European Radiology (2019) 29:4408–4417
https://doi.org/10.1007/s00330-018-5824-1

GASTROINTESTINAL

Radiomics model of contrast-enhanced computed tomography for predicting the recurrence of acute pancreatitis

Yong Chen[1] · Tian-wu Chen[1] · Chang-qiang Wu[2] · Qiao Lin[1] · Ran Hu[1] · Chao-lian Xie[1] · Hou-dong Zuo[1] · Jia-long Wu[3] · Qi-wen Mu[4] · Quan-shui Fu[5] · Guo-qing Yang[5] · Xiao Ming Zhang[1]

Received: 27 June 2018 / Revised: 19 September 2018 / Accepted: 9 October 2018 / Published online: 9 November 2018

Abstract

Objectives To predict the recurrence of acute pancreatitis (AP) by constructing a radiomics model of contrast-enhanced computed tomography (CECT) at AP first attack.

Methods We retrospectively enrolled 389 first-attack AP patients (271 in the primary cohort and 118 in the validation cohort) from three tertiary referral centers; 126 and 55 patients endured recurrent attacks in each cohort. Four hundred twelve radiomics features were extracted from arterial and venous phase CECT images, and clinical characteristics were gathered to develop a clinical model. An optimal radiomics signature was chosen using a multivariable logistic regression or support vector machine. The radiomics model was developed and validated by incorporating the optimal radiomics signature and clinical characteristics. The performance of the radiomics model was assessed based on its calibration and classification metrics.

Results The optimal radiomics signature was developed based on a multivariable logistic regression with 10 radiomics features. The classification accuracy of the radiomics model well predicted the recurrence of AP for both the primary and validation cohorts (87.1% and 89.0%, respectively). The area under the receiver operating characteristic curve (AUC) of the radiomics model was significantly better than that of the clinical model for both the primary (0.941 vs. 0.712, $p = 0.000$) and validation (0.929 vs. 0.671, $p = 0.000$) cohorts. Good calibration was observed for all the models ($p > 0.05$).

Conclusions The radiomics model based on CECT performed well in predicting AP recurrence. As a quantitative method, radiomics exhibits promising performance in terms of alerting recurrent patients to potential precautions.

Key Points

- *The incidence of recurrence after an initial episode of acute pancreatitis is high, and quantitative methods for predicting recurrence are lacking.*
- *The radiomics model based on contrast-enhanced computed tomography performed well in predicting the recurrence of acute pancreatitis.*
- *As a quantitative method, radiomics exhibits promising performance in terms of alerting recurrent patients to the potential need to take precautions.*

Yong Chen and Tian-wu Chen contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5824-1) contains supplementary material, which is available to authorized users.

✉ Xiao Ming Zhang
zhangxm@nsmc.edu.cn; cjr.zhxm@vip.163.com

[1] Sichuan Key Laboratory of Medical Imaging and Department of Radiology, Affiliated Hospital of North Sichuan Medical College, No. 63, Wenhua Road, Nanchong 637000, Sichuan, China

[2] Sichuan Key Laboratory of Medical Imaging and School of Medical Imaging, North Sichuan Medical College, Nanchong, Sichuan, China

[3] Department of Radiology, The Second Clinical Medical College of North Sichuan Medical College Nanchong Central Hospital, Nanchong, Sichuan, China

[4] Department of Medical Imaging & Imaging Institute of Rehabilitation and Development of Brain Function, The Second Clinical Medical College of North Sichuan Medical College Nanchong Central Hospital, Nanchong, Sichuan, China

[5] Department of Radiology, Suining Central Hospital, Suining, Sichuan, China

Keywords Radiomics · Tomography, X-ray computed · Acute pancreatitis · Recurrence

Abbreviations

AP	Acute pancreatitis
AUC	Area under the receiver operating characteristic curve
CECT	Contrast-enhanced computed tomography
CTSI	Computed tomography severity index
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
ICC	Intraclass correlation coefficient
LASSO	Least absolute shrinkage and selection operator
NPV	Negative predictive value
PACS	Picture archiving and communication system
PPV	Positive predictive value
RAC	Revised Atlanta Criteria
RAP	Recurrent acute pancreatitis
ROC	Receiver operating characteristic curve
ROI	Region of interest
SVM	Support vector machine

Introduction

Acute pancreatitis (AP) is one of the most common acute abdominal conditions and is the cause of up to 275,000 annual hospital admissions in the USA [1]. The incidence of recurrence after an initial episode of AP has reached approximately 22%, and some of these patients ultimately progress to chronic pancreatitis (CP) [2]. Few studies have reported on the recurrence of AP despite the term recurrent acute pancreatitis (RAP) being in use for nearly 70 years [3].

There is a paucity of quantitative methods for predicting AP recurrence. Factors that have been reported to be independent predictors of RAP include alcoholism, smoking, and pancreatic necrosis [4–6]. Recommendations for the prevention of AP recurrence based on the removal or management of risk factors are somewhat effective but seem not to be clinically reliable due to a lack of individual specificity [7]. Contrast-enhanced computed tomography (CECT) is the recommended imaging method for diagnosing AP but is not useful for predicting predisposition to recurrence. Radiomics is a new field describing methodologies that non-invasively provide information on diseases by quantitatively analyzing a large number of features extracted from traditional medical images [8, 9]. Radiomics offers insight into the heterogeneity of lesions that are unobservable with the naked eye; thus, such methods are becoming central to cancer research for achieving the potential of precision medicine [10]. Given the nature of textural analysis, the radiomics approach is applicable to all diseases, although its current main use lies in diagnosing cancer [8].

Underlying differences in RAP may exist after the first episode, and timely intervention for RAP after an initial attack rather than after a second attack should benefit high-risk patients. In this study, we aimed to create and validate a novel radiomics model based on CECT images incorporating the radiomics signature and clinical characteristics to quantitatively predict the risk of recurrence after an initial attack of AP.

Materials and methods

Patients

This study was conducted in three tertiary referral centers based in northeastern Sichuan Province. All procedures involving human participants adhered to the tenets of the Declaration of Helsinki. Ethical approval to perform this study was obtained from local IRBs, and informed consent was waved.

Based on recurrence outcomes, patients were classified into AP and RAP groups. Patients in the AP group were enrolled based on the following criteria: AP patients admitted with a first attack from January 2010 to June 2014 who were followed up by means of telephone and medical records through May 2018 to verify that no recurrent attacks occurred during the follow-up period. Patients in the RAP group were enrolled if they were AP patients admitted with a first attack and were followed up with a documented readmission for AP from January 2010 to May 2018. The mean follow-up period for the AP group was 62.8 ± 6.7 months (ranging from 48.3 to 88.5 months), and the mean interval between the first attack and second attack for the RAP group was 12.5 ± 3.6 months (ranging from 3.2 to 59.6 months).

Finally, 389 AP patients suffering initial attacks (mean age 49.20 ± 15.44 years; ranging from 18 to 85 years; males, 238) were recruited for the retrospective study (208 were assigned to the AP group and 181 were assigned to the RAP group). All patients were randomly allocated to the primary and validation cohorts at a ratio of 7:3. Therefore, 271 admissions of sentinel AP were allocated to the primary cohort, of whom 126 patients were in the RAP group. The remaining 118 first-attack AP patients, including 55 patients in the RAP group, were allocated to the validation group. Eight clinical patient characteristics, including gender, age, etiology, CT severity index (CTSI), disease severity based on the 2012 Revised Atlanta Criteria (RAC) [11], hospital stay, pancreatic necrosis, and smoking status, were collected from admission records. Detailed patient information is presented in Supplementary S1, and the flowchart of the study is shown in Fig. 1.

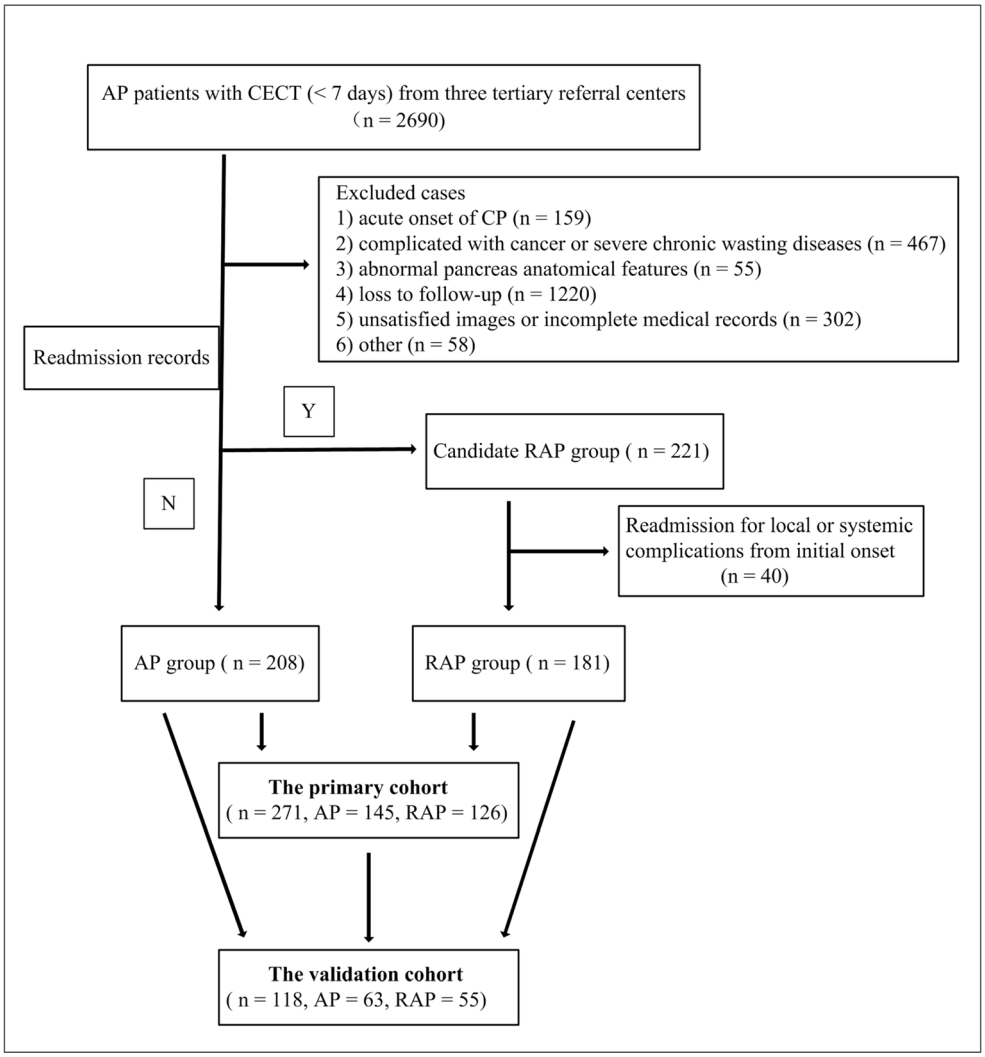

Fig. 1 Flow chart of patient recruitment in this study. CECT, contrast-enhanced computed tomography; AP, acute pancreatitis; RAP, recurrent acute pancreatitis

CECT image acquisition and retrieval procedure

All patients recruited in this study underwent a contrast-enhanced abdominal CT within one week of symptom onset using one of three multidetector row CT systems. Detailed information regarding the image acquisition is presented in Supplementary S2. Arterial and portal venous phase CT images were anonymously retrieved from a picture archiving and communication system (PACS) (Carestream) for feature extraction.

Image segmentation, preprocessing, and feature extraction

Two radiologists with 5 and 12 years of experience in abdominal imaging who were blinded to the patients' clinical outcomes delineated the region of interest (ROI) around the pancreatic parenchyma including pancreatic necrosis slice by slice and avoided vessels and the common bile duct. Manual delineation was performed using IBEX (β1.0, http://bit.ly/IBEX_MDAnderson) (Fig. 2) [12], an open-source software program that runs on MATLAB 2015b (MathWorks Inc). Four common feature groups, including intensity histogram, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and shape, were extracted from IBEX (Supplementary Table 1). In all, 412 radiomics features were extracted from the arterial and portal venous phase images of CECT. To guarantee the repeatability of the results, resampling and *z*-score normalization were performed as preprocessing steps for images and data, respectively (Supplementary S3).

Intra-observer and inter-observer agreement

The reproducibility of intra-observer and inter-observer agreement for radiomics features was measured using 50 randomly chosen samples drawn from the arterial and venous phases of three CT scanners by two blinded radiologists. To evaluate intra-observer reproducibility, reader 1 performed the ROI delineation twice within one week following the same

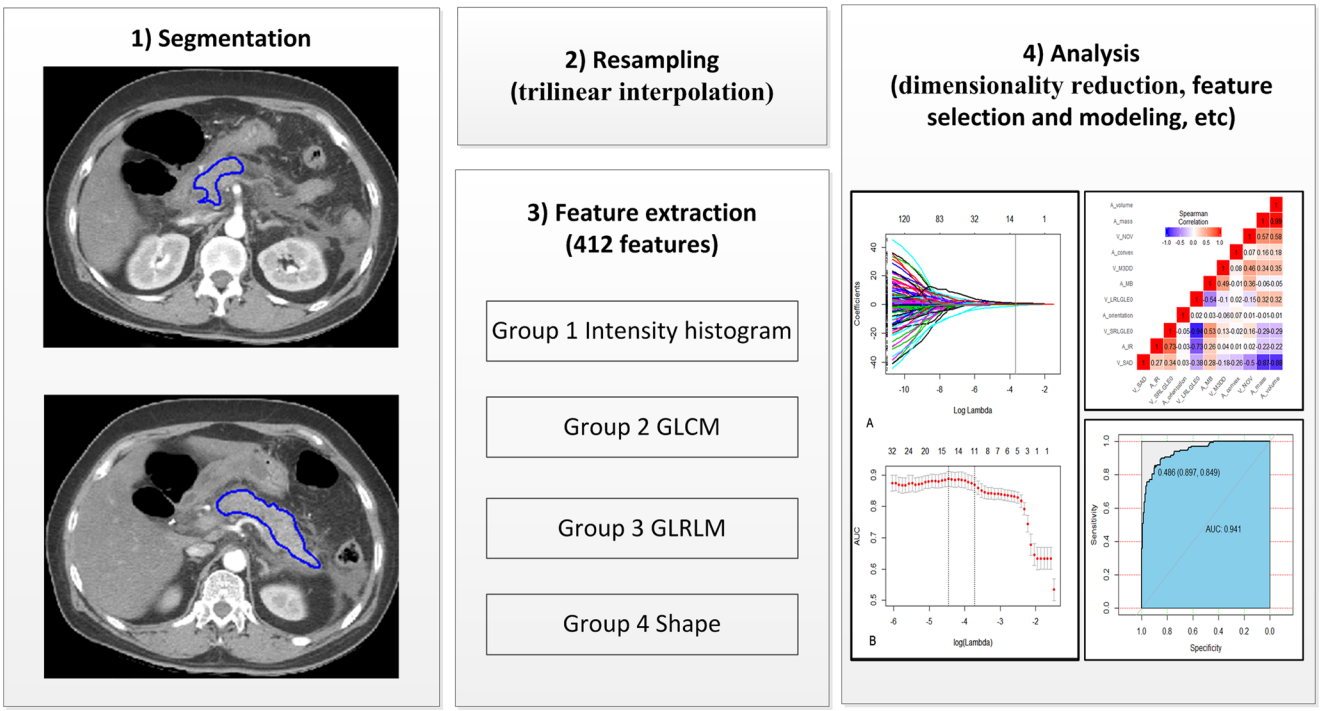

Fig. 2 Framework of this study. GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix

procedure. At the same time, reader 2 independently delineated the ROI once to assess inter-observer agreement by comparing the results with the radiomics features extracted from the first ROI delineation by reader 1. The intraclass correlation coefficient (ICC) was used to evaluate intra-observer and inter-observer agreement. An ICC score greater than 0.75 was considered to indicate satisfactory agreement. It is unrealistic that all radiomics features will fulfill the criteria for satisfactory agreement due to the nature of the variability resulting from the voxel-size and gray-level dependency [13]. Therefore, strong agreement was considered as 90% of the radiomics features having satisfactory agreement. Reader 1 accounted for the remaining samples if strong agreement was achieved.

Dimensionality reduction and radiomics feature selection

To avoid the curse of dimensionality and reduce the bias from radiomics features when modeling, three steps were adopted to select the features in the primary cohort. First, all features were tested by the independent samples *t* test or the Mann-Whitney *U* test to select potential important features for the primary cohort. Features that did not meet either of the above tests were excluded. Second, least absolute shrinkage and selection operator (LASSO) was used for dimensionality reduction and feature selection by performing variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model produced [14, 15]. The 1-standard error of the minimum criteria (the 1-SE criteria, a simpler model) was used to tune the regularization parameter (λ) and for feature selection using 10-fold cross-validation. Finally, Spearman correlation coefficients were calculated for the features selected by LASSO to avoid the underlying severe linear dependence. The strength of the correlation coefficients was established [16]. In this study, we considered features with very high positive correlations (0.90~1.00) to have severe linear dependence.

Development of the optimal radiomics signature

Two classical machine learning methods, logistic regression and the support vector machine (SVM), were used to produce a better predictive radiomics signature for the primary cohort in terms of the problem of whether the data were linearly separable. The radiomics signature based on selected radiomics features was developed using multivariable binary logistic regression with backward stepwise selection to build a linear classifier. A SVM with a Gaussian kernel was then used with the selected radiomics features to build a non-linear classifier. Optimizing the parameters of kernel size (γ, gamma, $\in [0.0001, 0.1]$) and the regularization parameters (C, cost, $\in [1, 10{,}000]$) for the SVM kernel function were realized by 10-fold cross-validation, which was capable of selecting the best performing signature. The assessment of the optimal radiomics signature mainly relied on the area under the receiver operating characteristic curve (AUC). Other discrimination metrics, including classification accuracy, positive predictive

value (PPV), negative predictive value (NPV), sensitivity, and specificity, were also calculated. Each radiomics signature was also assessed according to all of the metrics for the validation cohort.

Construction of the radiomics model

The optimal radiomics signature was then incorporated with the clinical characteristics to create a predictive radiomics model. The tuning of the optimal radiomics model followed the same procedures as those described above. For comparison, an independent clinical model was also created based on the same machine learning method used in the radiomics model. Model calibration was assessed using the Hosmer-Lemeshow test for logistic regression or kappa coefficients for the SVM. The discriminatory performance of the models was based on the classification metrics described above for both the primary and validation cohorts. An AUC comparison of the two models was also performed to determine the best model for predicting the recurrence of sentinel AP.

Statistics

All statistical analyses were conducted using R (Version 3.4.1, https://www.r-project.org/). Clinical characteristics were measured based on the variable type. The Shapiro–Wilk test was used to assess the normality of distributions, and the homogeneity of variance was tested using Bartlett's test. Continuous variables are presented as the means or medians and were compared using independent *t* tests or Wilcoxon Rank Sum tests based on their distributions. Categorical variables were measured as proportions and were compared using chi-square tests or Fisher's exact tests. LASSO regression based on multivariate binary logistic regression was performed using the "*glmnet*" package. The correlation coefficient matrix was plotted using the "*ggplot2*" package. SVM models and ROC curves were created using the "*e1071*" and "*pROC*" packages, respectively. Differences were considered significant at $p < 0.05$.

Results

Clinical characteristics

The clinical characteristics of sentinel AP patients of the primary and validation cohorts were recorded (Table 1). There was no significant difference between the two cohorts in terms of recurrence prevalence (46.5% and 46.6% in the primary and validation cohorts, respectively, $p = 0.983$). Of the eight characteristics measured, only age and etiology were significantly different for the primary and validation cohorts ($p < 0.05$). In the subgroup analysis for etiology, alcoholic, hypertriglyceridemic, and idiopathic etiologies were found to be independent risk factors for AP recurrence ($p < 0.05$). Moreover, hypertriglyceridemia was the most dangerous etiology (OR [95% CI], hypertriglyceridemic vs. alcoholic vs. idiopathic = 6.825 [3.128–15.658] vs. 2.318 [1.165–4.705] vs. 2.366 [1.142–4.990]). Male dominance was obvious in alcoholic patients. No significant difference was observed for recurrence prevalence according to sex or etiology in either of the cohorts (Supplementary Table 2). All of the clinical characteristics were applied to build the radiomics and clinical model.

Inter-observer and intra-observer agreement

For intra-observer agreement, the rate of satisfactory agreement for all 412 features reached 97.1% (mean ICC = 0.892, ranging from 0.004 to 0.997, Fig. 3a). For inter-observer agreement, the rate of satisfactory agreement for all 412 features reached 92.5% (mean ICC = 0.868, ranging from 0.001 to 0.997, Fig. 3b). For features with unsatisfactory agreement, the 12 features excluded from intra-observer agreement were also involved in the inter-observer agreement among the 31 features (Supplementary Table 3). Finally, 31 features were excluded, and the remaining 381 features were included in further analyses. Reader 1 completed the segmentation of all of the samples.

Dimensionality reduction and feature selection

Sixty-three features showed a Gaussian distribution with homoscedasticity, but none were significant based on independent samples *t* tests (*p* values ranging from 0.676 to 0.847). Among the remaining 349 features, Mann-Whitney *U* tests showed that 265 features were significantly different (*p* values ranging from 0.000 to 0.046). Therefore, a total of 265 features were used for LASSO regression. Eleven features were selected by LASSO, with the best tuned regularization parameter λ of 0.024 under the 1-SE criteria found by 10-fold cross-validation (Fig. 4, Supplementary Table 4). When Spearman correlation coefficients were calculated for the 11 features, one pair of features showed extremely strong positive correlations (mass and volume 0.99, $p = 0.000$, 95% CI 0.991~0.994; Supplementary Fig. 1). As a consequence, mass was excluded, and the remaining 10 features were used for further modeling.

Development of the optimal radiomics signature

Good performance of the radiomics signature based on multivariable logistic regression was observed in the primary cohort with an AUC of 0.919 (95% CI 0.886~0.953) and a classification accuracy of 86.0% (95% CI 0.813~0.899). The performance of the SVM radiomics signature was similar to that of the logistic signature, with a tuned best γ of 0.0001 and a *C* of 10,000 (AUC 0.913, 95% CI 0.876~0.949; classification accuracy 85.2%, 95% CI 0.805~0.892). Good

Table 1 Clinical characteristics of the primary and validation cohorts

	The primary cohort			The validation cohort		
	RAP ($n = 126$)	AP ($n = 145$)	p	RAP ($n = 55$)	AP ($n = 63$)	p
Gender			0.220			0.044*
Male	83 (65.9%)	85 (58.6%)		38 (69.1%)	32 (50.8%)	
Female	43 (34.1%)	60 (41.4%)		17 (30.9%)	31 (49.2%)	
Age (years)	50.42 ± 15.44	44.50 ± 13.91	0.020*	54.54 ± 16.55	46.66 ± 15.78	0.009*
Etiology			0.000*			0.001*
Biliary	20 (15.9%)	49 (33.8%)		12 (21.8%)	31 (49.2%)	
Alcoholic	35 (27.8%)	37 (22.5%)		15 (27.3%)	10 (15.9%)	
Hypertriglyceridemia	39 (31.0%)	14 (9.7%)		10 (18.2%)	5 (7.9%)	
Idiopathic	28 (22.2%)	29 (20.0%)		16 (29.1%)	8 (12.7%)	
Others	4 (3.2%)	16 (11.0%)		2 (3.6%)	9 (14.3%)	
CTSI	3 (2~4)	3 (2~4)	0.727	3(2~4)	3(2~4)	0.886
Disease severity			0.388			0.997
Mild	43 (34.1%)	40 (27.6%)		18 (32.7%)	21 (33.3%)	
Moderate	61 (48.4%)	82 (56.6%)		29 (52.7%)	33 (52.4%)	
Severe	22 (17.5%)	23 (15.9%)		8 (14.5%)	9 (14.3%)	
Hospital stay (days)	13 (10~20)	14 (10~20)	0.632	12 (9~15)	12 (10~17)	0.407
Pancreatic necrosis			0.321			0.972
Yes	16 (12.7%)	13 (9.0%)		6 (10.9%)	7 (11.1%)	
No	110 (87.3%)	132 (91.0%)		49 (89.1%)	56 (88.9%)	
Smoking			0.074			0.583
Yes	55 (43.7%)	48 (33.1%)		21 (38.2%)	21 (33.3%)	
No	71 (56.3%)	97 (66.9%)		34 (61.8%)	42 (66.7%)	

RAP recurrent acute pancreatitis, *AP* acute pancreatitis, *CTSI* computed tomography severity index

*p value < 0.05

performance and calibration were also observed for both radiomics signatures for the validation cohort (Table 2). Due to the similar performance, SVM required more complex computational processes, and the results indicated that the radiomics features that predicted recurrence of sentinel AP were linearly separable. Therefore, the optimal radiomics signature was based on a multivariable logistic regression.

Construction of the radiomics model

Good performance of the radiomics model for the primary cohort was observed, with an AUC of 0.941 (95% CI 0.915~0.967) and a classification accuracy of 87.1% (95% CI 0.825~0.908) (Table 3). No significance was found using the Hosmer–Lemeshow test in the primary cohort ($\chi^2 = 9.068$, $p = 0.337$), indicating good calibration of the radiomics model. The AUC of the independent clinical logistic model in the primary cohort was 0.712 (95% CI 0.651~0.774) and showed a classification accuracy of 67.5% (95% CI 0.616~0.731). The calibration of the clinical model also performed well ($\chi^2 = 3.770$, $p = 0.877$). The AUC of the radiomics model was significantly better than that of the clinical model ($p = 0.000$, Fig. 5a).

The radiomics logistic model also showed good performance for predicting the recurrence of AP in the validation cohort (AUC [95% CI], 0.929 [0.878~0.980]; classification accuracy [95% CI], 89.0% [0.819~0.940]) (Table 3). The AUC and classification accuracy of the clinical model for the validation cohort were 0.671 (95% CI, 0.572~0.771) and 61.0% (95% CI, 0.516~0.699), respectively. When comparing the AUCs between the two models, the radiomics model proved to be significantly better than the clinical model ($p = 0.000$, Fig. 5b). Good calibration was also observed for predicting AP recurrence in the validation cohort (both $p > 0.05$).

Discussion

In this study, we developed and validated a quantitative radiomics model based on CECT to provide a non-invasive and individualized prediction tool for the recurrence of AP after a sentinel attack. Our proposed radiomics model showed good

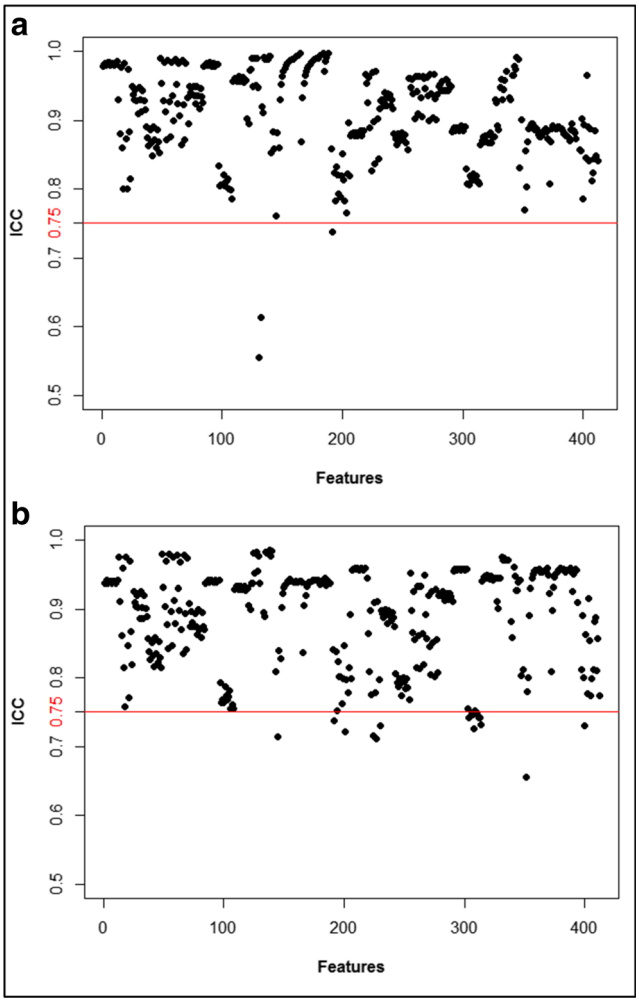

Fig. 3 Evaluation of feature stability and inter-observer and intra-observer agreement based on the interclass correlation coefficient (ICC). **a** All features presented good intra-observer agreement with ICCs of > 0.75 (above the red cutoff line). **b** All features presented good inter-observer agreement with ICCs of > 0.75

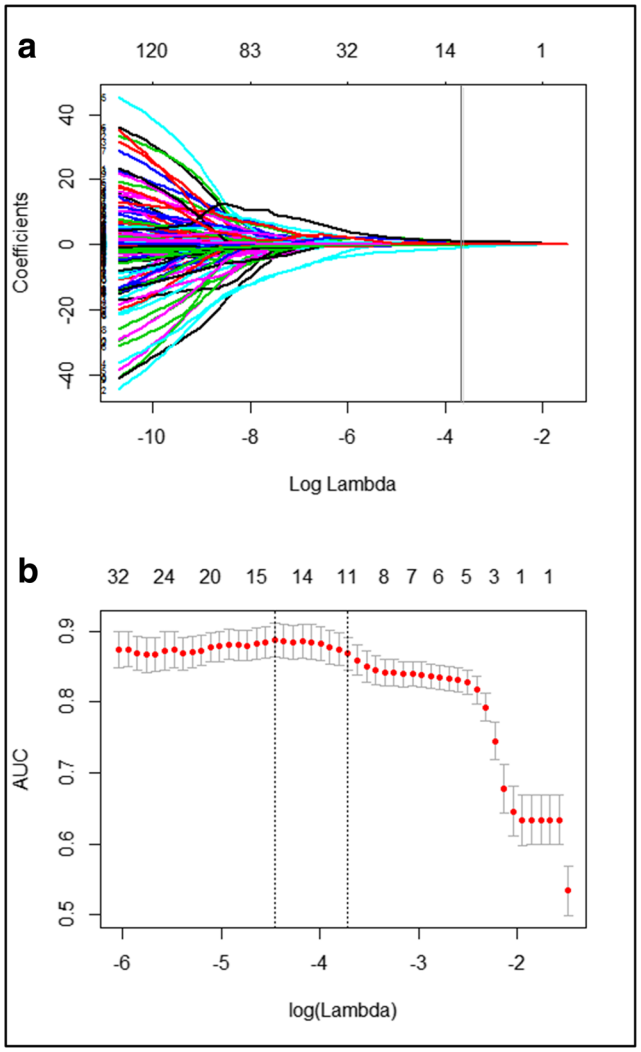

Fig. 4 Feature selection using the least absolute shrinkage and selection operator (LASSO) regression. **a** LASSO coefficient profiles for the 381 radiomics features. The vertical line shows the optimal value of λ (= 0.024) and 23 corresponding features with non-zero coefficients. **b** The AUC curve was plotted by tuning parameter (λ) selection performed by 10-fold cross-validation. Dotted lines on the left and right denote the minimum criterion and 1-standard error criterion (1-SE), respectively. The 1-SE criterion was applied

performance in both the primary and validation cohorts, with the classification accuracies of 87.1% and 89.0% respectively, suggesting that the clinical use of radiomics is promising in terms of alerting recurrent patients to potential episodes.

Quantitative methods for predicting the recurrence of AP are lacking. Independent risk factors for the recurrence of AP observed in both cohorts included etiology and a younger age, as also seen in previous studies [2, 4, 6, 17–20]. It should be noted that as an uncommon etiology for sentinel AP, hypertriglyceridemia appears to be the most dangerous. Hypertriglyceridemia has a high rate of recurrence, with reported rates of 30.1 to 44.2% [21–23]. In our study, hypertriglyceridemia was obviously more frequent in patients with recurrence in both cohorts. However, when a clinical model incorporating clinical characteristics was created to predict AP recurrence, it was found to have limited value for both cohorts.

In contrast, our proposed radiomics model performed well by integrating high-dimensional radiomics features and clinical characteristics. The optimal radiomics signature was tuned by multivariable logistic regression and showed an AUC of 0.919 for the primary cohort. By incorporating clinical characteristics, the performance of the radiomics model improved, and significant differences in AUC were observed for both the primary and validation cohorts when compared with the clinical model ($p = 0.000$); these results indicate the potential clinical use of our radiomics model as a quantitative tool for individually predicting recurrence in AP patients after a sentinel attack.

Table 2 The performance of radiomics signatures built using multivariable logistic regression and a support vector machine for the primary and validation cohorts

		AUC (95% CI)	Accuracy (95% CI)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
The primary cohort	MLR	0.919 (0.886~0.953)	86.0% (0.813~0.899)	84.8	88.6	91.0	80.2
	SVM	0.913 (0.876~0.949)	85.2% (0.805~0.892)	82.6	89.1	91.7	77.8
The validation cohort	MLR	0.897 (0.836~0.957)	84.8% (0.770~0.907)	80.8	91.1	93.7	74.6
	SVM	0.903 (0.843~0.964)	85.6% (0.779~0.914)	81.1	93.2	95.2	74.6

MLR multivariable logistic regression, *SVM* support vector machine, *PPV* positive predictive value, *NPV* negative predictive value

One of the robust aspects of our radiomics model with improved performance is the use of radiomics based on CECT images from three tertiary referral centers. The incorporation of the arterial and portal venous phases can provide more abundant information than either phase alone. Compared with MRI or ultrasound, scanning techniques based on non-ionizing procedures, such as CT, are found to have less intrinsic complexity that challenges the reproducibility of radiomics [24]. CECT is the recommended imaging technique for diagnosing AP, especially when local complications need to be differentiated [11]. The inclusion of patients from multiple regional centers enhanced the strengths of our findings for clinical use. The potential concern that the use of variable CT parameters may lead to the unsatisfactory reproducibility of radiomics features was maximally diminished using resampling as a preprocessing method, as noted previously [13, 25], a method that has been shown to maintain feature stability by optimizing gray-level discretization.

Another robust aspect of our radiomics model with improved performance was the optimization for feature selection and modeling. Univariate analysis, LASSO, and Spearman correlations were used for feature selection, thus guaranteeing the importance and independence of each feature in the final model. The performance of the radiomics signature was robust, as demonstrated by the similar results obtained when we compared two classical machine learning methods. The robustness of the modeling was guaranteed by 10-fold cross-validation and stepwise regression. The radiomics model performed well for the primary cohort, with an AUC of 0.941 and a classification accuracy of 87.1%, by incorporating the radiomics signature with clinical characteristics. Good discrimination of the radiomics model was also observed for the validation cohort (AUC, 0.929; classification accuracy, 89.0%). The calibration of the radiomics model was good, also indicating the robustness of our model.

Radiomics has a strong predictive power, and this is probably because radiomics features are more reflective of quantitative information drawn from images rather than those drawn by the naked eye. Evidence has shown that pancreatitis progression is gradual and that subsequent recurrence, which results in irreversible structural or functional changes to the pancreas, appears to occur abruptly without warning [26], indicating that the existence of changes is possible but that these changes are difficult to detect by traditional means. The individual propensity for recurrence implies that such changes may have occurred before the next episode. Based on this hypothesis, it is possible to predict recurrence after an initial onset of AP. Radiomics uncovers diagnostic information that is hidden in routine medical images, thus indicating an underlying biological basis [27, 28]. In a recent study on intraductal papillary mucinous neoplasms (IPMNs), Permuth et al [29] improved the diagnostic accuracy of a malignant pathology by creating a prediction model combining radiomics features with a miRNA classifier for 38 surgically resected IPMN patients. In our study, hidden information regarding the microenvironment was possibly revealed by radiomics, and further endeavors to disclose the pathophysiology of AP recurrence will contribute to decoding the radiomics findings.

Table 3 The performance of the clinical and radiomics models in the primary and validation cohorts

		Accuracy (95% CI)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC (95% CI)
The primary cohort	The clinical model	67.5% (0.616~0.731)	67.3	67.9	76.6	57.1	0.712 (0.651~0.774)
	The radiomics model	87.1% (0.825~0.908)	86.7	87.6	89.7	84.1	0.941 (0.915~0.967)
The validation cohort	The clinical model	61.0% (0.516~0.699)	60.5	62.2	77.8	41.8	0.671 (0.572~0.771)
	The radiomics model	89.0% (0.819~0.940)	83.8	97.7	98.4	78.2	0.929 (0.878~0.980)

PPV positive predictive value, *NPV* negative predictive value, *AUC* area under the receiver operating characteristic curve

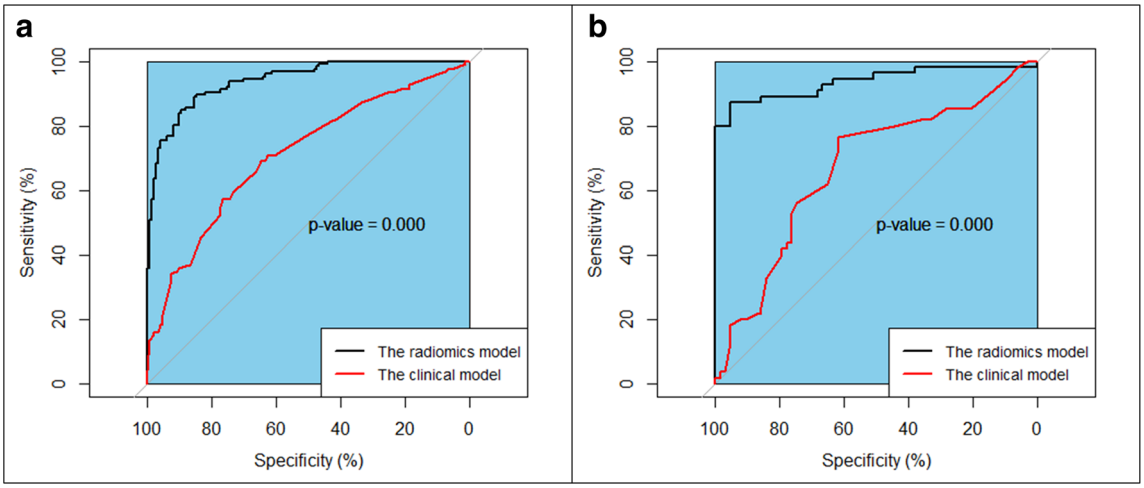

Fig. 5 ROC curves comparing the radiomics model and clinical model between the primary cohort (**a**) and the validation cohort (**b**)

Our study has several limitations. First, although there was a relatively long-term follow-up in our study, patients whose recurrence fell outside the scope of the follow-up period may have resulted in the overestimation of false positive predictions. Second, more clinical characteristics must be considered in a comprehensive evaluation. Finally, our hypothesis and interpretation of radiomics need to be proven by research on the pathophysiology of RAP.

In conclusion, radiomics serves as a promising biomarker for predicting the recurrence of AP. Proper intervention is required once a sentinel attack occurs.

Acknowledgements Thanks are due to Dr. Xin Li for the assistance with statistics and data visualization.

Funding This work was supported by the National Natural Science Foundation of China (Grant No. 81871440) and the Training Program for Science and Technology Innovation of Sichuan Province (Grant No. 2018036).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Xiao Ming Zhang, MD.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry Dr. Xin Li kindly provided statistical advice for this manuscript.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic or prognostic study
- multicenter study

References

1. Forsmark CE, Vege SS, Wilcox CM (2016) Acute pancreatitis. N Engl J Med 375:1972–1981
2. Sankaran SJ, Xiao AY, Wu LM, Windsor JA, Forsmark CE, Petrov MS (2015) Frequency of progression from acute to chronic pancreatitis and risk factors: a meta-analysis. Gastroenterology 149:1490–1500.e1
3. Doubilet H, Mulholland JH (1948) Recurrent acute pancreatitis: observations on etiology and surgical treatment. Ann Surg 128:609–636
4. Ahmed Ali U, Issa Y, Hagenaars JC et al (2016) Risk of recurrent pancreatitis and progression to chronic pancreatitis after a first episode of acute pancreatitis. Clin Gastroenterol Hepatol 14:738–746
5. Munigala S, Conwell DL, Gelrud A, Agarwal B (2015) Heavy smoking is associated with lower age at first episode of acute pancreatitis and a higher risk of recurrence. Pancreas 44:876–881
6. Gullo L, Migliori M, Pezzilli R et al (2002) An update on recurrent acute pancreatitis: data from five European countries. Am J Gastroenterol 97:1959–1962
7. Nordback I, Pelli H, Lappalainen-Lehto R, Järvinen S, Räty S, Sand J (2009) The recurrence of acute alcohol-associated pancreatitis can be reduced: a randomized controlled trial. Gastroenterology 136:848–855
8. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
9. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48:441–446
10. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762
11. Banks PA, Bollen TL, Dervenis C et al (2013) Classification of acute pancreatitis–2012: revision of the Atlanta classification and definitions by international consensus. Gut 62:102–111
12. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE (2015) IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. Med Phys 42:1341–1353
13. Shafiq-Ul-Hassan M, Zhang GG, Latifi K et al (2017) Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 44:1050–1062
14. Tibshirani R (2011) Regression shrinkage and selection via the Lasso: a retrospective. J R Stat Soc Series B Stat Methodology 73:273–282

15. Huang YQ, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 34:2157–2164
16. Hinkle DE, Wiersma W, Jurs SG (2003) Applied statistics for the behavioral sciences, 5th edn. Houghton Mifflin, Boston
17. Bertilsson S, Swärd P, Kalaitzakis E (2015) Factors that affect disease progression after first attack of acute pancreatitis. Clin Gastroenterol Hepatol 13:1662–1669.e3
18. Takeyama Y (2009) Long-term prognosis of acute pancreatitis in Japan. Clin Gastroenterol Hepatol 7:S15–S17
19. Yadav D, O'Connell M, Papachristou GI (2012) Natural history following the first attack of acute pancreatitis. Am J Gastroenterol 107:1096–1103
20. Lankisch PG, Breuer N, Bruns A, Weber-Dany B, Lowenfels AB, Maisonneuve P (2009) Natural history of acute pancreatitis: a long-term population-based study. Am J Gastroenterol 104:2797–2805 quiz 2806
21. Vipperla K, Somerville C, Furlan A et al (2017) Clinical profile and natural course in a large cohort of patients with hypertriglyceridemia and pancreatitis. J Clin Gastroenterol 51:77–85
22. Qiu L, Sun RQ, Jia RR et al (2015) Comparison of existing clinical scoring systems in predicting severity and prognoses of hyperlipidemic acute pancreatitis in Chinese patients: a retrospective study. Medicine (Baltimore) 94:e957
23. Huang YX, Jia L, Jiang SM, Wang SB, Li MX, Yang BH (2014) Incidence and clinical features of hyperlipidemic acute pancreatitis from Guangdong, China: a retrospective multicenter study. Pancreas 43:548–552
24. Limkin EJ, Sun R, Dercle L et al (2017) Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol 28:1191–1206
25. Larue RTHM, van Timmeren JE, de Jong EEC et al (2017) Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncol 56:1544–1553
26. Whitcomb DC (2013) Genetic risk factors for pancreatic disorders. Gastroenterology 144:1292–1302
27. Grossmann P, Stringfield O, El-Hachem N et al (2017) Defining the biological basis of radiomic phenotypes in lung cancer. Elife 6: e23421
28. Zhou M, Leung A, Echegaray S et al (2018) Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. Radiology 286:307–315
29. Permuth JB, Choi J, Balarunathan Y et al (2016) Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms. Oncotarget 7:85785–85797