Supplementary Table 1. Summary of all research articles included in the narrative review.

Reference	Article Title	Focus Area	Sample Size (N) / Dataset Size	Study Design / Type	ML or DL	Key Algorithms (Inferred)	External Validation (EV)	Potential Biases (Inferred)	Key Metric (Inferred)
29.	Esophageal Intelligence: Implementing AI-assisted analysis for high-resolution manometry in patients with achalasia	Neurogastroenterology	Small	AI Model Development (Manometry)	ML/DL	SVM/LSTM	Internal	Selection Bias	Accuracy, AUC
30.	Gemini-Assisted Deep Learning for the Diagnosis of Esophageal Neurogastroenterology Disorders	Neurogastroenterology	Moderate (Images)	DL Model Development & Validation	DL	CNN	Internal	Retrospective Bias	AUC, Sensitivity
31.	An artificial intelligence platform for high-resolution manometry diagnosis in esophageal disorders: a multi-center study	Neurogastroenterology	Moderate	ML/AI Model Development (Manometry)	ML	Random Forest	Yes (External)	Sample Bias	Accuracy, Kappa

32.	Comparison of measurement of esophageal Neurogastroenterology by artificial intelligence versus expert humans	Neurogastroenterology	Moderate	Comparative Study (Manometry)	ML	Measurement Algorithms	Yes (Inter-Method)	Operator Bias	Concordance (r, Kappa)
33.	Deep learning–based artificial intelligence system for automated diagnosis of esophageal Neurogastroenterology disorders	Neurogastroenterology	Moderate	DL Model Development (Manometry)	DL	Deep Neural Network	Internal	Generalization Bias	AUC, Sensitivity
34.	Enhancing the diagnostic yield of Chicago Classification v4.0 using machine and deep learning	Neurogastroenterology	Moderate	ML/DL Model Development	ML/DL	CNN/Random Forest	Yes (Robust Validation)	Labeling Bias	Accuracy, AUC
35.	Artificial intelligence facilitates high-resolution manometry diagnosis in esophageal disorders	Neurogastroenterology	Moderate	ML Model Development (Manometry)	ML	Supervised Classifier	Internal	Single-Center Bias	AUC, Sensitivity
36.	Artificial intelligence automates Chicago Classification v4.0 analysis of high-resolution manometry	Neurogastroenterology	Moderate	ML Model Development (Manometry)	ML	Supervised Classifier	Internal	Operator Bias	Accuracy, Time
37.	Automated software-derived supine vs upright high-resolution esophageal manometry parameters	Neurogastroenterology	Small	Software Development	ML	Linear Models/ ML	Internal	Sampling Bias	Concordance (r)

				(Manometry)					
39.	Neurogastroenterology dyspepsia and its subgroups: A machine learning approach	Neurogastroenterology	Large	Observational/Clinical Study	ML	Cluster Analysis	No	Classification Bias	Rome Criteria
40.	Artificial intelligence model on images from non-erosive reflux disease (NERD) for personalized treatment	Neurogastroenterology	Moderate	DL Model Development (Imaging)	DL	CNN	Internal	Image Bias	AUC, Accuracy
41.	Artificial intelligence-based diagnostic model for Neurogastroenterology dyspepsia using high-resolution manometry	Neurogastroenterology	Small	Diagnostic Study (ML)	ML	Supervised Classifier	No	Small Sample Bias	Accuracy, Sensitivity
42.	Discovering the key symptoms associated with Neurogastroenterology gastrointestinal disorders using machine learning and electronic health records	Neurogastroenterology	Large	Data Mining/ML Study	ML	Association Rules	No	Data Bias (EHR)	F1-score
43.	Pattern Identification in Patients with Neurogastroenterology Gastrointestinal Disorders using Machine Learning	Neurogastroenterology	Moderate	Data Mining/ML Study	ML	Predictive Models	No	Retrospective Bias	Classification
45.	Artificial intelligence model for analyzing symptoms of Neurogastroenterology gastrointestinal disorders: a deep learning approach	Neurogastroenterology	Moderate	DL Model Development (Symptoms)	DL	Recurrent Network (RNN)	Internal	Language/Context Bias	AUC
46.	Computerized auscultation applied to Neurogastroenterology bowel diseases: A pilot study	Neurogastroenterology	Small	Pilot Study (Bowel Sounds)	ML	Signal Processing	No	Instrumental Bias	Classification

47.	Noninvasive diagnosis of irritable bowel syndrome using exhaled volatile organic compounds and machine learning	Neurogastroenterology	Moderate	Diagnostic Study (AI)	ML	Supervised Classifiers	No	Cohort Bias	Accuracy, Sensitivity
49.	Gut microbiota associations with irritable bowel syndrome: a machine learning approach	Neurogastroenterology	Moderate	Microbiota Study (ML)	ML	Classification Models	No	Sequencing Bias	Rank, p-value
50.	Artificial intelligence-based personalized dietary intervention for irritable bowel syndrome: a randomized controlled trial	Neurogastroenterology	Moderate	Intervention Study (ML)	ML	Predictive Model	No	Adherence Bias	Symptom Change
51.	Usefulness of machine learning-based prediction models for irritable bowel syndrome	Neurogastroenterology	Moderate	Diagnostic Study (ML)	ML	Supervised Classifier	No	Diagnostic Bias	AUC
52.	A smartphone application using an artificial intelligence-based chatbot for personalized dietary guidance in irritable bowel syndrome	Neurogastroenterology	Moderate	Intervention Study (App/AI)	ML	Predictive Model	No	Self-Report Bias	Symptom Change
54.	Artificial Intelligence Model for Time Series Analysis of High-Resolution Esophageal Manometry	Neurogastroenterology	Moderate	DL Model Development (Manometry)	DL	RNN/LSTM	Internal	Signal Noise Bias	AUC, Accuracy
61.	Machine learning models are superior to conventional scoring systems for predicting significant liver fibrosis in patients with chronic hepatitis B	Hepatology	Large	ML Model Development (Fibrosis)	ML	Supervised Classifiers	Yes (External)	Gold Standard Bias	AUC, Accuracy
62.	Multicenter validation of FIB-6, a novel machine learning-based index for liver	Hepatology	Large	Validation Study (Fibrosis)	ML	Classification	Yes (Multicenter)	Retro-Validation Bias	AUC, C-index

	fibrosis assessment in patients with non-alcoholic fatty liver disease					Algorithms			
63.	Clinical prediction of HBV-associated hepatocellular carcinoma with machine learning algorithms	Hepatology	Large	ML Model Development (HCC)	ML	Supervised Classifiers	Internal	Single-Center Bias	AUC, Sensitivity
64.	Development of the AI-cirrhosis-ECG score using machine learning for predicting cirrhosis in patients with chronic hepatitis B	Hepatology	Large	ML Model Development (Cirrhosis)	ML	Supervised Classifier	Yes (Internal/External)	ECG Data Bias	AUC
65.	Ultrasound-based artificial intelligence for non-invasive diagnosis of cirrhosis	Hepatology	Moderate	DL Model Development (Cirrhosis/US)	DL	CNN (Image Analysis)	Internal	Image Bias	Accuracy, Sensitivity
66.	Using artificial intelligence to predict hepatocellular carcinoma risk after direct-acting antiviral therapy for hepatitis C virus infection	Hepatology	Large	ML Model Development (HCC)	ML	Predictive Models	Yes (External)	Retrospective Bias	AUC, C-index
67.	Detection of liver cirrhosis in standard computed tomography images using deep learning	Hepatology	Moderate	DL Model Development (CT Imaging)	DL	CNN	Internal	Image Bias	AUC, Accuracy
68.	Deep learning supports the differentiation of hepatocellular carcinoma from hepatic metastases on gadoxetate disodium-enhanced MRI	Hepatology	Moderate	DL Model Development (MRI/Masses)	DL	CNN	Internal	Image Bias	Accuracy, AUC

69.	Value of multimodal MRI radiomics integrated with clinical factors for predicting microvascular invasion in hepatocellular carcinoma	Hepatology	Moderate	Radiomics Study (AI)	ML	Supervised Classifiers	Internal	Radiomics Bias	AUC, C-index
70.	A machine learning approach to liver transplant candidate selection	Hepatology	Large	ML Model Development (Transplant)	ML	Random Forest	Yes (External)	Cohort Bias	C-index
71.	An imaging-based artificial intelligence model for predicting overall survival in patients with hepatocellular carcinoma	Hepatology	Moderate	DL Model Development (HCC/Imaging)	DL	CNN	Internal	Image Bias	AUC, Sensitivity
72.	Assessment of portal hypertension severity using radiomics features extracted from computed tomography images in patients with liver cirrhosis	Hepatology	Moderate	Radiomics Study (Portal Hypertension)	ML	Supervised Classifier	No	Reference Bias	Correlation
73.	Development and validation of machine learning models for predicting post-operative prognosis of hepatocellular carcinoma	Hepatology	Large	ML Model Development (Prognosis)	ML	Random Forest/GBM	Yes (External)	Retrospective Bias	C-index
75.	An interpretable artificial intelligence system for predicting prognosis in hepatocellular carcinoma patients after surgery	Hepatology	Moderate	ML Model Development (HCC/Prognosis)	ML	Explainable Models (XAI)	Internal	Interpretation Bias	C-index

76.	Deep learning to predict esophageal variceal bleeding in cirrhosis using endoscopic images	Hepatology	Moderate	DL Model Development (Endoscopic Image)	DL	CNN	Internal	Endoscopic Quality Bias	AUC
77.	Artificial intelligence-based detection of esophageal varices in upper endoscopy	Hepatology	Moderate	DL Model Development (Endoscopy)	DL	CNN	Internal	Detection Bias	Sensitivity
78.	Explainable machine learning model for predicting liver decompensation and mortality in patients with cirrhosis	Hepatology	Large	ML Model Development (Cirrhosis)	ML	XAI/Classifier	Yes (External)	Complexity Bias	AUC
81.	Deep learning radiomics model accurately predicts microvascular invasion in hepatocellular carcinoma	Hepatology	Moderate	Radiomics Study (DL)	DL	CNN/Radiomics	Internal	Image/ Segmentation Bias	AUC, C-index
83.	Machine learning models using non-invasive scores for diagnosis of steatohepatitis in patients with non-alcoholic fatty liver disease	Hepatology	Large	ML Model Development (Steatohepatitis)	ML	Supervised Classifier	Yes (External)	Reference Bias	AUC, Accuracy
84.	Assessment of a deep learning model for liver segmentation in CT images: a large multi-ethnic cohort study	Hepatology	Large	DL Model Development & Validation	DL	CNN (Image Analysis)	Yes (Multi-ethnic cohort)	Retrospective Bias	AUC, Sensitivity, Specificity

85.	Machine learning for individualized prediction of survival after liver resection for hepatocellular carcinoma	Hepatology	Large	ML Model Development (HCC)	ML	Random Forest/GBM	Yes (External)	Follow-up Bias	C-index
86.	An artificial intelligence model to predict hepatocellular carcinoma recurrence after liver transplantation	Hepatology	Very Large	ML Model Development (HCC)	ML	Random Forest, GBM	Yes (External Centers)	Single-Center Bias	AUC, C-index, Calibration
87.	Developing deep learning-based strategies for automated scoring of liver fibrosis in whole slide images	Hepatology	Moderate	DL Model Development (Microscopy)	DL	CNN (Digital Pathology)	Internal	Pathology Bias	Accuracy
90.	A novel multimodal deep learning model for predicting overall survival in patients with hepatocellular carcinoma	Hepatology	Moderate	DL Model Development (Multimodal)	DL	CNN + DNN	Internal	Data Fusion Bias	AUC, C-index
92.	Development of a computed tomography-based deep learning model for predicting microvascular invasion in hepatocellular carcinoma	Hepatology	Moderate	Radiomics Study (ML/CT)	ML	Supervised Classifiers	Internal	Image Bias	AUC, Sensitivity
93.	CT-radiomics and clinical risk scores for prediction of microvascular invasion in hepatocellular carcinoma: a machine learning approach	Hepatology	Moderate	Radiomics Study (ML/CT)	ML	Supervised Classifiers	Internal	Image Bias	AUC

94.	Deep learning–based multi-omics model for prognosis prediction in hepatocellular carcinoma	Hepatology	Moderate	Multi-Omics Study (DL)	DL	Deep Networks	Internal	Omics Integration Bias	C-index
95.	Exploring prognostic indicators in the pathological images of hepatocellular carcinoma using machine learning	Hepatology	Moderate	Digital Pathology Study (ML)	ML	Image Algorithms	Internal	Pathology Bias	C-index
96.	Classification and mutation prediction of hepatocellular carcinoma subtypes using deep learning analysis of whole-slide images	Hepatology	Moderate	Digital Pathology Study (ML)	ML	Image Classifiers	Internal	Sampling Bias	Accuracy
97.	Deep learning–enabled diagnosis of hepatic steatosis on computed tomography: a multicenter study	Hepatology	Large	Image Diagnosis (CT) using DL.	DL	Advanced CNN	Yes (Different Dataset)	Class Bias	Accuracy, Multi-class AUC
98.	Deep learning-based classification of focal liver lesions on ultrasound images	Hepatology	Moderate	DL Model Development (Imaging)	DL	CNN	Internal	Image Bias	AUC
99.	LiverNet: efficient and robust deep learning for automated liver segmentation in CT images	Hepatology	Moderate	DL Model Development (Imaging)	DL	CNN (Segmentation Analysis)	Internal	Segmentation Bias	Dice Score
100.	Development and validation of an optimized machine learning model for predicting post-transplant survival in hepatocellular carcinoma	Hepatology	Large	ML Model Development (Transplant)	ML	Supervised Classifier	Yes (External)	Single-Center Bias	C-index

103.	Using artificial intelligence for predicting recurrence of hepatocellular carcinoma after liver transplantation	Hepatology	Moderate	ML Model Development (Transplant)	ML	Supervised Classifiers	Internal	Clinical Data Bias	AUC, C-index
104.	Validation of the Toronto recurrence inference score for hepatocellular carcinoma (TORISE) in a multicenter cohort	Hepatology	Large	Validation Study (Transplant/ML)	ML	Specific Model	Yes (Multicenter)	Retro-Validation Bias	AUC, C-index
105.	A prospective multicenter validation of RETREAT score for predicting hepatocellular carcinoma recurrence after liver transplantation	Hepatology	Large	Validation Study (Transplant/ML)	ML	Specific Model	Yes (Prospective)	Implementation Bias	C-index
106.	An explainable supervised machine learning algorithm for predicting hepatocellular carcinoma recurrence after curative treatment	Hepatology	Moderate	ML Model Development (HCC)	ML	XAI/Classifier	Internal	Clinical Data Bias	AUC
107.	Predicting major adverse cardiovascular events in patients with cirrhosis using a machine learning model	Hepatology	Moderate	ML Model Development (Cirrhosis)	ML	Supervised Classifiers	Internal	Rare Event Bias	AUC
108.	DeePSC: a deep learning model for automated diagnosis and prognostication of primary sclerosing cholangitis on endoscopic retrograde cholangiopancreatography	Hepatology	Moderate	DL Model Development (Cholangiopathy)	DL	CNN (Image Analysis)	Internal	Image Bias	AUC, F1-score

109.	Primary sclerosing cholangitis risk estimate tool (P-SCORER) based on machine learning	Hepatology	Large	Risk Tool Development (ML)	ML	Supervised Classifier	Yes (External)	Cohort Bias	C-index
110.	Deep learning helps discriminate primary sclerosing cholangitis from secondary sclerosing cholangitis	Hepatology	Moderate	DL Model Development (Cholangitis)	DL	CNN	Internal	Image Bias	AUC
111.	Top-Down proteomics detection of potential biomarkers for primary biliary cholangitis using machine learning	Hepatology	Small	Proteomics Study (ML)	ML	Supervised Classifier	No	Small Sample Bias	Accuracy
112.	Development of decision forest models for the prediction of liver fibrosis stage in non-alcoholic fatty liver disease	Hepatology	Moderate	ML Model Development (Fibrosis)	ML	Random Forest	Internal	Clinical Data Bias	AUC, Sensitivity
114.	A Machine Learning-Based Method for Detecting Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD/MAFLD)	Hepatology	Moderate	ML Model Development (MASLD/MAFLD)	ML	Supervised Classifier	Internal	Gold Standard Bias	AUC
115.	Predicting metabolic dysfunction associated steatohepatitis (MASH/NASH) among patients with MASLD/NAFLD using machine learning	Hepatology	Large	ML Model Development (MASLD/MAFLD)	ML	Supervised Classifier	Yes (External)	Single-Center Bias	AUC, C-index
116.	Complexity of ballooned hepatocyte feature extraction on histopathological images of non-alcoholic steatohepatitis using machine learning	Hepatology	Moderate	Digital Pathology Study (ML)	ML	Image Algorithms	Internal	Pathology Bias	Correlation

118.	Noninvasive proteomic biomarkers for alcohol-related liver disease severity and prognosis using machine learning	Hepatology	Moderate	Proteomics Study (ML)	ML	Supervised Classifier	Internal	Sampling Bias	AUC
120.	Impact of pain, fatigue and bowel incontinence on inflammatory bowel disease patients: a machine learning approach	IBD	Large	Symptoms/ML Study	ML	Supervised Classifiers	No	Self-Report Bias	Classification
124.	Application of artificial intelligence modeling to improve personalized prediction of response to anti-TNF therapy in inflammatory bowel disease	IBD	Moderate	ML/DL Model Development	ML/DL	Supervised Classifier	Internal	Clinical Data Bias	AUC
125.	Artificial intelligence-aided colonoscopic image analysis for the assessment of endoscopic disease activity in ulcerative colitis	IBD/Endoscopy	Moderate	UC Lesion Detection (DL)	DL	CNN	Internal	Detection Bias	Sensitivity, Specificity
126.	Quantitative assessment of mucosal architecture by deep learning: a new paradigm for inflammatory bowel disease monitoring	IBD/Endoscopy	Small	DL Model Development (Imaging)	DL	CNN	No	Small Sample Bias	Concordance
127.	Building a trustworthy AI differential diagnosis model for inflammatory bowel disease: a multi-center study	IBD	Large	ML Model Development (Diagnosis)	ML	Explainable Models (XAI)	Yes (External)	EHR Data Bias	AUC
130.	Integrating radiomics with clinicoradiological features using machine learning to predict therapeutic response in Crohn's disease	IBD	Moderate	Radiomics Study (ML)	ML	Supervised Classifiers	Internal	Radiomics Bias	AUC, C-index

131.	Deep-Learning, Radiomics and Clinic Based Prediction Model for Therapeutic Response in Crohn's Disease	IBD	Moderate	Radiomics Study (DL)	DL	CNN/Radiomics	Internal	Image/Segmentation Bias	AUC, Accuracy
133.	Deployment of an artificial intelligence system for automated histologic grading in inflammatory bowel disease	IBD/Pathology	Moderate	Digital Pathology Study (ML)	ML	Image Classifiers	Yes (Prospective)	Pathology Bias	Time, Accuracy
135.	Artificial intelligence enabled histological scoring for ulcerative colitis: a multi-center development and validation study	IBD/Pathology	Moderate	DL Model Development (Pathology)	DL	CNN (Histology)	Yes (External)	Staining Bias	AUC, Concordance
136.	Deep learning analysis of histologic images for assessment of endoscopic and clinical disease activity in ulcerative colitis	IBD/Pathology	Moderate	DL Model Development (Pathology)	DL	CNN (Histology)	Internal	Pathology Bias	AUC
137.	Artificial intelligence program to predict response to vedolizumab and ustekinumab in inflammatory bowel disease	IBD	Moderate	ML Model Development (Prediction)	ML	Supervised Classifier	Internal	Retrospective Bias	AUC
138.	A series of genes for predicting responses to anti-TNF therapy in inflammatory bowel disease patients using machine learning	IBD	Small	Genomics Study (ML)	ML	Supervised Classifier	No	Small Sample Bias	Accuracy
139.	Development and validation of machine learning models for predicting clinical response to infliximab in children with Crohn's disease	IBD	Large	ML Model Development (Response Prediction)	ML	Random Forest/GBM	Yes (External)	EHR Data Bias	AUC, Sensitivity

140.	Precision medicine: Externally validated machine learning models for predicting response to anti-TNF therapy in inflammatory bowel disease	IBD	Large	ML Validation Study (Precision Medicine)	ML	Specific Model	Yes (External)	Cohort Bias	C-index
142.	Identification of gene expression profiles for predicting response to anti-TNF therapy in ulcerative colitis using machine learning	IBD	Small	Genomics Study (ML)	ML	Supervised Classifier	No	Small Sample Bias	Accuracy
143.	Post-operative Crohn's Disease Recurrence Prediction Using Machine Learning Algorithms	IBD	Moderate	ML Model Development (Prediction)	ML	Supervised Classifier	Internal	Follow-up Bias	AUC, C-index
145.	Artificial intelligence guided discovery of small molecule modulators for inflammatory bowel disease	IBD	Laboratory	Drug Discovery Study	ML	Search Algorithms	Yes (Experimental)	In Vitro Model Bias	Efficacy, Rank
146.	Artificial intelligence-rationalized balanced targeting of TNF-alpha and IL-1 beta for inflammatory bowel disease treatment	IBD	Laboratory	Drug Discovery Study	ML	Search Algorithms	Yes (Experimental)	In Vitro Model Bias	Efficacy, Rank
156.	Efficacy of real-time computer-aided detection system in colonoscopy: a prospective tandem study	Endoscopy	Large	Clinical Trial (Colonoscopy, CADe)	DL	CNN (Real-time Detection)	Yes (Prospective)	Operator Bias	ADR
157.	Use of computer-assisted detection (CADe) colonoscopy to improve adenoma detection rate: a randomized controlled trial	Endoscopy	Moderate	Clinical Trial	DL	CNN (Real-time	Yes (Prospective)	Operator Bias	ADR

				(Colonoscopy, CADe)		Detection)			
158.	A novel AI device for real-time optical diagnosis of colorectal polyps: A randomized clinical trial	Endoscopy	Moderate	Validation Study (Classification)	DL	CNN (Optical Classification)	Yes (External)	Image Bias	Accuracy, Sensitivity
162.	Optical diagnosis of colorectal polyp images using a deep learning model: an external validation study	Endoscopy	Moderate	Validation Study (Classification)	DL	CNN (Optical Classification)	Yes (External)	Image Bias	Accuracy, AUC
166.	Real-time computer-aided detection of colorectal neoplasia during colonoscopy: a randomized controlled trial	Endoscopy	Large	Randomized Clinical Trial (CADe)	DL	CNN (Real-time Detection)	Yes (Prospective)	Operator Bias	ADR, Polyp Rate
168.	Computer-aided detection versus advanced image enhancement techniques for adenoma detection: a randomized controlled trial	Endoscopy	Large	Clinical Trial (CADe vs NBI)	DL	CNN (Detection)	Yes (Prospective)	Operator Bias	ADR
169.	Impact of AI-aided colonoscopy in clinical practice: A large-scale observational study	Endoscopy	Large	Implementation/Observational Study	DL	CNN (Detection)	No	Single-Center Bias	ADR
170.	Usefulness of AI-equipped endoscopy for detecting colorectal neoplasia in clinical practice: a multicenter randomized controlled trial	Endoscopy	Moderate	Clinical Trial (CADe)	DL	CNN (Detection)	Yes (Prospective)	Operator Bias	ADR

171.	Use of a novel artificial intelligence system in colonoscopy: A randomized controlled trial of its impact on adenoma detection	Endoscopy	Moderate	Clinical Trial (CADe)	DL	CNN (Detection)	Yes (Prospective)	Operator Bias	ADR
172.	Artificial intelligence and colonoscopy experience: how AI affects the performance of endoscopists with different experience levels	Endoscopy	Moderate	Clinical Trial (CADe)	DL	CNN (Detection)	Yes (Prospective)	Experience Bias	ADR
173.	Use of artificial intelligence improves colonoscopy quality: a randomized controlled trial	Endoscopy	Moderate	Clinical Trial (CADe)	DL	CNN (Detection)	Yes (Prospective)	Operator Bias	ADR
174.	Improvement in adenoma detection rate using real-time computer-aided detection: an observational study	Endoscopy	Large	Observational Study (CADe)	DL	CNN (Detection)	No	Implementation Bias	ADR
177.	Performance of computer-aided detection in colonoscopy: a systematic review and meta-analysis of diagnostic accuracy	Endoscopy	Moderate	Detection Study (CADe)	DL	CNN (Detection)	Internal	Image Bias	Sensitivity
179.	Cost savings in colonoscopy with artificial intelligence: a decision analysis model	Endoscopy	Modeling	Pharmacoeconomic Model	N/A	N/A	Yes (Sensitivity)	Assumption Bias	Cost-Effectiveness
180.	Computer-aided diagnosis for optical diagnosis of colorectal polyps: A systematic review and meta-analysis	Endoscopy	Moderate	Classification Study (CADx)	DL	CNN (Classification)	Yes (External)	Image Bias	Accuracy
181.	Prevalence and clinical features of sessile serrated lesions: a machine learning approach	Endoscopy	Large	Prevalence Study (ML)	ML	Supervised Classifier	No	Single-Center Bias	Classification
182.	Artificial intelligence-based measurement of colorectal polyp size: a prospective validation study	Endoscopy	Moderate	Measurement Study (AI)	DL	CNN (Segmentation)	Internal	Gold Standard Bias	Measurement Error

183.	Computer-aided classification of colorectal segments for colonoscopy quality assessment: a machine learning and deep learning approach	Endoscopy	Moderate	Classification Study (ML/DL)	ML/DL	Supervised Classifier	Internal	Segmentation Bias	Accuracy
184.	Mapping the colon through the colonoscope's camera: a deep learning approach for anatomical location recognition	Endoscopy	Large	Database/Coordinate Model	ML	Localization Algorithms	Internal	Device Bias	Localization
185.	Polyp-Size: A Precise Endoscopic Dataset for Colorectal Polyp Size Estimation	Endoscopy	Large	Database (Polypometry)	N/A	N/A	N/A	Labeling Bias	N/A
186.	Artificial intelligence-based polyp size measurement for endoscopists: a prospective, multicenter study	Endoscopy	Moderate	Measurement Study (AI)	DL	CNN (Segmentation)	Internal	Gold Standard Bias	Measurement Error
187.	Influence of artificial intelligence on the diagnostic performance for Barrett's esophagus neoplasia: a randomized controlled trial	Endoscopy	Moderate	Diagnostic Trial (BE)	DL	CNN (Detection/Classification)	Yes (Video Trial)	Attention Bias	Sensitivity, Specificity
189.	A pilot study on automatic three-dimensional quantification of Barrett's esophagus length using machine learning	Endoscopy	Small	Pilot Study (3D BE)	ML	Reconstruction Algorithms	No	Small Sample Bias	Measurement Error
190.	Deep learning algorithm detection of Barrett's neoplasia during real-time endoscopy	Endoscopy	Moderate	DL Model Development (BE)	DL	CNN (Live Detection)	Yes (Live)	Endoscopic Quality Bias	Sensitivity, Specificity

191.	A deep learning system for detection of early Barrett's neoplasia: a multicentre retrospective study	Endoscopy	Moderate	DL Development and Validation	DL	CNN	Yes (Validation Sets)	Retrospective Bias	Sensitivity, Specificity, AUC
192.	Artificial intelligence system for the detection of Barrett's esophagus: a systematic review and meta-analysis	Endoscopy	Moderate	DL Model Development (BE)	DL	CNN	Internal	Image Bias	AUC, Sensitivity
193.	Towards a robust and compact deep learning system for computer-aided diagnosis of early Barrett's neoplasia: a multicenter study	Endoscopy	Moderate	DL Model Development (BE)	DL	CNN	Yes (Multicenter)	Transfer Bias	AUC
194.	Artificial intelligence using convolutional neural networks for real-time detection of early Barrett's neoplasia	Endoscopy	Moderate	DL Model Development (BE)	DL	CNN (Real-time Detection)	Yes (Video Trial)	Operator Bias	Sensitivity, Specificity
195.	Prospective development and validation of a volumetric laser endomicroscopy-based algorithm for the diagnosis of Barrett's esophagus	Endoscopy	Moderate	Algorithm Development (VLE)	ML	Classification Algorithms	Yes (Prospective)	Reference Bias	Accuracy
197.	Deep learning system compared with expert endoscopists for the diagnosis of early gastric cancer	Endoscopy	Moderate	Comparative Study (Gastric Cancer)	DL	CNN	Internal	Video Trial Bias	Accuracy, AUC
198.	Evaluation of the effects of an artificial intelligence-assisted system on the detection of early gastric cancer: a randomized controlled trial	Endoscopy	Moderate	Clinical Trial (Gastric Cancer, CADe)	DL	CNN (Detection)	Yes (Prospective)	Operator Bias	ADR

199.	Application of convolutional neural networks for evaluating the depth of invasion of early gastric cancer	Endoscopy	Moderate	DL Model Development (Invasion)	DL	CNN (Classification)	Internal	Image Bias	Accuracy
200.	Cooperation between artificial intelligence and endoscopists in the diagnosis of early gastric cancer: a prospective comparison study	Endoscopy	Moderate	Interaction Study (Gastric Cancer)	DL	CNN	Internal	Operator Bias	Accuracy, AUC
201.	An artificial intelligence system for comprehensive pathological prognosis in gastric cancer	Endoscopy	Moderate	DL Model Development (Prognosis)	DL	CNN	Internal	Pathological Data Bias	Accuracy, AUC
203.	Comparing blind spots of unsedated ultrafine, sedated standard, and sedated AI-assisted colonoscopy	Endoscopy	Moderate	Clinical Trial (Endoscopy)	DL	CNN (Blind Spot Detection)	Yes (Prospective)	Sedation Bias	ADR, Blind Spots
207.	Development and validation of a deep neural network for the assessment of endoscopic disease activity in ulcerative colitis	IBD	Moderate	DL Development and Validation (UC)	DL	Deep Neural Network	Yes (Multiple Datasets)	Lack of Prospective Application	Accuracy, Concordance (Kappa)
208.	Application of deep learning models to improve ulcerative colitis endoscopic activity scoring	IBD	Moderate	DL Model Development (Activity Score)	DL	CNN	Internal	Scoring Bias	Concordance (Kappa)

209.	Computer-aided diagnosis with monochromatic light for assessment of endoscopic activity in ulcerative colitis	IBD	Moderate	Diagnostic Study (Red Density)	ML	Image Processing	Internal	Illumination Bias	Classification
210.	An international multicenter real-life prospective study on the inter- and intra-observer agreement of the PICaSSO score in ulcerative colitis	IBD	Moderate	Validation Study (PICaSSO Score)	N/A	N/A	Yes (Multicenter)	Operator Bias	Concordance
211.	Fully automated diagnostic system with artificial intelligence for microscopic classification in inflammatory bowel disease	IBD	Moderate	System Development (Endocytoscopy)	DL	CNN	Internal	Image Bias	Sensitivity, Specificity
212.	Automatic, computer-aided determination of endoscopic disease activity in ulcerative colitis using Red Density	IBD	Moderate	Diagnostic Study (Red Density)	ML	Image Processing	Yes (External)	Reference Bias	AUC, Correlation
213.	Long-term follow-up of the red density pilot trial: Red Density correlates with histology and predicts relapse in ulcerative colitis	IBD	Moderate	Follow-up Study (Red Density)	ML	Image Processing	Yes (Follow-up)	Sampling Bias	Classification
214.	The AI-based Red Density score is correlated with the endoscopic subscore of the Mayo score in patients with ulcerative colitis	IBD	Moderate	Correlation Study (Red Density)	ML	Image Processing	Yes (Independent Cohort)	Reference Bias	Correlation
215.	Using computer vision to improve endoscopic disease activity assessment in ulcerative colitis	IBD	Moderate	DL Model Development (Quantification)	DL	Computer Vision (CV)	Yes (External)	Evaluation Bias	Correlation

217.	Gastroenterologist-level identification of small-bowel vascular lesions and polyps using deep learning for capsule endoscopy	Capsule Endoscopy	Large	DL Model Development (Capsule)	DL	CNN	Yes (Expert Comparison)	Gold Standard Bias	Accuracy, Sensitivity
218.	Artificial intelligence-assisted analysis of pan-enteric capsule endoscopy for Crohn's disease: A diagnostic study	Capsule Endoscopy	Moderate	Diagnostic Study (Capsule)	DL	CNN	Internal	Reading Bias	Sensitivity, Specificity
225.	Impact of an artificial intelligence-aided endoscopic imaging system on the training of novice endoscopists: a randomized controlled trial	Endoscopy	Moderate	Clinical Trial (Training)	DL	CNN (Detection)	Yes (Prospective)	Experience Bias	ADR
226.	Effectiveness of a novel artificial intelligence-assisted colonoscopy system in clinical practice: a multi-center retrospective cohort study	Endoscopy	Moderate	Cohort Study (CADe)	DL	CNN (Detection)	No	Single-Center Bias	ADR
228.	One-year review of real-time artificial intelligence (AI)-aided colonoscopy: experience from a large-volume academic center	Endoscopy	Large	Implementation Study (CADe)	DL	CNN (Detection)	No	Implementation Bias	ADR
229.	Cost-effectiveness of artificial intelligence for screening colonoscopy in the US	Endoscopy	Modeling	Pharmacoeconomic Model	N/A	N/A	Yes (Sensitivity)	Assumption Bias	Cost-Effectiveness
230.	PICaSSO Histologic Remission Index (PHRI) in ulcerative colitis: development and validation of an AI-assisted scoring system	IBD/Pathology	Moderate	Score Development/AI Application	ML	Supervised Classifiers	Yes (External	Pathology Bias	AUC, Concordance

							Cohort)		
236.	Endoscopist deskilling risk after exposure to artificial intelligence-assisted colonoscopy: an observational study	Endoscopy	Moderate	Observational Trial (Deskilling)	N/A	N/A	Yes (Multicenter)	Observation Bias	ADR
238.	Development of a multimodal machine-learning fusion model for predicting response to anti-TNF therapy in Crohn's disease	IBD	Moderate	ML Model Development (Multimodal)	ML	Fusion Classifier	Internal	Data Fusion Bias	AUC, Accuracy
239.	DeepGut: A collaborative multimodal large language model for gastrointestinal image analysis and report generation	General	Moderate	Multimodal LLM Development	LLM/ML	GPT-x / Multimodal	Internal	Language/Image Bias	AUC, Accuracy
242.	Optical classification of neoplastic colorectal polyps–a computer-assisted approach (the COACH study).	Endoscopy	Moderate	Classification Study (CADx)	ML	Supervised Classifier	Yes (Prospective)	Operator Bias	Accuracy, Sensitivity
243.	Prediction of polyp pathology using convolutional neural networks achieves “resect and discard” thresholds.	Endoscopy	Moderate	DL Development & Validation (CADx)	DL	CNN	Yes (External)	Image Bias	AUC, Accuracy
244.	Deep-learning system detects neoplasia in patients with Barrett’s esophagus with higher accuracy than endoscopists in a large multicenter video study.	Endoscopy	Large	DL Development & Validation	DL	CNN	Yes (Multi-Step Validation)	Retrospective Bias	Accuracy, AUC

245.	Development and validation of a real-time artificial intelligence-assisted system for detecting early gastric cancer: A multicentre retrospective diagnostic study.	Endoscopy	Large	DL Development & Validation (CADe)	DL	CNN	Yes (Multi center)	Retrospective Bias	Sensitivity, Specificity
246.	Automatic diagnosis of high-resolution esophageal manometry using artificial intelligence.	Neurogastroenterology	Small	ML Model Development (Manometry)	ML	Supervised Classifier	Internal	Small Sample Bias	Accuracy

This supplementary table provides a comprehensive summary of all research articles included in the review, detailing study design, dataset characteristics, machine learning approaches, and validation strategies.

The categorization of sample size used in this review reflects the typical data requirements for the development and evaluation of machine learning and deep learning models in clinical research. Datasets were classified as very small when they included fewer than 100 patients or primary samples, small for cohorts of 100 to 300, moderate for 301 to 1,000, large for 1,001 to 10,000, and very large when the number exceeded 10,000 patients or samples. Studies based exclusively on in vitro experiments or computational bioinformatics analyses without direct patient-level data were labeled as laboratory studies, while those relying on simulated or economic modelling approaches derived from previously published data were categorized as modeling studies.

Regarding validation strategies, external validation was defined as the evaluation of a model using an independent dataset, a multicenter cohort, or a prospective clinical design. Internal validation referred to studies in which performance was assessed using

the original dataset through hold-out partitions or cross-validation techniques. Studies that did not report any independent validation process were classified as having no external validation.

Acronyms: AI – Artificial Intelligence; ML – Machine Learning; DL – Deep Learning; CNN – Convolutional Neural Network; RNN – Recurrent Neural Network; LSTM – Long Short-Term Memory; SVM – Support Vector Machine; GBM – Gradient Boosting Machine; EHR – Electronic Health Record; AUC – Area Under the Curve (ROC); C-index – Concordance Index; ADR – Adenoma Detection Rate; EV – External Validation; CADe – Computer-Aided Detection; CADx – Computer-Aided Diagnosis; IBD – Inflammatory Bowel Disease; HCC – Hepatocellular Carcinoma; BE – Barrett's Esophagus; XAI – Explainable Artificial Intelligence; LLM – Large Language Model; N/A – Not Applicable/Not Stated.