

Supplementary Materials

Artificial intelligence in pancreatic cancer surgery: evidence maturity, boundaries, and clinical translation

Zhike Li[#], Gongze Peng[#], Liping Liu, Yongzhu He

Department of Hepatobiliary and Pancreas Surgery, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, Shenzhen 518020, Guangdong, China.

[#]These authors contributed equally to this work.

Correspondence to: Dr. Liping Liu, Dr. Yongzhu He, Department of Hepatobiliary and Pancreas Surgery, Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, Shenzhen 518020, Guangdong, China. E-mail: liuliping@mail.sustech.edu.cn; heyongzhu@szhospital.com

Supplementary Method 1. Literature search, study selection, and evidence-maturity appraisal

Literature sources and search strategy

PubMed/MEDLINE was used as the primary bibliographic database. The structured core-evidence search was organized across five predefined clinical domains along the pancreatic cancer surgical pathway: (S1) resectability, vascular involvement, and margin assessment; (S2) three-dimensional reconstruction, segmentation, and surgical planning; (S3) perioperative treatment stratification, including neoadjuvant and adjuvant treatment; (S4) postoperative complications; and (S5) lymph-node metastasis, recurrence, survival, and prognosis. Search concepts combined terms for pancreatic cancer or pancreatic ductal adenocarcinoma with artificial intelligence, machine learning, deep learning, radiomics, pathomics, and domain-specific clinical terms. Searches were restricted to English-language publications.

Reference lists of relevant studies were additionally examined to identify potentially missed reports. Because contemporaneous source-tracing records were not preserved in sufficient detail, the exact number of studies identified exclusively through reference tracing could not be reconstructed. Additional references were used where appropriate to support disease background, established clinical concepts, methodological appraisal, regulatory considerations, and broader translational discussion. These contextual sources were not included in evidence-maturity grading.

Core evidence set and study selection

The structured core-evidence search covered publications from January 1, 2020 to February 28, 2026 and yielded 718 records across the five clinical domains. After deduplication by PubMed identifier, 654 unique records remained, of which 112 were retained after relevance screening for more detailed eligibility assessment.

The final core evidence set used in this review comprised 34 studies, including 31 direct AI studies and three adjacent or protocol studies. Only the 31 direct AI studies were included in the formal evidence-maturity analysis.

Supplementary search for recent literature

To capture recent developments, an additional PubMed/MEDLINE search was conducted through September 17, 2026 using the same five clinical domains and general eligibility framework. Newly identified literature was screened for relevance to the scope of the review and was used, where appropriate, to update the narrative synthesis of rapidly evolving topics. This supplementary search was

used to maintain the currency of the narrative synthesis and did not automatically redefine the 31-study evidence-maturity cohort.

Eligibility criteria

Direct AI studies were eligible when they were original English-language investigations applying artificial intelligence, machine learning, deep learning, radiomics, pathomics, or related computational methods to clinically relevant tasks along the pancreatic cancer surgical pathway and reported model development, validation, or clinical evaluation using patient-level clinical, imaging, pathological, operative, or omics data.

Reviews, editorials, conference abstracts, animal or cell studies, non-AI treatment studies, conventional three-dimensional reconstruction without an AI component, and studies without a clinically relevant pancreatic surgical task were excluded from the direct-AI evidence set. Selected adjacent studies or study protocols were retained separately when they helped define technical boundaries or emerging translational pathways.

Z. L. and G. P. performed literature retrieval and evidence synthesis independently, with differences arising in those activities resolved through discussion. Full-text eligibility assessment was not performed independently in duplicate. The 112 records retained after relevance screening were candidates for detailed eligibility assessment. The historical records do not permit reliable reconstruction of the number of assessors at each screening stage or document a formal process for resolving eligibility disagreements. Subsequent study-level source verification during manuscript revision was distinct from the original screening process.

Data extraction and methodological appraisal

For studies in the core evidence set, information was extracted on study design, target population, data source, sample and event composition, clinical task, model type, model-development and validation strategy, principal performance results, calibration and threshold information, decision-utility analyses, workflow characteristics, clinical outcomes, and major methodological limitations.

Methodological appraisal was informed by principles from TRIPOD+AI, PROBAST, CLAIM, decision-curve analysis, and FUTURE-AI. These frameworks were used to structure critical appraisal rather than to generate a formal numerical risk-of-bias score. Particular attention was given to cohort design, sample and

event adequacy, separation of model-development and evaluation data, independent validation, calibration, decision utility, workflow integration, and clinical applicability.

Evidence-maturity grading

Each of the 31 direct AI studies was assigned to the highest cumulative evidence level for which all required criteria were fully satisfied.

Level 1 comprised model development and internal validation for the target clinical task. Internal random splitting, cross-validation, or an internal holdout set did not constitute independent Level 2 validation. Internal calibration, thresholds, or decision-curve analysis could not bypass the cumulative requirement for Level 2.

Level 2 required independent external, temporal, or prospective performance validation of a fixed model for the same clinical task. Evaluation data could not have been used for feature selection, parameter tuning, or operating-threshold selection. If model updating used the evaluation data, a separate independent evaluation was required. Multicenter data alone did not constitute external validation when all participating centers contributed to model development or tuning.

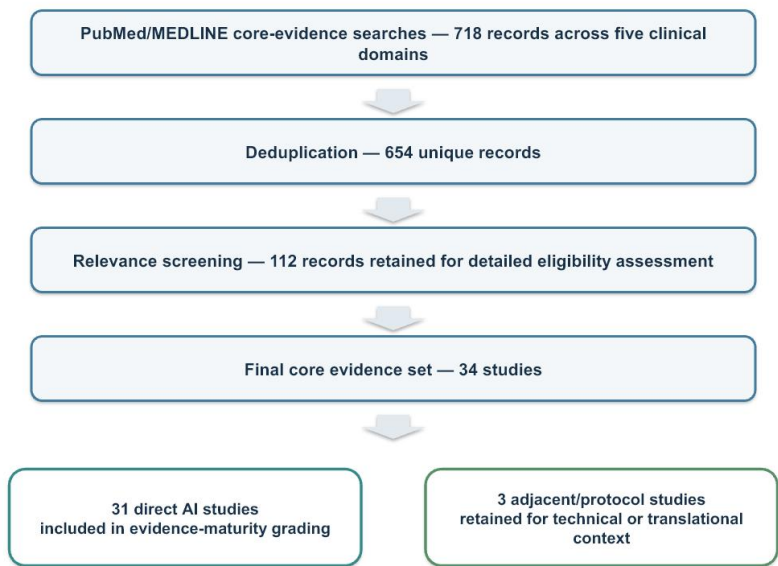
Level 3 required full satisfaction of Level 2 plus clinically interpretable decision-utility evidence appropriate to the intended clinical task and form of model output. For probabilistic risk-prediction models, calibration relevant to the intended use had to be demonstrated, and the risk output had to be linked to a defined clinical action through a prespecified or clinically interpretable threshold together with evidence of net benefit or incremental decision value over an appropriate clinical comparator. Calibration curves or decision-curve analysis alone were insufficient. For locked treatment-predictive classifiers, the classification rule and threshold had to be fixed before independent evaluation, and differential treatment value had to be demonstrated in independent data through a formal biomarker-treatment interaction test or an equivalent validated treatment-comparator analysis. Prognostic stratification, a significant association within a single treatment group, or significance in one treatment group but not another was insufficient to establish treatment-predictive utility. Level 3 did not imply validated prospective treatment guidance or improved patient outcomes.

Level 4 required prospective integration into a real clinical workflow with measured effects on clinical decisions, management, or workflow. Prospective performance validation alone did not satisfy Level 4.

Level 5 required demonstrated effects of an AI-assisted strategy on patient outcomes, resource use, efficiency, or safety.

Z. L. and G. P. jointly performed the study-level maturity assessment and evidence verification. Ambiguous or borderline classifications were reviewed with L.L. and Y.H. until a final classification was agreed. No inter-rater agreement statistic was calculated. Study-specific evidence supporting each maturity assignment is reported in Supplementary Table 2.

Supplementary Figure S1. Structured literature identification and core evidence-set selection



Supplementary Figure 1. Structured literature identification and core evidence-set selection. The structured PubMed/MEDLINE search yielded 718 records across five predefined clinical domains. After deduplication, 654 unique records remained, of which 112 were retained for detailed eligibility assessment. The final core evidence set comprised 34 studies, including 31 direct AI studies included in evidence-maturity grading and three adjacent or protocol studies retained for technical or translational context.

Supplementary Table 1. Characteristics and evidence appraisal of the 34 studies in the core evidence set

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Resectability/vascular involvement	[16]	Total $n = 2,130$: internal retrospective $n = 1,759$; external retrospective $n = 169$; prospective $n = 202$	3D deep learning vascular-invasion framework	Level 2. Yes - independent external validation across three hospitals and prospective performance evaluation in 202 patients; the model was fixed for evaluation	External accuracies exceeded 90% in several settings; prospective per-vessel accuracies about 87-90%; comparable/non-inferior to senior radiologists in several metrics	Strong adjunctive vascular quantification anchor	No replacement of radiologists/MDT; no proven R0, morbidity, or survival benefit
Resectability/vascular involvement	[17]	613 CT scans from 467 pancreatic tumor patients plus controls; balanced test set $n = 60$	Automated PDAC/vessel segmentation and tumor-vessel contact quantification	Level 1. No qualifying independent validation - model development and a 60-case test were	Vessel-involvement agreement 76%; arterial agreement 87%; venous agreement 59%; resectability classification	Automated quantification may standardize resectability assessment	Radiologist-assessment ground truth; weaker venous performance; no surgical outcome or MDT-impact proof

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Resectability/SMA margin	[14]	200 PDAC Whipple patients; 40 positive SMA margins	U-Net segmentation plus ResNet50 classification	reported, but the test source did not establish independent external validation. Level 1. No - single-center development with an internal holdout test only.	agreement 80% U-Net DSC 0.90; sensitivity 0.43; specificity 0.94; accuracy 0.85; F1 0.50	Feasibility with low-sensitivity warning	Not reliable for excluding positive margin or deciding therapy
Resectability/SMPV involvement	[18]	Bi-institutional retrospective PDAC cohorts; development $n = 173$, validation $n = 156$	CT radiomics risk score plus venous deformity	Level 2. Yes - the model was developed in one cohort and evaluated in an independent institutional	Validation AUC 0.928 versus radiologic evaluation AUC 0.768; sensitivity 84.2% versus 66.7%	Task-specific venous involvement risk estimation	Selected SMPV-suspected population; not all-vessel or general resectability decision evidence

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Resectability/human-machine fusion	[19]	Retrospective multicenter PC cohort; training $n = 263$, independent test $n = 86$	CECT ML models fused with NCCN-guideline expert assessment	validation cohort. Level 3. Independent testing plus calibration and explicit decision-utility evidence against appropriate comparators.	HMfML test AUC 0.8845, accuracy 82.56%, sensitivity 84.21%, specificity 82.09%	Human-machine fusion may complement guideline-based expert assessment	Manual segmentation; binary framing; no prospective decision-impact or outcome validation
Resectability/vascular involvement	[15]	87 PDAC patients; margin pathology in 61	Interpretable CAD tumor-vessel CT features	Level 1. No - single-center development and internal CAD evaluation only.	Vessel AUCs high; R0 AUC 0.87 without significant superiority	Adjunctive vascular assessment	No global expert superiority, radiologist replacement, or validated R0 decision benefit
Resectability/report structuring	[50]	Single-center retrospective; 180	GPT-3.5/GPT-4 free-text-to-structured reporting	Level 1. No - evaluation	GPT-4 report F1 0.997; surgeon	Supervised report	No source-image review or external

Decision node	Referen ce	Study design and population	Data/model type	Maturity level/validati on status	Key result	Interpretati on	Main methodological limitation/bounda ry
3D planning/segmentation	[21]	PDAC staging CT reports	Multiorgan auto- segmentation	used internal retrospective cases and an offline reader study, without independent external validation. Level 2. Yes - training and internal validation	classification 83% vs 76% with original reports; shorter review time	structuring and information- transfer support	validation; not autonomous resectability assessment
		275 pancreatic- cancer surgical patients; external validation $n = 40$		were followed by external validation using patients from two additional hospitals.	Overall DSC about 75%; pancreas DSC about 85- 86%; tumor DSC about 55-57%	AI-specific 3D visualization evidence	Not direct resectability stratification; tumor segmentation remains difficult
3D planning/SMA anatomy	[22]	124 pancreatic malignancy resections	Deep-learning virtual SMA modeling	Level 1. No qualifying independent	SMA branch/anatomy features associated	Anatomical quantificatio n and	No proof of improved operative approach, R0,

Decision node	Referen ce	Study design and population	Data/model type	Maturity level/validati on status	Key result	Interpretati on	Main methodological limitation/bounda ry
				external, temporal, or prospective performance validation of a fixed predictive model was established.	with POPF, severe complications, and blood loss	perioperative risk clue	survival, or outcomes
3D planning context	[23]	16 selected BR- PDAC patients	3D segmentation/rendering/pri nting; non-AI	Not graded. Small retrospective pilot	Printing took about 36 h and cost 700- 1000 euros; printed model added no extra information over segmentation	Adjacent 3D- planning context	Not AI model evidence; no routine 3D printing recommendation
Perioperative planning/major bleeding	[33]	Single-center retrospective resectable-PDAC cohort; $n = 175$; temporal test $n = 47$	Preoperative clinical variables; decision tree with neural-network comparator	Level 1. Unclear/not qualifying - a temporal split was reported, but public	Decision-tree accuracy 80.5% in training and 80.9% in testing; neural- network accuracy fell from 95.3% to	Exploratory bleeding-risk stratification and overfitting caution	Single-center retrospective; outcome included all abdominal fluid loss; no proof of reduced bleeding,

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Treatment/adjuvant stratification	[26]	343 resected PDAC RNAseq profiles from PRODIGE-24/CCTG PA6	AI-assisted transcriptomic drug-sensitivity signatures	code used validation data for parameter or random-seed selection and the mapping to the published test set could not be confirmed. Level 1.	54.1%		transfusion benefit, or improved survival
				Retrospective analysis of trial material; composite rule/cutoff was not independently locked and retested.	Matched group 47.8% had better DFS/CSS; mFFX matched DFS HR 0.31; GEM matched DFS HR 0.40	Predictive biomarker hypothesis for adjuvant stratification	Not deployment-ready chemotherapy assignment
Treatment/gemcitabine	[27]	350 resected PDAC	GemPred RNA signature	Level 3.	GemPred-positive	Predictive	No routine de-

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
sensitivity		RNAseq profiles from PRODIGE-24/CCTG PA6		Independent same-task validation plus clinically interpretable treatment-effect stratification meeting the cumulative Level 3 criteria.	in GEM arm: DFS 27.3 vs 10.2 mo, HR 0.43; CSS HR 0.42	biomarker hypothesis for GEM sensitivity	escalation or prospective treatment assignment
Treatment/advanced-PDAC treatment-selection evidence	[30]	477 advanced/unresectable PDAC patients; validation $n = 299$	H&E WSI computational histology biomarker	Direct AI study in the core evidence set; the biomarker and threshold were locked and validated in an	Biomarker predicted differential benefit from fluoropyrimidine-versus gemcitabine-based first-line therapy	AI-based treatment-selection evidence with indirect relevance to the perioperative surgical pathway.	Advanced/unresectable population; not direct neoadjuvant/adjuvant or resectable surgical-pathway evidence.

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
				independent cohort, with explicit biomarker-treatment interaction evidence.			
Treatment/prospective infrastructure	[31]	Planned 45 BR/LA PDAC after FOLFIRINOX	CT/MRI radiomics plus liquid biopsy multi-omics	Not graded. Protocol only	No completed results	Future prospective multimodal research design	No proven performance or clinical utility
Treatment/post-neoadjuvant ypN0 response	[28]	NCDB retrospective; 1053 cT1-4N0M0 PDAC patients completing NAC, radiotherapy, and surgery	LASSO logistic model using clinical variables	Level 1. No - random internal training and test split only.	AUC 0.718/0.725; predicted responders had ypN0 71.5% vs 29.7% and R0 87.3% vs 62.6%	Exploratory treatment-response stratification after completed multimodal therapy	Selected surgical cohort; no causal ENI/NAC benefit or pretreatment regimen selection
Treatment/fibrosis imaging biomarker	[29]	Multicenter retrospective;	AI pathology fibrosis quantification and CT	Level 2. Yes - the fibrosis-	External CT ensemble AUC	Adjacent imaging-	Unresectable treatment cohort;

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Treatment decisions/biological-risk stratification	[44]	resectable pathology cohorts $n = 361$ plus unresectable treatment cohort $n = 295$	radiomics	prediction task was tested in an independent external cohort. Level 1. External analysis addressed a different endpoint from the primary liver-metastasis task.	0.709; predicted high fibrosis associated with longer PFS/OS in the AG group	biomarker evidence linking fibrosis and treatment outcome	no causal AG benefit or perioperative treatment assignment
		MSK-MET training $n = 1,779$; MSK-IMPACT validation $n = 2,181$; resectable validation subgroup $n = 680$	Six-layer genomic deep-learning PanScore	Not graded. Prospective single-arm phase 1b/2	Resectable low/intermediate/high-risk median OS 35.4/29.0/17.9 months	Biological-risk stratification beyond radiographic resectability	Validation endpoint was OS, not future liver metastasis; no surgery or treatment-selection proof
Treatment/translational context	[32]	28 BR-PDAC patients	mFOLFIRINOX plus nivolumab trial; non-AI	Not graded. Prospective single-arm phase 1b/2	22/28 proceeded to surgery; R0 in 86% of resected	Neoadjuvant translational context	Not AI-guided treatment evidence or treatment recommendation
Complication/POPF after PD	[35]	8666 elective pancreatoduodenectomy	Preoperative-only ML	Level 1. No - NSQIP	XGBoost and random forest	Preoperative counseling/ri	Not deployment-ready drain,

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Complication/POPF digital pathology	[36]	omy patients; ACS-NSQIP	AI-assisted tissue composition quantification	development with cross-validation and an internal holdout only. Level 1. No - QuPath-based training with repeated elastic-net analysis and internal cross-validation only.	AUC 0.72	sk stratification	operative, or postoperative management
		134 H&E margin slides; 26 CR-POPF			Fibrotic tissue area associated with CR-POPF; ROC AUC about 0.58-0.73	Complement ary tissue quantification	Not standalone or real-time intraoperative prediction
Complication/postdischarge VTE	[38]	51,916 pancreatectomy patients; 743 pdVTE events (1.4%)	ML discharge-point risk model	Level 1. 80/20 stratified training-holdout internal validation	ROC-AUC 0.83 (95% CI, 0.80-0.87); AUPRC 0.34; threshold 0.10: sensitivity 39.6%, specificity 99.5%, PPV 51.8%, NPV	Key example that AUC does not equal clinical utility	Source-level sensitivity inconsistency is unexplained; no external/prospective validation or established benefit of model-guided

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Complication/POPF after distal pancreatectomy	[37]	230 radical distal pancreatectomy patients; small external test set	2D/3D body-composition plus clinical model	Level 2. Yes - internal tuning was followed by evaluation of the fixed model in an independent 42-patient cohort from another institution.	99.1%. The article text reports sensitivity 0.40%, whereas Supplement 2, Table 5 reports 39.6%.	Promising multimodal POPF prediction	prophylaxis.
					External combined model AUC 0.82; sensitivity 0.81; specificity 0.76		Not generalizable to all pancreatic operations
Prognosis/overall survival	[49]	1516 resected PDAC patients; external validation $n = 179$	Fully automated multiphase CT DL model	Level 2. Yes - multicenter independent external	Preoperative high-risk HR 2.03 internal and 2.47 external;	OS risk stratification after resection	No chemoradiotherapy selection or survival

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Prognosis/early recurrence	[42]	Multicenter retrospective PDAC surgery cohort; training $n = 320$, internal validation $n = 138$, external validation $n = 131$	Intratumoral/peritumoral radiomics, body composition, clinicopathological ML model	prognostic validation was reported. Level 1. External cohort reported; strict independence of the final model-selection process from that cohort was not fully verifiable.	postoperative HR 2.49 internal and 2.15 external Combined model external AUC 0.884; sensitivity 0.746; specificity 0.861; calibration/DCA and SHAP reported	Multimodal postoperative early-recurrence risk stratification	improvement proof Combined model is not purely preoperative; no treatment or surveillance decision proof
Prognosis/spatial immune pathology	[46]	Single-center retrospective; 304 upfront-surgery, R0, non-NAC PDAC patients	AI whole-slide TIL and immune-phenotype quantification	Level 1. No - a pre-existing Lunit SCOPE IO analyzer was applied in a single-center retrospective	Immune-inflamed phenotype and higher intratumoral TIL density associated with longer OS and RFS	Spatial pathology adds exploratory biological prognostic information	Selected cohort, one representative slide, no external validation or immunotherapy prediction

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
				cohort; no independent external, temporal, or prospective validation of the prognostic use was performed.			
Prognosis/CT feature fusion	[45]	Two independent resectable-PDAC hospital cohorts; training $n = 68$; test $n = 30$	CT radiomics and deep-learning feature fusion	Level 2. Yes - the model was tested in an independent hospital cohort.	Risk-score feature fusion yielded test AUC 0.84 versus 0.60 with principal component analysis	Complementary prognostic information from radiomics and transfer-learning features	Small retrospective cohorts; no treatment selection, surveillance threshold, decision-impact, or survival-benefit proof
Prognosis/LNM	[40]	Multiphase CT pancreatic-cancer LNM prediction study	Dual-transformation and contrastive-learning framework	Level 1. No - the pancreatic task had internal five-	Reported model AUC about 0.791 and accuracy about 0.744; public	Exploratory technical evidence for feature	Not deployable LNM prediction; not lymphadenectomy

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Prognosis/early recurrence	[41]	493 curative-intent resection patients; validation $n = 86$	CT deep-learning radiomics nomogram	fold cross-validation only; the purported external dataset addressed a non-pancreatic task. Level 2. Independent validation with calibration/DCA; no definitive clinical action threshold or appropriate current-care comparator.	external validation used non-pancreatic dataset Nomogram AUC 0.941 test and 0.920 validation; calibration/DCA reported	learning Early recurrence risk stratification	or pathology decision evidence No surveillance/adjuvant treatment threshold or outcome benefit

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
Prognosis/body-composition radiomics	[47]	Single-center retrospective; resected PDAC $n = 191$; temporal validation $n = 49$	L3 fat and muscle radiomics risk score	Level 2. Yes, temporal - the model was developed in the 2019-2022 cohort and independently tested in the earlier 2016-2018 cohort using the same threshold.	C-index 0.71/0.69; 12-month RFS AUC fell from 0.89 in training to 0.68 in validation	Host-state risk signal and temporal-generalization caution	No independent external validation or treatment/surveillance decision proof
Prognosis/occult liver metastasis	[43]	Five-center retrospective resected PDAC cohort $n = 438$; external $n = 103$	Preoperative CT radiomics with LASSO logistic regression	Level 2. Yes - independent external validation was performed across a five-center design.	AUC 0.73 training, 0.72 internal, and 0.71 external	Moderate biological-risk stratification beyond anatomic resectability	Manual segmentation; excluded NAT; cannot select NAT or cancel surgery
Prognosis/LNM	[39]	331 resectable PDAC patients; test	XGBoost using clinical, imaging, tumor markers	Level 1. No - data from two	Test AUC 0.795; sensitivity 95.3%;	Exploratory LNM risk	Not rule-in, not surgical extent

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
		<i>n</i> = 90		centers were pooled and then randomly split 3:1, constituting internal rather than independent external testing.	specificity 34.0%	stratification	decision, not pathology replacement
Prognosis/PET-CT overall survival	[48]	Single-center retrospective; 627 resectable PDAC patients undergoing preoperative 18F-FDG PET/CT	PET/CT tumor radiomics with LASSO	Level 1. No - single-center LASSO/Cox modeling with internal ten-fold cross-validation only.	Radiomics score independently associated with OS; HR 2.946; 12/24/36-month IAUC 0.729/0.710/0.705	PET/CT metabolic heterogeneity adds exploratory prognostic information	No external validation or treatment/follow-up decision proof
Resectability/vascular invasion	[20]	Retrospective multicenter; 5,930 tumor-vessel cases across six centers,	Routine venous-phase CT; clinician-centric invasion model	Level 1. External performance cohorts from	Overall external AUC 0.943 (95% CI 0.932-0.952); center AUCs	External-cohort discrimination was	No prospective decision-impact evaluation; no proof of improved

Decision node	Reference	Study design and population	Data/model type	Maturity level/validation status	Key result	Interpretation	Main methodological limitation/boundary
		including 3,575 cases in external cohorts from five centers		five centers were reported, but strict model-locking/data-isolation requirements for Level 2 were not fully verifiable.	0.939-0.991	reported, but strict Level 2 model-locking and data-isolation criteria could not be fully verified	resectability decisions, R0, morbidity, or survival.

The core evidence set comprised 31 direct AI studies and three adjacent or protocol studies retained to clarify technical boundaries or emerging translational pathways. Only the 31 direct AI studies were included in evidence-maturity grading.

Supplementary Table 2. Evidence-maturity mapping for the 31 direct AI studies

Reference	Clinical task	External/temporal/prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[16]	Preoperative CT-based quantification of multivessel	Yes - independent external validation across three hospitals and	Not reported as probability calibration	Radiologist performance was compared, but no clinically actionable	No - prospective performance testing did not place the model in a live	No AI-assisted patient, resource-use, or safety outcomes were	L2	Independent external and prospective performance

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[17]	invasion	prospective performance evaluation in 202 patients; the model was fixed for evaluation		threshold, DCA, net benefit, or complete same-task decision- utility chain was reported	clinical workflow or measure changes in decisions	evaluated		validation satisfies L2. Evidence does not extend to clinically interpretable decision utility or workflow impact
	Automated PDAC and vessel segmentation with tumor- vessel contact quantification	No qualifying independent validation - model development and a 60-case test were reported, but the test source did not establish independent external validation	Not applicable/not reported	Manual image interpretation served as a reference, but no actionable threshold, DCA, net benefit, or incremental clinical decision value was assessed	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	Development and internal testing support L1; independent external, temporal, or prospective performance validation was not demonstrated
	Prediction of positive superior mesenteric artery	No - single-center development with an internal holdout	No qualifying calibration analysis	Surgeon/radiologist comparison was reported, but no	No prospective workflow integration or	No AI-assisted patient, resource- use, or safety	L1	The study contains model development and

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[18]	margin after Whipple resection	test only	reported	actionable threshold, DCA, net benefit, or incremental decision value was established. DCA and comparison with radiologic assessment were reported; however, the 0.4 cutoff was a Youden-optimized statistical point rather than a prespecified clinical action threshold, and no executable decision rule was established	decision-impact evaluation	outcomes were evaluated.		internal holdout testing, but no independent validation
	CT radiomics prediction of superior mesenteric-portal vein involvement	Yes - the model was developed in one cohort and evaluated in an independent institutional validation cohort	Yes - calibration was reported in the validation cohort		No prospective workflow integration or measured decision impact	No AI-assisted patient, resource-use, or safety outcomes were evaluated	L2	Independent validation satisfies L2. Calibration and DCA alone did not establish clinically interpretable decision utility under the strict L3 definition
	Human-machine fusion classification of	Yes - training used centers 1 and 2, with independent testing	Yes - calibration was evaluated	Yes - clinically interpretable resectability	No - the model was not prospectively integrated into	No AI-assisted patient, resource-use, or safety	L3	The study satisfies L2 and adds calibration

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[15]	pancreatic cancer resectability	in centers 3 and 4 after model development	in the independent test set	probability ranges, DCA/net benefit, NCCN-guideline expert assessment and CT-signature comparators, and incremental same- task value were evaluated in independent testing Internal DCA and radiologist	routine workflow and actual decisions were not measured	outcomes were evaluated		plus explicit, clinically interpretable decision-utility evidence against appropriate comparators; it therefore meets L3 but not L4
	CT-based assessment of vascular invasion and R0 resection risk	No - single-center development and internal CAD evaluation only	Yes, but only in the internal evaluation	comparison were reported; threshold interpretation and incremental decision value remained unclear, and L2 was not met	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	Internal calibration or DCA cannot bypass the cumulative requirement for independent L2 validation
[50]	LLM conversion of pancreatic cancer staging	No - evaluation used internal retrospective cases	Not applicable/not reported	An offline surgeon reader study showed internal information-	No - the reader experiment was retrospective and	No patient or safety outcomes were evaluated;	L1	Internal model comparison and offline reader

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[21]	CT reports into structured resectability information	and an offline reader study, without independent external validation	Not applicable to the segmentation task	transfer gains. The 85% value was a tool-performance target, not a clinical action threshold; no DCA or net benefit was reported	offline, with no live workflow integration or measured management change	only review time was examined as an internal process measure	L2	testing support L1 but do not establish independent validation or prospective workflow impact
	Automated multiorgan and pancreatic tumor segmentation for surgical planning	Yes - training and internal validation were followed by external validation using patients from two additional hospitals		Manual segmentation and qualitative assessment were comparators, but downstream surgical decision utility, DCA, net benefit, and actionable thresholds were not evaluated	No prospective workflow integration or measured surgical decision impact	No AI-assisted patient, resource-use, or safety outcomes were evaluated		Independent external performance validation satisfies L2; the study did not assess downstream decision value
[22]	Deep-learning virtual superior mesenteric artery modeling for	No qualifying independent validation. The study	Not applicable/not reported for the anatomical	Anatomical features were associated with perioperative outcomes, but no	No prospective workflow integration or measured change in	Perioperative associations were observed, but effects of an AI-	L1	Retrospective deep-learning virtual SMA modeling in 124

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
	anatomical risk stratification in pancreas surgery	retrospectively analyzed 124 pancreatic resections from two institutions using deep-learning virtual SMA modeling; no external, temporal, or prospective validation chain of a fixed predictive model was established.	modeling task	clinical action threshold, DCA, net benefit, or incremental decision-utility analysis was reported	operative planning or decisions	assisted strategy on patient, resource-use, or safety outcomes were not tested.		pancreatic resections supports task-specific AI anatomical modeling, but no qualifying independent validation, actionable decision rule, prospective workflow integration, or AI-assisted outcome benefit was established
[33]	Preoperative prediction of major intraoperative bleeding	Unclear/not qualifying - a temporal split was reported, but public code used validation	No qualifying calibration analysis reported	Tree-model cut points were present, but their independent locking was unclear; no	No prospective workflow integration or measured decision impact	No AI-assisted patient, resource-use, or safety outcomes were evaluated	L1	Development and internal/temporal evaluation support L1, but

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[26]	Pancreas- View/drug- sensitivity signatures for adjuvant chemotherapy sensitivity	data for parameter or random-seed selection and the mapping to the published test set could not be confirmed	Not applicable/not reported for the signature rule.	DCA, net benefit, appropriate clinical comparator, or incremental decision value was reported	No prospective AI- guided treatment assignment or measured decision impact.	Treatment outcomes were analyzed retrospectively, but effects of an AI-guided strategy on patient outcomes, resources, or safety were not tested	L1	the independence needed for L2 could not be verified
		No qualifying independent validation - the composite modified- FOLFIRINOX cutoff was optimized in the same PRODIGE-24 clinical evaluation cohort		Treatment arms were compared and interaction analyses were reported, but the composite rule and cutoff were not independently locked and retested; no DCA or validated incremental decision rule was established				Because the evaluation data contributed to cutoff selection, the strict L2 data-isolation criterion was not fully satisfied.
[27]	GemPred prediction of differential	Yes - a previously fixed GemPred signature was	Not applicable to the locked binary	Yes - the fixed GemPred classification was	No prospective AI- guided treatment workflow or	Trial outcomes and subgroup toxicity were	L3	The fixed biomarker was independently

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[30]	benefit from adjuvant gemcitabine	retrospectively evaluated in independent randomized-trial material	biomarker rule	evaluated against randomized gemcitabine and modified- FOLFIRINOX arms with a prespecified biomarker-treatment interaction, providing explicit differential treatment-value evidence Yes - the locked	measured clinical decision impact	reported retrospectively; the study did not test patient, resource, or safety effects of an AI-guided treatment strategy		evaluated and the treatment interaction directly assessed decision utility, satisfying L3; no prospective deployment was tested
	CHAI pathology biomarker prediction of differential benefit from first-line chemotherapy	Yes - the biomarker and threshold were locked and validated in an independent cohort; prospective- source cohort material was evaluated offline	Not applicable to the locked binary treatment biomarker	gemcitabine- versus fluoropyrimidine- based comparator, validation-cohort treatment interaction, and target-trial emulation provided	No - the biomarker was applied offline and did not guide treatment in a prospective workflow	No patient, resource, or safety effect of an AI-guided treatment strategy was tested	L3	The study satisfies independent validation and explicit treatment decision utility. Its evidence is indirect for the perioperative

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[28]	Prediction of ypN0 response after completed neoadjuvant chemoradiothera py	No - random internal training and test split only	Yes, but only internally	explicit differential treatment-value evidence	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	surgical pathway because the population was advanced/unrese ctable PDAC
				An internal statistical threshold was used; no DCA, net benefit, appropriate clinical comparator, or validated incremental decision utility was established				Model development and internal testing satisfy L1; no independent validation was performed
[29]	CT-based prediction of tumor fibrosis	Yes - the fibrosis- prediction task was tested in an independent external cohort	Unclear/not sufficiently reported	A training-cohort fibrosis threshold was applied, but no DCA was reported. Treatment associations came from a different	No prospective workflow integration or measured treatment decision impact	No patient, resource, or safety effect of an AI-guided strategy was tested	L2	Independent validation supports L2 for the fibrosis- prediction task only; cross-task treatment

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
				task/population and could not be combined with the fibrosis validation to claim L3				associations do not establish L3
[44]	PanScore assessment of liver-metastasis propensity/biolo gical prognosis	No for the primary task - the external cohort evaluated overall survival/stage rather than the liver- metastasis task	No qualifying same-task calibration established	No clinically interpretable threshold, DCA, net benefit, appropriate same-task comparator, or incremental decision value was established	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	The external analysis addressed a different endpoint, so same-task independent validation required for L2 was not met Development and internal evaluation support L1; there was no independent validation
[35]	Prediction of postoperative pancreatic fistula after pancreatoduoden ectomy	No - NSQIP development with cross-validation and an internal holdout only	Unclear/not sufficiently reported	Only limited model/existing-risk- information comparisons were available; no validated action threshold, DCA, net	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[36]	Digital- pathology tissue composition assessment for postoperative fistula risk	No - QuPath-based training with repeated elastic-net analysis and internal cross-validation only	No qualifying calibration analysis reported	benefit, or incremental clinical decision value was established Surgical palpation/tissue composition provided context, but no clinical action threshold, DCA, net benefit, or incremental decision value was established	No prospective workflow integration or decision-impact evaluation; the tissue input was postoperative	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	The study reports development and internal validation only
	Prediction of postdischarge venous thromboembolis m after pancreatectomy	No - an 80/20 internal holdout was used without external, temporal, or prospective validation	Unclear/not sufficiently reported	At the 0.10 threshold, sensitivity was 39.6%, specificity 99.5%, PPV 51.8%, and NPV 99.1%; this operating point missed most pdVTE	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	Internal validation supports Level 1 only. The threshold- specific metrics do not establish decision utility,

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[37]	Prediction of postoperative fistula after distal pancreatectomy	Yes - internal tuning was followed by evaluation of the fixed model in an independent 42- patient cohort from another institution	No qualifying calibration analysis was identified	events. No decision- curve analysis, net benefit, or clinical comparator was reported. The article text separately reports sensitivity 0.40%, creating an unresolved source- level discrepancy A Youden cutoff, DCA, and single- modality/clinical- model comparisons were reported, but clinically interpretable incremental decision utility was not established	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L2	and the discrepancy between the article text and Supplement 2, Table 5 remains unexplained
				Training-derived median/quantile risk				Independent external testing satisfies L2. Missing calibration and insufficient decision-utility evidence prevent L3
[49]	DeepCT-PDAC prediction of	Yes - multicenter independent	Unclear/not sufficiently	Training-derived median/quantile risk	No prospective workflow	No AI-assisted patient, resource-	L2	Independent external

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[42]	overall survival after resection	external prognostic validation was reported	reported for the decision- utility framework	groups and TNM/clinical-factor comparators showed prognostic increment, but no executable decision threshold, DCA, net benefit, or management decision utility was evaluated	integration or decision-impact evaluation	use, or safety outcomes were evaluated		prognostic validation satisfies L2; prognostic increment alone did not meet the strict L3 decision-utility requirement
	Prediction of early recurrence after PDAC resection	Unclear/not fully qualifying - an external cohort existed, but isolation of the final algorithm/model- selection process from that cohort could not be confirmed	Yes - calibration was reported	DCA was reported and a self-built clinical model served as comparator, but the 0.450 threshold was statistical rather than an action rule and the comparator's representation of current care was	No prospective workflow integration or measured decision impact	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	Because the cumulative L2 model- locking/data- isolation criterion was not fully satisfied, calibration and DCA could not be used to skip directly to L3

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[46]	AI-derived spatial TIL/immune- phenotype assessment and association with OS and RFS	No - a pre-existing Lunit SCOPE IO analyzer was applied in a single- center retrospective cohort; no independent external, temporal, or prospective validation of the prognostic use was performed	Not applicable/not reported	unclear.				This boundary study applied an existing AI analyzer to explore prognostic associations. It was retained among the direct AI studies and classified as Level 1 because no qualifying independent validation, decision utility, workflow integration, or AI-assisted outcome benefit was
				No clinical action threshold, DCA, net benefit, appropriate clinical decision comparator, or incremental decision-value analysis was reported	No prospective workflow integration or measured management impact	OS and RFS associations were observed, but the study did not evaluate effects of an AI-assisted strategy on patient outcomes, resource use, or safety	L1	

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
								demonstrated
[45]	CT radiomics/deep- learning feature fusion for overall-survival prediction	Yes - the model was tested in an independent hospital cohort	Unclear/not sufficiently reported	Training-derived risk strata were used, but comparisons mainly concerned feature- fusion/dimensionality- reduction methods; no DCA, net benefit, or appropriate clinical comparator was evaluated	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L2	Independent- hospital testing satisfies L2; decision-value evidence was insufficient for L3
[40]	Multiphase CT contrastive- learning prediction of lymph-node metastasis	No - the pancreatic task had internal five-fold cross- validation only; the purported external dataset addressed a non-pancreatic task	No qualifying calibration analysis reported	No clinical action threshold, DCA, net benefit, appropriate clinical comparator, or incremental decision value was established	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	Same-task independent validation was absent; non- pancreatic external data could not satisfy L2 for pancreatic LNM prediction

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[41]	Preoperative CT deep-learning radiomics prediction of early recurrence	Yes - an 86-patient independent validation cohort was not used in preprocessing or tuning	Yes - calibration was reported in validation	DCA was reported against treat- all/treat-none, but the paper explicitly did not establish a definitive clinical threshold and did not compare against a current clinical decision standard; incremental decision value remained unproven	No prospective workflow integration or measured surveillance/treatme nt decision impact	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L2	Independent validation satisfies L2. Calibration and DCA without a clinical threshold or appropriate comparator did not meet strict L3
[47]	Body- composition radiomics prediction of recurrence-free survival	Yes, temporal - the model was developed in the 2019-2022 cohort and independently tested in the earlier 2016-2018 cohort using the same threshold	Unclear/not sufficiently reported	The development threshold was carried into temporal validation and a clinical model was compared, but no DCA, net benefit, or executable treatment/follow-up	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L2	Independent temporal validation satisfies L2. The direction must be described as testing in an earlier historical period; decision

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[43]	Preoperative CT prediction of occult liver metastasis	Yes - independent external validation was performed across a five-center design	Reported, but its adequacy for clinical decision use was unclear	decision utility was evaluated DCA was reported/partly documented and clinical- factor/imaging models were compared, but no clinically actionable surgery or neoadjuvant-therapy threshold or incremental decision rule was established	No prospective workflow integration or measured surgical/treatment decision impact	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L2	utility was not established Independent external validation satisfies L2; uncertainty around calibration/DCA and absence of an actionable decision rule prevent L3
		No - data from two centers were pooled and then randomly split 3:1, constituting internal rather than independent	No qualifying calibration analysis established	An internal classification threshold was used, but no DCA, net benefit, or appropriate clinical comparator was	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	The pooled random split supports internal validation only

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
[48]	Preoperative PET/CT radiomics prediction of overall survival	external testing No - single-center LASSO/Cox modeling with internal ten-fold cross-validation only	No qualifying calibration analysis reported	evaluated; low specificity/positive predictive value limited rule-in use Conventional PET parameters were compared and internal prognostic increment was reported, but no independent validation, clinical action threshold, DCA, or decision- utility evaluation was performed	No prospective workflow integration or decision-impact evaluation	No AI-assisted patient, resource- use, or safety outcomes were evaluated	L1	Internal model development and cross-validation support L1 only
	Routine venous- phase CT classification of tumor-vessel invasion	External and temporal performance cohorts were reported, including five external centers;	No calibration analysis was reported in the checked main article	No prespecified/clinical ly interpretable threshold, DCA, or net benefit was reported. An offline	No - the reader study was retrospective/offline and did not measure changes in real management	No AI-assisted patient, resource- use, or safety outcomes were evaluated; prospective	L1	Model development and internal evaluation fully support L1. Although

Reference	Clinical task	External/temporal/ prospective performance validation	Calibration	Decision utility/DCA/net benefit/comparator	Workflow/decision impact	Patient/resource /safety outcomes	Final level	Rationale/notes
		however, the main article did not verify that model weights, feature selection, hyperparameter tuning, and operating-threshold selection were fully isolated from all evaluation data		reader study assessed accuracy and agreement, not decision-level incremental utility		workflow and patient-outcome effects were stated as future work		external cohorts were present, strict L2 model- locking/data- isolation requirements could not be verified from the supplied main article; unknown items were not treated as satisfied

Each study was assigned to the highest cumulative evidence level for which all required criteria were fully satisfied. The final distribution was Level 1, $n = 18$; Level 2, $n = 10$; Level 3, $n = 3$; and Levels 4-5, $n = 0$. Components classified as unclear were not treated as satisfying the corresponding criterion.

References

- [14] Wang J, Ashraf Ganjouei A, Romero-Hernandez F, et al. A Computer Vision Algorithm to Predict Superior Mesenteric Artery Margin Status for Patients With Pancreatic Ductal Adenocarcinoma. *Ann Surg*. 2026;283(5):834–840. [DOI:10.1097/sla.0000000000006506]
- [15] Zhao T, Wen Y, Chen Y, et al. Utility of a CAD model for estimation of vascular involvement and R0 resection in pancreatic ductal adenocarcinoma. *Int J Surg*. 2025;111(11):7829–7841. [DOI:10.1097/js9.0000000000003048]
- [16] Zhang Y, Zhang H, Yang Y, et al. A clinically validated 3D deep learning approach for quantifying vascular invasion in pancreatic cancer. *NPJ Digit Med*. 2026;9(1):79. [DOI:10.1038/s41746-025-02260-3]
- [17] Bereska JJ, Janssen BV, Nio CY, et al. Artificial intelligence for assessment of vascular involvement and tumor resectability on CT in patients with pancreatic cancer. *Eur Radiol Exp*. 2024;8(1):18. [DOI:10.1186/s41747-023-00419-9]
- [18] Zhou Y, Wang J, Zhang SL, et al. A CT Radiomics-Based Risk Score for Preoperative Estimation of Intraoperative Superior Mesenteric-Portal Vein Involvement in Pancreatic Ductal Adenocarcinoma. *Ann Surg Oncol*. 2023;30(2):1206–1216. [DOI:10.1245/s10434-022-12664-0]
- [19] Yimamu A, Li J, Zhang H, et al. Computed tomography and guidelines-based human-machine fusion model for predicting resectability of the pancreatic cancer. *J Gastroenterol Hepatol*. 2024;39(2):399–409. [DOI:10.1111/jgh.16401]
- [20] Guo R, Zhang M, Zhu H, et al. A clinician-centric intelligent method towards reliable pancreatic cancer vascular invasion assessment: a retrospective, multi-centre study. *Lancet Reg Health West Pac*. 2026;71:101890. [DOI:10.1016/j.lanwpc.2026.101890]
- [21] Rhu J, Oh N, Choi GS, et al. 3D auto-segmentation of pancreas cancer and surrounding anatomical structures for surgical planning. *Int J Surg*. 2025;111(10):6715–6724. [DOI:10.1097/js9.0000000000002835]
- [22] Mellado S, Vega EA, Yamane K, et al. Deep-Learning Virtual Superior Mesenteric Artery Modeling for Risk Stratification in Pancreas Surgery. *Ann Surg Oncol*. 2026;33(2):1627–1635. [DOI:10.1245/s10434-025-18543-8]
- [23] Sguinzi R, Vidal J, Poroës F, et al. 3D modeling to predict vascular involvement in resectable pancreatic adenocarcinoma. *Heliyon*. 2025;11(1):e41473. [DOI:10.1016/j.heliyon.2024.e41473]
- [26] Fraunhofer N, Hammel P, Conroy T, et al. Development and validation of AI-assisted transcriptomic signatures to personalize adjuvant chemotherapy in patients with pancreatic ductal adenocarcinoma. *Ann Oncol*. 2024;35(9):780–791. [DOI:10.1016/j.annonc.2024.06.010]
- [27] Nicolle R, Bachet JB, Harlé A, et al. Prediction of Adjuvant Gemcitabine Sensitivity in Resectable Pancreatic Adenocarcinoma Using the GemPred RNA Signature: An Ancillary Study of the PRODIGE-24/CCTG PA6 Clinical Trial. *J Clin Oncol*. 2024;42(9):1067–1076. [DOI:10.1200/jco.22.02668]
- [28] Harada GK, Park J, Peterson N, et al. Predicting Benefit of Neoadjuvant Chemotherapy and Elective Nodal Irradiation in Pancreatic Adenocarcinoma: A Supervised Machine Learning Approach. *Cancer Med*. 2025;14(23):e71447. [DOI:10.1002/cam4.71447]

- [29] Yang Q, Mao Y, Han Y, et al. Noninvasive Computed Tomography-Based Quantification of Tumor Fibrosis Predicts Pancreatic Cancer Response to Gemcitabine/Nab-Paclitaxel. *Research (Wash D C)*. 2025;8:0937. [DOI:10.34133/research.0937]
- [30] Hendifar AE, Krishna V, Krishna V, et al. Development and Validation of a Computational Histology Artificial Intelligence-Powered Predictive Biomarker for Selection of Chemotherapy in Advanced Pancreatic Cancer. *J Clin Oncol*. 2026;44(26):2496–2506. [DOI:10.1200/JCO-25-02199]
- [31] Abreu de Carvalho LF, Gryspeerdt F, Ceelen W, et al. Prediction of surgical resectability after FOLFIRINOX chemotherapy for borderline resectable and locally advanced pancreatic cancer (PeRFormanCe): a multicenter prospective trial - trial protocol. *BMC Surg*. 2025;25(1):204. [DOI:10.1186/s12893-025-02938-1]
- [32] Wainberg ZA, Link JM, Premji A, et al. Neoadjuvant modified FOLFIRINOX plus nivolumab in borderline-resectable pancreatic ductal adenocarcinoma: a pilot phase 1 trial. *Nat Commun*. 2026;17(1):2232. [DOI:10.1038/s41467-026-68976-2]
- [33] Wakiya T, Ishido K, Kimura N, et al. Prediction of massive bleeding in pancreatic surgery based on preoperative patient characteristics using a decision tree. *PLoS One*. 2021;16(11):e0259682. [DOI:10.1371/journal.pone.0259682]
- [35] Ashraf Ganjouei A, Romero-Hernandez F, Wang JJ, et al. A Machine Learning Approach to Predict Postoperative Pancreatic Fistula After Pancreaticoduodenectomy Using Only Preoperatively Known Data. *Ann Surg Oncol*. 2023;30(12):7738–7747. [DOI:10.1245/s10434-023-14041-x]
- [36] Målyi A, Bronsert P, Schilling O, et al. Postoperative pancreatic fistula risk assessment using digital pathology based analyses at the parenchymal resection margin of the pancreas - Results from the randomized multicenter RECOPANC trial. *HPB (Oxford)*. 2025;27(3):393–401. [DOI:10.1016/j.hpb.2024.12.012]
- [37] Liang K, Miao Q, Jing Y, et al. Integrative deep learning analysis of 2D and 3D body composition features for predicting postoperative pancreatic fistula after distal pancreatectomy. *Ann Med*. 2025;57(1):2597067. [DOI:10.1080/07853890.2025.2597067]
- [38] Ahmed KS, Issaka SM, Marcinak CT, et al. Machine Learning-Driven Modeling to Predict Postdischarge Venous Thromboembolism After Pancreatectomy for Pancreas Cancer. *Ann Surg Oncol*. 2025;32(6):4085–4093. [DOI:10.1245/s10434-025-17032-2]
- [39] Daibo S, Homma Y, Ohya H, et al. Novel machine-learning model for predicting lymph node metastasis in resectable pancreatic ductal adenocarcinoma. *Ann Gastroenterol Surg*. 2025;9(1):161–168. [DOI:10.1002/ags3.12836]
- [40] Chen X, Wang W, Jiang Y, Qian X. A dual-transformation with contrastive learning framework for lymph node metastasis prediction in pancreatic cancer. *Med Image Anal*. 2023;85:102753. [DOI:10.1016/j.media.2023.102753]
- [41] Guan X, Liu J, Xu L, Jiang W, Wang C. A CT-Based Deep Learning Radiomics Nomogram for Early Recurrence Prediction in Pancreatic Cancer: A Multicenter Study. *Ann Surg Oncol*. 2025;32(10):7567–7576. [DOI:10.1245/s10434-025-17748-1]

- [42] Wu L, Cen C, Ouyang D, et al. Interpretable machine learning model for predicting early recurrence of pancreatic cancer: integrating intratumoral and peritumoral radiomics with body composition. *Int J Surg*. 2025;111(11):8198–8211. [DOI:10.1097/js9.00000000000003078]
- [43] Zhao B, Xia C, Xia T, et al. Development of a radiomics-based model to predict occult liver metastases of pancreatic ductal adenocarcinoma: a multicenter study. *Int J Surg*. 2024;110(2):740–749. [DOI:10.1097/js9.0000000000000908]
- [44] Zheng S, Liu Y, Troy Teo P, et al. A deep learning model for preoperative risk stratification of pancreatic ductal adenocarcinoma based on genomic predictors of liver metastasis. *Eur J Cancer*. 2025;226:115608. [DOI:10.1016/j.ejca.2025.115608]
- [45] Zhang Y, Lobo-Mueller EM, Karanicolas P, Gallinger S, Haider MA, Khalvati F. Improving prognostic performance in resectable pancreatic ductal adenocarcinoma using radiomics and deep learning features fusion in CT images. *Sci Rep*. 2021;11(1):1378. [DOI:10.1038/s41598-021-80998-y]
- [46] Kim H, Choi JH, Lim Y, et al. Artificial Intelligence-Powered Spatial Analysis of Immune Phenotypes in Resected Pancreatic Cancer. *JAMA Surg*. 2025;160(8):884–892. [DOI:10.1001/jamasurg.2025.1999]
- [47] Gu Q, Liu P, Hu X, Liu J, He Y. Development and validation of a CT-measured body composition radiomics model for prognostic assessment in resected pancreatic adenocarcinoma. *Sci Rep*. 2025;15(1):28722. [DOI:10.1038/s41598-025-14397-y]
- [48] Yoo J, Hyun SH, Lee J, et al. Prognostic Significance of ¹⁸F-FDG PET/CT Radiomics in Patients With Resectable Pancreatic Ductal Adenocarcinoma Undergoing Curative Surgery. *Clin Nucl Med*. 2024;49(10):909–916. [DOI:10.1097/RLU.00000000000005363]
- [49] Yao J, Cao K, Hou Y, et al. Deep Learning for Fully Automated Prediction of Overall Survival in Patients Undergoing Resection for Pancreatic Cancer: A Retrospective Multicenter Study. *Ann Surg*. 2023;278(1):e68–e79. [DOI:10.1097/sla.00000000000005465]
- [50] Bhayana R, Nanda B, Dehkharghanian T, et al. Large Language Models for Automated Synoptic Reports and Resectability Categorization in Pancreatic Cancer. *Radiology*. 2024;311(3):e233117. [DOI:10.1148/radiol.233117]