



OPEN ACCESS

EDITED BY

Frauke Alves,
Max Planck Institute for Experimental
Medicine, Germany

REVIEWED BY

Simone Maurea,
University of Naples Federico II, Italy
Stefano Trebeschi,
The Netherlands Cancer Institute (NKI),
Netherlands

*CORRESPONDENCE

Guohua Wang
✉ wangguohua89@163.com

RECEIVED 01 August 2023

ACCEPTED 10 October 2023

PUBLISHED 24 October 2023

CITATION

Qu X, Zhang L, Ji W, Lin J and Wang G
(2023) Preoperative prediction of tumor
budding in rectal cancer using multiple
machine learning algorithms based on
MRI T2WI radiomics.

Front. Oncol. 13:1267838.

doi: 10.3389/fonc.2023.1267838

COPYRIGHT

© 2023 Qu, Zhang, Ji, Lin and Wang. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Preoperative prediction of tumor budding in rectal cancer using multiple machine learning algorithms based on MRI T2WI radiomics

Xueting Qu¹, Liang Zhang², Weina Ji², Jizheng Lin²
and Guohua Wang^{1*}

¹Department of Radiology, Qingdao Municipal Hospital, Qingdao University, Qingdao, Shandong, China, ²Department of Radiology, the Affiliated Hospital of Qingdao University, Qingdao, Shandong, China

Objective: This study aimed to explore the radiomics model based on magnetic resonance imaging (MRI) T2WI and compare the value of different machine learning algorithms in preoperatively predicting tumor budding (TB) grading in rectal cancer.

Methods: A retrospective study was conducted on 266 patients with preoperative rectal MRI examinations, who underwent complete surgical resection and confirmed pathological diagnosis of rectal cancer. Among them, patients from Qingdao West Coast Hospital were assigned as the training group (n=172), while patients from other hospitals were assigned as the external validation group (n=94). Regions of interest (ROIs) were delineated, and image features were extracted and dimensionally reduced using the Least Absolute Shrinkage and Selection Operator (LASSO). Eight machine learning algorithms were used to construct the models, and the diagnostic performance of the models was evaluated and compared using receiver operating characteristic (ROC) curves and the area under the curve (AUC), as well as clinical utility assessment using decision curve analysis (DCA).

Results: A total of 1197 features were extracted, and after feature selection and dimension reduction, 11 image features related to TB grading were obtained. Among the eight machine learning models, the support vector machine (SVM) algorithm achieved the best diagnostic performance, with accuracy, sensitivity, and specificity of 0.826, 0.949, and 0.723 in the training group, and 0.713, 0.579, and 0.804 in the validation group, respectively. DCA demonstrated the clinical utility of this radiomics model.

Conclusion: The radiomics model based on MR T2WI can provide an effective and noninvasive method for preoperative TB grading assessment in patients with rectal cancer.

KEYWORDS

tumor budding 3, rectal cancer, magnetic resonance imaging, algorithm, radiomics

1 Introduction

Colorectal cancer (CRC) ranks second in terms of cancer-related deaths in the United States and is a leading cause of death for men under the age of 50 (1). The incidence of rectal cancer accounts for approximately a third of all cases (2). The treatment and prognosis assessment of rectal cancer primarily rely on TNM staging. However, patients with the same pathological stage may exhibit significantly different clinical treatment outcomes (3). Exploring additional factors that influence tumor prognosis can provide better guidance and treatment plans for patients, optimize patient management, and improve survival quality (4–6). It has been shown that perineural invasion, tumor deposits, and tumor budding are all adverse prognostic factors (5, 7, 8).

Tumor budding (TB), as a process of epithelial-mesenchymal transition in tumors (9, 10), was recommended by the International Tumor Budding Consensus Conference (ITBCC) in 2016 to be included in the guidelines/protocols and staging systems for rectal cancer pathology reporting. It also provided a definition for tumor budding (11). TB refers to the presence of scattered tumor cells or small clusters of tumor cells with poor differentiation at the invasive front of the tumor, observed under high-power microscopy. These cell groups are often single tumor cells or comprise less than 5 cells. TB is categorized into three grades based on the number of buds: low-grade budding (Budding1, Bd1) with 0–4 buds, intermediate-grade budding (Bd2) with 5–9 buds, and high-grade budding (Bd3) with ≥ 10 buds (within a 20x visual field, corresponding to an area of 0.785 mm²). High-grade budding has been significantly associated with reduced overall survival rate, adverse clinical-pathological features, invasive growth patterns, lymph node metastasis and tumor recurrence rate (4, 8) (11–14). The AJCC/UICC cancer staging guidelines have included tumor budding as an additional prognostic factor for rectal cancer (9). However, as a postoperative pathological characteristic, it cannot provide optimal assistance for preoperative surgical planning. Therefore, exploring preoperative tumor budding assessment methods is beneficial for implementing more accurate personalized treatment and improving tumor prognosis in a clinical setting.

The diagnosis of rectal cancer plays a crucial role in medical imaging, among which MRI has been recommended as the preferred method for diagnosing rectal cancer due to its excellent soft tissue resolution (3, 6, 15). However, traditional imaging diagnostics still have their limitations. With the advancement of scientific technology, artificial intelligence (AI) has also entered the field of medicine. Multiple studies have shown that AI-based imaging learning models can better predict tumor treatment response and prognosis without relying on pathological reports, thereby supporting accurate determination of clinical treatment plans (16–18). Radiomics is one of the AI imaging learning models (19) that has been used for preoperative TN staging prediction, gene typing, assessment of high-risk tissue histopathological variables, treatment efficacy, and prognosis evaluation in tumors (20–22). Current research on radiomics in rectal cancer mainly focuses on predicting aspects such as complete remission after neoadjuvant therapy for advanced rectal cancer, lymph node metastasis, KRAS/NRAS gene mutations,

and microsatellite instability (MSI). Few studies have reported on predicting tumor budding. In recent years, some scholars have begun to explore this area (23, 24). Currently, only Li et al. have conducted a study using T2WI, DWI, and contrast-enhanced imaging to predict tumor budding degree, with respective values for AUC, sensitivity, specificity, and accuracy of 0.796, 92.7%, 65.8%, and 81.2% (25). However, some patients are unable to undergo contrast-enhanced examinations due to issues such as contrast agent allergies, renal function abnormalities, or economic constraints. Therefore, this study aims to utilize various machine learning models based on T2WI radiomics to predict the degree of tumor budding in rectal cancer preoperatively. It also seeks to explore clinical risk factors in order to provide assistance in selecting preoperative surgical approaches for rectal cancer and improving patient prognosis.

2 Materials and methods

2.1 Participants

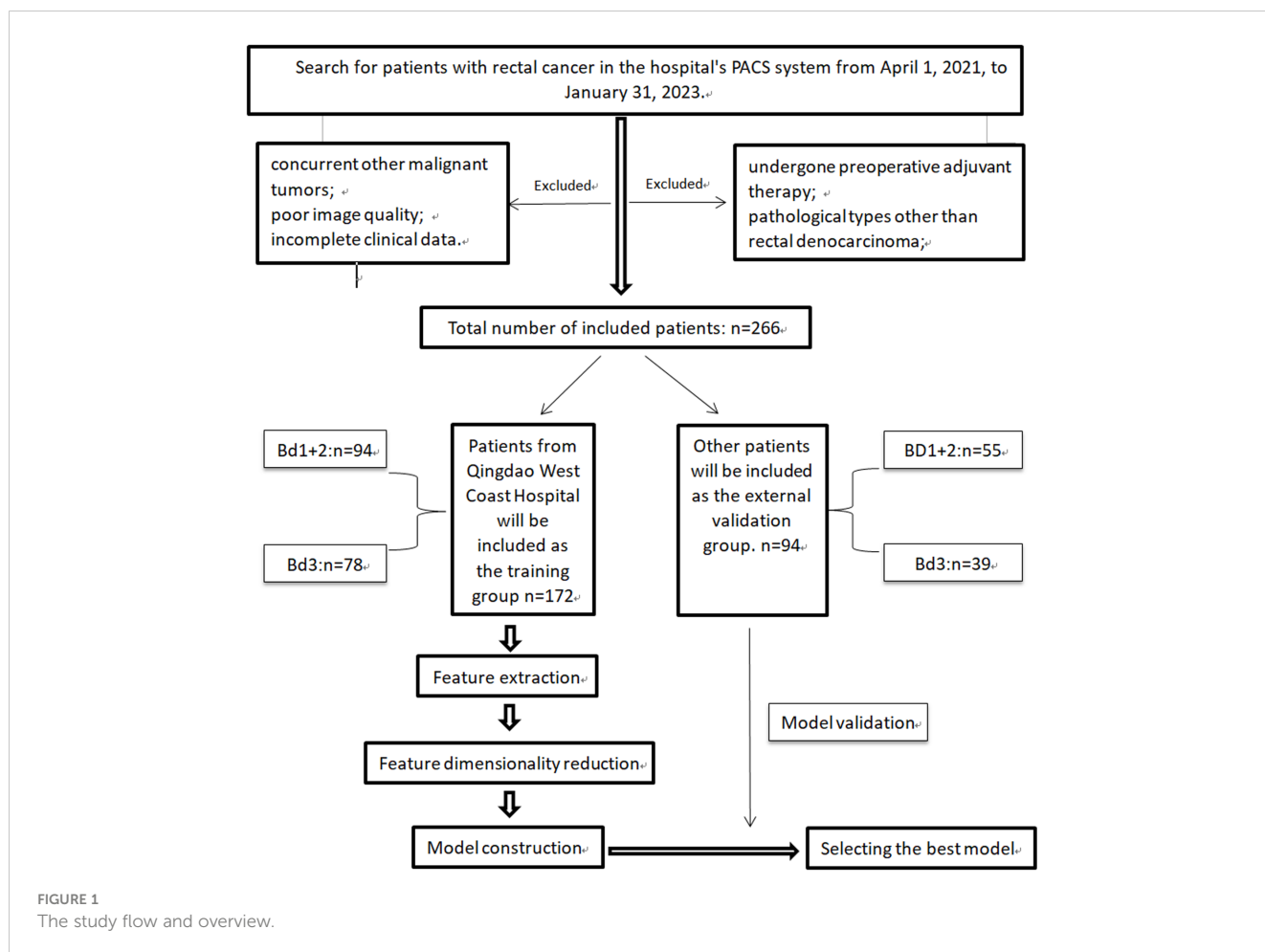
Patients who underwent rectal cancer surgery and parallel MRI scans from April 2021 to April 2023 were included in this study. The study flow and overview are shown in Figure 1. Patients from Qingdao West Coast Hospital were assigned as the training group, while patients from Qingdao Municipal Hospital and Qingdao University Affiliated Hospital were assigned as the validation group. The inclusion criteria were as follows (1): Pathological diagnosis of rectal adenocarcinoma with tumor budding grading (2); All patients underwent surgical treatment (3); Preoperative MRI scans performed within 2 weeks (4); Distance of the tumor's lower margin from the anal verge less than 15cm (26) (5); Complete clinical data. The exclusion criteria were as follows (1): Patients who received preoperative adjuvant therapies such as radiotherapy or immunotherapy (2); Exclusion of patients with mucinous adenocarcinoma (3); Patients with concurrent other malignant tumors (4); Patients with poor image quality (5); Patients with incomplete clinical data.

2.2 Clinical data

Clinical data collection primarily includes age, gender, carcinoembryonic antigen (CEA), carbohydrate antigen-199 (CA 199), distance from the tumor to the anal verge, length, depth of infiltration, circumferential growth, MR TN staging, distant metastasis, circumferential resection margin (CRM), and extramural vascular invasion (EMVI).

2.3 Image acquisition

All patients underwent imaging using 3.0T MR scanners after fasting for at least 4 hours. Conventional axial T2WI images were used to establish the radiomics model in this study. Detailed scanning parameters are provided in Supplementary Table 1.



2.4 Pathological evaluation

Avoid bias in tumor budding assessment, samples obtained by complete resection of the tumor lesion during surgery were used. The excised samples were stained with Hematoxylin and Eosin (H&E). According to the ITBCC (2016) recommendations and the AJCC/UICC Cancer Staging Guidelines (8th edition, 2017), the pathology report needs to include TN staging, tumor budding grading, pathological type, degree of differentiation, perineural invasion, vascular emboli, circumferential resection margin (CRM). This is necessary for data comparison and analysis in the study. In this research, tumor budding degree was categorized into the intermediate-low-grade group (Bd 1&2) and the high-grade group (Bd 3).

2.5 Image segmentation

The DICOM raw images were imported into the open-source software 3D Slicer (<https://www.slicer.org/>, version 5.0.3). One attending physician with 8 years of experience in abdominal MRI diagnosis, serving as the annotating physician, performed lesion segmentation on the axial T2WI images in conjunction with the

T2WI sagittal and DWI sequences. The regions of interest were manually outlined along the tumor edges layer by layer, and the software platform automatically reconstructed the volumes of interest (VOI). During the outlining process, one attending physician performs image segmentation along the outermost edge of the tumor visible in the T2WI images, while another senior physician reviews the ROI. When the tumor involves adjacent solid organs, draw along the maximum edge of the tumor to exclude the normal structures of other organs as much as possible. It is difficult to determine whether the nodule around the tumor is a tumor nodule or lymph node metastasis through imaging. When the nodule is closely connected to the tumor and the angle formed by the edges of the nodule and the tumor is obtuse, we believe that the nodule is a part of the tumor. On the contrary, if it is an acute angle, the nodule should be excluded. Any disagreements were resolved through consultation and discussion by two senior physicians. The following considerations were taken into account: (a) Exclusion of the intestinal lumen and its contents; (b) Exclusion of adjacent mesenteric fat tissue; (c) Exclusion of the uninvolved rectal wall. Furthermore, to evaluate the stability of all image features, 30 randomly selected cases were independently segmented by the two physicians again after a one-month interval. Inter- and intraobserver consistency evaluations were conducted (25). The

image segmentation process was performed in a single-blind manner, with the two physicians unaware of the tumor budding grading information of the patients.

2.6 Radiomics feature extraction and model construction

2.6.1 Image preprocessing

The segmented images obtained from the image segmentation process were imported into the Pyradiomics package (<https://github.com/Radiomics/pyradiomics>). Due to the presence of various scanning devices and imaging parameters, image preprocessing was required. The following steps were followed for image preprocessing (21, 27): Step 1: N4 bias field correction, which aims to eliminate magnetic field bias or intensity inhomogeneity in the images. Step 2: Resampling of the images using B-spline interpolation to standardize the voxel size to 1mm^3 . Step 3: As per the Image Biomarker Standardization Initiative (IBSI) guidelines (28), grayscale normalization was performed by scaling the gray values of MRI images based on $\mu \pm 3\sigma$ (μ : mean gray level within the VOI; σ : gray standard deviation) to minimize MRI signal intensity variations. This normalization step helps eliminate the influence of different field strengths and device parameters on the images. Image preprocessing reduces the effects caused by variations in image acquisition and parameters, ensuring the generalizability and robustness of the constructed models.

2.6.2 Feature extraction

After image preprocessing, image feature extraction for the training group is performed on the aforementioned platform. Features can be divided into three groups: (I) first-order statistics, (II) shape attributes, and (III) texture features. First-order statistics are features extracted directly from the raw data or image, typically including basic statistical measures to provide a basic description of the data. Shape attributes describe the external shape and structure of an object or region, relating to its contour, boundaries, and geometric properties. They are useful in distinguishing different objects, detecting edges and contours, and performing object recognition tasks. Texture features describe the second-order and higher-order spatial distribution of patterns or intensities. They provide information about the texture in the image, such as features related to repetition, variation, and local structure. In this case, several different methods are used for texture feature extraction, including Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), Gray-Level Size Zone Matrix (GLSZM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray-Level Dependence Matrix (GLDM).

Normalization using the z-score method was performed to ensure that all features have a mean of 0 and a standard deviation (SD) of 1, as described in previous studies (24). Feature selection, dimensionality reduction, and prediction model construction were conducted using the feature data from the training group. The

feature data from the external validation group were used to validate the effectiveness and performance of the constructed models.

2.6.3 Feature selection and dimensionality reduction

To ensure the robustness of the selected radiomics features, feature dimensionality reduction is performed. Firstly, the interclass/intraclass correlation coefficients (ICCs) are calculated. Features with ICC values greater than 0.75 are selected to exclude those with significant inter-observer/intra-observer variability. Statistical tests such as t-tests or Mann-Whitney U tests are then applied to retain features with p-values less than 0.05. To determine the correlation between features that exhibit high repetitiveness, the Spearman's rank correlation coefficient can be used. In this method, we calculate the correlation coefficient between features and retain one feature from any pair with a correlation coefficient greater than 0.9 (29, 30). To filter features and retain their descriptive capability to the maximum extent, we can use a greedy recursive elimination strategy. We can employ a nonlinear regression algorithm called LASSO regression, which utilizes the minimum absolute shrinkage and selection operator. By adjusting the weight parameter λ , we can shrink all regression coefficients towards zero. Through 10-fold cross-validation, we can find the optimal value for λ , and retain the features with non-zero coefficients, combining them into a radiomics signature. Finally, we can utilize the scikit-learn package in Python to build the LASSO regression model and calculate the radiomics score (Rad-score) for each patient by linearly combining the retained features. Prediction model building.

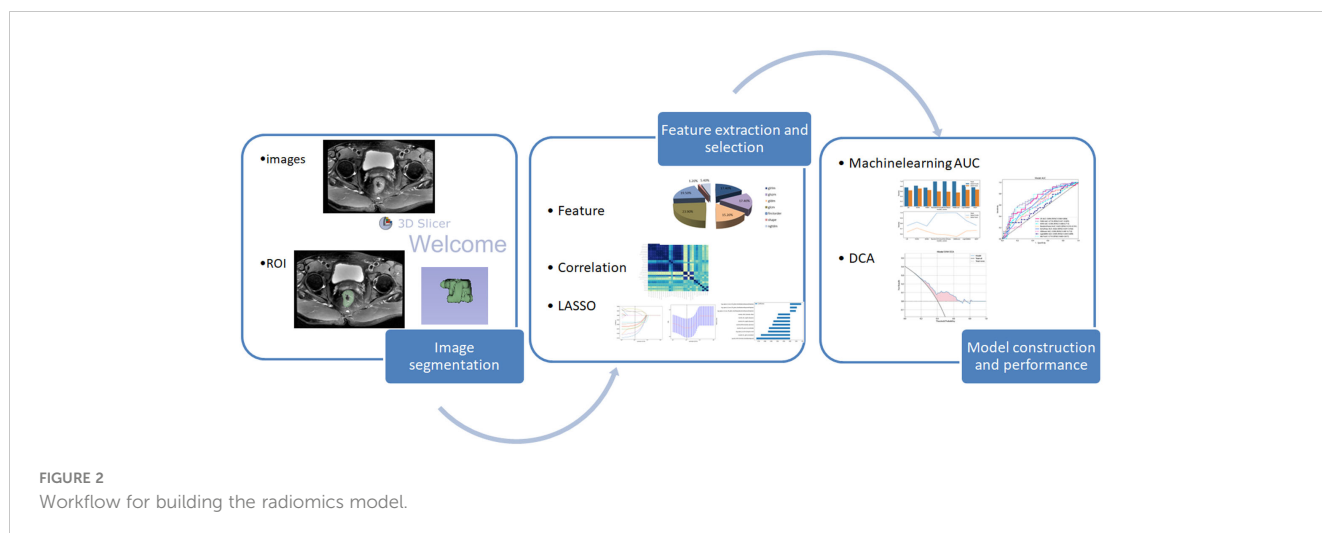
The selected radiomic features are individually used with eight classification algorithms: Logistic Regression (LR), K-Nearest Neighbor (KNN), Extra Trees (ET), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), Light Gradient Boosting Machine (LightGBM), and Multilayer Perceptron (MLP). These eight algorithms are used to build models. The best machine learning algorithm model is selected by calculating the area under the ROC curve. 5-fold cross-validation is employed to obtain the final Rad signature. Please refer to Figure 2 for the radiomics workflow.

2.7 Clinical risk factors

Statistical analysis was performed on clinical variables (age, gender, CEA, CA199 levels), tumor distance from the anal verge, length, depth of invasion, circumferential growth, MRI TN stage, distant metastasis, circumferential margin, and extramural vascular invasion. Independent sample t-tests, Mann-Whitney U-tests, or χ^2 tests were used to analyze these variables. Clinically significant risk factors that exhibited a statistically significant difference ($P < 0.05$) in predicting tumor budding grade were identified.

2.8 Statistical analysis

Statistical analysis was performed using Python 3.8 (<https://www.python.org>) and SPSS 19.0 software. Continuous data in



patient clinical information were tested for normality and homogeneity of variance. For comparisons, either t-tests or rank-sum tests were used. Categorical data were compared using the chi-square test. In the training group, radiomic features were first transformed to a normal distribution using regularization. Then, t-tests (when homogeneity of variance) and Mann-Whitney U-tests (when heterogeneity of variance) were conducted to compare the features and identify statistically significant differences. Single-factor and multi-factor logistic regression analyses were performed to evaluate clinical risk factors associated with high-grade tumor budding. Factors with a significance level of $P < 0.05$ were considered statistically significant. ROC and AUC were used to evaluate the predictive performance of each model in the training group and the external validation group. DCA was used to evaluate the clinical application value of the model.

3 Results

3.1 Characteristics of the patients

3.1.1 Clinical features

A total of 266 cases were included in the study, with 172 cases in the training group and 94 cases in the external validation group. The clinical characteristics of all patients in the training and external validation groups are shown in [Supplementary Table 2](#). Among the training group, 78 cases (45.35%) were classified as high-grade TB, while 38 cases (40.43%) in the external validation group were classified as high-grade TB. No clinical risk factors related to high-grade TB were found ($P > 0.05$).

3.1.2 Radiomics features

A total of 1197 features were extracted, detailed information can be found in [Table 1](#); [Figure 3](#). They are mainly divided into three categories (1): First-order statistics ($n=234$ cases) (2); Shape attributes ($n=14$ cases) (3); Texture features ($n=949$ cases), including GLCM, GLRLM, GLSZM, GLDM, and NGTDM.

Features were selected based on group-level and intra-group $ICC > 0.75$. The t-test or Mann-Whitney U test was applied for screening, selecting features with $P < 0.05$, resulting in 29 features being retained. The Spearman rank correlation coefficient was then used to keep only one feature from each pair of features with a correlation coefficient greater than 0.9. Next, a greedy recursive elimination strategy was employed to remove the feature with the highest redundancy at each step, resulting in only 23 features being retained. Pearson correlation coefficient was used to further screen and keep 16 features. LASSO was applied to select features with coefficients greater than 0, resulting in a final retention of 11 features, as shown in [Figure 4](#). The detailed feature names and their corresponding Rad-score values can be found in [Figure 5](#). The calculation formula of Rad-score is as follows:

$$\begin{aligned} \text{Rad-score} = & 0.45348837209300424 - 0.069649 \cdot \log_sigma_4_0_mm_3D_glcm_Imc2 \\ & + 0.021591 \cdot \log_sigma_4_0_mm_3D_glszm_SmallAreaLowGrayLevelEmphasis \\ & + 0.018396 \cdot \log_sigma_5_0_mm_3D_gldm_SmallDependenceLowGrayLevelEmphasis \\ & + 0.035346 \cdot \log_sigma_5_0_mm_3D_glszm_SmallAreaLowGrayLevelEmphasis \\ & - 0.038032 \cdot \text{wavelet_HHH_firstorder_Mean} \\ & - 0.106451 \cdot \text{wavelet_HHH_firstorder_RootMeanSquared} \\ & - 0.055620 \cdot \text{wavelet_HHH_firstorder_Skewness} \end{aligned}$$

TABLE 1 Feature extraction and classification.

Feature Type	Subtype	Number of Features
Firstorder	-	234
Texture	GLCM	286
	GLDM	182
	GLRLM	208
	GLSZM	208
	NGTDM	65
Shape	-	14
Total	-	1197

-, no value.

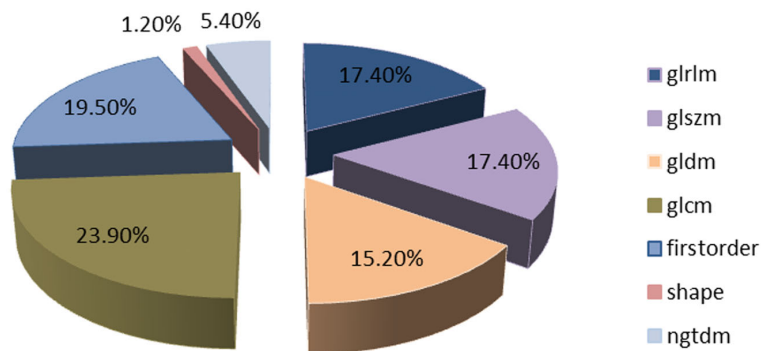


FIGURE 3
Number and ratio of handcrafted features.

3.2 Construction and diagnostic performance comparison of radiomics models

Radiomics models were constructed using eight different classification algorithms. The AUC values for the training group ranged from 0.807 to 1.000, while for the external validation group, they ranged from 0.569 to 0.718. Among all the algorithms, the SVM algorithm demonstrated the best diagnostic performance, with AUC values of 0.892 and 0.718 for the training and external validation groups, respectively (95% CI: 0.8413-0.9420, 0.6111-0.8246). Furthermore, the SVM model showed excellent performance in both the training and external validation groups, achieving better diagnostic accuracy with sensitivity, specificity, and accuracy of 57.9%, 80.4%, and 71.3%, respectively. Detailed data can be found in Table 2; Figure 6 illustrates the accuracy of all models and the AUC values of each rad signature model on the test set. The DCA demonstrated that the SVM model yielded the high net benefit for predicting rectal cancer TB.

4 Discussion

Due to the heterogeneity of tumors, the clinical treatment outcomes and prognosis of rectal cancer can vary significantly.

Multiple studies have explored various prognostic factors for tumors, such as histological type, infiltration pattern, differentiation degree, MSI, Aquaporins (AQP-1) (4, 5) thrombus formation, peripheral nerve invasion, tumor deposits, tumor budding, etc. However, most of these factors are based on invasive pathological reports and cannot be identified macroscopically. It poses a great challenge to non-invasive examination methods to determine the pathological prognostic factors of tumors before surgery. The emergence of new technologies like radiomics have provided possibilities in this regard. Many scholars have conducted value assessments of preoperative diagnosis of pathological prognostic factors using radiomics. For example, Li et al. found that a multi-modal MRI radiomics model can predict MSI in rectal cancer preoperatively (AUC values of 0.78 in both the training and validation sets) (31). Chen et al. found that a radiomics model and its nomogram can improve the predictive efficiency of AQP-1 overexpression (AUC values above 0.80 in both the training and validation sets) (32).

As a novel tumor prognostic factor, TB is gradually being recognized for its importance in personalized cancer treatment management. High levels of TB are considered significantly associated with adverse outcomes in patients (10, 33). The AJCC/ UICC Cancer Staging Guidelines have included TB as an additional prognostic factor for rectal cancer (8, 9).

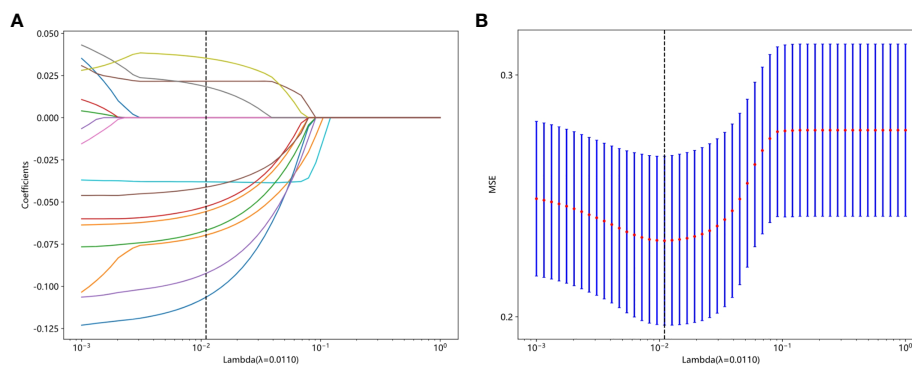
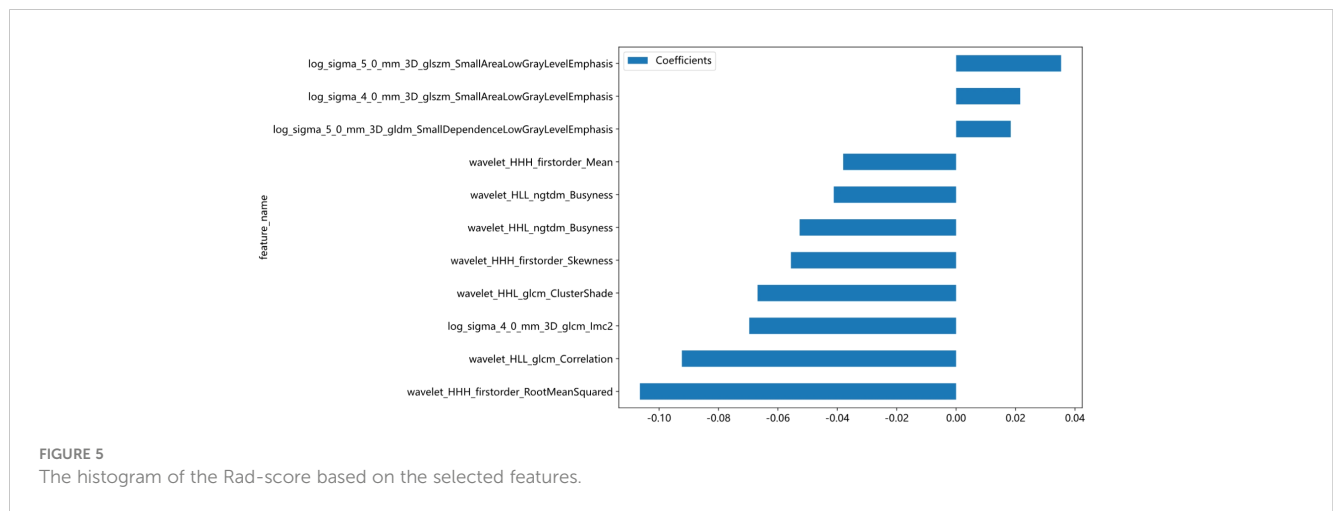


FIGURE 4
(A) The coefficient analysis of LASSO feature selection under 10-fold cross-validation, (B) MSE of 10 fold cross validation.



Currently, there are few reported studies on the preoperative non-invasive assessment of TB grading. Chen et al. utilized b-value threshold maps and ADC maps to evaluate the preoperative predictive value of TB grading and found that b-value threshold maps performed better than ADC values in assessing TB grading (34). However, b-value threshold maps are not part of the routine MRI imaging sequence, and their image resolution is inferior to T2WI sequences. Due to their status as special sequences, their application is limited. Li et al. developed a radiomics model based on MR plain and enhanced scans to predict TB grading, with an AUC value of 0.79625, demonstrating the feasibility of radiomics in preoperative TB grading prediction. However, in clinical practice,

the model’s applicability is somewhat limited by the inability of some patients to undergo the complete sequence acquisition due to issues such as examination costs, contrast agent allergies, and impaired renal function.

The T2WI sequence, as a routine sequence in MRI examinations, not only provides clearer visualization of lesion details compared to b-value threshold maps but also has lower scanning costs compared to dynamic contrast-enhanced sequences. It does not require consideration of factors such as impaired renal function or contrast agent allergies, making it more clinically applicable. Moreover, radiomics models can identify pathological features that are indistinguishable to the naked eye through

TABLE 2 Performance of machine learning classifiers for predicting TB status.

Model name	Task	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV
LR	train	0.756	0.807	0.7423 - 0.8720	0.756	0.755	0.720	0.789
LR	test	0.638	0.694	0.5840 - 0.8041	0.789	0.545	0.536	0.789
SVM	train	0.826	0.892	0.8413 - 0.9420	0.949	0.723	0.740	0.944
SVM	test	0.713	0.718	0.6111 - 0.8246	0.579	0.804	0.667	0.738
KNN	train	0.750	0.841	0.7863 - 0.8962	0.731	0.766	0.722	0.774
KNN	test	0.649	0.596	0.4796 - 0.7126	0.316	0.925	0.632	0.653
RF	train	0.994	0.999	0.9969 - 1.0000	1.000	0.989	0.987	1.000
RF	test	0.596	0.625	0.5102 - 0.7408	0.605	0.600	0.500	0.687
ET	train	1.000	1.000	1.0000 - 1.0000	1.000	1.000	1.000	1.000
ET	test	0.585	0.652	0.5413 - 0.7618	0.789	0.455	0.492	0.758
XGBoost	train	1.000	1.000	1.0000 - 1.0000	1.000	1.000	1.000	1.000
XGBoost	test	0.553	0.596	0.4810 - 0.7116	0.842	0.364	0.471	0.769
LightGBM	train	0.849	0.908	0.8653 - 0.9511	0.859	0.840	0.817	0.878
LightGBM	test	0.660	0.569	0.4496 - 0.6886	0.289	0.911	0.687	0.654
MLP	train	0.762	0.828	0.7674 - 0.8889	0.872	0.670	0.687	0.863
MLP	test	0.670	0.710	0.6029 - 0.8173	0.789	0.600	0.566	0.805

AUC, area under curve, CI, confidence interval, PPV, positive predictive value, NPV, negative predictive value, RF, RandomForest, ET, ExtraTrees.

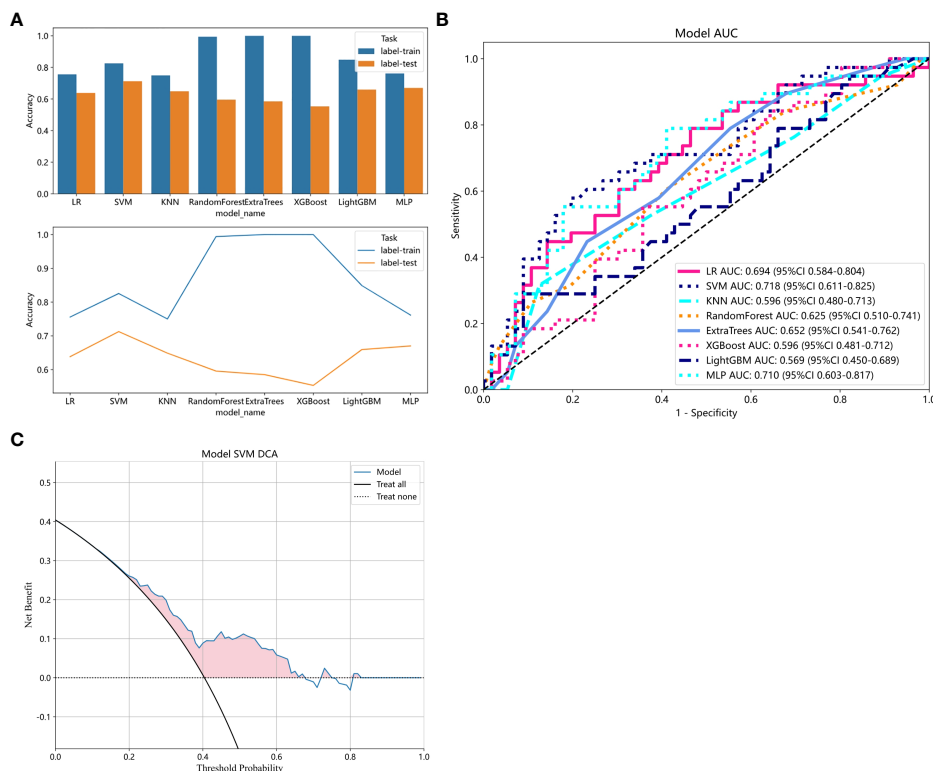


FIGURE 6 Comparison of accuracy among all models. **(B)** ROC analysis of different models on the validation set. **(C)** The decision curve of the SVM model. The y-axis represents standardized net benefit. The blue curve represents the radiomics model. The black curve represents the hypothetical scenario where all patients are successfully predicted. The black dashed line represents the hypothetical scenario where no patients are successfully predicted.

complex machine algorithms. Therefore, this study is based on the T2WI sequence and utilizes various machine learning methods to explore the optimal radiomics model for predicting tumor budding. This research aims to provide effective assistance in developing appropriate treatment plans and improving patient prognosis.

In this study, we established multiple radiomics predictive models based on T2WI and found that the SVM model exhibited a higher AUC value in predicting TB in rectal cancer. This indicates that the predictive model has good accuracy and stability. As a non-invasive examination method, it can provide assistance in preoperative treatment planning and prognosis assessment, contributing to personalized medical management for rectal cancer patients. The accuracy, sensitivity, and specificity of this model in the training group were 0.826, 0.949, and 0.723, respectively. In the validation group, these values were 0.713, 0.579, and 0.804, respectively. The model showed relatively lower sensitivity but higher specificity and accuracy, which may be attributed to the limited number of high-grade TB samples in the external validation group. In our research, we aimed to enhance the model’s credibility by selecting an external validation group while choosing radiomics verification methods. We then screened out 11 key features and constructed a radiomics model for predicting TB grading. Among these features, texture features were the most abundant (8/11), which is related to the arrangement and distribution patterns of medical image pixels. In MR images, the differences in tissue composition and internal tumor structures provide more detailed variations in local texture features. This is also

one of the reasons for using a large number of texture features. Moreover, texture features encompass various types, including GLCM, NGTDM, GLDM, etc. Multiple methods can be employed during feature extraction to obtain different types of texture features, which enhances prediction accuracy. By calculating all key features using eight algorithms, we found that the SVM algorithm yielded the highest diagnostic performance in the constructed predictive model (AUC=0.718). This may be due to the ability of the SVM algorithm to avoid overfitting issues to some extent, as it demonstrates robustness, stability, and good generalization ability, which are advantageous for high-dimensional data. The results of this study indicate that the radiomics model based on T2WI can accurately predict TB grading in rectal cancer, which is consistent with the findings of Li et al. (25). This could be attributed to the ability of radiomics to quantitatively analyze and link TB grading with the extracted image features, thus capturing the tumor heterogeneity factor that is indistinguishable to the naked eye through complex computational methods (35, 36).

No clinically significant risk factors related to TB grading ($p > 0.05$) were found in this study, which is consistent with the conclusions of most studies (7, 34, 37). However, there is controversy surrounding these results, and one study has indicated that age is a clinical risk factor. This discrepancy may be attributed to differences in the sampled population and sample size that could be collected.

The DCA indicates that the SVM model has the highest net clinical benefit, suggesting that the radiomics model developed may help in the formulation of personalized treatment management plans

for patients. This finding is similar to the results of Li et al., who used a radiomics model based on multimodal MRI to predict tumor budding in locally advanced rectal cancer (25). Both studies utilized MRI-based radiomics models to predict tumor heterogeneity factors that cannot be identified by the naked eye. Furthermore, both studies demonstrated that the models can predict the TB grade. This may be attributed to the fact that radiomics models are based on medical imaging examinations, which provide digitized features of the tumor's location. These models can identify tumor characteristics beyond the range of visual observation and imperceptible to the naked eye, and utilize machine learning algorithms to analyze, model, and predict these digitized features, thereby revealing the comprehensive situation of tumor growth and spread.

This study still has limitations. Firstly, as a retrospective study, it only included patients with rectal cancer who underwent curative surgery and had postoperative pathology indicating adenocarcinoma, resulting in selection bias. Secondly, there was sample bias, with a limited number of high-grade TB samples in the external validation group, leading to lower sensitivity. In future studies, it is necessary to increase the sample size. Thirdly, the area of interest was manually delineated, and subjective factors in the delineation process may have had some impact on the analysis, leading to subjective errors in the results. Fourthly, In terms of image segmentation, we devised our own exclusion criteria based on our experience regarding whether nodules closely connected to the tumor should be excluded as lymph node metastases. Although this exclusion criterion may not be completely rational, it improves the reproducibility in image segmentation. Fifthly, bowel preparation was not performed, which may have affected the quality and accuracy of the images.

In conclusion, our preliminary study suggests that the radiomics model based on T2WI can improve the preoperative prediction of tumor budding grading in rectal cancer and provide valuable reference for individualized treatment planning in clinical practice.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Qingdao Municipal Hospital Ethics Committee and Qingdao University Affiliated Hospital Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because This study is a retrospective study.

Author contributions

XQ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft. LZ: Data curation, Formal Analysis, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. WJ: Data curation, Funding acquisition, Writing – review & editing. JL: Conceptualization, Writing – review & editing. GW: Writing – review & editing, Conceptualization, Funding acquisition, Supervision, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

Thank you to the radiology technology teams at the hospital for providing rectal MRI scans for this research, and to the ONEKEY technology team for providing comprehensive support in data analysis for this work. Additionally, We would like to express our sincere gratitude to Mr. Zhang Zhiwei from the ONEKEY team for his assistance in the field of statistics.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1267838/full#supplementary-material>

References

1. Siegel RL, Wagle NS, Cercck A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin* (2023) 73(3):233–54. doi: 10.3322/caac.21772
2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin* (2023) 73(1):17–48. doi: 10.3322/caac.21763
3. Stanzione A, Verde F, Romeo V, Boccadifuoco F, Mainenti PP, Maurea S. Radiomics and machine learning applications in rectal cancer: current update and future perspectives. *World J Gastroenterol* (2021) 27(32):5306–21. doi: 10.3748/wjg.v27.i32.5306
4. Chen K, Collins G, Wang H, Toh JWT. Pathological features and prognostication in colorectal cancer. *Curr Oncol* (2021) 28(6):5356–83. doi: 10.3390/curroncol28060447
5. Chen J, Chen Y, Zheng D, Pang P, Zhang H, Zheng X, et al. Pretreatment mr-based radiomics nomogram as potential imaging biomarker for individualized assessment of perineural invasion status in rectal cancer. *Abdominal Radiol (New York)* (2021) 46(3):847–57. doi: 10.1007/s00261-020-02710-4
6. Fernandes MC, Gollub MJ, Brown G. The importance of mri for rectal cancer evaluation. *Surg Oncol* (2022) 43:101739. doi: 10.1016/j.suronc.2022.101739
7. Shin JK, Park YA, Huh JW, Yun SH, Kim HC, Lee WY, et al. Tumor budding as a prognostic marker in rectal cancer patients on propensity score analysis. *Ann Surg Oncol* (2021) 28(13):8813–22. doi: 10.1245/s10434-021-10286-6
8. Riffet M, Dupont B, Faisant M, Cerasuolo D, Menahem B, Alves A, et al. New histoprognostic factors to consider for the staging of colon cancers: tumor deposits, invasive tumor infiltration and high-grade budding. *Int J Mol Sci* (2023) 24(4):3573. doi: 10.3390/ijms24043573
9. Karjula T, Kemi N, Niskakangas A, Mustonen O, Puro I, Pohjanen VM, et al. The prognostic role of tumor budding and tumor-stroma ratio in pulmonary metastasis of colorectal carcinoma. *Eur J Surg Oncol* (2023) 49(7):1298–1306. doi: 10.1016/j.ejso.2023.02.009
10. Lugli A, Zlobec I, Berger MD, Kirsch R, Nagtegaal ID. Tumor budding in solid cancers. *Nat Rev Clin Oncol* (2021) 18(2):101–15. doi: 10.1038/s41571-020-0422-y
11. El Agy F, El Bardai S, Bouguenouch L, Lahmidani N, El Abkari M, Benjelloun EB, et al. Prognostic impact of tumor budding on moroccan colon cancer patients. *Int J Surg Oncol* (2022) 2022:9334570. doi: 10.1155/2022/9334570
12. Liu S, Zhang Y, Ju Y, Li Y, Kang X, Yang X, et al. Establishment and clinical application of an artificial intelligence diagnostic platform for identifying rectal cancer tumor budding. *Front Oncol* (2021) 11:626626. doi: 10.3389/fonc.2021.626626
13. Demir A, Alan O, Oruc E. Tumor budding for predicting prognosis of resected rectum cancer after neoadjuvant treatment. *World J Surg Oncol* (2019) 17(1):50. doi: 10.1186/s12957-019-1588-6
14. Cho SJ, Kakar S. Tumor budding in colorectal carcinoma: translating a morphologic score into clinically meaningful results. *Arch Pathol Lab Med* (2018) 142(8):952–7. doi: 10.5858/arpa.2018-0082-RA
15. Wang PP, Deng CL, Wu B. Magnetic resonance imaging-based artificial intelligence model in rectal cancer. *World J Gastroenterol* (2021) 27(18):2122–30. doi: 10.3748/wjg.v27.i18.2122
16. Cui Y, Yang W, Ren J, Li D, Du X, Zhang J, et al. Prognostic value of multiparametric mri-based radiomics model: potential role for chemotherapeutic benefits in locally advanced rectal cancer. *Radiother Oncol J Eur Soc Ther Radiol Oncol* (2021) 154:161–9. doi: 10.1016/j.radonc.2020.09.039
17. Cai C, Hu T, Gong J, Huang D, Liu F, Fu C, et al. Multiparametric mri-based radiomics signature for preoperative estimation of tumor-stroma ratio in rectal cancer. *Eur Radiol* (2021) 31(5):3326–35. doi: 10.1007/s00330-020-07403-6
18. Li ZY, Wang XD, Li M, Liu XJ, Ye Z, Song B, et al. Multi-modal radiomics model to predict treatment response to neoadjuvant chemotherapy for locally advanced rectal cancer. *World J Gastroenterol* (2020) 26(19):2388–402. doi: 10.3748/wjg.v26.i19.2388
19. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
20. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: A systematic review and meta-analysis. *BMC Cancer* (2021) 21(1):1058. doi: 10.1186/s12885-021-08773-w
21. Jing G, Chen Y, Ma X, Li Z, Lu H, Xia Y, et al. Predicting mismatch-repair status in rectal cancer using multiparametric mri-based radiomics models: A preliminary study. *BioMed Res Int* (2022) 2022:6623574. doi: 10.1155/2022/6623574
22. Cui Y, Wang G, Ren J, Hou L, Li D, Wen Q, et al. Radiomics features at multiparametric mri predict disease-free survival in patients with locally advanced rectal cancer. *Acad Radiol* (2022) 29(8):e128–e38. doi: 10.1016/j.acra.2021.11.024
23. Chong GO, Park SH, Jeong SY, Kim SJ, Park NJ, Lee YH, et al. Prediction model for tumor budding status using the radiomic features of F-18 fluorodeoxyglucose positron emission tomography/computed tomography in cervical cancer. *Diagnostics (Basel)* (2021) 11(8):1517. doi: 10.3390/diagnostics11081517
24. Chong GO, Park SH, Park NJ, Bae BK, Lee YH, Jeong SY, et al. Predicting tumor budding status in cervical cancer using mri radiomics: linking imaging biomarkers to histologic characteristics. *Cancers (Basel)* (2021) 13(20):5140. doi: 10.3390/cancers13205140
25. Li Z, Chen F, Zhang S, Ma X, Xia Y, Shen F, et al. The feasibility of mri-based radiomics model in presurgical evaluation of tumor budding in locally advanced rectal cancer. *Abdominal Radiol (New York)* (2022) 47(1):56–65. doi: 10.1007/s00261-021-03311-5
26. Salama E, Holland J, Boutros M. Surgical principles of rectal cancer. *Surg Oncol Clinics North America* (2022) 31(2):239–53. doi: 10.1016/j.soc.2021.11.005
27. Yu X, Song W, Guo D, Liu H, Zhang H, He X, et al. Preoperative prediction of extramural venous invasion in rectal cancer: comparison of the diagnostic efficacy of radiomics models and quantitative dynamic contrast-enhanced magnetic resonance imaging. *Front Oncol* (2020) 10:459. doi: 10.3389/fonc.2020.00459
28. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145
29. Shen J, Du H, Wang Y, Du L, Yang D, Wang L, et al. A novel nomogram model combining ct texture features and urine energy metabolism to differentiate single benign from Malignant pulmonary nodule. *Front Oncol* (2022) 12:1035307. doi: 10.3389/fonc.2022.1035307
30. Huang Y, Zhu T, Zhang X, Li W, Zheng X, Cheng M, et al. Longitudinal mri-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: A multicenter, retrospective study. *EClinicalMedicine* (2023) 58:101899. doi: 10.1016/j.eclinm.2023.101899
31. Li Z, Zhang J, Zhong Q, Feng Z, Shi Y, Xu L, et al. Development and external validation of a multiparametric mri-based radiomics model for preoperative prediction of microsatellite instability status in rectal cancer: A retrospective multicenter study. *Eur Radiol* (2023) 33(3):1835–43. doi: 10.1007/s00330-022-09160-0
32. Chen Y, Li B, Jiang Z, Li H, Dang Y, Tang C, et al. Multi-parameter diffusion and perfusion magnetic resonance imaging and radiomics nomogram for preoperative evaluation of aquaporin-1 expression in rectal cancer. *Abdominal Radiol (New York)* (2022) 47(4):1276–90. doi: 10.1007/s00261-021-03397-x
33. Zlobec I, Berger MD, Lugli A. Tumor budding and its clinical implications in gastrointestinal cancers. *Br J Cancer* (2020) 123(5):700–8. doi: 10.1038/s41416-020-0954-z
34. Chen F, Zhang S, Ma X, Chen Y, Wang Z, Zhu Y, et al. Prediction of tumor budding in patients with rectal adenocarcinoma using B-value threshold map. *Eur Radiol* (2023) 33(2):1353–63. doi: 10.1007/s00330-022-09087-6
35. Yang YS, Qiu YJ, Zheng GH, Gong HP, Ge YQ, Zhang YF, et al. High resolution mri-based radiomic nomogram in predicting perineural invasion in rectal cancer. *Cancer Imaging* (2021) 21(1):40. doi: 10.1186/s40644-021-00408-4
36. Mainenti PP, Stanzione A, Guarino S, Romeo V, Uggia L, Romano F, et al. Colorectal cancer: parametric evaluation of morphological, functional and molecular tomographic imaging. *World J Gastroenterol* (2019) 25(35):5233–56. doi: 10.3748/wjg.v25.i35.5233
37. Jurescu A, Dema A, Vaduva A, Gheju A, Vita O, Barna R, et al. Poorly differentiated clusters and tumor budding are important prognostic factors in colorectal carcinomas. *Bosn J Basic Med Sci* (2022) 22(2):164–77. doi: 10.17305/bjbm.2021.6110