



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

Zhike Li[#] , Gongze Peng[#], Liping Liu^{ID} , Yongzhu He^{ID}

Keywords:

Artificial intelligence, pancreatic ductal adenocarcinoma, surgical decision-making, resectability, perioperative risk, clinical translation

Citation: Li Z, Peng G, Liu L, He Y. Artificial intelligence in pancreatic cancer surgery: evidence maturity, boundaries, and clinical translation. *Art Int Surg.* 2026;6:556-78. <https://dx.doi.org/10.20517/ais.2026.75>

Received: 21 July 2026

First Decision: 19 Aug 2026

Revised: 4 Oct 2026

Accepted: 9 Oct 2026

Published: 27 Oct 2026

Academic Editor:

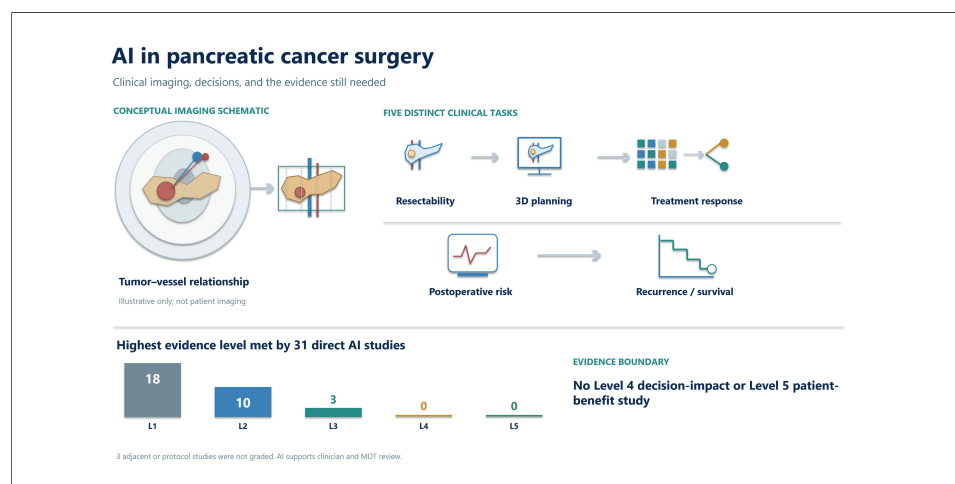
Andrew Gumbs

Copy Editor:

Tong Wang

Production Editor:

Tong Wang



Abstract

Pancreatic cancer surgery requires decisions about resectability, treatment sequencing, operative planning, complications, and prognosis. We mapped artificial intelligence (AI) studies to five surgical decision domains and evaluated study design, validation, performance, decision utility, workflow fit, and evidence limitations. This structured narrative review included 31 direct AI studies and three adjacent or protocol studies; only direct AI studies underwent evidence-maturity grading. Eighteen studies met Level 1, ten Level 2, and three Level 3; none reached Levels 4 or 5. Three-dimensional assessment of vascular involvement has undergone external and prospective performance validation, but its effect on surgical decisions or patient outcomes remains unproven. Automated segmentation can aid anatomical visualization, although tumor delineation and workflow demands limit use. Transcriptomic analyses of trial samples suggest differential adjuvant treatment benefit but do not establish prospective biomarker-guided treatment allocation. Models for complications, recurrence, and survival offer risk-stratification signals, yet calibration, actionable thresholds, class imbalance, and clinical utility remain concerns. AI should support radiologists, surgeons, and multidisciplinary teams rather than replace their



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

judgment. Clinical adoption requires robust external validation, prospective workflow assessment, and evidence that AI-assisted decisions improve patient outcomes.

INTRODUCTION

Pancreatic cancer has one of the poorest prognoses among gastrointestinal malignancies, with persistently high mortality and low rates of long-term survival^[1,2]. Surgery remains central to long-term survival for patients with resectable disease and for those who become candidates for resection after systemic treatment. Surgical care entails several interconnected decisions: assessing resectability and vascular involvement, sequencing treatment, planning the operation, managing perioperative risk, and evaluating postoperative prognosis^[1]. These decisions increasingly depend on combining imaging findings, pathology, clinical variables, and multidisciplinary expertise. Each offers a potential opportunity for artificial intelligence (AI) to support clinical judgment.

AI Applications in pancreatic cancer include image analysis, automated segmentation, radiomics, analysis of digital pathology, omics methods, and prediction of prognosis. Much of the wider literature has emphasized early detection and general diagnostic or therapeutic uses, including large-scale deep-learning studies based on computed tomography (CT) and comprehensive reviews^[3-6]. For surgical use, however, discrimination is only one requirement. Clinical relevance also depends on suitability for the intended population, external or prospective validation, calibration, actionable thresholds, integration into workflows, and, ultimately, the consequences of decisions made with model assistance.

This review follows successive surgical decision points throughout the pancreatic cancer care pathway, whereas earlier reviews have generally been organized by AI technique, data modality, or broad diagnostic and therapeutic application. Alongside model performance, we assess the maturity and limitations of the evidence, emphasizing external validation, clinical actionability, workflow integration, and the steps needed to turn predictions into usable decision support. [Figure 1](#) outlines this approach, connecting the main AI methods with five surgical decision nodes and using a consistent framework to assess evidence maturity across applications.

METHODS

Review design

We organized the review around successive decisions in pancreatic cancer surgery and examined five prespecified clinical domains: preoperative evaluation of resectability and vascular involvement; three-dimensional reconstruction and surgical planning; stratification for neoadjuvant and adjuvant treatment; management of perioperative complication risk; and models predicting lymph node metastasis, recurrence, and survival.

This approach enabled comparison of evidence maturity and limits to clinical use across tasks without combining findings into a single effect estimate. We adopted a structured narrative design and synthesized evidence according to the surgical decision being addressed. The review was not intended as a systematic review or meta-analysis; accordingly, we did not pool effect estimates or report according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. Our synthesis examined individual study designs and findings alongside similarities and differences in populations, data modalities, validation levels, and intended clinical applications.

PubMed/MEDLINE served as the main bibliographic database. The core-evidence search covered publications from 1 January 2020 to 28 February 2026. We performed a structured search covering five

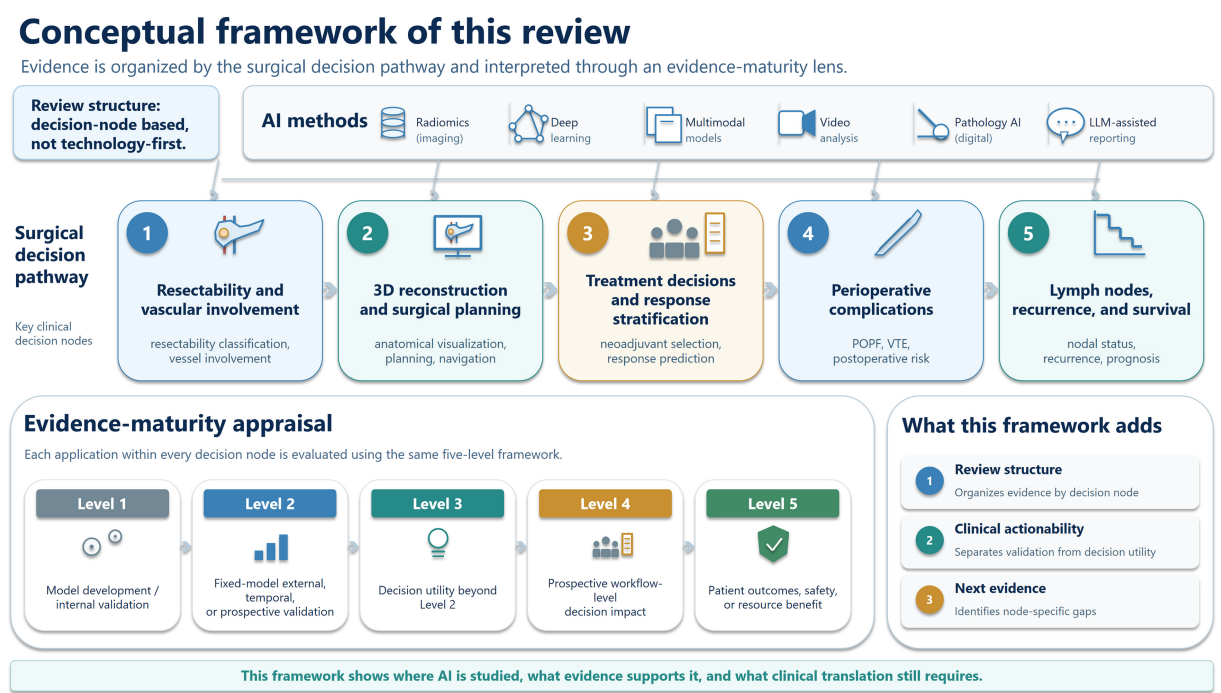


Figure 1. Conceptual framework of this review. Evidence is grouped into five surgical decision nodes; each application is assessed using the same cumulative five-level evidence-maturity framework. Source: Original diagram prepared for this review using native vector shapes. No third-party icons or image assets were incorporated. AI: Artificial intelligence.

predefined domains within surgical care for pancreatic cancer: evaluation of resectability and vascular involvement; three-dimensional reconstruction, segmentation, and operative planning; stratification for perioperative therapy; postoperative complications; and prediction of lymph node metastasis, recurrence, survival, and prognosis. We paired terms for pancreatic cancer and pancreatic ductal adenocarcinoma (PDAC) with domain-specific clinical terms and terms for AI, machine learning, deep learning, radiomics, and pathomics. We also screened the reference lists of relevant studies for reports that the search might have missed. A supplementary search of PubMed/MEDLINE, covering publications through 17 September 2026, aimed to identify recent developments. [Supplementary Method 1](#) and [Supplementary Figure 1](#) detail the search concepts, eligibility criteria, and study-selection process.

Study selection

We considered original English-language studies that used computational approaches such as AI, machine learning, deep learning, radiomics, or pathomics for clinically relevant tasks within the pancreatic cancer surgical pathway. Inclusion as a direct AI study required reporting of model development, validation, or clinical evaluation based on patient-level clinical, imaging, pathological, operative, or omics data. The direct-AI evidence set excluded reviews, editorials, conference abstracts, animal and cell studies, treatment studies without AI, conventional three-dimensional reconstruction lacking an AI component, and studies that did not address a relevant pancreatic surgical task. We retained selected adjacent studies and protocols in a separate category when they clarified technical limitations or emerging routes to clinical translation.

The core evidence set ultimately included 34 studies: 31 direct AI studies and 3 adjacent studies or protocols. Formal assessment of evidence maturity was restricted to the 31 direct AI studies. [Supplementary Table 1](#) summarizes their characteristics. References cited for disease background, clinical context, methodological frameworks, regulation, or wider discussion of translation were not graded for evidence maturity. Li Z (Zhike Li) and Peng G (Gongze Peng) performed literature retrieval and evidence synthesis independently,

and resolved differences 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 document a formal process for resolving eligibility disagreements.

Data extraction

For each study in the core set, we extracted information on study design, target population, data sources, sample characteristics and event composition, clinical task, type of model, methods of development and validation, key performance findings, calibration, decision thresholds, analyses of utility, workflow features, clinical outcomes, and major methodological limitations.

Our methodological appraisal drew on the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis-Artificial Intelligence (TRIPOD+AI) statement, the Prediction model Risk Of Bias Assessment Tool (PROBAST), the Checklist for Artificial Intelligence in Medical Imaging (CLAIM), principles of decision-curve analysis, and the Fairness, Universality, Traceability, Usability, Robustness, and Explainability for Artificial Intelligence (FUTURE-AI) guideline^[7-11]. We used these frameworks to guide critical assessment, without deriving a numerical risk-of-bias score.

We focused particularly on cohort design, sample and event adequacy, separation of development and evaluation datasets, independent validation, calibration, decision utility, integration into workflows, and clinical applicability. When source articles provided confidence intervals, we reported them; where they did not, we retained the published point estimates.

EVALUATION FRAMEWORK

Evaluation dimensions

AI research in pancreatic cancer surgery addresses diverse tasks, from segmentation and evaluation of vascular invasion to stratification of treatment sensitivity, prediction of complications, and estimation of survival. These tasks differ in their intended populations, consequences of error, and resulting clinical actions. Consequently, the same area under the receiver operating characteristic curve (AUC) or accuracy can have different clinical implications. Failure to detect vascular involvement may underestimate operative complexity and margin risk, while falsely predicting lymph node metastasis may classify additional patients as high risk unnecessarily. The acceptable trade-off between sensitivity and specificity therefore varies by task. Discrimination measures how well a model distinguishes higher-risk from lower-risk cases; calibration measures how closely predicted probabilities match observed event rates. Decision thresholds translate continuous estimates of risk into actions, including further testing, closer monitoring, or discussion of treatment options. For uncommon outcomes, assessment of an operating point must also account for positive predictive value (PPV), sensitivity, baseline risk, and the clinical consequences of errors.

Performance marks the start of the evidence chain. Internal validation mainly examines stability within the environment of the development data, while independent external or temporal validation tests generalizability across institutions, devices, populations, and periods. Prospective performance validation tests a fixed model on prospectively collected data for the same task. Collecting data prospectively does not, by itself, establish an independent evaluation or show that predictions alter clinical decisions. Actual use in prospective workflows and its influence on decisions are assessed separately at Level 4. In multistage systems that combine segmentation, feature extraction, and classification, reporting should distinguish errors at each step and show how they affect the final output. Identifying the main points of failure is equally necessary, since strong performance early in a pipeline can conceal weaknesses in its ultimate clinical task. Progress from predictive validity to clinical utility requires interpretable and actionable thresholds, compatibility with real-world workflows, and prospective comparative evidence of net decision benefit or better patient outcomes^[11,12].

Evidence maturity results

We assigned each of the 31 direct AI studies the highest cumulative evidence level it met by satisfying all required criteria. Level 1 covered development and internal validation of a model for the target clinical task. To reach Level 2, a fixed model required independent external, temporal, or prospective performance validation for that same task. Evaluation data must not have informed feature selection, parameter tuning, or the choice of operating threshold. When evaluation data were used to update a model, an additional independent evaluation was necessary.

Level 3 required all Level 2 criteria to be met, together with clinically interpretable evidence of decision utility appropriate to the task and output. Probabilistic risk models had to demonstrate calibration suitable for their intended use and connect predictions with a specific clinical action using a prespecified or clinically interpretable threshold. They also needed evidence of net benefit or additional decision value relative to an appropriate clinical comparator. Neither calibration nor decision-curve analysis was sufficient on its own. For locked treatment-predictive classifiers, investigators had to fix the classification rule and threshold before independent evaluation. Independent data also had to demonstrate differential treatment value using a formal test of biomarker-treatment interaction or an equivalent validated analysis incorporating a treatment comparator. Prognostic stratification, a significant association within one treatment group, or significance in one group but not another did not establish treatment-predictive utility.

Level 4 required prospective use within a clinical workflow and measurement of its effects on decisions, management, or workflow. Validating performance prospectively was insufficient to meet this level. Level 5 required evidence that an AI-assisted strategy affected patient outcomes, resource use, efficiency, or safety.

Li Z and Peng G jointly assessed study-level evidence maturity and verified the supporting evidence. Ambiguous or borderline classifications were reviewed with Liu L (Liping Liu) and He Y (Yongzhu He) until a final classification was agreed upon. No inter-rater agreement statistic was calculated.

Every direct AI study met the Level 1 criteria; of the 31 studies, 13 attained at least Level 2 and 3 attained Level 3. Classification by the highest level achieved placed 18 studies at Level 1, 10 at Level 2, and 3 at Level 3, with none at Levels 4 or 5. Most evidence thus concerned model development and internal validation or independent validation of performance. Only a few studies supplied sufficient clinically interpretable evidence of decision utility to exceed Level 2. No study prospectively measured the effect of integrating AI on clinical decisions, as required for Level 4, or demonstrated benefits for patient outcomes, resource use, or safety, as required for Level 5. [Table 1](#) gives an overview of evidence across decision points, while [Figure 2](#) displays its maturity distribution. [Table 2](#) presents representative studies, and [Figure 3](#) outlines the route to clinical translation.

RESECTABILITY ASSESSMENT

Preoperative evaluation of vascular involvement is among the more extensively investigated and relatively mature uses of AI in pancreatic cancer surgery. The tumor's anatomical relationship to major vessels is fundamental to staging and operative strategy: it informs resectability, the likelihood of achieving microscopically margin-negative (R0) resection, and decisions about neoadjuvant treatment. International consensus criteria for borderline resectability encompass anatomical, biological, and patient-related factors. Vascular involvement is therefore one important determinant of the treatment pathway and must be

Evidence maturity across surgical decision nodes



Figure 2. Evidence maturity across surgical decision nodes in pancreatic cancer. We classified each of the 31 direct AI studies at the highest cumulative evidence level for which it met all required criteria. The three adjacent studies or protocols were not graded. [Supplementary Table 2](#) lists individual assignments and the evidence supporting them. Circle areas represent study counts, which are also shown numerically inside the circles. Source: Original diagram prepared for this review from the evidence synthesis and evidence-maturity framework. AI: Artificial intelligence.

From model performance to clinical benefit: where current evidence stops

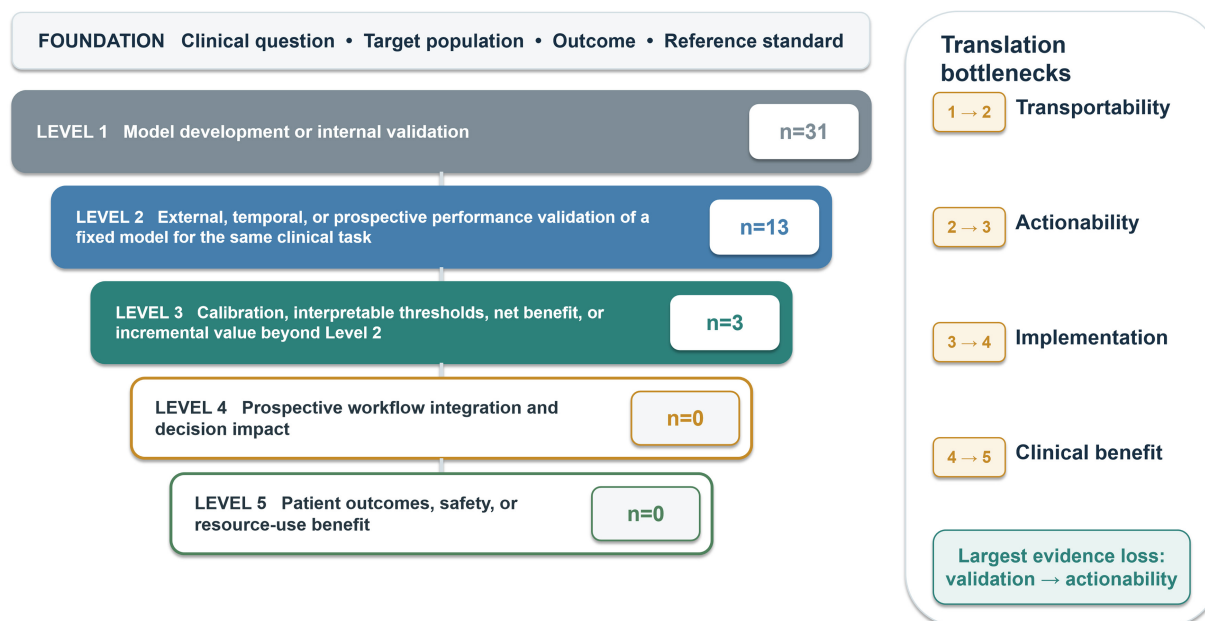


Figure 3. From model performance to clinical benefit: where current evidence stops. Counts show how many of the 31 direct AI studies attained each cumulative evidence level. The greatest gap lay between validation of performance and clinical actionability: 13 studies attained at least Level 2, only three attained Level 3, and none prospectively assessed decision impact or patient outcomes. Source: Original diagram prepared for this review from the evidence synthesis and evidence-maturity framework. AI: Artificial intelligence.

Table 1. Overview of evidence across surgical decision nodes in pancreatic cancer

Decision node	Clinical question	Primary modalities	Best available evidence	Adjunctive role	Main gap
Resectability and vascular involvement	Is the tumor resectable, are vessels involved, and what is the margin risk?	Contrast-enhanced CT; tumor-vessel features	External and prospective validation ^[16]	Per-vessel quantification and standardized risk discussion	Decision impact and benefits for R0 resection and survival remain unproven
3D reconstruction and surgical planning	How can complex anatomy be visualized and used for planning?	CT/MRI segmentation; 3D vascular models	Multicenter external validation for preoperative segmentation ^[21]	Anatomical visualization and procedural-risk alerts; emerging intraoperative recognition and AI-assisted visualization are discussed separately as ungraded context	Tumor segmentation, real-time surgeon-facing workflow integration, cost, and operative-outcome impact
Treatment decisions and response stratification	Who may be sensitive to a treatment or achieve a pathological response?	Transcriptomics, pathology, imaging, and clinical variables	Retrospective trial-material analyses and independent cohort evidence; selected studies reach explicit decision-utility criteria ^[26-27,30]	Sensitivity and response-stratification hypotheses	No prospective feature-guided treatment allocation
Perioperative complications	Can risks such as POPF and VTE be identified in advance?	Preoperative and discharge-time variables, body composition, and digital pathology	Large-sample internal or small-sample external validation ^[35,37,38]	Risk communication and research stratification	Thresholds, class imbalance, timeliness, and management utility
Lymph nodes, recurrence, and survival	How can biologically high-risk patients be identified?	CT, PET/CT, pathology, genomics, and body composition	Multicenter external or temporal validation, with calibration/DCA reported in selected studies ^[41,47,49]	Exploratory prognostic and trial stratification	No actionable treatment or surveillance thresholds or prospective impact evidence

AI: Artificial intelligence; CT: computed tomography; DCA: decision curve analysis; MRI: magnetic resonance imaging; PET/CT: positron emission tomography/computed tomography; POPF: postoperative pancreatic fistula; R0: microscopically margin-negative resection; VTE: venous thromboembolism.

considered together with tumor biology and the patient's general condition^[13]. In conventional imaging assessment, radiologists and surgeons combine information on tumor-vessel contact, the extent of encasement, vessel deformation, and distal reconstructability; their interpretations may vary. AI could translate experience-based descriptions into reproducible segmentations, measurements, and risk indicators for multidisciplinary team (MDT) discussion.

Studies assessing superior mesenteric artery margin risk have applied AI segmentation and classification to identify anatomical features linked to positive margins. One retrospective study at a single center developed a computer-vision pipeline linking these two tasks. Segmentation performed well, with a Dice similarity coefficient (DSC) of 0.90 for U-Net. Margin classification, however, combined high specificity with low sensitivity: sensitivity was 43%, specificity 94%, and accuracy 85%^[14]. Although the model identified negative margins with reasonable stability, it could miss positive margins, supporting its use as an adjunct to preoperative imaging assessment.

Evidence is more substantial for quantification of involvement in individual vessels. A single-center computer-aided diagnosis study used interpretable tumor-vessel features and achieved high internal discrimination for several vessels. Its assessment of portal vein and superior mesenteric vein involvement exceeded that of radiologists using encasement alone^[15]. A subsequent framework based on three-dimensional deep learning quantified involvement in five vessels: the celiac axis and common hepatic artery, the superior mesenteric artery and vein, and the portal vein. Diagnostic performance remained favorable in

Table 2. Selected anchor studies informing the main evidence synthesis

Decision node	Reference	Design and population	Validation	Key result	Clinical interpretation and boundary
Resectability/vascular involvement	[16]	Total $n = 2,130$: internal retrospective $n = 1,759$; external retrospective $n = 169$; prospective $n = 202$	External and prospective validation	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 Boundary: No replacement of radiologists/MDT; no proven R0, morbidity, or survival benefit
Resectability/SMA margin	[14]	200 PDAC Whipple patients; 40 positive SMA margins	Single-center retrospective pilot	U-Net DSC 0.90; sensitivity 0.43; specificity 0.94; accuracy 0.85; F1 0.50	Feasibility with low-sensitivity warning Boundary: Not reliable for excluding positive margin or deciding therapy
Resectability/human-machine fusion	[19]	Retrospective multicenter PC cohort; training $n = 263$, independent test $n = 86$	Independent test from two centers	HMfML test AUC 0.8845, accuracy 82.56%, sensitivity 84.21%, specificity 82.09%	Human-machine fusion showed incremental same-task value against guideline-based expert assessment and CT signatures. Boundary: no prospective workflow integration or patient-outcome validation
Resectability/report structuring	[50]	Single-center retrospective; 180 PDAC staging CT reports	Internal comparison against radiologist reference; surgeon reader study	GPT-4 report F1 0.997; surgeon classification 83% vs. 76% with original reports; shorter review time	Supervised report structuring and information-transfer support Boundary: No source-image review or external validation; not autonomous resectability assessment
3D planning/segmentation	[21]	275 pancreatic cancer surgical patients; external validation $n = 40$	Multicenter retrospective with external validation	Overall DSC about 75%; pancreas DSC about 85%-86%; tumor DSC about 55%-57%	AI-specific 3D visualization evidence Boundary: Not direct resectability stratification; tumor segmentation remains difficult
Treatment/adjuvant stratification	[26]	343 resected PDAC RNAseq profiles from PRODIGE-24/CCTG PA6	Retrospective validation in trial material	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 Boundary: Not deployment-ready chemotherapy assignment
Treatment/advanced-PDAC treatment-selection evidence	[30]	477 advanced/unresectable PDAC patients; validation $n = 299$	Development plus independent validation	Biomarker predicted differential benefit from fluoropyrimidine- versus gemcitabine-based first-line therapy	Adjacent computational pathology treatment-selection evidence Boundary: Not neoadjuvant/adjuvant or resectable surgical-pathway evidence
Treatment/post-neoadjuvant ypNO response	[28]	NCDB retrospective; 1053 cT1-4NOM0 PDAC patients completing NAC, radiotherapy, and surgery	Train $n = 600$; test $n = 453$; comparator cohort not equivalent external validation	AUC 0.718/0.725; predicted responders had ypNO 71.5% vs. 29.7% and R0 87.3% vs. 62.6%	Exploratory treatment-response stratification after completed multimodal therapy Boundary: Selected surgical cohort; no causal ENI/NAC benefit or pretreatment regimen selection

Treatment/fibrosis imaging biomarker	[29]	Multicenter retrospective; resectable pathology cohorts $n = 361$ plus unresectable treatment cohort $n = 295$	External CT test; retrospective treatment-subgroup analysis	External CT ensemble AUC 0.709; predicted high fibrosis associated with longer PFS/OS in the AG group	Adjacent imaging-biomarker evidence linking fibrosis and treatment outcome Boundary: Unresectable treatment cohort; no causal AG benefit or perioperative treatment assignment
Complication/POPF after PD	[35]	8666 elective pancreatoduodenectomy patients; ACS-NSQIP	Retrospective train/test split	XGBoost and random forest AUC 0.72	Preoperative counseling/risk stratification Boundary: Not deployment-ready drain, operative, or postoperative management
Complication/postdischarge VTE	[38]	51,916 pancreatectomy patients; 743 pdVTE events (1.4%)	80/20 stratified training-holdout internal validation	ROC-AUC 0.83 (95%CI: 0.80-0.87); AUPRC 0.34; at threshold 0.10, sensitivity 39.6%, specificity 99.5%, PPV 51.8%, NPV 99.1%	Threshold-specific risk stratification example. Boundary: source-level sensitivity inconsistency is unexplained; no validated prophylaxis-allocation rule or prospective benefit of model-guided prophylaxis
Complication/POPF after distal pancreatectomy	[37]	230 radical distal pancreatectomy patients; small external test set	Two-institution retrospective with external test	External combined model AUC 0.82; sensitivity 0.81; specificity 0.76	Promising multimodal POPF prediction Boundary: Not generalizable to all pancreatic operations
Prognosis/LNM	[39]	331 resectable PDAC patients; test $n = 90$	Two-center retrospective	Test AUC 0.795; sensitivity 95.3%; specificity 34.0%	Exploratory LNM risk stratification Boundary: Not rule-in, not surgical extent decision, not pathology replacement
Prognosis/early recurrence	[41]	493 curative-intent resection patients; validation $n = 86$	Multicenter retrospective validation	Nomogram AUC 0.941 test and 0.920 validation; calibration/DCA reported	Early recurrence risk stratification Boundary: No surveillance/adjuvant treatment threshold or outcome benefit
Prognosis/overall survival	[49]	1516 resected PDAC patients; external validation $n = 179$	Multicenter internal and external validation	Preoperative high-risk HR 2.03 internal and 2.47 external; postoperative HR 2.49 internal and 2.15 external	OS risk stratification after resection Boundary: No chemoradiotherapy selection or survival improvement proof
Prognosis/body-composition radiomics	[47]	Single-center retrospective; resected PDAC $n = 191$; temporal validation $n = 49$	Temporal validation	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 Boundary: No independent external validation or treatment/surveillance decision proof
Prognosis/occult liver metastasis	[43]	Five-center retrospective resected PDAC cohort $n = 438$; external $n = 103$	Internal and independent external validation	AUC 0.73 training, 0.72 internal, and 0.71 external	Moderate biological-risk stratification beyond anatomic resectability Boundary: Manual segmentation; excluded NAT; cannot select NAT or cancel surgery

Prognosis/spatial immune pathology	[46]	Single-center retrospective; 304 upfront-surgery, R0, non-NAC PDAC patients	Internal association analysis	Immune-inflamed phenotype and higher intratumoral TIL density associated with longer OS and RFS	Spatial pathology adds exploratory biological prognostic information. Boundary: Selected cohort, one representative slide, no external validation or immunotherapy prediction
Resectability/vascular invasion	[20]	Retrospective multicenter study; 5,930 tumor-vessel cases across six centers, including 3,575 in external cohorts from five centers	Five-center external cohorts reported; Level 2 model-locking and data-isolation criteria not fully verifiable	Overall external AUC 0.943; center-specific AUCs 0.939-0.991	External-cohort discrimination was reported, but the study remained Level 1 under our cumulative criteria. No prospective decision-impact evaluation or proof of improved resectability decisions, R0 resection, morbidity, or survival

For the postdischarge VTE study cited as reference^[38], the published article reports a sensitivity of 0.40%, whereas Table 5 in Supplement 2 (Results) of that same study reports a sensitivity of 0.396 (39.6%) at a threshold of 0.10. The reason for this discrepancy is unclear. **Table 2** presents the threshold-specific value from the original study's Supplement 2. ACS-NSQIP: American College of Surgeons National Surgical Quality Improvement Program; AG: nab-paclitaxel plus gemcitabine; AI: artificial intelligence; AUC: area under the receiver operating characteristic curve; AUPRC: area under the precision-recall curve; CSS: cancer-specific survival; CT: computed tomography; DCA: decision curve analysis; DFS: disease-free survival; DSC: Dice similarity coefficient; ENI: elective nodal irradiation; F1: F1 score; GEM: gemcitabine; GPT-4: Generative Pre-trained Transformer 4; HR: hazard ratio; LNM: lymph node metastasis; HmFmL: human-machine fusion machine learning; MDT: multidisciplinary team; mFFX: modified FOLFIRINOX; NAC: neoadjuvant chemotherapy; NAT: neoadjuvant therapy; NCDB: National Cancer Database; NPV: negative predictive value; OS: overall survival; PD: pancreaticoduodenectomy; PC: pancreatic cancer; PDAC: pancreatic ductal adenocarcinoma; pdVTE: postdischarge venous thromboembolism; PFS: progression-free survival; POPF: postoperative pancreatic fistula; PPV: positive predictive value; R0: microscopically margin-negative resection; RFS: recurrence-free survival; RNAseq: RNA sequencing; ROC-AUC: area under the receiver operating characteristic curve; SMA: superior mesenteric artery; TIL: tumor-infiltrating lymphocyte; VTE: venous thromboembolism; ypNO: pathologically node-negative after neoadjuvant therapy.

external and prospective evaluations, and several metrics were comparable or noninferior to those of senior radiologists^[16]. These findings indicate that AI could make descriptions of vascular involvement more consistent, especially when relationships involve multiple vessels and imaging planes.

Evaluation of venous involvement and overall resectability also demonstrates how AI can complement clinical judgment. An automated CT framework delineated PDAC and associated vessels and measured tumor-vessel contact. Overall resectability classifications matched radiologists' assessments in 80% of a balanced test set of 60 cases. Agreement was greater for arterial involvement than for venous involvement (87% vs. 59%)^[17]. In a retrospective study spanning two institutions, investigators combined CT radiomics and venous morphology into a risk score for intraoperative superior mesenteric-portal venous involvement. The score's AUC and sensitivity in the validation cohort exceeded those of conventional imaging assessment^[18]. Interpretation of these quantitative findings still requires consideration of operative exploration, the feasibility of reconstruction, and local inflammatory changes.

Combining model predictions with expert evaluation informed by guidelines may better approximate clinical practice. One retrospective multicenter study integrated machine-learning features from contrast-enhanced CT with expert assessments based on National Comprehensive Cancer Network guidelines. In an independent test set, this human-machine model surpassed both the CT signature and expert assessment alone in AUC, accuracy, sensitivity, and specificity^[19]. A subsequent retrospective multicenter study, published in 2026, developed a vascular-invasion model designed around clinicians' needs. The study included 5,930 tumor-vessel cases across six centers, including 3,575 cases in external cohorts from five centers; the pooled external AUC was 0.943^[20]. These external-cohort results do not, on their own, establish Level 2 under our framework: the published methods did not allow us to verify all model-locking and data-isolation requirements. We therefore classified the study as Level 1. AI may yield more objective

measurements and consistent risk indications, but its current role is to support assessment by radiologists, surgeons, and MDTs. Their effects on agreement within MDTs, rates of R0 resection, perioperative results, and long-term outcomes remain unknown.

SURGICAL PLANNING

Automated segmentation

Three-dimensional reconstruction, automated segmentation, and three-dimensional printing are connected technologies, but each requires different evidence. Segmentation algorithms delineate structures such as the pancreas, tumor, pancreaticobiliary ducts, and vessels. Reconstruction then represents these structures in an interactive spatial model, while printing creates a physical version. Segmentation and reconstruction can be components of an AI-driven imaging workflow; printing is not itself AI. Evaluation standards also differ. Segmentation is usually assessed against manual contours. Three-dimensional planning should be compared with conventional two-dimensional interpretation or manually constructed three-dimensional models, and physical printing must add information beyond a digital model viewed on screen. Keeping these distinctions clear helps identify whether a study validates segmentation, anatomical visualization, or clinical planning.

A multicenter study developed a model to segment pancreatic cancer and adjacent anatomy automatically. Overlap was favorable for the pancreas and several vessels, but tumor segmentation was less successful: DSCs were approximately 0.57 internally and 0.55 externally^[21]. Pancreatic tumors can be small, have poorly defined margins, and overlap with inflammatory or atrophic changes, making segmentation harder than for larger solid organs with clear contrast. These weaknesses matter for surgical planning. Missing tumor regions or displaced boundaries can distort measurements of tumor-vessel contact and remnant pancreatic volume, as well as three-dimensional displays. In addition to overall DSC, planning systems should therefore report failures involving small lesions, unclear boundaries, and differing acquisition protocols. Reporting should also cover how often manual correction is needed, how long it takes, and how consistently observers perform it. Extensive expert correction may erode the workflow gains from automation.

Vascular modeling

Three-dimensional vascular models connect anatomical displays with individualized indicators of risk. In a retrospective study at two institutions, deep learning enabled virtual models and measurements of superior mesenteric artery zones. Selected branching features were statistically associated with postoperative pancreatic fistula (POPF), other severe complications, and blood loss during surgery^[22]. The cohort included resections for several pancreatic tumor types. Accordingly, these anatomical features should be interpreted as indicators of operative complexity across diseases, rather than as oncological predictors specific to PDAC. They could draw attention to vascular variants, branching patterns, and technical challenges in preoperative discussions, helping with operative planning and communication of risk. The associations do not show that adapting the operative strategy to the model reduces complications. Prospective workflow evaluations should document whether the model alters the surgeon's plan and whether intraoperative findings support that alteration.

Physical printing requires more resources than digital three-dimensional visualization. A small retrospective study of 16 patients with borderline-resectable disease found that segmentation and three-dimensional reconstruction appeared to improve visualization of venous involvement, whereas a printed model supplied no additional information beyond the on-screen version^[23]. Producing the digital model took approximately 3 h, and printing took approximately 36 h, at a cost of about €700-1,000 per model. These requirements can limit access when treatment sequencing or operative scheduling demands prompt decisions. Although this is adjacent evidence without an AI component, it highlights a wider challenge in assessing surgical-planning tools. An intuitive visual display does not necessarily add clinical value. New approaches need direct

comparisons against two-dimensional interpretation, on-screen three-dimensional models, and standard MDT assessment.

Clinical application

The most immediate benefits of AI-assisted operative planning are likely to be faster segmentation, clearer depiction of complex spatial relationships, and quantitative description of individual vascular anatomy^[21-23]. Assessment can proceed at three levels. Technical assessment addresses segmentation errors and consistency between centers; process assessment covers manual correction, interpretation time, and agreement within teams; clinical assessment examines changes to operative plans and outcomes. Future studies should select endpoints appropriate to each level, including correction burden, time spent interpreting and discussing findings, interobserver agreement, changes in surgical approach, blood loss, complications, and margin status. They should also determine which complex vascular anatomies or borderline-resectable cases gain additional value, rather than applying three-dimensional processing routinely to all patients. For automated segmentation and three-dimensional modeling to function as evaluable decision support, they must deliver reliable, reproducible information that adds value within everyday workflows.

Intraoperative visualization

AI in pancreatic surgery is expanding beyond preoperative planning to recognize and visualize anatomy in operative video. One study trained a deep-learning model to identify two vascular structures: the superior mesenteric vein–portal vein axis and the splenic vein at pixel level. In videos of laparoscopic distal pancreatectomy and pancreatoduodenectomy, recall reached 85.0%, precision 91.1%, and mean DSC 0.754, with processing at roughly 11 frames per second^[24]. A separate study used 78 robotic pancreatoduodenectomy videos to develop a continuous AI score for superior mesenteric artery exposure and the extent of dissection. This score correlated strongly with expert surgeons' ratings (Spearman's $\rho = 0.767$, $P < 0.001$), with high agreement among surgeons [intraclass correlation coefficient (ICC), 0.81]. Displaying AI overlays also improved surgeons' confidence in judging the dissection extent ($P < 0.001$)^[25]. These studies illustrate emerging intraoperative recognition of anatomy, quantitative assessment of dissection, and assisted visualization. They are cited as contextual reports and were not among the 31 direct AI studies graded for evidence maturity.

TREATMENT DECISION-MAKING

Support for treatment decisions is an emerging use of AI in pancreatic cancer surgery. In particular, transcriptomic stratification of sensitivity to adjuvant chemotherapy has produced hypotheses that can be tested. Surgical care must consider more than anatomical resectability, since treatment sequence and the choice of adjuvant regimen also affect patient benefit. As neoadjuvant, conversion, and postoperative adjuvant approaches develop, methods for identifying treatment-sensitive and biologically high-risk subgroups may gain value. AI-assisted transcriptomic signatures initially developed in preclinical models were retrospectively validated using tumor samples from the PRODIGE-24/CCTG PA6 trial. Patients whose treatment matched the predicted subtype experienced longer disease-free and pancreatic cancer-specific survival^[26]. A related analysis using material from that trial identified a gemcitabine-sensitive subgroup using a ribonucleic acid signature; within that subgroup, gemcitabine was associated with improved disease-free and pancreatic cancer-specific survival^[27]. Both studies used randomized-trial samples and outcomes, but their molecular validation was retrospective, and treatment was not prospectively allocated according to the signatures. They therefore justify stratified treatment trials rather than routine use to choose adjuvant chemotherapy.

Prediction of response to perioperative therapy now extends beyond molecular signatures to phenotypes derived from clinical, pathological, and imaging data. A retrospective analysis of the National Cancer Database (NCDB) examined 1,053 patients treated with multiagent neoadjuvant chemotherapy, radiotherapy,

and surgery. Using least absolute shrinkage and selection operator (LASSO) modeling, investigators predicted pathologically node-negative status after neoadjuvant therapy (ypN0), achieving AUCs of 0.718 for training and 0.725 for testing. The response group identified by the model had higher ypN0 and R0 rates and longer survival^[28]. Predictions were additionally examined in 1,258 patients without elective nodal irradiation. Because treatment conditions differed, this supplemental analysis did not constitute independent external validation in the same intended-use setting^[28]. The findings suggest that routine clinical variables carry information about pathological response. However, the cohort was restricted to selected patients who eventually underwent surgery, and detailed chemotherapy and radiotherapy information was unavailable. The study cannot demonstrate a causal benefit of elective nodal irradiation or any particular neoadjuvant regimen. Another retrospective multicenter study used AI pathology to quantify fibrosis and developed a CT radiomics predictor of fibrosis. The ensemble achieved an AUC of 0.709 in an external CT test set; CT-predicted high fibrosis was associated with longer survival following first-line nab-paclitaxel plus gemcitabine in unresectable PDAC^[29]. The analysis connects pathological and imaging findings with treatment response. Nevertheless, its predictive analysis concerned unresectable disease, so it provides evidence from an adjacent population that cannot yet support allocation of neoadjuvant or adjuvant therapy.

Computational pathology and prospective multimodal designs offer additional approaches to treatment stratification. Independent validation of a hematoxylin-and-eosin whole-slide biomarker has shown differences in benefit from first-line chemotherapy among patients with advanced or unresectable pancreatic cancer^[30]. That population and treatment context differ from perioperative management of resectable disease. A prospective multicenter single-arm trial protocol proposes predicting changes in resectability after FOLFIRINOX (folinic acid, fluorouracil, irinotecan, and oxaliplatin) by combining radiomics from CT and magnetic resonance imaging (MRI) with liquid-biopsy multiomics. At present, it describes research infrastructure and provides no outcome data^[31]. A single-arm study combining modified FOLFIRINOX with an immune-checkpoint inhibitor supplies non-AI context for translational neoadjuvant research^[32]. For now, AI in neoadjuvant, conversion, and adjuvant treatment is most plausibly used to stratify treatment sensitivity and pathological response. Its use to select treatment will require prospective studies of feature-guided decisions that validate testing processes, actionable thresholds, and patient outcomes. Models must also separate prognostic features from predictive ones: identifying patients likely to fare poorly does not show that they will benefit differently from two regimens. Future work should prespecify treatment-interaction tests, explain how assay failures and missing samples are handled, and verify that transcriptomic, pathological, and imaging features are reproducible across centers.

COMPLICATION PREDICTION

Prediction tasks

The clinical purpose of a perioperative risk model depends on when its inputs become available. Preoperative imaging and variables can guide risk discussions, surgical preparation, and monitoring plans. In a single-center study, a decision tree based on preoperative characteristics predicted major bleeding with similar accuracy in training and temporal testing. A neural-network comparator performed less accurately on the test set, underscoring the importance of assessing overfitting^[33]. Anatomical information obtained during surgery may inform immediate management, while postoperative pathology and early laboratory findings better explain subsequent changes in risk. POPF, hemorrhage, infection, and venous thromboembolism (VTE) can prolong hospitalization, increase reintervention, and delay adjuvant therapy. Prediction endpoints should therefore capture clinically meaningful events affecting recovery. Under the International Study Group of Pancreatic Surgery classification, clinically relevant grade B and C POPF are distinguished from biochemical leakage^[34]. Consistent outcome definitions and explicit prediction times are essential for model comparison and for specifying intended use. They also help prevent bias from grouping minor biochemical changes with complications that alter management.

Pancreatic fistula prediction

Studies of POPF demonstrate how different sources contribute complementary information. An analysis of a national surgical quality database included 8,666 patients undergoing pancreatoduodenectomy and predicted clinically relevant POPF using variables available before surgery alone. XGBoost attained an AUC of 0.72^[35]. Although the large cohort reduces random error, the findings also indicate that routine preoperative variables offer only moderate initial stratification and incompletely characterize gland texture, duct features, and local histology. A multicenter digital-pathology study applied AI to resection-margin slides to quantify tissue components. The area of fibrosis showed an independent association with clinically relevant POPF; reported AUCs were approximately 0.58-0.73^[36]. The preoperative model may aid counseling before surgery; the slide-based model instead offers a postoperative histological explanation because its inputs are obtained after surgery. Intended use must respect this timing. A model requiring postoperative slides cannot serve as a preoperative decision tool.

Models incorporating body composition seek to capture systemic condition beyond what routine clinical variables describe. A two-institution study evaluated 230 distal pancreatectomy patients, including 42 in an external test cohort. Combining clinical variables with two- and three-dimensional body-composition features yielded an external AUC of 0.82, sensitivity of about 81%, and specificity of about 76%, exceeding the performance of a clinical-only model^[37]. The results suggest additional predictive information from body composition and demonstrate a role for automated segmentation in perioperative risk research. However, the cohort was limited to distal pancreatectomy, and external testing involved few patients. Differences between procedures in duct management, pancreatic remnants, and fistula mechanisms leave it uncertain whether the model transfers to pancreatoduodenectomy or other workflows for assessing body composition.

VTE prediction

VTE prediction demonstrates the importance of interpreting discrimination at a stated operating threshold. One study analyzed 51,916 pancreatectomy records, including 743 postdischarge VTE events (1.4%), using a stratified 80/20 split between training and holdout data^[38]. The leading gradient-boosting model produced two performance estimates: 0.83 (95%CI: 0.80-0.87) for the receiver operating characteristic AUC (ROC-AUC) and 0.34 for the area under the precision-recall curve (AUPRC). Supplementary results for a threshold of 0.10 showed 39.6% sensitivity, 99.5% specificity, 51.8% PPV, and 99.1% negative predictive value. These results neither identify a clinically optimal cutoff for extended prophylaxis nor demonstrate prospective benefit from model-guided prophylaxis.

A priority for complication models is to connect predicted risk with a defined action and the consequences of incorrect classification. False-negative results can deprive high-risk patients of closer monitoring or additional assessment. False-positive results can lead to more testing, inpatient observation, or prophylactic treatment. The balance of benefit and harm differs across these actions; consequently, drain management, nutritional support, and thromboembolic-risk assessment may need different thresholds even when based on one model^[37,38]. Protocols should also document whether and when predictions alter the team's initial judgment and why clinicians decline recommendations^[12]. The present findings support further assessment of risk stratification at the appropriate clinical time point. In particular, a discharge-based VTE model is not a preoperative tool. Prospective comparisons remain necessary to determine whether model use reduces complications or hospital stay, prevents unnecessary interventions, or improves completion of later treatment.

PROGNOSTIC PREDICTION

Perioperative risk assessment addresses immediate surgical safety, while models of lymph node metastasis, recurrence, and survival concern tumor biology and longer-term prognosis. Conventional evaluation relies mainly on pathological staging, margin assessment, tumor markers, and clinicians' judgment. AI could add

risk information from preoperative images, clinical variables, and biological features. A two-center XGBoost model for preoperative prediction of lymph node metastasis achieved test-set sensitivity of 95.3% but specificity of 34.0%. This high sensitivity limited missed positive cases, although low specificity restricted its value for confirmation^[39]. A study applying contrastive learning to multiphase CT reported a mean AUC of about 0.79. Its external data concerned another tumor type and prediction task, however, and did not independently validate the same clinical task^[40]. Such models offer exploratory risk indicators; they cannot replace pathological confirmation or determine the extent of lymphadenectomy on their own.

Prediction of early recurrence and occult metastasis broadens preoperative evaluation to include biological risk alongside anatomical resectability. In a multicenter study, a model integrating clinical variables with CT deep learning and radiomics achieved a validation-cohort AUC of 0.920 for early recurrence, with calibration and decision-curve analyses also reported^[41]. A further study combined intratumoral and peritumoral radiomics with body composition and clinicopathological variables, obtaining an external-cohort AUC of 0.884^[42]. A five-center study of occult liver metastasis defined the endpoint as metastasis found during surgery or within 6 months afterward; its AUC was 0.71 in an independent external cohort^[43]. Manual tumor segmentation was required, and patients treated with neoadjuvant therapy were excluded. The model therefore offers a basis for prospective stratification, rather than a tool for treatment selection. PanScore, a genomic deep-learning model, stratified radiographically resectable patients in an independent cohort. Median overall survival was 35.4 months in the low-risk group, 29.0 months in the intermediate-risk group, and 17.9 months in the high-risk group^[44]. Liver metastasis appearing soon after surgery is not necessarily equivalent to pre-existing micrometastasis. Moreover, independent validation of PanScore used overall survival, not subsequent liver metastasis, as its endpoint. These two studies support stratification of biological risk, but neither justifies choosing neoadjuvant therapy or canceling surgery.

Prognostic modeling now extends beyond CT to body composition and spatial pathology, as well as positron emission tomography/computed tomography (PET/CT). A CT investigation involving two cohorts showed that combining radiomic and deep-learning features improved prognostic discrimination in a test cohort from an independent hospital^[45]. A single-center spatial-pathology analysis included 304 patients treated with upfront surgery and R0 resection. AI-derived measurements of tumor-infiltrating lymphocytes and immune phenotypes were associated with overall and recurrence-free survival after surgery^[46]. However, the highly selected cohort, reliance on one representative slide, and lack of external validation restrict the findings to pathological prognostic stratification. They do not support conclusions about sensitivity to immunotherapy.

A body-composition radiomics model retained a moderate concordance index on temporal validation, but its AUC for 12-month recurrence-free survival fell from 0.89 in training to 0.68 in validation. This illustrates temporal drift and the influence of small validation cohorts^[47]. Among 627 patients at a single center, a radiomics score derived from preoperative fluorine-18 fluorodeoxyglucose (18F-FDG) PET/CT showed an independent association with postresection overall survival (hazard ratio, 2.95). Time-dependent AUCs at 12, 24, and 36 months were 0.729, 0.710, and 0.705, respectively^[48]. An entirely automated model based on multiphase CT has also undergone multicenter internal and external validation^[49]. These approaches extend risk evaluation to tumor, host, and microenvironmental features, but currently justify research stratification rather than alterations to adjuvant treatment or surveillance. The PET/CT results describe prognostic associations in one center and cannot establish superiority to CT, MRI, or models tested elsewhere. Clinical adoption requires additional value beyond tumor-node-metastasis stage, carbohydrate antigen 19-9, and standard imaging, with thresholds and net benefit tied to specific management changes.

CLINICAL TRANSLATION

Reporting and bias

Moving AI models for pancreatic cancer surgery into practice requires interpretable, reproducible findings applicable to the intended population. Gaps in reporting of populations, data sources, prediction timing, outcome definitions, missing-data methods, training and validation splits, feature selection, or optimization obscure whether bias, information leakage, or overfitting has inflated performance. Estimates can be overly optimistic when feature selection uses the entire dataset, images from one patient appear in both training and test sets, or investigators repeatedly consult the test set while tuning the model. Alongside point estimates, reports should provide confidence intervals, examples of failure, and additional value relative to a basic clinical model. This information helps separate random variation and technical improvements from clinically meaningful gains. TRIPOD+AI, PROBAST, and CLAIM offer complementary guidance on prediction-model reporting, assessment of bias, and reporting of medical-imaging AI^[7-9].

External validation

Independent external validation is necessary to assess generalizability between centers, but the strength of that evidence depends on where the external cohort originates and how heterogeneous it is. Testing a small cohort from the same region, using similar equipment or identical preprocessing, addresses a different question from testing across regions, devices, and healthcare systems. The 2026 multicenter vascular-invasion study reported discrimination in five external-center cohorts, but model-locking and data isolation could not be fully verified for Level 2^[20]. Conversely, substantial deterioration during temporal validation of a body-composition prognostic model^[47] highlights the need to monitor changes in data distributions over time. Reports of external validation should specify whether patient selection matched the development cohort, whether outcome definitions were consistent, and how missing inputs were managed. Otherwise, performance differences cannot reliably be attributed to the population, workflow, or algorithm. A practical route to implementation begins with silent prospective validation at the intended center, observing performance, calibration, and failure patterns without affecting care. Once local performance is understood, studies of clinician-AI collaboration can follow. Even consistently strong external performance does not establish that clinical decisions improve.

Calibration and thresholds

Translating predictions into action requires calibration and a clear understanding of the consequences of the chosen threshold. Decision-curve analysis adds to discrimination by evaluating the net benefit of model-assisted decisions with intervening in all patients or none across a range of threshold probabilities^[10]. The VTE model demonstrates how sensitivity and PPV vary with the threshold; the lymph node metastasis model demonstrates the burden associated with high sensitivity and low specificity^[38,39]. Both patterns must be judged in relation to the proposed action, rather than by AUC alone. Differences in event rates between institutions may alter predictive values and calibration, so a cutoff optimized in development cannot simply become a universal clinical threshold. Every threshold should correspond to a defined action. More false positives may be tolerable when the consequence is additional imaging, whereas changes to surgery or treatment demand stronger evidence. Before implementation, investigators should establish actionable thresholds in the target population and report calibration, predictive values, and the consequences of misclassification. Sensitivity analyses should assess whether net benefit remains stable across thresholds.

Human-AI collaboration

Interpretable outputs should enable clinicians to scrutinize the basis for a prediction and identify uncertainty. Models based on tumor-vessel features and the multicenter model of vascular invasion convert selected imaging signs into quantitative features that can be traced and reviewed clinically^[15,20]. Work on structured reporting assisted by large language models suggests that text tools can improve completeness,

although image review, external validation, and human oversight remain essential^[50]. Further assessment should test whether explanations remain stable when inputs are perturbed, accord with clinical knowledge, and help identify errors. Studies of clinician-AI collaboration can document acceptance, modification, or rejection of recommendations, compare benefits across levels of clinician experience, and establish procedures for resolving disagreement between experts and models. Linking data quality, predictions, clinical review, and escalation of abnormal findings within an accountable process would make interpretability an auditable part of practice rather than merely a display feature.

Workflow evaluation

A functioning model also needs to fit imaging, pathology, MDT, and perioperative processes. Three-dimensional modeling, digital pathology, and omics assays introduce additional image processing, specimen preparation, or sequencing. Their value in practice depends on whether the resulting information justifies the time, resources, and maintenance involved^[23,26,27]. Implementation studies should identify who obtains and checks inputs, when results become available, how missing or abnormal inputs are addressed, and whether recommendations are included in the medical record and MDT documentation. Evaluation should account for turnaround time, manual correction, repeat tests, and information-system maintenance costs. Following model updates, calibration and scope of use require reconfirmation. Monitoring measures and suspension criteria should address declining performance, safety incidents, and disparities between populations. The Developmental and Exploratory Clinical Investigations of Decision-support systems driven by Artificial Intelligence (DECIDE-AI) guideline highlights safety, human factors, workflows, and possible harms in early clinical evaluation^[12]. FUTURE-AI covers fairness, universality, traceability, usability, robustness, and explainability throughout the AI lifecycle^[11]. These frameworks distinguish successive, connected stages of evidence: predictive performance, deployment in practice, and patient benefit. Prospective assessment of effects on decisions and real-world outcomes remains the major gap for AI in pancreatic cancer surgery.

Implementation should reflect each institution's capacity. High-volume tertiary centers may be equipped to integrate multimodal platforms locally, monitor performance prospectively, and update models systematically. Regional centers and institutions with fewer resources may gain more from externally validated tools for specific tasks that accept standardized inputs, need little local infrastructure, and permit centralized support. Decisions about deployment should therefore account for case volume, data systems, specialist access, integration with information systems, technical assistance, and the ability to monitor performance over time.

Regulatory translation must also be assessed within each jurisdiction. In the United States, the U.S. Food and Drug Administration (FDA) issued draft guidance in January 2025 on lifecycle management and marketing submissions for AI-enabled device software functions and final guidance in August 2025 on predetermined change control plans^[51,52]. Within the European Union, devices enabled by AI may fall under complementary provisions of the Medical Device Regulation or In Vitro Diagnostic Medical Device Regulation and the Artificial Intelligence Act when relevant high-risk criteria apply^[53]. Comparisons between jurisdictions should examine intended use, requirements for clinical evidence, human oversight, procedures for software changes, cybersecurity, transparency, conformity assessment or marketing authorization, and post-market surveillance. Regulatory status in one jurisdiction should not be assumed to transfer directly to another.

Economic assessments should account for software purchase, annotation of data, system integration, upkeep, manual correction, and subsequent testing. They should also evaluate possible reductions in ineffective interventions, repeated examinations, and time spent making decisions. Conducting budget-impact and cost-effectiveness analyses alongside studies of decision impact can establish whether performance improvements justify routine adoption.

Declining performance on temporal validation, threshold-sensitive trade-offs for rare outcomes, and the tension between sensitivity and specificity pose different obstacles to clinical use^[38,39,47]. Within the evidence chain in [Figure 3](#), these obstacles relate to external validation, calibration and thresholds, and effects on decisions.

FUTURE DIRECTIONS

Preoperative assessment

Studies of vascular involvement have advanced beyond single-center development. A three-dimensional deep-learning framework demonstrated external and prospective performance validation; a separate multicenter vascular-invasion study reported external-cohort discrimination, but its published methods did not verify all model-locking and data-isolation criteria required for Level 2^[16,20]. The next question is whether these models improve observers' agreement on vascular invasion and resectability during actual MDT practice. Studies should capture effects on further testing, consideration of neoadjuvant therapy, preparation for vascular reconstruction, and surgical selection. Initial silent implementation can characterize local calibration and failure patterns. Before-and-after comparisons, crossover interpretation studies, or cluster designs can then evaluate conventional versus model-assisted assessment. Endpoints should include the proportion of judgments altered, futile operations, readiness for vascular reconstruction, R0 resection rates, and discussion time, extending evaluation beyond accuracy. Work on three-dimensional planning should improve segmentation of small tumors and poorly defined margins while measuring manual correction, planning time, and operative outcomes. It should separately assess the additional value of digital models and physical printing. When a model prompts a change in the operative plan, comparison with intraoperative or pathological findings is needed to establish whether that change was helpful.

Treatment decision-making

The main challenge for treatment-decision research is to test retrospective stratification hypotheses prospectively using feature-guided designs. Trial-derived transcriptomic signatures, predictors of ypNo after therapy, and imaging biomarkers of fibrosis capture molecular sensitivity, pathological response, and the tumor microenvironment, respectively^[26-29]. They differ, however, in the populations studied, timing of sampling, and actions they might inform. Future studies should prespecify testing procedures, decision thresholds, and allocation rules. Prospective comparative designs are needed to determine whether model-assisted selection alters treatment completion, opportunities for resection, or long-term outcomes. Evaluation of transcriptomic and pathological features should also cover tissue adequacy, agreement between laboratory platforms, turnaround times, and assay failure rates, establishing whether results are available within the treatment window. Studies comparing benefit between regimens should prespecify a biomarker-treatment interaction in their statistical design; showing a prognostic association is insufficient. Establishing value for treatment selection requires direct comparison of feature-guided allocation with a conventional strategy.

Risk prediction

Complication and prognostic models should move beyond stratifying risk to defining action pathways that can be tested. Models predicting POPF or VTE, lymph node metastasis or occult metastasis, and early recurrence or survival should report calibration, decision thresholds, predictive values, and decision curves for their intended population. Investigators should specify whether predictions inform risk communication, further testing, closer monitoring, or trial stratification^[37,38,41,49]. Every action needs clear criteria for initiation, duration, and cessation. For instance, a high-risk classification should not be given multiple unvalidated roles in postoperative monitoring, additional imaging, and research stratification. Tools intended to alter

anticoagulation, drain management, adjuvant treatment, or surveillance could be tested in cluster-randomized, stepped-wedge, or embedded prospective studies against current practice. Assessment should capture harms caused by false-positive and false-negative results, resource consumption, workload for clinicians, and patient outcomes, rather than focusing only on small AUC differences.

Multimodal research

Multimodal approaches can combine local tumor imaging, histology, molecular characteristics, and the patient's systemic condition. More modalities, however, do not automatically add clinical value. Published studies have combined body composition and clinical variables, CT findings and guideline-based information, or intratumoral and peritumoral radiomics and host features^[19,37,42]. These pairings suggest complementary information but add missing data, workflow dependence, and model complexity. Comparisons with unimodal models should use identical cohorts, endpoints, data partitions, and validation conditions. Reports should quantify gains in discrimination, shifts in calibration, net decision benefit, resilience when modalities are missing, and extra costs of acquisition and maintenance. Ablation analyses can establish which modalities account for gains, and comparing complete cases with those missing modalities can expose selection bias. Models also require a safe fallback when inputs from some modalities are unavailable. Multimodal deployment is warranted only if clinically meaningful net benefit over a simpler unimodal or baseline clinical model exceeds the extra cost and workflow burden. A simpler, more reproducible model may otherwise be the better practical choice.

Emerging methods

Emerging AI approaches may help overcome several current limitations. Self-supervised learning extracts general representations from unlabeled images or pathology data, potentially reducing reliance on scarce high-quality labels and laborious annotation. Generative AI may assist in controlled augmentation of data, annotation of images or text, and checks for consistency. Safeguards are nevertheless needed to keep synthetic bias and incorrect information out of training datasets. Large language models may be more suitable for combining complex findings from imaging, pathology, laboratory tests, and longitudinal records, as well as supporting structured reporting. A supervised approach to structured reporting of pancreatic cancer resectability has already been investigated^[50]. Broader reviews of surgical AI similarly identify the move from prediction to image guidance, clinical workflows, and ongoing evaluation as a common challenge^[54,55]. Before these methods inform surgical decisions, safeguards should address factual accuracy, privacy, updates, traceability of prompts and versions, and responsibility for human review. Evaluation datasets should cover incomplete structured information, contradictory test findings, unusual anatomy, and post-treatment changes. Omissions, fabricated content, erroneous attribution, and unsuitable recommendations should be counted separately, rather than compressed into a text-similarity metric. After changes to a model or prompt, records of versions, input sources, and human revisions should be retained to enable error investigation and monitoring for performance drift. FUTURE-AI offers a framework for monitoring data drift, safety incidents, and fairness after deployment^[11]. The priority is to develop auditable decision-support systems that work within clinical processes, remain under continuous surveillance, and demonstrate value prospectively, beyond achieving a high score in a single evaluation.

LIMITATIONS

Because PubMed/MEDLINE was the principal database, this review may have missed studies indexed only in other sources. Reference lists were examined for additional studies, but contemporaneous source-tracing records were insufficient to reconstruct the number identified exclusively through that route. This structured narrative review did not employ a systematic-review protocol or pool effect estimates. Differences in populations, clinical tasks, endpoints, and reporting also restricted direct comparison. Source articles did not consistently supply confidence intervals or calibration findings. We therefore retained performance estimates

as published, without retrospective harmonization. Full-text eligibility assessment was not conducted independently in duplicate, and the historical records do not document a formal eligibility-disagreement resolution procedure. Evidence-maturity levels were assigned jointly rather than by independent duplicate raters, and no inter-rater agreement statistic was calculated. Although borderline classifications received senior review, some grading judgments may remain subjective.

CONCLUSION

AI could contribute clinically at multiple points in the pancreatic cancer surgical pathway. Work on quantifying vascular involvement, visualizing anatomy in three dimensions, stratifying treatment sensitivity, estimating perioperative complication risk, and predicting recurrence and survival has advanced from exploratory development to initial external validation. Among these applications, preoperative measurement of vascular involvement has received relatively extensive validation.

Overall, the evidence remains in the early to intermediate stages of validation. Most studies involve retrospective development or internal validation, with fewer progressing to external or prospective evaluation. Substantial gaps remain in calibration, decision thresholds, workflow integration, and prospective measurement of decision impact. Decision support is therefore the appropriate present role for AI in pancreatic cancer surgery. At important decision points, quantitative outputs and risk estimates can supplement the assessments of radiologists, surgeons, and MDTs while leaving clinical judgment with those professionals.

Organizing the evidence around the surgical pathway clarifies how far AI research has progressed at each decision point, which clinical roles are currently supported, and what requires validation next. Separating applications with relatively strong evidence for adjunctive use from those still at an exploratory stage is more useful than labeling an entire technology mature. Claims affecting surgery or treatment continue to require prospective validation. For AI in pancreatic cancer surgery to become a demonstrably useful clinical tool, evidence must show successful integration into workflows, meaningful effects on decisions, and better outcomes in real-world practice.

DECLARATIONS

Authors' contributions

Conception and design of the review: Liu L, He Y

Literature retrieval, evidence synthesis, preparation of tables and figures, and writing of the original manuscript: Li Z, Peng G

Critical revision and supervision: Liu L, He Y

All authors participated in interpreting the evidence, reviewed and approved the final manuscript, and accepted accountability for every aspect of the work.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tools ChatGPT (OpenAI, version GPT-5.6 Sol, released 2026-07-09) and Claude (Anthropic, version Claude Opus 5, released 2026-07-24) was used to assist with manuscript preparation and language editing. For the revised graphical abstract and [Figure 1](#), OpenAI Codex using GPT-6 Sol (version gpt-6-sol, released 2026-09-22) assisted in creating original, editable vector diagrams. No third-party illustrations, stock icons, or patient images were used in these two revised graphics. All scientific judgments, evidence interpretation, and final decisions were made by the authors. The tools did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2024A1515013254). The funder had no role in the review design, literature selection, evidence interpretation, manuscript preparation, or the decision to submit the work for publication.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

Supplementary Materials

[Supplementary Materials](#)

REFERENCES

1. Stoop TF, Javed AA, Oba A, et al. Pancreatic cancer. *Lancet.* 2025;405:1182-202. [DOI PubMed](#)
2. National Cancer Institute. SEER cancer stat facts: pancreatic cancer. Available from: <https://seer.cancer.gov/statfacts/html/pancreas.html>. [Last accessed on 12 Oct 2026].
3. Huang B, Huang H, Zhang S, et al. Artificial intelligence in pancreatic cancer. *Theranostics.* 2022;12:6931-54. [DOI PubMed PMC](#)
4. Korfiatis P, Suman G, Patnam NG, et al. Automated artificial intelligence model trained on a large data set can detect pancreas cancer on diagnostic computed tomography scans as well as visually occult preinvasive cancer on prediagnostic computed tomography scans. *Gastroenterology.* 2023;165:1533-46.e4. [DOI PubMed PMC](#)
5. Chen PT, Wu T, Wang P, et al. Pancreatic cancer detection on CT scans with deep learning: a nationwide population-based study. *Radiology.* 2023;306:172-82. [DOI PubMed](#)
6. Sethi N, Varma CV, Reddy SS. Artificial intelligence in pancreatic cancer: applications in early detection, tumor staging, and survival prediction - a comprehensive review. *Clin Transl Oncol.* 2026;28:4308-27. [DOI PubMed](#)
7. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. [DOI PubMed PMC](#)
8. Wolff RF, Moons KGM, Riley RD, et al. ; PROBAST Group. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170:51-8. [DOI PubMed](#)
9. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell.* 2020;2:e200029. [DOI PubMed PMC](#)
10. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26:565-74. [DOI PubMed PMC](#)
11. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ.* 2025;388:e081554. [DOI PubMed PMC](#)
12. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ.* 2022;377:e070904. [DOI PubMed PMC](#)
13. Isaji S, Mizuno S, Windsor JA, et al. International consensus on definition and criteria of borderline resectable pancreatic ductal adenocarcinoma 2017. *Pancreatol.* 2018;18:2-11. [DOI PubMed](#)
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;283:834-40. [DOI PubMed](#)
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;111:7829-41. [DOI PubMed PMC](#)
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.* 2025;9:79. [DOI PubMed PMC](#)

17. Bereska JJ, Janssen BV, Nio CY, et al. ; Pancreatobiliary and Hepatic Artificial Intelligence Research (PHAIR) consortium. Artificial intelligence for assessment of vascular involvement and tumor resectability on CT in patients with pancreatic cancer. *Eur Radiol Exp.* 2024;8:18. DOI PubMed PMC
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;30:1206-16. DOI PubMed
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.* 2023;39:399-409. DOI PubMed
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 PubMed PMC
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;111:6715-24. DOI PubMed PMC
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;33:1627-35. DOI PubMed
23. Sguinzi R, Vidal J, Poroos F, et al. 3D modeling to predict vascular involvement in resectable pancreatic adenocarcinoma. *Heliyon.* 2025;11:e41473. DOI PubMed PMC
24. Shi J, Cui R, Wang Z, et al. Deep learning HRNet FCN for blood vessel identification in laparoscopic pancreatic surgery. *NPJ Digit Med.* 2025;8:235. DOI PubMed PMC
25. Takayama M, Takeuchi M, Tomita K, et al. Artificial intelligence-assisted estimation of superior mesenteric artery dissection extent during robotic pancreatoduodenectomy. *HPB.* 2026;28:976-86. DOI PubMed
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;35:780-91. DOI PubMed
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;42:1067-76. DOI PubMed PMC
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:e71447. DOI PubMed PMC
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.* 2025;8:0937. DOI PubMed PMC
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;44:2496-506. DOI PubMed PMC
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;25:204. DOI PubMed PMC
32. Wainberg ZA, Link JM, Premji A, et al. Neoadjuvant modified FOLFIRINOX plus nivolumab in borderline-resectable pancreatic ductal adenocarcinoma: a pilot phase I trial. *Nat Commun.* 2026;17:2232. DOI PubMed PMC
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:e0259682. DOI PubMed PMC
34. Bassi C, Marchegiani G, Dervenis C, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 years after. *Surgery.* 2017;161:584-91. DOI PubMed
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;30:7738-47. DOI PubMed
36. Mályi 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.* 2025;27:393-401. DOI PubMed
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;57:2597067. DOI PubMed PMC
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;32:4085-93. DOI PubMed
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;9:161-8. DOI PubMed PMC
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 PubMed

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;32:7567-76. DOI PubMed
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;111:8198-211. DOI PubMed PMC
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;110:740-9. DOI PubMed PMC
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 PubMed
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;11:1378. DOI PubMed PMC
46. Kim H, Choi JH, Lim Y, et al. Artificial intelligence-powered spatial analysis of immune phenotypes in resected pancreatic cancer. *JAMA Surg*. 2025;160:884-92. DOI PubMed PMC
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;15:28722. DOI PubMed PMC
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;49:909-16. DOI PubMed
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:e68-79. DOI PubMed
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:e233117. DOI PubMed
51. U.S. Food and Drug Administration. Artificial intelligence-enabled device software functions: lifecycle management and marketing submission recommendations. Draft Guidance for Industry and Food and Drug Administration Staff. January 2025. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing>. [Last accessed on 12 Oct 2026].
52. U.S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions. Guidance for Industry and Food and Drug Administration Staff. August 2025. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence>. [Last accessed on 12 Oct 2026].
53. Artificial Intelligence Board; Medical Device Coordination Group. AIB 2025-1/MDCG 2025-6: Interplay between the Medical Devices Regulation (MDR), In vitro Diagnostic Medical Devices Regulation (IVDR), and Artificial Intelligence Act (AIA). June 2025. Available from: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf. [Last accessed on 12 Oct 2026].
54. Müller PC, Erdem S, Kuemmerli C, et al. Artificial intelligence and image guidance in minimally invasive pancreatic surgery: current status and future challenges. *Art Int Surg*. 2025;5:170-81. DOI
55. Patel K, Connor AA, Kodali S, et al. From prediction to practice: a narrative review of recent artificial intelligence applications in liver transplantation. *Art Int Surg*. 2025;5:298-321. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.