



An MSCT-based radiomics nomogram combined with clinical factors can identify Crohn's disease and ulcerative colitis

Hui Li¹, Yan Mo², Chencui Huang², Qingguo Ren¹, Xiaona Xia¹, Xiaomin Nan¹, Xinyan Shuai¹, Xiangshui Meng¹

¹Department of Radiology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, Qingdao, China; ²Deepwise AI Lab, Beijing Deepwise & League of PHD Technology Co., Ltd., Beijing, China

Contributions: (I) Conception and design: H Li, X Meng; (II) Administrative support: X Meng; (III) Provision of study materials or patients: Y Mo, H Li; (IV) Collection and assembly of data: H Li, C Huang; (V) Data analysis and interpretation: Y Mo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiangshui Meng, MD. Department of Radiology, Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, 758 Hefei Road, Qingdao 266035, China. Email: qdqlfksk2021@163.com.

Background: We established and evaluated a radiomics nomogram based on multislice computed tomography (MSCT) arterial phase contrast-enhanced images to distinguish between Crohn's disease (CD) and ulcerative colitis (UC) objectively, quantitatively, and reproducibly.

Methods: MSCT arterial phase-enhancement images of 165 lesions (99 CD, 66 UC) in 87 patients with inflammatory bowel disease (IBD) confirmed by endoscopy or surgical pathology were retrospectively analyzed. A total of 132 lesions (80%) were selected as the training cohort and 33 lesions (20%) as the test cohort. A total of 1648 radiomic features were extracted from each region of interest (ROI), and the Pearson correlation coefficient and tree-based method were used for feature selection. Five machine learning classifiers, including logistic regression (LR), support vector machine (SVM), random forest (RF), stochastic gradient descent (SGD), and linear discriminative analysis (LDA), were trained. The best classifier was evaluated and obtained, and the results were transformed into the Rscore. Three clinical factors were screened out from 8 factors by univariate analysis. The logistic regression method was used to synthesize the significant clinical factors and the Rscore to generate the nomogram, which was compared with the clinical model and LR model.

Results: Among all machine learning classifiers, LR performed the best (AUC =0.8077, accuracy =0.697, sensitivity =0.8, specificity =0.5385), SGD model had the second best performance (AUC =0.8, accuracy =0.6667, sensitivity =0.75, specificity =0.5385), and the DeLong test results showed that there was no significant difference between LR and SGD ($P=0.465>0.05$), while the other models performed poorly. Texture features had the greatest impact on classification results among all imaging features. The significant features of the LR model were used to calculate the Rscore. The 3 significant clinical factors were perienteric edema or inflammation, CT value of arterial phase-enhancement (AP-CT value), and lesion location. Finally, a nomogram was constructed based on the 3 significant clinical factors and the Rscore, whose AUC (0.8846) was much higher than that of the clinical model (0.6154) and the LR model (0.8077).

Conclusions: The nomogram is expected to provide a new auxiliary tool for radiologists to quickly identify CD and UC.

Keywords: Crohn's disease (CD); ulcerative colitis (UC); nomogram; machine learning; computerized tomography (CT)

Submitted Jan 28, 2021. Accepted for publication Apr 04, 2021.

doi: 10.21037/atm-21-1023

View this article at: <http://dx.doi.org/10.21037/atm-21-1023>

Introduction

Inflammatory bowel disease (IBD) is a group of non-specific chronic and recurrent inflammatory diseases of the intestine mediated by abnormal immunity caused by the interaction of many factors, such as environment, infection, immunity, and genetics, amongst others (1). It mainly encompasses Crohn's disease (CD) and ulcerative colitis (UC). It has a lifelong tendency to relapse, showing a chronic process of repeated alternating infection and remission, and is called "green cancer". Moreover, IBD has evolved into a global disease with rising prevalence in every continent (2-4).

Since the diagnosis of IBD is multimodal, there is no single diagnostic gold standard for IBD (1,5). Endoscopy combined with pathology is currently recognized as the first-line investigation for the diagnosis of IBD. However, this procedure is complicated and invasive, with poor repeatability (6,7). Moreover, endoscopy can only observe mucosal lesions and the infiltrating depth of the intestinal wall, and extraenteric complications cannot be diagnosed. Furthermore, biopsy specimens are also extremely limited, thus the comprehensive assessment of IBD has certain limitations (8). Multislice computed tomography (MSCT) provides better visualization of the entire bowel wall, and the capability of MSCT to depict extraenteric diseases allows for the simultaneous diagnosis of complications associated with IBD. However, both CD and UC can involve the terminal ileum, colon, and rectum, and the CT signs have a certain degree of overlap, making differential diagnosis complicated. Individual patients often require specific therapeutic strategies, and personalized and precise treatment of IBD is the focus of research. Therefore, the distinction between CD and UC is of pivotal importance for tailored clinical management (9,10). The premise of precise treatment is based on precise diagnosis. To achieve precise imaging, it is necessary to break through the traditional medical imaging model based on morphology and semi-quantitative analysis. Radiomics can extract countless quantitative features from medical images with high-throughput calculations, and output objective classifications and diagnosis models through feature screening and machine learning algorithms, which is known as the bridge between medical imaging and personalized medicine (11). The application of radiomics for gastrointestinal diseases has achieved good research results in the diagnosis of small bowel tumors, colon cancer, rectal cancer, and other lesions (12-15), which greatly improves the diagnostic accuracy. However, very few studies on the diagnosis of IBD have

been reported (16,17). This study further explored the effectiveness of radiomics methods for the differential diagnosis of CD and UC, in order to accurately distinguish them quantitatively, objectively, and reproducibly.

We present the following article in accordance with the TRIPOD reporting checklist (available at <http://dx.doi.org/10.21037/atm-21-1023>).

Methods

Subject selection and patients

This retrospective study collected the clinical data of patients with IBD between July 2014 and September 2020 in Qilu Hospital (Qingdao), Cheeloo University of Medicine, Shandong University. In this study, quality control of the original images of patients was performed according to the standard of the "expert guidance on imaging examination and reporting of inflammatory bowel disease in China" (18).

The inclusion criteria of our study were as follows: (I) patients who could cooperate with the examination normally; (II) patients were diagnosed with CD or UC according to generally accepted recommendations (1,19,20); (III) no digestive tract cancer or other serious liver and kidney diseases; (IV) MSCT enhanced images were well tolerated and qualified.

The exclusion criteria were as follows: (I) pregnant women or nursing women, hyperthyroidism or iodine allergy, patients with other serious diseases involving unstable vital signs; patients with mental illness and low cognitive ability who could not cooperate with the examination; (II) patients with digestive tract malignant tumors or other severe liver and kidney disease; (III) unqualified images, such as missing, fuzzy, and incomplete images; (IV) images involving the distal ileum and outside the colon and rectum.

The inclusion and exclusion criteria of the data set in this study are detailed in *Figure 1*. A total of 87 patients with IBD confirmed by colonoscopy or enteroscopy and pathology were included. Among them, there were 61 CD patients (45.444±18.701 years old, 26 women and 35 men) and 26 UC patients (44.030±17.983 years old, 10 women and 16 men). All patients underwent an MSCT enhancement scan before and after endoscopy.

In order to develop and verify the performance of different machine learning models (LRs), 80% of the data were used as training data (n=132) and the remaining

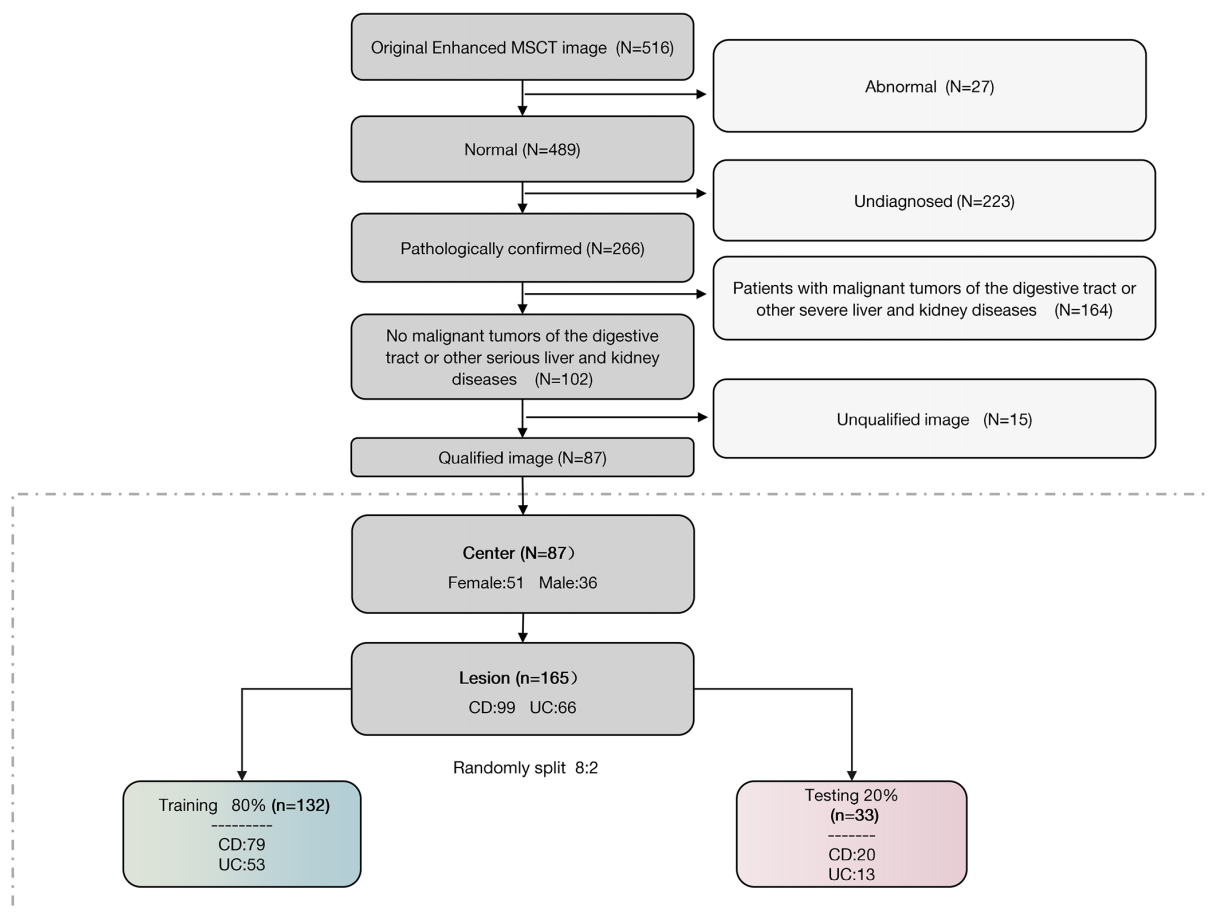


Figure 1 Process of inclusion and exclusion of images. MSCT, multislice computed tomography; CD, Crohn's disease; UC, ulcerative colitis.

20% as test data ($n=33$) using the random split method. In addition, clinical factor analysis and the logistic regression method were used to construct the nomogram classification prediction models for CD and UC, which integrated clinical factors and radiomic features.

Image data acquisition

On the day before CT examination, patients were given a solid free diet. Eight hours before the scan, patients fasted and took laxatives to clean the intestines, and took 2,000–3,000 mL 2.5% isotonic mannitol (400–500 mL every 15 minutes). Patients were injected with 10 mg anisodamine 20 minutes before the examination, and were scanned 10–15 minutes after the last oral mannitol to ensure the full expansion of the gastrointestinal tracts.

Enhanced scanning was performed with SOMATOM

definition flash CT, and the scanning ranged from the diaphragmatic crest to the pubic symphysis, with a thickness of 5 mm and thin layer reconstruction of 0.75 mm at 120 kV and 110–180 mA. A high pressure syringe was used to inject contrast medium from the anterior cubital vein. The iodine content was 300 mg/mL, the dosage was 60–80 mL, and the injection rate was 3.0–3.5 mL/s. Scanning was performed in the arterial phase (35 s after contrast injection) and intravenous phase (65 s after contrast injection).

Clinical diagnosis

The CT images of arterial phase enhancement were uploaded to the Deepwise multimodal research platform (<https://keyan.deepwise.com>, V1.6.2) for image annotation. The location of the inflamed intestinal segment was divided into the terminal ileum, cecum, ascending colon, transverse

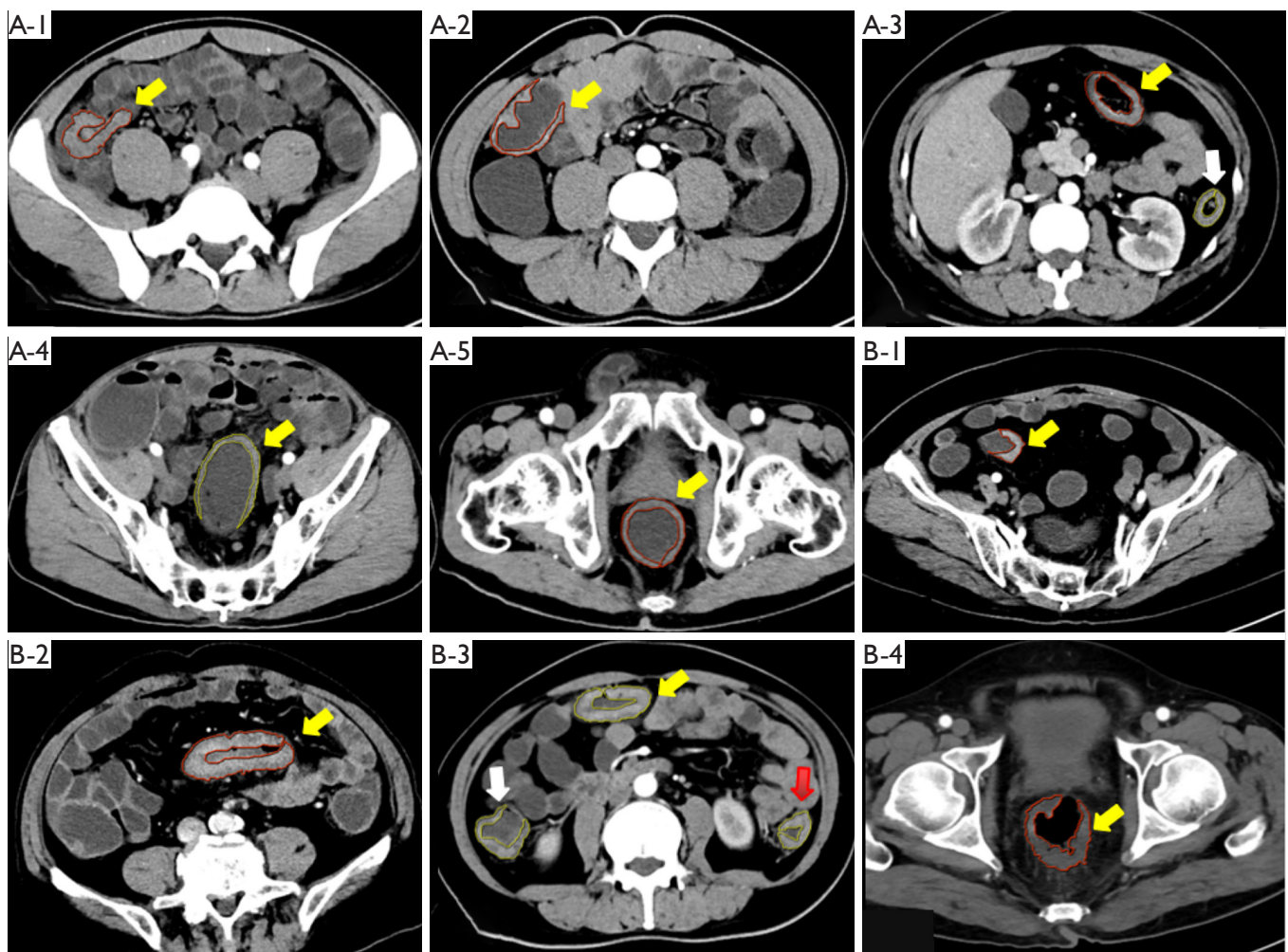


Figure 2 Radiologists drew lesions of different intestinal segments. (A-1) CD lesions located in the terminal ileum (yellow arrow); (A-2) CD lesions located in the ascending colon (yellow arrow); (A-3) CD lesions located in the transverse colon (yellow arrow) and descending colon (white arrow); (A-4) CD lesions located in the sigmoid colon (yellow arrow); (A-5) CD lesions located in the rectum (yellow arrow); (B-1) UC lesions located in the terminal ileum (yellow arrow); (B-2) UC lesions located in the sigmoid colon (yellow arrow); (B-3) UC lesions located in the transverse colon (yellow arrow) and in the ascending colon (white arrow), UC lesions located in the descending colon (red arrow); (B-4) UC lesions located in the rectum (yellow arrow). CD, Crohn's disease; UC, ulcerative colitis.

colon, descending colon, sigmoid colon, and rectum. Radiologists drew the lesions in each segment. The region of interest (ROI) was the lesion wall with segmental or diffuse intestinal thickening and marked enhancement. A radiologist with more than 10 years experience in abdominal diagnosis drew all lesions in the arterial phase-enhancement images, as shown in *Figure 2*, and this was then checked by another radiologist with 10 years experience.

The 8 clinical factors included age, gender, wall thickness, CT value of arterial phase-enhancement (AP-

CT value), perienteric edema or inflammation (increased attenuation of the mesenteric fat), engorged vasa recta (enlarged blood vessels that supply and drain an inflamed bowel loop), lymphadenopathy (the short axis diameter of mesenteric lymph nodes was greater than 1 cm), and lesion location. The wall thickness and AP-CT value were the average values of the 3 measured values of the thickest portion of the most distended segment or the site of the most severe inflammation. The mesenteric lymph nodes were measured in the short axis (21).

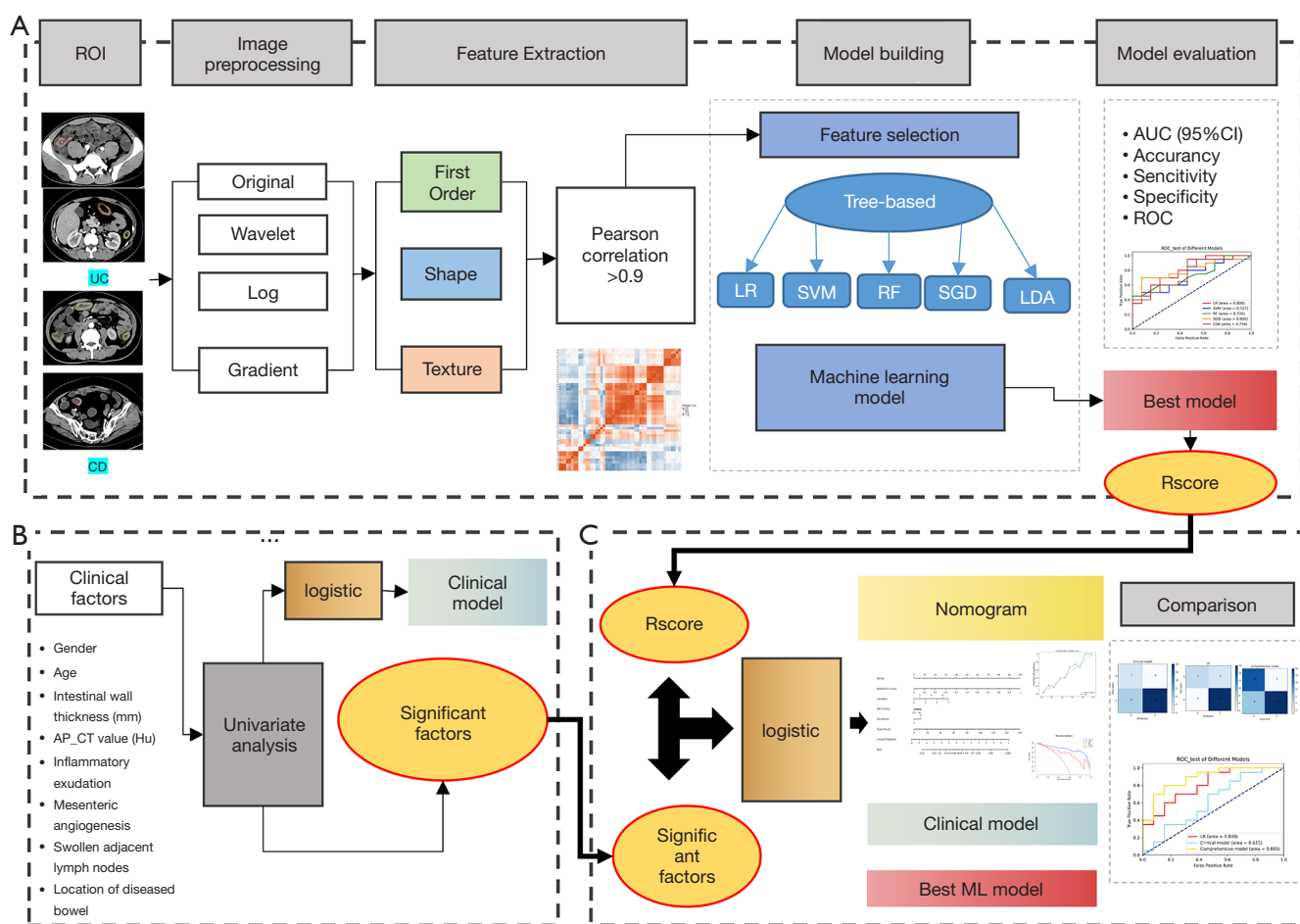


Figure 3 The flow chart showed that radiomics analysis and development of the nomogram, which consists of 3 parts: Part A, establishment of the machine learning model, calculation of Rscore imaging model establishment, Rscore calculation Part A, establishment of the machine learning model and calculation of Rscore imaging model establishment and Rscore calculation; Part B, establishment of the clinical model and screening of significant factors; Part C, establishment, comparison, and evaluation of the radiomics nomogram.

Radiomics analysis and development of the nomogram

As shown in the flow chart in *Figure 3*, the analysis and development of the nomogram consisted of 3 parts: Part A was the training process of the radiomics model, including loading original images and segmentation ROIs, image preprocessing, feature extraction, feature selection, training the machine LR, evaluating the model, and outputting an optimal model to calculate Rscores; Part B was the screening process of clinical factors, including single factor analysis, significant factor output, and the establishment of the logistic clinical model; Part C was the process of establishing a comprehensive nomogram combining clinical and radiomic features. Based on the logistic regression method, the nomogram was constructed using significant

clinical factors and Rscores, and 3 models were evaluated and compared.

Image preprocessing

According to the ROI labeled by the radiologist in the original DICOM image, the volume map of the two-dimensional ROI was obtained for feature extraction and quantification. The B-spline interpolation (22) sampling technique sitkbppline was used to resample the images with different resolutions, so that all the images had the same resolution after resampling [1, 1, 1]. Different image preprocessing methods (wavelet transform, Laplace transform, Gaussian filter transform, etc.) were used for

image transformation.

Feature extraction

The radiomics features of the transformed image were extracted, including the first order features based on the pixel values of the original image or the preprocessed image, the shape features describing the shape of the lesion, and the texture features describing the internal and surface texture of the lesion: gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and gray level dependence matrix (GLDM). A total of 1,648 radiomic features were extracted from each ROI and standardized by Z-score.

The three-level naming method was used to name the features, where the first level referred to the image preprocessing method and specified parameters (such as log-sigma-4-0-mm), the second level represented the feature type (such as firstorder, shape, and GLSZM), and the third level referred to the specific feature method (such as small area low gray level emphasis).

Feature extraction was performed by a comprehensive open source platform called PyRadiomics, which enables the processing and extraction of radiomic features from medical images and is implemented in Python.

Feature selection

Firstly, the Pearson correlation coefficient between all radiology features was calculated using feature correlation analysis. A total of 1,220 high correlation features were removed from the features with correlation coefficients greater than 0.9, and 428 low correlation features were retained. Then, further feature selection was carried out.

In this study, a tree-based feature selection (23) method was used, which utilized the principle of information gain. When the mutual information gain of class and feature in the training data was greater, the information content of the feature was more abundant, which also showed that the feature was more suitable for modeling. Most of the features with excellent classification results were selected, and the appropriate features were selected to make the prediction results more robust and to make the model more generalizable. Finally, the above features were incorporated into the modeling process of the machine LR.

The above method used the open source functions of “sklearn.feature_selection” and “pyradiomics” of the python language.

Machine learning algorithm

In order to obtain the optimal classifier for discriminating between CD and UC, all the data were randomly divided into two groups for training (n=132) and testing (n=33). Five machine learning algorithms (24,25) were tested, which were logistic regression (LR), support vector machine (SVM), random forest (RF), stochastic gradient descent (SGD), and linear discriminative analysis (LDA).

LR is a commonly used binary linear classifier (26). On the basis of LR, the sigmoid function is used for non-linear transformation, and the log maximum likelihood estimation function is used as the loss function to learn the posterior probability of a single sample. The SVM algorithm (27) can be used for linear or non-linear classification, and its purpose is to find the optimal segmentation hyperplane to make the points closer to the hyperplane have a larger distance. The RF algorithm (28) measures the relative importance of each feature for the prediction by looking at how many impure trees in the forest are reduced by using the feature, and automatically calculates the standardized score of each feature after training. The SGD algorithm (29) is an optimization algorithm, where in each update, a sample is randomly used for gradient descent, and the super parameters are adjusted to make the result near the global optimal solution. The LDA (30) algorithm projects the data on the low dimension, and its purpose is to select the projection direction with the best classification performance, so that the projection points of each category data are as close as possible, and the distance between the class centers of different categories of data is as large as possible.

The performance of these 5 classifiers was evaluated by area under the receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, and specificity. The optimal model was then selected to convert the output of the model results into a probability score (Rscore), which indicated the relative risk of CD in the test samples.

Establishment of the radiomics nomogram

Clinical factor analysis

In order to select the clinical factors significantly related to the outcome, single factor analysis was conducted on 8 clinical factors in the study to obtain the clinical factors with significant differences between the CD and UC groups, which were further involved in the establishment of clinical-radiomics nomogram. At the same time, logistic regression

was used to establish a simple clinical classification model and evaluate its performance.

For the purpose of providing a personalized classifier, a nomogram combining radiomic features and significant clinical factors was established to predict the risk of CD. Using the 3 selected clinical factors and the Rscore, the nomogram calibration model was established using the logistic regression method. The calibration curve was used to evaluate the calibration of the nomogram. The performance on the test set and training set was checked by a decision curve, and the ROC curve on the test set was displayed. Finally, the nomogram was compared with the clinical model and radiology model, and the ROC curve, accuracy, sensitivity, specificity, and other indicators on the test set were used to evaluate the performance of the clinical-radiomics nomogram.

Statistical analysis

The Deepwise DxAI platform (<http://dxonline.deepwise.com>) was used for statistical analysis, and the mean, variance, frequency, and percentage were used for statistical description. We tested whether the numerical variables were normally distributed, then used an independent sample *t*-test for normally distributed variables, and the Wilcoxon test for non-normally distributed variables. The chi-square test was used for unordered categorical variables. A two tailed *t*-test was utilized with P value of <0.05 was determined as significant.

Ethical considerations

The study was approved by the ethics committee of Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University. The ethical review committee did not require patients to apply in writing, but ensured that all patient data was processed anonymously. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). Individual consent for this retrospective analysis was waived.

Results

Patient characteristics

Among the 165 lesions included in this study, 99 lesions were CD (intestinal wall thickness 10.041 ± 4.450 mm, AP-

CT value 62.681 ± 15.541 Hu). Among them, 58 (58.6%) showed perienteric edema or inflammation, 68 (68.7%) showed engorged vasa recta, and 54 (54.5%) showed lymphadenopathy. There were 39 lesions located in the terminal ileum (39.4%), 16 in the cecum (16.2%), 11 in the ascending colon (11.1%), 7 in the transverse colon (7.1%), 8 in the descending colon (8.1%), 9 in the sigmoid colon (9.1%), and 9 in the rectum (9.1%). Furthermore, there were 66 UC lesions (intestinal wall thickness 9.597 ± 2.740 mm, AP-CT value 53.864 ± 19.458 Hu), of which 55 (83.3%) showed perienteric edema or inflammation, 48 (72.7%) showed engorged vasa recta, and 41 (62.1%) showed lymphadenopathy. There were 10 lesions located in the terminal ileum (4%), 8 in the cecum (12.3%), 8 in the ascending colon (12.3%), 8 in the transverse colon (12.3%), 11 in the descending colon (16.9%), 12 in the sigmoid colon (18.5%), and 8 in the rectum (12.3%).

In *Table 1*, statistical differences in clinical factors such as age, wall thickness, AP-CT value, perienteric edema or inflammation, engorged vasa recta, lymphadenopathy, and lesion location were analyzed between UC and CD patients.

Finally, we found that there were significant differences between UC and CD in the lesion location, AP-CT value, and perienteric edema or inflammation ($P < 0.05$), while there were no significant differences in other indicators between the two groups ($P \geq 0.05$).

Radiomics model

We extracted 1,648 radiomic features from each ROI region. Firstly, the Pearson correlation coefficient between different features was calculated, and the Pearson correlation heat map of 50 different types of features was generated (see [Figure S1](#)). The closer the color block was to red, the closer the correlation coefficient of features corresponding to the X and Y axis was to 1; the closer the color was to blue, the closer the correlation coefficient between features was to -1; and the closer the color was to white, the closer the correlation coefficient was to 0. Then, 1,200 high correlation features (Pearson coefficient >0.9) were removed and 448 low correlation image features were retained for further feature selection.

After the tree-based feature selection method, 5 machine LRs were trained, and the best performances of all models on the test set were obtained as the representative results. Finally, there were 156, 151, 40, 119, and 40 features in the LR, SVM, RF, SGD, and LDA models respectively. The relative coefficients of the 20 features with the highest

Table 1 Baseline characteristics of patients

Features	Total lesions (n=165)		P value
	UC (n=66)	CD (n=99)	
Gender, n (%)			0.472 ^c
Female	23 (34.8)	40 (40.4)	
Male	43 (65.2)	59 (59.6)	
Age (years ± SD)	44.030±17.983	45.444±18.701	0.630 ^a
Thickness of intestinal wall (mm)	9.597±2.740	10.041±4.450	0.894 ^b
AP-CT value (Hu)	53.864±19.458	62.681±15.541	0.002 ^{b**}
Perienteric edema or inflammation, n (%)	55 (83.3)	58 (58.6)	0.001 ^{c**}
Engorged vasa recta, n (%)	48 (72.7)	68 (68.7)	0.578 ^c
Lymphadenopathy, n (%)	41 (62.1)	54 (54.5)	0.335 ^c
Lesion location, n (%)			0.014 ^{c**}
Terminal ileum	10 (15.4)	39 (39.4)	
Cecum	8 (12.3)	16 (16.2)	
Ascending colon	8 (12.3)	11 (11.1)	
Transverse colon	8 (12.3)	7 (7.1)	
Descending colon	11 (16.9)	8 (8.1)	
Sigmoid colon	12 (18.5)	9 (9.1)	
Rectum	8 (12.3)	9 (9.1)	

^a, independent sample *t*-test; ^b, Wilcoxon test; ^c, Chi-square test; **, *P*<0.05. CD, Crohn's disease; UC, ulcerative colitis; AP-CT value, CT value of arterial phase-enhancement.

weight coefficients in the LR, SVM, RF, SGD, and LDA models were determined (see [Figure S2A,B,C,D,E](#)), and the proportions of the first-order, shape, and texture features in different models among the 20 important modeling features were shown (see [Figure S2F](#)). Texture features were found to have the greatest impact on classification results, followed by first-order features, and finally shape features.

Table 2 shows the performance of 5 machine learning classifiers on the test set and training set. The LR model showed the best performance [AUC (95% CI), 0.8077 (0.6553–0.96), accuracy =0.697, sensitivity =0.8, specificity =0.5385]. The SGD model had the second best performance [AUC (95% CI), 0.8 (0.646–0.954), accuracy =0.6667, sensitivity =0.75, specificity =0.5385]. The LDA, RF, and SVM models performed poorly. The performance details of all models are shown in *Table 2*. Among them, the RF and SVM models were over-fitted, and had strong learning ability in the training set but poor performance in the test set.

The ROC curves and AUCs of the 5 models on the test set and training set were determined (see [Figure S3](#)). The LR model was found to have performed better on the test set, and all models performed well on the training set. The DeLong test results showed that there was no significant difference between LR and SGD (*P*=0.465>0.05), but there was a significant difference between LR and other models (*P*<0.05). The difference between the remaining models was not statistically significant (*P*>0.05), and the ROC curves of all models in the training set were not statistically significant (*P*>0.05). Therefore, combining the results of *Table 2* and the DeLong test, it was found that the LR classifier performed best.

Finally, the 20 features with the highest weights in the LR model (see [Figure S4](#)) were selected, and the correlation coefficient between each feature and outcome was used to convert the results into a probability score, namely, a radiomic score (Rscore), which was used to represent the risk of CD.

Table 2 Performance of 5 machine learning models (LR, SVM, RF, SGD, LDA), the clinical model, and the nomogram on the test set and training set

Model	Test set (n=33)				Training set (n=132)			
	AUC (95% CI)	Accuracy	Sensitivity	Specificity	AUC (95% CI)	Accuracy	Sensitivity	Specificity
LR	0.8077 (0.6553–0.96)	0.697	0.8	0.5385	0.9879 (0.9719–1.0)	0.9621	0.9615	0.963
SVM	0.7269 (0.5528–0.901)	0.6061	0.6	0.6154	0.9998 (0.9991–1.0)	0.9848	0.9744	1
RF	0.7346 (0.5641–0.9052)	0.64	0.8	0.3077	1 (0.999–1.0)	1	1	1
SGD	0.8 (0.646–0.954)	0.6667	0.75	0.5385	0.9898 (0.9785–1.0)	0.9394	0.9615	0.9074
LDA	0.7538 (0.5844–0.9233)	0.7273	0.85	0.5385	0.9219 (0.8778–0.966)	0.8182	0.8846	0.7222
Clinical model	0.6154 (0.4052–0.8255)	0.5758	0.6	0.5385	0.7542 (0.671–0.8373)	0.6742	0.641	0.7222
Nomogram	0.8846 (0.7678–1.0)	0.7879	0.864	0.9231	0.9608 (0.9332–0.9884)	0.9939	0.8846	0.9794

LR, logistic regression; SVM, support vector machine; RF, random forest; SGD, stochastic gradient descent; LDA, linear discriminative analysis.

The radiomic score was attained with the following formula:

“Radiomic score = 0.3887*wavelet-HHL_glszm_GrayLevelVariance
+0.3376*log-sigma-2-0-mm-3D_firstorder_Kurtosis
+0.3285*wavelet-LLL_glcm_Imc2
+0.3155*exponential_glrmlm_ShortRunEmphasis
+0.3024*log-sigma-5-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis+0.2955*wavelet-HHL_firstorder_Kurtosis
+0.292*logarithm_gldm_LargeDependenceLowGrayLevelEmphasis
+0.2898*wavelet-LLL_glszm_LowGrayLevelZoneEmphasis
+0.2653*squareroot_glszm_LargeAreaEmphasis+0.2602*logarithm_glcm_Imc2
-0.2508*wavelet-HHH_firstorder_Median
-0.2522*wavelet-LHL_gldm_SmallDependenceHighGrayLevelEmphasis
-0.2555*original_shape_Maximum2DDiameterSlice
-0.2557*log-sigma-4-0-mm-3D_glszm_SmallAreaEmphasis
-0.2749*gradient_firstorder_Minimum
-0.2868*wavelet-LLL_gldm_LargeDependenceHighGrayLevelEmphasis
-0.3343*original_shape_Sphericity
-0.3376*log-sigma-5-0-mm-3D_glcm_Idn
-0.3718*gradient_glrmlm_RunLengthNonUniformity
-0.4156*square_glszm_SmallAreaLowGrayLevelEmphasis”

Clinical logistic regression model

Based on the univariate analysis in *Table 1*, we found that there were 3 clinical factors associated with UC and CD: AP-CT value, perienteric edema or inflammation, and lesion location. Firstly, we used the logistic regression method to construct a clinical classifier which only used clinical factors to classify UC and CD. The regression coefficient and odds ratio (OR) value of each factor were determined (see *Table S1*). Secondly, we incorporated these 3 significant clinical factors into the establishment of a radiomics nomogram.

In this study, UC was assigned as “0” and CD was assigned as “1”. The results in *Table 2* showed that the AP-CT value was positively correlated with CD. The higher the AP-CT value, the more likely the occurrence of CD (OR =1.024>1), and the results were statistically significant (P=0.022<0.05). There was a negative correlation between extraenteric inflammatory exudation and CD, indicating that UC was more likely to occur when there was extraenteric inflammatory exudation (OR =0.381<1). All intestinal segments were negatively correlated with CD, but the results were not statistically significant (OR <1, P>0.05).

Nomogram

Combined with clinical factors and machine learning (Rscore), a personalized comprehensive nomogram was constructed to predict the risk of CD. The nomogram is shown in *Figure 4A*, where “radiomic score” represents the Rscore, “location” represents the location of the diseased

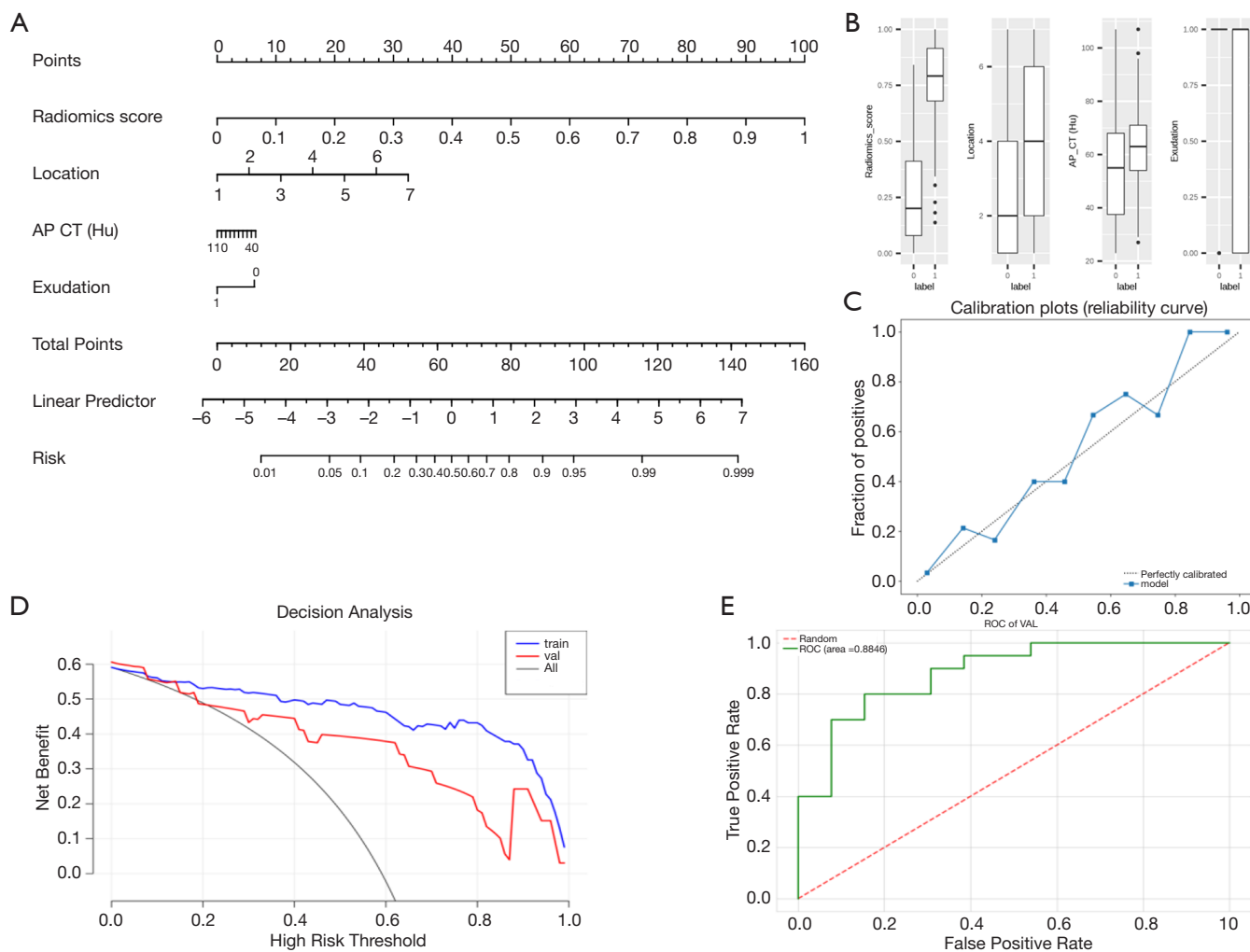


Figure 4 This diagram showed the personalized comprehensive nomogram was constructed and performance of nomogram. (A) Nomogram; (B) distribution of different factors among groups; (C) calibration curve; (D) decision curve; (E) ROC curve on the test set. ROC, receiver operating characteristic.

intestinal segment (1, distal ileum; 2, cecum; 3, ascending colon; 4, transverse colon; 5, descending colon; 6, sigmoid colon; and 7, rectum), “AP-CT (HU)” represents the CT value of arterial phase-enhancement, and “exudation” represents the perienteric edema or inflammation (yes 1, no 0). The Rscore plus each clinical factor score was equal to the total score. According to the total score of the ROI of the CT image, the probability value of the current lesion being CD was “risk”. Thus, a nomogram combining clinical factors and radiomics scores was established.

Figure 4B shows the distribution of different factors in the classification results of the comprehensive model (UC 0, CD 1). Figure 4C predicts the calibration of the nomogram

using a calibration curve. Figure 4D shows the model performance on the training set and test set using a decision curve. Figure 4E shows the ROC curve of the nomogram on the test set and the AUC (0.8846).

Furthermore, Table 2 compares the performances of the clinical model, the machine learning classifier, and the nomogram on the same training set and test set. The results showed that the nomogram had the best performance [AUC (95% CI), 0.8846 (0.7678–1.0), accuracy =0.7879, sensitivity =0.864, specificity =0.9231], followed by the LR model [AUC (95% CI), 0.8077 (0.6553–0.96), accuracy =0.697, sensitivity =0.8, specificity =0.5385], while the clinical model was the worst [AUC (95% CI), 0.6154 (0.4052–0.8255),

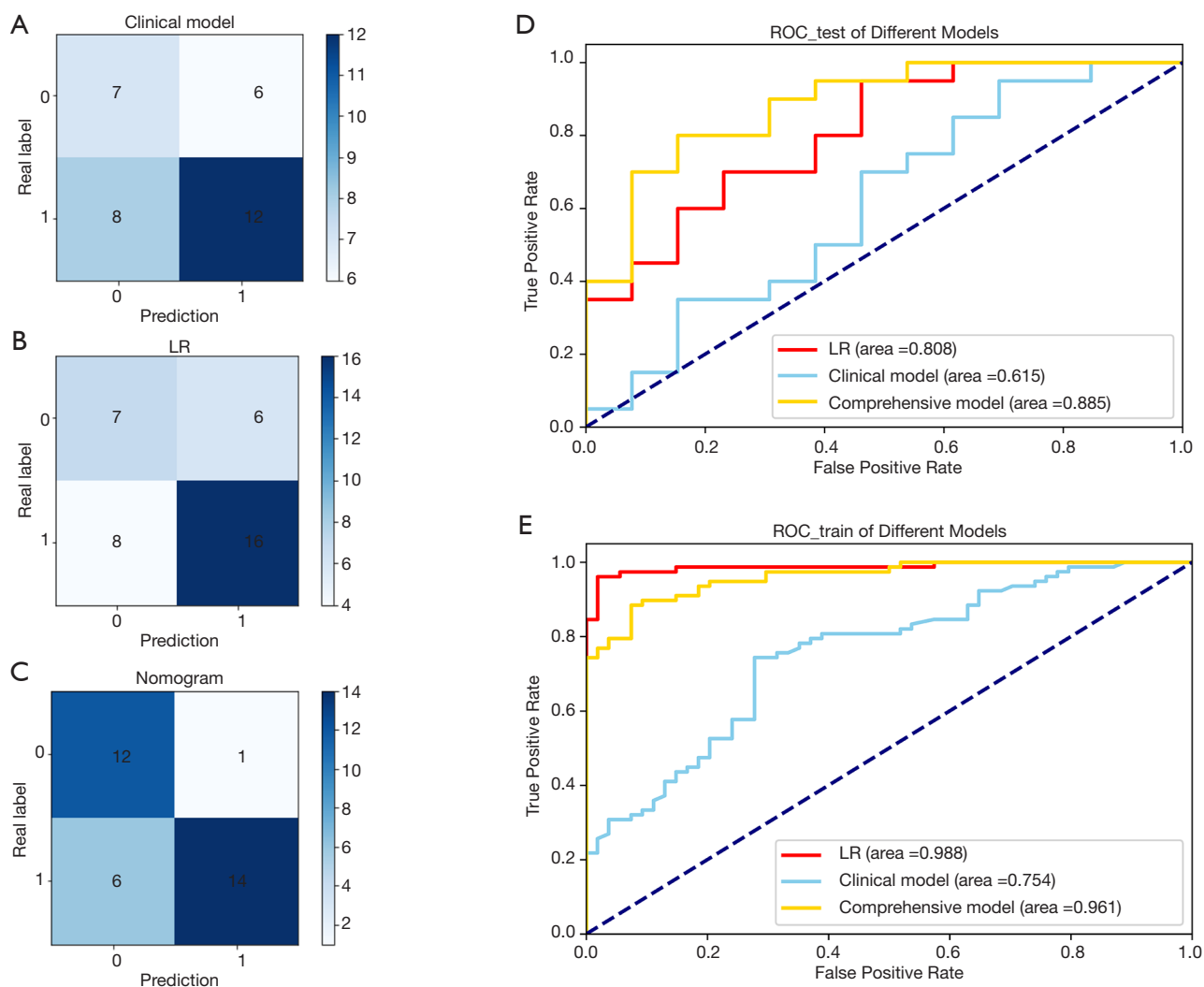


Figure 5 The figure compare the performances of the clinical model, the machine learning classifier, and the nomogram on the same training set. Confusion matrix of the clinical model (A), the LR model (B), and the nomogram (C) on the verification set, and the ROC curves on the test set (D) and training set (E). LR, logistic regression; ROC, receiver operating characteristic.

accuracy =0.5758, sensitivity =0.6, specificity =0.5385].

Finally, the confusion matrix of the 3 models on the test set was drawn as in *Figure 5A,B,C*, and the ROC curves of the 3 models on the test set and training set were drawn (*Figure 5D,E*). The ROC curves were analyzed by the pairwise DeLong test, which showed that in the test set, the ROC curves of the nomogram and the clinical model were significantly different ($P=0.035<0.05$). Although there were differences in AUC between the nomogram and the LR model, the difference was not statistically significant ($P=0.435>0.05$).

From all the indexes, the classification ability of the

nomogram was better than that of the machine LR and clinical model.

Discussion

Analysis of CT signs

So far, the pathogenesis of IBD lacks a clear mechanism. A variety of factors lead to varying degrees of inflammation in the intestine. Bowel wall thickness and segmental mural hyperenhancement are the significant CT signs (31,32). CT diagnosis of cavity organs has always been difficult. The quality of CT imaging mainly depends on adequate

intestinal distention and wall visibility. Radiologists have also studied various methods to better show intestinal lesions. Iso-osmotic mannitol as an oral contrast agent has been recognized as a better method (33-35). However, the intestine is nearly 10 meters long, and it is extremely difficult to fill all the intestines well. Therefore, measurement of intestinal wall thickness was inaccurate due to the degree of filling. This study also found that the wall thickness was not statistically significant in the discrimination between CD and UC.

However, the enhancement characteristics (AP-CT value) of the lesions were statistically significant between the two groups, and the AP-CT value of CD lesions was higher than that of UC. The inflammatory infiltration of UC was confined to the mucosal layer and submucosa, leading to mucosal congestion, edema, and erosion, which form superficial ulcers and crypt abscesses that do not generally involve the muscle layer. CD involves chronic proliferative inflammation of the whole wall, slit-like ulcers, lymphatic vasodilation, and fibrous tissue hyperplasia (36,37). Inflammation stimulates the proliferation and expansion of blood vessels in the mucosa and submucosa, which form the pathological basis for CT enhancement. The ROI measured by the CT value was the average of 3 points on the whole wall, and the layers of the intestinal wall could not be distinguished. This might have caused the difference in AP-CT value.

The ability to observe extraenteric findings as well as the advantages of CT over endoscopy facilitate its use as a supplement to endoscopy diagnosis (38). We found that perienteric edema or inflammation, engorged vasa recta, and lymphadenopathy were common changes that occurred in the mesentery related to IBD. The analysis results of this study showed that engorged vasa recta and lymphadenopathy were not statistically different between CD and UC, but there was a statistical difference in perienteric edema or inflammation. Some researchers have reported that the comb sign is a well-known sign of CD, which is attributed to mesenteric arterial arcades of affected segment changes including vascular dilatation, tortuosity, and conspicuous prominence and wide spacing of the vasa on the mesenteric side with the prominent comb-like arrangement (31,39). Other researchers have studied the correlation between quantitative measures of comb sign with disease activity in CD (40). Both indicate that the corresponding mesangial vascular proliferation of the affected bowel plays an important role in the diagnosis of CD. This was different from our research, which suggests

that we still need to discuss the role of such clinical factors in data from different centers.

In addition, we observed that the incidence of CD and UC had a certain tendency to increase, but there was no significant difference between them. A previous overview (5) showed that the first part that was involved in UC was the rectum, which gradually spread from far to near to the entire colon, and some might have avoided inflammation around the internal orifice of the appendix, inflammation of the ileocecal valve, and reflux ileitis. However, CD could occur throughout the entire digestive tract, and the terminal ileum, colon, and perianal regions were the most commonly affected sites. Therefore, UC was more common in the sigmoid colon and descending colon, and CD was more common in the terminal ileum and cecum. The results of our study were consistent with the previous studies (5).

Analysis of radiomic features

This study used 3 clinical factors (perienteric edema or inflammation, AP-CT value, and lesion location) that were significantly different between the UC and CD groups. Combined with the training Rscore of the LR model, a comprehensive radiomics nomogram based on MSCT imaging omics features and clinical factors was constructed [AUC of the test set (95% CI), 0.8846 (0.7678–1.0), accuracy =0.7879, sensitivity =0.864, specificity =0.9231]. In this study, the performance of the comprehensive nomogram was not only higher than the 5 machine LRs tested, but also much higher than the clinical prediction model. This suggests that the high-throughput features of our radiomics model can not only provide objective and quantifiable radiomics labels for UC and CD on CT images, but can also construct the most suitable model for the identification of UC and CD, which combined radiomics labels and clinical factor labels. This is more comprehensive and reliable than clinicians in distinguishing between the two groups of diseases through pure imaging or pure clinical factors, and the ability to distinguish between the two groups is stronger.

In the performance of the machine LRs, texture features had the greatest impact on the classification model, and texture features are often difficult to judge by the naked eyes of junior radiologists. Therefore, separate imaging diagnosis requires significant radiodiagnostic experience. Also, the subjective differences between observers are large. However, radiomics quantifies all image features and constructs an objective Rscore scoring method to make the

classification results more objective. Therefore, radiomics can be a potential auxiliary tool for doctors of different ages to identify these 2 diseases, and can assist the doctor to make a quick differential diagnosis.

Limitations

This study also had several limitations. Firstly, the total sample size of the study was small, which needs to be further expanded. On the one hand, this will provide more training data for machine learning to improve the learning ability of the model, and on the other hand, it may reduce the impact of data bias on machine learning algorithms. Secondly, there were many high-throughput image features in this study. We can further study whether there was a correlation between different features and clinical semantic factors. Thirdly, intestinal lesions were inevitably affected by the degree of filling, resulting in inaccurate measurement data. It is necessary to make full preparations for the gastrointestinal tract to reduce such effects. Fourthly, this study only performed feature analysis on the arterial phase-enhancement CT images, which might have resulted in incomplete data. It is not clear whether other phases such as plain scan, portal phase-enhancement, or delayed phase can also be used for differential diagnosis. In this regard, further research is still needed. Lastly, since the radiomics label was only for the lesion itself, yet the accompanying indirect signs were also an essential part of the diagnosis, this study combined the 2 for analysis. If the radiomics labeling can directly include the direct signs and indirect signs, it may be able to obtain higher diagnostic efficiency, which requires further development and upgrading of radiomics software.

Conclusions

In summary, compared with traditional imaging diagnostic methods, the radiomics nomogram of MSCT imaging features combined with clinical factors shared the advantages of clinical factors and radiomics at the same time, which was objective, non-invasive, and repeatable. The differential diagnosis of CD and UC using the nomogram was highly interpretable, and it is expected to become a new auxiliary tool for clinical diagnosis, providing accurate information for the clinical development of precise treatment.

Acknowledgments

Thanks to Dr. Qingguo Ren for his assistance in performing

a comprehensive search of the literature and Dr. Qingjun Jiang for his guidance and direction with the research project design.

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <http://dx.doi.org/10.21037/atm-21-1023>

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/atm-21-1023>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-21-1023>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was approved by the ethics committee of Qilu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). Individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD Part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019;13:144-64.
2. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st

- century: A systematic review of population-based studies. *Lancet* 2017;390:2769-78.
3. Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol* 2015;12:720-7.
 4. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet* 2017;389:1756-70.
 5. Tontini GE, Vecchi M, Pastorelli L, et al. Differential diagnosis in inflammatory bowel disease colitis: state of the art and future perspectives. *World J Gastroenterol* 2015;21:21-46.
 6. Torres J, Mehandru S, Colombel JF, et al. Crohn's disease. *Lancet*, 2017;389:1741-55.
 7. Calabrese E, Zorzi F, Onali S, et al. Accuracy of small-intestine contrast ultrasonography, compared with computed tomography enteroclysis, in characterizing lesions in patients with Crohn's disease. *Clin Gastroenterol Hepatol* 2013;11:950-5.
 8. Bollegala N, Griller N, Bannerman H, et al. Ultrasound vs Endoscopy, Surgery, or Pathology for the Diagnosis of Small Bowel Crohn's Disease and its Complications. *Inflamm Bowel Dis* 2019;25:1313-38.
 9. Ruummele FM, Veres G, Kolho KL, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. *J Crohns Colitis* 2014;8:1179-207.
 10. Bemelman WA, Warusavitarne J, Sampietro GM, et al. ECCO-ESCP consensus on surgery for Crohn's disease. *J Crohns Colitis* 2018;12:1-16.
 11. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749-62.
 12. Cui Y, Yang X, Shi Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 2019;29:1211-20.
 13. Horvat N, Veeraraghavan H, Khan M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. *Radiology* 2018;287:833-43.
 14. Liu Z, Zhang XY, Shi YJ, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* 2017;23:7253-62.
 15. Li JZ, Tang L. Advances in radiological studies of gastrointestinal stromal tumors. *Zhonghua Wei Chang Wai Ke Za Zhi* 2019;22:891-5.
 16. Goyal P, Shah J, Gupta S, et al. Imaging in discriminating intestinal tuberculosis and Crohn's disease: past, present and the future. *Expert Rev Gastroenterol Hepatol* 2019;13:995-1007.
 17. Stidham RW, Enchakalody B, Waljee AK, et al. Assessing Small Bowel Stricture and Morphology in Crohn's Disease Using Semi-automated Image Analysis. *Inflamm Bowel Dis* 2020;26:734-42.
 18. Li X, Feng S, Huang L, et al. Expert guideline on imaging examination and report specification of inflammatory bowel disease in China. *Chin J Inflamm Bowel Dis* 2021;5:17-21.
 19. Van Assche G, Dignass A, Panes J, et al. The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: Definitions and diagnosis. *J Crohns Colitis* 2010;4:7-27.
 20. Elsayes KM, Al-Hawary MM, Jagdish J, et al. CT enterography: principles, trends, and interpretation of findings. *Radiographics* 2010;30:1955-70.
 21. Guglielmo FF, Anupindi SA, Fletcher JG. Small Bowel Crohn Disease at CT and MR Enterography: Imaging Atlas and Glossary of Terms. *Radiographics* 2020;40:354-75.
 22. Svoboda M, Matiu-Iovan L, Frigura-Iliasa FM, et al. B-spline interpolation technique for digital signal processing. *International Conference on Information & Digital Technologies. IEEE*, 2015:366-71.
 23. Wiesaw P. Tree-based generational feature selection in medical applications. *Procedia Comput Sci* 2019;159:2172-8.
 24. Bishop CM. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York; 2006.
 25. Mitchell TM. *Machine Learning*. McGraw-Hill; 2003.
 26. Mount J. The equivalence of logistic regression and maximum entropy models. *win-vector.com*; 2011;1:2-5.
 27. Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min Knowl Discov* 1998;2:121-67.
 28. Breiman L. Random forest. *Machine Learning* 2001;45:5-32.
 29. Paras. Stochastic Gradient Descent. *Optimization* 2014;10:111-30.
 30. Keysers D, Ney H. Linear discriminant analysis and discriminative log-linear modeling. *Pattern Recognit* 2004;1:156-9.
 31. Carbo AI, Reddy T, Gates T, et al. The most characteristic lesions and radiologic signs of Crohn disease of the small bowel: air enteroclysis, MDCT, endoscopy, and pathology.

- Abdom Imaging 2014;39:215-34.
32. Gore RM, Balthazar EJ, Ghahremani GG, et al. CT features of ulcerative colitis and Crohn's disease. *AJR Am J Roentgenol* 1996;167:3-15.
 33. Wong J, Moore H, Roger M, et al. CT enterography: Mannitol versus VoLumen. *J Med Imaging Radiat Oncol* 2016;60:593-8.
 34. Wong J, Roger M, Moore H, et al. Performance of two neutral oral contrast agents in CT enterography. *J Med Imaging Radiat Oncol* 2015;59:34-8.
 35. Zhang LH, Zhang SZ, Hu HJ, et al. Multi-detector CT enterography with iso-osmotic mannitol as oral contrast for detecting small bowel disease. *World J Gastroenterol* 2005;11:2324-9.
 36. Guan Q. A Comprehensive Review and Update on the Pathogenesis of Inflammatory Bowel Disease. *J Immunol Res* 2019;2019:7247238.
 37. Gajendran M, Loganathan P, Catinella AP, et al. A comprehensive review and update on Crohn's disease. *Dis Mon* 2018;64:20-57.
 38. Magro F, Gionchetti P, Eliakim R, et al. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders. *J Crohns Colitis* 2017;11:649-70.
 39. Plastaras L, Vuitton L, Badet N, et al. Acute colitis: differential diagnosis using multidetector CT. *Clin Radiol* 2015;70:262-9.
 40. Wu YW, Tao XF, Tang YH, et al. Quantitative measures of comb sign in Crohn's disease: correlation with disease activity and laboratory indications. *Abdom Imaging* 2012;37:350-8.
- (English Language Editor: C. Betlazar-Maseh)

Cite this article as: Li H, Mo Y, Huang C, Ren Q, Xia X, Nan X, Shuai X, Meng X. An MSCT-based radiomics nomogram combined with clinical factors can identify Crohn's disease and ulcerative colitis. *Ann Transl Med* 2021;9(7):572. doi: 10.21037/atm-21-1023



Figure S2 This figure shows the features with the highest weight coefficient of five machine models. The distribution of the top 20 features (A,B,C,D,E) of the 5 ML models, and the weight ratios of the different types of features (F).

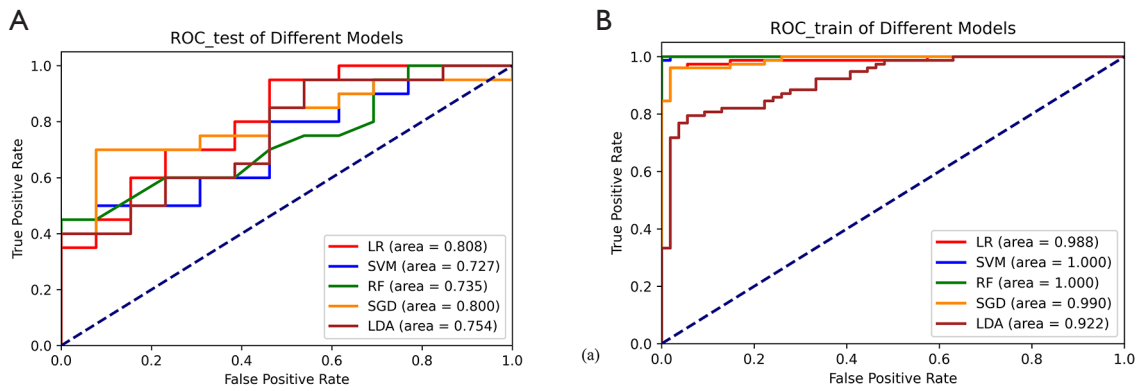


Figure S3 Comparison of diagnostic performance of different machine learning models. ROC curves of 5 machine learning models in the test set (A) and the training set (B).

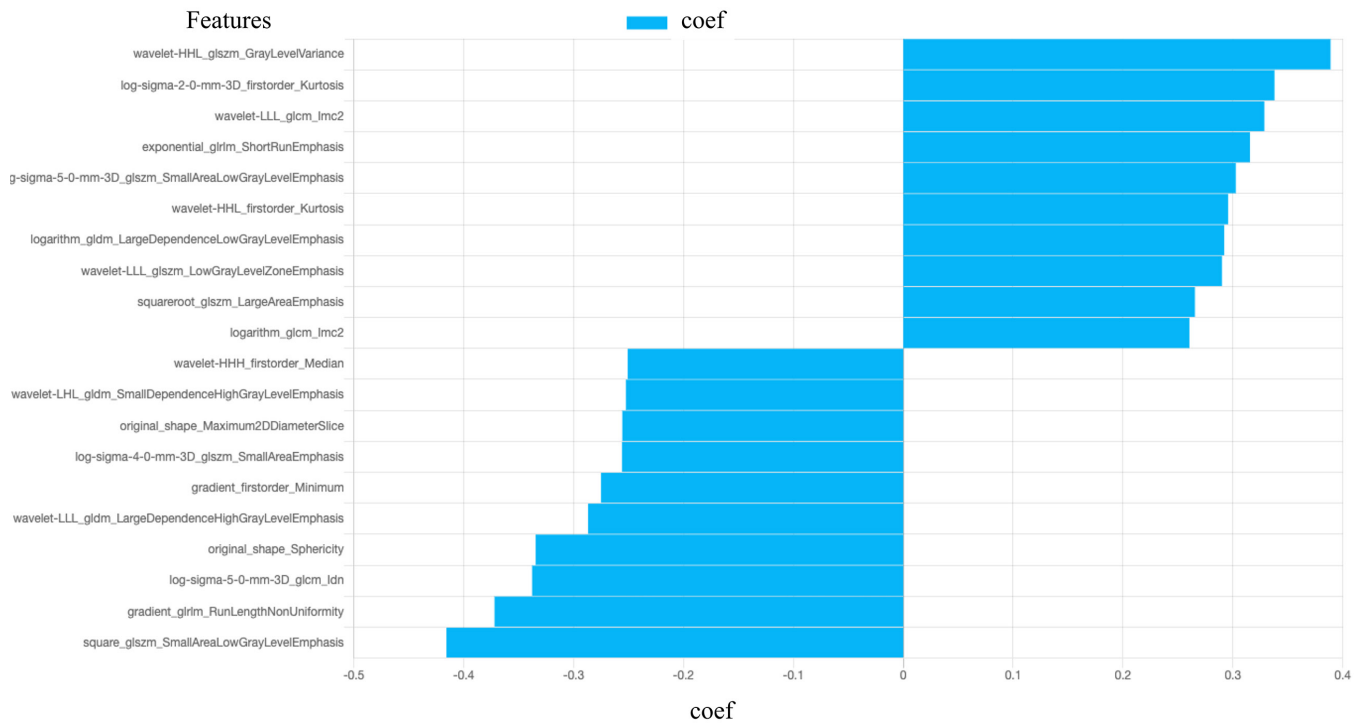


Figure S4 Coefficients of the top 20 features of the LR classifier.

Table S1 Clinical logistic regression model

Clinical factors	Regression coefficient	P value	OR	OR 95% CI
AP-CT value (Hu)	0.024	0.022	1.024	1.003-1.045
Perienteric edema or inflammation	-0.965	0.028	0.381	0.161-0.901
Lesion location				
Terminal ileum	-0.061	0.106	0.915	0.345-3.279
Cecum	-0.345	0.551	0.709	0.228-2.202
Ascending colon	-0.596	0.33	0.551	0.166-1.827
Transverse colon	-1.006	0.122	0.366	0.102-1.307
Descending colon	-0.849	0.178	0.428	0.125-1.470
Sigmoid colon	-1.05	0.078	0.35	0.109-1.123
Rectum	-0.486	0.446	0.615	0.176-2.149