurrence of local anisotropic gradient orientations (CoLIAGe); (5) 1080 Gabors; (6) 2944 shearlets; (7) 60 wavelet-based improved local binary patterns (WILBPs); and (8) 80 phased congruency-based local binary patterns (PLBPs). The ICC test results showed 5350 (90.1%) radiomic features with ICC coefficients of ≥ 0.75 (Fig. 3), which were considered stable and used for subsequent analysis. Through hypothesis testing, we found a total of 179 radiomic features associated with colorectal cancer.
LVI in the training set. A total of 15 features were used as the optimal subset for predicting LVI in RC after LASSO regression combined with tenfold cross-validation (Fig. 4). The correlation analysis showed that the maximum correlation coefficient between the radiomic features of the optimal feature subset was 0.68 (Fig. 5).

**Radiomics model establishment and efficacy assessment**

We used logistic regression on the optimal subset of features to build a radiomics model in this training set, and we also applied the validation set for model evaluation. The ROC curves (Fig. 6) showed that the radiomics model had good diagnostic efficacy in both sets (AUC of 0.849, sensitivity of 0.76, specificity of 0.75, and accuracy of 0.86 in the training set; AUC of 0.781, sensitivity of 0.75, specificity of 0.79, and accuracy of 0.83 in the validation set). The calibration curve suggested that the radiomics model in our study had good accuracy (Fig. 6). The results of the DCA showed that the radiomic features were clinically useful (Fig. 7).

**Discussion**

RC has a high incidence and poor prognosis. Microvascular invasion is an important factor in the residual and recurrence of RC [26]. At this stage of clinical practice, microvascular invasion is mainly identified by postoperative pathology, with a certain lag. Our study purpose was to develop a preoperative prediction model for LVI of RC using endorectal ultrasound-based radiomics. The model showed good predictive value in both training and validation sets, with AUC of 0.849 and 0.781, respectively, suggesting that endorectal ultrasound-based radiomics may predict the vascular invasion of RC to provide a reference for clinical treatment decisions.

Radiomics is useful in the diagnosis, staging, treatment response, and prognosis of many solid tumors. For example, radiomics can be applied to the differential diagnosis...
The ICC between reader1 and reader2

Fig. 3 The intraclass correlation coefficient (ICC) test between readers 1 and 2. Among 5,936 features, a total of 5,350 (90.1%) features had ICC values of ≥ 0.75.

Fig. 4 Selecting features via the least absolute shrinkage and selection operator (LASSO) logistic regression analysis. a Using tenfold cross-validation to select the tuning parameter (λ) in the LASSO model via minimum criteria. The correlation between the AUC curve and log(λ) was plotted in this chart. Dotted vertical lines were drawn at optimal values. b LASSO coefficient profiles of 179 features. The coefficient profile was plotted against the log(λ) sequence. After the tenfold cross-validation, the optimal λ resulted in 15 nonzero coefficients.
of HCC [24], the assessment of the efficacy of neoadjuvant chemotherapy for breast cancer [27], and the characterization, grading, and prediction of tumor outcome in kidney cancer [28]. Radiomics also has important value in diagnostic staging [29], treatment response [23, 30], and prognosis of RC [21]. For LVI of RC, CT-based radiomics and MRI-based radiomics have good diagnostic efficacy for microvascular infiltration of RC [18, 19]. However, endorectal ultrasound-based radiomics has not been reported for vascular invasion in RC.

Ultrasound is an inexpensive, quick, and noninvasive examination, which reflects its worth in offering TNM stage information of RC. Patients may gain benefits if this examination combining radiomics is used to predict the LVI of RC noninvasively before resection.

There have been some studies of radiomics predicting vascular invasion in RC. Zhang et al. [19] applied MR radiomics to predict LVI with an AUC of 0.884 in the training set and 0.876 in the validation set, but they included only 94 patients in this study, and this study extracted only 1188 radiomic features. A study using CT radiomics to predict RC LVI also extracted only 396 radiomic features [18]. Compared with the above two studies, our study had more patients and features, including a total of 203 patients and extracting 5936 radiomic features.
Fig. 6  

- **Fig. 6 a** AUC of the receiver operating characteristic (ROC) curve in the training set.  
- **Fig. 6 b** AUC of the ROC curve in the validation set.  
- **Fig. 6 c** Calibration curve in the training set.  
- **Fig. 6 d** Calibration curve in the validation set.

Fig. 7  

- **Fig. 7 a** Predictive performance of the model and its decision curve analysis.  
- **Fig. 7 b** are the decision curves of the probabilistic relationship between net income and threshold drawn based on the training and verification sets, respectively. The probability of the X-axis high-risk threshold is the probability of LVI in this study, ranging from 0 to 1. The Y-axis measures net income. The black line shows that all rectal cancer patients are free of LVI, and the yield is 0. The purple curve indicates that LVI occurred in all patients.
Nevertheless, radiomics has limitations, especially in its operator dependence on manual outlining of the ROIs and the generalizability of the model [31]. For example, Yao et al. [32], predicted HCC subtypes, Ki-67, and microvascular invasion (MVI) by using radiomics of multimodal ultrasound images; however, they only performed a model of the training set without the validation set and did not perform the ICC test. Li et al. [33] applied CT radiomics to develop a model for predicting distant metastasis in RC; in this study, the validation set was available, but the ICC test was not performed. The above studies did not perform ICC tests on the ROIs to assess the consistency between different operators and the reproducibility of these models. In our research, the prediction model is not only established by a training set but also verified by a validation set to ensure the universality of the model. At the same time, we performed an ICC test on the outlined portion of the ROIs to test the consistency between different operators. The results suggested that the ICC is above 0.75 in 90.1% (5350/5936), which guaranteed the reproducibility of our study, and that similar results can be obtained when the model is used by different analysts. Also, in the process of radiomics performance, we used CLAIM (Checklist for Artificial Intelligence in Medical Imaging) as the Radiomics Quality Score to make sure the methodological quality [34]. And this checklist was added in the Additional file 1.

This study has the following limitations. First, the case selection bias may exist in our retrospective study, which should be avoided by prospective studies in the future. Second, the ultrasound images used in this study were from different operators using different machines, and these confounding factors could potentially affect the predictive performance of the model, so the next step requires a prospective cohort study with strictly controlled variables (e.g., one operator, with one machine to acquire images). Third, ultrasound images only take the largest section for analysis. Although it can represent a lesion, it still cannot represent the whole lesion, and information may still be missed. Future studies need to analyze the overall lesions by forming three-dimensional images. Four, the limit number of cases is also one of our limitations. We need further research with large samples and multicenter data in the future.

Conclusions

In summary, we have developed a prediction model for endorectal ultrasound-based radiomics, which is expected to be used to predict the presence of RC LVI. This method can be promoted in clinical practice because it can noninvasively predict the presence or absence of LVI before surgery and thus influence clinical decision-making.
exCISION of recTAL cancer: a word of caution. J Nati Compr Canc Netw. 2018;16(7):822–8.

4. Lee L, de Lacy B, Gomez Ruiz M, Liberman AS, Albert MR, Monson JRT, Lacy A, Kim SH, Atallah SB. A multicenter matched comparison of transanal and robotic total mesorectal excision for mid and low rectal adenocarcinoma. Ann Surg. 2019;270(6):1110–6.

5. Bonnetain F, Bossert JF, Gerard JP, Calais G, Conroy T, Mineur L, Bouché O, Maingon P, Chapet O, Radosavljevic L, et al. What is the clinical benefit of preoperative chemoradiotherapy with SFU/leucovorin for T3–4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: surrogacy in question? Eur J Cancer. 2012;48(12):1781–90.

6. Kim YC, Kim JK, Kim MJ, Lee JH, Kim YB, Shin SJ. Feasibility of mesorectal vascular invasion in predicting early distant metastasis in patients with stage T3 rectal cancer based on rectal MRI. Eur Radiol. 2016;26(2):297–305.

7. Smith NJ, Barbachano Y, Norman AR, Swift RI, Abulafi AM, Brown G. Prognostic significance of magnetic resonance imaging-detected extramural vascular invasion in rectal cancer. Br J Surg. 2008;95(10):229–36.

8. Schaap D, Ogura A, Nederend J, Maas M, Crossen JS, Creemers GJ, van Linschoten I, Nieuwenhuijzen GA, Rutten JH, Kusters M. Prognostic implications of MRI-detected lateral nodal disease and extramural vascular invasion in rectal cancer. Br J Surg. 2018;105(13):1844–52.

9. Ogura A, Konishi T, Cunningham C, Garcia-Aguilar J, Iversen H, Toda S, Lee K, Lee HX, Rosara K, Lee B et al. Neoadjuvant (chemo)radiotherapy with total mesorectal excision only is not sufficient to prevent local recurrence in enlarged nodes: results of the multicenter lateral node subseries of patients with low cT3/4 rectal cancer. J Clin Oncol. 2019;37(1):33–43.

10. 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(2):367–87.

11. Van Calster B, Wynants L, Verbeeck JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol. 2018;74(6):796–804.

12. Chan BP, Patel R, Mbaugbaw L, Thabane L, Yaghoubi M. EUS versus magnetic resonance imaging in staging rectal adenocarcinoma: a diagnostic test accuracy meta-analysis. Gastrointest Endosc. 2019;90(2):196–203.

13. Xu Q, Xu Y, Sun H, Jiang T, Xie S, Ooi BY, Ding Y. MRI evaluation of complete response of locally advanced rectal cancer after neoadjuvant therapy: current status and future trends. Cancer Manag Res. 2021;13:4317–28.

14. Davey MS, Davey MG, Ryan EJ, Hogan AM, Kerin MJ, Joyce M. The use of radiomic analysis of magnetic resonance imaging in predicting distant metastases of rectal carcinoma following surgical resection: a systematic review and meta-analysis. Colorectal Dis. 2021;23(12):3065–72.

15. Hou M, Sun JH. Emerging applications of radiomics in rectal cancer: state of the art and future perspectives. World J Gastroenterol. 2021;27(25):3802–14.

16. Stanzione A, Verde F, Romeo V, Roccafuoco F, Mainenti PP, Maurea S. Radiomics and machine learning applications in rectal cancer: current update and future perspectives. World J Gastroenterol. 2021;27(25):3802–14.

17. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, Ma ZL, Liu ZY. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol. 2016;34(18):2157–64.

18. Ge YX, Wu XB, Wang Z, Zhang QJ, Zhou YY, Duan SF, Hu SD, Fei BJ. Prognostic value of CT radiomics in evaluating lymphovascular invasion in rectal cancer: diagnostic performance based on different volumes of interest. J Xray Sci Technol. 2021;29(4):663–74.

19. Zhang Y, He K, Guo Y, Liu X, Yang Q, Zhang C, Xie Y, Mu S, Guo X, Fu Y, et al. A novel multimodal radiomics model for preoperative prediction of lymphovascular invasion in rectal cancer. Front Oncol. 2020;10:457.

20. Yang YS, Qiu YJ, Zhang GH, Gong HP, Ge YQ, Zhang YF, Feng F, Wang YT. High resolution MRI-based radiomicnomogram in predicting perineural invasion in rectal cancer. Cancer Imaging. 2021;21(14):1.

21. Tiberiamehick H, Rouhanet P, Sbrara M, Fourghani R, Reinhold C, Nougaret S, Group GS. Radiomics modelling in rectal cancer to predict disease-free survival: evaluation of different approaches. Br J Surg. 2021;108(10):1243–50.