

Review

# Effectiveness of Radiomics-Based Machine Learning Models in Differentiating Pancreatitis and Pancreatic Ductal Adenocarcinoma: Systematic Review and Meta-Analysis

Lechang Zhang<sup>1</sup>, MM; Dewei Li<sup>2</sup>, MM; Tong Su<sup>1</sup>, MM; Tong Xiao<sup>1</sup>, MM; Shulei Zhao<sup>1</sup>, MD

<sup>1</sup>Department of Gastroenterology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

<sup>2</sup>Department of Infectious Diseases, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

## Corresponding Author:

Shulei Zhao, MD

Department of Gastroenterology

Shandong Provincial Hospital Affiliated to Shandong First Medical University

324 Jingwu Weiqi Rd

Jinan 250021

China

Phone: 86 13853121769

Email: [wenzhu24@126.com](mailto:wenzhu24@126.com)

## Abstract

**Background:** Pancreatic ductal adenocarcinoma (PDAC) and mass-forming pancreatitis (MFP) share similar clinical, laboratory, and imaging features, making accurate diagnosis challenging. Nevertheless, PDAC is highly malignant with a poor prognosis, whereas MFP is an inflammatory condition typically responding well to medical or interventional therapies. Some investigators have explored radiomics-based machine learning (ML) models for distinguishing PDAC from MFP. However, systematic evidence supporting the feasibility of these models is insufficient, presenting a notable challenge for clinical application.

**Objective:** This study intended to review the diagnostic performance of radiomics-based ML models in differentiating PDAC from MFP, summarize the methodological quality of the included studies, and provide evidence-based guidance for optimizing radiomics-based ML models and advancing their clinical use.

**Methods:** PubMed, Embase, Cochrane, and Web of Science were searched for relevant studies up to June 29, 2024. Eligible studies comprised English cohort, case-control, or cross-sectional designs that applied fully developed radiomics-based ML models—including traditional and deep radiomics—to differentiate PDAC from MFP, while also reporting their diagnostic performance. Studies without full text, limited to image segmentation, or insufficient outcome metrics were excluded. Methodological quality was appraised by means of the radiomics quality score. Since the limited applicability of QUADAS-2 in radiomics-based ML studies, the risk of bias was not formally assessed. Pooled sensitivity, specificity, area under the curve of summary receiver operating characteristics (SROC), likelihood ratios, and diagnostic odds ratio were estimated through a bivariate mixed-effects model. Results were presented with forest plots, SROC curves, and Fagan's nomogram. Subgroup analysis was performed to appraise the diagnostic performance of radiomics-based ML models across various imaging modalities, including computed tomography (CT), magnetic resonance imaging, positron emission tomography-CT, and endoscopic ultrasound.

**Results:** This meta-analysis included 24 studies with 14,406 cases, including 7635 PDAC cases. All studies adopted a case-control design, with 5 conducted across multiple centers. Most studies used CT as the primary imaging modality. The radiomics quality score scores ranged from 5 points (14%) to 17 points (47%), with an average score of 9 (25%). The radiomics-based ML models demonstrated high diagnostic performance. Based on the independent validation sets, the pooled sensitivity, specificity, area under the curve of SROC, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio were 0.92 (95% CI 0.91-0.94), 0.90 (95% CI 0.85-0.94), 0.94 (95% CI 0.74-0.99), 9.3 (95% CI 6.0-14.2), 0.08 (95% CI 0.07-0.11), and 110 (95% CI 62-194), respectively.

**Conclusions:** Radiomics-based ML models demonstrate high diagnostic accuracy in differentiating PDAC from MFP, underscoring their potential as noninvasive tools for clinical decision-making. Nonetheless, the overall methodological quality was moderate due to limitations in external validation, standardized protocols, and reproducibility. These findings support the

promise of radiomics in clinical diagnostics while highlighting the need for more rigorous, multicenter research to enhance model generalizability and clinical applicability.

**Trial Registration:** PROSPERO CRD42024575745; [https://www.crd.york.ac.uk/prospero/display\\_record.php?ID=CRD42024575745](https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42024575745)

*J Med Internet Res* 2025;27:e72420; doi: [10.2196/72420](https://doi.org/10.2196/72420)

**Keywords:** artificial intelligence; differential diagnosis; imaging modalities; systematic review; meta-analysis

## Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common histological subtype of pancreatic cancer and represents the deadliest solid malignancy, with a 5-year survival rate below 10%. It currently ranks as the third leading cause of cancer-related mortality worldwide, posing a substantial public health burden [1,2]. Surgical resection remains the only potentially curative treatment option. However, due to its insidious onset and lack of specific early clinical signs, most patients are diagnosed at an advanced stage when curative surgery is no longer feasible [3]. Notably, early-stage or incidentally detected PDAC is associated with markedly improved survival outcomes [4]. These observations highlight the urgent need for effective diagnostic strategies for PDAC [5].

Mass-forming pancreatitis (MFP), a focal pancreatic lesion that may arise from chronic pancreatitis or manifest as autoimmune pancreatitis, poses a diagnostic challenge, especially when it mimics the clinical features of PDAC [6,7]. The 2 conditions often share overlapping clinical manifestations, laboratory profiles, and imaging characteristics, thus complicating accurate differentiation in routine practice [8,9]. However, their prognoses and treatment approaches differ significantly [10]. PDAC is characterized by aggressive biological behavior, including neural invasion and hematogenous spread, often resulting in a poor prognosis. In contrast, MFP is a benign inflammatory condition, with approximately 80% of cases responding well to medical or interventional therapy. Misdiagnosis may therefore lead to either unnecessary surgery for benign MFP or delayed intervention for malignant PDAC, both carrying significant clinical consequences.

Traditionally, endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA) has served as a cornerstone for diagnosing pancreatic lesions by providing cytopathologic confirmation. Although generally safe and effective [11,12], EUS-FNA may yield inconclusive results if the samples are insufficient, particularly in the absence of rapid on-site evaluation [13,14]. In addition, tumor heterogeneity and sampling error can contribute to false-negative results [15-17]. As an invasive procedure, EUS-FNA also carries inherent procedural risks and limitations. Consequently, there

is growing interest in developing noninvasive, accurate diagnostic alternatives.

Radiomics has emerged as a promising noninvasive approach that extracts high-throughput quantitative imaging features, often imperceptible to the human eye, from standard-of-care radiological scans [18]. These high-dimensional data can be processed by machine learning (ML) algorithms to develop predictive models for disease classification and characterization. Compared with conventional imaging interpretation, which is often subjective and experience-dependent, radiomics offers an objective, reproducible method for capturing intratumoral heterogeneity [19]. Its capacity to capture subtle imaging patterns also positions it as a potential complement or alternative to invasive diagnostic techniques such as EUS-FNA. Recent studies have explored the application of radiomics-based ML models for differentiating PDAC from MFP. Despite the growing number of individual investigations, systematic evidence supporting the diagnostic performance of radiomics-based ML models for detecting PDAC remains lacking. This knowledge gap presents a challenge for the development of intelligent diagnostic tools. Therefore, this study aims to systematically review the diagnostic accuracy of radiomics-based ML models in distinguishing PDAC from MFP. By synthesizing current evidence, our findings may clarify the potential clinical utility of radiomics and provide an evidence-based foundation for the development of intelligent diagnostic technologies in pancreatic diseases.

## Methods

### Overview

This study was executed and documented in alignment with the guidelines set by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Checklist 1) [20]. Furthermore, this study followed an a priori-created protocol registered on the International Prospective Register of Systematic Reviews (CRD42024575745).

### Eligibility Criteria

The eligibility criteria are described in Textbox 1.

#### Textbox 1. Inclusion and exclusion criteria.

##### Inclusion criteria:

- Studies involving patients diagnosed with pancreatic ductal adenocarcinoma or mass-forming pancreatitis, including chronic pancreatitis and a focal type of autoimmune pancreatitis.
- Studies that fully developed and applied radiomics-based machine learning models—whether based on handcrafted features (traditional radiomics) or deep learning feature extraction (deep radiomics)—for differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis.
- Studies published in English.
- Studies with cohort, case-control, and cross-sectional designs.

#### Exclusion criteria:

- Conference abstracts not published in full.
- Studies that only performed image segmentation.
- Studies without any outcome measures to appraise the prediction accuracy of models, including c-index, sensitivity, specificity, accuracy, precision, confusion matrix,  $F_1$ -score, or calibration curves.

## Data Sources and Search Strategy

A comprehensive retrieval was carried out across PubMed, Embase, Cochrane, and Web of Science from their inception to June 29, 2024. The search strategy incorporated both Medical Subject Headings terms and keywords ([Multimedia Appendix 1](#)) to ensure a thorough identification of relevant studies.

## Study Selection and Data Extraction

The collected study reports were added to an EndNote library (EndNote 20.6 [Clarivate]), where duplicates were identified and removed. To recognize relevant studies, 2 reviewers (LZ and TS) independently filtered out the titles and abstracts. The entire texts of these articles were then reviewed to assess eligibility according to the pre-established criteria. A high level of agreement was achieved between the 2 reviewers (Cohen  $\kappa$  coefficient=0.89). Discrepancies, if any, were resolved through discussion or adjudication by a third reviewer (SZ).

The extracted data encompassed the title, first author's name, publication year, country, study design, patient source, source of radiomics, imaging protocol, number of imaging investigators, repeat measurement information, region of interest segmentation software, total number of PDAC cases, total number of cases, training set details (including number of PDAC cases and total number of cases), validation set details (including validation strategy, number of PDAC cases, and total number of cases), variable selection methods, model types, radiomics score construction, overfitting assessment, availability of code and data, and model evaluation metrics.

Two reviewers (LZ and TS) independently extracted the data and cross-checked the results for consistency. Any discrepancies were addressed through discussion. A third reviewer (SZ) would make the final decision if no consensus was reached.

## Assessment of Study Quality

Using the radiomics quality score (RQS), 2 investigators (DL and TX) independently appraised the methodological quality and risk of bias of the included studies [21]. The RQS assessed 16 dimensions across 6 domains, namely image and segmentation, feature selection, validation and utility, model performance, high-level evidence, and open science and data,

to appraise the methodological rigor of the construction of radiomics-based ML models. The total score, varying from  $-8$  to  $36$ , was derived from these 16 dimensions. Upon completion, 2 investigators (DL and TX) cross-checked their evaluations. A high level of agreement was observed between them (intraclass correlation coefficient=0.97; 95% CI 0.93–0.99). Any discrepancies were resolved through discussion or consultation with a third investigator (SZ).

## Synthesis Methods

The meta-analysis was conducted for diagnostic performance metrics, including sensitivity, specificity, area under the curve of summary receiver operating characteristic (SROC AUC), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR), using a bivariate mixed-effects model. This statistical approach considers both within-study variability (random effects) and between-study heterogeneity (fixed effects). It is particularly suitable for meta-analyses on the accuracy of diagnostic tests, as it allows for the simultaneous estimation of sensitivity and specificity, while addressing the inherent link between multiple measures. Unlike a hierarchical SROC model, a bivariate model directly yields pooled estimates of sensitivity and specificity with CIs, better aligning with our goal of producing clinically interpretable summary measures. By employing this model, reliable estimates of diagnostic performance metrics were obtained, providing a robust basis for evaluating radiomics-based ML models [22].

During this meta-analysis, sensitivity and specificity were analyzed using the diagnostic fourfold table. Nevertheless, the diagnostic fourfold table was not provided in most of the original studies. In such cases, we derived it from sensitivity, specificity, precision, and case numbers. Although only a minority of studies explicitly reported the probability threshold used for classification (like 0.5 or thresholds based on the Youden index), most provided diagnostic performance metrics. This suggested that some form of thresholding is applied, even if it was described in detail. To address this interstudy variability in threshold selection, we used a bivariate mixed-effects model, which inherently accounts for such differences and ensures the robustness of the pooled estimates.

The included studies used ML classifiers and reported results from various sets, including training, independent

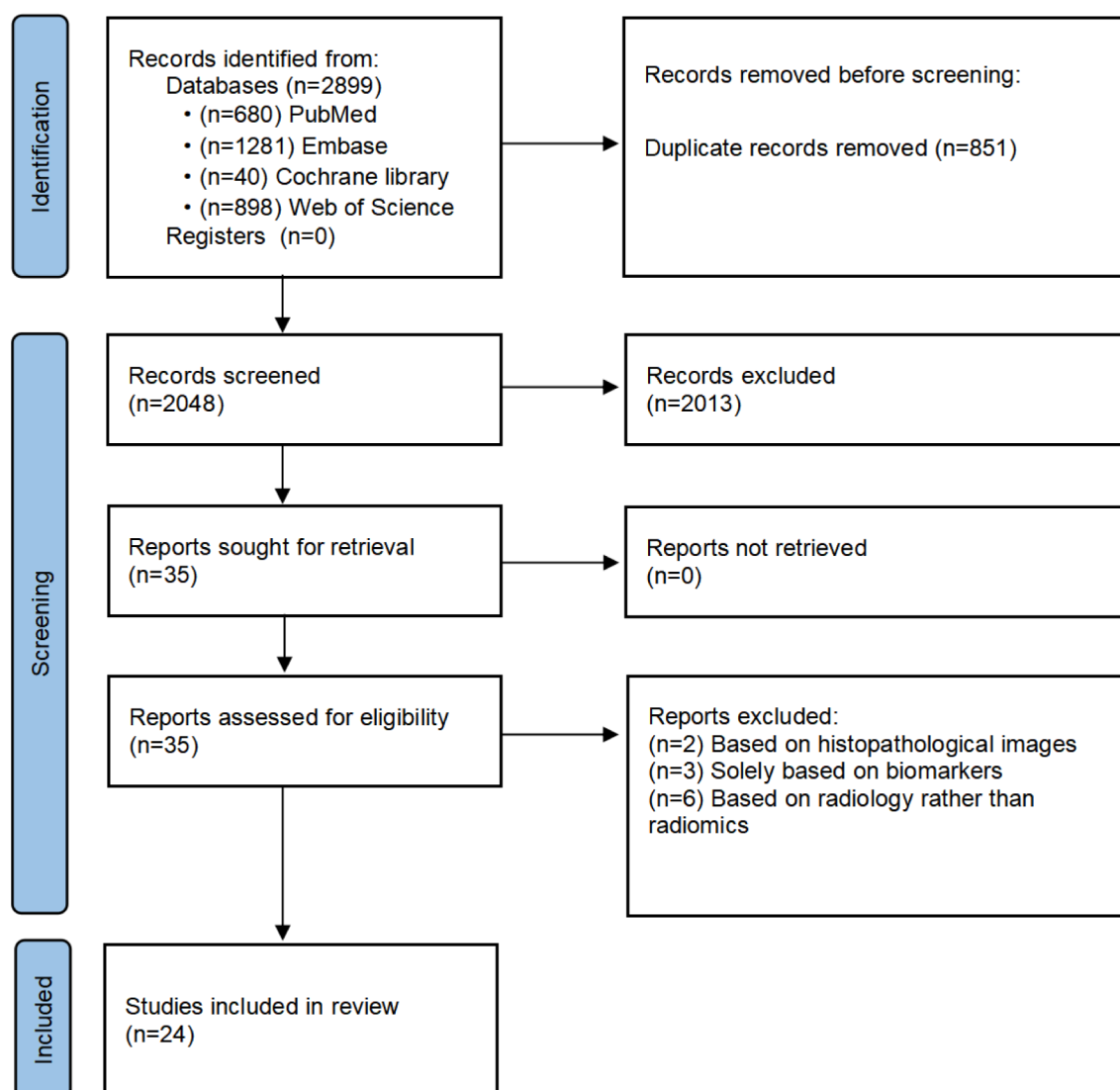
validation, and cross-validation sets. Therefore, these sets were analyzed separately. To further examine the predictive performance of radiomics-based ML models, subgroup analyses were executed based on various imaging modalities, including computed tomography (CT), endoscopic ultrasound (EUS), positron emission tomography-computed tomography (PET-CT), and magnetic resonance imaging (MRI). Each subgroup was examined separately using a bivariate mixed-effects model to obtain pooled sensitivity and specificity estimates. Heterogeneity among the included studies was appraised through the Cochrane Q-test ( $P$  value  $\leq .05$ ) and  $I^2$  statistic ( $>50\%$ ). Publication bias was examined using a funnel plot. 95% CIs were provided for all estimates. The meta-analysis was executed using Stata 16 software (Stata Corporation).

## Results

### Literature Search

Initially, 2899 articles were identified. Of the total, 851 articles were deleted due to duplicates. From the 2048 remaining articles, 2013 were removed after screening the titles and abstracts. Upon reviewing the entire text, 11 articles were removed (2 studies based on histopathological images, 3 focusing solely on biomarkers, and 6 relying on radiology rather than radiomics). Ultimately, 24 studies were included in this systematic review [23-46]. The literature screening process is presented in Figure 1.

**Figure 1.** Literature screening process.



### Study Characteristics

The 24 included studies were published between 2019 and 2024. In total, 14,406 participants were reported, with 7635 diagnosed as PDAC (Tables 1 and 2). All studies used a case-control design, with 5 conducted across multiple centers

[26,31,35,40,42] and the remaining at single centers. CT was the most frequently used imaging modality, featured in 13 studies [23,25,26,28-32,34,36-38,42]. EUS was used in 5 studies [40,43-46], MRI in 3 studies [27,35,41], and PET-CT in 3 studies [24,33,39].

**Table 1.** Study characteristics.

Author	Year	Country	Study design	Study setting	Imaging modality	ROI <sup>a</sup> segmenters, n	ROI segmentation software
Ren et al [23]	2020	China	Case-control	Single-center	CT <sup>b</sup>	— <sup>c</sup>	ITK-SNAP
Zhang et al [24]	2019	China	Case-control	Single-center	PET <sup>d</sup> -CT	6	3D Slicer
Zhang et al [25]	2022	China	Case-control	Single-center	CT	2	3D Slicer
Ye et al [26]	2023	China	Case-control	Multicenter	CT	2	3D Slicer
Shiraishi et al [27]	2022	Japan	Case-control	Single-center	MRI <sup>e</sup>	2	syngo
Ren et al [28]	2019	China	Case-control	Single-center	CT	3	ITK-SNAP
Qu et al [29]	2023	China	Case-control	Single-center	CT	2	3D Slicer
Park et al [30]	2020	United States	Case-control	Single-center	CT	7	Velocity AI
Ma et al [31]	2022	China	Case-control	Multicenter	CT	2	MITK
Lu et al [32]	2023	China	Case-control	Single-center	CT	2	LIFEx
Liu et al [33]	2021	China	Case-control	Single-center	PET-CT	2	3D Slicer
Li et al [34]	2022	China	Case-control	Single-center	CT	2	3D Slicer
Deng et al [35]	2021	China	Case-control	Multicenter	MRI	2	IBEX
E et al [36]	2020	China	Case-control	Single-center	CT	2	Weasis
Anai et al [37]	2022	Japan	Case-control	Single-center	CT	4	LIFEx
Ziegelmayr et al [38]	2020	Germany	Case-control	Single-center	CT	2	ITK-SNAP
Wei et al [39]	2023	China	Case-control	Single-center	PET-CT	3	3D Slicer
Tong et al [40]	2022	China	Case-control	Multicenter	EUS <sup>f</sup>	5	labelme
Chen et al [41]	2024	China	Case-control	Single-center	MRI	2	3D Slicer
Cao et al [42]	2023	China	Case-control	Multicenter	CT	48	—
Udriștoiu et al [43]	2021	Romania	Case-control	Single-center	EUS	—	—
Nakamura et al [44]	2024	Japan	Case-control	Single-center	EUS	7	—
Marya et al [45]	2021	United States	Case-control	Single-center	EUS	7	—
Kuwahara et al [46]	2023	Japan	Case-control	Single-center	EUS	—	—

<sup>a</sup>ROI: region of interest.<sup>b</sup>CT: computed tomography.<sup>c</sup>Not available.<sup>d</sup>PET: positron emission tomography.<sup>e</sup>MRI: magnetic resonance imaging.<sup>f</sup>EUS: endoscopic ultrasound.**Table 2.** Model details.

Author	Number of PDAC <sup>i</sup> cases	Number of cases	Number of PDAC cases in the training set	Number of cases in the training set	Validation strategy	Number of PDAC cases in the validation set	Number of cases in the validation set	Model type	Model calibration quality
Ren et al [23]	79	109	79	109	Cross-validation	79	109	RF <sup>b</sup>	— <sup>c</sup>
Zhang et al [24]	66	111	66	111	Cross-validation	66	111	RF, Adaboost <sup>d</sup> , and SVM <sup>e</sup>	—
Zhang et al [25]	71	138	59	103	Prospective external validation	12	35	LR <sup>f</sup>	Decision curve
Ye et al [26]	120	198	84	139	Prospective external validation	36	59	LR	Calibration curve and decision curve
Shiraishi et al [27]	77	105	77	105	Cross-validation	77	105	SVM	—



Author	Number of PDAC <sup>i</sup> cases	Number of cases	Number of PDAC cases in the training set	Number of cases in the training set	Validation strategy	Number of PDAC cases in the validation set	Number of cases in the validation set	Model type	Model calibration quality
Ren et al [28]	79	109	79	109	Random sampling	30	40	LR	—
Qu et al [29]	201	255	175	213	Random sampling	26	42	LR	Decision curve
Park et al [30]	93	182	60	120	Random sampling	33	62	RF	—
Ma et al [31]	151	175	151	175	Cross-validation	151	175	LR	Calibration curve and decision curve
Lu et al [32]	64	96	45	67	Random sampling	19	29	LR, RF, SVM, and DT <sup>g</sup>	Calibration curve and decision curve
Liu et al [33]	64	112	64	112	Cross-validation	64	112	SVM	—
Li et al [34]	42	97	42	97	Cross-validation	42	97	LR	—
Deng et al [35]	96	119	51	64	Institutional external validation	45	55	SVM	—
E et al [36]	51	96	51	96	Cross-validation	51	96	RF	—
Anai et al [37]	30	50	30	50	Cross-validation	30	50	SVM	—
Ziegelmayr et al [38]	42	86	42	86	Cross-validation	42	86	RF	—
Wei et al [39]	64	112	64	112	Cross-validation	64	112	DL <sup>h</sup>	—
Tong et al [40]	414	558	264	351	External validation	150	207	DL	—
Chen et al [41]	62	93	55	82	Random sampling	7	11	DL	—
Cao et al [42]	4706	9939	1431	3208	Random sampling and external validation	3275	6731	DL	—
Udriștoiu et al [43]	30	65	30	65	Cross-validation	30	65	DL	—
Nakamura et al [44]	61	85	61	85	Cross-validation	61	85	DL	—
Marya et al [45]	292	583	170	336	Random sampling	122	247	DL	—
Kuwahara et al [46]	680	933	518	694	Random sampling	162	239	DL	—

<sup>a</sup>PDAC: pancreatic ductal adenocarcinoma.

<sup>b</sup>RF: random forest.

<sup>c</sup>Not available.

<sup>d</sup>Adaboost: adaptive boosting.

<sup>e</sup>SVM: support vector machine.

<sup>f</sup>LR: logistic regression.

<sup>g</sup>DT: decision tree.

<sup>h</sup>DL: deep learning.

Regarding the software used for region of interest segmentation, the majority of studies [24,26,29,33,34,39,41] used 3D Slicer (Brigham and Women's Hospital). ITK-SNAP (University of Pennsylvania) was used in 3 studies [23,28,38], while 2 studies [32,37] used LIFEx (Institut Curie). Other software tools mentioned included syngo (Siemens Healthineers) [27], Velocity AI (Varian Medical Systems) [30],

MITK (German Cancer Research Center) [31], IBEX (MD Anderson Cancer Center) [35], Weasis (open source) [36], and labelme [MIT CSAIL] [40]. Most studies reported the methods used for variable selection, with least absolute shrinkage and selection operator regression and minimum redundancy maximum relevance being the most commonly employed techniques. In addition, the most frequently applied

ML for constructing radiomics-based models was logistic regression [25,26,28,29,31,32,34], decision tree [32], support vector machine (SVM) [24,32,33,35,37], adaptive boosting [24], and random forest (RF) [23,24,30,32,36,38]. Some studies also used deep learning (DL) [39-46] models. Only 4 studies [28,31,33,35] constructed models combining clinical features with radiomics features.

All studies described their validation methods. A total of 19 studies used internal validation [23,24,27-34,36-39,41,43-46], of which 7 used random sampling [28-30,32,41,45,46], while 12 used internal cross-validation [23,24,27,31,33,34,36-39,43,44]. Furthermore, 4 studies applied prospective or external institutional validation [25,26,35,40], and 1 incorporated both random sampling for internal validation and external institutional validation [42]. The training set from all included studies comprised a total of 6689 cases, with 3748 being PDAC cases. The validation set, on the other hand, contained 8960 cases, of which 4674 were PDAC cases.

### Assessment of Study Quality

The RQS was applied to assess the 24 studies, varying from 5 points (14%) to 17 points (47%), with an average score of 9 (25%) ([Multimedia Appendix 2](#)).

The primary factors contributing to the lower quality included no phantom models across scanners to evaluate interscanner variability, no imaging data at multiple time points to assess temporal stability, and a limited correlation between radiomic features and biology. In addition, few studies reported on model calibration or conducted cost-effectiveness analyses. Most studies were retrospective

in nature, with code and data generally not accessible to the public, limiting reproducibility.

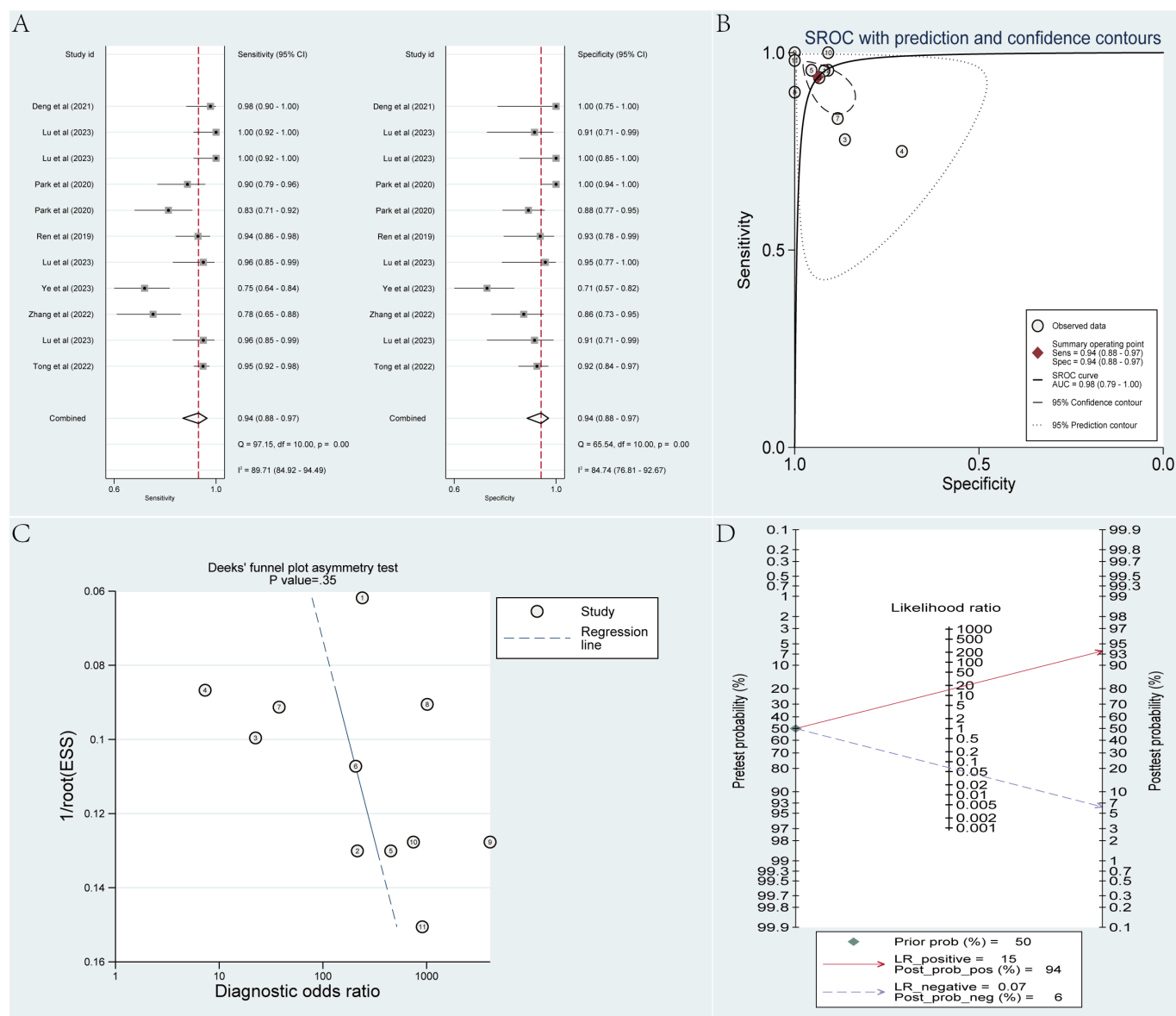
Conversely, the included studies demonstrated strengths in other areas. Many studies provided well-documented imaging protocols, and a substantial number of studies used multiple segmentation methods to enhance the robustness of segmentation. Statistical techniques, such as feature reduction to avoid overfitting and discrimination statistics, including SROC curves and AUC, were widely used across these studies. Furthermore, several studies validated their models using independent datasets, which supported the clinical applicability of their radiomics-based models.

### Meta-Analysis

#### Training Set

The training set included 11 radiomics-based ML models. By using a bivariate mixed-effects model, the pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR were calculated as 0.94 (95% CI 0.88-0.97), 0.94 (95% CI 0.88-0.97), 0.98 (95% CI 0.79-1.00), 14.8 (95% CI 7.3-29.8), 0.07 (95% CI 0.03-0.13), and 225 (95% CI 63-807), respectively. According to the funnel plot, no noticeable publication bias was observed across the included studies. In our analysis, PDAC accounted for approximately 50% (7635/14,406) of the cases. Taking 50% as the previous probability for the disease, given a PLR of 15, a positive result (PDAC) predicted by the models would correspond to a 94% probability of true PDAC. With an NLR of 0.07, a negative result (MFP) predicted by the models would imply a 94% probability of true MFP ([Figure 2](#)).

**Figure 2.** Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the training set. Data were derived from studies [25,26,28,30,32,35,40]. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram. AUC: area under the curve; ESS: effective sample size; LR: logistic regression; Prob: probability; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristic.



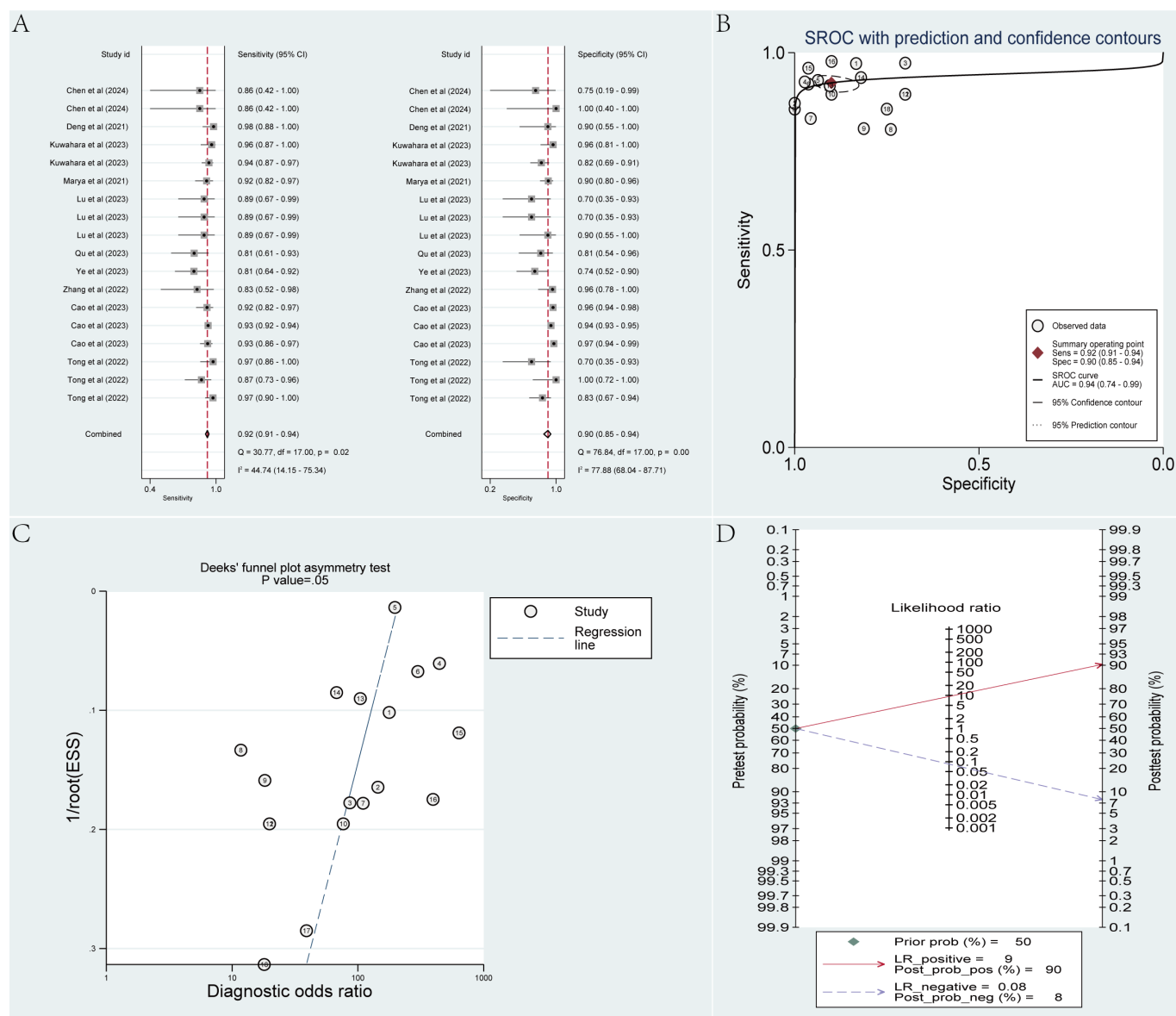
## Independent Validation Set

The independent validation set included 18 radiomics-based ML models. The pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR, derived from a bivariate mixed-effects model, were 0.92 (95% CI 0.91-0.94), 0.90 (95% CI 0.85-0.94), 0.94 (95% CI 0.74-0.99), 9.3 (95% CI 6.0-14.2), 0.08 (95% CI 0.07-0.11), and 110 (95% CI 62-194), respectively. According to the funnel plot, no noticeable publication bias

was noticed among the included studies. In our study, PDAC accounted for approximately 50% (7635/14,406) of the cases. Taking 50% as the previous probability for the disease, given a PLR of 9, a positive result (PDAC) predicted by the models would imply a 90% probability of true PDAC. With an NLR of 0.08, a negative result (MFP) predicted by the models would correspond to a 92% probability of true MFP (Figure 3).



**Figure 3.** Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the independent validation set. Data were derived from studies [25,26,29,32,35,40-42,45,46]. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram. AUC: area under the curve; ESS: effective sample size; LR: logistic regression; Prob: probability; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristic.

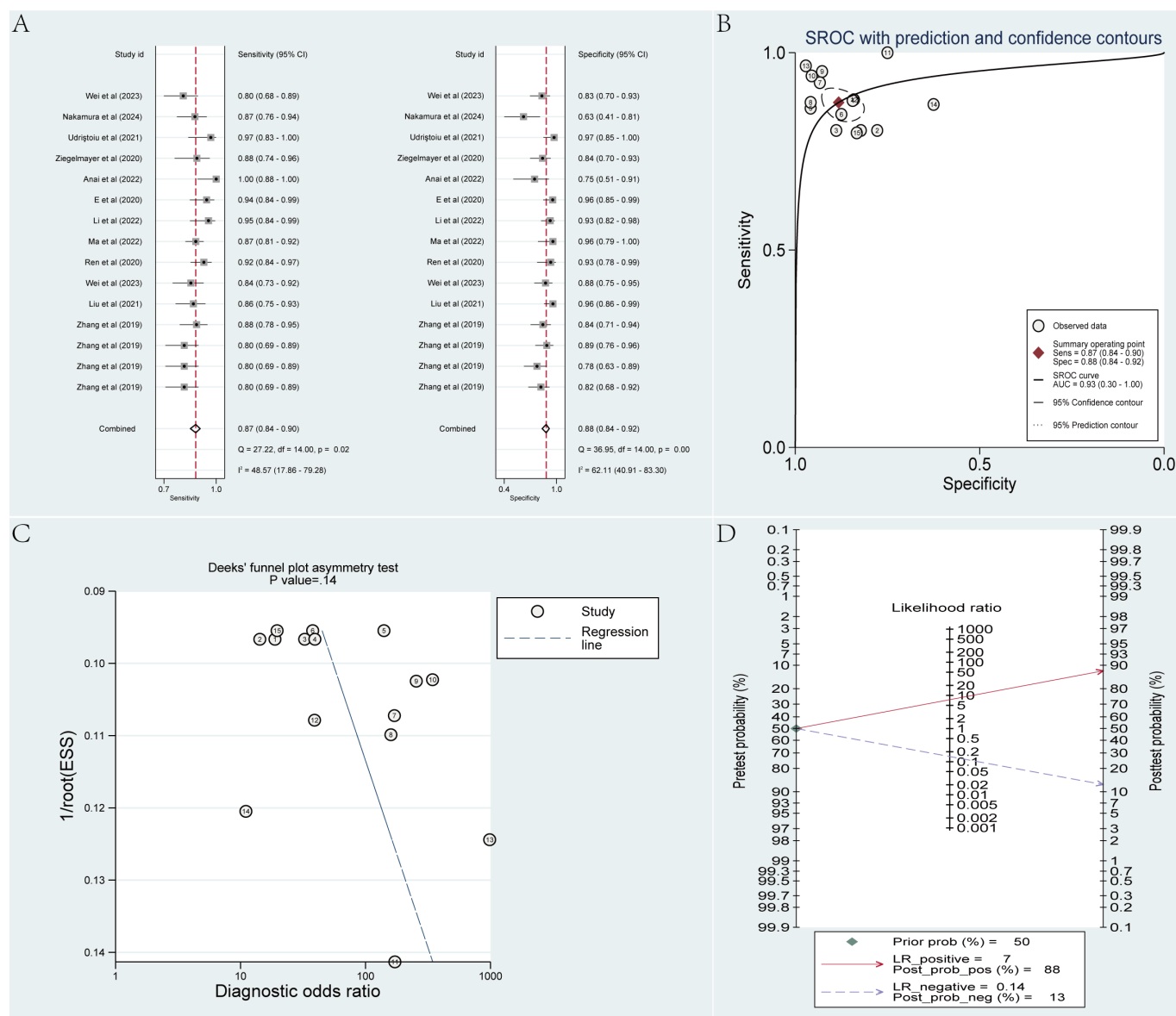


## Cross-Validation Set

The cross-validation set involved 15 radiomics-based ML models. The pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR, calculated employing a bivariate mixed-effects model, were 0.87 (95% CI 0.84-0.90), 0.88 (95% CI 0.84-0.92), 0.93 (95% CI 0.30-1.00), 7.4 (95% CI 5.2-10.6), 0.14 (95% CI 0.11-0.18), and 52 (95% CI 30-90), respectively. According to the funnel plot, no evident

publication bias was noted across the included studies. Approximately 50% (7635/14,406) of the cases in the analysis were PDAC. Taking 50% as the previous probability of the disease, given a PLR of 7, a positive result (PDAC) predicted by the models would have an 88% probability of true PDAC. With an NLR of 0.14, a negative result (MFP) predicted by the models would yield an 87% probability of true MFP (Figure 4).

**Figure 4.** Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the cross-validation set. Data were derived from studies [23,24,31,33,34,36-39,43,44]. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram. AUC: area under the curve; ESS: effective sample size; LR: logistic regression; Prob: probability; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristic.



## Subgroup Analysis by Different Imaging Modalities

### CT-Based ML Model

#### Training set

The training set included 8 CT-based ML models. By employing a bivariate mixed-effects model, the pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR were calculated as 0.90 (95% CI 0.84-0.95), 0.92 (95% CI 0.85-0.96), 0.97 (95% CI 0.41-1.00), 11.8 (95% CI 5.6-24.7), 0.10 (95% CI 0.06-0.19), and 113 (95% CI 33-389), respectively. According to the funnel plot, no noticeable publication bias was noted across the included studies. Approximately 50% (7635/14,406) of the cases in the analysis were PDAC. Taking 50% as the previous

probability for the disease, given a PLR of 12, a positive result (PDAC) predicted by the models would correspond to a 92% probability of true PDAC. With an NLR of 0.10, a negative result (MFP) predicted by the models would imply a 91% probability of true MFP (Multimedia Appendix 3).

#### Cross-validation training set

In total, 7 CT-based ML models were included in the cross-validation training set. The pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR, computed using a bivariate mixed-effects model, were 0.91 (95% CI 0.85-0.94), 0.89 (95% CI 0.84-0.93), 0.95 (95% CI 0.19-1.00), 8.6 (95% CI 5.6-13.0), 0.11 (95% CI 0.07-0.17), and 81 (95% CI 38-174), respectively. Based on the funnel plot, no evident publication bias was noted among the included studies. PDAC represented about 50% of the cases in the analysis. Taking 50% as the previous probability for the

disease, given a PLR of 9, a positive result (PDAC) predicted by the models would imply a 90% probability of true PDAC. With an NLR of 0.11, a negative result (MFP) predicted by the models would correspond to a 90% probability of true MFP ([Multimedia Appendix 4](#)).

### Independent Validation set

The independent validation set involved 8 CT-based ML models. The pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR, derived from a bivariate mixed-effects model, were 0.91 (95% CI 0.87-0.94), 0.91 (95% CI 0.84-0.95), 0.96 (95% CI 0.77-0.99), 9.9 (95% CI 5.3-18.8), 0.10 (95% CI 0.06-0.15), and 100 (95% CI 35-282), respectively. According to the funnel plot, no evident publication bias was noted across the included studies. In our analysis, PDAC accounted for approximately 50% (7635/14,406) of the cases. Taking 50% as the previous probability for the disease, given a PLR of 10, a positive result (PDAC) predicted by the models would correspond to a 91% probability of true PDAC. With an NLR of 0.10, a negative result (MFP) predicted by the models would correspond to a 91% probability of true MFP ([Multimedia Appendix 5](#)).

### EUS-Based ML Model

#### *Independent Validation Set*

A total of 6 EUS-based ML models were involved in the independent validation set. The pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR, calculated employing a bivariate mixed-effects model, were 0.94 (95% CI 0.91-0.96), 0.88 (95% CI 0.80-0.93), 0.97 (95% CI 0.78-1.00), 7.7 (95% CI 4.7-12.5), 0.07 (95% CI 0.04-0.11), and 114 (95% CI 60-217), respectively. The funnel plot implied no marked publication bias across the included studies. In our analysis, PDAC accounted for approximately 50% (7635/14,406) of the cases. Taking 50% as the previous probability of the disease, given a PLR of 8, a positive result (PDAC) predicted by the models would indicate an 88% probability of true PDAC. With an NLR of 0.07, a negative result (MFP) predicted by the models would correspond to a 94% probability of true MFP ([Multimedia Appendix 6](#)).

### PET-CT-Based ML Model

#### *Cross-Validation Training Set*

In total, 7 PET-CT-based ML models were included in the cross-validation training set. The pooled sensitivity, specificity, SROC AUC, PLR, NLR, and DOR, derived from a bivariate mixed-effects model, were 0.84 (95% CI 0.80-0.87), 0.88 (95% CI 0.82-0.92), 0.88 (95% CI 0.27-0.99), 7.1 (95% CI 4.7-10.9), 0.19 (95% CI 0.15-0.23), and 38 (95% CI 22-68), respectively. The funnel plot revealed no notable publication bias across the included studies. Approximately 50% (7635/14,406) of the cases in the analysis were PDAC. Taking 50% as the previous probability for the disease, given a PLR of 7, a positive result (PDAC) predicted by the models would correspond to an 88% probability of true PDAC. With an NLR of 0.19, a negative result (MFP) predicted by the

models would correspond to an 84% probability of true MFP ([Multimedia Appendix 7](#)).

### MRI-Based ML Model

According to Shiraishi et al [27], linear and nonlinear SVM models yielded AUCs of 0.89 and 0.82 in the training set, and 0.74 and 0.96 in the cross-validation set, respectively. In the testing set, the nonlinear SVM outperformed the linear SVM, with an AUC of 96.2% versus 74.4%. Deng et al [35] also employed an SVM model in their study, where the AUCs for the T1-weighted imaging, T2-weighted imaging, arterial phase, portal phase, and clinical models were 0.89, 0.91, 0.96, 0.99, and 0.52 in the training set, and 0.88, 0.90, 0.92, 0.96, and 0.65 in the independent validation set. All models surpassed radiologists and models based on clinical data in both sets. Chen et al [41] found that the accuracy and AUCs of the previous difference guidance network combined lesion and background information were 87.5% and 89.98% in conventional networks, and 89.77% and 92.8% in advanced networks.

## Discussion

### *Principal Findings*

This meta-analysis systematically evaluates the ability of radiomics-based ML models to distinguish PDAC from MFP. The results indicate that the pooled sensitivity, specificity, and SROC AUC in the independent validation set are 0.92, 0.90, and 0.94, respectively. These findings underscore the ability of radiomics-based ML models as a noninvasive diagnostic tool for distinguishing between PDAC and MFP, potentially decreasing the need for invasive procedures like EUS-FNA.

### *Comparison With Previous Work*

Several investigators have summarized the use of single imaging modalities, like CT or EUS, based on conventional imaging characteristics for differentiating PDAC from MFP. Their results indicate a promising diagnostic accuracy. For instance, Yoon et al [47] have reported that CT achieved a sensitivity of 0.83 (95% CI 0.75-0.90) and a specificity of 0.85 (95% CI 0.81-0.89), demonstrating its ability to detect pancreatic lesions. According to Yang et al [48], no noticeable difference in diagnostic performance is noted between EUS and CT. The sensitivity, specificity, and AUC are 0.82 (95% CI 0.73-0.88), 0.95 (95% CI 0.90-0.97), and 0.90 (95% CI, 0.87-92) for EUS, respectively, and 0.81 (95% CI 0.75-0.85), 0.94 (95% CI 0.90-0.96), and 0.92 (95% CI 0.90-0.94) for CT, respectively. However, these studies rely on traditional imaging features that are often subjectively interpreted. In contrast, our study applies radiomics-based approaches to the same imaging modalities, extracting high-dimensional quantitative data to provide a more objective and nuanced analysis. By incorporating multiple imaging sources and advanced diagnostic models, our study extends these findings and provides broader evidence, highlighting the superior predictive performance of

radiomics-based ML models in supporting clinical decision-making to distinguish PDAC from MFP.

## Future Directions

In this study, the diagnostic imaging modalities for PDAC primarily include CT, MRI, EUS, and PET-CT. According to the 2023 European Society for Medical Oncology Clinical Practice Guideline for pancreatic cancer [49], these imaging modalities play distinct roles in diagnosis and management. CT, as the recommended first-line imaging tool, provides a robust assessment of tumor size, vascular involvement, and metastatic disease. However, it has limitations in detecting isoattenuating tumors and small metastases. MRI, with superior soft-tissue contrast, complements CT, especially in complex cases or when contrast-enhanced imaging is contraindicated. EUS offers high-resolution images, making it essential for clarifying indeterminate findings. However, its application remains challenging due to its reliance on operators and limitations in evaluating distant metastases. PET-CT, although less commonly used for initial diagnosis, offers critical metabolic insights for identifying ambiguous or metastatic lesions. Each modality has its own strengths and limitations, and thus, the use of these modalities should be tailored for different patients in clinical practice.

Our meta-analysis highlights the effectiveness of radiomics-based ML models in differentiating PDAC from MFP. The pooled sensitivity and specificity for CT-based ML models exceed 90%, establishing CT as a reliable primary diagnostic tool. The AUCs for MRI-based ML models consistently are above 0.9, confirming their role in detailed tissue characterization. EUS-based ML models exhibited the highest sensitivity (94%) and were particularly effective for detecting subtle or challenging lesions. PET-CT-based ML models, despite moderate diagnostic metrics, provide valuable information on metabolism for ambiguous cases. According to these findings, CT could be used for initial screening of tumor size, vascular involvement, and metastatic disease, MRI for enhanced evaluation of complex cases, EUS for histological confirmation in indeterminate cases, and PET-CT for identifying ambiguous or metastatic lesions. This approach maximizes diagnostic precision and integrates radiomics effectively into clinical workflows.

Beyond diagnostic performance, real-world implementation of these imaging modalities also depends on factors like accessibility, cost-effectiveness, and ease of integration into routine practice. CT is widely accessible, relatively low-cost, and fast, making it suitable as a frontline diagnostic tool in most clinical settings. MRI offers superior tissue characterization but is more expensive and less available in resource-limited environments. EUS requires specialized equipment and expertise, often limiting its use to tertiary centers, while PET-CT is costly and not routinely employed in initial diagnostic workflows. Thus, while radiomics-based ML models show promising accuracy across modalities, their clinical utility must be contextualized based on resource availability, local expertise, and patient-specific considerations. Future implementation should align model development

with practical deployment strategies to optimize real-world adoption.

Radiomics alone has demonstrated considerable diagnostic performance in distinguishing PDAC from MFP. However, only a few studies use both clinical features and radiomics features to construct models. Clinical data (ie, patient age, serum markers, and medical history) can help reveal disease heterogeneity that may not be reflected in imaging features. Incorporating these features could significantly enhance model performance [21,50]. Such integration may also improve clinical interpretability and support decision-making, especially in borderline or ambiguous cases where radiomics alone might be insufficient. Nonetheless, combining multimodal data introduces challenges, like increased model complexity, the risk of overfitting with limited sample sizes, and the need for standardized clinical data collection. Future studies should focus on developing interpretable multidomain models and prospective data integration strategies that align with real-world clinical workflows, thereby enhancing both diagnostic accuracy and clinical applicability.

Multicenter and multiregional validations are crucial for enhancing the generalizability of radiomics-based ML models, as they provide a more comprehensive assessment of model performance across diverse clinical settings [21,51]. In this meta-analysis, most studies rely on single-center designs with internal validation, such as random sampling or cross-validation, which may not fully capture the variability encountered in different clinical settings. The use of single-center data can introduce bias, as it may not account for differences in imaging equipment, acquisition protocols, and patient populations that are present in real-world clinical practice.

Future studies should prioritize multicenter collaboration and the use of independent external validation sets to ensure broader representativeness. To address heterogeneity across centers, standardized imaging acquisition protocols (eg, harmonized contrast phases and MRI sequences) and consistent reporting of imaging parameters are essential. Image preprocessing methods (such as voxel resampling and intensity normalization) and harmonization techniques (eg, ComBat) can further mitigate scanner-related variability. In addition, the adoption of consensus-based guidelines for segmentation, feature extraction, and model evaluation would facilitate methodological consistency and improve reproducibility. These practical measures are essential to support the robust and clinically scalable deployment of radiomics-based ML models.

Given the relatively low incidence of PDAC, many studies included in this meta-analysis have limited sample sizes, which could contribute to selection bias and affect model robustness. The diagnostic performance of radiomics-based ML models is closely tied to sample sizes. Therefore, future multicenter studies with larger sample sizes are warranted. In addition, techniques, such as data augmentation (eg, image enhancement and synthetic data generation), can help expand datasets to improve the generalizability and stability of models in clinical practice.



In this meta-analysis, most included studies use traditional ML approaches, like SVM and RF, to construct radiomics-based ML models. While these methods can achieve high diagnostic accuracy, they often rely on manual feature extraction, which may lead to information loss [52]. This limitation can affect the generalizability of radiomics-based ML models across different clinical settings.

In contrast, DL techniques like convolutional neural networks can automatically learn features from raw images, offering a more comprehensive feature representation [52]. Although relatively few studies in our analysis use DL to construct models, the results are promising, indicating the potential advantages of DL in capturing complex imaging patterns. Future studies should explore advanced DL methods, like multimodal learning and transfer learning, to improve the diagnostic performance of radiomics-based ML models by integrating diverse imaging data [53]. This could enhance the adaptability and clinical utility of these models in differentiating PDAC from MFP. According to the RQS, the score for the included studies ranges from 5 points (14%) to 17 points (47%), with an average score of 9 (25%). This relatively low-to-moderate scoring reflects the stringency of the RQS framework. It places particular emphasis on prospective design, biological validation, repeatability assessment, and data or code availability. However, these factors are not widely adopted in radiomics research. While lower RQS scores suggest areas for improvement, they should be interpreted with caution. Most included studies fulfilled essential technical criteria, like detailed imaging protocols, appropriate feature selection methods, and validation, using independent datasets. These strengths underpin the robustness of the diagnostic accuracy estimates in this meta-analysis. However, the absence of phantom studies, limited use of longitudinal imaging, and inadequate reporting of model calibration or cost-effectiveness may limit model reproducibility and cross-center generalizability. Future radiomics studies should aim for improved adherence to reporting guidelines and quality frameworks to enhance transparency, clinical translatability, and model deployment across heterogeneous clinical settings.

## Strengths and Limitations

This study represents the first systematic appraisal of radiomics-based ML models in distinguishing PDAC from MFP across multiple imaging modalities, providing robust evidence for noninvasive diagnostic strategies. By synthesizing data from various studies, it offers critical insights into the clinical application and development of radiomics-based ML models in the future.

However, several limitations should be considered. First, despite a vast and systematic search, the number of included

studies is relatively small, limiting detailed subgroup analyses of different radiomics-based ML models and reducing the depth of discussions on specific diagnostic approaches. This limitation may affect the generalizability of the findings to diverse clinical settings. Second, substantial heterogeneity remains a challenge across the included studies. Differences in imaging protocols, segmentation methods (manual vs semiautomatic), feature selection strategies (such as least absolute shrinkage and selection operator and minimum redundancy maximum relevance), classifier types (such as logistic regression, decision tree, SVM, adaptive boosting, RF, and DL), and validation strategies all contribute to methodological variability. Many pipelines also rely on investigator-defined parameters, which may introduce bias. While most studies employ internal validation methods, like random sampling or cross-validation, only a limited number have implemented independent external or multicenter validation, limiting the generalizability of model performance across diverse clinical environments. These issues highlight the urgent need for standardized reporting guidelines and harmonized study designs to ensure the reproducibility and clinical applicability of radiomics-based ML models. Future studies should prioritize multicenter collaboration and adopt standardized radiomics pipelines to reduce heterogeneity and enhance generalizability. Third, although subgroup analyses are executed by imaging modality, these are descriptive in nature and employ separate bivariate models for each modality. No formal statistical comparisons between subgroups or adjustments for multiple comparisons are performed. This approach is chosen to examine diagnostic performance trends across modalities without overinterpreting differences, given the limited number of studies per subgroup and inherent heterogeneity. Future meta-analyses with larger sample sizes may benefit from formal comparative subgroup analyses. Finally, certain imaging modalities are less representative, such as MRI and PET-CT, making it challenging to fully appraise their diagnostic performance and potentially influencing the pooled diagnostic estimates. Finally, our meta-analysis does not describe imaging parameters across studies in detail, such as the specific phases used in contrast-enhanced CT (like arterial and venous) and the types of MRI sequences used (like T1-weighted and T2-weighted). This may hinder the assessment of consistency and reproducibility across studies.

## Conclusions

Radiomics-based ML models demonstrate excellent diagnostic accuracy in differentiating PDAC from MFP, supporting their potential role as noninvasive tools in clinical practice. However, to facilitate real-world adoption, future research should prioritize multicenter validation, standardized imaging protocols, and reproducible model pipelines.

## Acknowledgments

This work was supported by the Natural Science Foundation of Shandong Province (Grant ZR2022MH145).

## Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

---

**Authors' Contributions**

LZ and SZ contributed to the conception and design of the study. LZ, DL, TS, and TX were responsible for data acquisition. ZLC and TS performed data analysis and interpretation. LZ and DL drafted the manuscript. LZ, DL, TS, TX, and SZ critically revised the manuscript. SZ reviewed and approved the final submitted version of the manuscript and served as the corresponding author.

---

**Conflicts of Interest**

None declared.

---

**Multimedia Appendix 1**

Search strategy.

[\[DOCX File \(Microsoft Word File\), 16 KB-Multimedia Appendix 1\]](#)

---

**Multimedia Appendix 2**

Radiomics quality score of the included studies.

[\[DOCX File \(Microsoft Word File\), 31 KB-Multimedia Appendix 2\]](#)

---

**Multimedia Appendix 3**

Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the computed tomography training set. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram.

[\[PNG File \(Portable Network Graphics File\), 324 KB-Multimedia Appendix 3\]](#)

---

**Multimedia Appendix 4**

Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the computed tomography cross-validation set. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram.

[\[PNG File \(Portable Network Graphics File\), 312 KB-Multimedia Appendix 4\]](#)

---

**Multimedia Appendix 5**

Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the computed tomography independent validation set. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram.

[\[PNG File \(Portable Network Graphics File\), 319 KB-Multimedia Appendix 5\]](#)

---

**Multimedia Appendix 6**

Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the endoscopic ultrasound independent validation set. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram.

[\[PNG File \(Portable Network Graphics File\), 301 KB-Multimedia Appendix 6\]](#)

---

**Multimedia Appendix 7**

Meta-analysis results of radiomics-based machine learning models in differentiating pancreatic ductal adenocarcinoma from mass-forming pancreatitis within the positron emission tomography-computed tomography cross-validation set. (A) Coupled forest plots of pooled sensitivity and specificity. (B) Summary receiver operating characteristic curve. (C) Deeks' funnel plot. (D) Fagan nomogram.

[\[PNG File \(Portable Network Graphics File\), 312 KB-Multimedia Appendix 7\]](#)

---

**Checklist 1**

PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Checklist.

[\[DOCX File \(Microsoft Word File\), 30 KB-Checklist 1\]](#)

---

**References**

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. May 2021;71(3):209-249. [doi: [10.3322/caac.21660](https://doi.org/10.3322/caac.21660)] [Medline: [33538338](https://pubmed.ncbi.nlm.nih.gov/33538338/)]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. Jan 2018;68(1):7-30. [doi: [10.3322/caac.21442](https://doi.org/10.3322/caac.21442)] [Medline: [29313949](https://pubmed.ncbi.nlm.nih.gov/29313949/)]



3. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. *Lancet*. Jul 2, 2016;388(10039):73-85. [doi: [10.1016/S0140-6736\(16\)00141-0](https://doi.org/10.1016/S0140-6736(16)00141-0)] [Medline: [26830752](https://pubmed.ncbi.nlm.nih.gov/26830752/)]
4. Vaseen H, Ibrahim I, Ponce CG, et al. Benefit of surveillance for pancreatic cancer in high-risk individuals: outcome of long-term prospective follow-up studies from three European expert centers. *J Clin Oncol*. Jun 10, 2016;34(17):2010-2019. [doi: [10.1200/JCO.2015.64.0730](https://doi.org/10.1200/JCO.2015.64.0730)] [Medline: [27114589](https://pubmed.ncbi.nlm.nih.gov/27114589/)]
5. Gonda TA, Everett JN, Wallace M, Simeone DM, PRECEDE Consortium. Recommendations for a more organized and effective approach to the early detection of pancreatic cancer from the PRECEDE (Pancreatic Cancer Early Detection) consortium. *Gastroenterology*. Dec 2021;161(6):1751-1757. [doi: [10.1053/j.gastro.2021.08.036](https://doi.org/10.1053/j.gastro.2021.08.036)] [Medline: [34454916](https://pubmed.ncbi.nlm.nih.gov/34454916/)]
6. Matsubayashi H, Matsui T, Yabuuchi Y, et al. Endoscopic ultrasonography guided-fine needle aspiration for the diagnosis of solid pancreaticobiliary lesions: clinical aspects to improve the diagnosis. *World J Gastroenterol*. Jan 14, 2016;22(2):628-640. [doi: [10.3748/wjg.v22.i2.628](https://doi.org/10.3748/wjg.v22.i2.628)] [Medline: [26811612](https://pubmed.ncbi.nlm.nih.gov/26811612/)]
7. Lee H, Lee JK, Kang SS, et al. Is there any clinical or radiologic feature as a preoperative marker for differentiating mass-forming pancreatitis from early-stage pancreatic adenocarcinoma? *Hepatogastroenterology*. 2007;54(79):2134-2140. [Medline: [18251176](https://pubmed.ncbi.nlm.nih.gov/18251176/)]
8. Kim M, Jang KM, Kim JH, et al. Differentiation of mass-forming focal pancreatitis from pancreatic ductal adenocarcinoma: value of characterizing dynamic enhancement patterns on contrast-enhanced MR images by adding signal intensity color mapping. *Eur Radiol*. Apr 2017;27(4):1722-1732. [doi: [10.1007/s00330-016-4522-0](https://doi.org/10.1007/s00330-016-4522-0)] [Medline: [27510628](https://pubmed.ncbi.nlm.nih.gov/27510628/)]
9. Kobayashi G, Fujita N, Noda Y, et al. Lymphoplasmacytic sclerosing pancreatitis forming a localized mass: a variant form of autoimmune pancreatitis. *J Gastroenterol*. Aug 2007;42(8):650-656. [doi: [10.1007/s00535-007-2068-2](https://doi.org/10.1007/s00535-007-2068-2)] [Medline: [17701128](https://pubmed.ncbi.nlm.nih.gov/17701128/)]
10. Raimondi S, Lowenfels AB, Morselli-Labate AM, Maisonneuve P, Pezzilli R. Pancreatic cancer in chronic pancreatitis: aetiology, incidence, and early detection. *Best Pract Res Clin Gastroenterol*. Jun 2010;24(3):349-358. [doi: [10.1016/j.bpg.2010.02.007](https://doi.org/10.1016/j.bpg.2010.02.007)] [Medline: [20510834](https://pubmed.ncbi.nlm.nih.gov/20510834/)]
11. Kitano M, Yoshida T, Itonaga M, Tamura T, Hatamaru K, Yamashita Y. Impact of endoscopic ultrasonography on diagnosis of pancreatic cancer. *J Gastroenterol*. Jan 2019;54(1):19-32. [doi: [10.1007/s00535-018-1519-2](https://doi.org/10.1007/s00535-018-1519-2)] [Medline: [30406288](https://pubmed.ncbi.nlm.nih.gov/30406288/)]
12. Yoshinaga S, Itoi T, Yamao K, et al. Safety and efficacy of endoscopic ultrasound-guided fine needle aspiration for pancreatic masses: a prospective multicenter study. *Dig Endosc*. Jan 2020;32(1):114-126. [doi: [10.1111/den.13457](https://doi.org/10.1111/den.13457)] [Medline: [31166046](https://pubmed.ncbi.nlm.nih.gov/31166046/)]
13. Mohanty SK, Pradhan D, Sharma S, et al. Endoscopic ultrasound guided fine-needle aspiration: what variables influence diagnostic yield? *Diagn Cytopathol*. Apr 2018;46(4):293-298. [doi: [10.1002/dc.23883](https://doi.org/10.1002/dc.23883)] [Medline: [29280329](https://pubmed.ncbi.nlm.nih.gov/29280329/)]
14. Iglesias-Garcia J, Dominguez-Munoz JE, Abdulkader I, et al. Influence of on-site cytopathology evaluation on the diagnostic accuracy of endoscopic ultrasound-guided fine needle aspiration (EUS-FNA) of solid pancreatic masses. *Am J Gastroenterol*. Sep 2011;106(9):1705-1710. [doi: [10.1038/ajg.2011.119](https://doi.org/10.1038/ajg.2011.119)] [Medline: [21483464](https://pubmed.ncbi.nlm.nih.gov/21483464/)]
15. Mizuide M, Ryozaawa S, Fujita A, et al. Complications of endoscopic ultrasound-guided fine needle aspiration: a narrative review. *Diagnostics (Basel)*. Nov 17, 2020;10(11):964. [doi: [10.3390/diagnostics10110964](https://doi.org/10.3390/diagnostics10110964)] [Medline: [33213103](https://pubmed.ncbi.nlm.nih.gov/33213103/)]
16. Yane K, Kuwatani M, Yoshida M, et al. Non-negligible rate of needle tract seeding after endoscopic ultrasound-guided fine-needle aspiration for patients undergoing distal pancreatectomy for pancreatic cancer. *Dig Endosc*. Jul 2020;32(5):801-811. [doi: [10.1111/den.13615](https://doi.org/10.1111/den.13615)] [Medline: [31876309](https://pubmed.ncbi.nlm.nih.gov/31876309/)]
17. Sato N, Takano S, Yoshitomi H, et al. Needle tract seeding recurrence of pancreatic cancer in the gastric wall with para-gastric lymph node metastasis after endoscopic ultrasound-guided fine needle aspiration followed by pancreatectomy: a case report and literature review. *BMC Gastroenterol*. Jan 15, 2020;20(1):13. [doi: [10.1186/s12876-020-1159-x](https://doi.org/10.1186/s12876-020-1159-x)] [Medline: [31941458](https://pubmed.ncbi.nlm.nih.gov/31941458/)]
18. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. Feb 2016;278(2):563-577. [doi: [10.1148/radiol.2015151169](https://doi.org/10.1148/radiol.2015151169)] [Medline: [26579733](https://pubmed.ncbi.nlm.nih.gov/26579733/)]
19. Tourassi GD. Journey toward computer-aided diagnosis: role of image texture analysis. *Radiology*. Nov 1999;213(2):317-320. [doi: [10.1148/radiology.213.2.r99nv49317](https://doi.org/10.1148/radiology.213.2.r99nv49317)] [Medline: [10551208](https://pubmed.ncbi.nlm.nih.gov/10551208/)]
20. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
21. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. Dec 2017;14(12):749-762. [doi: [10.1038/nrclinonc.2017.141](https://doi.org/10.1038/nrclinonc.2017.141)] [Medline: [28975929](https://pubmed.ncbi.nlm.nih.gov/28975929/)]
22. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. Oct 2005;58(10):982-990. [doi: [10.1016/j.jclinepi.2005.02.022](https://doi.org/10.1016/j.jclinepi.2005.02.022)] [Medline: [16168343](https://pubmed.ncbi.nlm.nih.gov/16168343/)]

23. Ren S, Zhao R, Zhang J, et al. Diagnostic accuracy of unenhanced CT texture analysis to differentiate mass-forming pancreatitis from pancreatic ductal adenocarcinoma. *Abdom Radiol (NY)*. May 2020;45(5):1524-1533. [doi: [10.1007/s00261-020-02506-6](https://doi.org/10.1007/s00261-020-02506-6)] [Medline: [32279101](https://pubmed.ncbi.nlm.nih.gov/32279101/)]
24. Zhang Y, Cheng C, Liu Z, et al. Radiomics analysis for the differentiation of autoimmune pancreatitis and pancreatic ductal adenocarcinoma in <sup>18</sup>F-FDG PET/CT. *Med Phys Mex Symp Med Phys*. Oct 2019;46(10):4520-4530. [doi: [10.1002/mp.13733](https://doi.org/10.1002/mp.13733)]
25. Zhang H, Meng Y, Li Q, et al. Two nomograms for differentiating mass-forming chronic pancreatitis from pancreatic ductal adenocarcinoma in patients with chronic pancreatitis. *Eur Radiol*. 2022;32(9):6336-6347. [doi: [10.1007/s00330-022-08698-3](https://doi.org/10.1007/s00330-022-08698-3)]
26. Ye Y, Zhang J, Song P, et al. Clinical features and computed tomography radiomics-based model for predicting pancreatic ductal adenocarcinoma and focal mass-forming pancreatitis. *Technol Cancer Res Treat*. 2023;22:15330338231180792. [doi: [10.1177/15330338231180792](https://doi.org/10.1177/15330338231180792)] [Medline: [37287274](https://pubmed.ncbi.nlm.nih.gov/37287274/)]
27. Shiraishi M, Igarashi T, Hiroaki F, Oe R, Ohki K, Ojiri H. Radiomics based on diffusion-weighted imaging for differentiation between focal-type autoimmune pancreatitis and pancreatic carcinoma. *Br J Radiol*. Dec 1, 2022;95(1140):20210456. [doi: [10.1259/bjr.20210456](https://doi.org/10.1259/bjr.20210456)] [Medline: [35946923](https://pubmed.ncbi.nlm.nih.gov/35946923/)]
28. Ren S, Zhang J, Chen J, et al. Evaluation of texture analysis for the differential diagnosis of mass-forming pancreatitis from pancreatic ductal adenocarcinoma on contrast-enhanced CT images. *Front Oncol*. 2019;9:1171. [doi: [10.3389/fonc.2019.01171](https://doi.org/10.3389/fonc.2019.01171)] [Medline: [31750254](https://pubmed.ncbi.nlm.nih.gov/31750254/)]
29. Qu W, Zhou Z, Yuan G, et al. Is the radiomics-clinical combined model helpful in distinguishing between pancreatic cancer and mass-forming pancreatitis? *Eur J Radiol*. Jul 2023;164:110857. [doi: [10.1016/j.ejrad.2023.110857](https://doi.org/10.1016/j.ejrad.2023.110857)] [Medline: [37172441](https://pubmed.ncbi.nlm.nih.gov/37172441/)]
30. Park S, Chu LC, Hruban RH, et al. Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features. *Diagn Interv Imaging*. Sep 2020;101(9):555-564. [doi: [10.1016/j.diii.2020.03.002](https://doi.org/10.1016/j.diii.2020.03.002)]
31. Ma X, Wang YR, Zhuo LY, et al. Retrospective analysis of the value of enhanced CT radiomics analysis in the differential diagnosis between pancreatic cancer and chronic pancreatitis. *Int J Gen Med*. 2022;15:233-241. [doi: [10.2147/IJGM.S337455](https://doi.org/10.2147/IJGM.S337455)] [Medline: [35023961](https://pubmed.ncbi.nlm.nih.gov/35023961/)]
32. Lu J, Jiang N, Zhang Y, Li D. A CT based radiomics nomogram for differentiation between focal-type autoimmune pancreatitis and pancreatic ductal adenocarcinoma. *Front Oncol*. 2023;13:979437. [doi: [10.3389/fonc.2023.979437](https://doi.org/10.3389/fonc.2023.979437)] [Medline: [36937433](https://pubmed.ncbi.nlm.nih.gov/36937433/)]
33. Liu Z, Li M, Zuo C, et al. Radiomics model of dual-time 2-[<sup>18</sup>F]FDG PET/CT imaging to distinguish between pancreatic ductal adenocarcinoma and autoimmune pancreatitis. *Eur Radiol*. Sep 2021;31(9):6983-6991. [doi: [10.1007/s00330-021-07778-0](https://doi.org/10.1007/s00330-021-07778-0)] [Medline: [33677645](https://pubmed.ncbi.nlm.nih.gov/33677645/)]
34. Li J, Liu F, Fang X, et al. CT radiomics features in differentiation of focal-type autoimmune pancreatitis from pancreatic ductal adenocarcinoma: a propensity score analysis. *Acad Radiol*. Mar 2022;29(3):358-366. [doi: [10.1016/j.acra.2021.04.014](https://doi.org/10.1016/j.acra.2021.04.014)] [Medline: [34108115](https://pubmed.ncbi.nlm.nih.gov/34108115/)]
35. Deng Y, Ming B, Zhou T, et al. Radiomics model based on MR images to discriminate pancreatic ductal adenocarcinoma and mass-forming chronic pancreatitis lesions. *Front Oncol*. 2021;11:620981. [doi: [10.3389/fonc.2021.620981](https://doi.org/10.3389/fonc.2021.620981)] [Medline: [33842325](https://pubmed.ncbi.nlm.nih.gov/33842325/)]
36. E L, Xu Y, Wu Z, et al. Differentiation of focal-type autoimmune pancreatitis from pancreatic ductal adenocarcinoma using radiomics based on multiphasic computed tomography. *J Comput Assist Tomogr*. 2020;44(4):511-518. [doi: [10.1097/RCT.0000000000001049](https://doi.org/10.1097/RCT.0000000000001049)] [Medline: [32697521](https://pubmed.ncbi.nlm.nih.gov/32697521/)]
37. Anai K, Hayashida Y, Ueda I, et al. The effect of CT texture-based analysis using machine learning approaches on radiologists' performance in differentiating focal-type autoimmune pancreatitis and pancreatic duct carcinoma. *Jpn J Radiol*. Nov 2022;40(11):1156-1165. [doi: [10.1007/s11604-022-01298-7](https://doi.org/10.1007/s11604-022-01298-7)] [Medline: [35727458](https://pubmed.ncbi.nlm.nih.gov/35727458/)]
38. Ziegelmayer S, Kaissis G, Harder F, et al. Deep convolutional neural network-assisted feature extraction for diagnostic discrimination and feature visualization in pancreatic ductal adenocarcinoma (PDAC) versus autoimmune pancreatitis (AIP). *J Clin Med*. Dec 11, 2020;9(12):1-8. [doi: [10.3390/jcm9124013](https://doi.org/10.3390/jcm9124013)] [Medline: [33322559](https://pubmed.ncbi.nlm.nih.gov/33322559/)]
39. Wei W, Jia G, Wu Z, et al. A multidomain fusion model of radiomics and deep learning to discriminate between PDAC and AIP based on <sup>18</sup>F-FDG PET/CT images. *Jpn J Radiol*. Apr 2023;41(4):417-427. [doi: [10.1007/s11604-022-01363-1](https://doi.org/10.1007/s11604-022-01363-1)] [Medline: [36409398](https://pubmed.ncbi.nlm.nih.gov/36409398/)]
40. Tong T, Gu J, Xu D, et al. Deep learning radiomics based on contrast-enhanced ultrasound images for assisted diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis. *BMC Med*. Mar 2, 2022;20(1):74. [doi: [10.1186/s12916-022-02258-8](https://doi.org/10.1186/s12916-022-02258-8)] [Medline: [35232446](https://pubmed.ncbi.nlm.nih.gov/35232446/)]
41. Chen LD, Bao KZ, Chen Y, Hao JA, He JF. A deep learning framework for mass-forming chronic pancreatitis and pancreatic ductal adenocarcinoma classification based on magnetic resonance imaging. *CMC*. 2024;79(1):409-427. [doi: [10.32604/cmc.2024.048507](https://doi.org/10.32604/cmc.2024.048507)]

42. Cao K, Xia Y, Yao J, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. *Nat Med*. Dec 2023;29(12):3033-3043. [doi: [10.1038/s41591-023-02640-w](https://doi.org/10.1038/s41591-023-02640-w)] [Medline: [37985692](https://pubmed.ncbi.nlm.nih.gov/37985692/)]
43. Udriștoiu AL, Cazacu IM, Gruionu LG, et al. Real-time computer-aided diagnosis of focal pancreatic masses from endoscopic ultrasound imaging based on a hybrid convolutional and long short-term memory neural network model. *PLoS One*. 2021;16(6):e0251701. [doi: [10.1371/journal.pone.0251701](https://doi.org/10.1371/journal.pone.0251701)] [Medline: [34181680](https://pubmed.ncbi.nlm.nih.gov/34181680/)]
44. Nakamura H, Fukuda M, Matsuda A, et al. Differentiating localized autoimmune pancreatitis and pancreatic ductal adenocarcinoma using endoscopic ultrasound images with deep learning. *DEN Open*. Apr 2024;4(1):e344. [doi: [10.1002/deo2.344](https://doi.org/10.1002/deo2.344)] [Medline: [38434146](https://pubmed.ncbi.nlm.nih.gov/38434146/)]
45. Marya NB, Powers PD, Chari ST, et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut*. Jul 2021;70(7):1335-1344. [doi: [10.1136/gutjnl-2020-322821](https://doi.org/10.1136/gutjnl-2020-322821)] [Medline: [33028668](https://pubmed.ncbi.nlm.nih.gov/33028668/)]
46. Kuwahara T, Hara K, Mizuno N, et al. Artificial intelligence using deep learning analysis of endoscopic ultrasonography images for the differential diagnosis of pancreatic masses. *Endoscopy*. Feb 2023;55(2):140-149. [doi: [10.1055/a-1873-7920](https://doi.org/10.1055/a-1873-7920)] [Medline: [35688454](https://pubmed.ncbi.nlm.nih.gov/35688454/)]
47. Yoon SB, Jeon TY, Moon SH, Shin DW, Lee SM, Choi MH. Sa1141 differentiation of autoimmune pancreatitis from pancreatic adenocarcinoma using ct characteristics: a systematic review and meta-analysis. *Gastroenterology*. May 2023;164(6):S-299. [doi: [10.1016/S0016-5085\(23\)01674-8](https://doi.org/10.1016/S0016-5085(23)01674-8)]
48. Yang J, Huang J, Zhang Y, Lu Q. Contrast-enhanced ultrasound (CEUS) and contrast-enhanced computed tomography (CECT) for differentiating between mass-forming pancreatitis and pancreatic ductal adenocarcinoma: a meta-analysis. *Insights Imaging*. 2022;14:352-353. [doi: [10.1186/s13244-022-01337-x](https://doi.org/10.1186/s13244-022-01337-x)]
49. Conroy T, Pfeiffer P, Vilgrain V, et al. Pancreatic cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. *Ann Oncol*. Nov 2023;34(11):987-1002. [doi: [10.1016/j.annonc.2023.08.009](https://doi.org/10.1016/j.annonc.2023.08.009)] [Medline: [37678671](https://pubmed.ncbi.nlm.nih.gov/37678671/)]
50. Kozikowski M, Suarez-Ibarrola R, Osiecki R, et al. Role of radiomics in the prediction of muscle-invasive bladder cancer: a systematic review and meta-analysis. *Eur Urol Focus*. May 2022;8(3):728-738. [doi: [10.1016/j.euf.2021.05.005](https://doi.org/10.1016/j.euf.2021.05.005)] [Medline: [34099417](https://pubmed.ncbi.nlm.nih.gov/34099417/)]
51. Abbaspour E, Karimzadgh S, Monsef A, Joukar F, Mansour-Ghanaei F, Hassanipour S. Application of radiomics for preoperative prediction of lymph node metastasis in colorectal cancer: a systematic review and meta-analysis. *Int J Surg*. Jun 1, 2024;110(6):3795-3813. [doi: [10.1097/JS9.0000000000001239](https://doi.org/10.1097/JS9.0000000000001239)] [Medline: [38935817](https://pubmed.ncbi.nlm.nih.gov/38935817/)]
52. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. Jun 21, 2017;19(221-48):221-248. [doi: [10.1146/annurev-bioeng-071516-044442](https://doi.org/10.1146/annurev-bioeng-071516-044442)] [Medline: [28301734](https://pubmed.ncbi.nlm.nih.gov/28301734/)]
53. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. Jan 2019;25(1):24-29. [doi: [10.1038/s41591-018-0316-z](https://doi.org/10.1038/s41591-018-0316-z)] [Medline: [30617335](https://pubmed.ncbi.nlm.nih.gov/30617335/)]

## Abbreviations

**AUC:** area under the curve  
**CT:** computed tomography  
**DL:** deep learning  
**DOR:** diagnostic odds ratio  
**EUS:** endoscopic ultrasound  
**EUS-FNA:** endoscopic ultrasound-guided fine-needle aspiration  
**MFP:** mass-forming pancreatitis  
**ML:** machine learning  
**MRI:** magnetic resonance imaging  
**NLR:** negative likelihood ratio  
**PDAC:** pancreatic ductal adenocarcinoma  
**PET-CT:** positron emission tomography-computed tomography  
**PLR:** positive likelihood ratio  
**PRISMA:** Preferred Reporting Items for Systematic reviews and Meta-Analyses  
**RF:** random forest  
**RQS:** radiomics quality score  
**SROC:** summary receiver operating characteristic  
**SROC AUC:** area under the curve of summary receiver operating characteristic  
**SVM:** support vector machine

*Edited by Javad Sarvestan; peer-reviewed by Peng Wu, Shadi Afyouni; submitted 09.02.2025; final revised version received 05.06.2025; accepted 05.06.2025; published 31.07.2025*

Please cite as:

Zhang L, Li D, Su T, Xiao T, Zhao S

*Effectiveness of Radiomics-Based Machine Learning Models in Differentiating Pancreatitis and Pancreatic Ductal Adenocarcinoma: Systematic Review and Meta-Analysis*

*J Med Internet Res* 2025;27:e72420

URL: <https://www.jmir.org/2025/1/e72420>

doi: [10.2196/72420](https://doi.org/10.2196/72420)

© Lechang Zhang, Dewei Li, Tong Su, Tong Xiao, Shulei Zhao. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 31.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.