ORIGINAL RESEARCH PAPER
Volume - 14 | Issue - 08 | August - 2025 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

ARTIFICIAL INTELLIGENCE-DRIVEN DIAGNOSTIC TOOLS IN GASTROENTEROLOGY: A SYSTEMATIC REVIEW OF CLINICAL APPLICATIONS AND ACCURACY

Gastroenterology

Dr. Payal Kaw Assistant Professor, Surgical Gastroenterology, SGPGIMS, Lucknow.

Vipin Kumar Pandey* Assistant Professor, Gastroenterology, SGPGIMS, Lucknow. *Corresponding Author

ABSTRACT

Background: Artificial intelligence (AI) has rapidly emerged as a transformative tool in gastroenterology, providing significant enhancements in diagnostic accuracy and clinical efficiency. **Aim:** This systematic review aims to investigate AI-based diagnostic tools utilized in gastroenterology, with a focus on their clinical integration, usability, and validation status. **Materials And Methods:** A comprehensive literature search was conducted using PubMed and Medline, following the PRISMA 2020 flowchart. The focus was on AI-based diagnostic tools and their applications in gastroenterology over the past decade. **Results:** We identified and analysed 22 studies published over the past ten years that evaluated artificial intelligence (AI)–based models for diagnosis, prognostication, and quality improvement across a wide range of gastrointestinal (GI) diseases. With a few exceptions, the majority of the studies indicated that integrating AI into existing diagnostic modalities enhanced diagnostic accuracy. **Conclusion:** AI-based diagnostic tools demonstrate significant advancements in the detection of gastrointestinal (GI) diseases, improving accuracy across various modalities. Although the majority of studies indicate favourable performance metrics, wider clinical adoption necessitates standardized validation, multicentre trials, and enhanced transparency of the models.

KEYWORDS

Artificial Intelligence, Gastrointestinal Diseases, Diagnostic Accuracy, Endoscopy, Machine Learning.

INTRODUCTION

Gastroenterology, a field heavily dependent on visual diagnostics and intricate data interpretation, has undergone a transformative shift with the integration of artificial intelligence (AI) [1]. The global burden of gastrointestinal (GI) diseases—including colorectal cancer, inflammatory bowel disease, gastritis, and Barrett's oesophagus—continues to increase, particularly in low- and middle-income countries [2]. Despite advancements in diagnostic modalities such as endoscopy and histopathology, challenges remain in achieving timely, accurate, and cost-effective diagnoses due to resource limitations, interobserver variability, and data overload [3].

AI, particularly machine learning (ML) and deep learning (DL) algorithms, presents a promising solution to these challenges. These tools have the potential to reduce diagnostic delays, enhance consistency, and assist clinicians in high-volume settings with real-time decision-making. In gastroenterology, AI has been utilized for tasks such as polyp detection during colonoscopy, classification of gastritis from endoscopic images, and prediction of tumour invasion depth in gastric cancer [4-6]. Studies have demonstrated that AI models frequently outperform human experts, achieving diagnostic accuracies exceeding 90% in specific applications [5].

The potential of AI extends far beyond image recognition. It includes clinical data analysis, biomarker integration, and predictive modelling, all of which can significantly enhance personalized medicine approaches. However, despite a surge in research and development, the clinical translation of these technologies remains fragmented and inconsistent. This inconsistency is primarily due to methodological discrepancies, a lack of standardized validation protocols, and concerns regarding the generalizability of findings across diverse populations [7].

This systematic review examines AI-based diagnostic tools utilized in gastroenterology, with an emphasis on their clinical integration, usability, and validation status.

By synthesizing existing literature, this review offers a critical evaluation of the role of AI in gastroenterology and its potential to transform diagnostic workflows, enhance patient outcomes, and mitigate healthcare disparities.

MATERIALS AND METHODS

A comprehensive literature search was conducted across PubMed and Medline for published literature following the PRISMA 2020 flowchart, focusing on AI-based diagnostic tools and their application in gastroenterology over the past decade. A literature search was conducted using a combination of MeSH terms and keywords:
("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning") AND ("Gastroenterology" OR "Endoscopy" OR "Colonoscopy" OR "Capsule Endoscopy" OR "GI Imaging") AND ("Diagnostic Accuracy" OR "Sensitivity" OR "Specificity" OR "AUROC" OR "Clinical Utility")

The search was supplemented by manually screening the reference lists of eligible articles. Two reviewers independently screened titles, abstracts, and full texts. Disagreements were resolved by consensus. A PRISMA flow diagram (Figure 1) summarizes the selection process. To generate a high-quality review with a low risk of bias, we included only clinical trials, including randomized clinical trials (RCTs) and prospective studies that met the following criteria:

- **Population:** Patients undergoing diagnostic procedures for GI disorders
- **Intervention:** AI-based tools like CADx (Computer-Aided Diagnosis), CADe (Computer-Aided Detection), radiomics, NLP (Natural Language Processing), and LLMs (Large Language Models).
- **Outcomes:** Diagnostic accuracy, clinical impact, workflow integration

A total of 2,181 articles were identified. After removing articles that didn't meet the above criteria, 50 records were screened for assessment. Twenty-two studies met all eligibility criteria and were included in the final synthesis.

RESULTS

We identified and analysed 22 studies published over the past ten years that evaluated artificial intelligence (AI)–based models for diagnosis, prognostication, and quality improvement across a wide range of gastrointestinal (GI) diseases (Tables 1-3) [8-29]. The modalities examined included endoscopic imaging (white light, narrow-band imaging, fluorescence, optical coherence tomography), radiology (computed tomography, magnetic resonance imaging, endoscopic ultrasound, computed tomography enterography), histopathology (whole-slide images), and multimodal clinical-laboratory datasets. Table 4 summarizes the various AI tools and modalities utilized in these studies. Studies spanned infectious, inflammatory, premalignant, neoplastic, and organ-specific applications (Table 5). Below, we synthesize the key findings categorized by disease type.

Infectious and Inflammatory Disorders

Endoscopy is routinely performed in patients for the diagnosis and confirmation of infectious conditions like H. pylori infection and inflammatory conditions of the bowel. However, the diagnostic accuracy can vary significantly depending on the experience level of the endoscopist [30, 31]. Moreover, special equipment is required for the accurate diagnosis of these conditions. The developments in AI

may resolve some of these issues. AI can learn rules from sample data, such as endoscopic images, through machine learning and deep learning and develop tools with higher detection accuracy across all experience levels.

Zou et al. [10] developed a convolutional neural network (CNN) system for the detection of H. pylori during white-light endoscopy, achieving an accuracy of 89.6%, sensitivity of 90.9%, specificity of 88.9%, and area under the receiver operating characteristic curves (AUROCs) of 84.1% for positive cases and 90.3% for negative cases. This significantly enhanced the performance of endoscopists in a RCT. While the study utilized internal multicentre RCT data, it lacked external validation across diverse populations and clinical settings. Similarly, Iacucci et al. [19] introduced a multimodal fusion model that integrates endoscopic and histological features for precise assessment of ulcerative colitis (UC) remission using video-frame deep learning.

Premalignant Lesions And Malignancies

AI can serve as a valuable tool for the precise detection of premalignant lesions, such as Barrett's oesophagus, gastric atrophy, intestinal metaplasia, and colorectal polyps. This capability enables timely intervention, ultimately improving patient prognosis. Meinikheim et al. [8] utilized endoscopic images to train an AI-based decision support system that surpassed both expert and non-expert endoscopists in the detection of Barrett's neoplasia. Similarly, Chen Jing et al. [9] utilized a novel near-infrared peptide heterodimer combined with deep learning segmentation to detect dysplasia with high sensitivity and specificity. Fang et al. [15] employed the GasMIL model to grade gastric atrophy and intestinal metaplasia in whole-slide histopathology, achieving 80% sensitivity, 85% specificity, and an area under the receiver operating characteristic curve (AUROC) of 0.953, thereby enhancing pathologist accuracy in a multicentre validation study. Various AI-assisted models have also been developed and validated to improve the detection of polyps and adenomas, either through direct AI-assisted real-time colonoscopy or by providing image assistance to enhance bowel preparation, which indirectly increases the polyp detection rate [20-22]. However, a study by Zippelius et al. [23] using the GI Genius computer-aided detection (CADe) system in a tandem study reported adenoma miss rates of 4.5% for AI versus 5.5% for routine methods, with adenoma detection rates (ADRs) of approximately 51% when using AI, demonstrating performance comparable to that of experienced endoscopists.

AI is revolutionizing malignancy detection by enhancing precision, scalability, and early intervention. Although challenges such as bias and interpretability remain, the integration of federated learning and quantum computing holds the potential to unlock next-generation oncology tools. Guo et al. [12] developed a real-time computer-aided detection (CAD) system that achieved near-perfect area under the receiver operating characteristic (AUROC) for detecting early oesophageal cancer and precursor lesions using narrow-band imaging (NBI). ENDOANGEL and GRAIDS are advanced convolutional neural network (CNN) tools with significant capabilities in detecting oesophageal and gastric precursors, cancer, and monitoring blind spots [13, 17, 18]. Similarly, CNN models designed for detecting pancreaticobiliary neoplasia using computed tomography (CT) images and endoscopic ultrasound (EUS) have outperformed conventional diagnostic tools [24, 27-29]. The CH-EUS MASTER [29] provided real-time guidance for EUS-guided fine-needle aspiration (EUS-FNA), thereby improving the first-pass diagnostic yield in pancreatic solid organ lesions (SOL).

Yang et al. [11] developed a deep-learning endoscopic ultrasound (EUS) system to differentiate gastrointestinal stromal tumours (GISTs) from leiomyomas. When combined with expert review, the joint diagnostic accuracy increased to 88.8%. However, the study did not investigate how the AI tool integrates into real-time clinical workflows or its usability for endoscopists.

Liver Fibrosis In NAFLD

Liver fibrosis in non-alcoholic fatty liver disease (NAFLD) is a dynamic process driven by metabolic dysfunction and inflammation. Early risk stratification may facilitate the implementation of preventive strategies to halt its progression. Effective methods for detecting fibrosis have been developed using machine learning techniques. Feng G et al. [25] created a machine learning model that significantly outperformed traditional fibrosis scores, such as FIB-4 and APRI, in predicting fibrosis stages of ≥F2. Ma H et al. [26] applied multiple ML techniques to a large health screening cohort and found that Bayesian networks were the most effective for predicting NAFLD.

Overall Insights And Key Concepts

The overall accuracy of the AI diagnostic tools exceeded 90%, with the top performers being those that detect pancreatic cancer, colon capsule completion, and oesophageal cancer, all achieving accuracy rates surpassing 94% to 96%. Most of these tools concentrate on complex conditions such as Barrett's neoplasia and gastric pathologies, while multimodal and histology-integrated AI systems are emerging as valuable clinical tools.

- **Enhanced Diagnostic Performance:** AI systems exhibit high accuracy, sensitivity, and specificity across various conditions, often matching or surpassing the performance of non-expert endoscopists and, in some cases, even that of expert practitioners, particularly for subtle or hard-to-detect lesions.
- **Real-time Assistance:** Numerous systems are engineered for real-time application during endoscopic procedures, delivering immediate feedback (e.g., bounding boxes, risk stratification) with minimal latency, which is essential for clinical utility.
- **Enhanced Detection Rates:** Artificial Intelligence (AI) has been shown to significantly improve detection rates for critical lesions, including high-risk oesophageal lesions (HrELs) and colorectal polyps/adenomas.
- **Quality Improvement:** In addition to lesion detection, artificial intelligence (AI) enhances the quality of endoscopy by minimizing blind spots and improving bowel preparation. These improvements indirectly influence detection rates.
- **Multimodal Integration:** The development of models that integrate various data types (e.g., endoscopic images and histological slides for ulcerative colitis, as well as endoscopic ultrasound images and clinical data for pancreatic lesions) demonstrates superior performance compared to single-modality approaches, as they capture a more comprehensive range of disease characteristics.
- **Guidance for Invasive Procedures:** Artificial Intelligence (AI) is being developed to assist in procedures such as Endoscopic Ultrasound-Fine Needle Aspiration (EUS-FNA), enhancing the diagnostic yield on the first attempt.
- **Objectivity and Standardization:** AI systems provide more consistent criteria for diagnosis and assessment, thereby reducing inter-observer variability and the subjectivity that is inherent in human interpretation.

Several studies represent prospective, multicentre randomized controlled trials, demonstrating progress toward real-world clinical integration, with some systems already implemented in clinical workflows. Despite demonstrating the transformative potential of artificial intelligence (AI) in gastrointestinal diagnostics, these studies exhibit several limitations. Most are single-centre or pilot investigations with relatively small sample sizes, which restrict the generalizability of their findings to broader patient populations. Many utilize case-control or video-review designs instead of fully prospective, real-time clinical workflows, introducing selection bias and limiting the assessment of real-world performance. The heterogeneity in endoscopic platforms, imaging protocols, and AI model training data further hampers reproducibility across different institutions. Few studies incorporate robust external or multicentre validation cohorts or longitudinal follow-up to evaluate long-term clinical impact, cost-effectiveness, and operator variability. Finally, the opaque, “black-box” nature of many deep learning algorithms raises concerns about interpretability and clinician trust, underscoring the need for transparent models and standardized reporting before widespread adoption.

DISCUSSION

The burgeoning field of AI in gastroenterology has demonstrated remarkable potential in improving diagnostic accuracy and streamlining clinical workflows across a wide range of GI conditions. The studies reviewed consistently emphasize the ability of AI-driven tools to enhance human performance, especially for less experienced clinicians, while offering objective, real-time decision support.

Across the spectrum of gastrointestinal AI research, validation approaches have included single-centre prospective trials [8, 12, 14], multicentre randomized controlled studies [10, 22], and retrospective training/test splits [11, 15]. A prominent theme across various applications, such as Barrett's oesophagus, H. pylori infection, upper

GI cancers, and pancreatic lesions, is the significant enhancement in diagnostic performance (accuracy, sensitivity, specificity) observed when AI assists human endoscopists. For instance, AI assistance notably increased the accuracy of non-expert endoscopists in diagnosing H. pylori and high-risk oesophageal lesions, often bringing their performance closer to that of experts [10]. This is particularly valuable in settings where access to highly specialized experts is limited. Beyond assisting human interpretation, several AI systems have demonstrated robust stand-alone diagnostic capabilities [8, 9]. The studies also emphasize the real-time applicability of many of these AI systems, enabling immediate feedback during procedures. This real-time capability is crucial for guiding targeted biopsies and enhancing the completeness of examinations, as evidenced by ENDOANGEL for upper GI cancer detection and CH-EUS MASTER for EUS-FNA guidance [13, 17, 29]. Furthermore, the AI-driven smartphone application for bowel preparation exemplifies a practical, patient-centred application of AI that improves pre-procedure quality and, consequently, polyp detection rates [21].

Despite these encouraging results, several limitations and challenges consistently arise in the reviewed literature.

Heterogeneity and Standardization: A significant challenge in the field is the heterogeneity of endoscopic platforms, imaging protocols, and AI model training data. This variability complicates direct comparisons and limits generalizability. While some studies utilized multicentre data, others were conducted as single-centre or pilot investigations with relatively small sample sizes. The absence of standardized validation, multicentre trials, and transparency in model development is identified as a barrier to wider clinical adoption.

Study Design Biases: Many studies employed retrospective or video-review designs, which can introduce selection bias and may not accurately reflect real-time clinical workflows. Some studies acknowledged potential biases, such as the use of unblinded statisticians, the absence of "washout" periods in tandem designs, and the influence of specific endoscope types and image compression.

External Validation: Few studies included robust external or multicenter validation cohorts, which are essential for determining generalizability and reliability across diverse clinical settings. Studies that conducted external validation, such as the pancreatic cancer CT tool and GRAIDS for upper GI cancers, demonstrated promising results; however, they still revealed limitations, including predominantly homogeneous populations.

Human-AI Interaction: While AI can improve performance, the integration of AI recommendations into human decision-making processes is complex. The factors that influence endoscopists' decisions to either follow or disregard AI advice remain unclear. The "black box" nature of deep learning models can foster scepticism, particularly among experienced clinicians. However, incorporating explainable AI (XAI) features, such as Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP), can help mitigate this scepticism and promote greater acceptance.

Real-World Challenges and "Noise": AI models trained on high-quality retrospective images may encounter difficulties in real-world environments due to reflections, blurring, foam, or inadequate air insufflation. Some studies have reported that AI systems failed to identify lesions in blurry or transitional images, as well as in cases with significant inflammation and bleeding. This highlights the necessity for more diverse training data that captures these real-world variations. Limitations of the Gold Standard: The "gold standard" for diagnosis, typically histology, may not always provide a comprehensive assessment due to sampling errors or the impracticality of obtaining biopsies for all suspected lesions. This can result in an underestimation of true negatives or an increase in false negatives when evaluating AI systems.

Computational Resources: While some AI systems are designed to operate on entry-level computers, others may still necessitate costly backend servers and complex technical support. This requirement could hinder widespread adoption, particularly in resource-limited environments.

Future Perspectives

To truly revolutionize clinical practice, future research on AI-driven diagnostic tools in gastroenterology should concentrate on several key areas [3, 32-40]:

- Large-scale, multicentre, prospective RCTs are urgently needed to conduct more rigorous studies with robust external validation. This will help confirm the efficacy and generalizability of AI systems across diverse patient populations and clinical settings.
- **Standardized Validation and Reporting:** Adhering to guidelines such as DECIDE-AI and PRISMA-DTA is essential for ensuring transparency, reproducibility, and comparability across studies.
- **Focus on Real-Time Integration:** Future development should prioritize the seamless integration of AI into existing clinical workflows, offering real-time assistance that is both swift and user-friendly. This includes addressing issues related to latency and compatibility with various existing endoscopic equipment.
- **Multimodal AI Models:** The successful integration of various data modalities, such as endoscopic images combined with clinical information or histological data, represents a promising avenue for future models. This approach aims to enhance diagnostic accuracy and more comprehensively replicate human decision-making processes.
- **Explainable AI (XAI):** Ongoing research aimed at enhancing the interpretability of AI models is essential for building trust and acceptance among clinicians, thereby moving beyond the "black box" perception. By offering visual cues and clear explanations for AI predictions, we can empower endoscopists to interact more effectively with the technology.
- **Addressing Real-World Variability:** Training datasets should be expanded to encompass a broader range of diverse and challenging real-world scenarios, including variations in image quality, artifacts, and atypical lesion presentations, in order to enhance model robustness.
- **Impact on Patient Outcomes:** Future studies should examine not only the diagnostic accuracy of AI assistance but also its long-term effects on patient outcomes, including reduced miss rates, earlier diagnoses, and improved survival rates.
- **Broader Applicability:** Expanding the applications of AI to encompass a wider range of GI diseases, including rare conditions, and validating models across diverse ethnic groups and geographical regions will be essential for global clinical adoption.

In conclusion, AI-driven diagnostic tools demonstrate significant advancements in the detection of GI diseases, improving accuracy across various modalities. Although most studies indicate favourable performance metrics, wider clinical adoption necessitates standardized validation, multicentre trials, and enhanced model transparency.

Table 1: Published series on role of Artificial Intelligence-Driven Diagnostic Tools in upper GI pathologies.

S No.	Author/Year (N) [Ref.]	Pathology	AI Tool	Modality	Accuracy Sensitivity Specificity	AUROC	Validation
1.	Meinikheim et al./ 2024 (N=96 images) [8]	Barrett's Oesophagus	AI-based decision support system using deep learning model	Endoscopy	81.3% 92.2% 68.9%	NR	Internal, RCT
2.	Chen Jing et al./ 2022 (N=31) [9]	Barrett's Oesophagus	Deep Learning (CNN + near-infrared peptide heterodimer imaging)	Endoscopy + Fluorescence Imaging	NR 94.1% 92.6%	0.95	Internal, Pilot study
3.	Zou et al. / 2024 (N=639) [10]	H. pylori infection	CNN	Endoscopy (White Light Imaging)	89.6% 90.9% 88.9%	H. pylori positivity: 84.1% (95%CI: 73.0%, 95.2%)	Internal, Multicentric RCT

						and negativity: 90.3% (95%CI: 82.2%, 98.4%)	
4.	Yang et al./ 2022 (N= 752) [11]	GIST	Deep Learning (AI-based EUS system)	EUS	78.8% 88.9% 95.10%	NR	Internal & External (Retrospective training &test split)
5.	Guo et al./ 2019 (N=6473 images) [12]	Oesophageal SCC & Precancerous Lesions	Deep Learning CAD System	NBI	NR 98.04% 95.03%	0.989	Internal, Prospective real-time trial
6.	Li et al./ 2024 (N=3117) [13]	Oesophageal Cancer and Precursor lesions	Deep CNN (ENDOANGEL-ELD system)	Endoscopy (White Light + NBI)	98.2% 89.7% 98.5%	NR	Prospective randomized controlled
7.	Tani et al./ 2023 (N=237 lesions) [14]	Oesophageal SCC	Deep Learning (AI-based system using BiT-HyperRule)	Magnifying Endoscopy -NBI	80.6% 68.2% 83.4%	NR	Prospective single-center study
8.	Fang et al./ 2024 (N=545) [15]	Gastric atrophy and intestinal metaplasia	Semi-supervised CNN (GasMIL)	Histology	NR 80% 85%	0.953	Internal & External
9.	Namikawa et al./ 2020 (N=220) [16]	Gastric Cancer and ulcer	Advanced-CNN	Endoscopy	95.9% 99.0% 93.3%	NR	Internal
10.	Wu et al./ 2021 (N=1050) [17]	Gastric Cancer	CNN (ENDOANGEL)	Endoscopy (White Light + NBI)	84.7% 100% 84.3%	NR	Multicentre RCT
11.	Luo et al./ 2019 (N=84,424) [18]	Upper GI Cancer (Oesophageal + Gastric)	Deep CNN (GRAIDS)	Endoscopy (White Light)	95.5% 94.2% 98.5%	0.954	Multicentre, case-control

*N= Number of patients/lesions/images, Ref.= Reference, AI=Artificial Intelligence, AUROC = Area under Receiver operator curve, NR = Not Reported, CNN= Convolutional neural networks, EUS= Endoscopic Ultrasonography, NBI= Narrow-Band Imaging, H. pylori=Helicobacter pylori, GIST= Gastrointestinal stomal tumour, SCC= Squamous cell cancer, GI= Gastrointestinal, RCT= Randomised control trial.

Table 2: Published series on role of Artificial Intelligence-Driven Diagnostic Tools in Lower GI pathologies.

S No.	Author/Year (N) [Ref.]	Pathology	AI Tool	Modality	Accuracy Sensitivity Specificity	AUROC	Validation
1.	Iacucci et al./ 2025 (N=102) [19]	UC (Histologic Remission)	Endo-Histo Fusion Model (BioMedCLIP + CONCH)	Endoscopy + Histology	89.69% 89.72% 89.67%	NR	Internal, clinical trial
2.	Deding et al., 2020 (N=97) [20]	CCE compared with CTC following incomplete OC	AI Localization Algorithm	Capsule Endoscopy	Completion Rate: 75–80% Relative Sensitivity vs. CTC: 2.67 (polyps >5mm), 1.91 (polyps ≥9mm)	NR	Internal, comparative study
3.	Huang Zhong et al./2025 (N=232) [21]	Quality of bowel preparation for colonoscopy using smart phone app	CNN (EfficientNetV2 and ShuffleNet model)	Colonoscopy	87.0% 91% 83.0%	0.86	Internal, RCT
4.	Park et al./2024 (N=805) [22]	Colorectal Polyps	CNN (RetinaNet)	Colonoscopy (Real-time Video)	NR 97.87% NR	NR	External, multicentre RCT
5.	Zippelius et al./ 2022 (N=150) [23]	Colorectal Adenomas	CADe System (GI Genius)	Colonoscopy (Real-time Tandem)	No significant difference in AMR or ADR AI-assisted and conventional colonoscopy.	NR	Internal, Prospective nonrandomized comparative

*N= Number of patients/lesions/images, Ref.= Reference, AI=Artificial Intelligence, AUROC = Area under Receiver operator curve, NR = Not Reported, CNN= Convolutional neural networks, CADe= Computer-Aided Detection, UC= Ulcerative Colitis, CD= Crohn's Disease, OC= Optical colonoscopy, CCE= Colon capsule endoscopies, CTC= Computed tomography colonography, AMR= Adenoma miss rates, ADR= Adenoma detection rates, pCR= pathological complete response, ML= Machine Learning, CTE= CT Enterography, MRI= Magnetic resonance imaging, , RCT= Randomised control trial.

Table 3: Published series on role of Artificial Intelligence-Driven Diagnostic Tools in HPB pathologies.

S No.	Author/Year (N) [Ref.]	Pathology	AI Tool	Modality	Accuracy Sensitivity Specificity	AUROC	Validation
1.	Robles-Medranda et al./ 2023 (N=170 videos) [24]	Biliary Neoplasia	CNN	DSOC	NR 90.5% 68.2%	0.848 (vs. expert) 0.794 (vs. nonexpert)	External, multicentre
2.	Feng G et al./ 2021 (N=553) [25]	Significant Liver Fibrosis in NAFLD	ML (LASSO)	Clinical + Biopsy Data	NR 72.6% 82.5%	0.893	Internal, multi-cohort
3.	Ma H et al./	NAFLD Screening	ML (Bayesian	Clinical +	83% 67.5%	NR	Internal, cross-

	2018 (N=10, 508) [26]		network model)	Lab + Ultrasound	87.8%	NR	-sectional
4.	Chen et al./ 2023 (N=1473) [27]	Pancreatic Cancer	CNN	CT Imaging	91% 89.7% 92.8%	0.95	Internal + External, Nationwide population-based study
5.	Cui et al./ 2024 (N=758) [28]	Pancreatic SOL	Joint Multimodal AI (EUS-CNN + Clinical Fusion)	Clinical information and EUS	98% 88–99% 77–98%	0.996 (internal), 0.955 -0.976 (external)	Internal + External, Randomised cross-over
6.	Tang et al./ 2023 (N=29) [29]	Pancreatic SOL	CNN + Random Forest (CH-EUS MASTER)	CH-EUS	88.9% 90.9% 100%	0.848 (vs expert), 0.794 (vs nonexpert)	Prospective + Independent Video Test

*N= Number of patients/lesions/images, Ref.= Reference, AI=Artificial Intelligence, AUROC = Area under Receiver operator curve, NR = Not Reported, CNN= Convolutional neural networks, CT= Computed tomography, EUS=Endoscopic Ultrasonography, CH-EUS= Contrast-Enhanced Harmonic EUS, ML= Machine Learning, DOCS= Digital Single-Operator Cholangioscopy.

Table 4: Ai Tools/Modalities Used In Various Studies

S No.	AI tool/ technique	Study Reference
1.	Convolutional Neural Network	Meinikheim et al. [8] Chen et al. [9] Zou et al. [10] Yang et al. [11] Guo et al. [12] Li SW et al. [13] Tani et al. [14] Fang et al. [15], Namikawa et al. [16] Wu et al. [17] Luo et al. [18] Robles-Medranda et al. [24] Tang et al. [29]
2.	Multimodal AI / Fusion Models	Iacucci et al. [19] Cui et al. [28]
3.	Machine Learning	Feng et al. [24] Ma et al. [26]
4.	Smartphone App + AI	Huang Zhong et al. [21]
5.	AI-Assisted Colonoscopy	Park et al. [22] Zippelius et al. [23]
6.	Capsule Endoscopy + AI	Deding et al. [20]

*AI= Artificial Intelligence

Table 5: Various studies based on disease focus.

S No.	Disease type	Study focus	Reference
1.	Helicobacter pylori Infection	AI-assisted endoscopy system for *H. pylori* detection	Zou et al. [10]
2.	Ulcerative Colitis	Endo-Histo fusion model for remission and therapy response	Iacucci et al. [19]
3.	Barrett's Esophagus / Neoplasia	AI-enhanced diagnosis and fluorescence detection of neoplasia	Meinikheim et al. [8] Chen et al. [9]
4.	Gastric Atrophy / Intestinal Metaplasia	Semi-supervised deep learning for grading from pathology	Fang et al. [15]
5.	Colorectal Imaging / Capsule Endoscopy	colonoscopy imaging (capsule vs CT colonography)	
6.	Colonoscopy / Polyp Detection	AI-enhanced tools for bowel prep, polyp, and adenoma detection	Huang Zhong et al. [21] Park et al. [22] Zippelius et al. [23]
7.	Gastrointestinal Stromal Tumor	AI system differentiating GIST vs Leiomyoma via EUS	Yang et al. [11],
8.	Esophageal Squamous Cell Carcinoma	Deep learning-based and real-time AI diagnosis	Guo et al. [12] Li SW et al. [13] Tani et al. [14]
9.	Gastric Cancer & Ulcers	AI systems for detection, classification, and image quality improvement	Namikawa et al. [16] Wu et al. [17] Luo et al. [18]
10.	Cholangioscopy / Biliary Neoplasia	CNN model for diagnosing neoplasia on digital cholangioscopy	Robles-Medranda et al. [24]
11.	Pancreatic Cancer / Lesions	CNN and multimodal AI systems for detecting solid lesions and differentiating masses	Chen et al. [27] Cui et al. [28] Tang et al. [29]
12.	Liver Disease / Fibrosis	ML algorithms predicting fibrosis and clinical risk in NAFLD	Feng G et al. [25] Ma et al. [26]

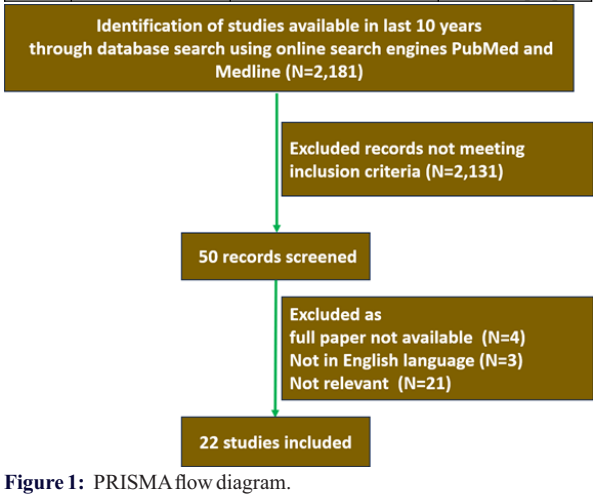

Figure 1: PRISMA flow diagram.

REFERENCES

1 **Kröner PT**, Engels MM, Glicksberg BS, Johnson KW, Mzaik O, van Hooft JE, Wallace MB, El-Serag HB, Krittanawong C. Artificial intelligence in gastroenterology: A state-of-the-art review. World J Gastroenterol 2021; **27**: 6794-6824 [PMID: 34790008 DOI: 10.3748/wjg.v27.i40.6794]

2 **Danpanichkul P**, Suparan K, Tothanarungroj P, Dejvajara D, Rakwong K, Pang Y, Barba R, Thongpiya J, Fallon MB, Harnois D, Lui RN, Wallace MB, Yang JD, Roberts LR, Wijarnpreecha K. Epidemiology of gastrointestinal cancers: a systematic analysis from the Global Burden of Disease Study 2021. Gut 2024; **74**: 26-34 [PMID: 39242191 DOI: 10.1136/gutjnl-2024-333227]

3 **Ramoni D**, Scuricini A, Carbone F, Liberale L, Montecucco F. Artificial intelligence in gastroenterology: Ethical and diagnostic challenges in clinical practice. World J Gastroenterol 2025; **31**: 102725 [PMID: 40093670 DOI: 10.3748/wjg.v31.i10.102725]

4 **Klang E**, Sourosh A, Nadkarni GN, Sharif K, Lahat A. Deep Learning and Gastric Cancer: Systematic Review of AI-Assisted Endoscopy. Diagnostics (Basel) 2023; **13**: 3613 [PMID: 38132197 DOI: 10.3390/diagnostics13243613]

5 **Krenzer A**, Heil S, Fitting D, Matti S, Zoller WG, Hann A, Puppe F. Automated classification of polyps using deep learning architectures and few-shot learning. BMC Med Imaging 2023; **23**: 59 [PMID: 37081495 DOI: 10.1186/s12880-023-01007-4]

6 **Turtoi DC**, Brata VD, Incze V, Ismaiel A, Dumitrascu DI, Militaru V, Munteanu MA, Botan A, Toc DA, Duse TA, Popa SL. Artificial Intelligence for the Automatic Diagnosis of Gastritis: A Systematic Review. J Clin Med 2024; **13**: 4818 [PMID: 39200959 DOI: 10.3390/jcm13164818]

7 **Siontis GCM**, Sweda R, Noseworthy PA, Friedman PA, Siontis KC, Patel CJ. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. BMJ Health Care Inform 2021; **28**: e100466 [PMID: 34969668 DOI: 10.1136/bmjhci-2021-100466]

8 **Meinikheim M**, Mendel R, Palm C, Probst A, Muzalyova A, Scheppach MW, Nagl S, Schnoy E, Römmele C, Schulz DAH, Schlottmann J, Prinz F, Rauber D, Rückert T, Matsumura T, Fernández-Esparrach G, Parsa N, Byrne MF, Messmann H, Ebigbo A. Influence of artificial intelligence on the diagnostic performance of endoscopists in the assessment of Barrett's esophagus: a tandem randomized and video trial. Endoscopy

2024; **56**: 641-649 [PMID: 38547927 DOI: 10.1055/a-2296-5696]
9 **Chen J**, Jiang Y, Chang TS, Rubenstein JH, Kwon RS, Wamsteker EJ, Prabhu A, Zhao L, Appelman HD, Owens SR, Beer DG, Turgeon DK, Seibel EJ, Wang TD. Detection of Barrett's neoplasia with a near-infrared fluorescent heterodimeric peptide. Endoscopy 2022; **54**: 1198-1204 [PMID: 35299273 DOI: 10.1055/a-1801-2406]
10 **Zou PY**, Zhu JR, Zhao Z, Mei H, Zhao JT, Sun WJ, Wang GH, Chen DF, Fan LL, Lan CH. Development and application of an artificial intelligence-assisted endoscopy system for diagnosis of Helicobacter pylori infection: a multicenter randomized controlled study. BMC Gastroenterol 2024; **24**: 335 [PMID: 39350033 DOI: 10.1186/s12876-024-03389-3]
11 **Yang X**, Wang H, Dong Q, Xu Y, Liu H, Ma X, Yan J, Li Q, Yang C, Li X. An artificial intelligence system for distinguishing between gastrointestinal stromal tumors and leiomyomas using endoscopic ultrasonography. Endoscopy 2022; **54**: 251-261 [PMID: 33827140 DOI: 10.1055/a-1476-8931]
12 **Guo L**, Xiao X, Wu C, Zeng X, Zhang Y, Du J, Bai S, Xie J, Zhang Z, Li Y, Wang X, Cheung O, Sharma M, Liu J, Hu B. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). Gastrointest Endosc 2020; **91**: 41-51 [PMID: 31445040 DOI: 10.1016/j.gie.2019.08.018]
13 **Li SW**, Zhang LH, Cai Y, Zhou XB, Fu XY, Song YQ, Xu SW, Tang SP, Luo RQ, Huang Q, Yan LL, He SQ, Zhang Y, Wang J, Ge SQ, Gu BB, Peng JB, Wang Y, Fang LN, Wu WD, Ye WG, Zhu M, Luo DH, Jin XX, Yang HD, Zhou JJ, Wang ZZ, Wu JF, Qin QQ, Lu YD, Wang F, Chen YH, Chen X, Xu SJ, Tung TH, Luo CW, Ye LP, Yu HG, Mao XL. Deep learning assists detection of esophageal cancer and precursor lesions in a prospective, randomized controlled study. Sci Transl Med 2024; **16**: eadk5395 [PMID: 38630847 DOI: 10.1126/scitranslmed.adk5395]
14 **Tani Y**, Ishihara R, Inoue T, Okubo Y, Kawakami Y, Matsueda K, Miyake M, Yoshii S, Shichijo S, Kanesaka T, Yamamoto S, Takeuchi Y, Higashino K, Uedo N, Michida T, Kato Y, Tada T. A single-center prospective study evaluating the usefulness of artificial intelligence for the diagnosis of esophageal squamous cell carcinoma in a real-time setting. BMC Gastroenterol 2023; **23**: 184 [PMID: 37231330 DOI: 10.1186/s12876-023-02788-2]
15 **Fang S**, Liu Z, Qiu Q, Tang Z, Yang Y, Kuang Z, Du X, Xiao S, Liu Y, Luo Y, Gu L, Tian L, Liang X, Fan G, Zhang Y, Zhang P, Zhou W, Liu X, Tian J, Wei W. Diagnosing and grading gastric atrophy and intestinal metaplasia using semi-supervised deep learning on pathological images: development and validation study. Gastric Cancer 2024; **27**: 343-354 [PMID: 38095766 DOI: 10.1007/s10120-023-01451-9]
16 **Namikawa K**, Hirasawa T, Nakano K, Ikenoyama Y, Ishioka M, Shiroma S, Tokai Y, Yoshimizu S, Horiuchi Y, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T. Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: comparison between the original and newly developed systems. Endoscopy 2020; **52**: 1077-1083 [PMID: 32503056 DOI: 10.1055/a-1194-8771]
17 **Wu L**, He X, Liu M, Xie H, An P, Zhang J, Zhang H, Ai Y, Tong Q, Guo M, Huang M, Ge C, Yang Z, Yuan J, Liu J, Zhou W, Jiang X, Huang X, Mu G, Wan X, Li Y, Wang H, Wang Y, Zhang H, Chen D, Gong D, Wang J, Huang L, Li J, Yao L, Zhu Y, Yu H. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial. Endoscopy 2021; **53**: 1199-1207 [PMID: 33429441 DOI: 10.1055/a-1350-5583]
18 **Luo H**, Xu G, Li C, He L, Luo L, Wang Z, Jing B, Deng Y, Jin Y, Li Y, Li B, Tan W, He C, Seeruttun SR, Wu Q, Huang J, Huang DW, Chen B, Lin SB, Chen QM, Yuan CM, Chen HX, Pu HY, Zhou F, He Y, Xu RH. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol 2019; **20**: 1645-1654 [PMID: 31591062 DOI: 10.1016/S1470-2045(19)30637-0]
19 **Iacucci M**, Santacroce G, Meseguer P, Diéguez A, Del Amor R, Kolawole BB, Chaudhari U, Zammarchi I, Hayes B, Crotty R, Zardo D, Maeda Y, Puga-Tejada M, Ditonno I, Vadori V, Burke L, D'Amico F, Ghosh S, Grisan E, Naranjo V. Endo-Histo foundational fusion model: a novel artificial intelligence for assessing histologic remission and response to therapy in ulcerative colitis clinical trial. J Crohns Colitis 2025; **19**: jjaf108 [PMID: 40581339 DOI: 10.1093/ecco-jcc/jjaf108]
20 **Deding U**, Herp J, Havshoei AL, Kobaek-Larsen M, Buijs MM, Nadimi ES, Baatrup G. Colon capsule endoscopy versus CT colonography after incomplete colonoscopy. Application of artificial intelligence algorithms to identify complete colonic investigations. United European Gastroenterol J 2020; **8**: 782-789 [PMID: 32731841 DOI: 10.1177/2050640620937593]
21 **Zhong H**, Hou C, Huang Z, Chen X, Zou Y, Zhang H, Wang T, Wang L, Huang X, Xiang Y, Zhong M, Hu M, Xiong D, Wang L, Zhang Y, Luo Y, Guan Y, Xia M, Liu X, Yang J, Gan T, Wei W, Chen H, Gong H. A clinical pilot trial of an artificial intelligence-driven smart phone application of bowel preparation for colonoscopy: a randomized clinical trial. Scand J Gastroenterol 2025; **60**: 116-121 [PMID: 39709551 DOI: 10.1080/00365521.2024.2443520]
22 **Park DK**, Kim EJ, Im JP, Lim H, Lim YJ, Byeon JS, Kim KO, Chung JW, Kim YJ. A prospective multicenter randomized controlled trial on artificial intelligence assisted colonoscopy for enhanced polyp detection. Sci Rep 2024; **14**: 25453 [PMID: 39455850 DOI: 10.1038/s41598-024-77079-1]
23 **Zippelius C**, Alqahtani SA, Schedel J, Brookman-Amissah D, Muehlenberg K, Federle C, Salzberger A, Schorr W, Pech O. Diagnostic accuracy of a novel artificial intelligence system for adenoma detection in daily practice: a prospective nonrandomized comparative study. Endoscopy 2022; **54**: 465-472 [PMID: 34293812 DOI: 10.1055/a-1556-5984]
24 **Robles-Medranda C**, Baquerizo-Burgos J, Alcivar-Vasquez J, Kahaleh M, Raijman I, Kunda R, Puga-Tejada M, Egas-Izquierdo M, Arevalo-Mora M, Mendez JC, Tyberg A, Sarkar A, Shahid H, Del Valle-Zavala R, Rodriguez J, Merfea RC, Barreto-Perez J, Saldaña-Pazmiño G, Calle-Loffredo D, Alvarado H, Lukashok HP. Artificial intelligence for diagnosing neoplasia on digital cholangioscopy: development and multicenter validation of a convolutional neural network model. Endoscopy 2023; **55**: 719-727 [PMID: 36781156 DOI: 10.1055/a-2034-3803]
25 **Feng G**, Zheng KI, Li YY, Rios RS, Zhu PW, Pan XY, Li G, Ma HL, Tang LJ, Byrne CD, Targher G, He N, Mi M, Chen YP, Zheng MH. Machine learning algorithm outperforms fibrosis markers in predicting significant fibrosis in biopsy-confirmed NAFLD. J Hepatobiliary Pancreat Sci 2021; **28**: 593-603 [PMID: 33908180 DOI: 10.1002/jhbp.972]
26 **Ma H**, Xu CF, Shen Z, Yu CH, Li YM. Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. Biomed Res Int 2018; **2018**: 4304376 [PMID: 30402478 DOI: 10.1155/2018/4304376]
27 **Chen PT**, Wu T, Wang P, Chang D, Liu KL, Wu MS, Roth HR, Lee PC, Liao WC, Wang W. Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-based Study. Radiology 2023; **306**: 172-182 [PMID: 36098642 DOI: 10.1148/radiol.220152]
28 **Cui H**, Zhao Y, Xiong S, Feng Y, Li P, Lv Y, Chen Q, Wang R, Xie P, Luo Z, Cheng S, Wang W, Li X, Xiong D, Cao X, Bai S, Yang A, Cheng B. Diagnosing Solid Lesions in the Pancreas With Multimodal Artificial Intelligence: A Randomized Crossover Trial. JAMA Netw Open 2024; **7**: e2422454 [PMID: 39028670 DOI: 10.1001/jamanetworkopen.2024.22454]
29 **Tang A**, Tian L, Gao K, Liu R, Hu S, Liu J, Xu J, Fu T, Zhang Z, Wang W, Zeng L, Qu W, Dai Y, Hou R, Tang S, Wang X. Contrast-enhanced harmonic endoscopic ultrasound (CH-EUS) MASTER: A novel deep learning-based system in pancreatic mass diagnosis. Cancer Med 2023; **12**: 7962-7973 [PMID: 36606571 DOI: 10.1002/cam4.5578]
30 **Nguyen VX**, Le Nguyen VT, Nguyen CC. Appropriate use of endoscopy in the diagnosis and treatment of gastrointestinal diseases: up-to-date indications for primary care providers. Int J Gen Med 2010; **3**: 345-357 [PMID: 21116340 DOI: 10.2147/IJGM.S14555]
31 **Watanabe K**, Nagata N, Shimbo T, Nakashima R, Furuhata E, Sakurai T, Akazawa N, Yokoi C, Kobayakawa M, Akiyama J, Mizokami M, Uemura N. Accuracy of endoscopic diagnosis of Helicobacter pylori infection according to level of endoscopic experience and the effect of training. BMC Gastroenterol 2013; **13**: 128 [PMID: 23947684 DOI: 10.1186/1471-230X-13-128]
32 Simonetto, Douglas. "Artificial Intelligence and the Future of Gastroenterology and Hepatology." Hepatol Pancreat Sci 7 (2023): 211.
33 **Li G**, Mabe K, Haseyama M. Clarifying diagnostic scope and future perspectives in AI-assisted diagnosis of Helicobacter pylori infection. J Gastroenterol 2025 [PMID: 40517341 DOI: 10.1007/s00535-025-02277-1]
34 **Alagappan M**, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. World J Gastrointest Endosc 2018; **10**: 239-249 [PMID: 30364792 DOI: 10.4253/wjge.v10.i10.239]
35 **Lee K**, Kim ES. Generative Artificial Intelligence in the Early Diagnosis of Gastrointestinal Disease. Applied Sciences 2024; **14**: 11219 [DOI: 10.3390/app142311219]
36 **Zhou J**, Hu N, Huang ZY, Song B, Wu CC, Zeng FX, Wu M. Application of artificial intelligence in gastrointestinal disease: a narrative review. Ann Transl Med 2021; **9**: 1188 [PMID: 34430629 DOI: 10.21037/atm-21-3001]
37 **Ali H**, Muzammil MA, Dahiya DS, Ali F, Yasin S, Hanif W, Gangwani MK, Aziz M, Khalaf M, Basuli D, Al-Haddad M. Artificial intelligence in gastrointestinal endoscopy: a comprehensive review. Ann Gastroenterol 2024; **37**: 133-141 [PMID: 38481787 DOI: 10.20524/aog.2024.0861]
38 Tanabe S, Perkins EJ, Ono R, Sasaki H. Artificial intelligence in gastrointestinal diseases. Artif Intell Gastroenterol 2021; 2(3): 69-7
39 **Tanabe S**, Perkins EJ, Ono R, Sasaki H. Artificial intelligence in gastrointestinal diseases. AIG 2021; **2**: 69-76 [DOI: 10.35712/aig.v2.i3.69]
40 **Patel V**, Khan MN, Shrivastava A, Sadiq K, Ali SA, Moore SR, Brown DE, Syed S. Artificial Intelligence Applied to Gastrointestinal Diagnostics: A Review. J Pediatr Gastroenterol Nutr 2020; **70**: 4-11 [PMID: 31567886 DOI: 10.1097/MPG.0000000000002507]