ASIDE

ASIDE Gastroenterology

ASIDE GASTROENTEROLOGY

Review Article

Artificial Intelligence Applications Along the ERCP Pathway: An Evidence Map and Relevance to Pancreatic Cancer-Related Biliary Obstruction

Ahmed Abdallah Salman[1,*]

1-Internal Medicine Department, Faculty of Medicine, Cairo University, Cairo, Egypt

ARTICLE INFO

Article history:
Received 1 Aug. 2026
Received in revised form 10 Oct. 2026
Accepted 8 Nov. 2026
Published 28 Nov. 2026

Keywords:
Artificial intelligence
Endoscopic retrograde cholangiopancreatography
Malignant biliary obstruction
Pancreatic cancer
Prediction model

ABSTRACT

Introduction: Artificial intelligence (AI) is being applied to isolated tasks along the endoscopic retrograde cholangiopancreatography (ERCP) pathway, but its relevance to pancreatic cancer-related biliary obstruction is often inferred from mixed malignant obstruction or general ERCP cohorts. We mapped clinical applicability and evidence maturity.

Methods: This narrative evidence map used PubMed and Europe PMC searches covering inception through 30 June 2026, supplemented by Scopus and Google Scholar coverage checks, without a pancreatic-cancer disease block. A second clinician independently verified study identities, key performance data, and applicability; independent duplicate screening and extraction were not documented. Learned imaging and structured-data models were separated from conventional regression/nomograms and non-learned image processing. Applicability was classified as direct, partly direct, indirect, or adjacent; prediction studies received structured critical appraisal.

Results: Of 234 unique primary-database records, 73 reports were included. Supplementary searches added 10 reports, for a total of 83. Directly applicable evidence was sparse. One small computed tomography convolutional model was predominantly pancreatic cancer-specific (65/70 patients); other evidence came from mixed malignant obstruction, general ERCP, or adjacent diagnostic populations. External or center-independent testing was uncommon, and performance often declined on external data. No model demonstrated improved durable drainage or cancer-treatment delivery in pancreatic cancer.

Conclusions: Current evidence supports technical feasibility for individual tasks. Clinical utility and patient benefit for pancreatic-cancer biliary drainage have not been established and require prospective, center-independent evaluation with clinically meaningful outcomes.

1. Introduction

Pancreatic cancer remains one of the most lethal solid tumors and a leading cause of cancer death [1]. Tumors of the pancreatic head and periampullary region frequently compress or invade the distal common bile duct (CBD), producing obstructive jaundice that delays systemic therapy, impairs nutrition, predisposes to cholangitis, and erodes quality of life. Endoscopic retrograde cholangiopancreatography (ERCP) is central to management because it permits transpapillary decompression, tissue sampling when required, and stent placement for palliation or optimization before oncological treatment [2, 3]. This review concerns distal malignant biliary obstruction caused by pancreatic cancer; hilar obstruction, which raises distinct questions of segmental drainage, is outside its scope.

Decision-making in this setting is demanding. The endoscopist must weigh stricture level and length, upstream dilation, papillary accessibility, duodenal invasion, expected survival, surgical candidacy, the chemotherapy schedule, and the risk of stent dysfunction, often under time pressure. Much of this reasoning is tacit, and outcomes vary with operator experience and center volume. Endoscopic ultrasound (EUS)-guided biliary drainage is increasingly used when ERCP fails or is technically unsuitable, adding a further decision node [4, 5].

Artificial intelligence (AI) is proposed as a way to make selected judgments more consistent, but the term is used loosely. In this review, a learned model estimates a mapping from data; machine learning (ML) includes non-linear algorithms such as random forests and gradient boosting, while deep learning (DL) includes multilayer neural networks commonly applied to images. Conventional or penalized regression and nomograms are reported as statistical comparators rather than counted as AI. Image processing or three-dimensional reconstruction without a learned component is also treated separately.

* Corresponding author: Ahmed Abdallah Salman, Internal Medicine Department, Faculty of Medicine, Cairo University, Cairo, Egypt. Emails: awea84@kasralainy.edu.eg, awea844@gmail.com

Published and owned by PubPorta Publishing LLC. Academic oversight and scholarly guidance are provided by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print): 3066-4004, ISSN (Online): 3066-4012. ©2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosting by ASIDE Journals.

Citation: Salman AA. Artificial Intelligence Applications Along the ERCP Pathway: An Evidence Map and Relevance to Pancreatic Cancer-Related Biliary Obstruction. ASIDE GI. 2026;2(4):86-98, doi:10.71079/ASIDE.GI.1128261077

Table 1: Structured comparison with recent narrative reviews of AI in pancreaticobiliary endoscopy.

Review	Scope	Search transparency	Evidence organization	Study-level distinction in current review
Araújo et al. 2025 [7]	EUS and ERCP across pancreaticobiliary disease	Narrative overview; search yield and record-level exclusions not reported	AI applications grouped by modality and disease	No ERCP-only model taxonomy, PC composition, split-unit audit, or study-level domain-informed appraisal
Jiang et al. 2024 [8]	Current and proposed AI applications in pancreaticobiliary endoscopy	Narrative review; no reproducible disease-block-free ERCP search	Broad technology and future-direction synthesis	No model-level direct/indirect applicability or cohort-overlap audit
Jain et al. 2025 [9]	Impact of AI across pancreaticobiliary endoscopy	Narrative review; heterogeneous diagnostic and therapeutic evidence	Descriptive survey of EUS, ERCP, and cholangioscopy	No separation of learned models from conventional regression/non-learned processing at study level
Current review	ERCP-pathway evidence mapped to pancreatic cancer-related biliary obstruction	Reproducible PubMed and Europe PMC flow through 30 Jun 2026; 234 primary-flow records screened; 444 Scopus records and 200 Google Scholar results checked; 10 additional reports	Technical/imaging, learned patient prediction, conventional comparator, and non-learned processing separated	Adds PC composition, directness, split unit, overlap, domain-informed appraisal, and figure source data in a narrative evidence map

AI, artificial intelligence; ERCP, endoscopic retrograde cholangiopancreatography; EUS, endoscopic ultrasound; PC, pancreatic cancer. This comparison describes scope and method rather than assigning quality scores.

Computer-aided detection is established in colonoscopy, and reported pancreaticobiliary applications now span cannulation, fluoroscopic interpretation, cholangioscopy, stent measurement, and prediction of post-procedural outcomes [6–9]. Three recent narrative reviews surveyed AI across pancreaticobiliary endoscopy, but differed in scope, search transparency, treatment of non-learned techniques, and study-level appraisal. Table 1 provides a structured comparison. The present review therefore maps ERCP-specific evidence, distinguishes direct from indirect relevance to pancreatic cancer, and asks what each task-specific system has actually demonstrated.

2. Methods

This is a narrative evidence map, not a systematic review or meta-analysis, and it is reported with reference to the Scale for the Assessment of Narrative Review Articles (SANRA) criteria for justification, aims, literature search, referencing, reasoning, and presentation [10]. No protocol was registered. We updated the search to map the evidence reproducibly and correct the evidence base. An exhaustive synthesis of effect estimates was outside the scope.

PubMed and Europe PMC were searched from inception through 30 June 2026. Scopus and Google Scholar were used as supplementary coverage checks with the same broad concepts. Scopus (444 records) and the first 200 Google Scholar results sorted by relevance were screened. The disease block used in the original search was removed to identify general ERCP evidence, which was then classified for applicability. The PubMed strategy combined ERCP, cholangiography, cholangioscopy, biliary stent, or biliary drainage terms with AI, ML, DL, neural-network, tree-based, radiomics, or computer-aided terms; the concepts were adapted to each source. PubMed and Europe PMC returned 224 and 176 records, respectively. After removal of 164 cross-database duplicates and two records whose exact first-publication dates were after the cutoff, 234 unique records were screened. After deduplication against the primary flow, 10 supplementary reports were retained in the evidence map. One conference report was excluded because full-report availability by the cutoff could not be established; its exclusion is recorded in Supplementary Data. We provide complete source-specific strings, the original cutoff, a flow summary, the primary screening log, the supplementary-inclusion audit, and full-text exclusions in Supplementary Appendix S1 and Supplementary Data.

We separated original learned-model studies from reviews, guidelines, and contextual clinical evidence. We also separated (i) learned imaging or technical systems, (ii) learned structured-data patient-prediction models, (iii) conventional or penalized regression and nomograms, and (iv) non-learned image processing. We classified applicability to pancreatic cancer-related ERCP as direct (predominantly pancreatic cancer), partly direct (mixed malignant obstruction with a pancreatic-cancer subset), indirect (general ERCP), or adjacent (endobiliary diagnostic or technical work that did not address the target therapeutic decisions). Overlapping cohorts were identified by center, date, sample size, and authors and were not treated as independent replication.

A second clinician independently verified the included-study identities, key performance data, and applicability to pancreatic cancer-related biliary obstruction against primary reports, and reviewed the final corrected dataset. The author and the second clinician checked all final inclusion decisions and extracted values against the primary reports. This verification pass was separate from screening and extraction;

independent duplicate screening and extraction were not documented. Technical imaging studies were examined for patient-, image-, and center-level separation and repeated-image leakage. Patient-prediction studies received structured critical appraisal informed by the four domains of the Prediction Model Risk of Bias Assessment Tool plus Artificial Intelligence (PROBAST+AI): participants and data sources, predictors, outcome, and analysis. Without a completed signaling-question record, no formal risk-of-bias grades are assigned. Decision-curve analysis was interpreted as estimation of net benefit under specified thresholds, not proof of clinical utility. No quantitative synthesis was attempted because tasks, populations, outcomes, and validation designs were heterogeneous. AI disclosure: OpenAI Codex assisted with figure coding and formatting.

3. Clinical Context: ERCP in Pancreatic Cancer-Related Biliary Obstruction

The immediate aims of ERCP are to restore bile flow, relieve jaundice and pruritus, treat or prevent cholangitis, and enable oncological treatment. Cross-sectional imaging and EUS generally define and stage the tumor beforehand, so ERCP here is therapeutic rather than diagnostic [2, 11]. Where the diagnosis is uncertain, brush cytology and intraductal biopsy have limited sensitivity, and digital single-operator cholangioscopy permits direct visualization and targeted sampling. In an established pancreatic head cancer with a typical distal stricture, it is rarely indicated; its role is confined to strictures that remain indeterminate after cross-sectional imaging and failed tissue sampling.

Treatment intent governs the decision to drain. In resectable disease managed by upfront surgery, routine preoperative drainage is avoided; the trial most often cited in support of this position used plastic stents in patients with moderate hyperbilirubinemia and predates both routine metal stenting and contemporary neoadjuvant pathways, so its conclusions should not be extended uncritically to those settings [12]. Drainage remains appropriate when surgery is delayed, when jaundice is severe or symptomatic, when cholangitis is present, or when neoadjuvant therapy is planned [11, 13].

Stent choice determines drainage durability. Plastic stents are inexpensive and easily exchanged but occlude sooner; self-expandable metal stents (SEMS) provide a larger lumen, longer patency, and fewer repeat procedures, and are preferred for unresectable disease and prolonged neoadjuvant treatment [14, 15]. In pooled comparisons of unresectable malignant biliary obstruction, SEMS reduced stent dysfunction from 46.8% to 21.6% and reintervention from 56.6% to 21.6% relative to plastic stents [2]. The choice between covered and uncovered SEMS remains uncertain: covered stents resist tumor ingrowth and are removable but migrate more often and may occlude the cystic duct, whereas uncovered stents migrate less but admit ingrowth, and pooled analyses have not established a decisive patency advantage for either design [16, 17].

Difficult biliary cannulation can prevent successful drainage in pancreatic cancer because of periampullary infiltration, duodenal narrowing, and distorted papillary orientation; it also increases the risk of post-ERCP pancreatitis (PEP) [18, 19]. Escalation options include double-guidewire cannulation, transpancreatic sphincterotomy, and needle-knife fistulotomy. PEP prophylaxis includes rectal non-steroidal anti-inflammatory drugs, pancreatic duct stenting in high-risk cases, and appropriate periprocedural hydration [19, 20]. The 2018 European Society of Gastrointestinal Endoscopy (ESGE) performance measures specify a minimum native-papilla cannulation rate of 90% in normal anatomy and a 95% target for expert centers. They set successful stent placement below the hilum at a minimum of 95%, with PEP rates below 10% as the minimum standard and below 5% as the target [21]. These standards provide context for evaluating procedural performance. Gastric outlet obstruction from duodenal invasion may make the papilla inaccessible and require enteral stenting or EUS-guided gastroenterostomy [5]. When biliary and duodenal obstruction coexist, biliary drainage is usually secured first or in the same session: a duodenal stent across the papilla can hinder later transpapillary access, whereas EUS-guided gastroenterostomy preserves that route [5]. EUS-guided biliary drainage provides rescue after failed ERCP and can serve as an alternative primary strategy in expert centers [2, 22–24].

These are the decisions that a proposed AI-supported pathway would have to improve. Figure 1 maps published models at the model level, separates direct from indirect applicability, and defines temporal, geographic, center-based, and investigator-independent validation.

4. Results

The broadened search identified a much larger general-ERCP and adjacent diagnostic literature than the original pancreatic-cancer-constrained search. The balance is clinically important: only a small minority of learned systems were developed in cohorts dominated by pancreatic cancer, and none evaluated an integrated ‘smart ERCP’ system. Table 2 summarizes representative learned models and explicitly records model taxonomy, pancreatic-cancer composition, split unit, validation type, operating metrics, and critical-appraisal considerations. Complete model-level data and the record-level audit are provided in Supplementary Data.

4.1. Pre-ERCP Planning and Procedural Difficulty

The most directly relevant cannulation study applied DL to endoscopic images from 531 patients at two Korean centers, including 104 patients with malignant strictures [25]. The ResNet classifier had a mean area under the receiver operating characteristic curve (AUROC) of 0.680 across fivefold cross-validation (standard deviation, 0.021; VGG, 0.657; DenseNet, 0.626). Difficult cannulation comprised 163 prolonged cannulations and six failed procedures, for a total of 169 difficult cases. Recall for difficult cannulation was 0.611, so approximately four in ten difficult cases were missed. These were cross-validation estimates, rather than

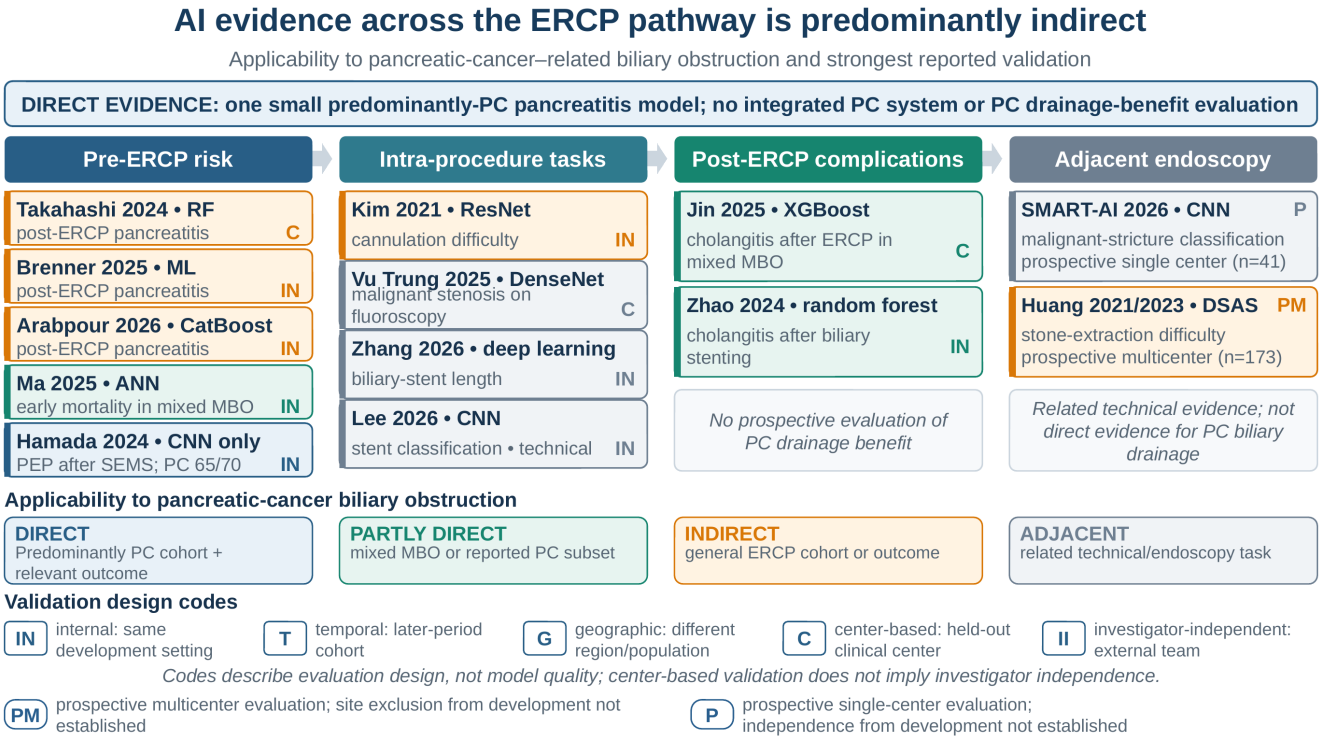


Figure 1: Selected models along the ERCP pathway and their applicability to pancreatic cancer-related biliary obstruction. Direct, predominantly pancreatic cancer; partly direct, mixed malignant biliary obstruction with a pancreatic-cancer subset; indirect, general ERCP; adjacent, related diagnostic or technical tasks. Validation codes: IN, internal evaluation in the development setting; T, temporal evaluation in later patients; G, geographic evaluation; C, evaluation at a center excluded from development; II, evaluation by investigators independent of development; P, prospective single-center evaluation with independence from development not established; PM, prospective multicenter evaluation without established exclusion of participating centers from development. These properties are not interchangeable, and none alone proves clinical benefit. The Hamada card refers to the pre-procedure CNN; its combined regression model also used procedure time. AI, artificial intelligence; ANN, artificial neural network; CNN, convolutional neural network; DSAS, difficulty scoring and assistance system; ERCP, endoscopic retrograde cholangiopancreatography; MBO, malignant biliary obstruction; ML, machine learning; PC, pancreatic cancer; PEP, post-ERCP pancreatitis; RF, random forest; SEMS, self-expandable metal stent. AI disclosure: OpenAI Codex assisted with formatting this schematic; its content is based on the cited studies.

results from a separate external test cohort. Pancreatic-cancer composition was not reported.

A prospective two-center comparison in 72 patients with pancreatic cancer-related obstruction found shorter cannulation and fluoroscopy times after manual three-dimensional reconstruction [26]. The report states that patients were randomly assigned, but does not describe sequence generation or allocation concealment. Differences in clinical success and adverse events were not statistically significant. Trained technicians created the reconstructions in commercial software without a learned component; the study is therefore classified as non-AI anatomical preparation.

Learned difficulty systems for bile-duct stone extraction have undergone multicenter development and prospective validation, and a later DL study addressed cannulation mechanics [27–29]. These studies are relevant to technical design and validation, but their stone-disease or general-ERCP populations make them indirect for pancreatic cancer. No published model predicts, before the procedure, whether ERCP, EUS-guided drainage, or percutaneous drainage should be selected for an individual patient with pancreatic cancer.

4.2. Cholangiographic and Cholangioscopic Interpretation

Fluoroscopic cholangiography defines the level of obstruction, stricture morphology, upstream dilation, guidewire position, and drainage adequacy, yet interpretation is operator-dependent and degraded by incomplete opacification, overlapping devices, respiratory motion, and variable image quality [6].

A DenseNet121 network distinguished malignant from benign bile-duct stenosis in 251 patients [30]. The cohort table lists 1,072 fluoroscopy images; the Results narrative implies 1,074 because it reports two more Leipzig images than the cohort table. The discrepancy is retained in the extraction notes. Pancreatic ductal adenocarcinoma accounted for 13 patients (5.2%). Internal-test AUROC was 0.90, but AUROC fell to 0.72 and 0.76 at two external centers, with recall falling to 0.55 and 0.44. Confidence intervals were not reported. Development and evaluation patients were disjoint, while reported metrics were calculated per image. Repeated images within patients therefore affect the effective precision of image-level estimates, without demonstrating train/test patient leakage.

A systematic review of AI for malignant biliary-stricture imaging found small, retrospective, heterogeneous studies [31]. More recent cholangioscopy systems have undergone multicenter clinical validation and prospective real-time evaluation, including the SMART-AI trial [32, 33]. These are important benchmarks for multi-device and live evaluation, but the task is diagnosis of indeterminate strictures rather than procedural management of established pancreatic cancer. They are therefore classified as adjacent evidence rather than proof of a smart therapeutic pathway.

For pancreatic cancer, useful intraprocedural outputs would include confirming guidewire passage, estimating stricture geometry, identifying incomplete drainage, and documenting stent position. No such function has been shown to improve durable drainage, reduce reintervention, reduce adverse events, or shorten time to oncological therapy.

4.3. Stent Measurement, Classification, and Selection

A DL system was developed using a dataset of 794 images from 431 patients to segment the duodenoscope, bile duct, and guidewire and estimate stent length [34]. Among 121 correctly detected strictures, 104 length estimates were within 1 cm of the fluoroscopic reference (86.0%; mean absolute error, 0.81 cm). External patient- or center-level evaluation could not be established from the available primary report. This demonstrates measurement agreement against a two-dimensional reference, but does not establish that the selected length reduces dysfunction, migration, inadequate drainage, or reintervention. Projectional foreshortening may affect both model and reference.

A separate DL classifier identified biliary-stent types in radiographs and used transfer learning to add another device class [35]. The original dataset contained 412 images from 151 patients with percutaneously placed stents; expansion added 76 images from 34 patients with endoscopically placed stents. The resulting 488-image, 185-patient dataset represents adjacent device-classification evidence across different drainage routes. It does not select an appropriate stent for a patient. No learned model has compared plastic with metal, covered with uncovered metal, or transpapillary with alternative drainage using patient-centered outcomes. Those decisions remain supported by clinical comparative evidence and patient context [14–17, 36].

4.4. Prediction of Adverse Events and Prognosis

The most directly applicable learned model used pre-procedure computed tomography (CT) images from 70 patients undergoing SEMS placement for distal malignant obstruction, 65 (92.9%) with pancreatic cancer; 21 developed pancreatitis [37]. The convolutional neural network (CNN) alone had an AUROC of 0.67. Adding its output to logistic regression yielded an AUROC of 0.74. Still, the study used only internal resampling, and the combined model included procedure time, which is unavailable for a purely pre-procedure decision. The small event count and absence of external validation limit inference despite direct clinical relevance.

Two learned models addressed post-ERCP cholangitis in malignant obstruction [38, 39]. In the multicenter study of 1,421 patients, the XGBoost AUROC fell from 0.965 in training to 0.767 in the internal test set and 0.727 at the external center [38]. The internal test cohort had 35 events among 308 patients, and the external cohort had 35 among 395. The development cohort contained 272 pancreatic cancers among 1,026 patients; the external cohort contained 89 pancreatic cancers among 395 patients (22.5%). In the single-center study of 218 patients, 62 (28.4%) had pancreatic cancer, and the optimized random forest achieved an AUROC of 0.89 in the designated test set [39]. However, test-set AUROC was used as the fitness function for feature selection, so this estimate does not represent evaluation on an untouched test set. Both studies used published cholangitis criteria (Tokyo Guidelines 2018 in both, with ESGE criteria also used by Zhao et al.). Procedure-dependent predictors also limit claims of pre-procedure use.

An artificial neural network (ANN) for early mortality after stenting was developed in 285 patients, 58 (20.4%) with pancreatic cancer, using a random same-center validation split. The report gives an ANN AUROC of 0.813 without specifying which cohort underlies that estimate [40]. A later multicenter survival report used a 245-patient Jiaxing development cohort and 92 patients at external centers [41]. Its development cohort probably overlaps substantially with the earlier 245-patient Jiaxing mortality report, given the shared center, recruitment period, and sample size [42]. Exact identity remains unresolved: the pancreatic-cancer counts in the two reports are incompatible with simply adding the 92 external patients. The reports are therefore not treated as independent replication. The earlier model used least absolute shrinkage and selection operator (LASSO) regression followed by logistic regression, not a non-linear learned algorithm; its test AUROC of 0.919 (95% confidence interval (CI), 0.818–1.000) came from 41 patients, including four deaths, and included post-ERCP complications and successful drainage. It cannot support a pre-ERCP decision and is retained as a conventional statistical comparator.

The general-ERCP literature is substantially larger for post-ERCP pancreatitis. It includes proof-of-concept, multicenter, registry, radiomics, multimodal, and externally evaluated models [43–53]. Several used patient-level or center-disjoint validation, but populations were dominated by stone disease or mixed indications and are indirect for pancreatic-cancer drainage. Reported discrimination ranged widely across datasets; the multicenter 12-trial calculator, for example, had AUROC 0.70 in cross-validation and 0.74 in a prospective pilot [47], whereas a 12-center external study reported marked between-center variation [48]. Such heterogeneity argues against importing a single headline AUROC into pancreatic-cancer care.

Recurrent biliary obstruction, migration, reintervention, drainage durability, and uninterrupted oncological therapy are the outcomes that would validate stent-related clinical usefulness. They have been characterized in conventional clinical studies [36, 54], but no AI model has prospectively improved them.

Table 2: Model-level evidence, applicability to pancreatic cancer, validation unit, and principal appraisal.

Study / task	Taxonomy	Cohort and PC composition	Split / validation	Key reported performance	Applicability and appraisal
Kim 2021 [25] Cannulation difficulty	Learned imaging / technical	531 patients; 104 malignant strictures; 169 difficult cases (163 prolonged, 6 failed); PC composition not established	Fivefold CV; no center-independent external cohort	ResNet mean AUROC 0.680 (SD 0.021); difficult-class recall 0.611	Indirect. Limited recall and no external test; image-based clinical prediction requires patient-grouped validation.
Huang 2021 → 2023 [27, 28] Stone extraction	Learned imaging / technical	Development: 1,954 cholangiograms; prospective validation: 173 patients; PC not target	Same model lineage; multicenter prospective validation	Predicted-difficult cases had longer extraction and lower single-session clearance	Indirect stone-disease evidence. Prospective validation, not independent model lineage.
Liu 2025 [29] Cannulation guidance	Learned imaging / technical	1,840 annotated images; PC composition not established	Image-level evaluation; external patient/center test not established	mAP 0.932	Indirect. Guidance mechanics, not demonstrated clinical outcome.
Vu Trung 2025 [30] Malignant stenosis	Learned fluoroscopy classification	251 patients; cohort table 1,072 images (Results implies 1,074); PDAC 13/251 (5.2%)	Patient-disjoint split; metrics per image; external cohorts 17 and 32 patients	AUROC 0.90 internal; 0.72 and 0.76 external; CI NR	Adjacent diagnostic evidence. External decline; small centers and correlated images affect precision.
Mascarenhas 2026 / SMART-AI 2026 [32, 33] Cholangioscopy	Learned endoscopic-video diagnosis	129 patients/135 examinations; SMART-AI prospective n=41; PC composition not established	Mascarenhas: multicenter clinical evaluation; SMART-AI: prospective single-center n=41, held-out-center independence not established	Patient-level AUROC 0.904; SMART-AI accuracy 0.878	Adjacent diagnostic evidence, not therapeutic PC ERCP.
Zhang 2026 [34] Stent length	Learned segmentation / measurement	431 patients; 794 images; Malignant/PC composition not established	Internal evaluation; patient-disjoint split and external patient/center evaluation not established	Within 1 cm in 104/121 detected strictures (86.0%); MAE 0.81 cm	Adjacent. Technical measurement agreement only; no dysfunction, migration, drainage, or reintervention outcome.
Lee 2026 [35] Stent identification	Learned radiographic classification	Original 151 patients/412 percutaneous-stent images; expanded with 34 patients/76 endoscopic-stent images	Fivefold CV and transfer learning; no independent external center	Vendor macro-AUROC 0.9828 (bootstrap 95% CI 0.9711–0.9923)	Adjacent mixed-route device task. Expanded dataset is not synthetic augmentation. Does not select a stent for a patient.
Hamada 2024 [37] PEP after SEMS	CNN + structured prediction	70 distal MBO; PC 65/70 (92.9%); PEP 21/70	Patient-grouped fivefold/leave-one-out internal validation; no external cohort	CNN AUROC 0.67; logistic+CNN: AUROC 0.74, PPV 0.85, accuracy 0.83, specificity 0.96	Direct. Small cohort and few events; no external validation. Combined model includes procedure time.
Jin 2025 [38] Post-ERCP cholangitis	XGBoost structured prediction	1,421 MBO; 135 events; development PC 272/1,026; external PC 89/395 (22.5%)	Center 1 train/test; Center 2 external n=395	AUROC 0.965/0.767/0.727; test 35/308 events; external 35/395; external Brier 0.0922	Partly direct. External performance decline and decision-time availability require attention.
Zhao 2024 [39] Post-stent cholangitis	Random forest structured prediction	218 malignant obstruction; PC 62/218 (28.4%); events 49	Same-center random split; test-set AUROC used in feature-selection fitness	Optimized test AUROC 0.89; earlier tuned random forest: AUROC 0.87, accuracy 0.86, F1 0.57	Partly direct. Optimized test-set performance is not an untouched evaluation; no external cohort.

Table 2: (continued)

Study / task	Taxonomy	Cohort and PC composition	Split / validation	Key reported performance	Applicability and appraisal
Ma 2025 [40] Early mortality	Artificial neural network	285 MBO; PC 58/285 (20.4%); deaths 57	Random same-center development 200 / validation 85	ANN AUROC 0.813 (evaluation cohort not specified); separate validation sensitivity 0.611, specificity 0.955	Partly direct. Same-center split; no center-independent evaluation; the report's external label does not establish externality.
Zhu 2026 [41] Survival	XGBoost accelerated-failure-time	337 MBO; PC 158/337 (46.9%)	Development 245; external centers 40 and 52	C-index 0.902 development; 0.722/0.705 external	Partly direct. Probable overlap with Zhu 2023; exact identity is unresolved because cancer composition differs.
Zhu 2023 [42] 30-day mortality	LASSO + logistic nomogram (conventional comparator)	245 MBO; PC 15/245 (6.1%); deaths 23	Single center; test n=41, including four deaths	Test AUROC 0.919 (95% CI 0.818–1.000)	Partly direct conventional comparator. Small test cohort and post-ERCP predictors limit pre-procedure use.
Archibugi 2023 [43] PEP	Gradient boosting	1,150 general ERCP; PEP 6.1%; PC composition not established	Prospective multicenter; internal hold-out	Test AUROC 0.671	Indirect. General ERCP; limited events and no independent external cohort.
Takahashi 2024 [44] PEP	Random forest	615 development (PC 107/615, 17.4%) plus 544 external (PC 41/544, 7.5%); no PC-specific performance	Patient-level external center	AUROC 0.821 cross-validation; 0.770 external	Indirect. Critical-appraisal notes in Supplementary Data; general ERCP population.
Chen 2024 [45] Papilla radiomics / PEP	DL-derived papilla radiomic score + logistic regression	2,038 plus external 334; PC composition not established	Multicohort; center-level external evaluation	AUROC 0.825–0.857	Indirect hybrid model. Learned imaging feature plus conventional final prediction; possible cohort overlap, exact identity unconfirmed.
Xu 2024 [46] Multimodal PEP	Multimodal random forest	1,916 multicenter ERCP cases; PC composition not established	Internal and external patient cohorts	Threshold metrics reported; full extraction in Supplementary Data	Indirect. General ERCP population; no PC-specific clinical-impact evaluation.
Brenner 2025 [47] PEP	Gradient boosting risk calculator	7,389 trial participants; PEP 8.6%; prospective pilot n=135; PC composition not established	Twelve RCT datasets; fivefold CV; prospective pilot	AUROC 0.70 CV (PPV 0.15, NPV 0.95); 0.74 prospective pilot	Indirect. Prospective technical evaluation, not randomized AI-impact evaluation.
Wang 2025 [48] PEP	Random forest	2,247 development (1,798 train/449 test); 674 patients at 12 external centers; stone disease	Center-level external evaluation	Internal test AUROC 0.947; external 0.576–0.913 (mean 0.768)	Indirect. Multiple external centers reveal heterogeneous performance; no PC-specific impact study.
Chen 2025 [49] PEP	ML feature selection + final logistic model	917 train, 392 internal, 366 and 83 external; stone disease	Patient/center cohorts	AUROC 0.845 reported for final model	Indirect; final prediction model is conventional logistic regression, not counted in AI numerator.
De 2025 [50] PEP	LightGBM and comparators	688; PEP 88; PC composition not established	Random 70/30 same-center split	Learned-model test AUROC 0.807–0.840; simplified score 0.75 in text/0.76 in Figure 5B	Indirect. Few events and same-center split; no external evaluation.

Table 2: (continued)

Study / task	Taxonomy	Cohort and PC composition	Split / validation	Key reported performance	Applicability and appraisal
Arabpour 2026 [51] PEP	CatBoost	Prospective registry n=1,190; PEP 170; PC composition not established	Tenfold cross-validation; no external center	AUROC 0.688; sensitivity 0.704, specificity 0.672, PPV 0.265, NPV 0.920	Indirect. Prospective data, internal resampling only.
Ping 2026 [52] PEP under NSAIDs	Fusion ML	Retrospective 991/57 events; prospective 124; PC composition not established	Prospective evaluation from same study setting	AUROC 0.84 pre-procedure; 0.86 post-procedure	Indirect. Prediction time must be separated; prospective technical evaluation is not clinical utility.
Jamil 2026 [53] PEP	Tabular, image, multimodal ML	Cleaned tabular cohort: 446 patients/27 PEP; multimodal test n=20 (5 PEP); PC composition not established	Fivefold CV plus small held-out multimodal set	XGBoost CV AUROC 0.786; multimodal test 0.57	Indirect. Very small multimodal test cohort; unstable discrimination.

ANN, artificial neural network; AUROC, area under the receiver operating characteristic curve; C-index, concordance index; CI, confidence interval; CNN, convolutional neural network; CV, cross-validation; DL, deep learning; ERCP, endoscopic retrograde cholangiopancreatography; F1, harmonic mean of precision and recall; LASSO, least absolute shrinkage and selection operator; MAE, mean absolute error; mAP, mean average precision; MBO, malignant biliary obstruction; ML, machine learning; N/A, not applicable; NPV, negative predictive value; NR, not reported; not established, not confirmed from the accessed primary material; NSAID, non-steroidal anti-inflammatory drug; PC, pancreatic cancer; PDAC, pancreatic ductal adenocarcinoma; PEP, post-ERCP pancreatitis; PPV, positive predictive value; RCT, randomized controlled trial; RF, random forest; SD, standard deviation; SEMS, self-expandable metal stent. Applicability: direct, predominantly PC; partly direct, mixed MBO with a PC subgroup; indirect, general ERCP; adjacent, related diagnostic or technical evidence. These are critical-appraisal notes informed by PROBAST+AI domains, not completed formal grades. Full domain notes, operating metrics and source links are in Supplementary Data.

: 10.71079/ASIDE.GI.112826107

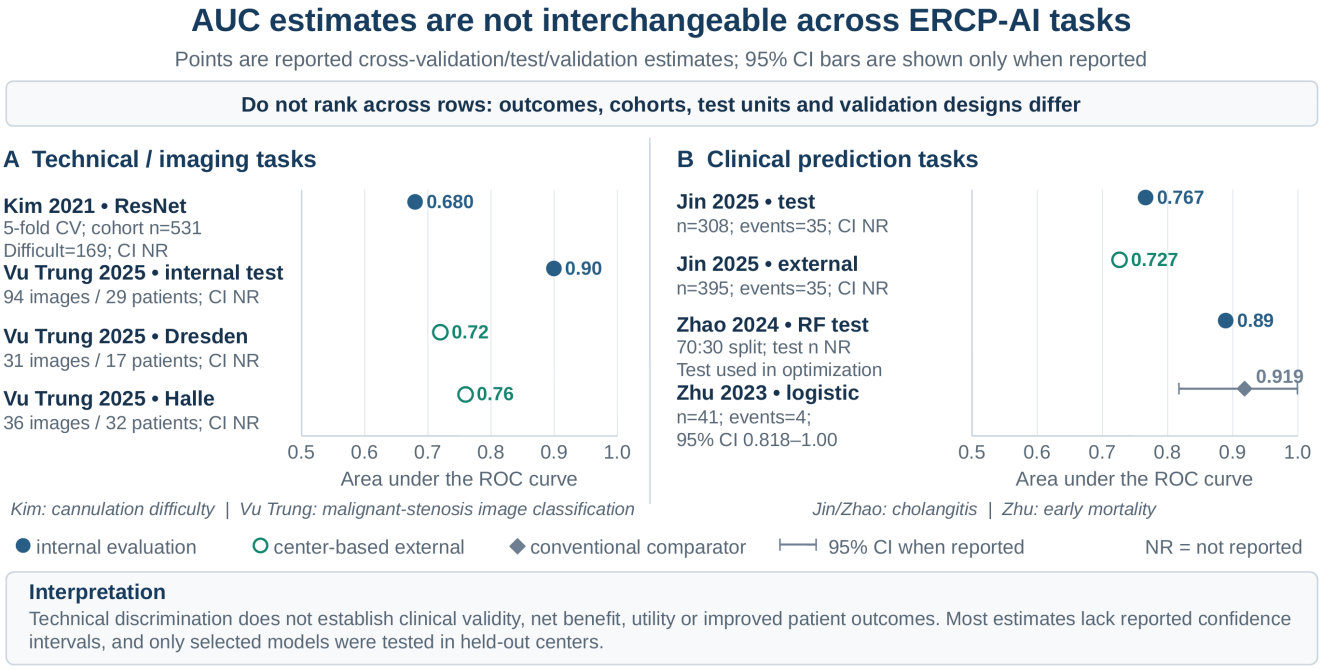


Figure 2: Selected reported discrimination across different tasks. The display includes the endoscopic cannulation classifier, the fluoroscopic malignant-stricture classifier, and mixed malignant-obstruction models for cholangitis or early mortality with extractable cross-validation or test estimates. Eight estimates are shown; Ma et al. are omitted because the report does not identify the cohort underlying the ANN AUROC. This is a descriptive subset, not an exhaustive comparison or pooled analysis. Technical/imaging and patient-prediction tasks are shown separately. Evaluation size, unit, event counts where available, validation design, and reported 95% confidence intervals accompany the estimates. Kim's point is a mean across fivefold cross-validation; 531 is the parent cohort, not a held-out test size. Vu Trung's estimates are per image; both image and patient denominators are displayed. Zhao's designated test set was used during feature optimization. Missing confidence intervals are not estimated. The logistic nomogram is a conventional comparator. ANN, artificial neural network; AUC, area under the receiver operating characteristic curve; CI, confidence interval; CV, cross-validation; ERCP, endoscopic retrograde cholangiopancreatography; MBO, malignant biliary obstruction; NR, not reported; PC, pancreatic cancer; RF, random forest; ROC, receiver operating characteristic. AI disclosure: OpenAI Codex assisted with chart formatting; plotted estimates derive from the cited publications.

4.5. Study-Level Appraisal

Technical systems and patient-prediction models require different forms of evaluation. Technical systems may show segmentation, classification, or measurement accuracy without establishing clinical validity. Patient-prediction models require reliable outcome definitions, predictor availability at the intended decision time, adequate events, calibration, and evaluation outside the development setting. Figure 2 therefore reports a defined subset of selected estimates and displays evaluation size and unit, event count, uncertainty, and validation type without ranking unlike tasks by AUROC.

The domain-informed appraisal identified limitations that require study-specific interpretation. Small event counts, same-center random splits, incomplete missing-data reporting, model-selection optimism, and sparse calibration limit confidence in several patient-prediction studies. In Zhao et al., explicit reuse of test-set AUROC during feature selection compromises the independence of the reported test estimate [39]. Retrospective design alone does not establish high risk of bias, and multicenter recruitment alone does not establish low risk. A predictor unavailable before ERCP restricts preprocedure applicability but is not automatically biased if the intended prediction time is later. Diagnostic prediction from images is within the scope of PROBAST+AI; pure device recognition, segmentation, or geometric measurement requires a separate technical appraisal. Patient-grouped splitting, locked preprocessing, and center-disjoint evaluation remain essential safeguards.

Calibration was reported less consistently than discrimination. When present, decision-curve analysis estimated net benefit across assumed thresholds; it did not show that clinicians used the model, changed management, or improved outcomes. Zhao et al. reported a training accuracy of 1.00 for the decision-tree and AdaBoost models [39]. Lower performance on test data raises concern about overfitting, although perfect training accuracy alone does not establish it. Likewise, absence of random allocation is not a defect of model development: randomization becomes relevant when a mature tool is tested for clinical impact. No identified system had demonstrated a randomized benefit for pancreatic-cancer biliary drainage or oncological-treatment delivery.

Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis plus Artificial Intelligence (TRIPOD+AI) is reporting guidance, PROBAST+AI is an appraisal tool for model studies, and Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (DECIDE-AI) addresses early live clinical evaluation of AI-based decision support [55–57]. They are complementary, not interchangeable. Future

reports should state the prediction time, unit of analysis, repeated-image handling, center and investigator independence, thresholds, prevalence, confusion matrices, sensitivity, specificity, predictive values, calibration, and model availability. Supplementary Data distinguishes information confirmed as unreported from information that could not be verified because source access or extraction was incomplete. Missing values are not inferred.

4.6. What a Credible Study of AI in ERCP Should Report

A study intended to change pancreatic-cancer practice should define the target population, treatment intent, decision point, and clinical action. Outcomes should include technical and clinical success, recurrent obstruction, migration, cholangitis, reintervention, and time to systemic therapy, with accepted definitions. Model development should be distinct from locked external evaluation; discrimination should include confidence intervals and be accompanied by calibration and threshold-specific operating data. Early live evaluation should describe the clinician-algorithm pair, overrides, failures, workflow burden, and safety [55–57]. Randomized comparison is justified only after those stages show a usable and stable intervention.

5. Discussion

ERCP poses distinct difficulties for AI development and deployment.

Volumes are far smaller than in colonoscopy. Fluoroscopic images are grayscale, projectional, cluttered by instruments, and heterogeneous across equipment and technique. Much of the decisive information is haptic and never recorded. Some labels are operator-dependent, and many structured predictors become available only during or after the procedure. An intraprocedural output must also be delivered without adding cognitive load, alert fatigue, or automation bias.

The colonoscopy precedent also warrants caution. In a multicenter observational study, non-AI-assisted adenoma detection declined after routine AI exposure, raising a possible deskilling signal [58]. The design cannot establish causality, but it shows why standalone accuracy and the performance of a clinician-algorithm system are different questions. Society guidance similarly emphasizes realistic expectations, training, and structured evaluation [6].

Regulatory classification, data governance, cybersecurity, post-deployment monitoring, liability, consent for data use, and the effect of automation on training must be addressed before clinical use. These implementation requirements do not convert offline model performance into clinical readiness.

5.1. Limitations and Future Directions

This is a narrative evidence map, not a registered systematic review. The primary database flow was built from PubMed and Europe PMC, with Scopus and Google Scholar used as supplementary coverage checks. Google Scholar screening was limited to the first 200 results sorted by relevance. The search was restricted to English-language full reports first published by 30 June 2026, and later studies are not captured. The review has one author. A second clinician independently verified the identities of included studies, key performance data, and pancreatic-cancer applicability against primary reports, and reviewed the final corrected dataset. The author and the second clinician checked all final inclusion decisions and extracted values against primary reports. Independent duplicate screening, extraction, and appraisal were not documented; selection error, missed studies, transcription error, and subjective judgments remain possible. The domain-informed critical appraisal is not a completed formal PROBAST+AI assessment.

The evidence base is heterogeneous in task, population, split unit, outcome, and validation design. Some full-length publications were accessible only as abstracts, which limited appraisal; publication type did not guarantee access to the complete text. Formal pooling would create a misleading summary. Pancreatic-cancer composition could not be established for several cohorts, and applicability cannot be inferred from a biliary task alone. The supplementary dataset distinguishes verified values, unreported items, and unresolved extractions. One supplementary conference report was excluded because full-report availability by the historical cutoff could not be established.

The practical value of a model would be realized within multidisciplinary care. Biliary drainage is planned against a treatment calendar: patients receiving neoadjuvant therapy need drainage that survives chemotherapy, whereas those proceeding rapidly to surgery weigh removability and speed differently [13, 15]. A risk estimate is actionable only if it reaches the team scheduling treatment and is linked to a defined response. Defining who acts, at what threshold, and with what alternative is part of the intervention.

Priorities follow directly: build multicenter, multi-device datasets with patient-grouped partitions; separate development from temporal, geographic, and investigator-independent evaluation; report calibration and threshold-specific operating data; and evaluate endpoints that matter to patients. Locked prospective evaluation should precede any impact trial; larger retrospective cohorts alone cannot establish clinical utility.

6. Conclusions

AI research now spans several isolated tasks along the ERCP pathway, but an integrated smart system for pancreatic cancer does not exist. The most directly applicable evidence comes from a small, internally resampled CT model; many other studies draw on mixed malignant obstruction, stone disease, or adjacent cholangioscopy. Technical feasibility has been shown for particular tasks such as image classification, segmentation, and measurement. Clinical utility and patient benefit for pancreatic-cancer biliary drainage have not been established. Progress requires center-independent prospective evaluation, explicit control of patient and image leakage, calibrated, threshold-specific reporting, and endpoints defined by durable drainage, fewer adverse events, fewer reinterventions, and timely cancer treatment.

7. Positioning Relative to Prior Reviews

This review does not claim an integrated pancreatic-cancer evidence base. Its added contribution is the structured comparison in Table 1, the search update without a disease block, explicit separation of learned models from conventional regression and non-learned processing, study-level pancreatic-cancer applicability, cohort-overlap checks, split-unit reporting, and structured critical appraisal informed by PROBAST+AI domains. The supplementary Scopus and Google Scholar checks added 10 reports retained in the evidence map; these expanded the evidence map but did not alter its clinical conclusion [59–68]. The resulting conclusion is deliberately narrower than a claim of clinical readiness. A prospective study reported lower radiation exposure with AI-assisted fluoroscopy [68]; this adjacent technical-safety result does not establish improved drainage or oncological outcomes in pancreatic cancer.

Conflicts of Interest

The author declares no conflict of interest.

Funding Source

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

The author thanks Ahmed Elewa, MD, for independently verifying the included-study identities, key performance data, and pancreatic-cancer applicability against primary reports, and reviewing the final corrected dataset.

Informed Consent

Not applicable. No identifiable individual patient data are reported.

Institutional Review Board (IRB)

Not applicable. This review uses previously published aggregate information and involved no recruitment, intervention, or collection of new patient data; therefore, institutional review board approval was not required.

Large Language Model

Claude (Anthropic; Opus 5) assisted with language editing, proofreading, search-query development, and finding papers. Claude use ended on 25 July 2026. It did not assist with screening, eligibility decisions, or data extraction. OpenAI Codex (OpenAI; GPT-6) assisted with figure formatting. The author and Ahmed Elewa checked all final inclusion decisions and extracted values against primary reports. The author retains responsibility.

Authors Contribution

The author conceived the review, undertook the literature search and evidence appraisal, and drafted and revised the manuscript. The author approved the final manuscript and is responsible for its content.

Data Availability

No new primary patient-level data were generated. Supplementary Appendix S1 contains the complete search strategies, flow summary, eligibility and classification rules, and verification approach. Supplementary Data contains the record-level screening log, included-study extraction, cohort-overlap audit, domain-informed critical-appraisal notes, and source data for Figures 1 and 2; all source information is traceable to the cited publications.

References

1. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin. 2025;75(1):10-45. [PMID: 39817679, https://doi.org/10.3322/caac.21871].
2. Itonaga M, Kitano M. Endoscopic biliary drainage for distal bile duct obstruction due to pancreatic cancer. Clin Endosc. 2025;58(1):40-52. [PMID: 39322288, PMCID: PMC11837563, https://doi.org/10.5946/ce.2023.294].
3. Paik WH, Park DH. Endoscopic management of malignant biliary obstruction. Gastrointest Endosc Clin N Am. 2024;34(1):127-40. [PMID: 37973224, https://doi.org/10.1016/j.giec.2023.07.004].
4. Barbosa EC, Santo PADE, Baraldo S, Nau AL, Meine GC. EUS-versus ERCP-guided biliary drainage for malignant biliary obstruction: a systematic review and meta-analysis of randomized controlled trials. Gastrointest Endosc. 2024;100(3):395-405.e8. [PMID: 38648989, https://doi.org/10.1016/j.gie.2024.04.019].
5. Chantarojanasiri T, Ratanachu-Ek T. Biliary drainage in pancreatic cancer with duodenal invasion: which route is the best? Clin Endosc. 2025;58(1):82-4. [PMID: 39895272, PMCID: PMC11837551, https://doi.org/10.5946/ce.2024.301].
6. Messmann H, Bisschops R, Antonelli G, Libânio D, Sinonquel P, Abdelrahim M, et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) position statement. Endoscopy. 2022;54(12):1211-31. [PMID: 36270318, https://doi.org/10.1055/a-1950-5694].
7. Araújo CC, Frias J, Mendes F, Martins M, Mota J, Almeida MJ, et al. Unlocking the potential of AI in EUS and ERCP: a narrative review for pancreaticobiliary disease. Cancers (Basel). 2025;17(7):1132. [PMID: 40227709, PMCID: PMC11988021, https://doi.org/10.3390/cancers17071132].
8. Jiang H, Ye LS, Yuan XL, Luo Q, Zhou NY, Hu B. Artificial intelligence in pancreaticobiliary endoscopy: current applications and future directions. J Dig Dis. 2024;25(9-10):564-72. [PMID: 39740251, https://doi.org/10.1111/1751-2980.13324].
9. Jain A, Pabba M, Jain A, Singh S, Ali H, Vinayek R, et al. Impact of artificial intelligence on pancreaticobiliary endoscopy. Cancers (Basel). 2025;17(3):379. [PMID: 39941748, PMCID: PMC11815774, https://doi.org/10.3390/cancers17030379].
10. Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4:5. [PMID: 30962953, https://doi.org/10.1186/s41073-019-0064-8].
11. Conroy T, Pfeiffer P, Vilgrain V, Lamarca A, Seufferlein T, O'Reilly EM, et al. Pancreatic cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023;34(11):987-1002. [PMID: 37678671, https://doi.org/10.1016/j.annonc.2023.08.009].
12. van der Gaag NA, Rauws EA, van Eijck CH, Bruno MJ, van der Harst E, Kubben FJ, et al. Preoperative biliary drainage for cancer of the head

of the pancreas. N Engl J Med. 2010;362(2):129-37. [PMID: 20071702, https://doi.org/10.1056/NEJMoa0903230].

13. Inoue T, Nakamura M, Ito K. Biliary drainage during neoadjuvant chemotherapy in pancreatic cancer: evidence and practical recommendations. Cancers (Basel). 2026;18(3):467. [PMID: 41681939, PMCID: PMC12896535, https://doi.org/10.3390/cancers18030467].
14. Dumonceau JM, Tringali A, Papanikolaou IS, Blero D, Mangiavillano B, Schmidt A, et al. Endoscopic biliary stenting: indications, choice of stents, and results: European Society of Gastrointestinal Endoscopy (ESGE) clinical guideline - updated October 2017. Endoscopy. 2018;50(9):910-30. [PMID: 30086596, https://doi.org/10.1055/a-0659-9864].
15. Lyu Y, Ye S, Wang B. Comparison of metal versus plastic stent for preoperative biliary drainage in patients with pancreatic cancer undergoing neoadjuvant therapy: a meta-analysis and systematic review. BMC Gastroenterol. 2023;23(1):235. [PMID: 37438761, PMCID: PMC10337088, https://doi.org/10.1186/s12876-023-02874-5].
16. Tringali A, Hassan C, Rota M, Rossi M, Mutignani M, Aabakken L. Covered vs. uncovered self-expandable metal stents for malignant distal biliary strictures: a systematic review and meta-analysis. Endoscopy. 2018;50(6):631-41. [PMID: 29342491, https://doi.org/10.1055/s-0043-125062].
17. Nakai Y, Isayama H, Wang HP, Rerknimitr R, Khor C, Yasuda I, et al. International consensus statements for endoscopic management of distal biliary stricture. J Gastroenterol Hepatol. 2020;35(6):967-79. [PMID: 31802537, PMCID: PMC7318125, https://doi.org/10.1111/jgh.14955].
18. Noda J, Takano Y, Tamai N, Yamawaki M, Azami T, Niiya F, et al. Difficult biliary cannulation during endoscopic retrograde cholangiopancreatography for distal malignant biliary obstruction caused by pancreatic cancer: an observational study. DEN Open. 2025;5(1):e70092. [PMID: 40046941, PMCID: PMC11880624, https://doi.org/10.1002/deo2.70092].
19. Testoni PA, Mariani A, Aabakken L, Arvanitakis M, Bories E, Costamagna G, et al. Papillary cannulation and sphincterotomy techniques at ERCP: European Society of Gastrointestinal Endoscopy (ESGE) clinical guideline. Endoscopy. 2016;48(7):657-83. [PMID: 27299638, https://doi.org/10.1055/s-0042-108641].
20. Dumonceau JM, Kapral C, Aabakken L, Papanikolaou IS, Tringali A, Vanbiervliet G, et al. ERCP-related adverse events: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy. 2020;52(2):127-49. [PMID: 31863440, https://doi.org/10.1055/a-1075-4080].
21. Domagk D, Oppong KW, Aabakken L, Czakó L, Gyökeres T, Manes G, et al. Performance measures for ERCP and endoscopic ultrasound: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy. 2018;50(11):1116-27. [PMID: 30340220, https://doi.org/10.1055/a-0749-8767].
22. Paik WH, Lee TH, Park DH, Choi JH, Kim SO, Jang S, et al. EUS-guided biliary drainage versus ERCP for the primary palliation of malignant biliary obstruction: a multicenter randomized clinical trial. Am J Gastroenterol. 2018;113(7):987-97. [PMID: 29961772, https://doi.org/10.1038/s41395-018-0122-8].
23. ASGE Standards of Practice Committee, Pawa S, Marya NB, Thiruvengadam NR, Ngamruengphong S, Baron TH, et al. American Society for Gastrointestinal Endoscopy guideline on the role of therapeutic EUS in the management of biliary tract disorders: summary and recommendations. Gastrointest Endosc. 2024;100(6):967-79. [PMID: 39078360, https://doi.org/10.1016/j.gie.2024.03.027].
24. Lambin T, Leblanc S, Napoléon B. Advances in EUS-guided biliary drainage for the management of pancreatic cancer. Cancers (Basel). 2025;17(21):3428. [PMID: 41228222, PMCID: PMC12608985, https://doi.org/10.3390/cancers17213428].
25. Kim T, Kim J, Choi HS, Kim ES, Keum B, Jeen YT, et al. Artificial intelligence-assisted analysis of endoscopic retrograde cholangiopancreatography image for identifying ampulla and difficulty of selective cannulation. Sci Rep. 2021;11(1):8381. [PMID: 33863970, PMCID: PMC8052314, https://doi.org/10.1038/s41598-021-87737-3].
26. Fan M, Shang G, Liu S, Li T, Li D, Ding Z, et al. 3D-assisted ERCP enhances cannulation efficiency in malignant biliary obstruction: a prospective two-center study. Int J Surg. 2026;112(1):808-14. [PMID: 41092424, PMCID: PMC12825914, https://doi.org/10.1097/JS9.0000000000003302].
27. Huang L, Lu X, Huang X, Zou X, Wu L, Zhou Z, et al. Intelligent difficulty scoring and assistance system for endoscopic extraction of common bile duct stones based on deep learning: multicenter study. Endoscopy. 2021;53(5):491-8. [PMID: 32838430, https://doi.org/10.1055/a-1244-5698].
28. Huang L, Xu Y, Chen J, Liu F, Wu D, Zhou W, et al. An artificial intelligence difficulty scoring system for stone removal during ERCP: a prospective validation. Endoscopy. 2023;55(1):4-11. [PMID: 35554877, https://doi.org/10.1055/a-1850-6717].
29. Liu Y, Chen X, Zuo S. A deep learning-driven method for safe and effective ERCP cannulation. Int J Comput Assist Radiol Surg. 2025;20(5):913-22. [PMID: 39920403, https://doi.org/10.1007/s11548-025-03329-w].
30. Vu Trung K, Hollenbach M, Veldhuizen GP, Saldanha OL, Garbe J, Rosendahl J, et al. Deep learning-based detection of malignant bile duct stenosis in fluoroscopy images of endoscopic retrograde cholangiopancreatography. Digestion. 2025;106(4):287-302. [PMID: 39675349, PMCID: PMC12324797, https://doi.org/10.1159/000543049].
31. Njei B, McCarty TR, Mohan BP, Fozo L, Navaneethan U. Artificial intelligence in endoscopic imaging for detection of malignant biliary strictures and cholangiocarcinoma: a systematic review. Ann Gastroenterol. 2023;36(2):223-30. [PMID: 36864938, PMCID: PMC9932867, https://doi.org/10.20524/aog.2023.0779].
32. Mascarenhas M, Widmer J, Mendes F, Ribeiro T, Pinto da Costa A, Agudo B, et al. Clinical validation of AI-assisted evaluation of indeterminate biliary strictures in digital-single operator cholangioscopy: a transcontinental multicentric study. Clin Transl Gastroenterol. 2026;17(6):e01015. [PMID: 41805080, https://doi.org/10.14309/ctg.0000000000001015].
33. Marya NB, Powers PD, Marcello M, Rau P, Nasser-Ghodsi N, Marshall C, et al. Classification of biliary strictures using real-time cholangioscopy artificial intelligence: the SMART-AI trial. Clin Gastroenterol Hepatol. 2026;24(11):3011-21. [PMID: 41966465, https://doi.org/10.1016/j.cgh.2026.03.033].
34. Zhang WL, Shao XJ, Dong XY, Shao HT, Li GC, Li Z, et al. Artificial intelligence-assisted biliary stent length selection for common bile duct strictures in endoscopic retrograde cholangiopancreatography: model development and validation. Hepatobiliary Pancreat Dis Int. 2026;25(1):76-82. [PMID: 41271517, https://doi.org/10.1016/j.hbpd.2025.11.004].
35. Lee J, Shin BK, Yu GH, Dang TV, Yoon CJ, Lee JH, et al. Deep learning-based biliary stent classification and transfer learning adaptation to an additional stent type. Eur Radiol Exp. 2026;10(1):85. [PMID: 42257782, PMCID: PMC13247018, https://doi.org/10.1186/s41747-026-00749-4].
36. Tamura T, Yamai T, Uza N, Yamasaki T, Masuda A, Tomooka F, et al. Adverse events of self-expandable metal stent placement for malignant distal biliary obstruction: a large multicenter study. Gastrointest Endosc. 2024;99(1):61-72.e8. [PMID: 37598864, https://doi.org/10.1016/j.gie.2023.08.004].
37. Hamada T, Yasaka K, Nakai Y, Fukuda R, Hakuta R, Ishigaki K, et al. Computed tomography-based prediction of pancreatitis following biliary metal stent placement with the convolutional neural network. Endosc Int Open. 2024;12(6):E772-80. [PMID: 38904060, PMCID: PMC11188753, https://doi.org/10.1055/a-2298-0147].
38. Jin H, Sun X, Fu C, Fan C, Chen J, Zhang Z, et al. Machine learning-based prediction model for post-ERCP cholangitis in patients with malignant biliary obstruction: a retrospective multicenter study. Surg Endosc. 2025;39(8):5107-26. [PMID: 40634730, PMCID: PMC12287249, https://doi.org/10.1007/s00464-025-11937-5].
39. Zhao R, Gu L, Ke X, Deng X, Li D, Ma Z, et al. Risk prediction of cholangitis after stent implantation based on machine learning. Sci Rep. 2024;14(1):13715. [PMID: 38877118, PMCID: PMC11178872, https://doi.org/10.1038/s41598-024-64734-w].
40. Ma Y, Qi J, Zhang X, Liu K, Liu Y, Yu X, et al. Development and application of an early warning model for predicting early mortality following stent placement in malignant biliary obstruction: a comparative analysis of logistic regression and artificial neural network approaches. Oncol Lett. 2025;29(5):237. [PMID: 40166368, PMCID: PMC11956143, https://doi.org/10.3892/ol.2025.14983].

41. Zhu Z, Bian C, Zhao L, Li W, Hu K, Song Q, et al. Predicting postoperative survival in patients with malignant biliary obstruction using an interpretable machine learning model: a multicenter study. Cancer Med. 2026;15(3):e71692. [PMID: 41789602, PMCID: PMC12964315, https://doi.org/10.1002/cam4.71692].
42. Zhu Z, Hu K, Zhao F, Liu W, Zhou H, Zhu Z, et al. Machine learning-based nomogram for 30-day mortality prediction for patients with unresectable malignant biliary obstruction after ERCP with metal stent: a retrospective observational cohort study. BMC Surg. 2023;23(1):260. [PMID: 37649027, PMCID: PMC10470194, https://doi.org/10.1186/s12893-023-02158-5].
43. Archibugi L, Ciarfaglia G, Cárdenas-Jaén K, Poropat G, Korpela T, Maisonneuve P, et al. Machine learning for the prediction of post-ERCP pancreatitis risk: a proof-of-concept study. Dig Liver Dis. 2023;55(3):387-93. [PMID: 36344369, https://doi.org/10.1016/j.dld.2022.10.005].
44. Takahashi H, Ohno E, Furukawa T, Yamao K, Ishikawa T, Mizutani Y, et al. Artificial intelligence in a prediction model for postendoscopic retrograde cholangiopancreatography pancreatitis. Dig Endosc. 2024;36(4):463-72. [PMID: 37448120, https://doi.org/10.1111/den.14622].
45. Chen K, Lin H, Zhang F, Chen Z, Ying H, Cao L, et al. Duodenal papilla radiomics-based prediction model for post-ERCP pancreatitis using machine learning: a retrospective multicohort study. Gastrointest Endosc. 2024;100(4):691-702.e9. [PMID: 38583542, https://doi.org/10.1016/j.gie.2024.03.031].
46. Xu Y, Dong Z, Huang L, Du H, Yang T, Luo C, et al. Multistep validation of a post-ERCP pancreatitis prediction system integrating multimodal data: a multicenter study. Gastrointest Endosc. 2024;100(3):464-72.e17. [PMID: 38583541, https://doi.org/10.1016/j.gie.2024.03.033].
47. Brenner T, Kuo A, Sperna Weiland CJ, Kamal A, Elmunzer BJ, Luo H, et al. Development and validation of a machine learning-based, point-of-care risk calculator for post-ERCP pancreatitis and prophylaxis selection. Gastrointest Endosc. 2025;101(1):129-38.e0. [PMID: 39147103, https://doi.org/10.1016/j.gie.2024.08.009].
48. Wang G, Sun Q, Zhu H, Qiao S, Xu P, He X, et al. Development and external validation of a model for post-endoscopic retrograde cholangiopancreatography pancreatitis. iScience. 2025;28(6):112570. [PMID: 40491489, PMCID: PMC12146526, https://doi.org/10.1016/j.isci.2025.112570].
49. Chen K, Wang L, Wang X, Yang L, Zhang X, Lin Y, et al. Machine learning-derived predictive model for post-ERCP pancreatitis in patients with common bile duct stones: a retrospective multicenter study. Surg Endosc. 2025;39(12):8171-83. [PMID: 40962918, https://doi.org/10.1007/s00464-025-12169-3].
50. De T, Du G, Yin H, Wang H, Wang W, Ma T, et al. Development and validation of a practical prediction model for post-ERCP pancreatitis using machine learning. Front Surg. 2025;12:1628956. [PMID: 41256833, PMCID: PMC12620498, https://doi.org/10.3389/fsurg.2025.1628956].
51. Arabpour E, Sadeghi A, Rastegar R, Mohammadi P, Safavi-Naini SAA, Moghadam PK, et al. Predicting post-ERCP pancreatitis using machine learning: risk stratification and feature importance analysis. J Hepatobiliary Pancreat Sci. 2026;33(5):418-30. [PMID: 41492120, https://doi.org/10.1002/jhbp.70053].
52. Ping W, Cao P, Li J, Zhao B, Lin Y, Liu C, et al. Fusion prediction model for post-ERCP pancreatitis under NSAIDs prophylaxis. Surg Endosc. 2026;40(8):6766-79. [PMID: 42298037, https://doi.org/10.1007/s00464-026-12960-w].
53. Jamil A, Nazir W, Altaf A, Niaz SK. AI-based prediction of post-ERCP pancreatitis: a comparative study using tabular, image, and multimodal data. Diagnostics (Basel). 2026;16(12):1824. [PMID: 42351482, PMCID: PMC13298505, https://doi.org/10.3390/diagnostics16121824].
54. Takeda T, Sasaki T, Mie T, Furukawa T, Kanata R, Kasuga A, et al. Novel risk factors for recurrent biliary obstruction and pancreatitis after metallic stent placement in pancreatic cancer. Endosc Int Open. 2020;8(11):E1603-10. [PMID: 33140016, PMCID: PMC7581474, https://doi.org/10.1055/a-1244-1989].
55. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. [PMID: 38626948, PMCID: PMC11019967, https://doi.org/10.1136/bmj-2023-078378].
56. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. [PMID: 40127903, https://doi.org/10.1136/bmj-2024-082505].
57. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. [PMID: 35585198, https://doi.org/10.1038/s41591-022-01772-9].
58. Budzyń K, Romańczyk M, Kitala D, Kołodziej P, Bugajski M, Adami HO, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol. 2025;10(10):896-903. [PMID: 40816301, https://doi.org/10.1016/S2468-1253(25)00133-5].
59. Blum J, Hunn S, Smith J, Chan FY, Turner R. Using artificial intelligence to predict choledocholithiasis: can machine learning models abate the use of MRCP in patients with biliary dysfunction? ANZ J Surg. 2024;94(7-8):1260-5. [PMID: 38525849, https://doi.org/10.1111/ans.18950].
60. Shao F, Huang Q, Wang C, Qiu L, Hu YG, Zha SY. Artificial neural networking model for the prediction of early occlusion of bilateral plastic stent placement for inoperable hilar cholangiocarcinoma. Surg Laparosc Endosc Percutan Tech. 2018;28(2):e54-8. [PMID: 29252936, https://doi.org/10.1097/SLE.0000000000000502].
61. Wang Z, Yuan H, Lin K, Zhang Y, Xue Y, Liu P, et al. Artificial intelligence-empowered assessment of bile duct stone removal challenges. Expert Syst Appl. 2024;258:125146. [https://doi.org/10.1016/j.eswa.2024.125146].
62. Sun K, Li M, Shi Y, He H, Li Y, Sun L, et al. Convolutional neural network for identifying common bile duct stones based on magnetic resonance cholangiopancreatography. Clin Radiol. 2024;79(7):553-8. [PMID: 38616474, https://doi.org/10.1016/j.crad.2024.02.018].
63. Sugimoto Y, Kurita Y, Kuwahara T, Satou M, Meguro K, Hosono K, et al. Diagnosing malignant distal bile duct obstruction using artificial intelligence based on clinical biomarkers. Sci Rep. 2023;13(1):3262. [PMID: 36828831, PMCID: PMC9958195, https://doi.org/10.1038/s41598-023-28058-5].
64. Phuangthongkham P, Angsuwatcharakon P, Kulpatcharapong S, Vateekul P, Rerknimitr R. Real-time identification of malignant biliary strictures on cholangioscopy images using explainable convolutional neural networks with heatmaps. IEEE Access. 2023;11:49943-56. [https://doi.org/10.1109/ACCESS.2023.3276642].
65. Martin G, Becq A, Bloch I, Camus M, Pinna A, Szewczyk J. Comparison of loss functions for guidewire automatic detection on a limited dataset of fluoroscopic images for endoscopic retrograde cholangiopancreatography. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). IEEE; 2024. p. 1-4. [https://doi.org/10.1109/ISBI56570.2024.10635427].
66. Wang Z, Lin K, Zheng M, Gong L, Chen Z, Wu M. Accurate measurement of key structures in CBD patients using deep learning. Biomed Signal Process Control. 2025;100:106979. [https://doi.org/10.1016/j.bspc.2024.106979].
67. Thomas L, Manoj Kumar MV, Annappa B, Arun S, Mubin A. Prediction of gallstone disease progression using modified cascade neural network. In: Somani AK, Srivastava S, Mundra A, Rawat S, editors. Proceedings of First International Conference on Smart System, Innovations and Computing. Singapore: Springer; 2018. p. 729-38. [https://doi.org/10.1007/978-981-10-5828-8_69].
68. Bang JY, Hough M, Hawes RH, Varadarajulu S. Use of artificial intelligence to reduce radiation exposure at fluoroscopy-guided endoscopic procedures. Am J Gastroenterol. 2020;115(4):555-61. [PMID: 32195731, https://doi.org/10.14309/ajg.0000000000000565].