

Artificial Intelligence in Inflammatory Bowel Disease Endoscopy: Implications for Clinical Trials

Harris A. Ahmad,^a James E. East,^b Remo Panaccione,^{c, iD} Simon Travis,^{b, iD} James B. Canavan,^a Keith Usiskin,^a Michael F. Byrne^{d, e}

^aBristol Myers Squibb, Princeton, NJ, USA

^bTranslational Gastroenterology Unit, Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK

^cInflammatory Bowel Disease Clinic, University of Calgary, Calgary, AB, Canada

^dUniversity of British Columbia, Division of Gastroenterology, Department of Medicine, Vancouver, BC, Canada

^eSatisfai Health, Vancouver, BC, Canada

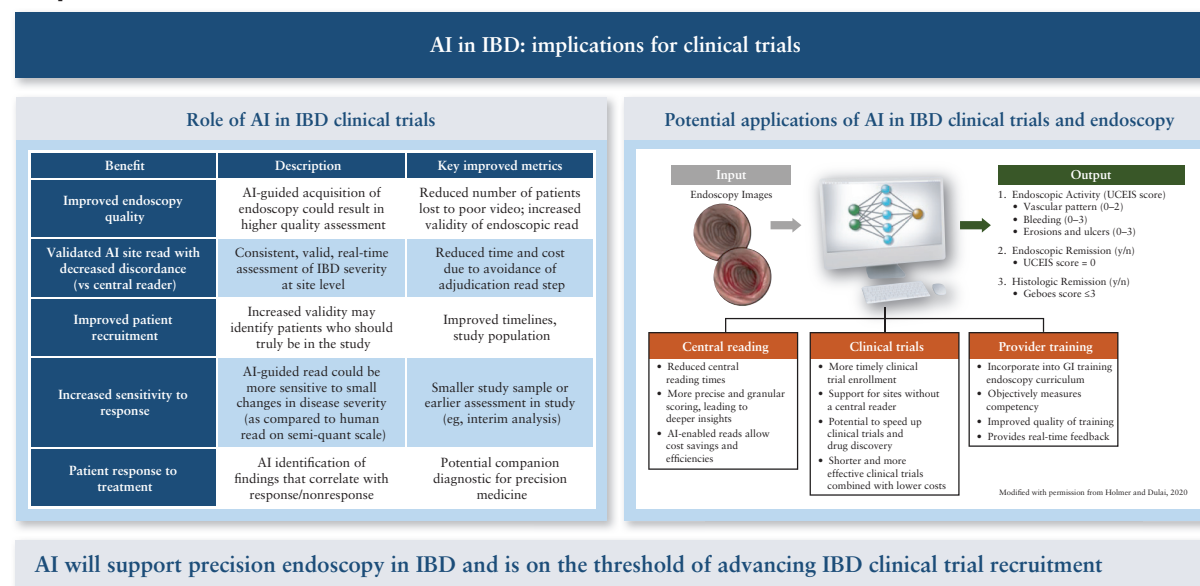
Corresponding author: Michael F. Byrne, MD, MRCP, FRCPC, University of British Columbia, Division of Gastroenterology, Department of Medicine, CEO and Founder, Satisfai Health, Vancouver, BC, Canada, Tel.: +1.604.512.1388; Fax: +1.604.628.2419; E-mail: Michael.byrne@vch.ca

Abstract

Artificial intelligence shows promise for clinical research in inflammatory bowel disease endoscopy. Accurate assessment of endoscopic activity is important in clinical practice and inflammatory bowel disease clinical trials. Emerging artificial intelligence technologies can increase efficiency and accuracy of assessing the baseline endoscopic appearance in patients with inflammatory bowel disease and the impact that therapeutic interventions may have on mucosal healing in both of these contexts. In this review, state-of-the-art endoscopic assessment of mucosal disease activity in inflammatory bowel disease clinical trials is described, covering the potential for artificial intelligence to transform the current paradigm, its limitations, and suggested next steps. Site-based artificial intelligence quality evaluation and inclusion of patients in clinical trials without the need for a central reader is proposed; for following patient progress, a second reading using AI alongside a central reader with expedited reading is proposed. Artificial intelligence will support precision endoscopy in inflammatory bowel disease and is on the threshold of advancing inflammatory bowel disease clinical trial recruitment.

Keywords: Endoscopy; imaging

Graphical Abstract



AI, artificial intelligence; GI, gastrointestinal; IBD, inflammatory bowel disease; UCEIS, Ulcerative Colitis Endoscopic Index of Severity.

H.A. Ahmad, et al. Artificial Intelligence in Inflammatory Bowel Disease Endoscopy: Implications for Clinical Trials



1. Introduction

Artificial intelligence [AI] refers to computer algorithms capable of learning, problem solving, and decision making [Table 1].^{1–3} Machine learning is a subfield of AI where algorithms are trained to perform tasks by recognising patterns from data without being explicitly programmed.⁴ Machine learning can evaluate large datasets and detect patterns to assess disease characteristics, such as severity or prognosis.² Clinical applications of AI have expanded across medical fields, including gastroenterology, radiology, pathology, and cardiology.⁵ Importantly, AI is expected to transform endoscopy and image interpretation.⁶

Deep learning, a subset of machine learning, uses multi-layered artificial neural networks to mimic the human brain and includes convolutional neural networks [CNNs] which are widely used in image and pattern recognition.^{5,7} Network interconnections allow algorithms to optimise classification during training by determining weights and adjusting for factors such as inherent biases or diversity.¹ Several AI algorithms have been applied to gastroenterology to support computer vision techniques, such as computer-aided detection [CADE] and computer-aided diagnosis [CADx].^{6,8,9} Machine learning algorithms have particular potential for scoring disease activity, refining endpoints, and recruiting patients for trials in inflammatory bowel disease [IBD].^{10–13} Current AI algorithms developed for IBD assessment, and their benefits and limitations in clinical trials, can be found in Table 2.^{14–17}

2. Current Status of Endoscopy in IBD Clinical Trials

Endoscopic mucosal healing is a therapeutic target for IBD¹⁸ since it is associated with lower rates of corticosteroid dependency, hospitalisation, and surgery¹⁹; however, endoscopy has inherent limitations in clinical trials [Table 3].^{10,20–23}

2.1. Patient recruitment

Endoscopic assessment is central for clinical trial patient selection, including enrolment, stratification, re-randomisation,

and open-label drug eligibility.¹⁰ However, patient recruitment is a significant challenge in IBD clinical trials, partly because physicians are focused on procedures rather than recruitment.²⁴ Interobserver variability and endoscopist inexperience may also lead to misvaluation of disease severity, resulting in inappropriate patient enrolment or incorrect treatment arm assignment.²⁰ Recruitment inefficiencies may result in the loss of participants from a relatively small pool of eligible patients, creating the need for larger cohorts and increasing clinical trial costs.^{10,21,22}

2.2. Local versus central reading

To overcome the subjective variability in endoscopic scoring, central reading of endoscopic videos has become commonplace in IBD clinical trials and has been extended to interpretation of histopathological samples.

Local readers tend to overscore the screening endoscopy and underscore the outcome endoscopy. A clinical trial in patients with ulcerative colitis [UC] found that data from local readers supported a marginal difference [30.0% vs 20.6%; $p = 0.069$] in clinical remission between mesalazine compared with placebo.²⁰ However, when endoscopic images were reviewed by a single central reader, remission rates were 29.0% versus 13.8% [$p = 0.011$] for mesalazine and placebo, respectively.²⁰ Independent assessment excluded 31% of enrolled patients who did not have sufficient endoscopic disease, highlighting the objectivity introduced by central review.²⁰ Similar trends in the objectivity of local readers relative to central readers have been reported in a clinical trial in patients with Crohn's disease [CD].²⁵

2.3. Endoscopy acquisition and quality

Although central reading decreases interobserver variability and adjudication mitigates subjectivity, these steps are costly.¹⁰ Machine learning might replace one, two, or all human central readers, resulting in decreased costs and more accurate and consistent reporting.¹⁰ Additionally, central reading

Table 1. Artificial intelligence terminology^{1–3}

Artificial intelligence	The field of computer science which concerns the theory and development of computers to perform tasks that usually require human intelligence, such as image classification, speech recognition, and decision making
Machine learning	A field of artificial intelligence that refers to the computers' ability to learn to make decisions or detect patterns [without explicitly being programmed] from data
Deep learning	Subfield of machine learning that exploits many layers of nonlinear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification using various neural network frameworks
Neural networks	Model of layers consisting of connected nodes broadly similar to neurons in a biological nervous system
Convolutional neural networks	Deep learning architecture that adaptively learns hierarchies of features through back-propagation and is used for detection and recognition tasks in images [eg, face recognition]
Computer-aided detection/diagnosis	Describes use of a computer algorithm to provide detection [CADE] or a diagnosis [CADx] of a specified object/region of interest
Supervised learning	The task of an algorithm learning a function that maps an input to an output based on provided example data
Unsupervised learning	The task of a machine learning algorithm to learn the underlying data structure of unlabelled example data—for example, finding commonalities—leading to insights and therefore a greater understanding of the example data
Classification	The process of predicting a class/subcategory of given data points from known example data
Support vector machine	A discriminative classifier that determines classes from a separating hyperplane; through the use of a kernel, support vector machines can be adapted to suit nonlinear problems

Adapted from Seyed Tabib *et al.* [2020]² and Pannala *et al.* [2020].³

Table 2. Examples of current AI algorithms for IBD assessment in clinical trials^{14–17}

AI algorithm	Benefits	Limitations
Bayesian additive regression trees	Can establish cause-effect relationship	May not accurately represent the true data generating distribution and therefore may misrepresent the relationship between variables
Gradient boosting machine	Can capture complex relationships between variables to predict events	Clinicians likely not familiar with this methodology
Clustering	Can discover patterns and structure in labelled and unlabelled datasets; unsupervised model	Clustering of clinical data can be hindered by missing variables; can be difficult to cluster multivariate and relatively short time series
Decision tree	Can classify treatment response and predict outcomes	Simplification errors may occur when measuring the benefit of treatment decisions on outcomes such as quality-adjusted life-years; performing a time-consuming analysis adequately in a busy clinical environment may be difficult; various factors in decision making cannot be accurately reflected in a decision tree
Neural network	Can help predict clinical outcomes or make a diagnosis	Difficult to interpret
Random forest	Can predict survival outcome	Not suitable to predict benefit for a specific treatment
Regression trees	Can define prognostic groups for patients due to simplicity and intuitive interpretation	Intrinsic limitations in predictive performance
Support vector machine	Can classify and predict high-dimensional data, including diagnosis, disease course, disease severity, disease subtypes, and medication adherence	Eliminates factors/parameters based on conditional relevance

AI, artificial intelligence; IBD inflammatory bowel disease.

Table 3. Challenges of IBD endoscopy reading in IBD studies^{10,20–23}

Challenge	Description	Implication
Limited local reader expertise	IBD endoscopy evaluation of disease severity varies greatly in expertise across global sites	Improper eligibility/efficacy read, central reader discordance, adjudication reading
Inconsistencies across local reads	Local reads can vary in assessment consistency even within the same site and patient examination	Improper eligibility/efficacy read, central reader discordance, adjudication reading
Poor endoscopy quality	Endoscopies can vary greatly in quality across global sites	Not readable endoscopy assessment, lost time in screening, excluded patient
Local vs central read discordance	Discordance on reads leads to greater costs, long turnaround times, delayed reads	Patient lost to being out of screening window, lost study budget

IBD, inflammatory bowel disease.

incurs a delay [typically 2 to 3 days], which could invalidate a patient’s eligibility.²⁶ The video is sent to a central laboratory for quality control; it is then edited and uploaded to the central reader, who assesses it [usually within 24 h] and returns the reading. Immediate, objective assessments would decrease the delay of central reading.²⁶

Central reading is a step forward, but is not the answer to improving the quality of the image and/or data capture. Technique, false interpretations, variability among readers, and missing data contribute to endoscopy quality.²³ Moreover, inadequate bowel preparation leaving debris that obscures video quality, and endoscope slipping [cinematography] causing blind spots, affect acquisition and quality.^{12,23} Imaging artefacts created by motion, bright pixel areas due to specularly or pixel saturation, or underexposure can limit assessment of underlying tissue.²⁷ More than 60% of

an endoscopy video frame and nearly 70% of an endoscopy video sequence can be corrupted by artefacts.²⁸ An ideal model would be a system that improves the quality of data capture, which in turn improves the performance of endoscopy and provides site-level reading.

2.4. Assessment and endpoints

Endoscopic remission in CD and UC and histological remission in UC correlate with improved outcomes in IBD, and both are primary or key secondary endpoints in clinical trials.¹⁹ Human evaluation of colonoscopy and biopsy interpretation is, however, subjective.^{11,29,30}

Scoring systems attempt to provide consistency but were developed using older technologies, often without item-response theory, and are subject to performance limitations.^{8,23} The first

endoscopic scores were designed to assess severity rather than extent of endoscopic activity in UC. The Baron Score uses a four-point scale based on severity of mucosal friability and bleeding.³¹ The Modified Mayo Endoscopic Score [MMES] is also used for UC but combines severity of the Mayo Endoscopic Subscore [MES] with extent of disease.³² The Ulcerative Colitis Endoscopic Index of Severity [UCEIS] is a validated system used to score vascular pattern, bleeding, and erosions/ulcers in the worst affected area.^{33,34} Unlike the UCEIS, the MES levels overlap, using descriptive terms that are not mutually exclusive, and neither index scores disease extent.

The Crohn's Disease Endoscopic Index of Severity [CDEIS] assesses ulceration on colonic segments and stenosis using a score from 0 to 44, with higher scores indicating increased severity.³⁵ The Simple Endoscopic Score for Crohn's Disease [SES-CD] evaluates four endoscopic variables [presence and size of ulcers, proportion of surface covered by ulcers, proportion of surface affected by disease, and severity of stenosis] in each of the five ileocolonic segments.³⁶ Endoscopists score the variables on a scale of 0 to 3.

Several endoscopic scoring systems have been validated in clinical studies, including the Modified Multiplier Simple Endoscopic Score for Crohn's disease [MM-SES-CD], Rutgeerts score in CD, and Paddington International virtual ChromoendoScopy ScOre [PiCaSSO] in UC. The MM-SES-CD assesses endoscopic severity to predict 1-year endoscopic remission in patients with CD who are on active therapy.³⁷ A specialised CD scoring system, the Rutgeerts score, is used for predicting recurrence of disease in patients undergoing ileo-colonic resection.³⁸ In a recent post hoc analysis, MM-SES-CD had similar performance to the Rutgeerts score for predicting subsequent clinical recurrence of postoperative CD.³⁹ In UC, the PiCaSSO score used virtual electronic chromoendoscopy to assess vascular and mucosal features of healing and demonstrated the highest correlation with histology compared with the MES and UCEIS.⁴⁰

The reliability of scoring instruments is measured by intraclass correlation coefficients [ICCs] where: 1 is perfect reliability; 0.9 to <1 indicates excellent reliability; 0.75 to 0.9 indicates good reliability; 0.5 to 0.75 indicates moderate reliability; and <0.5 indicates poor reliability.⁴¹ The interrater ICC for SES-CD, for example, lies between 0.6 and 0.8 and is heavily dependent upon the level of training [ICC 0.68 for untrained vs 0.93 for trained physicians].⁴²

Scoring systems are limited by reader subjectivity and incomplete visualisation of the mucosa, along with inadequate validation and complexity of the scoring instrument.^{23,43} AI can provide objective and consistent assessment of mucosal disease activity, translating into more accurate clinical trial data.

3. The Role of AI in IBD Clinical Trials

AI can improve patient recruitment, enhance endoscopy quality, provide a validated site read, increase sensitivity to response, and improve patient treatment response [Table 4].^{4,10-13,23,44-51} Advances in predictive modelling are expected to improve decision making in clinical research programmes and to streamline drug development pathways.⁴⁵ Machine learning techniques have been adopted in trial design to ensure consistent and objective assessment, including patient recruitment.

3.1. Patient recruitment

AI can help identify appropriate candidates for trial enrolment by matching electronic medical record [EMR] information and other patient data against selection criteria.^{4,44,45} Machine learning algorithms can make a real-time enrolment decision at the site level, with concordance similar to central reading.^{10,23,46} AI could identify patients meeting selection criteria during routine endoscopy who might otherwise not be captured, assuming that consent enables the use of the recorded video. Enhancement of patient cohort selection through AI can increase recruitment efficiency with a smaller, less heterogeneous sample size.^{4,10} In addition, predicting patient response to placebo could lead to increased confidence in patient selection decisions, with the potential for a synthetic control arm.^{21,52} With more reliable prediction of patient outcomes, AI could also support early 'go/no-go' decisions for drug development.

3.2. Rapid endoscopic results

Additional tools to expedite central reading are needed to reduce delays in clinical trials.^{10,11} AI-assisted assessment of disease activity is expected to decrease variability and minimise the need for second reader/adjudication.²³ With AI, endoscopic assessment would instantly be available at the local site to provide a score upon the site's submission to the

Table 4. Potential benefits of AI application in IBD trials^{4,10-13,23,44-51}

Benefit	Description	Key improved metrics
Improved endoscopy quality	AI-guided acquisition of endoscopy could result in higher quality assessment	Reduced number of patients lost to poor video; increased validity of endoscopic read
Validated AI site read with decreased discordance [vs central reader]	Consistent, valid, real-time assessment of IBD disease severity at site level	Reduced time and cost due to avoidance of adjudication read step
Improved patient recruitment	Increased validity may identify patients who should truly be in the study	Improved timelines, study population
Increased sensitivity to response	AI-guided read could be more sensitive to small changes in disease severity [as compared with human read on semi-quantitative scale]	Smaller study sample or earlier assessment in study [eg, interim analysis]
Patient response to treatment	AI identification of findings that correlate with response/nonresponse	Potential companion diagnostic for precision medicine

central laboratory. This would eliminate the central reading delay, which would be consigned to ‘over-reading’ rather than primary reading.

3.3. Cost savings

The estimated average cost for IBD clinical trials ranges from \$30 million for a pivotal clinical trial to \$55 million [US dollars] for phase I through phase IV trials.^{53,54} AI has the potential to reduce central reading cost, which accounts for a considerable portion of trial budgets. AI can reduce the cost of video equipment and digitalisation of histology slides necessary to perform offsite analysis.²⁶ One study estimated that AI-assisted optical biopsy for colon polyps would decrease the costs of colonoscopy by 10.9% or by \$85.2 million per year in the USA alone.⁵⁵ In the absence of a cost savings value in IBD, the savings seen for colon polyps may provide a perspective.

3.4. Improving endoscopic assessment with AI

Clinical trials have successfully used AI in IBD endoscopy, including CAdE [eg, polyp detection], CAdx [eg, polyp classification], and improvement [eg, scoring bowel preparation], demonstrating its ability to advance endoscopic quality while decreasing interobserver variability.^{12,47–51} An example of an AI-assisted endoscopy interface can be found in Figure 1. AI can outperform humans, does not get tired or impatient, and does not have a limited attention span; therefore, it is less likely to miss subtleties.^{2,8,10,13}

Endoscopic techniques for polyp detection and assessment of IBD differ. This matters for both CD and UC because biopsy bleeding, friability, or scope trauma may be scored as a

consequence of disease activity. AI algorithms can be trained to differentiate this type of bleeding from disease severity.

Examples of AI with the potential to improve endoscopic assessment of disease include Red Density, EndoBRAIN, and PICaSSO. An operator-independent, computer-based tool, Red Density can score disease activity in UC using a redness map and vascular pattern recognition.⁵⁶ This score had significant correlation with the histological scoring systems [Robarts histopathology index] and with MES and UCEIS endoscopic scores. Due to its high level of performance and algorithm structure, Red Density does not require as much information as the CNN and presents an important application of AI. Another example where AI has improved the assessment ability of endoscopy is the EndoBRAIN system, which has demonstrated the ability to detect high-grade dysplasia in patients with long-standing UC who subsequently underwent an endoscopic submucosal dissection.⁵⁶ Because diagnosis of colitis-associated colorectal cancer may be difficult due to inflammation-associated consequences on mucosal appearance, the use of EndoBRAIN could help less experienced endoscopists with identification of lesions. The PICaSSO is the first validated endoscopic score using the new generation of virtual chromoendoscopy endoscopes in UC. This score had a very good interobserver agreement in the pre-test and post-test evaluations that could reflect the full spectrum of mucosal and vascular changes, including mucosal healing in UC.⁵⁷

3.5. Quality of examination metrics

Automated quality of examination [QoE] metrics can improve endoscopy examination and provide real-time feedback.⁵⁸ For example, AI can alert the endoscopist if the withdrawal time

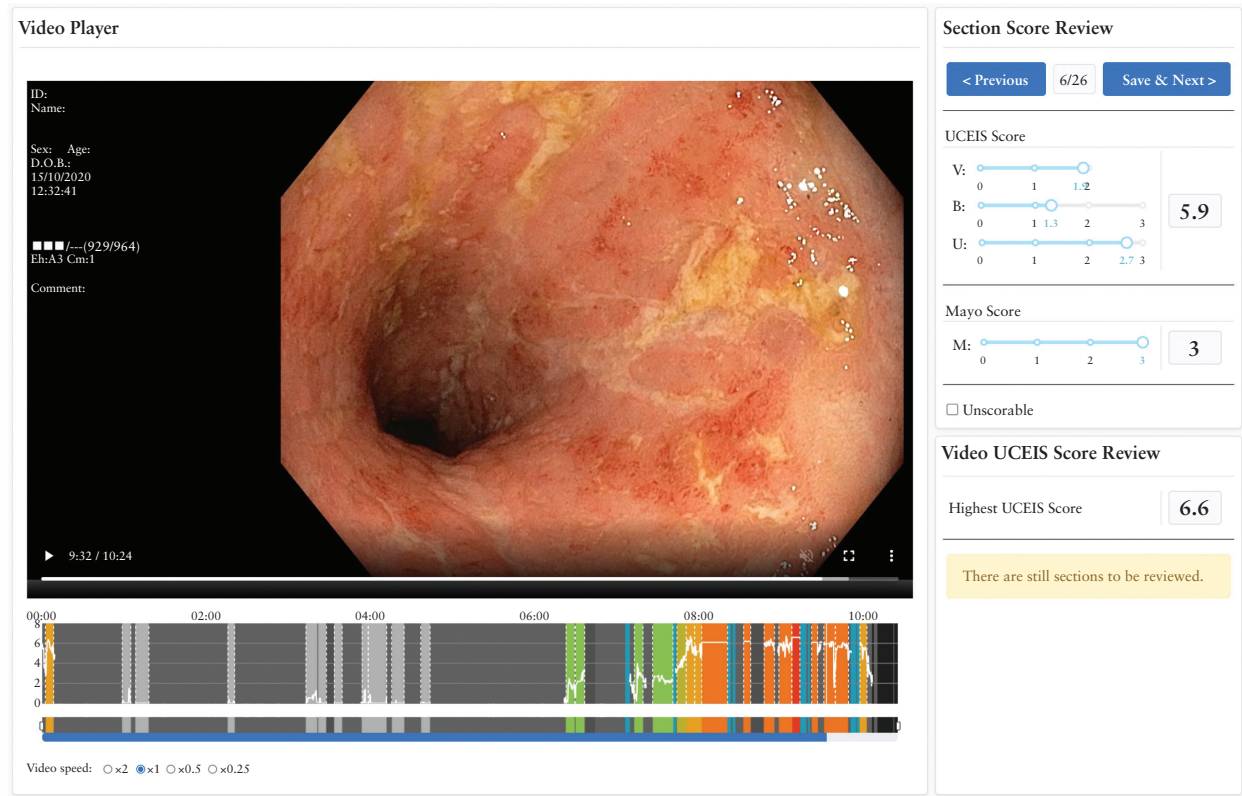


Figure 1. Example of AI-assisted endoscopy interface. Image provided by Dr Michael Byrne on behalf of Satisfai. AI, artificial intelligence; B, bleeding; U, ulceration; UCEIS, Ulcerative Colitis Endoscopic Index of Severity; V, vascular pattern.

[a quality metric for polyp detection] is below a predefined threshold.^{12,59,60} Meta-analysis of prospective trials found that AI-based polyp detection systems increased the detection of non-advanced adenomas and polyps, compared with standard colonoscopy.⁶⁰ Since IBD clinical trials exclude patients with neoplasia, AI could be useful for excluding ineligible patients.⁶¹

QoE metrics can be incorporated into machine learning algorithms designed to prevent the collection of poor-quality videos, by alerting the user and reducing the need for a patient to return for re-evaluation. For quality assurance, AI can report on the total percentage of colonic surface area visualised, bowel preparation, and resolution of the endoscopic image.^{12,58,62} This facilitates a thorough examination, which is in everyone's interests, including the patient's.

By way of example, AI helps real-time differentiation of adenomas from post-inflammatory polyps.⁹ A deep CNN applied to 125 consecutive colonoscopy videos was able to differentiate between hyperplastic polyps and adenomatous polyps, with an accuracy of 94%, a sensitivity of 98% [95% CI 92%–100%], and specificity of 83% [95% CI 67%–93%].

A real-time quality improvement system [WISENSE] was developed to monitor blind spots, record procedure time, and generate photodocumentation during 324 consecutive oesophagogastroduodenoscopies.⁴⁸ WISENSE had a 90% accuracy for monitoring blind spots and significantly decreased the rate of blind spots compared with the control group [5.9% in WISENSE group vs 22.5% in control group; $p < 0.001$]. A deep learning model has been shown to assess missed areas during colonoscopy, using depth and pose estimation, providing segment by segment coverage with 93% agreeing with the physician reviewer.⁶²

AI can restore corrupted data in endoscopic imaging. Using a dataset of 1290 endoscopy images, a CNN detector processed artefacts with indefinable shapes and generated a quality score for each video frame. The model had a mean average precision [mAP at 5% threshold] of up to 49.0 and a computational time as low as 88 ms, allowing real-time processing. The detector was also able to restore approximately 25% of the video frames to increase the overall frame retention rate to nearly 70%.²⁷

The European Society of Gastrointestinal Endoscopy recently developed the key performance indicators [KPIs] that should be part of and adopted in every IBD endoscopy unit.⁶³ Important KPIs are bowel preparation, photodocumentation, number of biopsies, standardised endoscopic scores, and detection rate of dysplasia associated with IBD. These quality metrics should also be incorporated in future clinical trials. AI may play a role in automating KPI metrics and improve the quality and ability of clinical trials to meet their objective.

3.6. Assessment of clinical trial endpoints

Image-based endpoint detection using machine learning capabilities has led to more reliable and efficient endpoint assessment.⁴ Deep learning algorithms can analyse large volumes of imaging data, enabling objective evaluation of endoscopy.²

Current scoring systems are limited by design—the UCEIS evaluates the worst segment of the lesion as opposed to integrating multiple areas, and SES-CD is affected by significant subjectivity.²³ Computer vision algorithms can provide cumulative quantification of erosions/ulcers, of the affected

area or normal mucosa, or of endoscopy quality.^{2,11,64} In an analysis of the mirikizumab phase 2 trial, the ability of a recurrent neural network [RNN] to predict central reader scores was compared with the UCEIS and MES scoring systems.¹⁰ A total of 795 full-length endoscopy videos from 249 patients was analysed by central readers and used to train the RNN. The study showed excellent agreement with human central reading scores, with an endoscopic healing accuracy of 97.0% and 95.5% for the UCEIS and MES, respectively.¹⁰

Recognising the comprehensive assessment by AI algorithms, some models exploit spectral characteristics and tissue colour to detect inflammation over a larger area compared with conventional scoring systems. For example, a model was trained to differentiate epithelial tissue of IBD and control patients from other tissue, as a first step, using Raman spectra as a second step to classify the sample as CD, UC, or healthy.⁶⁵ In a cross-sectional analysis of 38 patients [14 patients with CD, 13 patients with UC, and 11 healthy controls], Raman spectroscopy classified each group with 98.9% accuracy, 99.1% sensitivity, and 98.1% specificity for detecting healthy controls.⁶⁵ Furthermore, a trained neural network using Raman spectroscopy has been developed that can accurately differentiate mucosal healing from active inflammation in CD and UC.⁶⁶

Although AI can already match image interpretation by experienced gastroenterologists, the aim should be to exceed the abilities of skilled physicians.¹³ A real-time, operator-independent tool based on Red Density can now accurately identify inflammation and assess UC disease activity.¹¹ Red Density uses an algorithm built from 29 patients with UC and six healthy controls, based on the red channel of the red-green-blue pixel values and pattern recognition. The Red Density score significantly correlated with the Robarts histological index [$r = 0.65$, $p < 0.0001$], MES [$r = 0.61$, $p < 0.0001$], and UCEIS [$r = 0.56$, $p < 0.001$].¹¹ Another study used 8000 images to train and validate three different CNN models. The models distinguished eight classes of anatomical gastrointestinal landmarks and diseases, with accuracies approaching 99%.⁶⁷

Using improved endoscopic assessment tools to predict long-term clinical outcomes is a critically important role of AI in IBD. A study by Maeda and colleagues was the first that analysed the relationship between real-time AI-assisted colonoscopy outputs and the long-term prognosis of patients with UC. The findings showed that this fully automated AI system was able to assess the risk of clinical relapse in patients with UC in clinical remission, therefore enabling the clinicians to make real-time treatment decisions.⁶⁸

AI has the potential to become the gold standard for assessing disease severity.⁸ Systems are being developed to standardise scoring of difficult parameters, such as endoscopic healing.⁶⁴ AI-derived endoscopic assessment in clinical trials can be expected to lead to predictive scoring measures and to evolve into a machine that produces scores for the endoscopist related to outcomes, which may reduce the heterogeneity of treatment decisions.

3.7. Defining remission

Remission matters, and an accurate definition is an extension of improved disease activity scoring. One deep learning algorithm used a CNN-graded endoscopic severity rating in 3082 patients with UC to discriminate between disease remission [MES 0 or 1] and moderate to severe disease activity

[MES 2 or 3].¹³ Weighted kappa scores showed almost perfect agreement between the deep learning model and human reviewers in grading endoscopic severity [0.86, 95% CI 0.85–0.87].¹³ A study of 841 patients with UC was able to identify MES scores of 0 and 0 to 1 with area under the receiver operating characteristic curves [AUROCs] of 0.86 and 0.98, respectively, using a CNN-based CAD system for endoscopic severity.⁶⁹ Notably, the CNN performed better in the rectum than in the right and left colon for an MES score of 0 [AUROCs = 0.92, 0.83, and 0.83, respectively].⁶⁹ Using data from a single-centre retrospective cohort, a machine learning algorithm predicted remission using laboratory values and patient age in 1080 patients receiving thiopurine therapy.⁷⁰ The five most important predictor variables included haemoglobin, lymphocytes, haematocrit, neutrophils, and platelets. The algorithm differentiated remission from non-remission in the validation dataset, with an AUROC of 0.79, versus 0.49 using 6-thioguanine nucleotide metabolite levels.

Beyond the human eye, AI can explore new definitions of remission, such as quantifying vascular pattern, light reflex, or the pallor of normal mucosa. AI can also assist real-time histological evaluation. A deep neural network based on endoscopic images of UC [DNUC] was developed to predict histological remission.⁴⁶ For endoscopic remission, the DNUC was sensitive [93.3%] and specific [87.8%], with a diagnostic accuracy of 90.1%. For histological remission, the DNUC demonstrated 92.4% sensitivity, 93.5% specificity, and 92.9% diagnostic accuracy.⁴⁶ Since histological remission is associated with a better long-term outcome, detection in real time by the DNUC has immediate implications for clinical trials and practice.¹⁹

3.8. Integration of data

The potential to assimilate data sources from IBD datasets [including clinical symptoms, endoscopic read-outs, histopathology, gene expression values, and other outcomes] represents multiparametric data analysis that can provide further insight from clinical trials.^{10,45,52,71} Analysis of large genomic, transcriptomic, proteomic, and microbiomic [multiomic] datasets by machine learning could lead to the discovery of novel, clinically relevant biomarkers.⁷² Enormous opportunities to transform the IBD field lie at the intersection of multiomics, pathology, and endoscopy with AI solutions.⁷³

Indeed, multiomics potentially predicts IBD treatment outcomes.^{74,75} In the precision medicine era, AI could provide detailed insight into a patient's molecular profile and inform prognosis, disease aetiology, and/or therapeutic response.^{2,74} Faced with the challenge of unevenly sampled and sparse clinical time series data, a novel approach founded in extreme value theory [EVT] was deployed to convert these measurements into interpretable metrics of patient abnormality. Machine learning techniques and EVT methods were able to compare adalimumab and infliximab over several years in terms of relative effectiveness, predicting patient response and characterization of this response.⁷⁶

3.9. Training opportunities

AI can support the education and training of endoscopists. Experienced endoscopists already gain valuable knowledge from the feedback of central readers. Real-time feedback delivered by AI can serve as an extension of training and can bolster examination quality by guiding endoscopists as they

perform the procedure, assisted by an accurate tool that can provide on-the-job education and constructive feedback.²⁶

4. Limitations of AI in Clinical Trials

Datasets are selected and categorised by humans, which may bias AI algorithms.²⁶ Model accuracy depends on the degree to which endoscopists correctly provide scoring information.⁷ Current methods use supervised learning during which the algorithm is trained to make the same decision as physicians. Using unsupervised learning would identify clinically relevant patterns within data without ground truth information, which may be an effective strategy to avoid bias.⁷ Other approaches to improve ground truth measurements include agreement with a central reader IBD group using a Delphi procedure, correlation with histology/transcriptomic data, and correlation with longer-term clinical outcomes as ground truth. AI should decrease intra-observer variability and the assumption might be that the kappa for intra-observer variation in AI is 1.00, but this has yet to be examined and will be a determinant of the ability of AI to reduce variability. Models need to be trained on the differences between CD and UC, where the latter assessment of endoscopic findings and histology from mucosal biopsies might give a less complete picture [and ability to prognosticate outcomes] than the former, because of the transmural process. Additionally, AI models will need to be trained to ignore biopsy bleeding/friability; this is particularly true for UC assessment, because mucosal bleeding needs to be assessed ahead of the scope. That includes education of endoscopists regarding technique, to optimise views during scope insertion, in contrast to spotting polyps on withdrawal.

Endoscopic interpretation is potentially biased by clinical information, but that depends on the clinical context [discriminating ischaemic from UC, for example], although some scoring systems are not biased by clinical information.^{20,77} Rare clinical scenarios also challenge AI systems, since they have less representation within training datasets. An example is distinguishing Behcet's [more common in East Asia] from CD. Thus, high-quality datasets are needed to ensure geographical, technical, and patient demographic diversity.¹ The American Society for Gastrointestinal Endoscopy proposes a professionally managed image library,⁶ but the requirements to ensure correct diagnosis or ground truth for publicly available datasets are not always clear. Some datasets are clearly annotated [eg, the SUN Colonoscopy Video Database⁷⁸].

Pharmaceutical groups, by the nature of regulatory requirements, are likely to hold high-quality datasets that are specific to IBD trial populations and would be optimal for AI development. Collaboration with pharmaceutical groups [eg, endoscopic video resource from their anonymised trial videos] would complement the work of the Foundation for the National Institutes of Health [FNIH] Biomarkers Consortium on mucosal healing [sponsored by contributing Pharma [Bristol Myers Squibb, Lilly, Johnson & Johnson, Takeda]]. The large number of trial patients in endoscopic remission would provide potential videos and linked histology to assist in a deep dive into the definition of remission. EMR systems would provide a useful resource for algorithm development.⁷⁹

Standardisation of image capture is necessary to train algorithms to reflect clinical scenarios.^{6,26} Examination quality can be determined by automated analysis of endoscopy videos, facilitating the exclusion of poor-quality data.

The use of video capsule technology has been evolving in the field of CD. It has been applied to diagnosis and assessment of mucosal healing in the small bowel. However, limitations of this technology include the large amounts of data collected and consequent long duration of the analysis, both of which may be overcome with AI.⁵⁶ AI could enable selection of the frame or the section of video needed for the assessment, shortening the time for diagnosis and requiring a limited amount of data storage. However, obstacles remain in the development of AI for video capsule endoscopy [CE] that must be overcome in order for this technology to be implemented in the clinic. Current obstacles include: [a] use of retrospective data from single centres or small patient cohorts that restrict the generalisation of the established CNN systems and lead to lack of validation of the AI system; [b] use of single images, not the entire video, so that the analysis is not able to provide an overall evaluation of the validated scores for video capsule [eg, the Lewis score]; [c] uncertain performance of the CE in real-world practice due to potentially low quality of CE images; and [d] lack of use of CE data from various kinds of CE systems and diverse clinical situations.⁸⁰ Despite enthusiasm, problems with implementing AI in CD assessment remain and can be attributed to lack of education and knowledge among IBD providers, as well as hesitancy related to the potential of AI to replicate or replace expert clinical judgement.

Whereas early attempts to investigate AI for diagnosing dysplasia and identifying neoplastic lesions in IBD show promise,⁵⁶ the need for human assessment is likely to remain in the immediate future, even as algorithms advance in the use of inflammation assessment. Development of an AI algorithm for digital pathology that is capable of recognising and characterising dysplasia in IBD remains a challenge, as further improvements in diagnostic performance are needed.

Technological risks are inherent to AI due to the large amounts of data involved. Considerations regarding the nature of the data, patient privacy, cyber security, and potential roles of these algorithms are all of paramount importance in AI design. Many AI systems run as updatable software on a hardware platform. Updates are likely to be downloaded from the internet, making AI systems hackable or allowing unscrupulous actors to introduce errors that may reduce performance or even install ransomware. Commercial, malicious, or fictitious attacks on AI may invalidate a trial even if there are no patient safety risks.⁸¹ These are material concerns. Some facilities have technical limitations that present barriers to using AI tools [eg, hospital Wi-Fi networks may limit the ability to upload and assimilate data, although 5G may provide adequate bandwidth].

Clinical trial inclusion and exclusion criteria present barriers to enrolment. Sadly, some investigators may be tempted by financial conflicts of interest to override trial entry barriers for monetary gain.⁸² Central reading mitigates this possibility and AI may prevent such non-adherence. AI already supports existing models that use data to identify non-adherence and inappropriate subject enrolment.⁸²

IBD clinical trials typically cost between \$30 and \$55 million [US dollars],^{53,54} with a considerable portion of trial budgets spent on central reading costs. Whereas there is potential for AI to decrease costs of clinical trials in IBD, particularly with regard to central reading, the true cost savings remain unclear. The operational challenges of integrating AI technologies into existing setups may be a burden on sites that require additional hardware, personnel, or time. In addition, there has not yet been a prospective, multicentre, clinical trial

in IBD where the sponsor implemented AI reading at trial initiation. All studies using AI in IBD have been post hoc evaluations or single-centre, prospective studies. Full incorporation of AI in a global clinical trial will be an important step in promoting widespread use.

5. Regulatory Opportunities

FDA guidance for industry on UC clinical trial endpoints recommends that endoscopic assessment is made by both the endoscopist and the blinded central reader.⁸³ The FDA also recommends that the study protocol specifies how discrepancies between the assessments of the endoscopist and the central reader will be handled in the efficacy analysis. Charters that standardise the methodology underlying the scoring of endoscopic characteristics that may have subjective elements are particularly important. The FDA recommends the involvement of central reading for histological evaluations of biopsy specimens, including charters that standardise procedures and assessments. Machine learning techniques can increase consistency, objectivity, and accuracy of assessments, so their incorporation could help meet FDA recommendations.

The FDA regulatory framework distinguishes between technologies that ‘drive’ or that ‘inform’ clinical management.^{84,85} Within this framework, CADE and CADx are positioned as technologies to drive clinical management through predicting disease risk or aiding in diagnosis. Many AI applications have already received regulatory approval, including approaches for detecting atrial fibrillation, diagnosing diabetic retinopathy, interpreting magnetic resonance imaging, and diagnosing intracranial haemorrhage.⁸¹ Technologies with the ability to improve consistency in objective assessment are likely to be well received by regulatory agencies, in the same manner that central reading in IBD clinical trials was adopted because of its impact on consistency.²⁰

6. Future Directions

In the future, more prospective studies that will allow for better definition of the role of AI, implementation of AI into multicentre, global, prospective, IBD clinical trials, and development of optimal algorithms for both endoscopic and histological assessment of IBD in clinical trials will be needed. Up to now, almost all studies using AI in IBD have been post hoc evaluations or single-centre, prospective studies. One of the few examples of a multicentre, international, prospective study with a large cohort of patients, and with many AI endoscopy videos as well as histological slides, was the study in which PICaSSO endoscopy and histology AI were developed.⁴⁰ The AI was able to predict endoscopic inflammation/remission and long-term clinical outcomes in white-light endoscopy and virtual electronic chromoendoscopy. In a recent publication,⁸⁶ a new UC histological score that can be incorporated into an AI algorithm was developed, the PICaSSO Histologic Remission Index [PHRI]. This AI algorithm based on PHRI differentiated active from quiescent UC with high accuracy, sensitivity, and specificity; it also had the highest correlation with endoscopic activity and clinical outcomes. However, further validation of this approach, as well as the pairing of AI endoscopy with digital pathology in multicentre studies, is needed for the future direction for AI, as these improved assessments will be tied closely to important clinical patient outcomes.

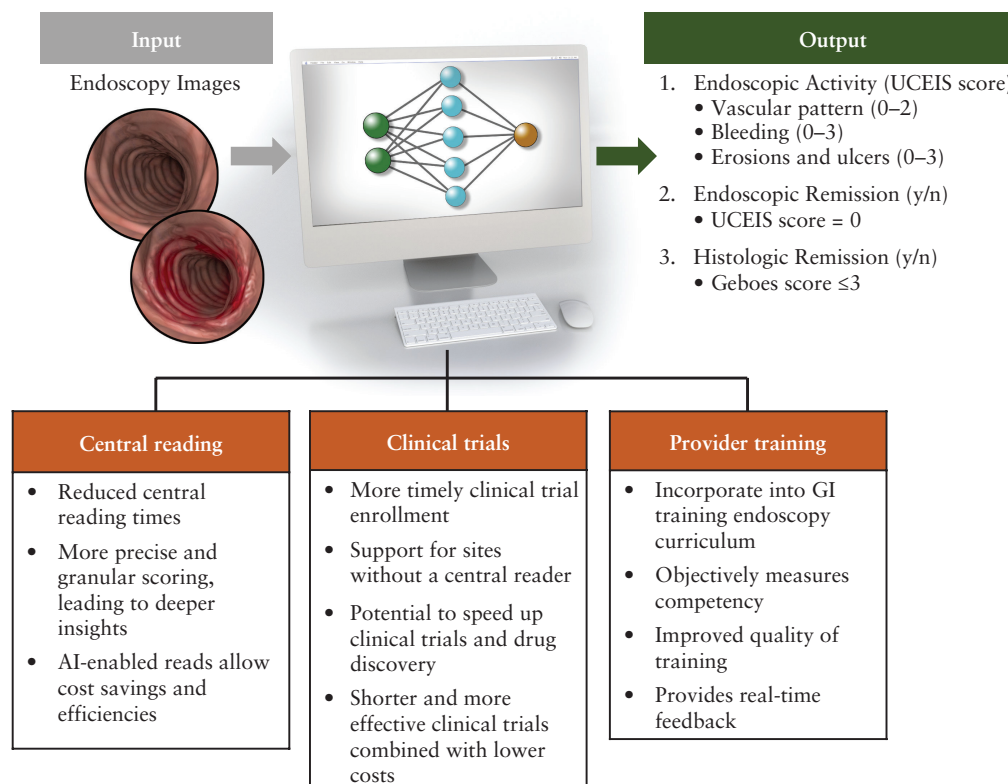


Figure 2. Potential applications of AI in inflammatory bowel disease clinical trials and endoscopy.²⁶ AI, artificial intelligence; GI, gastrointestinal; UCEIS, Ulcerative Colitis Endoscopic Index of Severity. Modified with permission from Holmer and Dulai, 2020.²⁶

AI has the potential to advance IBD clinical trials and support the quality of IBD endoscopy [Figure 2]. In the near future, regulatory applications will be filed to embed AI into the trial process at all levels, including assessment of primary endpoints [possibly as second or third readings]. By improving the quality of clinical assessment and allowing greater sensitivity between treatment groups, AI has the potential to decrease sample sizes and costs. Combining natural language processing of EMR data or endoscopy request forms will help pre-identify patients suspected of having IBD, with or without inflammation scoring, to facilitate [and possibly automate] trial enrolment. It would be possible to match eligible patients directly to study teams, including those at other sites, to expand the trial site footprint. With its ability to considerably enhance trial efficiency and reduce costs, AI technology is on the cusp of transforming IBD clinical trials.

Data Availability Statement

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

Funding

No funding or grant was received.

Conflict of Interest

HAA and JBC are employees of Bristol Myers Squibb. JEE reports personal fees from Boston Scientific, Falk, Lumendi, Paion, and Satisfai Health, outside the submitted work; in addition, JEE has a patent Methods and framework for as-

sessing image quality issued, and a patent Quantification of Barrett's oesophagus issued. RP 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, 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, Theravance Biopharma, Trellus Health, and UCB. ST has served as a paid consultant to AbbVie, Allergan, Amgen, Asahi, Bioclinica, Biogen, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, ChemoCentryx, Cosmo, Enterome, Equillium, Ferring, GSK, Genentech, Genzyme, Giuliani SpA, Immunocore, Immunometabolism, Janssen, Lilly, MSD, Merck, Mestag, 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, UKIERI, and Vifor; he has received honoraria from AbbVie, Amgen, Biogen, Ferring, Lilly, Pfizer, and Takeda; and he has had travel/accommodation expenses covered or reimbursed by AbbVie, Amgen, Biogen, Ferring, Lilly, Johnson & Johnson, Pfizer, and Takeda. KU was an employee of Bristol Myers Squibb at the time of manuscript initiation; he reports personal fees from Arena, Bristol Myers Squibb, Crinetics Pharmaceuticals, Insmed, and Locust Walk Capital. MFB is CEO and Founder of Satisfai Health.

Author Contributions

Concept or design: HAA, MFB, JEE, KU, JBC. Manuscript review and revisions: all authors. Final approval of manuscript: all authors. All authors confirm that they had full access to the underlying data and accept responsibility to submit for publication.

Acknowledgements

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. This manuscript, including related data, figures, and tables, has not been previously published and the manuscript is not under consideration elsewhere.

References

1. Le Berre C, Sandborn WJ, Aridhi S, *et al.* Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020;158:76–94.e2.e72.
2. Seyed Tabib NS, Madgwick M, Sudhakar P, Verstockt B, Korcsmaros T, Vermeire S. Big data in IBD: big progress for clinical practice. *Gut* 2020;69:1520–32.
3. Pannala R, Krishnan K, Melson J, *et al.* Artificial intelligence in gastrointestinal endoscopy. *VideoGIE* 2020;5:598–613.
4. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci* 2019;40:577–91.
5. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* 2019;7:e7702.
6. 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;92:951–9.
7. Ruffle JK, Farmer AD, Aziz Q. Artificial Intelligence-assisted gastroenterology: promises and pitfalls. *Am J Gastroenterol* 2019;114:422–8.
8. Bossuyt P, Vermeire S, Bisschops R. Scoring endoscopic disease activity in IBD: artificial intelligence sees more and better than we do. *Gut* 2020;69:788–9.
9. Byrne MF, Chapados N, Soudan F, *et al.* Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94–100.
10. Gottlieb K, Requa J, Karnes W, *et al.* Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology* 2021;160:710–9.e2.
11. 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;69:1778–86.
12. Gong D, Wu L, Zhang J, *et al.* Detection of colorectal adenomas with a real-time computer-aided system [ENDOANGEL]: a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020;5:352–61.
13. Stidham RW, Liu W, Bishu S, *et al.* Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open* 2019;2:e193963.
14. Stafford IS, Gosink MM, Mossotto E, Ennis S, Hauben M. A systematic review of artificial intelligence and machine learning applications to inflammatory bowel disease, with practical guidelines for interpretation. *Inflamm Bowel Dis* 2022;28:1573–83.
15. Ghimire K, Lai W, Omar Y, Schwebke T, Vo J. Machine learning approach to distinguish ulcerative colitis and Crohn's disease using SMOTE [Synthetic Minority Oversampling Technique] methods. *SMU Data Sci Rev* 2021;5:9.
16. Chen G, Shen J. Artificial intelligence enhances studies on inflammatory bowel disease. *Front Bioeng Biotechnol* 2021;9:635764.
17. Gubatan J, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: emerging technologies and future directions. *World J Gastroenterol* 2021;27:1920–35.
18. Turner D, Ricciuto A, Lewis A, *et al.* 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;160:1570–83.
19. Bryant RV, Burger DC, Delo J, *et al.* Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up. *Gut* 2016;65:408–14.
20. Feagan BG, Sandborn WJ, D'Haens G, *et al.* The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;145:149–57.e2.
21. Lee CS, Lee AY. How artificial intelligence can transform randomized controlled trials. *Transl Vis Sci Technol* 2020;9:9.
22. Reinisch W, Mishkin DS, Oh YS, *et al.* Impact of various central endoscopy reading models on treatment outcome in Crohn's disease using data from the randomized, controlled, exploratory cohort arm of the BERGAMOT trial. *Gastrointest Endosc* 2021;93:174–82.e2.
23. Gottlieb K, Daperno M, Usiskin K, *et al.* Endoscopy and central reading in inflammatory bowel disease clinical trials: achievements, challenges and future developments. *Gut* 2021;70:418–26.
24. Dubinsky MC, Collins R, Abreu MT; International Organization for the Study of Inflammatory Bowel Diseases [IOIBD]. Challenges and opportunities in IBD clinical trial design. *Gastroenterology* 2021;161:400–4.
25. Rutgeerts P, Reinisch W, Colombel JF, *et al.* Agreement of site and central readings of ileocolonoscopy scores in Crohn's disease: comparison using data from the EXTEND trial. *Gastrointest Endosc* 2016;83:188–97.e1.e181–183.
26. Holmer AK, Dulai PS. Using artificial intelligence to identify patients with ulcerative colitis in endoscopic and histologic remission. *Gastroenterology* 2020;158:2045–7.
27. Ali S, Zhou F, Bailey A, *et al.* A deep learning framework for quality assessment and restoration in video endoscopy. *Med Image Anal* 2020;68:101900.
28. Ali S, Zhou F, Braden B, *et al.* An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Sci Rep* 2020;10:2748.
29. Peyrin-Biroulet L, Sandborn W, Sands BE, *et al.* Selecting therapeutic targets in inflammatory bowel disease [STRIDE]: Determining therapeutic goals for treat-to-target. *Am J Gastroenterol* 2015;110:1324–38.
30. Cushing KC, Ananthakrishnan AN. Editorial: histologic normalisation in ulcerative colitis. Authors' reply. *Aliment Pharmacol Ther* 2020;51:401.
31. Baron JH, Connell AM, Lennard-Jones JE. Variation between observers in describing mucosal appearances in proctocolitis. *Br Med J* 1964;1:89–92.
32. Lobatón T, Bessissow T, De Hertogh G, *et al.* The Modified Mayo Endoscopic Score [MMES]: A new index for the assessment of extension and severity of endoscopic activity in ulcerative colitis patients. *J Crohns Colitis* 2015;9:846–52.
33. Travis SP, Schnell D, Krzeski P, *et al.* Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity [UCEIS]. *Gut* 2012;61:535–42.
34. Travis SP, Schnell D, Krzeski P, *et al.* Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology* 2013;145:987–95.
35. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective

- multicentre study. Groupe d'Etudes Thérapeutiques des Affections Inflammatoires du Tube Digestif [GETAID]. *Gut* 1989;30:983–9.
36. Koutroumpakis E, Katsanos KH. Implementation of the simple endoscopic activity score in crohn's disease. *Saudi J Gastroenterol* 2016;22:183–91.
 37. Narula N, Wong ECL, Colombel JF, et al. Predicting endoscopic remission in Crohn's disease by the modified multiplier SES-CD [MM-SES-CD]. *Gut* 2022;71:1078–87.
 38. Parigi TL, Mastroiocco E, Da Rio L, et al. Evolution and new horizons of endoscopy in inflammatory bowel diseases. *J Clin Med* 2022;11:872.
 39. Narula N, Wong ECL, Dulai PS, Marshall JK, Jairath V, Reinisch W. The performance of the Rutgeerts score, SES-CD, and MM-SES-CD for prediction of postoperative clinical recurrence in Crohn's disease. *Inflamm Bowel Dis* 2022;Jun 28. doi: 10.1093/ibd/izac130. Online ahead of print.
 40. Iacucci M, Smith SCL, Bazarova A, et al. An international multicenter real-life prospective study of electronic chromoendoscopy score PI-CaSSO in ulcerative colitis. *Gastroenterology* 2021;160:1558–69. e8.e1558.
 41. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
 42. Hart L, Bessissow T. Endoscopic scoring systems for the evaluation and monitoring of disease activity in Crohn's disease. *Best Pract Res Clin Gastroenterol* 2019;38–39:101616.
 43. Mohammed Vashist N, Samaan M, Mosli MH, et al. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database Syst Rev* 2018;1:Cd011450.
 44. Archer M, Germain S. The integration of artificial intelligence in drug discovery and development. *Int J Digital Health* 2021;1:5.
 45. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3.
 46. 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;158:2150–7.
 47. 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;68:1813–9.
 48. 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;68:2161–9.
 49. Chen D, Wu L, Li Y, et al. Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: a prospective, single-blind, 3-parallel-group, randomized, single-center trial. *Gastrointest Endosc* 2020;91:332–9.e3.e333.
 50. Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy [CADE-DB trial]: a double-blind randomised study. *Lancet Gastroenterol Hepatol* 2020;5:343–51.
 51. Liu WN, Zhang YY, Bian XQ, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol* 2020;26:13–9.
 52. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther* 2018;47:763–72.
 53. Sertkaya A, Birkenbach A, Berling A, Eyraud J. Examination of Clinical Trial Costs and Barriers for Drug Development. <https://aspe.hhs.gov/report/examination-clinical-trial-costs-and-barriers-drug-development> Accessed December 15, 2022.
 54. Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated costs of pivotal trials for novel therapeutic agents approved by the US Food and Drug Administration, 2015–2016. *JAMA Intern Med* 2018;178:1451–7.
 55. 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;92:905–11.e1.
 56. Solitano V, Zilli A, Franchellucci G, et al. Artificial endoscopy and inflammatory bowel disease: Welcome to the future. *J Clin Med* 2022;11:569.
 57. Iacucci M, Daperno M, Lazarev M, et al. Development and reliability of the new endoscopic virtual chromoendoscopy score: the PICaSSO [Paddington International Virtual ChromoendoScopy ScOre] in ulcerative colitis. *Gastrointest Endosc* 2017;86:1118–27. e5.
 58. 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;158:1219–21.e2.
 59. Shaikat A, Rector TS, Church TR, et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. *Gastroenterology* 2015;149:952–7.
 60. Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* 2021;53:277–84.
 61. Maeda Y, Kudo SE, Ogata N, et al. Can artificial intelligence help to detect dysplasia in patients with ulcerative colitis? *Endoscopy* 2021;53:E273–4.
 62. Freedman D, Blau Y, Katzir L, et al. Detecting deficient coverage in colonoscopies. *IEEE Trans Med Imaging* 2020;39:3451–62.
 63. 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;54:904–15.
 64. Nakase H, Hirano T, Wagatsuma K, et al. Artificial intelligence-assisted endoscopy changes the definition of mucosal healing in ulcerative colitis. *Dig Endosc* 2021;33:903–11.
 65. 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.
 66. Smith SCL, Banbury C, Zardo D, et al. Raman spectroscopy accurately differentiates mucosal healing from non-healing and biochemical changes following biological therapy in inflammatory bowel disease. *PLoS One* 2021;16:e0252210.
 67. Cogan T, Cogan M, Tamil L, MAPGI. Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. *Comput Biol Med* 2019;111:103351.
 68. Maeda Y, Kudo SE, Ogata N, et al. Evaluation in real-time use of artificial intelligence during colonoscopy to predict relapse of ulcerative colitis: a prospective study. *Gastrointest Endosc* 2022;95:747–56.e2.
 69. Ozawa T, Ishihara S, Fujishiro M, et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019;89:416–21.e1.
 70. Waljee AK, Sauder K, Patel A, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J Crohns Colitis* 2017;11:801–10.
 71. Johnson T, Steere B, Zhang P, et al. Mirikizumab-induced transcriptome changes in patient biopsies at Week 12 are maintained through Week 52 in patients with Ulcerative Colitis [abstract DOP09]. *J Crohns Colitis* 2021;15:S047–8.
 72. Friedrich M, Pohin M, Jackson MA, et al. IL-1-driven stromal-neutrophil interaction in deep ulcers defines a pathotype of therapy non-responsive inflammatory bowel disease. *Nat Med* 2021;27:1970–81.
 73. Fiocchi C, Iliopoulos D. What's new in IBD therapy: An 'omics network' approach. *Pharmacol Res* 2020;159:104886.
 74. Zarringhalam K, Enayattallah A, Reddy P, Ziemek D. Robust clinical outcome prediction based on Bayesian analysis of transcriptional profiles and prior causal networks. *Bioinformatics* 2014;30:i69–77.
 75. Douglas GM, Hansen R, Jones CMA, et al. Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. *Microbiome* 2018;6:13.

76. Niehaus K. *Phenotypic modelling of Crohn's disease severity: a machine learning approach*. PhD thesis. Department of Engineering Science, Trinity College, University of Oxford, 2016.
77. Travis SP, Schnell D, Feagan BG, *et al*. The impact of clinical information on the assessment of endoscopic activity: characteristics of the Ulcerative Colitis Endoscopic Index Of Severity [UCEIS]. *J Crohns Colitis* 2015;9:607–16.
78. Itoh H, Misawa M, Mori Y, Oda M, Kudo S-E, Mori K. *SUN Colonoscopy Video Database*. <http://amed8k.sundatabase.org/> Accessed December 15, 2022.
79. Ananthakrishnan AN, Cai T, Savova G, *et al*. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis* 2013;19:1411–20.
80. Yang YJ. The future of capsule endoscopy: The role of artificial intelligence and other technical advancements. *Clin Endosc* 2020;53:387–94.
81. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
82. Shiovitz TM, Bain EE, McCann DJ, *et al*. Mitigating the effects of nonadherence in clinical trials. *J Clin Pharmacol* 2016;56:1151–64.
83. U.S. Food and Drug Administration. Ulcerative Colitis: Clinical Trial Endpoints Guidance for Industry. <https://www.federalregister.gov/documents/2016/08/08/2016-18716/ulcerative-colitis-clinical-trial-endpoints-draft-guidance-for-industry-availability#:~:text=The%20Food%20and%20Drug%20Administration%20%28FDA%20or%20Agency%29,ulcerative%20colitis%20%28UC%29%20in%20adult%20and%20pediatric%20patients> Accessed December 15, 2022.
84. 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. <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf> Accessed December 15, 2022.
85. Walradt T, Glissen Brown JR, Alagappan M, Lerner HP, Berzin TM. Regulatory considerations for artificial intelligence technologies in GI endoscopy. *Gastrointest Endosc* 2020;92:801–6.
86. Gui X, Bazarova A, Del Amor R, *et al*. PICaSSO Histologic Remission Index [PHRI] in ulcerative colitis: development of a novel simplified histological score for monitoring mucosal healing and predicting clinical outcomes and its applicability in an artificial intelligence system. *Gut* 2022;71:889–98.