



Artificial intelligence in inflammatory bowel disease: implications for clinical practice and future directions

Harris A. Ahmad¹, James E. East², Remo Panaccione³, Simon Travis², James B. Canavan¹, Keith Usiskin¹, Michael F. Byrne^{4,5}

¹Bristol Myers Squibb, Princeton, NJ, USA; ²Translational Gastroenterology Unit, Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK; ³Inflammatory Bowel Disease Clinic, University of Calgary, Calgary, AB; ⁴Division of Gastroenterology, Department of Medicine, University of British Columbia, Vancouver, BC; ⁵Satisfai Health, Vancouver, BC, Canada

Inflammatory bowel disease encompasses Crohn's disease and ulcerative colitis and is characterized by uncontrolled, relapsing, and remitting course of inflammation in the gastrointestinal tract. Artificial intelligence represents a new era within the field of gastroenterology, and the amount of research surrounding artificial intelligence in patients with inflammatory bowel disease is on the rise. As clinical trial outcomes and treatment targets evolve in inflammatory bowel disease, artificial intelligence may prove as a valuable tool for providing accurate, consistent, and reproducible evaluations of endoscopic appearance and histologic activity, thereby optimizing the diagnosis process and identifying disease severity. Furthermore, as the applications of artificial intelligence for inflammatory bowel disease continue to expand, they may present an ideal opportunity for improving disease management by predicting treatment response to biologic therapies and for refining the standard of care by setting the basis for future treatment personalization and cost reduction. The purpose of this review is to provide an overview of the unmet needs in the management of inflammatory bowel disease in clinical practice and how artificial intelligence tools can address these gaps to transform patient care. (**Intest Res 2023;21:283-294**)

Key Words: Artificial intelligence; Inflammatory bowel diseases; Endoscopy

INTRODUCTION

Inflammatory bowel disease (IBD) is a global disease with over 6.8 million cases worldwide reported in 2017, representing an 85% increase in prevalence since 1990.¹ IBD is a debilitating condition that involves recurring or chronic inflammation of the gastrointestinal tract and encompasses Crohn's disease (CD) and ulcerative colitis (UC). Currently, there is no single diagnostic criterion.² Instead, diagnosis and assessment of disease severity require a combination of a patient's history, examination, endoscopic, histologic, radiologic, and

biochemical investigations.³⁻⁵ Endoscopic evaluation defines the location and pattern of inflammation (e.g., segmental or continuous, mild or severe) and can also identify noninflammatory pathology (e.g., dysplasia).³ Endoscopic and histologic improvements correlate with better outcomes and are treatment targets in the management of IBD.^{6,7} Using an inflammatory pathway-based approach to inform treatment decisions may be more useful than classifying disease by organ involvement alone.⁸

The complexity of IBD management lends itself to artificial intelligence (AI) as a means of improving clinical practice (Fig. 1).³ Comprised of machine learning, deep learning, and neural networks, AI can help optimize IBD diagnosis, refine assessment of macroscopic and microscopic disease severity, and improve disease monitoring.⁹⁻¹¹

Received February 8, 2023. Revised March 10, 2023. Accepted March 11, 2023.

Correspondence to Michael F. Byrne, Division of Gastroenterology, Department of Medicine, Vancouver General Hospital/University of British Columbia, 5135-2775 Laurel Street, Vancouver, BC V5Z 1M9, Canada.

Tel: +1-604-875-5474, Fax: +1-604-628-2419, E-mail: Michael.byrne@vch.ca

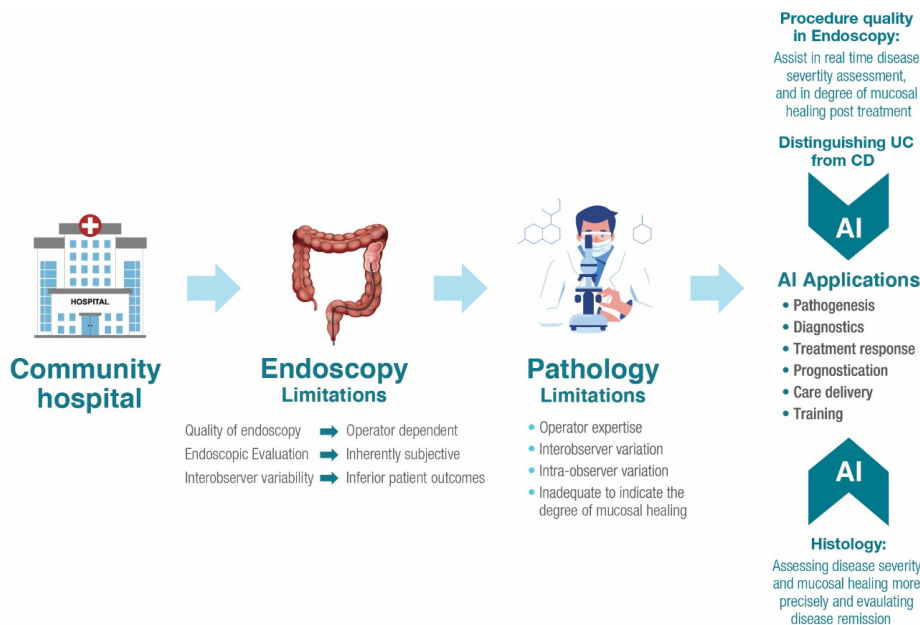


Fig. 1. Potential applications of artificial intelligence (AI) in inflammatory bowel disease diagnosis and management.³ CD, Crohn's disease; UC, ulcerative colitis.

ENDOSCOPY AND HISTOLOGY IN IBD CLINICAL ASSESSMENT: CURRENT USE AND UNMET NEEDS

Data collected from endoscopic and histologic assessments are used by physicians to determine the severity of IBD. Nevertheless, protocols are not standardized. Implementation of protocols may reflect lack of expertise, which presents an opportunity to develop AI tools, including histological scores, that may optimize IBD assessment.

1. Endoscopy

Endoscopic evaluation of IBD identifies inflammation, characterizes lesions, and assesses mucosal healing, but challenges remain.¹⁰ Visual evaluation of the mucosa relies on human interpretation, which is inherently subjective.^{3,12} The quality of endoscopy is dependent on the operator and training.¹³ Endoscopic expertise varies greatly, which is especially important in parts of the world where IBD is emerging while experience and equipment are limited.

Since therapy is influenced by endoscopic assessment,^{7,13} interobserver variability contributes to inferior patient outcomes.^{12,14} Scoring of disease severity is only semi-quantitative and not often practiced.¹⁵ For example, variability in endoscopic scoring among 58 gastroenterologists revealed an interrater agreement of only 0.47 for Mayo endoscopic subscore

ratings in patients with UC and 0.33 for Rutgeerts score in patients with CD.¹⁴ Based on this variability, study authors estimated that one-third of patients would be managed differently based on endoscopic data alone. In another study, novel computer vision-enabled endoscopic disease distribution measures were able to better detect the significant therapeutic effect of ustekinumab over placebo in UC compared with traditional endoscopic scoring instruments.¹⁵ This illustrates that objective computer-aided activity measures may improve endoscopic IBD data quality by refining therapeutic efficacy assessment compared with conventional scoring, enhancing efficiency in clinical trials and revamping therapeutic disease monitoring in the care of UC.

2. Histology

In the last few decades, endoscopic remission was considered the most important treatment target. However, in recent years it was learned that microscopic inflammation is seen in 16% to 100% of biopsies from colonic mucosa of patients with endoscopic healing. Furthermore, in patients with UC, persistent histologic inflammation can exist in the presence of endoscopic mucosal healing and is associated with increased risks of dysplasia. In a retrospective study, a computer-aided diagnosis (CADx) system predicted persistent histologic inflammation in patients with UC using endocytoscopic images with an accuracy of 91.0%, a sensitivity of 74.0%, and a specificity of

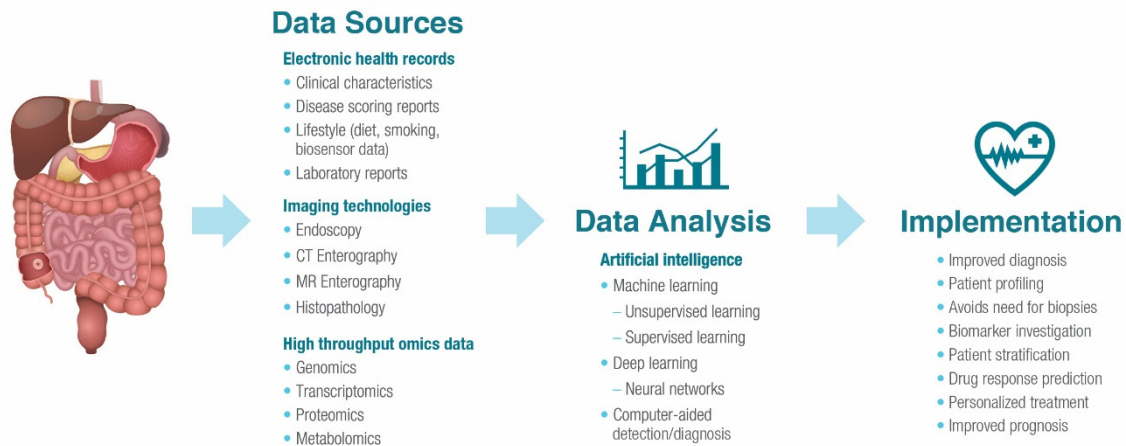


Fig. 2. Potential benefits of the application of artificial intelligence in inflammatory bowel disease clinical practice. CT, computed tomography; MR, magnetic resonance. Modified from Seyed Tabib NS, et al. Gut 2020;69:1520-1532.¹⁰

97.0%.¹⁶ Therefore, endoscopic healing alone is limited in predicting long-term disease outcomes.⁶

Histologic remission is associated with a sustained clinical remission and a lower risk of colectomy or colorectal cancer, and is recognized as a potential new therapeutic target.⁶ However, mucosal biopsies involve arbitrary decisions on biopsy location and delays in pathologist reading, and are invasive and costly. Optical biopsy, such as confocal laser endomicroscopy with higher magnification and resolution of the mucosa, is appealing since it has been shown to predict histologic inflammation and outcomes.¹⁷ Nevertheless, as with endoscopy, such techniques depend on the operator and human interpretation, and remain vulnerable to interobserver variation.¹⁶

Consequently, there has been growing evidence that supports the use of histological scoring systems in IBD. Several scores have been developed and include the Robarts histopathology index and the Nancy index, the only 2 scores recommended by the European Crohn's and Colitis Organisation for use in patients with UC.¹⁸ Most scores for CD have not been validated due to their complexity and the discontinuous nature of lesions in CD. However, the recent guidelines support the adoption of the same scoring systems for CD and UC as the main histological features of UC and CD activity are shared, and the recent evidence demonstrates that UC and CD responses to treatment can be interchangeably measured.¹⁸

ROLE OF AI IN IBD CLINICAL PRACTICE

AI can support physicians managing patients with IBD in making more informed, real-time treatment decisions during en-

doscopy and in assessing histopathology. Applications of AI exist at multiple points in the patient care pathway, including increasing procedure quality, differentiating between CD and UC, assessing disease severity, assessing histologic severity and remission, and identifying bleeding sources, among others (Fig. 2).^{10,19}

1. Opportunities for AI in IBD Assessment

1) Procedure Quality

Recently, there has been much focus on quality of colonoscopy in IBD. In 2022, the European Society of Gastrointestinal Endoscopy released guidelines on performance measures for colonoscopy in IBD, including general measures such as bowel preparation and cecal and ileal intubation; endoscopic evaluation of disease activity, with the use of validated scores; and colonoscopic surveillance for dysplasia.²⁰ AI can assist endoscopists in real time, including decreasing sampling errors, assisting with photodocumentation to improve the completeness of examinations, preventing missed blind spots, and monitoring quality metrics, thus enabling endoscopists to focus on areas of interest.²¹ Such improvements in procedure quality may optimize both the current daily practice of and training in IBD assessment, ultimately leading to optimized patient care.

Examination quality metrics incorporated into machine-learning algorithms prevent poor-quality videos by alerting the endoscopist in real time to issues that would otherwise require the patient to return for re-evaluation. An algorithm to detect and alert the endoscopist to missed areas in real time demonstrated agreement with the physician reviewer in 93%

of cases.²² AI can identify blood or stool as well as alert the endoscopist if they are moving too quickly through the examination, thus helping reduce the chance for blind spots. A set of 5,476 images from 2,000 colonoscopy patients was used to train a deep convolutional neural network (CNN) model to assess bowel preparation every 30 seconds during the withdrawal phase, providing the endoscopist with a real-time, accurate (accuracy range, 80.0%–93.3%) method for evaluating bowel preparation.²³ AI can also be used to quantify the percentage of colonic surface area visualized, report on the clarity of the endoscopic view,²⁴ and identify artefacts and restore corrupted visual data.²⁵ A CNN model detected and classified visual artefacts, generated a quality score for each video frame in nearly real time, and restored, on average, 25% of video frames in a dataset of 1,290 endoscopy images.²⁵

2) Distinguishing CD versus UC

AI can improve the ability to discriminate CD from UC.³ One study investigated deep-learning classifiers for processing gene-expression data in IBD, combining a deep neural network with a support vector machine.²⁶ The deep-learning system distinguished CD from UC with 95.0% accuracy. Support vector machines have also been used to create spectral histopathology, an automatic calculation of tissue morphology from Raman spectroscopy, that can reveal the same morphological information as a classic hematoxylin and eosin staining.²⁷ In a second step, this support vector machine used Raman spectral signatures to differentiate IBD subtypes with an accuracy of 98.9%. In another example, a natural language processing algorithm using random forest and CNN approaches discerned among patients with CD, UC, and intestinal tuberculosis using a description of the endoscopic image in the form of free text.²⁸ This approach distinguished CD from UC with an area under the curve of 0.936, a sensitivity of 0.890, and a specificity of 0.837.

3) Capsule Endoscopy

A deep-learning framework was evaluated for its ability to detect CD lesions in the small bowel and colon, determine lesion localization, and assess lesion severity using images collected with pan-enteric capsule endoscopy.²⁹ A total of 7,744 images from 38 patients with suspected or known CD were included, and the automated framework detected ulcerations consistent with CD with 95.7% sensitivity, 99.8% specificity, and diagnostic accuracy of 98.4%, demonstrating high efficiency and robustness. In addition, the diagnostic accuracy was similar

for ulcerations located in the small bowel (98.5%) and colon (98.1%).

4) Assessment of Disease Remission

Computer-assisted support systems can assess IBD disease activity, and the standardization of image capturing will improve the accuracy of algorithms in reflecting clinical scenarios.¹⁹ Deep-learning models can accurately identify remission using the Mayo endoscopic score (MES). Recently, a CAD model discriminated between endoscopic mucosal healing (MES 0, 1) and non-mucosal healing (MES 2, 3) with 94.5% accuracy, 84.6% sensitivity, and 96.9% specificity.³⁰ An automated MES scoring system that used a CNN model distinguished remission (MES 0, 1) versus active disease (MES 2, 3) in 84% (221 of 264) of videos.³¹ The model also identified informative versus noninformative images with an area under the curve of 0.961. A deep-learning CADx system was developed to mimic the assessment done by a gastroenterologist: colonoscopy image assessment, lesion identification, and Mayo score generation. Based on an evaluation of 1,672 endoscopic videos, the system achieved high accuracy with an area under the curve of 0.84 for a Mayo subscore ≥ 1 , 0.85 for a subscore of ≥ 2 , and 0.85 for a subscore of ≥ 3 .³²

AI can infer pathologic activity from endoscopic image analysis. In a prospective study of 875 patients with UC, a deep neural network assessed endoscopic images for endoscopic and histologic remission with 90.1% and 92.9% accuracy, respectively, allowing for the identification of remission without the need for mucosal biopsy.³³ Histologic remission is the best predictor of long-term outcomes in patients with UC, but histopathology scoring is cumbersome. Red density is an operator-independent tool that uses machine learning to calculate a score based on red pixel values and vascular pattern detection in endoscopic images.³⁴ The red density algorithm was optimized to correlate with endoscopic and histologic disease activity. In a validation study, the algorithm showed a significant correlation with Roberts histopathology index ($r=0.65$, $P<0.0001$), the Ulcerative Colitis Endoscopic Index of Severity ($r=0.56$, $P=0.0004$), and MES ($r=0.61$, $P<0.0001$). While the red density score needs further validation, it is a promising new tool that provides objective operator-independent digital scoring of endoscopic and histologic disease. In practical terms, AI-assisted assessment of endoscopic findings that correlate with histopathology will improve the predictive value of endoscopy and reduce the need for invasive and costly biopsy procedures.

2. Opportunities for AI to Support Clinical Management of IBD

AI can provide decision support to optimize the treatment of patients with IBD through predicting response to treatment or the need for surgery. AI that predicts treatment outcomes can aid in clinical decision-making related to patient selection, switching treatment, and determining response or relapse. Both therapeutic response and lack of response are useful insights into the efficacy of an intervention.³⁵ Machine learning can predict patient responses to treatment based on clinical data. For patients with UC, vedolizumab is an effective therapy with clinical improvements continuing after 30 weeks, albeit slow to produce clinical results. Using phase 3 clinical trial laboratory results up to week 6, a machine-learning model predicted corticosteroid-free endoscopic remission at week 52, with an area under the curve of 0.73 (95% confidence interval [CI], 0.65–0.82).³⁵ In another study, machine learning predicted the efficacy of vedolizumab at week 22 in patients with UC using clinical data collected at baseline.³⁶ The model revealed high negative predicted value (92.3%) for corticosteroid-free clinical remission with vedolizumab, indicating that other treatment options might be considered.³⁶ For patients with CD, a machine-learning model predicted response to ustekinumab treatment beyond week 42, with an area under the curve of 0.78 (95% CI, 0.69–0.87), using laboratory data up to week 8 from three phase 3 clinical trials.³⁷ Predicting the probability of nonresponse after a short trial is valuable as patients can quickly switch therapies, aiding in overall disease management and decreasing costs.

Furthermore, machine-learning approaches can aid in personalizing medication dosing. Thiopurines are widely used immunomodulators for the treatment of UC and CD; however, dose optimization is difficult, with physicians relying on patterns in the complete blood count to monitor clinical response and titrate dosing. In one study, machine-learning algorithms predicted thiopurine response from age and laboratory data (area under the curve of 0.79 compared with 0.49 with 6-thioguanine nucleotide metabolite measurement).³⁸ In another study, a neural network predicted the need for surgery after cytopheresis therapy in UC with a sensitivity of 0.96 and specificity of 0.97.³⁹

Therapies that neutralize tumor necrosis factor (TNF) are efficacious in the treatment of IBD; however, many patients do not respond to anti-TNF therapy and are exposed to side effects such as infections, skin disorders, and lupus-like autoimmunity.⁴⁰ AI can predict anti-TNF therapeutic responders,

enhancing the safety and cost effectiveness of this treatment. Fecal calprotectin measurements taken after induction of anti-TNF therapy (infliximab) were used to predict clinical response and endoscopic remission (mucosal healing) after 1 year.⁴¹ Clinical response was predicted with 83.0% sensitivity and 74.0% specificity; mucosal healing was predicted with 79.0% sensitivity and 57.0% specificity. A neural network machine-learning model used baseline parameters to predict UC disease activity and risk of relapse at 1-year with anti-TNF therapy (infliximab/adalimumab). The model demonstrated excellent performance with 90.0% accuracy on the test set and 100.0% accuracy on the validation set.⁴² In another study, confocal laser endomicroscopy was used for *in vivo* molecular imaging of membrane-bound TNF (mTNF) expressing cells in the gastrointestinal mucosa of patients with CD.⁴⁰ The investigators found patients with a high number of mTNF-positive cells in the colon had significantly higher probability of clinical response to anti-TNF therapy compared with patients who had a low number of mTNF-positive cells (92.0% vs. 15.0%).

IBD pathogenesis involves environmental, genetic, microbial, and immune factors, thereby warranting a comprehensive approach to care. The integration of omics data into clinical practice using AI can provide personalized medicine in real time to improve quality of care and patient outcomes.^{10,43} A machine-learning model predicted endoscopic response to ustekinumab in patients with CD by integrating genomics and transcriptomics data.⁴⁴ The study identified 10- and 15-feature transcriptomic and genomic panels, respectively, that can predict endoscopic response to therapy. Additionally, multi-omics profiling can identify proteomic, metabolomic, and microbial biomarkers associated with relapse in patients with quiescent IBD.⁴⁵ CDPATH is a patient-facing web-based program that uses biomarkers and clinical data to create a personalized prediction of CD prognosis over a 3-year period.⁴⁶ By stratifying low-, medium-, and high-risk patients, treatment escalation can be tailored to individual patients. CDPATH allows both providers and patients to visualize individualized risks of complications over time, facilitating patient empowerment and shared decision-making. Recently, an algorithm- and biomarker-based test, called the endoscopic mucosal healing index (EHI), was developed to identify patients with CD in endoscopic remission and measures 13 proteins in the blood, encompassing 6 categories of mucosal healing.⁴⁷ The biomarkers include C-reactive protein and SAA1 (inflammation); ANG1 and ANG2 (angiogenesis); MMP1, MMP2, MMP3, MMP9, and EMMPRIN (matrix remodeling); TGF α (proliferation and re-

pair growth factor); IL-7 (immune recruitment modulation); and CEACAM1 and VCAM1 (cell adhesion). The EHI test demonstrated favorable accuracy for identifying endoscopic inflammation and reasonable diagnostic accuracy for identifying histologic inflammation. In addition, EHI performance was comparable to fecal calprotectin and superior to C-reactive protein.

Another opportunity for AI in IBD management resides with wearable devices with biosensors that can correlate daily function with disease activity in IBD patients and play a role in disease management.⁴⁸ Fitbit metrics, including daily steps, heart rate, and sleep data, were collected from 56 patients 1 week prior to a disease assessment. During the assessment, clinical visit data, C-reactive protein and fecal calprotectin values, and colonoscopy data were obtained. A total of 132 disease assessments were obtained, with 66 assessments showing active disease and 66 showing quiescent disease. Patients with active disease had fewer daily steps compared with patients with quiescent disease (6,331 steps vs. 8,241 steps, respectively; $P < 0.001$). Daily number of steps was predictive of elevated biomarkers of inflammation, with an area under the curve of 0.65 (95% CI, 0.61–0.69).

ROLE OF AI ACROSS DISCIPLINES

Medical imaging, including pathology, radiology, and endoscopy, has been an early adopter of AI across disciplines.¹²

1. Pathology

AI can recognize regions of interest in histology slides. A multi-instance, deep-learning network model designed for whole slide image classification identified gastric cancer with an 86.5% accuracy using a dataset of 608 images.⁴⁹ AI algorithms designed for colorectal cancer can distinguish normal tissue, hyperplasia, adenoma, adenocarcinoma, and histologic subtypes of polyps or adenocarcinomas.¹¹ For example, a deep-learning model classified colorectal polyp subtypes (i.e., hyperplastic, sessile serrated, traditional serrated, tubular, and tubulovillous/villous) in a set of 239 whole slide images with an accuracy of 93.0%.⁵⁰ Another deep neural network classified the 4 most common colorectal polyp types with an accuracy of 87.0%, which was comparable with that of local pathologists (86.6%).⁵¹ Using a murine model of gut inflammation, researchers trained a deep-learning algorithm to recognize key features of inflamed and noninflamed mucosa from microscopic images of pathological sections.⁵²

AI applications have utility in breast cancer pathology for rare-event identification, tumor percentage calculation, and mitosis detection, all of which are time consuming and susceptible to interobserver variability.⁵³ Deep-learning algorithms designed to detect lymph node metastases in tissue sections from women with breast cancer found the algorithms outperformed a panel of pathologists in a simulated time-constrained diagnostic setting (area under the curve of 0.99 for the best algorithm vs. 0.88 for the best pathologist).⁵⁴ In another study, a supervised deep-learning model for mitosis detection from whole slide images was trained using handcrafted features (morphological, textural, and intensity features) extracted from datasets of previous AI medical challenges and reported high precision (92.0%), recall (88.0%), and F-score (90.0%).⁵⁵

2. Radiology

AI is used in radiology for screening, disease classification, and disease characterization. Computer-aided detection (CADe) systems help analyze screening mammograms by marking suspicious regions for further review. Improvements upon this technology include a CADe system based on deep CNNs to classify malignant or benign lesions in mammograms.⁵⁶ This system reported an area under the curve of 0.95 with high sensitivity (0.9) and low false-positive rate (0.3 false-positive marks per image) compared with commercially available systems (up to 1.25 false-positive marks per image). The utility of AI in reducing workload was explored in a retrospective evaluation consisting of 15,987 mammograms from the Córdoba Tomosynthesis Screening Trial and found AI-supported breast cancer screening strategies could reduce workload by up to 70.0% without reducing cancer detection.⁵⁷ Compared with the original double reading of digital mammography images, AI-based digital tomography screening was associated with a 29.7% reduction in workload, a 25.0% improvement in sensitivity, and a 27.1% reduction in recall rate.

3. Endoscopy

CADe and CADx models used to detect polyps have improved procedure performance.¹⁹ A deep-learning CADe system developed for real-time use in clinical practice detected neoplasms with 88.0% accuracy, 93.0% sensitivity, and 83.0% specificity, achieving higher accuracy compared with general endoscopists (88.0% vs. 73.0%, respectively).⁵⁸ In a case-control diagnostic study of 1 million endoscopy images from more than 84,000 patients, a real-time CADe system for upper gastrointestinal cancer achieved high accuracy (95.5%) in the internal

validation set, with comparable sensitivity to experienced endoscopists.⁵⁹ Moreover, a CNN model differentiated between mucosal and submucosal invasive Barrett’s cancer with accuracy of 71.0%, sensitivity of 77.0%, and specificity of 64.0%.⁶⁰

Real-time endohistologic visualizing systems provide histologic inference during colonoscopy. A multicenter study examining the diagnostic accuracy of a CNN model in distinguishing neoplasms from non-neoplasms using endocytoscopic images found the model performed similarly to well-trained specialists (96.0% vs. 94.6% accuracy; $P=0.141$) and significantly better than general gastroenterologists (96.0% vs. 70.4% accuracy; $P<0.0001$).⁶¹ In April 2021, the U.S. Food and Drug Administration (FDA) approved an AI system that detects colonic lesions (adenomas or carcinomas) during colonoscopy at a higher rate compared to standard colonoscopy (55.1% vs. 42.0%).⁶²

BARRIERS TO AI IMPLEMENTATION

Barriers to the adoption of AI in clinical practice include lack of standardized data, data-sharing limitations, educational barriers and physician hesitancy, regulatory hurdles, and cost barriers (Table 1).⁵³ Heterogeneity of data sources used for training and validation (e.g., missing or irrelevant data) can decrease performance of the AI model in a real-world setting. In addition, the complexity of real-world conditions may not be adequately incorporated into AI algorithms and, thus, AI tools in practice may yield lower accuracies than reported in the literature.⁵³ Rare clinical scenarios could also challenge AI models, as these scenarios will have less representation in training datasets. To address this shortcoming, high-quality, standardized datasets are needed to ensure geographic, technical, and patient diversity.⁹ The American Society for Gastrointestinal Endoscopy has proposed a professionally managed image library¹⁹; however, the requirements to ensure verified

diagnoses for publicly available datasets are ambiguous.

Physician distrust, technophobia, liability concerns, and a fear that AI may replace physicians could lead to hesitancy in the adoption of AI tools.^{53,63} In a survey of 487 pathologists from 59 countries, most respondents (72.0%) felt AI would have a positive effect on diagnostic efficiency, although the majority also thought the diagnostic decision-making process should remain predominantly a human task.⁶⁴ The FDA regulatory approval process for software as a medical device is evolving, and the gastroenterology field will play a key role.⁶⁵ In January 2021, the FDA published the “Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan.”⁶⁶ As noted in the action plan, stakeholder concerns for AI include the labeling for AI/ML-based devices and the need for manufacturers to clearly describe the data that were used to train the algorithm, the logic employed, the role intended to be served by its output, and the evidence of the device’s performance. Therefore, ensuring an alliance among physicians, clinicians, bioinformaticians, and regulatory authorities to develop protocols and guidelines applicable to AI algorithms may be the first step to improving real-world practicality of AI.

Although data on cost effectiveness of AI in health care are limited, AI tools are expected to reduce overall costs due to reduction in endoscopic procedure burden.⁶⁷ For example, using AI to classify diminutive colonic polyps *in vivo*, rather than sending samples for pathology, resulted in an estimated cost savings of \$85.2 million in the United States.⁶⁸ Nonetheless, substantial up-front investment may be required to incorporate AI into clinical practice.⁶⁹ In the current fee-for-service reimbursement framework, the adoption of AI may be difficult; however, in a value-based model where improving quality at decreased costs is important, AI will likely become a valuable adjunct. It is possible that payers will cover a new drug or drug continuation only if an accurate assessment of a patient’s dis-

Table 1. Barriers to AI Implementation⁵³

Barrier	Comment
Lack of standardized data	Heterogeneity of data sources used for training and validation
Data-sharing limitations	High-quality datasets needed to ensure geographic, technical, and patient diversity
Educational barriers and physician hesitancy	Physician distrust, technophobia, liability concerns, and a fear that AI may replace physicians
Regulatory hurdles	Evolving regulatory approval process for software as a medical device; concerns with labeling for AI/ML-based devices
Cost barriers	Substantial up-front investment may be required to incorporate AI into clinical practice; financial incentives provided through reimbursement fee codes will be needed

AI, artificial intelligence; ML, machine learning.

ease activity using AI-based technology is available. Thus, payer requirements for reimbursement will be key to the adoption of AI in health care. Indeed, it is expected that AI-assisted optical biopsy will not achieve widespread use in clinical practice unless there are financial incentives provided through reimbursement fee codes.⁷⁰ In recent studies, AI models with insurance claims data were capable of accurately predicting IBD-related hospitalization and steroid use within a 6-month period in patients with IBD, while AI models with gene-based data outperformed more costly biomarker analyses for predicting outcomes.⁷¹ Additionally, AI confirmation of investigator scoring without the need for central reading may lead to potential cost savings on drug development. Through consistent evaluation of endoscopic disease severity, AI can relieve the time burden of extensive procedure assessments while also improving quality of IBD endoscopy and patient care, a theme congruent to the information burden borne by physicians resulting from the dramatic increase in medical literature.⁷² As many AI studies are retrospective, or done in limited settings, more real-world data are needed.⁵³ The widespread adoption of AI tools in clinical practice hinges on the ability to demonstrate improvements in efficiency and accuracy that generate sufficient return on investment.

CONCLUSIONS

Although challenges remain, research clearly supports the application of AI in improving the quality of IBD diagnosis and management. AI-based tools can maintain consistent, objective, accurate, and accelerated clinical assessments, predict treatment responses, and improve the quality of endoscopy at all levels. Inevitably, AI models will continually improve as the technology becomes widely available and more data are incorporated in the algorithms. Therefore, the current challenges in the diagnosis and management of IBD present ideal future opportunities for transforming patient care using AI.

ADDITIONAL INFORMATION

Funding Source

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

Ahmad HA and Canavan J are employees of Bristol Myers Squibb.

East JE reports personal fees from Boston Scientific, Falk, Lumendi, Paion, and Satisfai, outside the submitted work. In addition, he has a patent Methods and framework for assessing image quality issued, and a patent Quantification of Barrett's esophagus issued. Panaccione R reports personal fees from Abbott, AbbVie, Alimentiv (formerly Robarts), Amgen, Arena Pharmaceuticals, AstraZeneca, Biogen, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Celltrion, Cosmos Pharmaceuticals, Eisai, Elan, Eli Lilly, Ferring, Fresenius Kabi, Galapagos, Genentech, Gilead Sciences, GlaxoSmithKline, HC3 Communications, Janssen, Meducom, Merck, Mylan, Oppilan, Organon, Pandion Pharma, Pfizer, Progenity, Protagonist Therapeutics, Receptos, Roche, Sandoz, Satisfai Health, Schering-Plough, Shire, Sublimity Therapeutics, Takeda Pharmaceuticals, Theravance Biopharma, Trellus Health, and UCB. Travis S has served as a paid consultant to AbbVie, Allergan, Amgen, Asahi, Bioclinica, Biogen, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, ChemoCentryx, Cosmo, Enterome, Equillum, Ferring, GSK, Genentech, Genzyme, Giuliani SpA, Immunocore, Immunometabolism, Janssen, Lilly, MSD, Merck, Metastag, Neovacs, Novo Nordisk, NPS Pharmaceuticals, Pfizer, Proximagen, Receptos, Roche, Satisfai Health, Sensyne Health, Shire, Sigmoid Pharma, Sorriso, Takeda, Topivert, UCB, VHSquared, Vifor, and Zeria. He has received grants and/or has grants pending from AbbVie, ECCO, Helmsley Trust, IOIBD, Janssen, Lilly, Norman Collisson Foundation, Pfizer, UCB, UK-IERI, and Vifor. He has received honoraria from AbbVie, Amgen, Biogen, Ferring, Lilly, Pfizer, and Takeda. He has had travel/accommodation expenses covered or reimbursed by AbbVie, Amgen, Biogen, Ferring, Lilly, JNJ, Pfizer, and Takeda. Usiskin K was an employee of Bristol Myers Squibb at the time of manuscript initiation. He reports personal fees from Arena, Bristol Myers Squibb, Crinetics Pharmaceuticals, Insmad, and Locust Walk Capital. Byrne MF is CEO and Founder of Satisfai Health.

Panaccione R and Travis S are editorial board members of the journal but were not involved in the peer reviewer selection, evaluation, or decision process of this article. No other potential conflicts of interest relevant to this article were reported.

Data Availability Statement

No new data were generated or analyzed in support of this research.

Author Contributions

Conceptualization: Ahmad HA, Byrne MF, East JE, Usiskin K,

Canavan JB. Writing - original draft: all authors. Writing - review & editing: all authors. Approval of final manuscript: all authors.

Additional Contributions

Professional medical writing support from Gorica Malisanovic, MD, PhD, and editorial assistance were provided by Peloton Advantage, LLC, an OPEN Health company, Parsippany, NJ, USA, and were funded by Bristol Myers Squibb.

ORCID

Ahmad HA <https://orcid.org/0000-0002-0890-2071>
 East JE <https://orcid.org/0000-0001-8035-3700>
 Panaccione R <https://orcid.org/0000-0002-5247-940X>
 Travis S <https://orcid.org/0000-0002-2690-4361>
 Canavan JB <https://orcid.org/0000-0003-0701-8396>
 Usiskin K <https://orcid.org/0000-0003-4806-6554>
 Byrne MF <https://orcid.org/0000-0002-7740-7795>

REFERENCES

1. GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* 2020;5:17-30.
2. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019;13:144-164.
3. Cohen-Mekelburg S, Berry S, Stidham RW, Zhu J, Waljee AK. Clinical applications of artificial intelligence and machine learning-based methods in inflammatory bowel disease. *J Gastroenterol Hepatol* 2021;36:279-285.
4. Gajendran M, Loganathan P, Catinella AP, Hashash JG. A comprehensive review and update on Crohn's disease. *Dis Mon* 2018;64:20-57.
5. Gajendran M, Loganathan P, Jimenez G, et al. A comprehensive review and update on ulcerative colitis. *Dis Mon* 2019;65: 100851.
6. Dal Buono A, Roda G, Argollo M, Peyrin-Biroulet L, Danese S. Histological healing: should it be considered as a new outcome for ulcerative colitis? *Expert Opin Biol Ther* 2020;20:407-412.
7. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: an update on the selecting therapeutic targets in inflammatory bowel disease (STRIDE) initiative of the International Organization for the Study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. *Gastroenterology* 2021;160: 1570-1583.
8. Schett G, McInnes IB, Neurath MF. Reframing immune-mediated inflammatory diseases through signature cytokine hubs. *N Engl J Med* 2021;385:628-639.
9. Le Berre C, Sandborn WJ, Aridhi S, et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020;158:76-94.
10. Seyed Tabib NS, Madgwick M, Sudhakar P, Verstockt B, Korcsmaros T, Vermeire S. Big data in IBD: big progress for clinical practice. *Gut* 2020;69:1520-1532.
11. Pannala R, Krishnan K, Melson J, et al. Artificial intelligence in gastrointestinal endoscopy. *VideoGIE* 2020;5:598-613.
12. Chahal D, Byrne MF. A primer on artificial intelligence and its application to endoscopy. *Gastrointest Endosc* 2020;92:813-820.
13. Sinonquel P, Eelbode T, Bossuyt P, Maes F, Bisschops R. Artificial intelligence and its impact on quality improvement in upper and lower gastrointestinal endoscopy. *Dig Endosc* 2021; 33:242-253.
14. Fernandes SR, Pinto JS, Marques da Costa P, Correia L; GEDII. Disagreement among gastroenterologists using the Mayo and Rutgeerts Endoscopic Scores. *Inflamm Bowel Dis* 2018;24: 254-260.
15. Stidham R, Yao H, Soroushmehr R, et al. 796 Computer vision measurement of disease severity distribution outperforms traditional endoscopic scoring for detecting therapeutic response in a clinical trial of ustekinumab for ulcerative colitis. *Gastroenterology* 2022;162:S-193.
16. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019; 89:408-415.
17. Kiesslich R, Duckworth CA, Moussata D, et al. Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease. *Gut* 2012;61: 1146-1153.
18. Vespa E, D'Amico F, Sollai M, et al. Histological scores in patients with inflammatory bowel diseases: the state of the art. *J Clin Med* 2022;11:939.
19. Berzin TM, Parasa S, Wallace MB, Gross SA, Repici A, Sharma P. Position statement on priorities for artificial intelligence in GI endoscopy: a report by the ASGE Task Force. *Gastrointest Endosc* 2020;92:951-959.

20. Dekker E, Nass KJ, Iacucci M, et al. Performance measures for colonoscopy in inflammatory bowel disease patients: European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *Endoscopy* 2022;54:904-915.
21. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut* 2019;68:2161-2169.
22. Freedman D, Blau Y, Katzir L, et al. Detecting deficient coverage in colonoscopies. *IEEE Trans Med Imaging* 2020;39:3451-3462.
23. Zhou J, Wu L, Wan X, et al. A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020;91:428-435.
24. Thakkar S, Carleton NM, Rao B, Syed A. Use of artificial intelligence-based analytics from live colonoscopies to optimize the quality of the colonoscopy examination in real time: proof of concept. *Gastroenterology* 2020;158:1219-1221.
25. Ali S, Zhou F, Bailey A, et al. A deep learning framework for quality assessment and restoration in video endoscopy. *Med Image Anal* 2021;68:101900.
26. Smolander J, Dehmer M, Emmert-Streib F. Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders. *FEBS Open Bio* 2019;9:1232-1248.
27. Bielecki C, Bocklitz TW, Schmitt M, et al. Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells. *J Biomed Opt* 2012;17:076030.
28. Tong Y, Lu K, Yang Y, et al. Can natural language processing help differentiate inflammatory intestinal diseases in China? Models applying random forest and convolutional neural network approaches. *BMC Med Inform Decis Mak* 2020;20:248.
29. Majtner T, Brodersen JB, Herp J, Kjeldsen J, Halling ML, Jensen MD. A deep learning framework for autonomous detection and classification of Crohn's disease lesions in the small bowel and colon with capsule endoscopy. *Endosc Int Open* 2021; 9:E1361-E1370.
30. Huang TY, Zhan SQ, Chen PJ, Yang CW, Lu HH. Accurate diagnosis of endoscopic mucosal healing in ulcerative colitis using deep learning and machine learning. *J Chin Med Assoc* 2021;84:678-681.
31. Yao H, Najarian K, Gryak J, et al. Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2021;93:728-736.
32. Gutierrez Becker B, Arcadu F, Thalhammer A, et al. Training and deploying a deep learning model for endoscopic severity grading in ulcerative colitis using multicenter clinical trial data. *Ther Adv Gastrointest Endosc* 2021;14:2631774521990623.
33. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology* 2020;158:2150-2157.
34. Bossuyt P, Nakase H, Vermeire S, et al. Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density. *Gut* 2020;69:1778-1786.
35. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther* 2018;47:763-772.
36. Miyoshi J, Maeda T, Matsuoka K, et al. Machine learning using clinical data at baseline predicts the efficacy of vedolizumab at week 22 in patients with ulcerative colitis. *Sci Rep* 2021;11: 16440.
37. Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and validation of machine learning models in prediction of remission in patients with moderate to severe Crohn disease. *JAMA Netw Open* 2019;2:e193721.
38. Waljee AK, Sauder K, Patel A, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J Crohns Colitis* 2017;11:801-810.
39. Takayama T, Okamoto S, Hisamatsu T, et al. Computer-aided prediction of long-term prognosis of patients with ulcerative colitis after cytoapheresis therapy. *PLoS One* 2015;10:e0131197.
40. Atreya R, Neumann H, Neufert C, et al. In vivo imaging using fluorescent antibodies to tumor necrosis factor predicts therapeutic response in Crohn's disease. *Nat Med* 2014;20:313-318.
41. Guidi L, Marzo M, Andrisani G, et al. Faecal calprotectin assay after induction with anti-tumour necrosis factor α agents in inflammatory bowel disease: prediction of clinical response and mucosal healing at one year. *Dig Liver Dis* 2014;46:974-979.
42. Popa IV, Burlacu A, Mihai C, Prelipcean CC. A machine learning model accurately predicts ulcerative colitis activity at one year in patients treated with anti-tumour necrosis factor α agents. *Medicina (Kaunas)* 2020;56:628.
43. Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. *Nat Rev Gastroenterol Hepatol* 2019;16:312-321.
44. Verstockt B, Sudahakar P, Creyns B, et al. DOP70 An integrated multi-omics biomarker predicting endoscopic response in ustekinumab treated patients with Crohn's disease. *J Crohns*

- Colitis 2019;13(Supplement_1):S072-S073.
45. Borren NZ, Plichta D, Joshi AD, et al. Multi-“-Omics” profiling in patients with quiescent inflammatory bowel disease identifies biomarkers predicting relapse. *Inflamm Bowel Dis* 2020; 26:1524-1532.
 46. Siegel CA, Horton H, Siegel LS, et al. A validated web-based tool to display individualised Crohn’s disease predicted outcomes based on clinical, serologic and genetic variables. *Aliment Pharmacol Ther* 2016;43:262-271.
 47. D’Haens G, Kelly O, Battat R, et al. Development and validation of a test to monitor endoscopic activity in patients with Crohn’s disease based on serum levels of proteins. *Gastroenterology* 2020;158:515-526.
 48. Sossenheimer PH, Yvellez OV, Andersen MJ, et al. 539 Wearable devices can predict disease activity in inflammatory bowel disease patients. *Gastroenterology* 2019;156:S-111.
 49. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019;68:1813-1819.
 50. Korbar B, Olofson AM, Miralflor AP, et al. Deep learning for classification of colorectal polyps on whole-slide images. *J Pathol Inform* 2017;8:30.
 51. Wei JW, Suriawinata AA, Vaickus LJ, et al. Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides. *JAMA Netw Open* 2020;3:e203398.
 52. Bédard A, Westerling-Bui T, Zuraw A. Proof of concept for a deep learning algorithm for identification and quantification of key microscopic features in the murine model of DSS-induced colitis. *Toxicol Pathol* 2021;49:897-904.
 53. Cheng JY, Abel JT, Balis UG, McClintock DS, Pantanowitz L. Challenges in the development, deployment, and regulation of artificial intelligence in anatomic pathology. *Am J Pathol* 2021;191:1684-1692.
 54. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199-2210.
 55. Saha M, Chakraborty C, Racoceanu D. Efficient deep learning model for mitosis detection using breast histopathology images. *Comput Med Imaging Graph* 2018;64:29-40.
 56. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with deep learning. *Sci Rep* 2018;8:4165.
 57. Raya-Povedano JL, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M. AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: a retrospective evaluation. *Radiology* 2021;300:57-65.
 58. de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-learning system detects neoplasia in patients with Barrett’s esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology* 2020;158:915-929.
 59. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol* 2019;20:1645-1654.
 60. Ebigbo A, Mendel R, Rückert T, et al. Endoscopic prediction of submucosal invasion in Barrett’s cancer with the use of artificial intelligence: a pilot study. *Endoscopy* 2021;53:878-883.
 61. Kudo SE, Misawa M, Mori Y, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. *Clin Gastroenterol Hepatol* 2020;18:1874-1881.
 62. U.S. Food and Drug Administration. FDA authorizes marketing of first device that uses artificial intelligence to help detect potential signs of colon cancer [Internet]. c2021 [cited 2021 Jul 23]. <https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-first-device-uses-artificial-intelligence-help-detect-potential-signs-colon>
 63. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* 2019;7:e7702.
 64. Sarwar S, Dent A, Faust K, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *NPJ Digit Med* 2019;2:28.
 65. Walradt T, Glissen Brown JR, Alagappan M, Lerner HP, Berzin TM. Regulatory considerations for artificial intelligence technologies in GI endoscopy. *Gastrointest Endosc* 2020;92:801-806.
 66. U.S. Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD): discussion paper and request for feedback [Internet]. c2021 [cited 2022 Dec 15]. <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>
 67. El Hajjar A, Rey JF. Artificial intelligence in gastrointestinal endoscopy: general overview. *Chin Med J (Engl)* 2020;133:326-334.
 68. Mori Y, Kudo SE, East JE, et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). *Gastrointest Endosc*

- 2020;92:905-911.
69. Chen MM, Golding LP, Nicola GN. Who will pay for AI? *Radiol Artif Intell* 2021;3:e210030.
70. Rex DK. Making a resect-and-discard strategy work for diminutive colorectal polyps: let's get real. *Endoscopy* 2022;54:364-366.
71. Zand A, Stokes Z, Sharma A, van Deen WK, Hommes D. Artificial intelligence for inflammatory bowel diseases (IBD); accurately predicting adverse outcomes using machine learning. *Dig Dis Sci* 2022;67:4874-4885.
72. Schoeb D, Suarez-Ibarrola R, Hein S, et al. Use of artificial intelligence for medical literature search: randomized controlled trial using the Hackathon Format. *Interact J Med Res* 2020; 9:e16606.