BMC Medical Imaging

RESEARCH

Open Access

Deep learning radiomics nomogram for preoperatively identifying moderate-to-severe chronic cholangitis in children with pancreaticobiliary maljunction: a multicenter study

Hui-min Mao[1], Kai-ge Chen[2], Bin Zhu[3], Wan-liang Guo[1*] and San-li Shi[4*]

Abstract

Background Long-term severe cholangitis can lead to dense adhesions and increased fragility of the bile duct, complicating surgical procedures and elevating operative risk in children with pancreaticobiliary maljunction (PBM). Consequently, preoperative diagnosis of moderate-to-severe chronic cholangitis is essential for guiding treatment strategies and surgical planning. This study aimed to develop and validate a deep learning radiomics nomogram (DLRN) based on contrast-enhanced CT images and clinical characteristics to preoperatively identify moderate-to-severe chronic cholangitis in children with PBM.

Methods A total of 323 pediatric patients with PBM who underwent surgery were retrospectively enrolled from three centers, and divided into a training cohort ($n=153$), an internal validation cohort (IVC, $n=67$), and two external test cohorts (ETC1, $n=58$; ETC2, $n=45$). Chronic cholangitis severity was determined by postoperative pathology. Handcrafted radiomics features and deep learning (DL) radiomics features, extracted using transfer learning with the ResNet50 architecture, were obtained from portal venous-phase CT images. Multivariable logistic regression was used to establish the DLRN, integrating significant clinical factors with handcrafted and DL radiomics signatures. The diagnostic performances were evaluated in terms of discrimination, calibration, and clinical usefulness.

Results Biliary stones and peribiliary fluid collection were selected as important clinical factors. 5 handcrafted and 5 DL features were retained to build the two radiomics signatures, respectively. The integrated DLRN achieved satisfactory performance, achieving area under the curve (AUC) values of 0.913 (95% CI, 0.834–0.993), 0.916 (95% CI, 0.845–0.987), and 0.895 (95% CI, 0.801–0.989) in the IVC, and two ETCs, respectively. In comparison, the clinical model, handcrafted signature, and DL signature had AUC ranges of 0.654–0.705, 0.823–0.857, and 0.840–0.872 across the same cohorts. The DLRN outperformed single-modality clinical, handcrafted radiomics, and DL radiomics models, with all integrated discrimination improvement values > 0 and $P<0.05$. The Hosmer–Lemeshow test and calibration curves

*Correspondence:
Wan-liang Guo
gwlsuzhou@163.com
San-li Shi
sanlishi1982@163.com
Full list of author information is available at the end of the article

showed good consistency of the DLRN ($P > 0.05$), and the decision curve analysis and clinical impact curve further confirmed its clinical utility.

Conclusions The integrated DLRN can be a useful and non-invasive tool for preoperatively identifying moderate-to-severe chronic cholangitis in children with PBM, potentially enhancing clinical decision-making and personalized management strategies.

Keywords Children, Chronic cholangitis, Deep learning, Nomogram, Pancreaticobiliary maljunction, Radiomics

Introduction

Pancreaticobiliary maljunction (PBM) is a congenital defect in which the pancreatic and bile ducts join outside of the duodenal wall, usually forming a long common channel [1]. There is a higher incidence of PBM in Asian populations, which is 100 to 1000 times higher than in other parts of the world [2]. This anatomical anomaly induces prolonged reflux of pancreatic juice into the common bile duct (CBD), leading to chronic inflammation of the CBD [3]. PBM is considered to involve a hyperplasia-dysplasia-carcinoma sequence caused by chronic inflammation [3, 4]. Therefore, for children diagnosed with PBM, surgery is recommended regardless of the presence of symptoms [5].

Chronic cholangitis is a critical factor influencing the clinical management and prognosis of PBM. Long-term severe cholangitis can cause close adhesion between the CBD wall and the surrounding tissues, which increases the difficulty of operation [5–7]. In addition, the infiltration of a large number of chronic inflammatory cells can thicken the CBD wall and increase fragility, thereby elevating the risk of intraoperative injury [5, 6]. Importantly, such chronic inflammatory changes are significantly associated with postoperative anastomotic stricture [8]. Therefore, accurate preoperative diagnosis of moderate-to-severe chronic cholangitis in PBM is crucial for optimizing surgical planning, minimizing intraoperative risks, and improving patient outcomes.

Magnetic resonance imaging (MRI) is the preferred modality in evaluating PBM [9]. However, due to the abdominal breathing artifacts in pediatric MRI examinations, contrast-enhanced computed tomography (CE-CT) remains an effective diagnostic tool for PBM, particularly in younger children [10, 11], and for preoperatively assessing critical anatomical variations in PBM, such as the aberrant right hepatic artery [12, 13]. While CE-CT images are routinely acquired, radiologists often encounter difficulties in diagnosing and grading PBM-associated chronic cholangitis due to the lack of specific imaging criteria. Currently, diagnosis largely relies on postoperative histopathological evaluation. This challenge is particularly pronounced in moderate-to-severe cases, where precise preoperative diagnosis is critical for surgical planning. Thus, it is necessary to develop an objective, reliable, noninvasive and accurate diagnostic approach to improve the preoperative evaluation of chronic cholangitis in children with PBM.

Radiomics is a promising technique that transforms conventional medical images into high-throughput quantitative features, which has been increasingly used in disease diagnosis and prognostic prediction [14, 15]. In addition, when deep learning (DL) features derived from convolutional neural networks are incorporated into handcrafted radiomics, superior performance has been achieved in lesion decoding and prognostic prediction, such as gastric cancer, radiation esophagitis, hepatocellular carcinoma and so on [16–18]. Previous studies have established the feasibility of identifying chronic cholangitis in PBM children using either handcrafted radiomics models or DL models separately [19, 20]. However, these studies were limited by small sample sizes and lack of independent external validation. Notably, no research has yet combined clinical characteristics with both handcrafted and DL radiomics features for preoperative diagnosis of cholangitis.

Therefore, this study aimed to develop and validate a deep learning radiomics nomogram (DLRN) for preoperative identification of moderate-to-severe chronic cholangitis in children with PBM using multicenter datasets, offering a non-invasive and effective alternative to histopathological examination for early detecting choledochal inflammation.

Methods

Study patients

This retrospective study was approved by the review boards of each participating center, with waived patient informed consent. Pediatric patients with PBM who received preoperative CE-CT examination followed by pancreaticobiliary surgery between January 2011 and February 2023 from three independent hospitals were retrospectively reviewed. In children, PBM was diagnosed based on an abnormally long common channel (> 5 mm) between the pancreatic and bile ducts, as confirmed by surgery or imaging examination [21]. The inclusion criteria were as follows: (1) age ≤ 18 years old; (2) availability of histopathological slides of CBD; (3) abdominal CE-CT examination performed within one

month before surgery; and (4) availability of clinicopathological and imaging data. The exclusion criteria were as follows: (1) poor CT image quality impeding lesion segmentation; (2) incomplete clinicopathological and imaging data; (3) prior history of biliary drainage; and (4) previous abdominal surgery for other diseases. The flow of patient selection is shown in Fig. S1.

A total of 323 eligible children with PBM were included in the final analysis. Of these, 220 were from center I (Children's hospital of Soochow university) and were randomly divided into a training cohort (TC, $n=153$) and an internal validation cohort (IVC, $n=67$) at a 7:3 ratio. 58 patients from center II (Xuzhou Children's Hospital) and 45 patients from center III (Shandong Provincial Hospital) were collected as external test cohort 1 (ETC1) and external test cohort 2 (ETC2), respectively. Details of the sample size calculation are provided in Supplementary Method S1. The study flowchart is presented in Fig. 1.

Assessment of moderate-to-severe chronic cholangitis

The severity of chronic cholangitis was evaluated by two experienced pathologists through histopathological examination of resected CBD specimens. The grading criteria were as follows: grade 0, no inflammation; grade 1, infiltration of a small number of lymphocytes without aggregation; grade 2, moderate lymphocyte infiltration with multifocal aggregation; and grade 3, extensive lymphocyte infiltration with epithelial erosion [8]. Grades 0 and 1 were categorized as non-moderate-to-severe chronic cholangitis, while grades 2 and 3 were classified as moderate-to-severe chronic cholangitis. Both pathologists were blinded to the clinical and imaging data of all subjects, and any disagreements were resolved through discussion to reach a consensus.

Clinical characteristics

Clinical data of the patients, including sex, age, clinical symptoms, and laboratory examination, were obtained from medical records. Based on preoperative CT images, the imaging features of Todani classification [22], diameter of CBD, biliary stones, peribiliary fluid collection were evaluated. PBM is commonly associated with choledochal cysts, being present in almost all Todani type I cysts (excluding type Ib) and type IVa cysts, while it is rarely observed in other Todani types (Ib, II, III, IVb, or V) [5, 23].

Imaging features were assessed independently by two experienced radiologists blinded to clinical and pathological data (A and B, with 3 and 10 years of experience in abdominal imaging, respectively). For all categorical variables, consensus between the two radiologists must be reached for inclusion in the final analysis. For continuous variable, intraclass correlation coefficient (ICC) was used to evaluate the consistency between the two radiologists. The ICC for the diameter of CBD was 0.911 (95% CI, 0.883–0.944), indicating excellent agreement between the

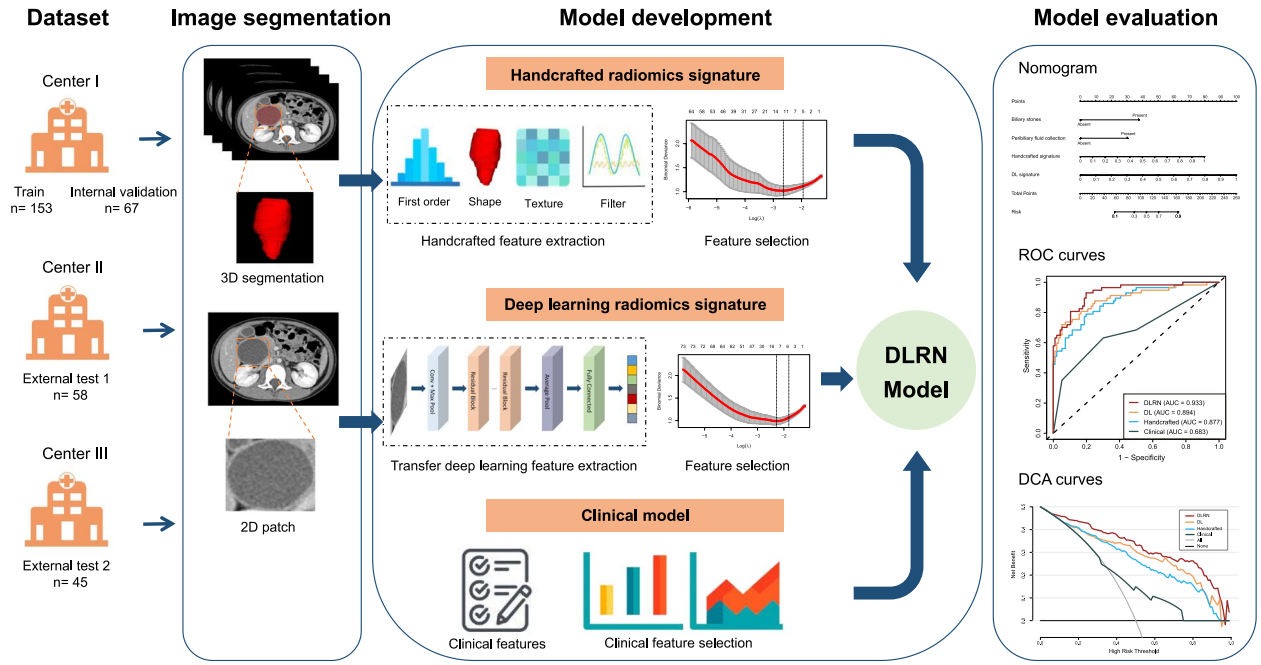

Fig. 1 Workflow of this study. Abbreviations: DCA, decision curve analysis; DLRN, deep learning radiomics nomogram; ROC, receiver operating characteristic

two observers. Therefore, the mean value of the two radiologists' measurements were used for the final analysis.

Image acquisition and standardization

All included patients underwent abdominal CE-CT examination within one month before surgery. Portal venous-phase CT images were retrieved from Picture Archiving and Communication Systems for further analysis. The CT image acquisition settings of the three centers are presented in Supplementary Method S2. In order to reduce data variability among multi-center cohorts and ensure the comparability of DL and traditional radiomics signature, image standardization was undertaken. This involved a two-step process: First, the CT images were preprocessed by setting the window level to 50 Hounsfield Units (HU) and the window width to 400 HU, thereby optimizing image contrast. Subsequently, the images were resampled to a voxel spacing of $1 \times 1 \times 1$ mm^3, ensuring consistent spatial representation across the dataset.

Image segmentation

The extrahepatic CBD is the primary site affected by PBM and the main surgical target for its management. Pathological evaluation of PBM-related chronic cholangitis is also based on the resected CBD tissue. Therefore, three-dimensional regions of interest (ROIs) of the CBD segmentations were manually delineated on axial reconstructed thin-slice portal venous-phase CT images by Radiologist A using ITK-SNAP software (version 3.8.0; http://www.itksnap.org). In patients with PBM presenting with Todani type IVa, segmentation was similarly restricted to the CBD, excluding intrahepatic dilated bile ducts. The segmentation encompassed the entire visible CBD, including both the bile duct wall and bile fluid, across all contiguous slices. This methodology aligns with previous studies in PBM, which similarly analyzed the full structure of the CBD [19, 20].

After one month, 30 cases from the TC were randomly selected, and their ROIs were segmented again by radiologist A and radiologist B to evaluate the intra/inter-observer reproducibility of handcrafted features. Both intra- and inter- ICCs > 0.85 were retained for subsequent analysis.

DL radiomics feature extraction

Handcrafted features of the three-dimensional ROIs were extracted using Pyradiomics Module (http://pyradiomics.readthedocs.io) [24], with gray level discretization performed at a fixed bin width of 25 HU. For each patient, 1834 quantitative features were obtained, including seven categories: (1) 14 shape-based features; (2) 360 first-order features; (3) 440 Gy-level co-occurrence matrix features; (4) 280 Gy-level dependence matrix features; (5) 320 Gy-level run-length matrix features; (6) 320 Gy-level size zone matrix features; (7) 100 neighboring gray-tone difference matrix features.

ResNet50 architecture was adapted to develop the convolutional neural network for DL feature extraction [25]. A total of 2048 DL features were extracted for each patient (Supplementary Method S3).

Feature selection and radiomics signature construction

Before feature selection, all handcrafted and DL radiomics features were normalized using the Z-score method to ensure comparability across features with diverse scales. Then, a four-step procedure involving univariable analysis, Spearman correlation analysis, the least absolute shrinkage and selection operator algorithm, and multivariable logistic regression was used for feature selection and signature construction (Supplementary Method S4).

Clinical model and DLRN construction

In the TC, statistically significant clinical characteristics ($P < 0.05$) were selected using univariate analysis, and subsequently incorporated into the multivariable logistic regression to construct the clinical model. Additionally, a DLRN, which combined the significant clinical characteristics along with the handcrafted and DL signatures from the TC, was developed using multivariable logistic regression. A backward stepwise selection approach was used with the akaike information criterion (AIC) serving as the stopping rule. The variance inflation factor (VIF) was employed to assess multicollinearity among the features. DLRN score for estimating the risk of moderate-to-severe chronic cholangitis in the TC was also calculated and applied to patients in the IVC, ETC1 and ETC2.

Model evaluation

The performances of all generated models were calculated using the receiver operating characteristic (ROC) area under the curve (AUC), and other metrics measuring performances, including sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were also recorded. To compare the performance of the DLRN with other models, the DeLong test was used to evaluate differences in AUC and integrated discrimination improvement (IDI) was implemented to assess the improvement in diagnostic accuracy. The calibration of the DLRN was evaluated using Hosmer–Lemeshow test and the calibration curve, derived through bootstrapping with 1000 resamples. Decision curve analysis (DCA) and clinical impact curve (CIC) were employed to evaluate the clinical value of the DLRN in the total cohorts.

Statistical analysis

All statistical analyses were conducted using R software (version 4.2.1), Python (version 3.9.12), and IBM SPSS software (version 26.0). Statistically significant threshold was set as $P < 0.05$. The detailed statistical analysis is shown in Supplementary Method S5.

Results

Baseline characteristics

The baseline clinical characteristics of all 323 pediatric patients with PBM are presented in Table 1 and Supplementary Table S1. Of all included patients, the median age was 3.0 years (IQR, 1.4–5.7 years) and 231 (71.5%) were female. The rate of moderate-to-severe chronic cholangitis was comparable across the four cohorts, with the rate of 37.3% (57/153) in TC, 37.3% (25/67) in IVC, 34.5% (20/58) in ETC1, and 40.0% (18/45) in ETC2. Additionally, the distribution of clinical characteristics was similar among these four cohorts. In the TC, using univariate analysis, significant differences between non-moderate-to-severe chronic cholangitis and moderate-to-severe chronic cholangitis groups were showed in biliary stones and peribiliary fluid collection ($P < 0.05$; Table 1). Therefore, these two characteristics were selected for subsequent analyses and clinical model building.

Feature selection and radiomics signature development

In this study, two types of radiomics features, including 1834 handcrafted and 2048 DL features were extracted. After feature selection (process shown in Supplementary Result S1 and Fig. S2), 5 handcrafted and 5 DL features were acquired to build the two radiomics signatures, respectively. The final features are shown in Supplementary Table S2. As indicated in Table 2, the handcrafted radiomics signature yielded AUCs of 0.844 (95% CI, 0.750–0.938), 0.857 (95% CI, 0.741–0.972), and 0.823 (95% CI, 0.698–0.948) in the IVC, ETC1, and ETC2 respectively. The DL radiomics signature achieved AUCs of 0.872 (95% CI, 0.768–0.977), 0.854 (95% CI, 0.758–0.950), and 0.840 (95% CI, 0.715–0.964) in the IVC, ETC1, and ETC2, respectively.

Clinical model and DLRN construction

In the TC, two characteristics, biliary stones and peribiliary fluid collection, were selected to construct the clinical model using multivariate logistic regression analysis with the lowest AIC value of 189.857 (Supplementary Table S2). Then, the DLRN was established using these two clinical features, the handcrafted signature, and DL signature as the independent predictors, achieving the lowest AIC value of 106.554 (Fig. 2a and Table 3). The VIFs of the four factors ranged from 1.135 to 1.735, indicating minimal multicollinearity. In addition, the DLRN scores in the moderate-to-severe chronic cholangitis group were significantly higher than those in the non-moderate-to-severe group across all cohorts (all $P < 0.001$, Fig. 2b).

Model evaluation and comparison

The DLRN exhibited good discrimination in the IVC, ETC1 and ETC2, with AUCs of 0.913 (95% CI, 0.834–0.993), 0.916 (95% CI, 0.845–0.987), and 0.895 (95% CI, 0.801–0.989), respectively (Table 2 and Fig. 3). The AUC of the DLRN was significantly higher than that of the clinical model (AUC = 0.681 [95% CI: 0.555–0.807; $P < 0.001$] in the IVC; 0.705 [95% CI: 0.564–0.845; $P = 0.005$] in the ETC1; 0.654 [95% CI: 0.496–0.812; $P = 0.011$] in the ETC2) (Table 2 and Supplementary Table S3). Furthermore, compared with other three single-modality models, the DLRN performed significant improvement in predictive accuracy (IDI > 0, all $P < 0.05$; Supplementary Table S3). The calibration curves and Hosmer–Lemeshow test showed good agreements between the DLRN prediction and actual observation in each cohort (all $P > 0.05$; Fig. 2c). In the whole cohorts, DCA indicated that the DLRN obtained the highest net benefit than other models across all risk threshold (Fig. 4a). A visual inspection of CIC suggested that the DLRN could accurately diagnose moderate-to-severe chronic cholangitis within a relatively wide range of threshold probabilities (Fig. 4b).

Discussion

In this study, we developed and independently validated four diagnostic models: a clinical model that consisted of biliary stones and peribiliary fluid collection, a handcrafted radiomics model, a DL radiomics model, and a DLRN that integrated clinical variables with two types of radiomics signatures. The results demonstrated that the DLRN enhanced the preoperative identification of moderate-to-severe chronic cholangitis in children with PBM, offering high accuracy and robustness.

Severe chronic cholangitis secondary to PBM can lead to preoperative or intraoperative biliary perforation, and even malignant transformation [5, 26, 27]. Early identification of the severity of chronic cholangitis is crucial for the precise and safe management of pancreaticobiliary surgery and for the short- and long-term prognosis of PBM. In a previous study, a handcrafted radiomics score based on T2-weighted MR images showed good performance for the diagnosis of cholangitis in PBM, with an AUC of 0.834 in the validation cohort [19]. In our study for the preoperative identification of moderate-to-severe chronic cholangitis, the handcrafted radiomics model demonstrated comparable performance with AUCs

Table 1 Clinical characteristics of patients with or without moderate-to-severe chronic cholangitis in the four cohorts

Variable	Training cohort		*P* value	Internal validation cohort		*P* value	External test cohort 1		*P* value	External test cohort 2		*P* value
	Non-moderate-to-severe chronic cholangitis (*n* = 96)	Moderate-to-severe chronic cholangitis (*n* = 57)		Non-moderate-to-severe chronic cholangitis (*n* = 42)	Moderate-to-severe chronic cholangitis (*n* = 25)		Non-moderate-to-severe chronic cholangitis (*n* = 38)	Moderate-to-severe chronic cholangitis (*n* = 20)		Non-moderate-to-severe chronic cholangitis (*n* = 27)	Moderate-to-severe chronic cholangitis (*n* = 18)	
Age (y), median (IQR)	2.7 (1.4–5.3)	3.0 (1.3–4.8)	0.896	2.9 (1.5–6.0)	3.4 (2.0–6.0)	0.604	3.0 (2.0–8.5)	2.4 (1.1–8.4)	0.330	4.0 (1.0–8.0)	3.0 (1.0–5.6)	0.422
Female, No. (%)	66 (68.8)	41 (71.9)	0.678	32 (76.2)	19 (76.0)	0.986	24 (60.5)	16 (85.0)	0.188	19 (70.4)	14 (77.8)	0.836
Abdominal pain, No. (%)	64 (66.7)	37 (64.9)	0.825	25 (59.5)	17 (68.0)	0.488	21 (55.3)	7 (35.0)	0.142	15 (55.6)	12 (66.7)	0.456
Jaundice, No. (%)	21 (21.9)	18 (31.6)	0.183	8 (19.1)	7 (28.0)	0.395	9 (23.7)	7 (35.0)	0.359	7 (25.9)	8 (44.4)	0.197
Fever, No. (%)	11 (11.5)	11 (19.3)	0.181	5 (11.9)	4 (16.0)	0.916	7 (18.4)	5 (25.0)	0.805	6 (22.2)	3 (16.7)	0.939
Vomiting, No. (%)	48 (50.0)	26 (45.6)	0.600	22 (52.4)	12 (48.0)	0.729	17 (44.7)	8 (40.0)	0.729	9 (33.3)	8 (44.4)	0.451
Abdominal mass, No. (%)	5 (5.2)	4 (7.0)	0.917	3 (7.1)	3 (12.0)	0.817	2 (5.3)	2 (10.0)	0.895	3 (11.1)	1 (5.6)	0.915
Todani classifcation, No. (%)			0.075			0.359			0.071			0.591
I	58 (60.4)	26 (45.6)		25 (59.5)	12 (48.0)		28 (73.7)	10 (50.0)		20 (74.1)	12 (66.7)	
IVa	38 (39.6)	31 (54.4)		17 (40.5)	13 (52.0)		10 (26.3)	10 (50.0)		7 (25.9)	6 (33.3)	
Diameter of CBD (mm), median (IQR)	19.5 (13.2–32.9)	26.2 (15.9–34.7)	0.097	19.3 (14.1–32.1)	19.8 (10.5–29.2)	0.685	26.6 (16.5–35.7)	37.4 (22.3–43.9)	0.036	24.3 (11.6–36.4)	32.2 (21.9–39.6)	0.121
Biliary stones, No. (%)	29 (30.2)	36 (63.2)	< 0.001	16 (38.1)	16 (64.0)	0.040	11 (28.9)	13 (65.0)	0.008	7 (25.9)	8 (44.4)	0.197
Peribiliary fluid collection, No. (%)	24 (25.0)	23 (40.4)	0.047	11 (26.2)	14 (56.0)	0.015	10 (26.3)	9 (45.0)	0.150	5 (18.5)	8 (44.4)	0.060
Elevated WBC count, No. (%)	32 (33.3)	25 (43.9)	0.193	15 (35.7)	8 (32.0)	0.757	10 (26.3)	10 (50.0)	0.071	6 (22.2)	7 (38.9)	0.227
Elevated AST, No. (%)	28 (29.2)	22 (38.6)	0.229	10 (23.8)	11 (44.0)	0.085	18 (47.4)	12 (60.0)	0.360	12 (44.4)	6 (33.3)	0.456
Elevated ALT, No. (%)	42 (43.8)	27 (47.4)	0.664	23 (54.8)	11 (44.0)	0.394	19 (50.0)	14 (70.0)	0.144	15 (55.6)	8 (44.4)	0.465
Elevated GGT, No. (%)	53 (55.2)	40 (70.2)	0.067	27 (64.3)	17 (68.0)	0.757	25 (65.8)	15 (75.0)	0.471	18 (66.7)	11 (61.1)	0.703
Elevated TBil, No. (%)	24 (25.0)	21 (36.8)	0.120	12 (28.57)	9 (36.00)	0.526	12 (31.6)	8 (40.0)	0.521	8 (29.6)	6 (33.3)	0.793

Table 1 (continued)

Variable	Training cohort		*P* value	Internal validation cohort		*P* value	External test cohort 1		*P* value	External test cohort 2		*P* value
	Non-moderate-to-severe chronic cholangitis (*n* = 96)	Moderate-to-severe chronic cholangitis (*n* = 57)		Non-moderate-to-severe chronic cholangitis (*n* = 42)	Moderate-to-severe chronic cholangitis (*n* = 25)		Non-moderate-to-severe chronic cholangitis (*n* = 38)	Moderate-to-severe chronic cholangitis (*n* = 20)		Non-moderate-to-severe chronic cholangitis (*n* = 27)	Moderate-to-severe chronic cholangitis (*n* = 18)	
Elevated serum amylase, No. (%)	16 (16.7)	13 (22.8)	0.349	10 (23.8)	4 (16.0)	0.447	5 (13.2)	5 (25.0)	0.442	6 (22.2)	5 (27.8)	0.944

Abbreviation: *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *CBD* common bile duct, *GGT* gamma-glutamyl transferase, *IQR* interquartile range, *TBil* total bilirubin, *WBC* white blood cell

Table 2 Performances of clinical model, handcrafted signature, DL signature, and DLRN in the four cohorts

Model	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Training cohort						
Clinical model	0.683 (0.594–0.772)	0.673	0.632	0.698	0.554	0.761
Handcrafted signature	0.877 (0.821–0.933)	0.797	0.789	0.802	0.703	0.865
DL signature	0.894 (0.837–0.951)	0.856	0.702	0.948	0.889	0.843
DLRN	0.933 (0.893–0.973)	0.850	0.930	0.802	0.736	0.951
Internal validation cohort						
Clinical model	0.681 (0.555–0.807)	0.627	0.640	0.619	0.500	0.743
Handcrafted signature	0.844 (0.750–0.938)	0.776	0.840	0.738	0.656	0.886
DL signature	0.872 (0.768–0.977)	0.866	0.760	0.929	0.864	0.867
DLRN	0.913 (0.834–0.993)	0.866	0.920	0.833	0.767	0.946
External test cohort 1						
Clinical model	0.705 (0.564–0.845)	0.690	0.650	0.711	0.542	0.794
Handcrafted signature	0.857 (0.741–0.972)	0.828	0.600	0.947	0.857	0.818
DL signature	0.854 (0.758–0.950)	0.776	0.750	0.789	0.652	0.857
DLRN	0.916 (0.845–0.987)	0.862	0.800	0.895	0.800	0.895
External test cohort 2						
Clinical model	0.654 (0.496–0.812)	0.622	0.667	0.593	0.522	0.727
Handcrafted signature	0.823 (0.698–0.948)	0.756	0.833	0.704	0.652	0.864
DL signature	0.840 (0.715–0.964)	0.800	0.722	0.852	0.765	0.821
DLRN	0.895 (0.801–0.989)	0.822	0.833	0.815	0.750	0.880

Abbreviation: *AUC* area under the receiver operating characteristic curve, *CI* confidence interval, *DL* deep learning, *DLRN* deep learning radiomics nomogram, *NPV* negative predictive value, *PPV* positive predictive value

ranging from 0.823 to 0.857, and our model was validated in two independent centers. Notably, all five selected handcrafted radiomics features were higher-order, including the exponential, gradient, laplacian of gaussian and wavelet features. This is consistent with prior studies that higher-order features can reveal the deep attributes of images, capturing more detailed insights into lesion heterogeneity [28, 29]. These features achieved superiority not only in neoplastic disease [28, 30], but also in non-neoplastic diseases such as inflammatory bowel disease, pancreatitis, and liver fibrosis [31–33]. Similarly, in our study, the higher-order handcrafted radiomics features proved effective in identifying imaging heterogeneity associated with moderate-to-severe chronic cholangitis in PBM, demonstrating their potential to enhance diagnostic performance in non-neoplastic diseases.

However, the reliance on predefined feature extraction in handcrafted radiomics inherently limits its ability to fully capture the complexity of medical images. In contrast, DL has the capacity to discern complex patterns and capture intricate details from the hidden layers of neural networks without relying on predetermined features, which complements the existing practices of handcrafted radiomics and opens new avenues for nuanced medical analysis [28, 34]. The application of DL has extended to various medical domains, encompassing not only diagnosis but also the prediction of treatment responses and the assessment of disease outcomes [35–38]. In the current study, ResNet50 architecture was used in the current study to extract DL radiomics features and the resulting DL signature demonstrated noteworthy performance in the diagnosis of moderate-to-severe cholangitis. Our developed DL signature yielded AUCs ranging from 0.840 to 0.872, outperforming the clinical model in all cohorts and the handcrafted signature in most cohorts. Based on the high-capacity and multi-layered network structure, DL algorithm can extract nonlinear and abstract high-dimensional data. A prior study combined UNet++and ResNeSt18 to classify chronic cholangitis severity in PBM using CE-CT images, achieving a Dice coefficient of 0.839 ± 0.150 for CBD segmentation and an AUC of 0.711 for classification [20]. While both studies employed DL methods, the observed differences in performance may be attributed to errors associated with automated segmentation in the prior study and its reliance on an end-to-end output method, in contrast to the current study's approach of DL feature extraction and rigorous feature selection. Additionally, the prior study included a smaller patient cohort (76 vs. 323), which likely impacted the model's accuracy and robustness. By leveraging feature-level DL extraction and selection, this approach enhanced the ability to identify critical features,

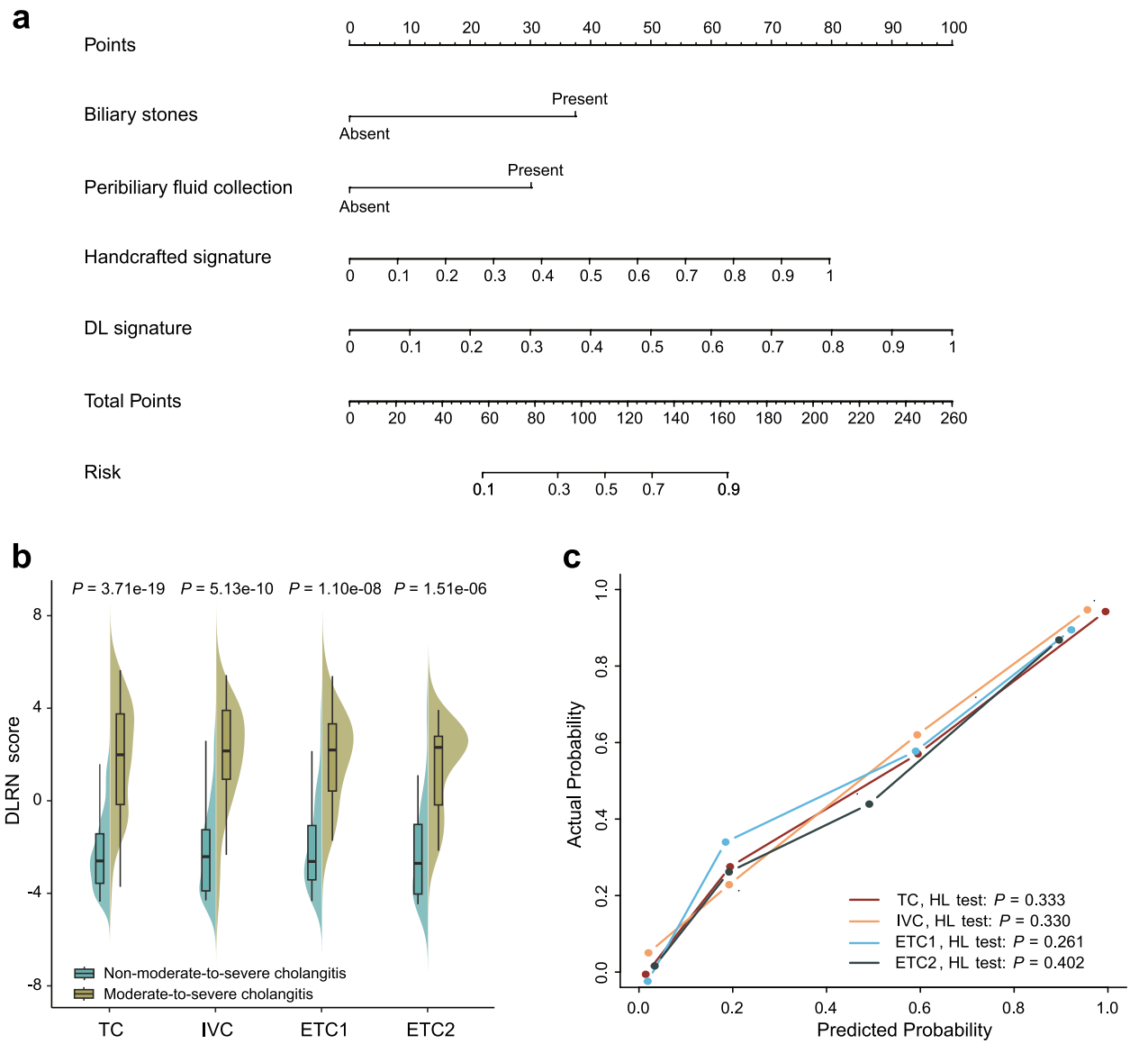

Fig. 2 Deep learning radiomics nomogram (DLRN) and its performance. **a** DLRN for identifying moderate-to-severe chronic cholangitis developed based on the training dataset. **b** Violin plots of the differences in DLRN score between non-moderate-to-severe cholangitis and moderate-to-severe cholangitis groups in all cohorts. **c** Calibration curves of DLRN in the four cohorts. Abbreviations: ETC1, external test cohort 1; ETC2, external test cohort 2; HL test, Hosmer–Lemeshow test; IVC, internal validation cohort; TC, training cohort

Table 3 Multivariable logistic regression analysis of the DLRN

Intercept and variable	β	OR (95% CI)	*P* value	AIC
Intercept	− 4.584	-	< 0.001	106.554
Biliary stones (present)	1.559	4.755 (1.666–15.192)	0.005	
Peribiliary fluid collection (present)	1.253	3.501 (1.125–11.762)	0.034	
Handcrafted signature (per 0.1 increase)	3.313	27.473 (2.124–433.374)	0.014	
DL signature (per 0.1 increase)	4.160	64.089 (7.121–769.009)	< 0.001	

DLRN score = −4.584 + (1.559 × Biliary stones) + (1.253 × Peribiliary fluid collection) + (3.313 × Handcrafted signature) + (4.160 × DL signature)

Abbreviation: *AIC* Akaike information criterion, *CI* confidence interval, *DL* deep learning, *DLRN* deep learning radiomics nomogram, *OR* odds ratio

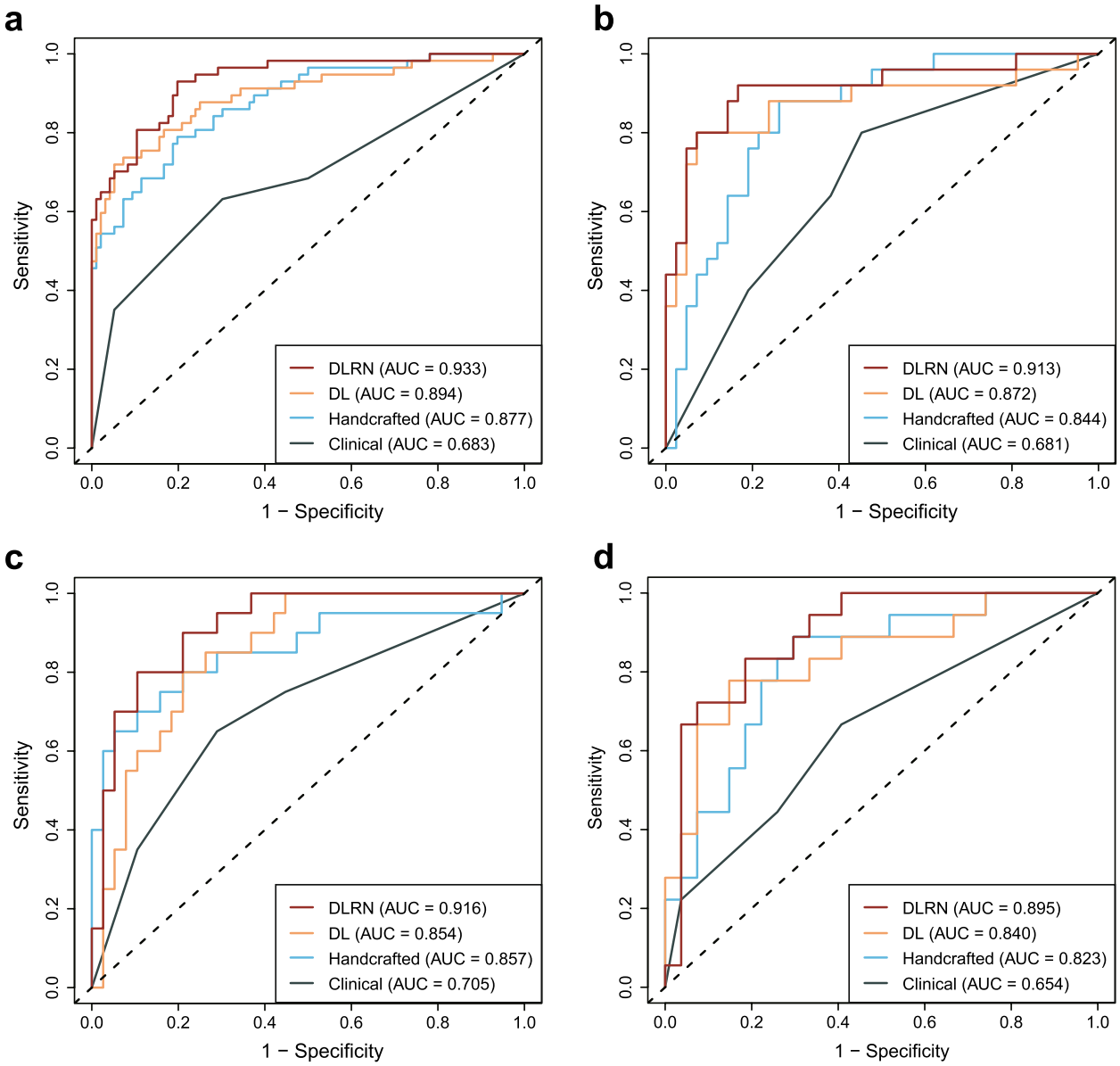

Fig. 3 Receiver operating characteristic (ROC) curves of the four models. Graphs show ROC curves of the deep learning radiomics nomogram (DLRN), deep learning (DL) signature, handcrafted signature, and clinical model in the training cohort (**a**), internal validation cohort (**b**), external test cohort 1 (**c**), and external test cohort 2 (**d**). Abbreviations: AUC, area under the receiver operating characteristic curve

minimize redundancy, and enable the multi-scale integration of features from various sources.

Furthermore, the influence of clinical factors on cholangitis was considered. A combined DLRN was developed and validated, incorporating biliary stones, peribiliary fluid collection, as well as handcrafted and DL radiomics signatures. Biliary stones and peribiliary fluid collection have been shown to be associated with cholangitis [39–41]. This is consistent with our study, thereby justifying their incorporation within the diagnostic model. The DLRN presented good calibration and excellent diagnostic performance with AUCs ranging from 0.895 to 0.916, which was higher than the other three models alone across all cohorts. This finding aligns with previous research, highlighting that multi-scale information fusion can significantly enhance the predictive performance of models [18, 28]. For instance, Yin et al. [42] developed a multimodal model that integrated handcrafted and deep learning-based radiomics features from CT images along with key clinical characteristics to predict severe acute

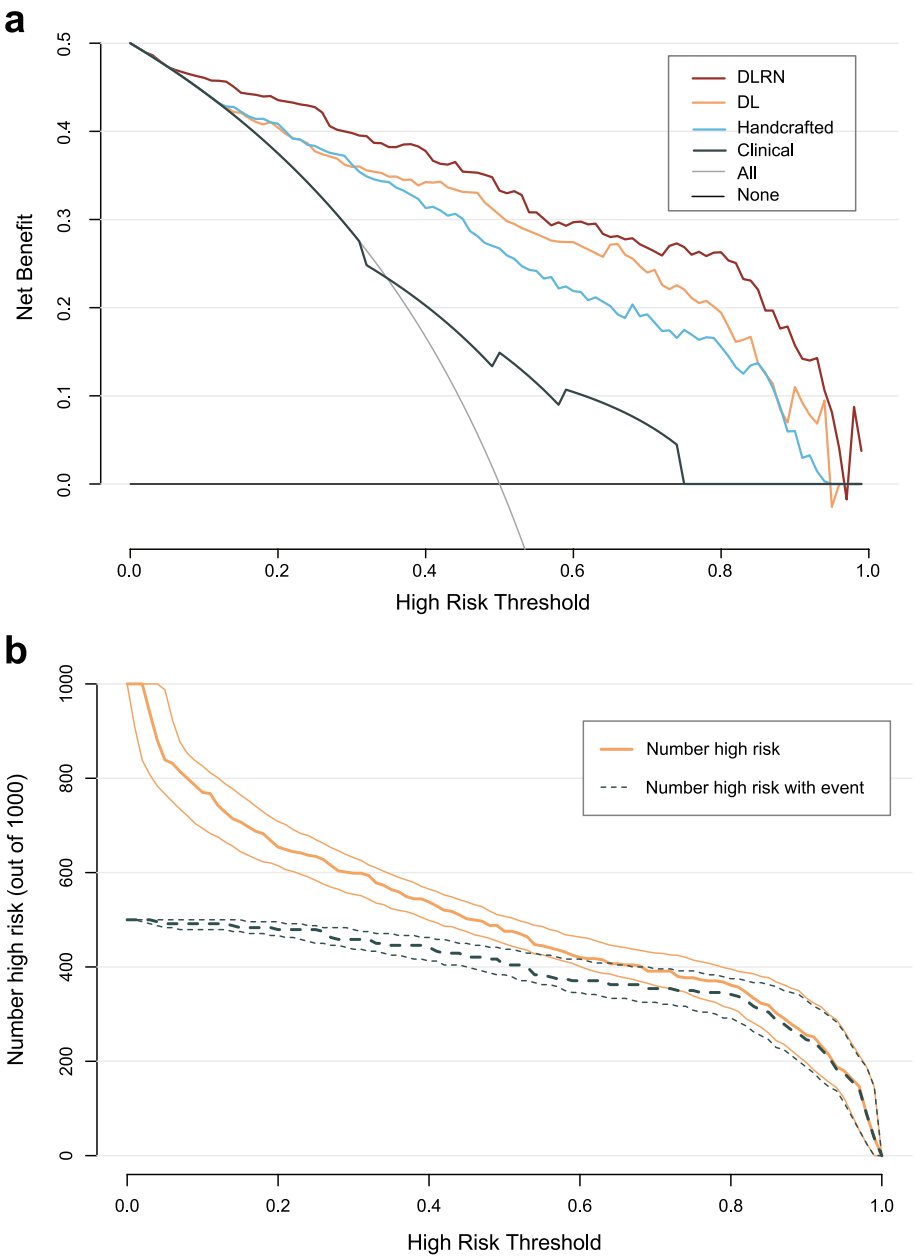

Fig. 4 Decision Curve Analysis (DCA) and Clinical Impact Curve (CIC). **a** DCA of deep learning radiomics nomogram (DLRN), deep learning (DL) signature, handcrafted signature, and clinical model. **b** CIC of DLRN in the whole cohorts

pancreatitis. Their results demonstrated that the multimodal model outperformed single-modality models and traditional scoring systems, achieving AUCs of 0.874 and 0.916 in internal and external test sets, respectively. Similarly, another study employed a feature integration approach and showed that the combined nomogram model exhibited strong discriminatory ability for IgA nephropathy, achieving an ROC of 0.884 in the testing cohort [43]. The consistently superior performance of combined models across studies indicated the feasibility and importance of integrating radiomics with clinical features to improve diagnostic accuracy in inflammatory diseases. In addition, the minimal AIC and improved IDI in this study suggested that the improved discriminatory ability of the DLRN is the result of effective feature amalgamation rather than model overfitting. The promising performance underscored the potential of the DLRN in enhancing individualized preoperative risk

assessments for moderate-to-severe chronic cholangitis, thereby assisting clinicians in selecting appropriate treatment strategies for pediatric patients with PBM to minimize intraoperative injuries. Furthermore, these patients should be closely monitored and actively followed up to prevent serious perioperative and postoperative complications.

This study has some limitations. First, this was a retrospective study, which was limited by inherent bias. Therefore, further prospective studies are required to validate the robustness and the clinical utility. Second, the differences in CT systems and scan parameters across different centers could influence the radiomics results. However, these differences were partially addressed through image standardization. Importantly, these disparities also reflect the diversity of real-world clinical practice, where similar inconsistencies are often encountered. Such variations validate the reproducibility and generalizability of multi-center findings, highlighting the robustness of the DLRN and its potential for broader clinical implementation. Third, limitations in imaging resolution and segmentation techniques prevented the bile duct wall from being delineated as an independent ROI. Consequently, the inclusion of bile fluid within the ROI may have introduced confounding factors, potentially affecting the model's reliability. We will pay attention to advancements in fine structure segmentation techniques and continue to explore improved strategies for separate feature extraction. Finally, our study primarily employed portal venous-phase CT images to extract radiomics features. Future research is needed to determine whether integrating multimodal images could further enhance model performance.

Conclusions

In conclusion, the proposed DLRN, incorporating clinical characteristics with both handcrafted and DL radiomics signatures, exhibited excellent performance in identifying moderate-to-severe chronic cholangitis. It can be a useful and non-invasive tool to preoperatively assess the severity of choledochal inflammation, thereby facilitating individualized treatment and management strategies in children with PBM.

Abbreviations

AIC	Akaike information criterion
AUC	Area under the curve
CBD	Common bile duct
CE-CT	Contrast-enhanced computed tomography
CIC	Clinical impact curve
DCA	Decision curve analysis
DL	Deep learning
DLRN	Deep learning radiomics nomogram
ETC	External test cohort
ICC	Intraclass correlation coefficient
IDI	Integrated discrimination improvement
IQR	Interquartile range
IVC	Internal validation cohort
MRI	Magnetic resonance imaging
NPV	Negative predictive value
PBM	Pancreaticobiliary maljunction
PPV	Positive predictive value
ROC	Receiver operating characteristic
ROI	Region of interest
TC	Training cohort
VIF	Variance inflation factor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01579-3.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements

We would like to thank our colleagues in the department of pathology, Children's Hospital of Soochow University for helping with histopathological analysis. We also sincerely thank Platform Onekey AI and their developers' help in this research work. The code can be accessed from the following link: https://gitee.com/wangqingbaidu/OnekeyCompo/tree/master.

Authors' contributions

Conceiving the study and design: W.G. and S. S. Acquisition, analysis, or interpretation of data: all authors. Drafting of the manuscript: H.M. Statistical analysis: H.M, K.C., and B.Z. Critical revision of the manuscript for important intellectual content: W.G., S. S., K.C., B.Z., and S.H. All authors read and approved the final manuscript.

Funding

This study was supported by National Natural Science Foundation of China (No.81971685), Scientific Research Project of Jiangsu Provincial Health Commission (No.ZD2022015), Science and Technology Development Project of Suzhou (SKY2022054), Suzhou Clinical Key Disease Diagnosis and Treatment Technology Special Project (No.LCZX202311), Soochow University Medical + X project (No.ML13101823), and "Science, Education, and Health Strengthening" Key Project of Suzhou (No. ZDXM2024007).

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to ethical restrictions but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Boards of each participating center, which waived the informed consent requirement because of the retrospective nature of the study and minimal risk to participants. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Radiology, Children's Hospital of Soochow University, Suzhou 215025, China. [2]Department of Ultrasound, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021,

China. [3]Department of Interventional Therapy, Xuzhou Children's Hospital, Xuzhou 221000, China. [4]Department of Radiology, The 8th Hospital of Xi'an, Xi'an 710000, China.

Received: 7 November 2024 Accepted: 2 February 2025
Published online: 05 February 2025

References

1. Kamisawa T, Takuma K, Anjiki H, Egawa N, Kurata M, Honda G, et al. Pancreaticobiliary maljunction. Clin Gastroenterol Hepatol. 2009;7:S84–8.
2. Ragot E, Mabrut J-Y, Ouaïssi M, Sauvanet A, Dokmak S, Nuzzo G, et al. Pancreaticobiliary Maljunctions in European Patients with Bile Duct Cysts: results of the multicenter study of the French Surgical Association (AFC). World J Surg. 2017;41:538–45.
3. Kamisawa T, Kaneko K, Itoi T, Ando H. Pancreaticobiliary maljunction and congenital biliary dilatation. Lancet Gastroenterol Hepatol. 2017;2:610–8.
4. Kamisawa T, Kuruma S, Chiba K, Tabata T, Koizumi S, Kikuyama M. Biliary carcinogenesis in pancreaticobiliary maljunction. J Gastroenterol. 2017;52:158–63.
5. Kamisawa T, Ando H, Suyama M, Shimada M, Morine Y, Shimada H, et al. Japanese clinical practice guidelines for pancreaticobiliary maljunction. J Gastroenterol. 2012;47:731–59.
6. Guo W-L, Zhan Y, Fang F, Huang S-G, Deng Y-B, Zhao J-G, et al. Factors affecting the operating time for complete cyst excision and Roux-en-Y hepaticojejunostomy in paediatric cases of congenital choledochal malformation: a retrospective case study in Southeast China. BMJ Open. 2018;8:e022162.
7. Qiao G, Li L, Li S, Tang S, Wang B, Xi H, et al. Laparoscopic cyst excision and Roux-Y hepaticojejunostomy for children with choledochal cysts in China: a multicenter study. Surg Endosc. 2015;29:140–4.
8. Kim JH, Choi TY, Han JH, Yoo BM, Kim JH, Hong J, et al. Risk factors of postoperative anastomotic stricture after excision of choledochal cysts with hepaticojejunostomy. J Gastrointest Surg. 2008;12:822–8.
9. Ono A, Arizono S, Isoda H, Togashi K. Imaging of pancreaticobiliary maljunction. Radiographics. 2020;40:378–92.
10. Okada T, Sasaki F, Honda S, Naitou S, Onodera Y, Todo S. Usefulness of axial planes of helical computed tomography for diagnosis of pancreaticobiliary maljunction in early infants with negative findings on magnetic resonance cholangiopancreatography. J Pediatr Surg. 2008;43:579–82.
11. Ono S, Fumino S, Iwai N. Diagnosis and treatment of pancreaticobiliary maljunction in children. Surg Today. 2011;41:601–5.
12. Takada S, Uchida H, Hinoki A, Shirota C, Sumida W, Tainaka T, et al. Variations of the hepatic artery and bile duct in patients with pancreaticobiliary maljunction: impact on postoperative outcomes. J Hepatobiliary Pancreat Sci. 2023;30:1241–8.
13. Chen W, Zhao L, Wang J, Guo W. Hepatic vascular variations and visual three-dimensional reconstruction technique in pediatric patients with choledochal cyst. Surg Radiol Anat. 2020;42:1489–99.
14. Braman N. Radiomics for surgical planning and prognostication. JAMA Netw Open. 2020;3:e2028608.
15. Magnuska ZA, Roy R, Palmowski M, Kohlen M, Winkler BS, Pfeil T, et al. Combining radiomics and autoencoders to distinguish benign and malignant breast tumors on US images. Radiology. 2024;312:e232554.
16. Xie C, Yu X, Tan N, Zhang J, Su W, Ni W, et al. Combined deep learning and radiomics in pretreatment radiation esophagitis prediction for patients with esophageal cancer underwent volumetric modulated arc therapy. Radiother Oncol. 2024;199:110438.
17. Li M, Fan Y, You H, Li C, Luo M, Zhou J, et al. Dual-energy CT deep learning radiomics to predict macrotrabecular-massive hepatocellular carcinoma. Radiology. 2023;308:e230255.
18. Guo L, Hao X, Chen L, Qian Y, Wang C, Liu X, et al. Early warning of hepatocellular carcinoma in cirrhotic patients by three-phase CT-based deep learning radiomics model: a retrospective, multicentre, cohort study. EClinicalMedicine. 2024;74:102718.
19. Yang Y, Zhang X-X, Zhao L, Wang J, Guo W-L. Development of a simplified model and nomogram in preoperative diagnosis of pediatric chronic cholangitis with pancreaticobiliary maljunction using clinical variables and MRI radiomics. Insights Imaging. 2023;14:41.
20. Guo W-L, Geng A-K, Geng C, Wang J, Dai Y-K. Combination of UNet++ and ResNeSt to classify chronic inflammation of the choledochal cystic wall in patients with pancreaticobiliary maljunction. Br J Radiol. 2022;95:20201189.
21. Kamisawa T, Ando H, Hamada Y, Fujii H, Koshinaga T, Urushihara N, et al. Diagnostic criteria for pancreaticobiliary maljunction 2013. J Hepatobiliary Pancreat Sci. 2014;21:159–61.
22. Todani T, Watanabe Y, Narusue M, Tabuchi K, Okajima K. Congenital bile duct cysts: classification, operative procedures, and review of thirty-seven cases including cancer arising from choledochal cyst. Am J Surg. 1977;134:263–9.
23. Kamisawa T, Ando H, Shimada M, Hamada Y, Itoi T, Takayashiki T, et al. Recent advances and problems in the management of pancreaticobiliary maljunction: feedback from the guidelines committee. J Hepatobiliary Pancreat Sci. 2014;21:87–92.
24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77:e104–7.
25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;770–8.
26. Kamisawa T, Kuruma S, Tabata T, Chiba K, Iwasaki S, Koizumi S, et al. Pancreaticobiliary maljunction and biliary cancer. J Gastroenterol. 2015;50:273–9.
27. Fukuzawa H, Urushihara N, Miyakoshi C, Kajihara K, Kawahara I, Isono K, et al. Clinical features and risk factors of bile duct perforation associated with pediatric congenital biliary dilatation. Pediatr Surg Int. 2018;34:1079–86.
28. Cui Y, Zhang J, Li Z, Wei K, Lei Y, Ren J, et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: a multicenter cohort study. EClinicalMedicine. 2022;46:101348.
29. Ma A, Yan X, Qu Y, Wen H, Zou X, Liu X, et al. Amide proton transfer weighted and diffusion weighted imaging based radiomics classification algorithm for predicting 1p/19q co-deletion status in low grade gliomas. BMC Med Imaging. 2024;24:85.
30. Wang M, Perucho JAU, Hu Y, Choi MH, Han L, Wong EMF, et al. Computed tomographic radiomics in differentiating histologic subtypes of epithelial ovarian carcinoma. JAMA Netw Open. 2022;5:e2245141.
31. Zhao Y, Wei J, Xiao B, Wang L, Jiang X, Zhu Y, et al. Early prediction of acute pancreatitis severity based on changes in pancreatic and peripancreatic computed tomography radiomics nomogram. Quant Imaging Med Surg. 2023;13:1927–36.
32. Yang Y, Zhang X, Zhao L, Mao H, Cai T, Guo W. Development of an MRI-based radiomics-clinical model to diagnose liver fibrosis secondary to pancreaticobiliary maljunction in children. J Magn Reson Imaging. 2023;58:605–17.
33. Zhu C, Yu Y, Wang S, Wang X, Gao Y, Li C, et al. A novel clinical radiomics nomogram to identify Crohn's disease from intestinal tuberculosis. J Inflamm Res. 2021;14:6511–21.
34. Ma X, Xia L, Chen J, Wan W, Zhou W. Development and validation of a deep learning signature for predicting lymph node metastasis in lung adenocarcinoma: comparison with radiomics signature and clinical-semantic model. Eur Radiol. 2022;33:1949–62.
35. Huang Y, Zhu T, Zhang X, Li W, Zheng X, Cheng M, et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: a multicenter, retrospective study. eClinicalMedicine. 2023;58:101899.
36. Ahmad T, Guida A, Stewart S, Barrett N, Jiang X, Vincer M, et al. Can deep learning classify cerebral ultrasound images for the detection of brain injury in very preterm infants? Eur Radiol. 2024. https://doi.org/10.1007/s00330-024-11028-4.
37. Jiang X, Zhao H, Saldanha OL, Nebelung S, Kuhl C, Amygdalos I, et al. An MRI deep learning model predicts outcome in rectal cancer. Radiology. 2023;307:e222223.
38. Zhu J, Zou L, Xie X, Xu R, Tian Y, Zhang B. 2.5D deep learning based on multi-parameter MRI to differentiate primary lung cancer pathological subtypes in patients with brain metastases. Eur J Radiol. 2024;180:111712.
39. Pötter-Lang S, Ba-Ssalamah A, Bastati N, Messner A, Kristic A, Ambros R, et al. Modern imaging of cholangitis. Br J Radiol. 2021;94:20210417.

40. Mohammad Alizadeh AH. Cholangitis: diagnosis, treatment and prognosis. J Clin Transl Hepatol. 2017;5:404–13.
41. Seo N, Kim SY, Lee SS, Byun JH, Kim JH, Kim HJ, et al. Sclerosing cholangitis: clinicopathologic features, imaging spectrum, and systemic approach to differential diagnosis. Korean J Radiol. 2016;17:25–38.
42. Yin M, Lin J, Wang Y, Liu Y, Zhang R, Duan W, et al. Development and validation of a multimodal model in predicting severe acute pancreatitis based on radiomics and deep learning. Int J Med Inf. 2024;184:105341.
43. Qin X, Xia L, Ma Q, Cheng D, Zhang C. Development of a novel combined nomogram model integrating deep learning radiomics to diagnose IgA nephropathy clinically. Ren Fail. 2023;45:2271104.

Publisher's Note