Original Research

Development and validation of a novel model incorporating MRI-based radiomics signature with clinical biomarkers for distinguishing pancreatic carcinoma from mass-forming chronic pancreatitis

Jingjing Liu [a], Lei Hu [a], Bi Zhou [a],[*], Chungen Wu [a], Yingsheng Cheng [a]

[a] Department of Radiology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai 200233, People's Republic of China

A B S T R A C T

Purpose: It is difficult to make a clear differential diagnosis of pancreatic carcinoma (PC) and mass-forming chronic pancreatitis (MFCP) via conventional examinations. We aimed to develop a novel model incorporating an MRI-based radiomics signature with clinical biomarkers for distinguishing the two lesions.
Methods: A total of 102 patients were retrospectively enrolled and randomly divided into the training and validation cohorts. Radiomics features were extracted from four different sequences. Individual imaging modality radiomics signature, multiparametric MRI (mp-MRI) radiomics signature, and a final mixed model based on mp-MRI and clinically independent risk factors were established to discriminate between PC and MFCP. The diagnostic performance of each model and model discrimination were assessed in both the training and validation cohorts.
Results: ADC had the best predictive performance among the four individual radiomics models, but there were no significant differences between the pairs of models (all $p > 0.05$). Six potential radiomics features were finally selected from the 960 texture features to formulate the radiomics score (rad-score) of the mp-MRI model. In addition, the boxplot results of the distributions of rad-scores identified the rad-score as an independent predictive factor for the differentiation of PC and MFCP ($p < 0.001$). Notably, the nomogram integrating rad-score and clinically independent risk factors had a better diagnostic performance than the mp-MRI and clinical models. These results were further confirmed by the validation group.
Conclusion: The mixed model was developed and preliminarily validated to distinguish PC from MFCP, which may benefit the formulation of treatment strategies and nonsurgical procedures.

Introduction

Pancreatic carcinoma (PC), a malignant tumor of the pancreatic exocrine glands, is currently recognized as one of the deadliest malignant tumors worldwide [1, 2]. It is characterized by a high degree of malignancy, rapid progression, and extremely poor prognosis [3, 4]. Because of the extremely low five-year survival rate of patients with PC, it may surpass breast cancer as the third leading cause of cancer death by 2025. Mass-forming chronic pancreatitis (MFCP) is a type of chronic pancreatitis [5]. Long-term inflammation of the pancreas causes the pancreatic parenchyma to be replaced by fibrous tissue; eventually, a lump of tissue with chronic inflammatory cell infiltration is formed locally [6, 7]. PC and MFCP usually occur in the head of the pancreas.

Both are extremely similar in terms of clinical symptoms, serum tumor markers, and imaging features [8–10]. Thus, it is very difficult to make a clear diagnosis of PC and MFCP before surgery and appropriate treatment. However, their management is completely different; misdiagnosis will cause patients with PC to miss the best surgical opportunity, while patients with MFCP may receive unnecessary surgical treatment. According to statistics, at least 2.0% - 5.0% of the postoperative pathology of the pancreatic head mass is inflammatory [11, 12]. Therefore, accurate diagnosis is important for the survival and prognosis of patients.

Currently, there are many clinical diagnostic methods for distinguishing PC from MFCP [13–15]. Tumor markers, such as serum carbohydrate antigen 19–9 (CA19–9) and carcinoembryonic antigen (CEA), can be used for early screening of pancreatic cancer, but its false

---

negative rate was high [16]. Imaging examinations, including B-ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI), help to provide valuable information for differential diagnosis; however, their accuracy is limited due to the substantial overlap in images, so it is still difficult to clearly differentiate MFCP and PC. Biopsy is the most precise diagnostic method for discriminating between PC and MFCP lesions. However, it is invasive and has many complications [17]. Radiomics can convert traditional digital images into mineable high-throughput features through objective calculations on a computer to extract and analyze texture signatures that cannot be observed by the naked eye [18, 19]. Currently, the field of radiomics has grown exponentially and has been applied in the diagnosis of many diseases [20–22]. Multiparametric magnetic resonance imaging (MRI) is a non-invasive procedure involving non-ionizing radiation that provides exceptional diagnostic performance in the differentiation of pancreas-related diseases [14]. A rare study has reported an MRI-based radiomics model for distinguishing PC from MFCP.

In this study, we constructed and validated a comprehensive model combining MRI-based radiomics signature with known clinical biomarkers (CA19–9 and CEA) to enhance diagnostic accuracy in differentiating PC from MFCP, and vice versa.

Materials and methods

Patient population

This retrospective study was approved by the local institutional ethics committee; informed consent was waived owing to the retrospective nature of the study. The final diagnostic criterion for PC is: pancreatic ductal adenocarcinoma confirmed by surgical pathology. The final diagnostic criteria for MFCP are: (1) chronic pancreatitis confirmed by surgical pathology; (2) chronic pancreatitis confirmed by needle biopsy, and the lesions have shrunk during three-month follow-up after conservative treatment. First, the departmental database of Shanghai Jiaotong University Affiliated Sixth People's Hospital between January 2017 and March 2021 was searched, and 146 patients meeting the diagnostic criteria were included in the primary cohort of our study. Next, 127 cases were selected according to the inclusion criteria: 1) patients who underwent an MRI scan within two weeks before the pathological diagnosis; 2) MR images which contained the following sequences: T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI) with $b = 800$ s/mm^2, and apparent diffusion coefficient (ADC) maps; and 3) patients who had complete clinical information and pathologic examination results. Finally, as shown in Fig. 1, the 102 samples (including 54 patients with PC and 48 patients with MFCP) were enrolled after the implementation of the exclusion criteria: 1) no definite mass was found on MR images; 2) the image quality was poor; and 3) patients who received various treatments prior to the MRI examination.

The included dataset was randomly divided into two groups: a training cohort (n = 72) and a validation cohort (n = 30). Some clinical and imaging characteristics, including age at diagnosis, sex, location and size of the lesion, and the presence of CA19–9 and CEA were recorded. We defined the CA19–9 level as normal from 0 to 37 U/ml and CEA level as normal from 0 to 5 ng/ml; if otherwise, it was considered abnormal.

MR Image acquisition

All patients received a 3.0-T MRI (MAGNETOM Skyra, Siemens Healthcare); the signal was received using a phased-array 18-channel body coil combined with an integrated 32-channel spine coil. The examination consisted of different sequences: (1) a transversal pre-contrast volumetric-interpolated breath-hold examination (VIBE) T1-weighted sequence, (2) a fat-suppressed, transversal half-Fourier acquisition single-shot turbo spin-echo (HASTE) T2-weighted sequence, and (3) a fat-suppressed, single-shot EPI DWI sequence with b-values of 50 and
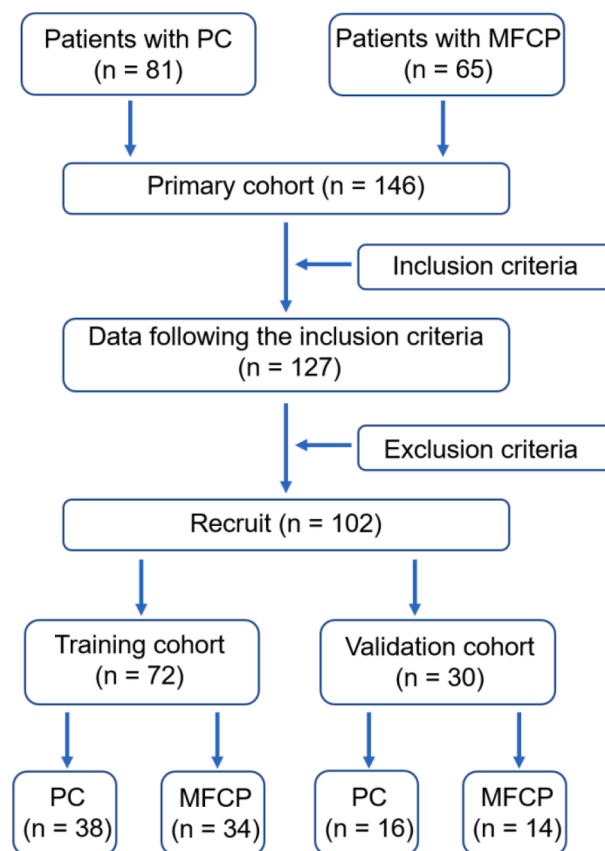

Fig. 1. Flowchart of patient enrollment in our study.

800 s/mm^2. ADC maps were inline calculated with the b-values acquired using a monoexponential function. Detailed information on the acquisition parameters is provided in Table 1.

Image segmentation and feature extraction

Axial T1WI, T2WI, DWI$_{b800}$, and ADC images were acquired for image segmentation and feature extraction. For patients with multiple significant lesions, only those larger than 1 cm^3 were analyzed. Regions of interest (ROIs) were manually delineated slice-by-slice along the margin of the lesion on all images of different sequences by using dedicated software (ITK-Snap, Version 3.6.0; www.itksnap.org). Subsequently, the volume of interest (VOI) delineation was generated. The ADC maps and corresponding DWI images were intrinsically co-registered; thus, the segmented VOIs were directly replicated from DWI to ADC.

Intra- and interclass correlation coefficients (ICC) were employed to evaluate the intra- and inter-observer reproducibility of the volume segmentation and radiomics feature extraction by two radiologists with

Table 1
The parameters of the magnetic resonance imaging sequence.

Parameter	T1WI	T2WI	DWI
Field of view (mm^2)	260 × 320	240 × 320	216 × 268
Acquisition matrix	320 × 195	320 × 197	134 × 108
Slice Thickness (mm)	3.4	4.5	5
Flip angle	9	160	90
Echo train length (mm)	1	96	43
Echo time (ms)	1.3	80	43
Repetition time (ms)	3.3	1600	5300
Pixel Bandwidth (Hz/px)	445	870	2490
b-value (s/mm^2)	n.a.	n.a.	50, 800

n.a., not applicable.

8 (observer 1) and 15 years (observer 2) of abdominal imaging experience. First, the MR images of 40 randomly selected patients underwent image segmentation and feature extraction by observers 1 and 2, respectively. Then, observer 1 repeated the same procedure, with an interval of more than one week between delineations. An ICC greater than 0.75 was determined to have good feature extraction consistency [23, 24]. Finally, the remaining image segmentation and feature extraction were performed by observer 1. Consequently, all studies were confirmed by observer 2.

An open-source Python package for quantitative data from medical images (Pyradiomics, v2.1.2) was used to extract texture features, including both low- and high-order radiomic features. Image preprocessing was performed to achieve image normalization and discretization. Pyradiomics provided a normalization method, which normalized the image by taking the mean with standard deviation as the center. The minimum-redundancy maximum-relevance (mRMR) algorithm and least absolute shrinkage and selection operator (LASSO) were used to select the features. First, the mRMR was conducted to maximize the relevance between features and categorical variables, and minimize the relevance between features and features, finally eliminate multicollinearity and prevent over-fitting. For mRMR algorithm, the relevance between features and categories is calculated by the mean value of the information gain, and the redundancy between features and features is calculated by dividing the sum of mutual information by the square of the number of features. Then, LASSO, trained by a ten-fold cross-validation method, was performed to select the optimized subset of features. The most reliable features were used to construct a radiomics model using multivariable logistic regression analysis.

Model building

In this study, a three-step operation was conducted to build a novel mixed-prediction model. First, the diagnostic efficacy of each imaging modality to distinguish PC from MFCP was evaluated. To compare the efficiency of differential diagnosis of each modality, four predictive signatures were built as follows: T1WI, T2WI, DWI, and ADC. Next, because T1WI and T2WI can predominantly provide high-resolution images of the abdominal anatomy, whereas physiologic and functional data are provided by DWI and ADC, a multiparametric MRI (mp-MRI) radiomics signature model based on T1WI, T2WI, DWI, and ADC was built. Finally, by combining the mp-MRI signature with clinically independent risk factors, a mixed-prediction model was established to distinguish PC from MFCP. The workflow of the development of the radiomic signature and the comprehensive model is shown in Fig. 2.

Statistical analysis

Concerning the clinical data, continuous variables, such as the age of the patient and the size of the lesion, are described as means ± SD. Categorical variables, such as patient sex, lesion location, and CA19-9 and CEA levels, were described as numbers and percentages. Normally distributed continuous variables were assessed using an independent samples *t*-test; non-normally distributed continuous variables were assessed using Mann-Whitney U tests; and categorical variables were assessed using Pearson chi-square test or Fisher exact test. The receiver operating characteristic (ROC) curve was used to quantify the predictive accuracy, sensitivity, and specificity of each imaging modality in both the training and validation sets. The corresponding values of the area under the ROC curve (AUC) were calculated to evaluate the model performance. The AUCs of the ROC curves between different radiomics models were compared using the DeLong test. Binomial exact test was used for computing the confidence interval of AUC. The accuracy, sensitivity, and specificity in Table 3 were calculated as the following formula: Accuracy = (True Positive + True Negative) / (True Positive + True Negative + False Positive + False Negative); Sensitivity = True Positive / (True Positive + False Negative); Specificity = True Negative /
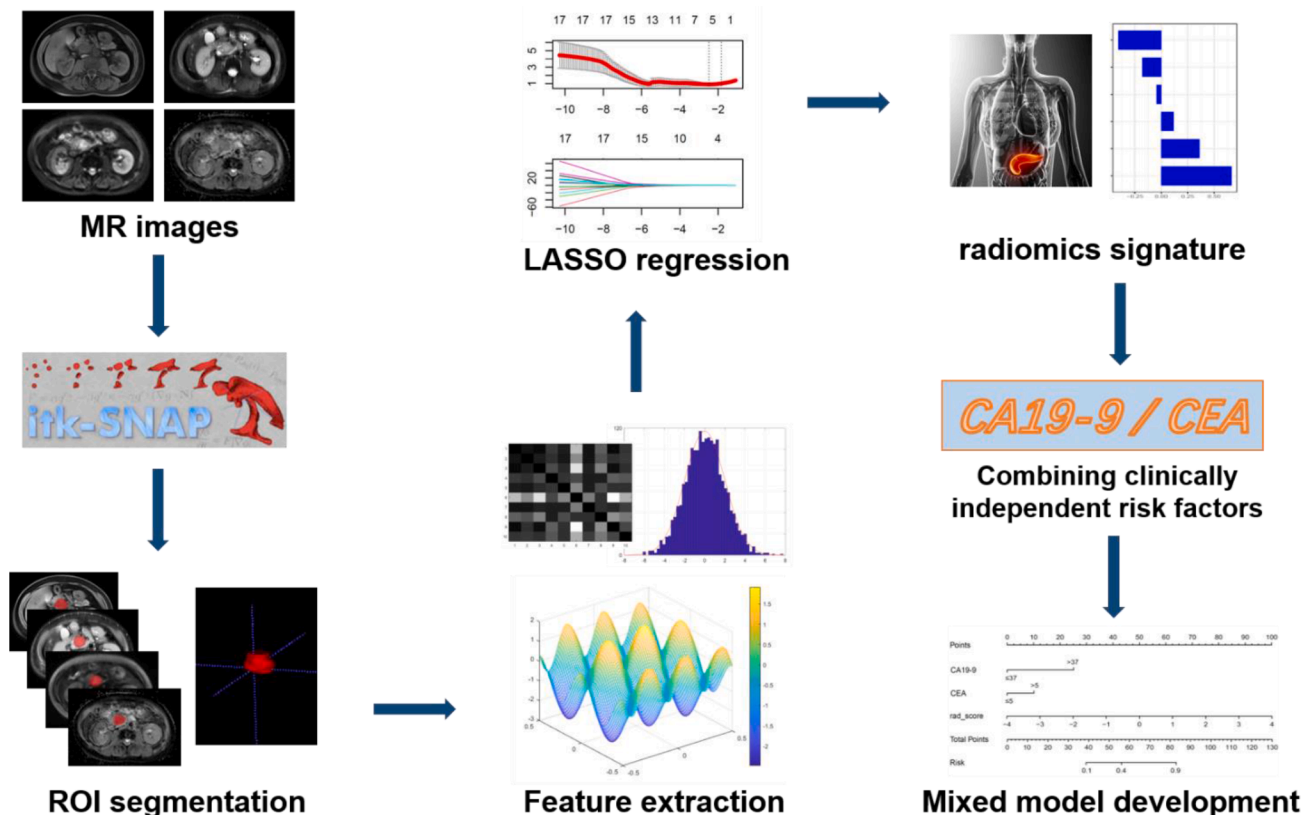

Fig. 2. Workflow showing the development of the radiomic signature and the comprehensive model.

(True Negative + False Positive).

The performance of the mp-MRI radiomics signature model and the clinical model were evaluated using the ROC and calibration curves. The efficiency of fit of the models was assessed using the Hosmer–Lemeshow test. Decision curve analysis (DCA) was used to estimate the clinical utility of the established model. The net benefits vs. risk thresholds in the training and validation sets were also calculated [25].

The LASSO logistic regression model, which combined with penalty parameter tuning, were performed using a ten-fold cross-validation based on the minimum criteria. The likelihood ratio test, used Akaike's information criterion (AIC) as the stopping rule, was applied for backward stepwise selection. The "glmnet" package in R was used to perform LASSO logistic regression. The "cv. glmnet" function was used to select the tuning parameter (λ). All statistical analyses were conducted using the R statistical software (Version 3.6.3). Nomogram and calibration plots were constructed using the "rms" package. The "pROC" package was used for ROC plotting. DCA curve plots were constructed using the "rmda" package. The "rms" package in R was used to calibrate the radiomic signature. Statistical significance was set at $p < 0.05$.

Results

Clinical data of patients

This study consisted of 102 patients; the flowchart of the study is illustrated in Fig. 1. Patient characteristics are summarized in Table 2; there were no statistical differences in age ($p = 0.882, 0.549$), sex ($p = 0.637, 0.707$), lesion location ($p = 0.259, 0.657$), and lesion size ($p = 0.192, 0.414$) between the PC and MFCP groups in the training and validation cohorts, respectively. However, both levels of CA19-9 and CEA were significantly different in patients with PC compared with those in patients with MFCP; p values were 0.013 for CA19-9 and 0.033 for CEA in the validation dataset.

Table 2
Characteristics of the study population and MR imaging findings.

	The training cohort		p-value	The validation cohort		p-value
	PC (n = 38)	MFCP (n = 34)		PC (n = 16)	MFCP (n = 14)	
Age (Y)	61.6 ± 14.4	62.1 ± 14.1	0.882	63.3 ± 13.5	60.5 ± 11.5	0.549
Size (cm²)	6.84 ± 2.57	6.13 ± 1.91	0.192	7.07 ± 2.48	6.38 ± 2.01	0.414
Sex, n (%)			0.637			0.707
Male	21 (55.3)	16 (47.1)		11 (68.7)	8 (57.1)	
Female	17 (44.7)	18 (52.9)		5 (31.3)	6 (42.9)	
Location, n (%)			0.259			0.657
Head or neck	28 (73.7)	29 (85.3)		12 (75)	12 (85.7)	
Body or tail	10 (26.3)	5 (14.7)		4 (25)	2 (14.3)	
CA19–9, n (%)			<0.001*			0.013*
0~37 U/ml	11 (28.9)	28 (82.4)		5 (31.3)	11 (78.6)	
>37 U/ml	27 (71.1)	6 (17.6)		11 (68.7)	3 (21.4)	
CEA, n (%)			0.037*			0.033*
0~5 ng/ml	15 (39.5)	22 (64.7)		6 (37.5)	11 (78.6)	
>5 ng/ml	23 (60.5)	12 (35.3)		10 (62.5)	3 (21.4)	

*Data are statistically significant with $p < 0.05$.
Y, years; PC, pancreatic carcinoma; MFCP, mass-forming chronic pancreatitis; CA19–9, carbohydrate antigen 19–9; CEA, carcinoembryonic antigen.

Intra- and inter-observer reproducibility

The intra- and inter-observer feature extraction reproducibility was evaluated using ICCs. Regarding the intra-observer agreement of radiomics features, the ICCs ranged from 0.771 to 0.995; for the inter-observer agreement, the ICCs ranged from 0.802 to 0.971. These results were all greater than 0.75, confirming their advantages for feature extraction reproducibility.

Predictive performance of the individual radiomics signature

The radiomics signature of each separate imaging modality was selected in the training set using the least absolute shrinkage and selection operator (LASSO) algorithm to perform dimensionality reduction and was employed in the subsequent modeling analysis. There were eight, nine, nine, and eight radiomics signatures included in T1WI, T2WI, DWI, and ADC, respectively, for model building (**Supplementary Material 1**). We used ROC curves to demonstrate the predictive performances of the four models in the differential diagnosis of PC and MFCP (**Fig. 3**A and B). The AUCs of the T1WI, T2WI, DWI and ADC models were 0.885 [95% confidence interval (CI): 0.788–0.948], 0.898 (95%CI: 0.804–0.957), 0.872 (95%CI: 0.773–0.939), 0.917 (95%CI: 0.828–0.969) in the training cohorts, and 0.871 (95% CI: 0.698–0.964), 0.888 (95%CI: 0.720–0.973), 0.848 (95%CI: 0.670–0.952), 0.908 (95% CI: 0.749–0.983) in the validation cohorts. In addition, the accuracy, sensitivity, and specificity of all MRI radiomics signature-based models were calculated and recorded. The detailed results of the predictive performance are presented in Table 3. Applying the DeLong test to compare the AUCs across the four radiomics models, we observed that there were no significant differences between the pairs of models (all $p > 0.05$). However, the ADC model had the best predictive performance among the four radiomics models; the T1WI model showed lower performance than the other individual sequences.

Establishment of mp-MRI radiomics signature and clinical model

Based on the predictive performance of the radiomics signature of each imaging modality, T1WI, T2WI, DWI, and ADC maps were employed to build a multiparametric MRI radiomics signature-based model. Finally, six radiomics features with non-zero coefficients (one from T1WI, two from T2WI, one from DWI, and two from ADC maps) were selected out of 960 texture features after LASSO logistic regression and were used to construct the mp-MRI radiomic signature (**Fig. 4**). The calculation formula of the radiomics score (rad-score) was established; the details are shown in **Supplementary Material 2**. Compared with the radiomics signature of individual imaging modalities, the mp-MRI model showed the best diagnostic efficiency for distinguishing PC from MFCP. Boxplots showing the rad-score values were significantly higher in patients with PC than in patients with MFCP in both the training and validation cohorts (**Fig. 5**). Next, CA19–9 and CEA were chosen as clinically independent risk factors for final mixed-model building because their levels were significantly different between patients with PC and those with MFCP. Finally, we built a quantitative mixed-prediction model incorporating the mp-MRI rad-score and clinically independent risk factors. A combined nomogram was developed incorporating CEA and CA19–9 levels. Each factor is distributed in a weighted number of points. The risk of PC was associated with the total number of points for each patient, which was calculated using the nomogram (**Fig. 6**A).

Development and validation of the comprehensive prediction models

The ROC curve was used to assess the discriminative ability of the three models (**Fig. 3** C and D). As shown in Table 3, the AUCs of the mixed, the mp-MRI, and the clinically independent risk factors models were 0.853 (95% CI: 0.750–0.925), 0.950 (95%CI: 0.872–0.988), 0.973
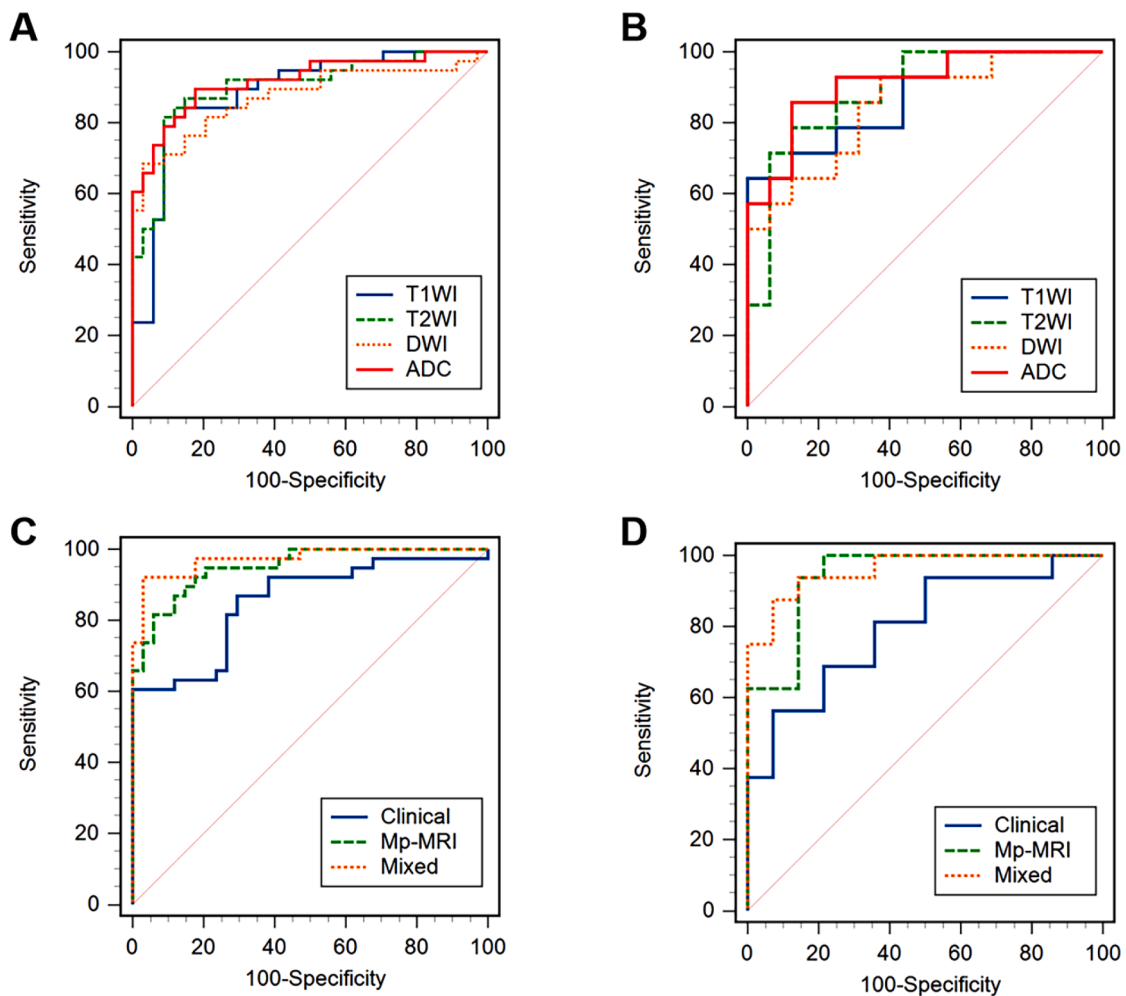
Fig. 3. Receiver operating characteristic (ROC) curves of four single radiomics signature in the training group (A) and the validation group (B). ROC curves of three prediction models in the training group (C) and the validation group (D).

Table 3
Predictive performance of different models.

	AUC	95%CI	ACC	SEN	SPE
T1WI					
Training	0.885	0.788–0.948	0.833	0.790	0.882
Validation	0.871	0.698 to 0.964	0.789	0.714	0.875
T2WI					
Training	0.898	0.804–0.957	0.847	0.816	0.880
Validation	0.888	0.720 to 0.973	0.827	0.786	0.879
DWI					
Training	0.872	0.773–0.939	0.778	0.711	0.853
Validation	0.848	0.670 to 0.952	0.767	0.837	0.688
ADC					
Training	0.917	0.828–0.969	0.842	0.857	0.824
Validation	0.908	0.749 to 0.983	0.886	0.895	0.875
Clinical					
Training	0.853	0.750–0.925	0.694	0.632	0.765
Validation	0.799	0.613–0.922	0.734	0.688	0.786
Mp-MRI					
Training	0.950	0.872–0.988	0.875	0.921	0.824
Validation	0.942	0.791–0.994	0.894	0.928	0.857
Mixed					
Training	0.973	0.904–0.997	0.898	0.922	0.871
Validation	0.960	0.817–0.998	0.909	0.939	0.875

AUC, area under the curve; CI, confidence interval; ACC, accuracy; SPE, specificity; SEN, sensitivity.

(95%CI: 0.904–0.997) in the training cohort, respectively, and 0.799 (95% CI: 0.613–0.922), 0.942 (95%CI: 0.791–0.994), and 0.960 (95% CI: 0.817–0.998) in the validation cohort, respectively. Between the three models, the comprehensive model displayed the best evaluation performance in both the training and validation cohorts and was statistically different from the clinical model ($p = 0.011$ and 0.029. respectively). The mixed model also possessed the highest accuracy (0.898 and 0.909, respectively), the highest specificity (0.871 and 0.875, respectively), and the highest sensitivity (0.922 and 0.939, respectively) in the differential diagnosis of the two cohorts.

A calibration curve was used to illustrate the consistency between the predicted risks and the actual observed outcomes. The red dotted line fitted the gray line, which represents the reference line showing the "ideal" prediction. As shown in **Fig. 6B**, the Hosmer–Lemeshow test also showed good calibration of the mixed model in the validation cohorts ($p = 0.935$), confirming the excellent predictive accuracy of the mixed model. Furthermore, a DCA curve was used to intuitively show the net benefits of the potential population under different risk thresholds and to determine whether the final mixed model could help with clinical treatment strategies (**Fig. 6C**). According to the DCA, when the risk thresholds varied from 0 to 1, the mixed model achieved the highest net benefit compared with the "treat all" and "treat none" strategies as well as with the clinically independent risk factors and mp-MRI models in both the training and validation cohorts. For example, in the validation cohort, if we defined the risk threshold as 50%, the standardized net benefit of patients was 0.82 and 0.58 for the mp-MRI model and
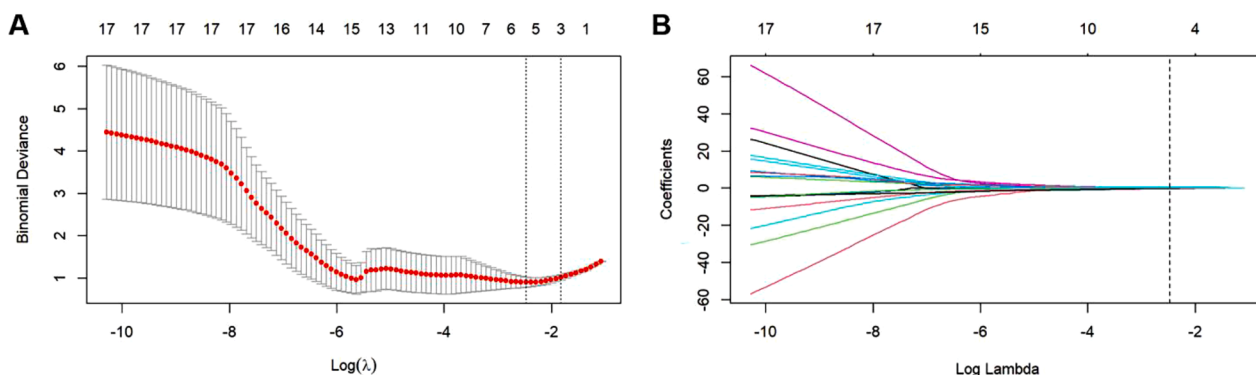
Fig. 4. LASSO logistic regression for texture feature selection. (A) Selection of the tuning parameter (λ) in the LASSO model. (B) LASSO coefficient profiles of the 17 texture features.
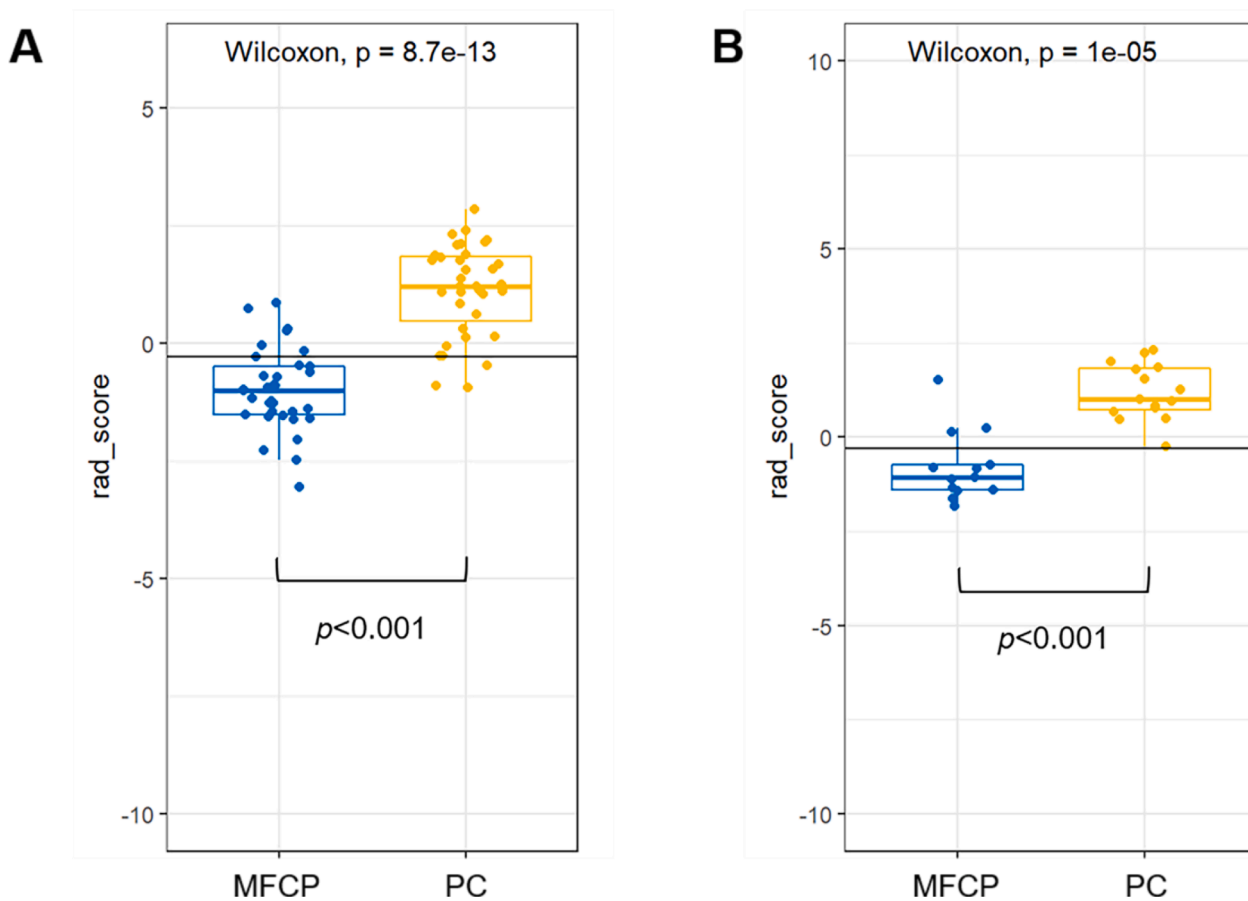

Fig. 5. Boxplots of the distributions of radiomics scores to distinguish MFCP from PC group according to mp-MRI prediction model in the training dataset (A), and validation dataset (B).

clinically independent risk factors model, respectively, whereas the standardized net benefit was 0.84 for the mixed model.

Discussion

With the increasing occurrence of alcohol- and smoking-related habits, the incidence of chronic pancreatitis has also increased gradually, to which MFCP has become more common clinically. However, it is extremely difficult to make a clear diagnosis between MFCP and PC before surgery and appropriate treatment according to existing imaging technology and clinical data. Pathological biopsy is the gold standard for the diagnosis of solid pancreatic lesions. In the last few years,

endoscopic ultrasound-guided fine-needle biopsy (EUS-FNB) has been used for clinical practice of sampling pancreatic masses, which can obtain histological tissue samples to easily perform immunohisto-chemistry [26]. The diagnostic accuracy of independent EUS-FNB without rapid on-site evaluation (ROSE) can reach 0.974 (95% CI: 0.953–0.988) [27]. Multiparametric MRI-based radiomics analysis is non-invasive and efficient, providing a quantitative measure of intrale-sional heterogeneity that may help in distinguishing benign and malignant lesions, assessing tumor aggressiveness, and evaluating treatment response [19]. Multiparametric MRI has already been used for the differentiation of PC from MFCP in some studies [15, 28, 29], but those applying the mp-MRI radiomics signature remain scarce. In the
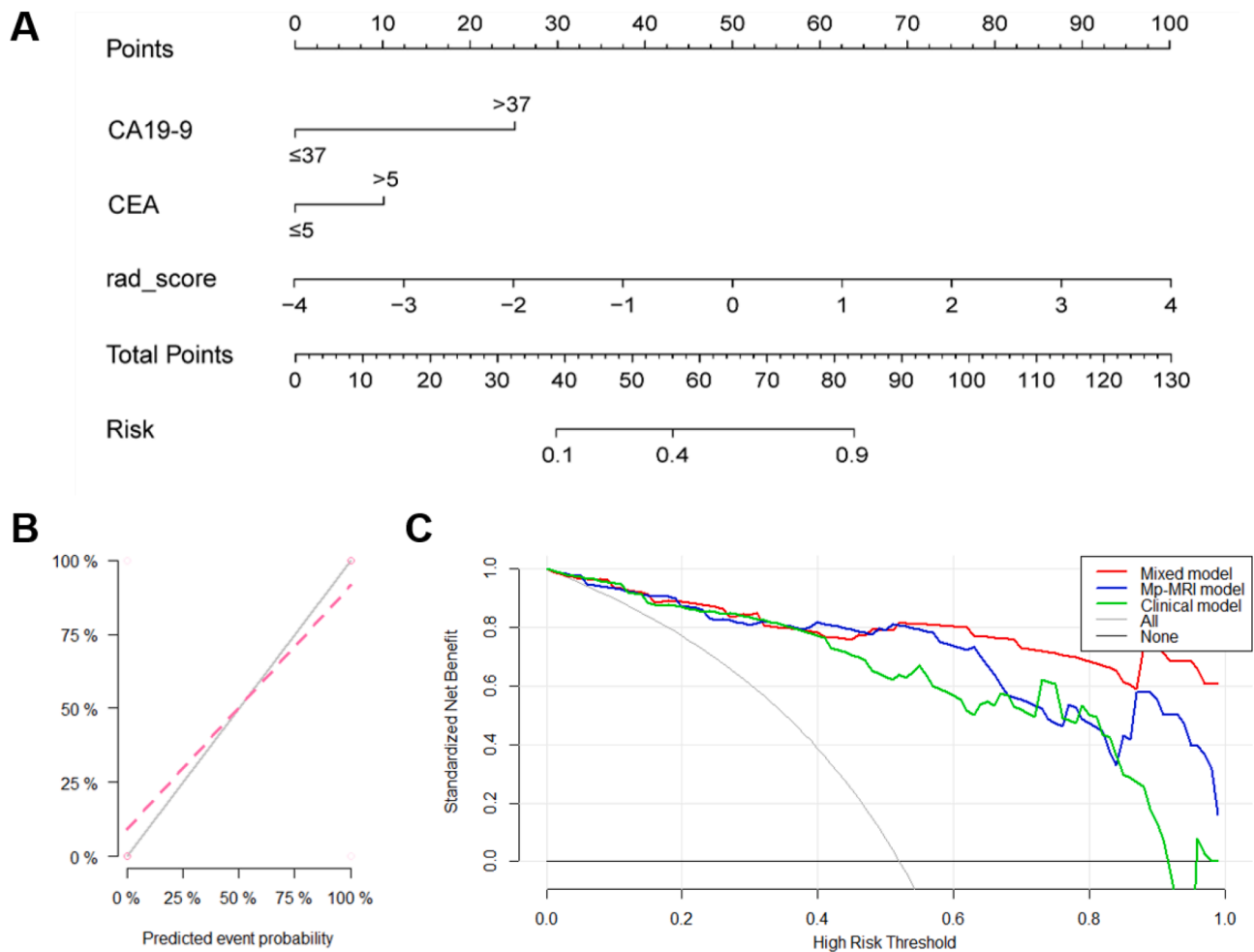
Fig. 6. (A)The nomogram of the mixed model incorporating the radiomic signature, the CA19–9 level, and the CEA level. (B) The calibration curve of the mixed model in the validation group. (C) The decision curve analysis (DCA) curve of clinical use assessment of three prediction models in the validation group.

present study, we first developed and validated a comprehensive model that combined MRI-based radiomics signature with clinically independent risk factors to enhance the diagnostic accuracy in differentiating PC and MFCP lesions.

T1WI and T2WI are regular sequences for MR inspection. The DWI and ADC measurements used for discriminating PC from MFCP have been suggested in some studies [30, 31]. Therefore, we first compared the diagnostic performance of the single-imaging modality, involving radiomics signatures extracted from T1WI, T2WI, DWI, and ADC sequences. We found that although there was no significant difference between the four single-imaging modalities (all $p > 0.05$), the radiomics signature of ADC had the best AUC, accuracy, specificity, and sensitivity for differentiating PC and MFCP. Furthermore, the differences in histological features between PC and MFCP might be attributed to their large variance in ADC. PC is a highly fibrotic malignancy, which may consist of more intense cellular density and denser fibrin matrix than MFCP, thus presenting with an even lower ADC value [32]. In a study by Lee et al., ADC images analyzed by abdominal radiologists showed a specificity of 69.2% and a sensitivity of 87.2% for differentiating between PC and MFCP [33]. Compared with the traditional manual imaging evaluation, the radiomics signature of ADC not only showed a high sensitivity of 0.895, but also displayed a high specificity of 0.875, thereby suggesting a lower misdiagnosis rate. Our results confirmed that radiomics signature was more effective and reliable for distinguishing PC from MFCP, as the imaging texture features could distinguish differences in tumor microarchitecture, intratumor heterogeneity, subtle phenotypic, etc., which are difficult to observe with the naked eye.

A previous study has also used radiomics analysis to distinguish PC from MFCP, but this research only compared the diagnostic efficiency of four single-imaging modalities, including T1WI, T2WI, and the artery and portal phases of dynamic contrast-enhanced MRI. In our study, we built a multiparametric MRI radiomics signature model based on T1WI, T2WI, DWI, and ADC. The mp-MRI model showed the best predictive performance compared to the individual models. The texture features in the mp-MRI radiomic signature included first-order statistics, shape-based features (3D), gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), and gray level size zone matrix (GLSZM). Based on the weight of radiomics features, we suggest that high-order features, such as GLCM, GLRLM, and GLSZM, better reflect tumor heterogeneity and biology, which is consistent with previous studies [34, 35]. The boxplots of the distributions of radiomics scores indicated that the multiparametric MRI-based rad-score was an independent diagnostic factor for the differentiation of PC and MFCP.

In addition, we wanted to explore whether predictive performance would be better when combining MRI-based radiomics signature with clinically independent risk factors. Several imaging and clinical characteristics were analyzed in this study. We found that only serum CA19–9 and CEA levels were significantly different between patients with PC and MFCP in both the training and validation cohorts. Previous studies demonstrated that serum CA19–9 levels could be helpful for differentiating PC from MFCP; however, its false negative rate was high [36, 37]. Sakamoto et al. applied CEA, which is a glycoprotein that can be elevated in PC, to monitor the prognosis of patients with PC, but its specificity was low [38]. These results suggested that the clinically

independent risk factors perform poorly in differentiating PC and MFCP. In our study, we established a clinical model by combining serum CA19–9 and CEA levels to identify PC and MFCP; however, the results showed that the AUC, accuracy, specificity, and sensitivity were low, which was consistent with previous studies. For these reasons, separate clinically independent risk factors could be used as a reference for the diagnosis of these two diseases.

Therefore, we constructed a comprehensive model incorporating clinical factors (CA19-9 and CEA levels) and MRI-based radiomics signature, which showed excellent performance and superior diagnostic accuracy in the differentiation of PC and MFCP in both training and validation cohorts. Finally, the calibration curve indicated adequate consistency between the predicted risk of the mixed model and the actual outcome. DCA showed that the mixed model surpassed the solely radiomic signature across a wide range of threshold probabilities, which revealed that clinically independent risk factors added incremental value to diagnostic accuracy. Our mixed model had several advantages. The data used in the model showed beneficial results, which were easily accessible and cost less. Compared with invasive biopsies, non-invasive radiomic analysis can be widely applied to patients. Furthermore, biopsies could lead to sampling bias because of intra-tumoral heterogeneity, whereas the radiomic signature represents a comprehensive evaluation of the whole tumor.

Our study had some limitations. First, the dynamic contrast-enhanced MRI, including the arterial and portal venous phases, was disregarded because we wanted to ensure sufficient samples to develop the model. Second, this was a retrospective study, which might have resulted in selection bias. Third, the clinical characteristics analyzed were not sufficient. Finally, although our set is one of the largest cohorts regarding radiomics and differential diagnosis of PC and MFCP lesions, the sample size is relatively small, and the study lacks of external validation. Large number of samples and external validation are necessary for the development of a model for clinical application. Large-scale and multicenter studies should be conducted in the future.

Conclusions

We developed and preliminarily validated a novel model integrating a multiparametric MRI-based radiomic signature and clinically independent risk factors to distinguish PC from MFCP. An accurate differential diagnosis may aid in formulating treatment strategies and may help to avoid unnecessary surgical operations. Although the advantages and results are promising, this prediction model still needs to be explored in a larger sample size.

Author contributions

Jingjing Liu: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing - original draft. Lei Hu: Conceptualization, Formal analysis, Methodology. Bi Zhou: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. Chuangen Wu: Conceptualization, Formal analysis, Writing - review & editing. Yingsheng Cheng: Conceptualization, Formal analysis, Writing - review & editing.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgment

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2022.101357.

References

[1] P. Rawla, T. Sunkara, V. Gaduputi, Epidemiology of pancreatic cancer: global trends, etiology and risk factors, World J. Oncol. 10 (1) (2019) 10–27.
[2] C. Neuzillet, et al., Pancreatic cancer: French clinical practice guidelines for diagnosis, treatment and follow-up (SNFGE, FFCD, GERCOR, UNICANCER, SFCD, SFED, SFRO, ACHBT, AFC), Dig. Liver Dis. 50 (12) (2018) 1257–1271.
[3] H. Sung, et al., Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (3) (2021) 209–249.
[4] G. Pandya, et al., A comprehensive review of the multifaceted role of the microbiota in human pancreatic carcinoma, Semin. Cancer Biol. (2021).
[5] Y. Zen, V. Deshpande, Tumefactive inflammatory diseases of the pancreas, Am. J. Pathol. 189 (1) (2019) 82–93.
[6] J. Mössner, et al., Chronic pancreatitis (2011).
[7] T. Ichikawa, et al., Duct-penetrating sign at MRCP: usefulness for differentiating inflammatory pancreatic mass from pancreatic carcinomas, Radiology 221 (1) (2001) 107–116.
[8] W.I. Chang, et al., The clinical and radiological characteristics of focal mass-forming autoimmune pancreatitis: comparison with chronic pancreatitis and pancreatic cancer, Pancreas 38 (4) (2009) 401–408.
[9] S. Aslan, et al., Efficacy of perfusion CT in differentiating of pancreatic ductal adenocarcinoma from mass-forming chronic pancreatitis and characterization of isoattenuating pancreatic lesions, Abdom. Radiol. 44 (2) (2019) 593–603.
[10] Y. Deng, et al., Radiomics model based on MR images to discriminate pancreatic ductal adenocarcinoma and mass-forming chronic pancreatitis lesions, Front. Oncol. 11 (2021), 620981.
[11] Y. Sugiyama, et al., Characteristic magnetic resonance features of focal autoimmune pancreatitis useful for differentiation from pancreatic cancer, Jpn J Radiol 30 (4) (2012) 296–309.
[12] Z. Ruan, et al., Multi-modality imaging features distinguish pancreatic carcinoma from mass-forming chronic pancreatitis of the pancreatic head, Oncol. Lett. 15 (6) (2018) 9735–9744.
[13] Q. Yin, et al., Pancreatic ductal adenocarcinoma and chronic mass-forming pancreatitis: differentiation with dual-energy MDCT in spectral imaging mode, Eur. J. Radiol. 84 (12) (2015) 2470–2476.
[14] K. Sandrasegaran, et al., Use of diffusion-weighted MRI to differentiate chronic pancreatitis from pancreatic cancer, AJR Am. J. Roentgenol. 201 (5) (2013) 1002–1008.
[15] M. Klauss, et al., Intravoxel incoherent motion MRI for the differentiation between mass forming chronic pancreatitis and pancreatic carcinoma, Invest. Radiol. 46 (1) (2011) 57–63.
[16] H. Nasief, et al., Improving treatment response prediction for chemoradiation therapy of pancreatic cancer using a combination of delta-radiomics and the clinical biomarker CA19-9, Front. Oncol. 9 (2019) 1464.
[17] S. Yoshinaga, et al., Role of endoscopic ultrasound-guided fine needle aspiration (EUS-FNA) for diagnosis of solid pancreatic masses, Dig. Endosc. 23 (Suppl 1) (2011) 29–33.
[18] G.D. Tourassi, Journey toward computer-aided diagnosis: role of image texture analysis, Radiology 213 (2) (1999) 317–320.
[19] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2) (2016) 563–577.
[20] B. Zhang, S. Zhang, The potential of radiomics in the assessment of intestinal fibrosis in Crohn's disease, Gastroenterology (2021).
[21] B. Badic, et al., Prediction of recurrence after surgery in colorectal cancer patients using radiomics from diagnostic contrast-enhanced computed tomography: a two-center study, Eur. Radiol. (2021).
[22] A. Blazevic, et al., Predicting symptomatic mesenteric mass in small intestinal neuroendocrine tumors using radiomics, Endocr. Relat. Cancer 28 (8) (2021) 529–539.
[23] L.R. Felicio, et al., Correlation between trochlear groove depth and patellar position during open and closed kinetic chain exercises in subjects with anterior knee pain, J. Appl. Biomech. 28 (3) (2012) 335–342.
[24] J. Owen, D. Stephens, J.G. Wright, Reliability of hip range of motion using goniometry in pediatric femur shaft fractures, Can. J. Surg. 50 (4) (2007) 251–255.
[25] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making 26 (6) (2006) 565–574.
[26] S.F. Crinò, et al., Comparison between EUS-guided fine-needle aspiration cytology and EUS-guided fine-needle biopsy histology for the evaluation of pancreatic neuroendocrine tumors, Pancreatology 21 (2) (2021) 443–450.
[27] S.F. Crinò, et al., Endoscopic ultrasound-guided fine-needle biopsy with or without rapid on-site evaluation for diagnosis of solid pancreatic lesions: a randomized controlled non-inferiority trial, Gastroenterology 161 (3) (2021), 899–909.e5.
[28] J.H. Kwon, et al., Differentiating focal autoimmune pancreatitis and pancreatic ductal adenocarcinoma: contrast-enhanced MRI with special emphasis on the arterial phase, Eur. Radiol. 29 (11) (2019) 5763–5771.
[29] R. De Robertis, et al., Intravoxel incoherent motion diffusion-weighted MR imaging of solid pancreatic masses: reliability and usefulness for characterization, Abdom Radiol 44 (1) (2019) 131–139.

[30] R. Fattahi, et al., Pancreatic diffusion-weighted imaging (DWI): comparison between mass-forming focal pancreatitis (FP), pancreatic cancer (PC), and normal pancreas, J. Magn. Reson. Imaging 29 (2) (2009) 350–356.

[31] W.C. Huang, et al., Differentiation between pancreatic carcinoma and mass-forming chronic pancreatitis: usefulness of high b value diffusion-weighted imaging, J. Dig. Dis. 12 (5) (2011) 401–408.

[32] A. Lemke, et al., Differentiation of pancreas carcinoma from healthy pancreatic tissue using multiple b-values: comparison of apparent diffusion coefficient and intravoxel incoherent motion derived parameters, Invest. Radiol. 44 (12) (2009) 769–775.

[33] S.S. Lee, et al., Quantitative analysis of diffusion-weighted magnetic resonance imaging of the pancreas: usefulness in characterizing solid pancreatic masses, J. Magn. Reson. Imaging 28 (4) (2008) 928–936.

[34] T.Y. Tang, et al., Development of a novel multiparametric MRI radiomic nomogram for preoperative evaluation of early recurrence in resectable pancreatic cancer, J. Magn. Reson. Imaging 52 (1) (2020) 231–245.

[35] S. Park, et al., Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features, Diagn. Interv. Imaging 101 (9) (2020) 555–564.

[36] S.B. Su, et al., Carbohydrate antigen 19-9 for differential diagnosis of pancreatic carcinoma and chronic pancreatitis, World J. Gastroenterol. 21 (14) (2015) 4323–4333.

[37] X. Gu, R. Liu, Application of 18F-FDG PET/CT combined with carbohydrate antigen 19-9 for differentiating pancreatic carcinoma from chronic mass-forming pancreatitis in Chinese elderly, Clin. Interv. Aging 11 (2016) 1365–1370.

[38] Y. Sakamoto, et al., Post-chemotherapeutic CEA and CA19-9 are prognostic factors in patients with colorectal liver metastases treated with hepatic resection after oxaliplatin-based chemotherapy, Anticancer Res. 35 (4) (2015) 2359–2368.