Cancer Biol Med 2026. doi: 10.20892/j.issn.2095-3941.2026.0566

ORIGINAL ARTICLE

OPEN ACCESS

An integrated CT/MRI diagnostic system for pancreatic cystic lesions that supports pathological classification and risk stratification

Xiaonan Cui[1,2*], Houpu Liu[1,2*], Sunyi Zheng[1,2], Hongwei Wang[2,3], Jiaxin Liu[1,2], Ziyi Gao[4], Tianye Liu[1], Jing Liang[5], Xiaoxu Wang[6], Xinru Zheng[1], Jing Wang[7], Keyi Bian[8], Xiaomeng Yang[1,2], Chunlin Zhang[1,2], Tianshuo Jiao[1,2], Yufei Yang[1,2], Shuxuan Fan[1], Xiuchao Wang[3], Song Gao[3], Chuntao Gao[3], Yukuan Feng[3], Hui Li[9], Guoping Xu[10], Xiang Zhang[11], Wen Shen[12], Haoran Sun[13], Baocun Sun[14], Yan Sun[14], Jianyu Xiao[1,2], Yalin Zheng[15], Michael Ying[16], Geertruida H. de Bock[17], Zhaoxiang Ye[1,2], Jihui Hao[3]

[1]Department of Radiology, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, State Key Laboratory of Druggability Evaluation and Systematic Translational Medicine, Tianjin Key Laboratory of Digestive Cancer, Tianjin Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China; [2]Medical Artificial General Intelligence for Computation (MAGIC) Lab, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, State Key Laboratory of Druggability Evaluation and Systematic Translational Medicine, Tianjin Key Laboratory of Digestive Cancer, Tianjin Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China; [3]Pancreas Center, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, State Key Laboratory of Druggability Evaluation and Systematic Translational Medicine, Tianjin Key Laboratory of Digestive Cancer, Tianjin Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China; [4]Queen Mary School Hainan, Beijing University of Posts and Telecommunications, Lingshui 570203, China; [5]Department of Radiology, First Affiliated Hospital of Hainan Medical University, Haikou 570102, China; [6]Center for Health Statistics and Information, National Health Commission of the PRC, Beijing 100044, China; [7]School of Public Health, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China; [8]Department of Medical Imaging, Jinan Third People's Hospital, Jinan 250101, China; [9]Department of Gastrointestinal Cancer Biology, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, State Key Laboratory of Druggability Evaluation and Systematic Translational Medicine, Tianjin Key Laboratory of Digestive Cancer, Tianjin Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China; [10]Department of Radiology, Second Hospital of Tianjin Medical University, Tianjin 300211, China; [11]Department of Radiology, Tianjin NanKai Hospital, Tianjin Medical University, Tianjin 300100, China; [12]Department of Radiology, Tianjin First Central Hospital, Tianjin 300192, China; [13]Department of Radiology, Tianjin Medical University General Hospital, Tianjin 300052, China; [14]Department of Pathology, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, State Key Laboratory of Druggability Evaluation and Systematic Translational Medicine, Tianjin Key Laboratory of Digestive Cancer, Tianjin Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China; [15]Department of Eye and Vision Science, University of Liverpool, United Kingdom, Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool L7 8TX, United Kingdom; [16]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon 999077, Hong Kong SAR, China; [17]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 9713 GZ, The Netherlands

ABSTRACT

Objective: To develop a pathology-based preoperative decision-support system (Structured Imaging Report System for Pancreatic Cystic Lesions, SIRS-PCL) for subtype classification and risk stratification of PCLs.

Methods: We conducted a multicentre retrospective study involving 1,102 patients treated between 1 January 2011 and 31 December 2024. The study cohort was divided into a development dataset (n = 858) and a temporal validation dataset (n = 244) from the periods 2011–2022 and 2023–2024, respectively. Three logistic prediction models were used to classify the five PCL subtypes and to perform separate three-tier risk stratifications (benign/low, medium, and high) for intraductal papillary mucinous neoplasm (IPMN) and mucinous cystic neoplasm (MCN) cases. SIRS-PCL was prospectively implemented at Tianjin Medical University Cancer Institute and Hospital from 25 June 2025 to 25 June 2026, and the diagnostic accuracy, sensitivity, and specificity were evaluated in a cohort of 76 patients.

Results: A total of 1,102 patients with PCLs were included in the study (median age 56 years, interquartile range 44–64), of whom 714 were female (64.8%). The five-category model included in SIRS-PCL achieved a mean area under the receiver operating characteristic curve (AUC) of 0.93 [95% confidence interval (CI) 0.91–0.95] in temporal validation. The IPMN model showed good discrimination for risk stratification (mean AUC 0.86, 95% CI 0.79–0.93), with the highest overall accuracy of 77.1% exceeding current guidelines (range 50.0%–58.6%). The MCN risk model also showed high performance (mean AUC 0.97, 95% CI 0.91–1.00). In comparison to current practice, SIRS-PCL use reduced unnecessary surgeries 11.5% *vs.* 23.1% and over-examinations 6.7% *vs* 33.7%. In the prospective implementation cohort, SIRS-PCL achieved accuracy of 86.2% for five-category pathological classification, 73.7% for IPMN risk stratification, and 100% for MCN risk stratification.

Conclusions: SIRS-PCL is a pathology-based, quantitative decision-support system for preoperative pathological subtype classification and pathological risk stratification of surgically evaluated neoplastic PCLs.

KEYWORDS Pancreatic cystic lesions; machine learning; abdominal imaging, noninvasive; cyst-related malignancy

Introduction

Pancreatic cystic lesions (PCLs) represent a growing diagnostic and therapeutic challenge in clinical practice. Once considered rare, these lesions are now common incidental findings owing to the widespread use of high-resolution cross-sectional imaging[1-4]. The increase in detection has revealed a broad spectrum of PCLs, including neoplastic PCLs such as serous cystic neoplasm (SCN), intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasm (MCN), and cystic or cystic-solid presentations of solid pseudopapillary neoplasm (SPN) and pancreatic neuroendocrine tumor (NET), all of which require accurate differential diagnosis because of their distinct biological behaviors and management strategies[3,5]. In this study we specifically focus on neoplastic PCLs with pathological confirmation. Non-neoplastic cystic lesions, particularly pancreatic pseudocysts, were intentionally excluded because they lack neoplastic potential and were outside the scope of pathological subtype prediction and pathological risk stratification.

Current preoperative subtype classification relies primarily on invasive diagnostic procedures, including endoscopic ultrasound and cyst-fluid cytology. However, these methods have limited diagnostic accuracy, with misclassification rates reaching up to 30% in surgically resected specimens[6]. The inherent diagnostic uncertainty is further compounded by the high PCL detection rate in the general population, which has raised concerns regarding potential overdiagnosis and overtreatment, particularly for IPMN and MCN[2]. As resection represents the only curative option for cystic neoplasia, it is inevitably associated with substantial perioperative risks, including morbidity rates of 20%–45% and mortality rates of 2%–3%[7,8]. Therefore, accurate pathological subtype prediction and pathological risk stratification before surgery are essential to support individualized surgical decision-making and avoid unnecessary surgical intervention[2,9,10].

Current international guidelines, including those from the International Association of Pancreatology (Kyoto), the American College of Gastroenterology (ACG), the European Study Group on Cystic Tumours of the Pancreas, and the Hong Kong consensus recommendations, have highlighted several high-risk stigmata and worrisome imaging features, with a predominant focus on IPMN, to guide preoperative risk assessment and surgical decision-making[7,8,11,12]. However, there is substantial heterogeneity among the current guidelines with respect to diagnostic criteria and indications for surgical resection[13]. Consequently, clinical application of these guidelines has resulted in significant practice variability: nearly half of patients undergoing pancreatic resection for presumed IPMN may receive unnecessary or premature surgery, and approximately one-quarter of PCLs are misclassified preoperatively[6]. These findings underscore the need for a standardized imaging-based framework to improve preoperative pathological subtype prediction and pathological risk assessment for surgically evaluated neoplastic PCLs.

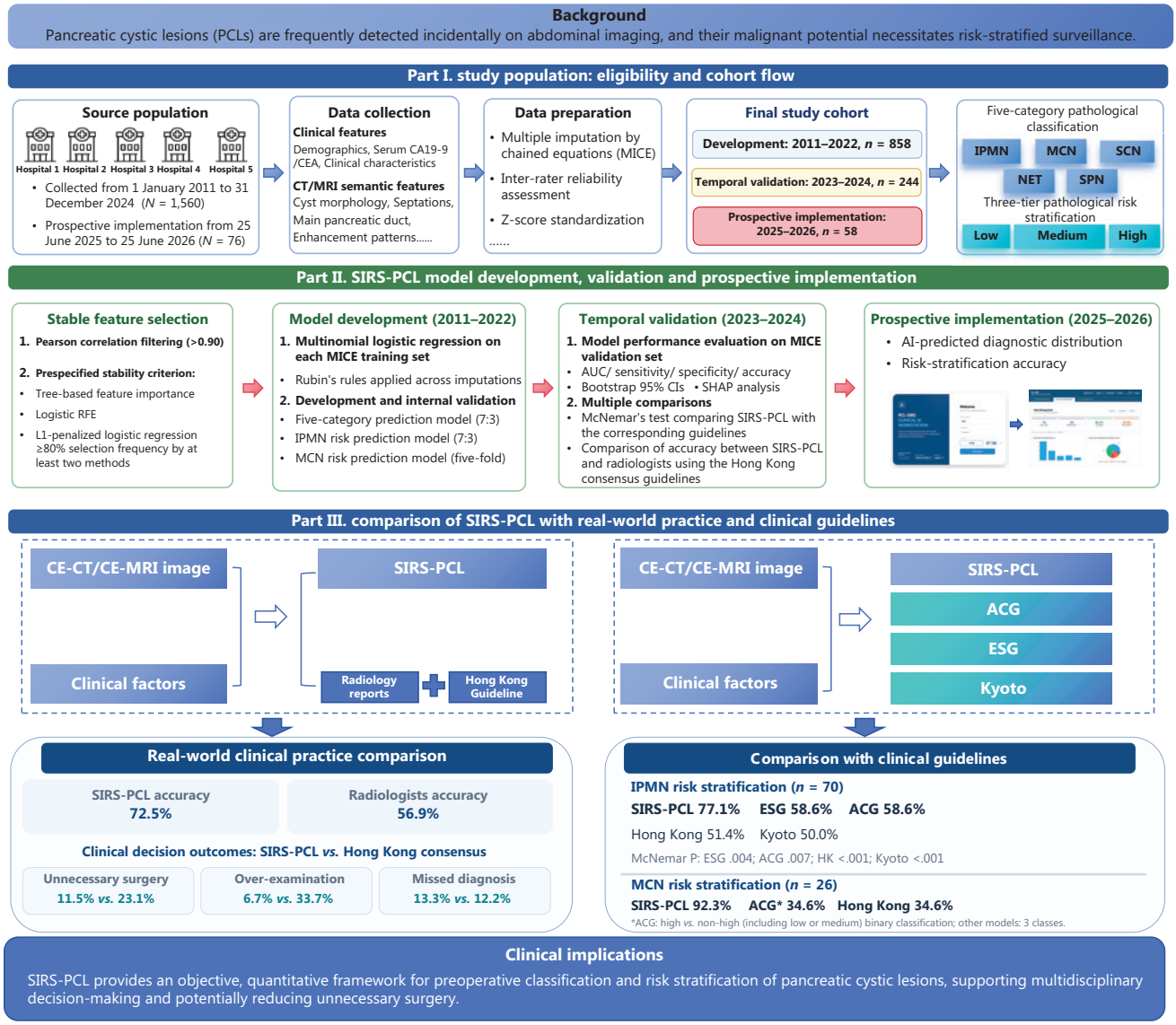
The primary objective of this study was to develop and validate a pathology-based structured imaging reporting system for preoperative pathological subtype classification and pathological risk stratification of surgically evaluated neoplastic PCLs (Structured Imaging Report System for Pancreatic Cystic Lesions, SIRS-PCL). The secondary objective was to compare the diagnostic performance and clinical utility of SIRS-PCL with imaging assessments based on established guideline criteria using a common pathology-defined reference standard.

Materials and methods

Study design and patients

This retrospective multicentre study was conducted using the Tianjin Medical Imaging and Pathology Data Lake for

*These authors contributed equally to this work.
Correspondence to: Jihui Hao
E-mail: haojihui@tjmuch.com
ORCID ID: https://orcid.org/0000-0002-1607-1730
Received July 15, 2026; accepted September 7, 2026.
Available at www.cancerbiomed.org
©2026 The Authors. Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)



Study flow Part I. Establishment of multicenter retrospective and prospective cohorts for data collection, including 76 patients prospectively enrolled, of whom 58 had definitive pathological diagnoses within the five target categories and were included in the prospective evaluation of pathological classification; Part II. Development, validation, and prospective implementation of SIRS-PCL for pancreatic cystic lesion classification and risk stratification; and Part III. Clinical evaluation through comparisons with real-world clinical practice and established guidelines to assess its potential clinical utility. SIRS-PCL, structured imaging report system for pancreatic cystic lesions; PCL, pancreatic cystic lesion; IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; SCN, serous cystic neoplasm; SPN, solid pseudopapillary neoplasm; NET, pancreatic neuroendocrine tumor; CT, computed tomography; MRI, magnetic resonance imaging; MICE, multiple imputation by chained equations; RFE, recursive feature elimination; SHAP, shapley additive explanations; AUC, area under the receiver operating characteristic curve; CI, confidence interval; ACG, American college of gastroenterology; ESG, European study group on cystic tumours of the pancreas.

Pancreatic Cystic Lesions (TMIPDL-PCLs), which spans 1 January 2011 to 31 December 2024 and comprises data from five hospitals in Tianjin, China (Tianjin Medical University Cancer Institute and Hospital; Tianjin Medical University General Hospital; Tianjin Nankai Hospital; Tianjin First Central Hospital; and Second Hospital of Tianjin Medical University). Patients from the TMIPDL-PCLs cohort were stratified into development and temporal validation datasets based on enrollment periods: 2011–2022 and 2023–2024, respectively. A prospective clinical evaluation was then conducted at Tianjin Medical University Cancer Institute from 25 June 2025 to 25 June 2026 among patients with PCLs who

underwent SIRS-PCL diagnostic classification and, when applicable, risk stratification. The detailed study design is shown in **Figure 1**.

Patients were eligible if they had undergone surgical resection of PCLs with a definitive postoperative histopathological diagnosis and preoperative contrast-enhanced (CE) computed tomography (CT) or magnetic resonance imaging (MRI) performed within 1 month before surgery. Exclusion criteria were: previous or concurrent malignancy; any preoperative anticancer treatment; pseudocysts and purely solid lesions on imaging; no contrast-enhanced imaging sequences; and no definitive pathological diagnosis. The institutional ethics committee of Tianjin Medical University Cancer Institute and Hospital approved the study (registration number bc20255885).

Diagnostic definitions for PCL classification and reference standard for risk stratification

The final pathological diagnosis for each patient was extracted from pathology reports[5]. Because the study spanned a 14-year period during which pathological classification criteria evolved, all original pathological diagnoses were independently reviewed by two experienced pancreatic pathologists. Any discrepancies were resolved *via* consensus with a third senior pancreatic pathologist. Pathological subtype classification was standardized according to the fifth edition of the World Health Organization (WHO) Classification of Digestive System Tumours[5], and pathological risk stratification for IPMN and MCN according to the World Gastroenterology Organization (WGO) criteria before model development (**Appendix 1**)[14]. For IPMN, risk classification

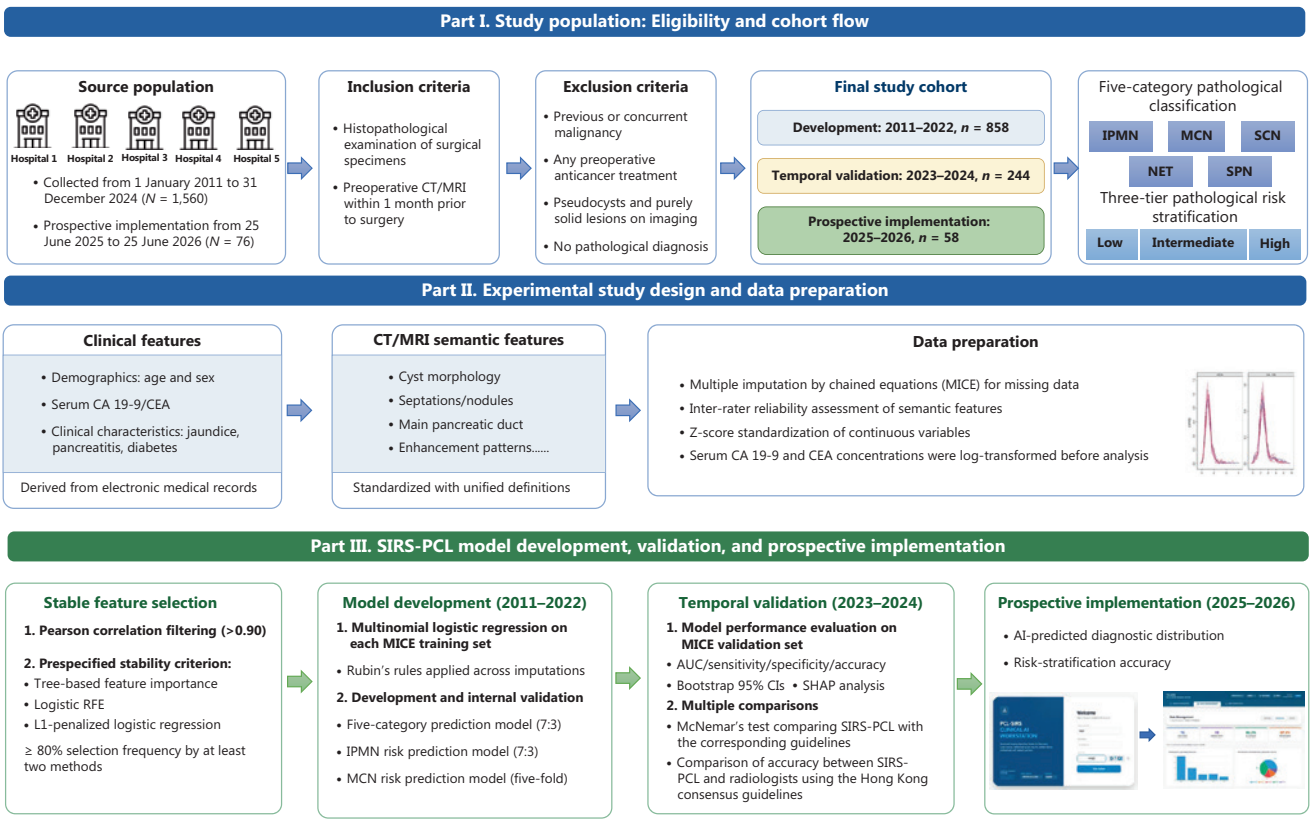


Figure 1 Study design and analytical workflow for the development and validation of SIRS-PCL. The workflow summarizes (I) the study population, eligibility criteria, and cohort allocation; (II) imaging and clinical data preparation; and (III) SIRS-PCL model development, temporal validation, and prospective implementation. SIRS-PCL comprises a five-category pathological classification model and a three-tier pathological risk-stratification model for IPMN and MCN. AUC, area under the receiver operating characteristic curve; CA 19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; CI, confidence interval; CT, computed tomography; IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; MICE, multiple imputation by chained equations; MRI, magnetic resonance imaging; PCL, pancreatic cystic lesion; NET, pancreatic neuroendocrine tumor; RFE, recursive feature elimination; SCN, serous cystic neoplasm; SHAP, Shapley Additive Explanations; SIRS-PCL, Structured Imaging Report System for Pancreatic Cystic Lesions; SPN, solid pseudopapillary neoplasm.

comprised three tiers: (1) low risk was defined as low-grade dysplasia managed with annual follow-up; (2) medium risk was defined as high-grade dysplasia or carcinoma *in situ* requiring further diagnostic evaluation; and (3) high risk was defined as invasive carcinoma warranting surgical intervention. For MCN, risk classification similarly comprised three tiers: (1) benign/low risk was defined as benign lesions or low-grade dysplasia managed with annual surveillance; (2) medium risk was defined as high-grade dysplasia or carcinoma *in situ*; and (3) high risk was defined as invasive carcinoma warranting surgical resection[14]. These categories represent study-defined pathology-based analytical tiers derived from standardized pathological diagnoses and were used for model development and comparative analyses rather than as substitutes for formal WHO pathological grading terminology; the detailed definitions and corresponding mapping criteria are provided in **Tables S1** and **S4**.

Imaging protocol

Abdominal imaging was performed using CE-MRI or CE-CT to evaluate PCLs in enrolled patients. For patients who underwent surgical resection, preoperative imaging studies obtained within 1 month before surgery were anonymized and independently reviewed by any two of four abdominal radiologists (X. Zheng, T. Liu, J. Liang, J. Liu, each with ≥5 years of abdominal imaging experience). All reports were subsequently verified by two senior radiologists with more than 10 years of experience (X. Cui or S. Fan) who were blinded to pathological information. Radiology features included 20 categorical imaging variables and nine continuous variables; specific radiology features are presented in **Figure 2A, B**. The main variable definitions and illustrative diagrams are provided in **Appendix 1**.

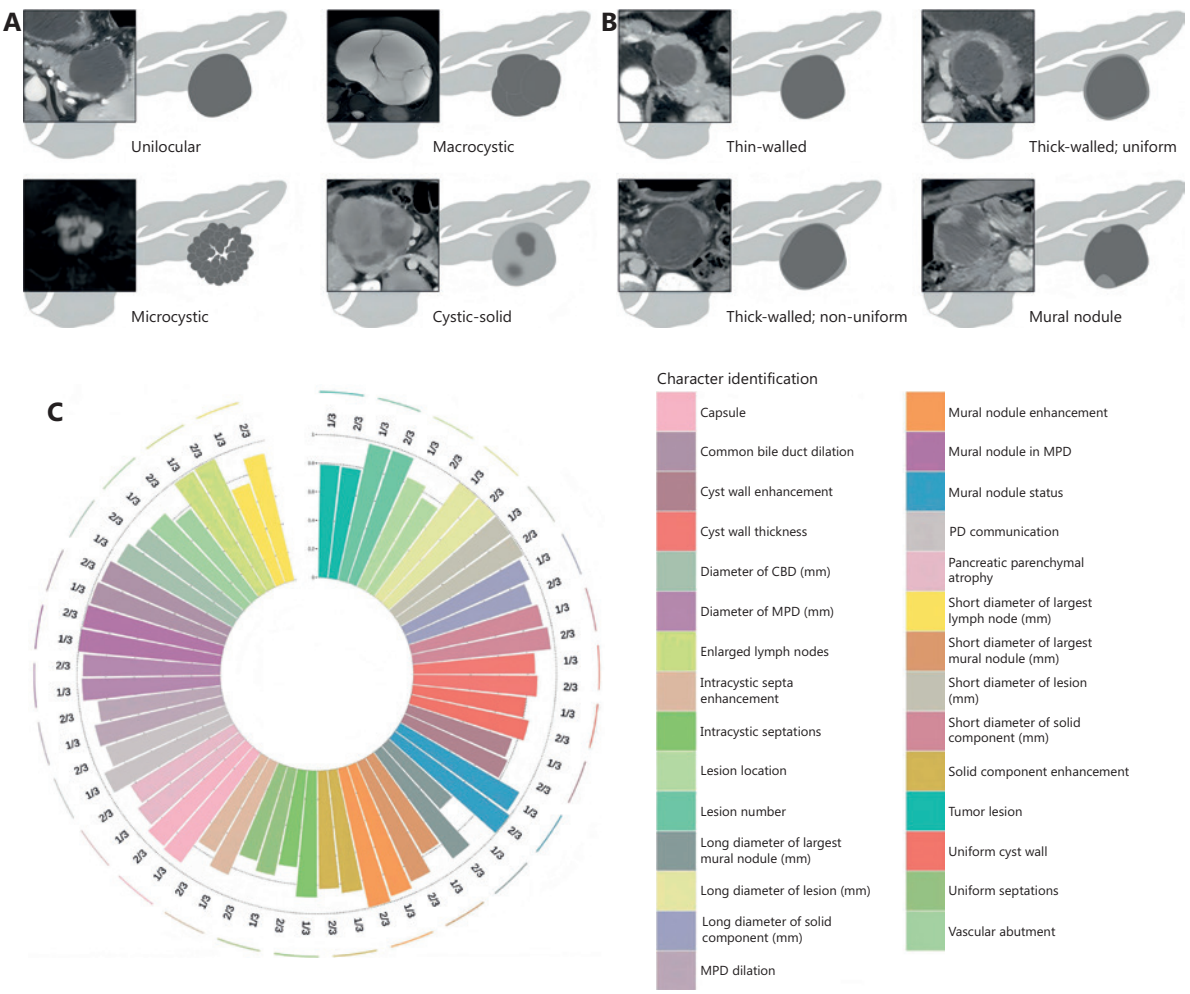


Figure 2 Main PCL imaging features and inter-rater reliability coefficients for imaging variables. (A) Four morphology types for tumor lesions. (B) Definitions for the status of cyst walls. (C) Inter-rater reliability coefficients for the imaging variables assessed. CBD, common bile duct; CT, computed tomography; MPD, main PD; MRI, magnetic resonance imaging; PD, pancreatic duct; SIRS-PCL, Structured Imaging Report System for Pancreatic Cystic Lesions.

Clinical data

Clinical information, including demographic characteristics, medical history, and laboratory data, were extracted from electronic medical records. The demographic information collected included patient age and sex. Medical history documented the presence or absence of three conditions: diabetes mellitus, jaundice, and pancreatitis. Laboratory data included serum carbohydrate antigen 19-9 (CA 19-9) and carcinoembryonic antigen (CEA) measured within 1 month before surgery. The rate of missing data was 14.9% for CEA and 14.7% for CA 19-9 in the development cohort, and 13.3% for both markers in the temporal validation cohort. Missing values were handled using multiple imputation by chained equations (MICE), as detailed in **Appendix 2**.

Model development and temporal validation

A structured imaging report system for PCLs (SIRS-PCL) was developed and validated in accordance with established guidance for clinical prediction modeling. The process comprised the development of three complementary prediction models: (1) a five-category pathological classification model to distinguish five PCLs subtypes (IPMN, MCN, SCN, SPN, and NET); (2) a three-tier IPMN risk stratification model (low, medium, and high risk); and (3) a three-tier MCN risk stratification model (benign/low, medium, and high risk). The complete analytical workflow, including data preprocessing, feature selection, model development, and validation, is summarized in **Figure 1**.

We assessed inter-rater reliability among radiologists to quantify measurement consistency (**Table S2**). Two-stage feature selection was then used to identify informative predictors while limiting multicollinearity, imputation uncertainty, and method-specific selection bias. First, highly correlated predictors were identified using a Pearson correlation threshold of > 0.90 and filtered within the development data before model fitting; when two or more predictors showed high correlation, clinical expertise informed retention of the most clinically relevant or reproducibly measured variable. Feature selection was then performed across the ten imputed datasets using a multimethod ensemble strategy. Candidate predictors were independently screened using recursive feature elimination, tree-based feature importance, and L1-regularised regression, and were retained within each selection run if selected by at least two methods. Selection frequencies were aggregated across imputations and resampling runs to assess predictor stability. On the basis of established methodology for stability selection and previous applications, a prespecified selection-frequency threshold of 80% was applied consistently across models, corresponding to selection in at least eight of 10 imputed datasets for the five-category and IPMN models, and at least 40 of 50 imputation-fold runs for the MCN model. This criterion was used to retain predictors consistently selected across repeated analyses rather than being optimized according to model performance (**Table S3**)[15,16]. Categorical predictors were one-hot encoded, and selected encoded features were mapped back to their corresponding source variables to preserve clinical interpretability. When multiple selected encoded levels represented the same underlying absence or negative imaging state (such as an absent cyst wall, no mural nodule, or no enhancement), these overlapping levels were reconciled at the source-variable level, where applicable, before final modeling. The final models therefore used the retained source variables and their corresponding encoded representations, together with selected continuous predictors, as the modeling feature set.

Model generalizability and performance were externally validated in an independent cohort recruited between January 2023 and December 2024. Primary temporal validation was performed *via* multiple imputation, and predictions were generated within 10 imputed temporal validation datasets using the pooled coefficients from model development. Final temporal validation predictions were obtained by averaging predicted probabilities across all imputed datasets for each patient; the class with the highest averaged probability was assigned as the final prediction.

Statistical analysis

Results for continuous variables with a skewed distribution are reported as the median and interquartile range (IQR). Results for categorical variables are reported as the frequency and percentage [*n* (%)]. Continuous variables were standardized using Z-score normalization using the parameters derived for model development and validation, with serum CA 19-9 and CEA concentrations log-transformed before standardization. Inter-rater reliability among three radiologists was assessed using Gwet's agreement coefficient 1 (AC1) statistic for categorical variables and Lin's concordance correlation coefficient (CCC) for continuous variables with a non-normal distribution[17,18]. Pairwise comparisons were conducted between the senior radiologist and each junior radiologist, and variables were retained if at least one comparison exceeded the predefined thresholds (CCC > 0.85; AC1 > 0.75).

Multinomial logistic regression for all three prediction models was applied to the MICE dataset[19-21]. The final pooled model was evaluated on the test set from the last imputed dataset using Rubin's rules[21]. For the five-category subtype discrimination model and IPMN risk stratification model, imputed data were partitioned into a development dataset at a 7:3 ratio. For the MCN risk stratification model, fivefold cross-validation was applied. Temporal validation was conducted in the temporal validation dataset (from 2023 to 2024). Decision curve analysis (DCA) was used to compare the clinical utility of the SIRS-PCL system against the "radiologist diagnosis plus Hong Kong consensus" approach in the temporal validation cohort with complete data[7,8,11,12]. Model complexity was

determined after stability-based feature selection, and the validation strategy was selected according to the sample size available for each prediction task to maximize model stability and generalizability while minimizing the risk of overfitting[22,23]. Detailed information presented in **Appendix 2**.

Model performance was evaluated in terms of the area under the receiver operating characteristic curve (AUC), precision, sensitivity, and specificity. Uncertainty in performance metrics was quantified *via* bootstrap resampling (1,000 iterations) with 95% confidence intervals (CIs). Predictor effects are reported as the odds ratio (OR) with 95% CI, and SHAP (Shapley additive explanation) analysis was conducted to identify the contribution of each predictor. The 95% CIs for the end-to-end accuracy of SIRS-PCL were calculated using the Wilson score method based on the proportion of correctly classified cases among all patients evaluated. Comparisons of the IPMN or MCN prediction model against the current four guidelines were performed using McNemar's test. Bonferroni correction was applied separately to the prespecified disease-specific families of guideline comparisons. Because four guidelines provided applicable risk-management criteria for IPMN and two for MCN, the corresponding significance thresholds were $\alpha = 0.0125$ (0.05/4) and $\alpha = 0.025$ (0.05/2). These comparisons were based on overall diagnostic accuracy rather than separate hypothesis tests across the three pathological risk categories[7,8,11,12].

All statistical analyses were performed using Python (version 3.9.25)[24]. Data imputation and visualization were implemented in R (version 4.4.1). Reporting is in accordance with the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence) guidelines for the development and validation of prediction models.

Patient and public involvement

We sought input and suggestions on result visualization and model interpretability from both patients with PCLs and surgical oncologists. The goal was to provide probability data for the different subtypes and to demonstrate how different indicators contribute to prediction outcomes when making decisions about surgery. Considering that different radiological features may affect the surgical approach, our interface incorporates recommendations that allow real-time comparison to multiple clinical guidelines, and thus provides comprehensive insights for patient information and multidisciplinary team communication.

Results

Patient characteristics

Among 1,560 patients identified, 458 (29.3%) are excluded (**Figure 1**). A total of 1,102 patients with PCLs were thus included in the study (median age 56 years, IQR 44–64), of whom 714 were female (64.8%). The development cohort comprised 858 patients (median age, 55 years (IQR, 44–64 years), of whom 567 were female (66.1%); the temporal validation cohort comprised 244 patients median age, 58 years (IQR, 45–66 years), of whom 147 were female (60.2%). All 29 imaging variables had acceptable inter-rater agreement (CCC/AC1 coefficients, 0.598–0.998) and were retained for analysis (**Figure 2C**). Baseline demographic and clinical characteristics are presented in **Table 1**.

Among the 1,102 patients, the distribution of pathological subtypes was 314 (28.5%) with IPMN, 185 (16.8%) with MCN, 114 (10.3%) with NET, 355 (32.2%) with SCN, and 134 (12.2%) with SPN. The detailed distribution across cohorts is presented in **Table 1**. For IPMN ($n = 314$), risk stratification identified 116 (36.9%) low-risk, 50 (15.9%) medium-risk, and 148 (47.1%) high-risk cases. For MCN ($n = 185$), risk stratification identified 135 (73.0%) benign/low-risk, 14 (7.6%) medium-risk, and 36 (19.5%) high-risk cases. Full demographic data for the entire dataset and risk stratification details for IPMN and MCN are presented in **Appendix 3**.

In the prospective implementation cohort, 76 patients with PCLs were evaluated using SIRS-PCL. The median age was 60 years (IQR 52–66), and 49 patients (64.5%) were female. Pathological diagnoses were available for 60 patients and included SCN in 20, IPMN in 19, MCN in seven, NET in seven, SPN in five, benign/other in one, and malignant/other in one; 16 patients had no pathological diagnosis. Finally, 58 patients with complete pathological diagnoses were enrolled in a study to assess the diagnostic and risk stratification performance of SIRS-PCL. The baseline characteristics of this prospective cohort are presented in **Appendix 3**.

Feature selection, model development, and model performance

The final model for five-category subtype classification incorporates 17 predictors, while the IPMN and MCN risk stratification models each incorporate nine key imaging features, as shown in **Appendix 2**. Full equations for all three models are presented in **Appendices 4–6**. Results for sensitivity analyses conducted using complete-case datasets without MICE imputation across all three models are presented in **Appendix 7**.

Pathological classification model

For the pathological classification model in the temporal validation cohort, the mean AUC was 0.93 (95% CI 0.91–0.95). The performance metrics for each pathological subtype were as follows: AUC of 0.96 (95% CI 0.93–0.98) for IPMN, 0.98 (95% CI 0.95–0.99) for SPN, 0.91 (95% CI 0.87–0.95) for MCN, 0.90 (95% CI 0.87–0.94) for SCN, and 0.87 (95% CI 0.78–0.95) for NET. Detailed information is presented in **Table 2** and

Table 1 Descriptive statistics for patients in the entire cohort, including clinical and demographic characteristics

Characteristic	Development cohort						Validation cohort					
	Total *n* = 858	IPMN *n* = 236	MCN *n* = 152	NET *n* = 96	SCN *n* = 281	SPN *n* = 93	Total *n* = 244	IPMN *n* = 78	MCN *n* = 33	NET *n* = 18	SCN *n* = 74	SPN *n* = 41
Female, *n* (%)	567 (66.1)	95 (40.2)	125 (82.2)	51 (53.1)	217 (77.2)	79 (84.9)	147 (60.2)	31 (39.7)	28 (84.8)	5 (27.8)	50 (67.6)	33 (80.5)
Age (years)	55 (44–64)	63 (56–67)	55 (41–62)	54 (45–59)	55 (45–63)	29 (18–39)	58 (45–66)	65 (58–69)	51 (40–58)	55 (40–63)	60 (52–66)	30 (22–41)
Imaging modality, *n* (%)												
CE-CT	590 (68.8)	153 (64.8)	113 (74.3)	64 (66.7)	197 (70.1)	63 (67.7)	166 (68.0)	53 (67.9)	29 (87.9)	11 (61.1)	49 (66.2)	24 (58.5)
CE-MRI	268 (31.2)	83 (35.2)	39 (25.7)	32 (33.3)	84 (29.9)	30 (32.2)	78 (32.0)	25 (32.1)	4 (12.1)	7 (38.9)	25 (33.8)	17 (41.5)
Risk category, *n* (%)												
Benign	313 (36.5)	0 (0.0)	49 (32.2)	0 (0.0)	264 (94.0)	0 (0.0)	70 (28.7)	0 (0.0)	9 (27.3)	0 (0.0)	61 (82.4)	0 (0.0)
Low risk	153 (17.8)	86 (36.4)	60 (39.5)	0 (0.0)	7 (2.5)	0 (0.0)	59 (24.1)	30 (38.5)	17 (51.5)	0 (0.0)	12 (16.2)	0 (0.0)
Medium risk	160 (18.6)	41 (17.3)	12 (7.9)	30 (31.3)	10 (3.6)	67 (72.0)	48 (19.7)	9 (11.5)	2 (6.1)	5 (27.8)	1 (1.4)	31 (75.6)
High risk	232 (27.0)	109 (46.2)	31 (20.4)	66 (68.7)	0 (0.0)	26 (28.0)	67 (27.5)	39 (50.0)	5 (15.1)	13 (72.2)	0 (0.0)	10 (24.4)
Diabetes, *n* (%)												
Yes	154 (17.9)	56 (23.7)	21 (13.8)	20 (20.8)	51 (18.1)	6 (6.5)	31 (12.7)	18 (23.1)	2 (6.1)	0 (0.0)	9 (12.2)	2 (4.9)
No	704 (82.1)	180 (76.3)	131 (86.2)	76 (79.2)	230 (81.9)	87 (93.5)	213 (87.3)	60 (76.9)	31 (93.9)	18 (100.0)	65 (87.8)	39 (95.1)
Jaundice, *n* (%)												
Yes	47 (5.5)	19 (8.1)	8 (5.3)	6 (6.3)	10 (3.6)	4 (4.3)	9 (3.7)	6 (7.7)	0 (0.0)	1 (5.6)	2 (2.7)	0 (0.0)
No	811 (94.5)	217 (91.9)	144 (94.7)	90 (93.7)	271 (96.4)	89 (95.7)	235 (96.3)	72 (92.3)	33 (100.0)	17 (94.4)	72 (97.3)	41 (100.0)
Pancreatitis, *n* (%)												
Yes	23 (2.7)	13 (5.5)	5 (3.3)	3 (3.1)	1 (0.4)	1 (1.1)	3 (1.2)	2 (2.6)	1 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
No	835 (97.3)	223 (94.5)	147 (96.7)	93 (96.9)	280 (99.6)	92 (98.9)	241 (98.8)	76 (97.4)	32 (97.0)	17 (100.0)	74 (100.0)	41 (100.0)
CA19-9 (U/mL)	12.6 (7.2–27.5)	19.6 (10.3–83.8)	14.1 (7.4–39.1)	14.1 (8.0–31.6)	10.6 (6.2–17.9)	9.8 (5.9–14.0)	11.0 (6.2–21.6)	18.2 (8.7–71.3)	9.2 (6.4–17.5)	17.1 (6.9–25.8)	8.5 (4.9–11.6)	8.6 (4.1–17.7)
CEA (ng/mL)	1.9 (1.2–3.0)	2.8 (1.8–4.0)	1.9 (1.0–3.1)	1.8 (1.4–2.7)	1.7 (1.2–2.3)	1.1 (0.7–1.7)	1.8 (1.2–3.1)	2.7 (1.9–4.1)	1.8 (1.0–3.8)	2.1 (0.9–2.6)	1.6 (1.2–2.5)	1.1 (0.6–1.5)

[a]CE-CT, contrast-enhanced computed tomography; CE-MRI, contrast-enhanced magnetic resonance imaging; range; IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; SCN, serous cystic neoplasm; SPN, solid pseudopapillary neoplasm; NET, neuroendocrine tumor.

Table 2 Classification performance of the five-category pathological classification model in four cohorts

Cohort	Histologic type	Prediction performance					
		AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Accuracy
Training (*n* = 600)	IPMN	0.95 (0.93–0.96)	0.81	0.91	0.77	0.93	0.88
	MCN	0.85 (0.80–0.88)	0.48	0.91	0.54	0.89	0.83
	SCN	0.90 (0.87–0.93)	0.73	0.86	0.71	0.86	0.81
	NET	0.97 (0.95–0.98)	0.78	0.97	0.74	0.97	0.94
	SPN	0.99 (0.97–0.99)	0.85	0.99	0.90	0.98	0.97
	Mean	0.92 (0.90–0.94)	0.73	0.93	0.73	0.93	0.89
Internal validation (*n* = 258)	IPMN	0.95 (0.93–0.97)	0.79	0.92	0.79	0.92	0.88
	MCN	0.87 (0.81–0.94)	0.61	0.92	0.61	0.92	0.86
	SCN	0.90 (0.86–0.94)	0.75	0.86	0.72	0.88	0.83
	NET	0.96 (0.88–0.99)	0.76	0.98	0.81	0.97	0.95
	SPN	0.98 (0.95–1.00)	0.82	0.98	0.85	0.98	0.97
	Mean	0.93 (0.90–0.95)	0.75	0.93	0.76	0.93	0.90
Temporal validation (*n* = 244)	IPMN	0.96 (0.93–0.98)	0.90	0.90	0.80	0.95	0.90
	MCN	0.91 (0.87–0.95)	0.54	0.90	0.46	0.93	0.85
	SCN	0.90 (0.87–0.94)	0.62	0.91	0.75	0.85	0.82
	NET	0.87 (0.78–0.95)	0.50	0.94	0.39	0.96	0.91
	SPN	0.98 (0.95–0.99)	0.73	0.98	0.88	0.95	0.94
	Mean	0.93 (0.91–0.95)	0.66	0.93	0.66	0.93	0.88
Prospective validation (*n* = 58)	IPMN	—	1.00	0.85	0.76	1.00	0.90
	MCN	—	0.86	0.98	0.86	0.98	0.97
	SCN	—	0.65	0.97	0.93	0.84	0.86
	NET	—	1.00	1.00	1.00	1.00	1.00
	SPN	—	1.00	1.00	1.00	1.00	1.00
	Mean	—	0.90	0.96	0.91	0.96	0.86

AUC, area under the receiver operating characteristic curve; CI, confidence interval; IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; SCN, serous cystic neoplasm; SPN, solid pseudopapillary neoplasm; NET, neuroendocrine tumor; NPV, negative predictive value; PPV, positive predictive value.

Figure 3A. The same feature set was applied to develop a prediction model using the full dataset (**Figure 3B**).

Each pathological subtype demonstrated distinct predictive features. IPMN was most strongly associated with pancreatic duct (PD) communication (OR 15.96, 95% CI 9.37–27.18). SCN showed characteristic microcystic morphology (OR 4.40, 95% CI 2.31–8.37). MCN was associated with absent cyst-wall enhancement (OR 2.24, 95%CI 1.24–4.06) and elevated serum CA 19-9 levels (OR 1.42, 95% CI 1.01–1.99). SPN was characterized by younger patient age (OR 0.17, 95% CI 0.10–0.28) and capsule presence (OR 4.33, 95% CI 1.97–9.50). NET demonstrated arterial-phase enhancement (OR 5.58, 95% CI 3.43–9.06) and rarely showed PD communication (OR 0.28, 95% CI 0.14–0.55). Complete predictor details and subgroup analyses according to imaging modality are presented in **Appendix 4**; the latter demonstrated comparable diagnostic performance between CT-based and MRI-based examinations.

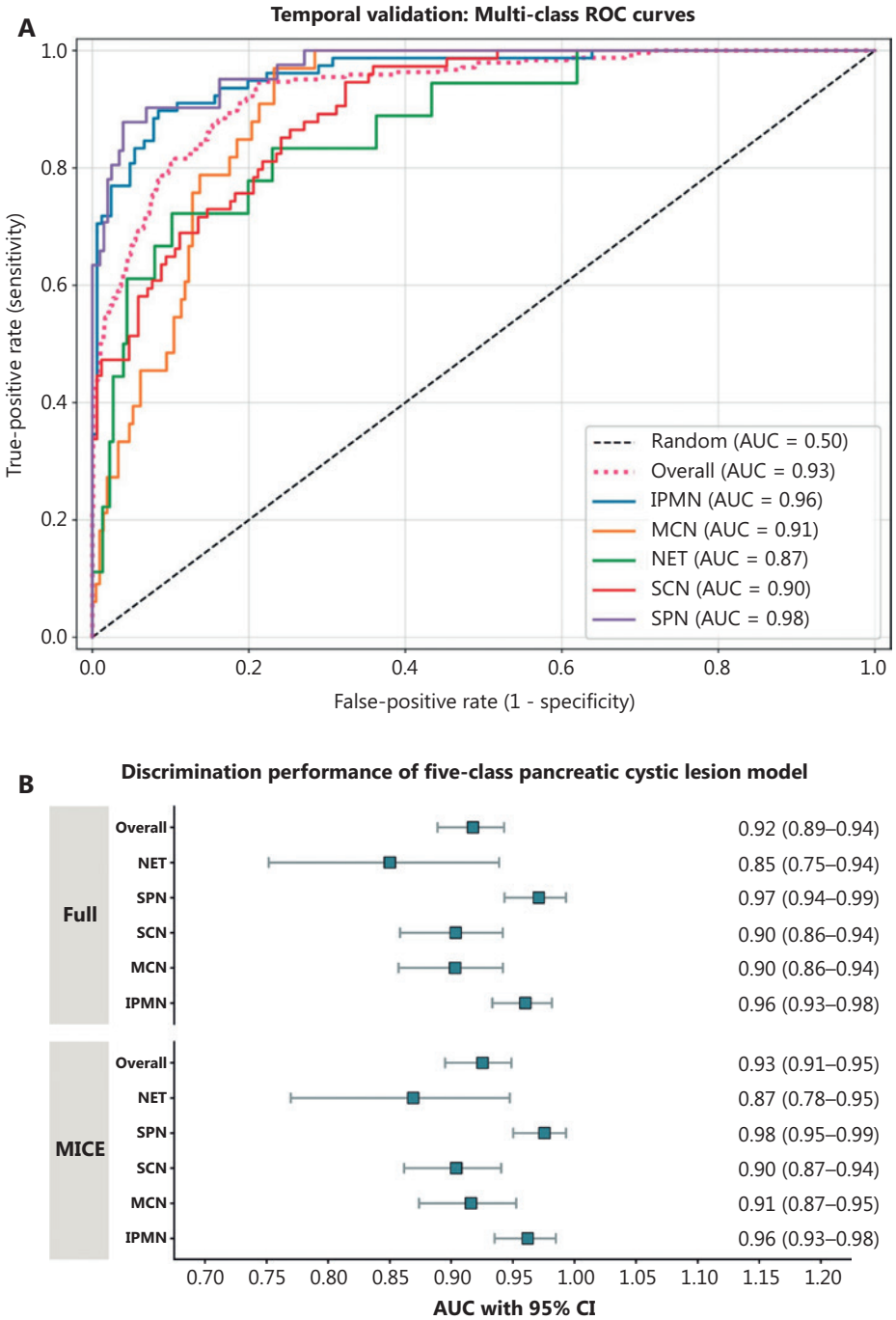


Figure 3 Discrimination performance of the SIRS-PCL five-category pathological classification model. (A) ROC curves for five-category pathological classification in the temporal validation cohort. (B) Class-specific and overall AUCs with 95% CIs for complete-case and MICE analyses. AUC, area under the ROC curve; CI, confidence interval; IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; MICE, multiple imputation by chained equations; NET, neuroendocrine tumor; ROC, receiver operating characteristic; SCN, serous cystic neoplasm; SPN, solid pseudopapillary neoplasm.

IPMN risk stratification model

The mean AUC for predicting three-tier IPMN risk stratification was 0.86 (95% CI 0.79–0.93), with performance varying by risk tier: AUC 0.92 (95% CI 0.85–0.97) for high risk, 0.71 (95% CI 0.56–0.86) for medium risk, and 0.83 (95% CI 0.73–0.92) for low risk (**Table 3**). Each risk category

Table 3 Prediction performance of the IPMN risk stratification model in four cohorts

Cohort	Pathological risk stratification	Prediction performance					
		AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Accuracy
Training (*n* = 165)	Low risk	0.90 (0.85–0.94)	0.90	0.79	0.71	0.93	0.83
	Medium risk	0.81 (0.72–0.90)	0.38	0.96	0.69	0.88	0.86
	High risk	0.92 (0.87–0.96)	0.82	0.88	0.85	0.85	0.85
	Mean	0.89 (0.85–0.93)	0.70	0.88	0.75	0.89	0.85
Internal validation (*n* = 71)	Low risk	0.96 (0.91–0.99)	0.88	0.89	0.82	0.93	0.89
	Medium risk	0.70 (0.52–0.86)	0.08	0.97	0.33	0.84	0.82
	High risk	0.93 (0.86–0.98)	0.94	0.76	0.78	0.94	0.85
	Mean	0.90 (0.83–0.96)	0.64	0.87	0.64	0.90	0.85
Temporal validation (*n* = 78)	Low risk	0.83 (0.73–0.92)	0.73	0.81	0.71	0.83	0.78
	Medium risk	0.71 (0.56–0.86)	0.33	0.91	0.33	0.91	0.85
	High risk	0.92 (0.85–0.97)	0.87	0.90	0.89	0.88	0.88
	Mean	0.86 (0.79–0.93)	0.65	0.87	0.65	0.87	0.84
Prospective validation (*n* = 19)	Low risk	–	0.91	0.63	0.77	0.83	0.79
	Medium risk	–	0.00	0.94	0.00	0.89	0.84
	High risk	–	0.67	0.92	0.80	0.86	0.84
	Mean	–	0.53	0.83	0.52	0.86	0.74

AUC, area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

demonstrated distinctive imaging characteristics. High-risk IPMN was characterized by vascular abutment (OR 2.56, 95% CI 1.10–5.94) and larger lesion diameter (OR 1.59, 95% CI 1.06–2.40). Low-risk IPMN was associated with uniform cyst-wall thickness (OR 1.82, 95% CI 1.14–2.89), thin walls, and smaller size (OR 0.57, 95% CI 0.38–0.86). Medium-risk IPMN had thick cyst walls (OR 1.42, 95% CI 0.82–2.47) and main PD dilatation. Detailed OR values, SHAP results, and subgroup analyses by imaging modality are provided in **Appendix 5**.

MCN risk stratification model

The MCN model achieved high discriminatory performance across all risk categories (mean AUC 0.97, 95% CI 0.91–1.00). High-risk MCN was associated with a short diameter for mural nodules (OR 1.85, 95% CI 1.22–2.81) and solid components (OR 1.66, 95% CI, 1.23–2.24). Low-risk MCN was characterized by unilocular morphology (OR 2.01, 95% CI 1.24–3.28), uniform cyst-wall thickness (OR 2.25, 95% CI 1.49–3.41), and a pancreatic body or tail location (OR 2.27, 95% CI 1.06–4.83). Complete model results are presented in **Table 4**, with comprehensive OR and SHAP values, and subgroup analyses by imaging modality detailed in **Appendix 6**.

Comparison of SIRS-PCL *versus* real-world clinical practice for end-to-end assessment

Performance results for SIRS-PCL in comparison to radiologists using the Hong Kong consensus guidelines, as well as clinical examples of its application, are presented in **Appendix 8**. SIRS-PCL achieved overall accuracy of 72.5% (95% CI 66.1–78.1%) across the five pathological classifications. By contrast, radiologists achieved overall accuracy of 56.9% (95% CI 50.1–63.4%). Furthermore, SIRS-PCL demonstrated higher sensitivity than radiologists across the IPMN (90.0% *vs.* 67.1%), MCN (53.8% *vs.* 46.2%), SCN (64.4% *vs.* 40.7%), SPN (76.3% *vs.* 73.7%), and NET (50.0% *vs.* 27.8%) categories. The model also exhibited robust specificity (IPMN 92.2%, MCN 90.8%, SCN 90.8%, SPN 98.3%, NET 93.3%) in comparison to the radiologists (IPMN 90.8%, MCN 89.2%, SCN 88.2%, SPN 96.0%, NET 98.4%). Overall, SIRS-PCL would result in rates of 11.5% (12/104) for unnecessary surgery, 6.7% (7/104) for overexamination, and 13.3%

Table 4 Prediction performance of the MCN risk stratification model cross-validation, temporal validation, and prospective validation datasets

Cohort	Pathological risk stratification	Prediction performance					
		AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	Accuracy
Cross-validation	Benign/low risk	0.88 (0.86–0.90)	0.94	0.68	0.88	0.83	0.87
	Medium risk	0.66 (0.60–0.72)	0.03	0.97	0.08	0.92	0.90
	High risk	0.93 (0.92–0.95)	0.69	0.92	0.68	0.92	0.87
	Mean	0.88 (0.85–0.90)	0.55	0.86	0.55	0.89	0.88
Temporal validation (n = 33)	Benign/low risk	0.97 (0.90–1.00)	0.88	1.00	1.00	0.70	0.91
	Medium risk	1.00 (1.00–1.00)	1.00	0.97	0.67	1.00	0.97
	High risk	0.97 (0.91–1.00)	1.00	0.93	0.71	1.00	0.94
	Mean	0.97 (0.91–1.00)	0.96	0.97	0.79	0.90	0.94
Prospective validation (n = 7)	Benign/low risk	–	1.00	1.00	1.00	1.00	1.00
	Medium risk	–	–	–	–	–	–
	High risk	–	1.00	1.00	1.00	1.00	1.00
	Mean	–	1.00	1.00	1.00	1.00	1.00

AUC, area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

(13/98) for missed diagnosis. By contrast, the Hong Kong consensus led to substantially higher rates for unnecessary surgery 23.1% (24/104) and overexamination 33.7% (35/104), and a slightly lower rate for missed diagnosis 12.2% (12/98). The original confusion matrix is presented in **Figure 4**.

DCA revealed relatively small differences in net benefit between the two strategies at lower threshold probabilities. From approximately 0.10 onwards, SIRS-PCL demonstrated a greater net benefit than radiologist assessment plus guideline-based interpretation across most of the threshold range evaluated (**Appendix 4**).

Comparison of SIRS-PCL with clinical guidelines for IPMN and MCN risk prediction

In a subset of 70 IPMN cases with complete data (no imputation), the proposed model achieved the highest overall accuracy (77.1%), which was significantly higher than with the European (58.6%), Hong Kong (51.4%), Kyoto (50.0%), and ACG (58.6%) guidelines. McNemar's test with Bonferroni correction (α = 0.0125) confirmed statistical significance for all pairwise guideline comparisons (P =0.007 *vs*. ACG; $P < 0.001$ *vs* Hong Kong, and Kyoto; $P = 0.004$ *vs*. European guidelines). Detailed comparison results are shown in **Figure 5A**.

In a subset of 26 complete MCN cases, the model had superior overall accuracy (92.3%) in comparison to the ACG (34.6%) and Hong Kong (34.6%) guidelines. McNemar's test showed statistically significant differences for both comparisons (ACG $P < 0.001$; Hong Kong $P < 0.001$). Detailed comparison results are shown in **Figure 5B**.

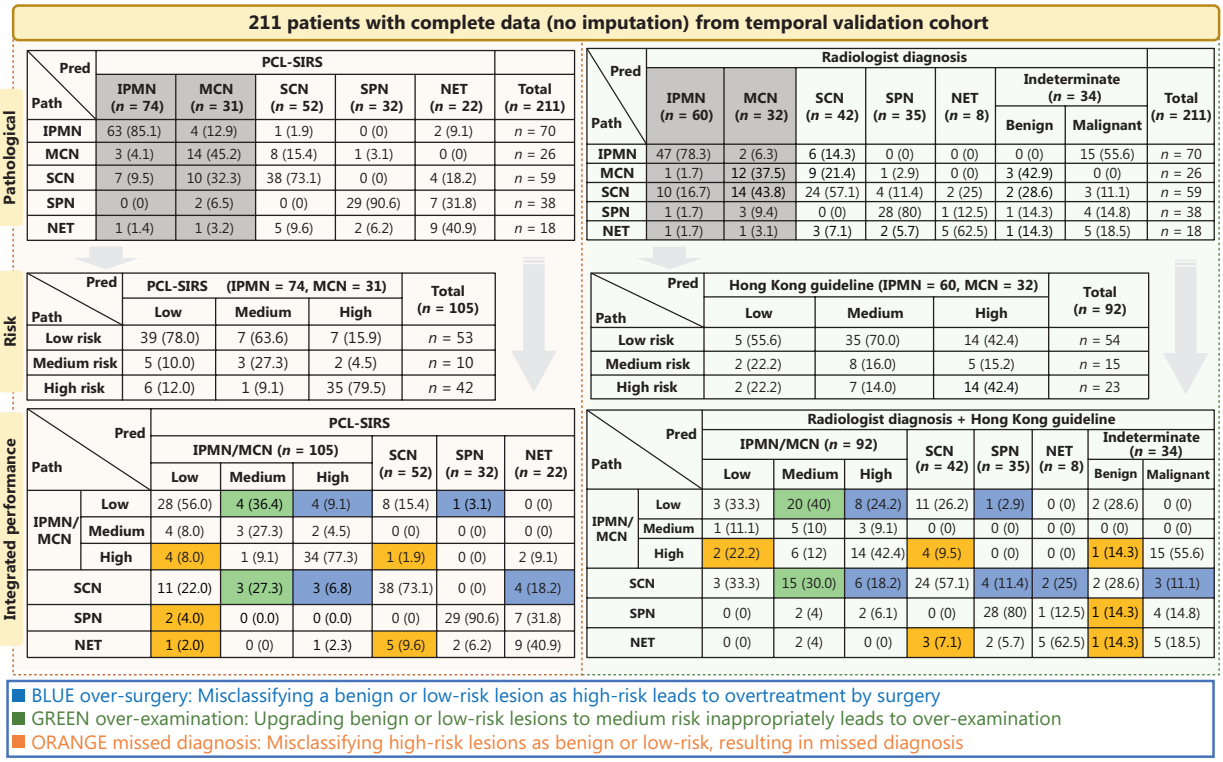
Prospective clinical application of SIRS-PCL

Among 58 prospective cases with a pathological diagnosis, SIRS-PCL achieved overall five-class diagnostic accuracy of 86.2%. Class-specific sensitivity was 100.0% for IPMN, 85.7% for MCN, 65.0% for SCN, 100.0% for NET, and 100.0% for SPN; the corresponding specificity values were 84.6%, 98.0%, 97.4%, 100.0%, and 100.0% (**Table 2**, **Figure 6A**).

For IPMN pathological risk stratification, 19 cases were evaluable, with overall accuracy of 73.7%. Sensitivity was 90.9% for the low-risk, 0% for the medium-risk, and 66.7% for the high-risk category, with corresponding specificity values of 62.5%, 94.1%, and 92.3% (**Table 3**, **Figure 6B**). For MCN pathological risk stratification, only seven cases were evaluable, all of which were correctly classified; however, these cases represented only the low-risk and high-risk categories, with no intermediate-risk MCN (**Table 4**, **Figure 6C**). Details for the prospective implementation cohort and platform-level summary statistics, including the number of cases registered, pathological reference availability, AI diagnostic accuracy, risk-stratification accuracy, and diagnostic distributions, are shown in **Appendix 8**.

Discussion

Our results demonstrate consistent SIRS-PCL performance for PCL diagnosis and risk stratification. In the temporal



Pathological	PCL-SIRS					
Path \ Pred	IPMN (*n* = 74)	MCN (*n* = 31)	SCN (*n* = 52)	SPN (*n* = 32)	NET (*n* = 22)	Total (*n* = 211)
IPMN	63 (85.1)	4 (12.9)	1 (1.9)	0 (0)	2 (9.1)	*n* = 70
MCN	3 (4.1)	14 (45.2)	8 (15.4)	1 (3.1)	0 (0)	*n* = 26
SCN	7 (9.5)	10 (32.3)	38 (73.1)	0 (0)	4 (18.2)	*n* = 59
SPN	0 (0)	2 (6.5)	0 (0)	29 (90.6)	7 (31.8)	*n* = 38
NET	1 (1.4)	1 (3.2)	5 (9.6)	2 (6.2)	9 (40.9)	*n* = 18

Pathological	Radiologist diagnosis							
Path \ Pred	IPMN (*n* = 60)	MCN (*n* = 32)	SCN (*n* = 42)	SPN (*n* = 35)	NET (*n* = 8)	Indeterminate (*n* = 34)		Total (*n* = 211)
						Benign	Malignant	
IPMN	47 (78.3)	2 (6.3)	6 (14.3)	0 (0)	0 (0)	0 (0)	15 (55.6)	*n* = 70
MCN	1 (1.7)	12 (37.5)	9 (21.4)	1 (2.9)	0 (0)	3 (42.9)	0 (0)	*n* = 26
SCN	10 (16.7)	14 (43.8)	24 (57.1)	4 (11.4)	2 (25)	2 (28.6)	3 (11.1)	*n* = 59
SPN	1 (1.7)	3 (9.4)	0 (0)	28 (80)	1 (12.5)	1 (14.3)	4 (14.8)	*n* = 38
NET	1 (1.7)	1 (3.1)	3 (7.1)	2 (5.7)	5 (62.5)	1 (14.3)	5 (18.5)	*n* = 18

Risk	PCL-SIRS (IPMN = 74, MCN = 31)			Total (*n* = 105)
Path \ Pred	Low	Medium	High	
Low risk	39 (78.0)	7 (63.6)	7 (15.9)	*n* = 53
Medium risk	5 (10.0)	3 (27.3)	2 (4.5)	*n* = 10
High risk	6 (12.0)	1 (9.1)	35 (79.5)	*n* = 42

Risk	Hong Kong guideline (IPMN = 60, MCN = 32)			Total (*n* = 92)
Path \ Pred	Low	Medium	High	
Low risk	5 (55.6)	35 (70.0)	14 (42.4)	*n* = 54
Medium risk	2 (22.2)	8 (16.0)	5 (15.2)	*n* = 15
High risk	2 (22.2)	7 (14.0)	14 (42.4)	*n* = 23

Integrated performance		PCL-SIRS					
Path \ Pred		IPMN/MCN (*n* = 105)			SCN (*n* = 52)	SPN (*n* = 32)	NET (*n* = 22)
		Low	Medium	High			
IPMN/MCN	Low	28 (56.0)	4 (36.4)	4 (9.1)	8 (15.4)	1 (3.1)	0 (0)
	Medium	4 (8.0)	3 (27.3)	2 (4.5)	0 (0)	0 (0)	0 (0)
	High	4 (8.0)	1 (9.1)	34 (77.3)	1 (1.9)	0 (0)	2 (9.1)
SCN		11 (22.0)	3 (27.3)	3 (6.8)	38 (73.1)	0 (0)	4 (18.2)
SPN		2 (4.0)	0 (0.0)	0 (0.0)	0 (0)	29 (90.6)	7 (31.8)
NET		1 (2.0)	0 (0)	1 (2.3)	5 (9.6)	2 (6.2)	9 (40.9)

Integrated performance		Radiologist diagnosis + Hong Kong guideline							
Path \ Pred		IPMN/MCN (*n* = 92)			SCN (*n* = 42)	SPN (*n* = 35)	NET (*n* = 8)	Indeterminate (*n* = 34)	
		Low	Medium	High				Benign	Malignant
IPMN/MCN	Low	3 (33.3)	20 (40)	8 (24.2)	11 (26.2)	1 (2.9)	0 (0)	2 (28.6)	0 (0)
	Medium	1 (11.1)	5 (10)	3 (9.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	High	2 (22.2)	6 (12)	14 (42.4)	4 (9.5)	0 (0)	0 (0)	1 (14.3)	15 (55.6)
SCN		3 (33.3)	15 (30.0)	6 (18.2)	24 (57.1)	4 (11.4)	2 (25)	2 (28.6)	3 (11.1)
SPN		0 (0)	2 (4)	2 (6.1)	0 (0)	28 (80)	1 (12.5)	1 (14.3)	4 (14.8)
NET		0 (0)	2 (4)	0 (0)	3 (7.1)	2 (5.7)	5 (62.5)	1 (14.3)	5 (18.5)

Figure 4 Comparison of SIRS-PCL with radiologist- and guideline-based assessment in the complete-data temporal validation cohort. Confusion matrices show (top) pathological classification, (middle) pathological risk stratification, and (bottom) integrated classification and risk-stratification performance for SIRS-PCL (left) and radiologist- and guideline-based assessment (right). IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; NET, neuroendocrine tumor; PCL, pancreatic cystic lesion; SCN, serous cystic neoplasm; SIRS-PCL, Structured Imaging Report System for Pancreatic Cystic Lesions; SPN, solid pseudopapillary neoplasm.

validation cohort, SIRS-PCL achieved diagnostic accuracy of 90% for IPMN, 85% for MCN, 91% for NET, 82% for SCN, and 94% for SPN. The disease-specific three-tier risk stratification models for IPMN and MCN also showed clinically relevant sensitivity and specificity for identifying low-risk and high-risk lesions. For IPMN risk stratification, sensitivity and specificity were 73% and 81% for low-risk disease, and 87% and 90% for high-risk disease, respectively. For MCN risk stratification, sensitivity and specificity were 88% and 100% for benign or low-risk disease, and 100% and 93% for high-risk disease, respectively. In comparison to radiologists using the Hong Kong guidelines, SIRS-PCL improved the diagnostic accuracy by 15.6 percentage points and reduced the rate of pathology-defined unnecessary surgery by 11.6 percentage points. In the prospective implementation cohort, SIRS-PCL maintained favorable five-category classification performance, with overall accuracy of 86.2%. Risk stratification performance was more variable: 73.7% of the 19 pathology-confirmed IPMN cases were correctly classified, whereas all seven evaluable MCN cases were correctly classified, although only the low-risk and high-risk MCN categories were represented. Collectively, these prospective findings demonstrate the potential of SIRS-PCL for clinical implementation and provide preliminary evidence supporting its applicability to pathological subtype classification and risk stratification in routine clinical practice.

Our multicenter study developed and validated a pathology-based structured imaging reporting system for preoperative evaluation of surgically evaluated neoplastic PCLs. Three principal findings merit further discussion. First, SIRS-PCL provides a pathology-based, end-to-end prediction framework that integrates preoperative classification of the pathological subtype with subsequent stratification of pathological risk. Deng et al. developed a clinical-radiological model in a cohort of 317 pathologically confirmed PCLs with worrisome or high-risk features to distinguish malignant from nonmalignant lesions[25]. Worrisome features and high-risk stigmata were established primarily for IPMN, but different PCL subtypes have distinct pathological origins, malignant potential, and clinical management strategies. Applying a common risk-feature framework across heterogeneous PCL subtypes may therefore limit subtype-specific interpretation and clinical applicability[25]. By contrast, SIRS-PCL incorporates pathological subtype identification as the initial step in the diagnostic

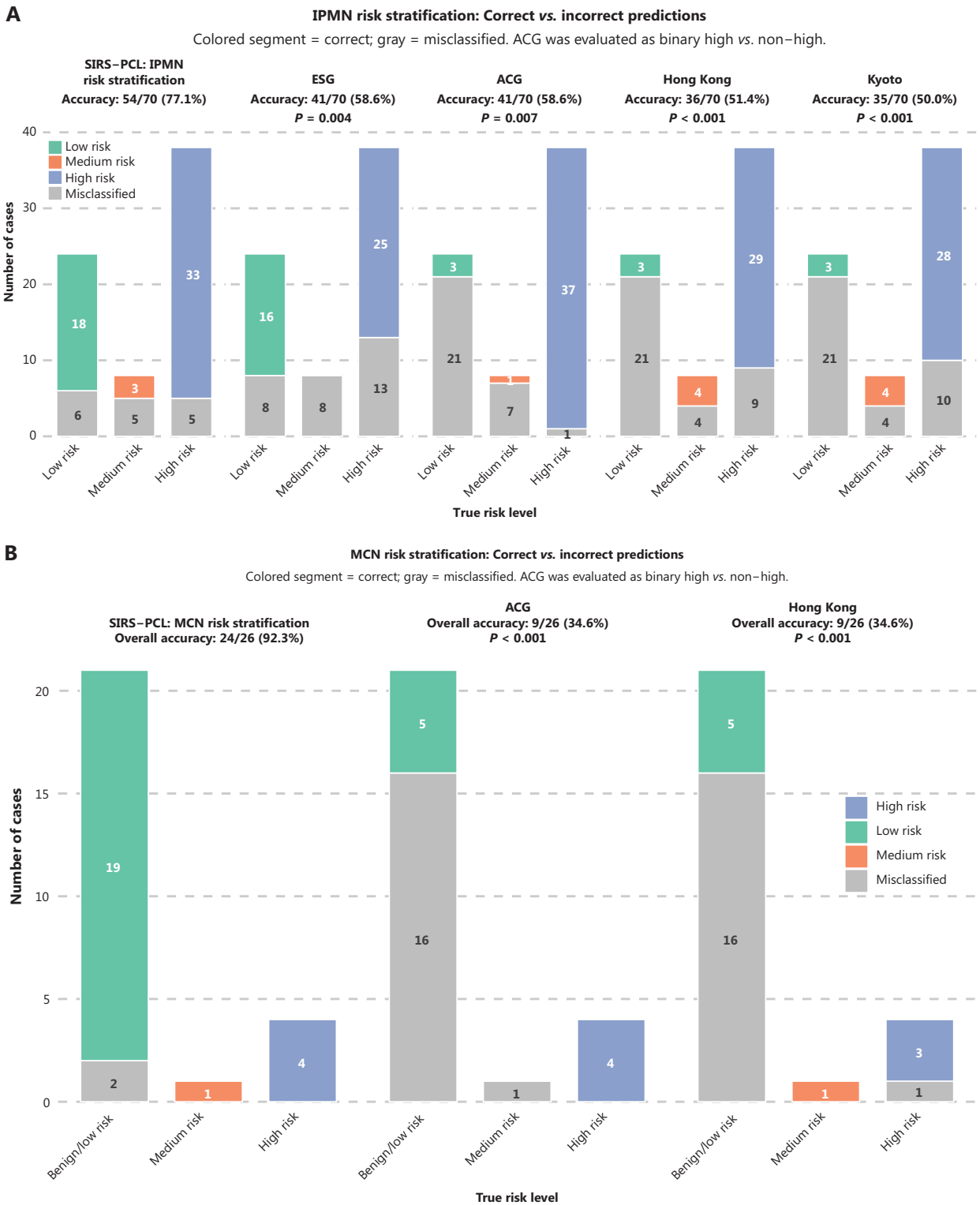


Figure 5 (A) Comparison of the proposed IPMN risk stratification model and guideline-based classification in complete-data temporal validation (n = 70). (B) Comparison of the proposed MCN risk stratification model and guideline-based classification in complete-data temporal validation (n = 26). ACG, American College of Gastroenterology; ESG, European Study Group; IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm.

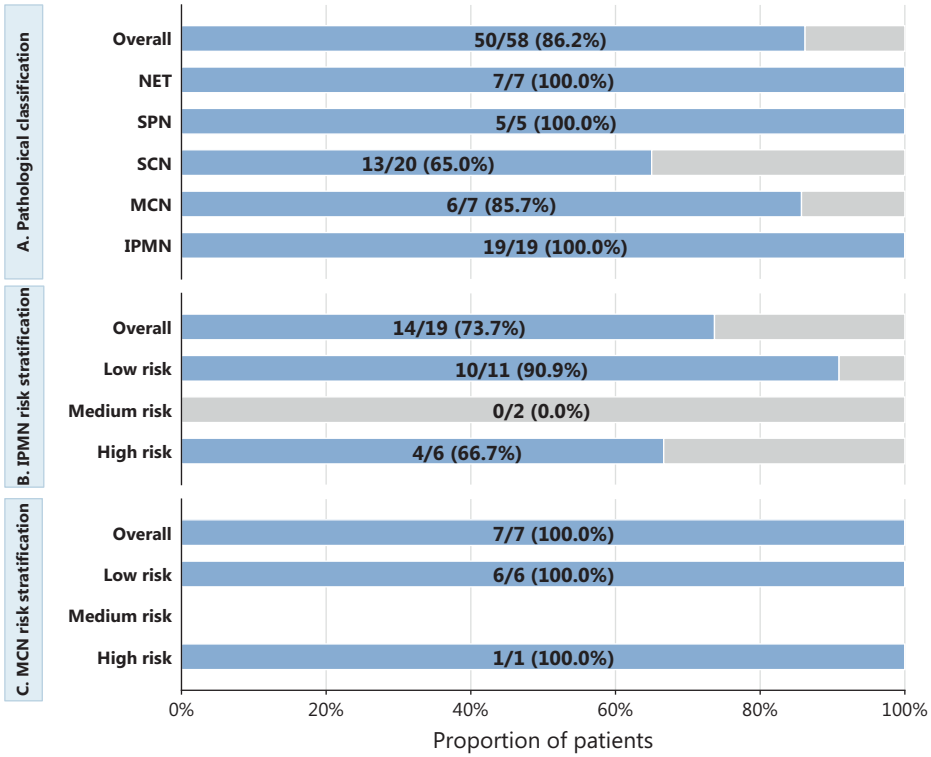


Figure 6 Proportions of correct classification and misclassification for (A) five-category pathological classification, (B) IPMN pathological risk stratification, and (C) MCN pathological risk stratification. Colored and gray segments indicate correct classification and misclassification, respectively; values are *n/N* (%). IPMN, intraductal papillary mucinous neoplasm; MCN, mucinous cystic neoplasm; NET, pancreatic neuroendocrine neoplasm; SCN, serous cystic neoplasm; SIRS-PCL, Structured Imaging Report System for Pancreatic Cystic Lesions; SPN, solid pseudopapillary neoplasm.

pathway and subsequently performs three-tier pathological risk stratification for IPMN and MCN, and thus extends prediction beyond a binary malignant *versus* nonmalignant framework. In the temporal validation cohort, the five-category classification model achieved a mean AUC of approximately 0.93, whereas the IPMN and MCN risk-stratification models achieved AUC values of approximately 0.86 and 0.97, respectively. This distinction is clinically important, as accurate subtype classification is the basis for subsequent pathological risk prediction, and errors at this stage may propagate throughout the entire decision-making process. The improvement in diagnostic performance observed in our study suggests that a structured imaging reporting framework may reduce variability in preoperative interpretation and provide a more standardized foundation for pathology-based prediction.

The second important finding is that SIRS-PCL demonstrated greater potential clinical utility than guideline-based imaging assessment according to our DCA results. Current international guidelines are primarily designed to maximize sensitivity and therefore inevitably sacrifice specificity in selected clinical scenarios, which results in unnecessary investigations or surgical interventions for a proportion of patients[7,11,12]. The individualized probability estimates generated by SIRS-PCL may facilitate more balanced preoperative pathological assessment by reducing unnecessary interventions while maintaining identification of high-risk lesions. Importantly, the threshold probabilities were derived from the DCA and should not be interpreted as recommended intervention thresholds. Instead, these findings suggest that pathology-based quantitative prediction may complement existing guideline-based assessment and improve individualized decision-making. Future studies incorporating additional clinical information, molecular biomarkers, radiomics, and AI-derived imaging signatures may further enhance the predictive performance and clinical applicability of SIRS-PCL[26].

The third important finding is that SIRS-PCL was developed using routinely available clinical variables and standardized imaging semantic features that can be obtained during routine CT or MRI interpretation without any requirement for additional invasive procedures, radiomics, or specialized molecular testing. Recent advances have demonstrated the potential of more technically intensive approaches. For example, Yuan et al. developed PCN-AI, a deep learning–based CT model, using 63 automatically quantified imaging features from 1,835

patients across four institutions. The model achieved AUC values of 0.845–0.988 across hierarchical classification tasks and improved the diagnostic performance of radiologists in a multireader study[27]. Notably, despite relying only on routinely interpretable semantic imaging features rather than automated quantitative image analysis, SIRS-PCL achieved favorable discrimination in the temporal validation cohort, with AUC values of approximately 0.93 for five-category pathological classification, 0.86 for IPMN risk stratification, and 0.97 for MCN risk stratification. This design may reduce the technical requirements for implementation while preserving a clinically interpretable and standardized assessment pathway.

Our study has several limitations. First, only surgically treated patients with histopathological confirmation were included, and surgical indications were determined according to routine multidisciplinary practice at each participating institution rather than a standard study protocol. Although pathological subtype classification and pathological risk categories were subsequently standardized according to the WHO and WGO criteria, differences in institutional surgical practice may have introduced selection bias. Consequently, the pathological subtype distribution should not be interpreted as reflecting the epidemiological prevalence of PCLs, which limits the generalizability of SIRS-PCL to surveillance and broader nonsurgical populations. Accordingly, SIRS-PCL should be regarded as a pathology-based preoperative decision-support system rather than a standalone indication for surgery, and its predictions should be interpreted in conjunction with multidisciplinary clinical assessment[5,14].

Second, temporal validation rather than institution-based external validation was adopted because surgically confirmed cases, particularly uncommon pathological subtypes, were unevenly distributed across participating centers. Although this strategy maintained a relatively balanced case composition for model development and evaluation, temporal validation cannot fully substitute for independent external validation and may also be influenced by temporal changes in imaging acquisition protocols and scanner technology. Thus, there is still a need for further prospective implementation cohorts across independent institutions and broader clinical settings. Finally, several uncommon pathological categories, particularly medium-risk MCN, were limited in number despite the relatively large overall multicenter cohort. Although stability-based feature selection was used to restrict model complexity, and five-fold cross-validation was applied to improve data utilization in the smaller MCN cohort, these approaches cannot compensate for sparse outcome information. Therefore, class-specific performance estimates, particularly for medium-risk MCN, and the positive predictive values for uncommon subtypes should be interpreted with caution. Larger independent prospective multicenter cohorts with adequate representation of all pathological categories are required to further establish the stability and generalizability of the model[3].

Conclusion

In conclusion, SIRS-PCL is a pathology-based structured imaging reporting system for preoperative subtype classification and risk stratification of surgically evaluated pancreatic cystic neoplasms. By combining standard clinical and imaging features in a quantitative format, SIRS-PCL supports objective preoperative assessment and complements multidisciplinary clinical judgment. Future work may further improve prediction and individualized management by incorporating radiomics, AI-derived imaging signatures, molecular biomarkers for cyst fluid, and longitudinal follow-up data. Prospective implementation validation in surveillance cohorts, together with temporal validation across independent institutions and broader clinical settings, will be essential to confirm the robustness, generalizability, and clinical usefulness of SIRS-PCL.

Acknowledgements

Dr Robert Sykes (www.doctored.org.uk) provided technical editing services for the final draft of this manuscript. Shanghai United Imaging Healthcare provided technical assistance for integrated development of the SIRS-PCL system.

Grant support

This work was supported by grants from the National Key Laboratory of Druggability Evaluation and Systematic Translational Medicine (24ZXZSSS00440) and the Independent Project Fund of the Key Laboratory of Drug Efficacy Evaluation and Systematic Translational Medicine, Tianjin Medical University Cancer Institute and Hospital (QZ23-11) to Jihui Hao; from the National Natural Science Foundation of China (82302180), Tianjin Key Laboratory of Digestive Cancer (XH24-3), and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-009A) to Xiaonan Cui; from the General Program of Tianjin Health Commission Science and Technology Project (TJWJ2022MS007) and Tianjin Science and Technology Major Special Projects and Programs (22ZXJBSY00030) to Chuntao Gao; and from the project fund of Tianjin Medical University Cancer Institute & Hospital (Y2505) to Tianye Liu.

Conflict of interest statement

The authors have no potential conflicts of interest to disclose.

Author contributions

Conceived and designed the study and developed the protocol: Jihui Hao, Xiaonan Cui, Houpu Liu.

Wrote the article: Xiaonan Cui, Houpu Liu.
Reviewed the methodology and contributed to the analysis: Houpu Liu, Sunyi Zheng, Jing Wang, Michael Ying, ZYL, Xiaoxu Wang, Geertruida H. de Bock.
Enrolled patients and acquired data and performed the radiological image assessment: Xiaonan Cui, Hongwei Wang, Jiaxin Liu, Ziyi Gao, Tianye Liu, Jing Liang, Xinru Zheng, Keyi Bian, Xiaomeng Yang, Chunlin Zhang, Tianshuo Jiao, Yufei Yang, Shuxuan Fan.
Contributed to data interpretation: Shuxuan Fan, Xiuchao Wang, Song Gao, Chuntao Gao, Yukuan Feng, Hui Li.
Contributed to the imaging evaluation: Guoping Xu, Xiang Zhang, Wen Shen, Haoran Sun, Baocun Sun, Yan Sun, Jianyu Xiao, Yalin Zheng, Tianye Liu.
Critically reviewed the study design and revised the manuscript: Jihui Hao, Zhaoxiang Ye.

Data availability statement

SIRS-PCL is publicly accessible at https://sirs-pcl.tjmucih.cn:16433/. For demonstration and system testing, we have provided a test account (Username: TEST; Password: Test1234@). The source code generated in this study is publicly available in the SIRS-PCL GitHub repository at https://github.com/MAGIC-LAB-TJMUCH/Development-of-SIRS-PCL.git.

Appendices

Appendix 1. Main variable definitions and schematic depicting the representative main imaging features of PCL.
Appendix 2. Data processing, feature selection, and study size.
Appendix 3. Demographic tables for the three models.
Appendix 4. Development and temporal validation of the pathological classification model .
Appendix 5. Development and temporal validation of the IPMN risk stratification model.
Appendix 6. Development and temporal validation of the MCN risk stratification model.
Appendix 7. Subgroup analysis conducted in full datasets without MICE.
Appendix 8. Clinical application of SIRS-PCL: prospective data management and case demonstrations of pathological five-category classification and risk stratification.

References

1. Pozzi-Mucelli RM, Rinta-Kiikka I, Wünsche K, Laukkarinen J, Labori KJ, Ånonsen K, et al. Pancreatic MRI for the surveillance of cystic neoplasms: comparison of a short with a comprehensive imaging protocol. Eur Rad. 2017; 27: 41-50.
2. Van Huijgevoort NC, Del Chiaro M, Wolfgang CL, van Hooft JE, Besselink MG. Diagnosis and management of pancreatic cystic neoplasms: current evidence and guidelines. Nat Rev Gastroenterol Hepatol. 2019; 16: 676-89.
3. Gonda TA, Cahen DL, Farrell JJ. Pancreatic cysts. N Engl J Med. 2024; 391: 832-43.
4. Gardner TB, Park WG, Allen PJ. Diagnosis and management of pancreatic cysts. Gastroenterology. 2024; 167: 454-68.
5. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO classification of tumours of the digestive system. Histopathology. 2020; 76: 182-8.
6. Tanner S, Dutta P, Quingalahua E, Chalhoub JM, Hawa F, Facciorusso A, et al. High burden of pancreatic surgery for intraductal papillary mucinous neoplasia with low-grade dysplasia and benign cysts: a systematic review and meta-analysis. United European Gastroenterol J. 2026; 14: e70145.
7. Cheung TT, Lee YT, Tang RSY, She WH, Cheng KC, Cheung CC, et al. The Hong Kong consensus recommendations on the diagnosis and management of pancreatic cystic lesions. Hepatobiliary Surg Nutr. 2023; 12: 715-35.
8. European Study Group on Cystic Tumours of the Pancreas. European evidence-based guidelines on pancreatic cystic neoplasms. Gut. 2018; 67: 789-804.
9. Farrell JJ, Fernández-del Castillo C. Pancreatic cystic neoplasms: management and unanswered questions. Gastroenterology. 2013; 144: 1303-15.
10. Das A, Ngamruengphong S, Nagendra S, Chak A. Asymptomatic pancreatic cystic neoplasm: a cost-effectiveness analysis of different strategies of management. Gastrointest Endos. 2009; 70: 690-9.e6.
11. Elta GH, Enestvedt BK, Sauer BG, Lennon AM. ACG Clinical Guideline: diagnosis and management of pancreatic cysts. Am J Gastroenterol. 2018; 113: 464-79.
12. Ohtsuka T, Fernandez-Del Castillo C, Furukawa T, Hijioka S, Jang JY, Lennon AM, et al. International evidence-based Kyoto guidelines for the management of intraductal papillary mucinous neoplasm of the pancreas. Pancreatology. 2024; 24: 255-70.
13. Zhao W, Liu S, Cong L, Zhao Y. Imaging features for predicting high-grade dysplasia or malignancy in branch duct type intraductal papillary mucinous neoplasm of the pancreas: a systematic review and meta-analysis. Ann Surg Oncol. 2022; 29: 1297-312.
14. Malagelada J, Guda N, Goh K, Hackert T, Layer P, Molero X, et al. Pancreatic cystic lesions. WGO practice guideline. Milwaukee, WI: World Gastroenterology Organization; 2019.
15. Long Q, Johnson BA. Variable selection in the presence of missing data: resampling and imputation. Biostatistics. 2015; 16: 596-610.
16. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Ser B: Stat Methodol. 2010; 72: 417-73.
17. Vach W, Gerke O. Gwet's AC1 is not a substitute for Cohen's kappa – a comparison of basic properties. MethodsX. 2023; 10: 102212.
18. Kusunoki T, Matsuoka J, Ohtsu H, Kagimura T, Nakamura H. Relationship between intraclass and concordance correlation coefficients: similarities and differences. Jpn J Biometrics. 2009; 30: 35-53.
19. Heymans MW, Twisk JWR. Handling missing data in clinical research. J Clin Epidemiol. 2022; 151: 185-8.

20. Wolbers M, Noci A, Delmar P, Gower-Page C, Yiu S, Bartlett JW. Standard and reference-based conditional mean imputation. Pharm Stat. 2022; 21: 1246-57.
21. Nakagawa S, De Villemereuil P. A general method for simultaneously accounting for phylogenetic and species sampling uncertainty via Rubin's rules in comparative analysis. Syst Biol. 2019; 68: 632-41.
22. Pate A, Riley RD, Collins GS, van Smeden M, Van Calster B, Ensor J, et al. Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. Stat Methods Med Res. 2023; 32: 555-71.
23. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II – binary and time-to-event outcomes. Stat Med. 2019; 38: 1276-96.
24. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems 30* (Proceedings of the 31st International Conference on Neural Information Processing Systems); 2017 Dec 4–9; Long Beach, CA, USA. Red Hook, NY: Curran Associates; 2017; p. 4765-74.
25. Deng W, Liu J, Wang X, Xie F, Wang S, Zhang X, et al. Should all pancreatic cystic lesions with worrisome or high-risk features be resected? A clinical and radiological machine learning model may help to answer. Acad Radiol. 2024; 31: 1889-97.
26. He Y, Guo J, Gao P, Li Z, Wang W, Chen K. Advances in plasma metabolomics detection technology and its clinical applications in lung cancer and other malignancies. Holist Integr Oncol. 2026; 5: 33.
27. Yuan X, Chen C, Shi Z, Liu W, Zhang X, Yang M, et al. Deep learning CT model for stratified diagnosis of pancreatic cystic neoplasms: multicenter development, validation, and real-world clinical impact. NPJ Digit Med. 2025; 8: 609.

Cite this article as: Cui X, Liu H, Zheng S, Wang H, Liu J, Gao Z, et al. An integrated CT/MRI diagnostic system for pancreatic cystic lesions that supports pathological classification and risk stratification. Cancer Biol Med. 2026; 23: 1495-1512. doi: 10.20892/j.issn.2095-3941.2026.0566