

Artificial intelligence in gastrointestinal endoscopy for inflammatory bowel disease: a systematic review and new horizons

Gian Eugenio Tontini¹, Alessandro Rimondi, Marta Venero, Helmut Neumann, Maurizio Vecchi, Cristina Bezzio* and Flaminia Cavallaro*

Abstract

Introduction: Since the advent of artificial intelligence (AI) in clinical studies, luminal gastrointestinal endoscopy has made great progress, especially in the detection and characterization of neoplastic and preneoplastic lesions. Several studies have recently shown the potential of AI-driven endoscopy for the investigation of inflammatory bowel disease (IBD). This systematic review provides an overview of the current position and future potential of AI in IBD endoscopy.

Methods: A systematic search was carried out in PubMed and Scopus up to 2 December 2020 using the following search terms: artificial intelligence, machine learning, computer-aided, inflammatory bowel disease, ulcerative colitis (UC), Crohn's disease (CD). All studies on human digestive endoscopy were included. A qualitative analysis and a narrative description were performed for each selected record according to the Joanna Briggs Institute methodologies and the PRISMA statement.

Results: Of 398 identified records, 18 were ultimately included. Two-thirds of these (12/18) were published in 2020 and most were cross-sectional studies (15/18). No relevant bias at the study level was reported, although the risk of publication bias across studies cannot be ruled out at this early stage. Eleven records dealt with UC, five with CD and two with both. Most of the AI systems involved convolutional neural network, random forest and deep neural network architecture. Most studies focused on capsule endoscopy readings in CD ($n=5$) and on the AI-assisted assessment of mucosal activity in UC ($n=10$) for automated endoscopic scoring or real-time prediction of histological disease.

Discussion: AI-assisted endoscopy in IBD is a rapidly evolving research field with promising technical results and additional benefits when tested in an experimental clinical scenario. External validation studies being conducted in large and prospective cohorts in real-life clinical scenarios will help confirm the added value of AI in assessing UC mucosal activity and in CD capsule reading.

Ther Adv Gastroenterol

2021, Vol. 14: 1–12

DOI: 10.1177/
17562848211017730

© The Author(s), 2021.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Alessandro Rimondi

Department of
Pathophysiology and
Organ Transplantation,
Università degli Studi di
Milano, Via Francesco
Sforza 35, Milano 20122,
Italy

Gastroenterology
and Endoscopy Unit,
Fondazione IRCCS Ca'
Granda Ospedale Maggiore
Policlinico, Milan, Italy
alessandro.rimondi5@gmail.com

Gian Eugenio Tontini

Gastroenterology
and Endoscopy Unit,
Fondazione IRCCS Ca'
Granda Ospedale Maggiore
Policlinico, Milan, Italy

Department of
Pathophysiology and
Transplantation, University
of Milan, Milan, Italy

Marta Venero

Gastroenterology Unit, Rho
Hospital, ASST Rhodense,
Milan, Italy

Helmut Neumann

Department of
Interdisciplinary
Endoscopy, University
Hospital Mainz, Mainz,
Germany

Maurizio Vecchi

Gastroenterology
and Endoscopy Unit,
Fondazione IRCCS Ca'
Granda Ospedale Maggiore
Policlinico, Milan, Italy

Department of
Pathophysiology and
Transplantation, University
of Milan, Milan, Italy

Cristina Bezzio

Gastroenterology Unit, Rho
Hospital, ASST Rhodense,
Milan, Italy

Flaminia Cavallaro

Gastroenterology
and Endoscopy Unit,
Fondazione IRCCS Ca'
Granda Ospedale Maggiore
Policlinico, Milan, Italy

*These authors
contributed equally.

Plain language summary

Artificial intelligence for inflammatory bowel disease endoscopy

- Artificial intelligence (AI) is a promising technology in many areas of medicine. In recent years, AI-assisted endoscopy has been introduced into several research fields, including inflammatory bowel disease (IBD) endoscopy, with promising applications that have the potential to revolutionize clinical practice and gastrointestinal endoscopy.

- We have performed the first systematic review of AI and its application in the field of IBD and endoscopy.
- A formal process of paper selection and analysis resulted in the assessment of 18 records. Most of these (12/18) were published in 2020 and were cross-sectional studies (15/18). No relevant biases were reported. All studies showed positive results concerning the novel technology evaluated, so the risk of publication bias cannot be ruled out at this early stage.
- Eleven records dealt with UC, five with CD and two with both. Most studies focused on capsule endoscopy reading in CD patients ($n=5$) and on AI-assisted assessment of mucosal activity in UC patients ($n=10$) for automated endoscopic scoring and real-time prediction of histological disease.
- We found that AI-assisted endoscopy in IBD is a rapidly growing research field. All studies indicated promising technical results. When tested in an experimental clinical scenario, AI-assisted endoscopy showed it could potentially improve the management of patients with IBD.
- Confirmatory evidence from real-life clinical scenarios should be obtained to verify the added value of AI-assisted IBD endoscopy in assessing UC mucosal activity and in CD capsule reading.

Keywords: artificial intelligence, computer-aided diagnosis, Crohn's disease, endoscopy, inflammatory bowel disease, machine learning, ulcerative colitis

Received: 7 February 2021; revised manuscript accepted: 26 April 2021.

Introduction

Recent advances in artificial intelligence (AI) are driving important transformations in medicine, including digestive endoscopy. AI can assess a huge amount of morphometric data in real time, revealing details that are often overlooked by clinicians and thus providing more precise and objective endoscopic diagnosis.¹ In addition, computer-aided diagnosis (CAD) systems can also help improve the performance of endoscopists by acting as quality controllers or training vectors. Within the next few years, AI is widely expected to become a new reference standard for improved diagnostic performance during lesion detection, characterization and classification in luminal gastrointestinal endoscopy (e.g. when investigating colorectal polyps, chronic gastritis and early gastric cancer) and in capsule enteroscopy.²

AI has recently been applied to the investigation of inflammatory bowel disease (IBD), in which luminal endoscopy plays a pivotal role in diagnosis, assessment of disease activity, extent or complications, and therapeutic decision making.³⁻⁵

The aim of our research was to review the implications of the use of AI in IBD endoscopy. No area of interest was excluded. Any reports that included endoscopic evaluation, endoscopy reporting, and clinical decisions linked to endoscopy evaluation were considered in our review. Within this systematic review, we provide an overview of the emerging applications and future potential of AI systems in the field of IBD endoscopy.

Methods

This systematic review was performed according to the Joanna Briggs Institute methodologies⁶ and the PRISMA statement.⁷ A systematic search was done in PubMed (US National Library of Medicine National Institutes of Health) and Scopus up to 2 December 2020 by one researcher (GET) using the following search terms: artificial intelligence, machine learning, computer-aided, inflammatory bowel disease, ulcerative colitis, Crohn's. All study types providing original data were included (i.e. prospective and retrospective

observational studies, case-control or cohort studies, case reports, case series reports, and population-based and experimental studies), except for abstracts and conference papers published only in a short format. The search was limited to studies conducted in human subjects with an established diagnosis of IBD and focusing on gastrointestinal endoscopy. No restrictions were placed on language. To identify additional relevant missed publications, a manual search was conducted in the reference lists of all included studies, as well as of reviews, consensus statements or guidelines published in 2015–2020. Two independent researchers (CB and FC) separately performed the screening search, excluding irrelevant studies and extracting all relevant data from each eligible study on specifically designed spreadsheets. All potentially eligible articles were further analysed by two independent researchers (CB and FC) who reviewed the full text based on the Joanna Briggs Institute criteria and appraisal tool for cohort studies (<https://jbi.global/critical-appraisal-tools>), and assessed the methodological study quality, the risk of bias of individual studies at the study level, and the risk of bias that may affect the cumulative evidence. Disagreements between investigators were resolved through discussion. A third independent researcher (GET) was consulted to solve any residual disagreement during either the screening or the full text assessment phase. Finally, one investigator (GET) provided the final list of eligible and excluded records and the PRISMA diagram.

Qualitative results are presented following a narrative format. Quantitative analyses were not performed according to sample size and heterogeneity of study aims and extracted data.

Results

A total of 398 records were identified after duplicate removal. Of these, 371 were excluded during the screening phase, and another nine were excluded after full text assessment by the independent pair of researchers (Figure 1). Six additional studies were identified through other sources (i.e. manual searching in the reference lists of all included studies and of published reviews, consensus statements and guidelines). Eighteen records were finally included. Two-thirds of these were published in 2020, while the oldest record was published in 2003. The included records were cross-sectional studies (15/18) with

a prospective ($n=7$) or retrospective ($n=8$) image collection, a retrospective cohort study ($n=1$), a meta-analysis of cross-sectional studies ($n=1$) and a single case report ($n=1$). Eleven records dealt with ulcerative colitis (UC), five with Crohn's disease (CD) and two with both UC and CD patients; six were from Asia (five from Japan), six from Europe, three from the USA, and three were international.

Among the cohort studies, the methodological study quality was judged as moderate to very high and the risk of relevant bias at the study level was not reported (the data extraction sheet is available in Supplemental file 1). Also, reporting biases were not identified across selected records. However, all selected records reported positive results, thereby implying a potential risk of publication bias across studies.

Mucosal activity

Precise and reproducible IBD mucosal activity assessment is crucial in the era of treat-to-target IBD management.^{8,9} Endoscopic and histological healing is a main clinical and research aim in the treatment of IBD, especially for UC patients.¹⁰ AI was first used in 2003 in the field of IBD to assess endoscopic severity in patients with UC. In this pioneering work, Sasaki *et al.*¹¹ characterized the Matts score for grading endoscopic severity using pictorial parameters of mucosal redness derived by gray scale analysis from 133 digital colonoscopy fixed images of 55 UC patients. Starting from the assumption that mucosal redness is proportional to histological microvascular bed area, the visual parameters changing along with disease severity will reflect enlarged microvessels and a more heterogeneous spatial distribution of microvessels. Consistently, the degree of mucosal redness was quantified as an index of hemoglobin through a Bayesian-driven CAD algorithm. This algorithm was able to differentiate the Matts grades based on the kurtosis of the index of hemoglobin with high sensitivity and specificity when discriminating Matts 1 from Matts 2, Matts 2 from Matts 3, and Matts 3 from Matts 4 (84% and 96%, 94% and 70%, and 100% and 85%, respectively).¹¹

More recently, Ozawa *et al.*¹² first trained and then evaluated a novel convolutional neural network (CNN) assessing the Mayo endoscopic subscore (MES) in 3981 standard endoscopic fixed

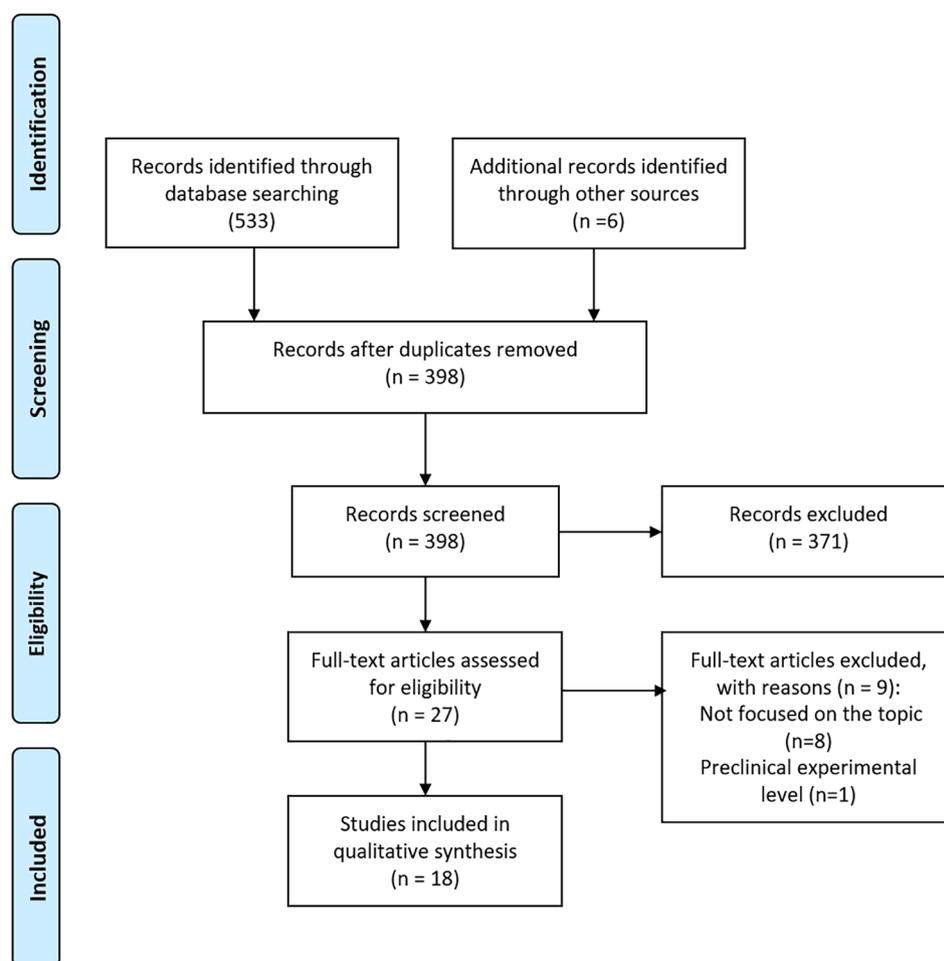


Figure 1. PRISMA flow diagram.

images from 114 UC patients. The system showed a high level of performance with areas under the receiver operating characteristic curve (AUROCs) of 0.86 and 0.98 to identify Mayo 0 and 0–1, using expert endoscopists as the reference standard for the ‘true Mayo endoscopic score’.¹²

In other research from Japan, Takenaka *et al.*¹³ trained and validated a deep neural network (DNN) to assess both endoscopic and histopathological disease activity using the ulcerative colitis endoscopic index of severity score (UCEIS; 40,758 colonoscopy images) and the Geboes score of histology (20,655 histological images). Results from the real-life validation phase (875 patients, 4187 colonoscopy images and 4104 biopsy specimens) showed 90% and 93% accuracy for endoscopic and histological remission, respectively; the intraclass correlation coefficients between the DNN and expert endoscopists and

pathologists were also remarkable (0.917 and 0.859, respectively).¹³ These findings have recently been supported by the same group in a follow-up study of 875 UC patients (median follow-up of 20 months), showing for the first time ever that CAD-driven endoscopic assessment of UC could predict the patient’s prognosis.¹⁴

Recently, the first operator-independent score for endoscopy in IBD patients has been developed (29 UC and six healthy patients), tested and partially validated (10 UC patients) in a multicenter study from Belgium and Japan.¹⁵ The ‘red density’ score is calculated by a computer algorithm based on the red channel of the red–green–blue pixel values and pattern recognition from endoscopic images using a high-definition prototype endoscope with white-light illumination delivered by a 300W xenon lamp (Pentax Medical, HOYA Corporation, Tokyo, Japan). The red density

imaging can be activated on a standard endoscopic monitor on demand and in real time. Based on preliminary results, the red density score significantly correlates with both endoscopic (MES and UCEIS) and histopathological disease activity (Robarts histological index). Larger, prospective studies are ongoing to corroborate these findings and explore whether this AI could integrate (or even replace) the standard human-driven endoscopic and pathological assessment of mucosal activity in an objective way.

Interestingly, AI-driven endoscopic assessment has also been adopted to overcome human subjectivity during the on-site and off-site revision of digital records such as fixed images or videos in both research and routine practice. In 2019, Stidham *et al.*¹⁶ explored this field for the first time with a CNN constructed as a deep learning model which was trained on and categorized 16,514 images (from 3082 UC patients) into two clinically relevant groups: endoscopic remission (MES 0 or 1) and moderate to severe disease (MES 2 or 3). A set of 30 additional full-motion colonoscopy videos was used for external validation to mimic real-world application. The CNN was excellent for distinguishing MES 0–1 from MES 2–3 (AUROC 0.966). Weighted κ agreement between the CNN and the adjudicated reference score was also good for identifying exact MES subscores ($\kappa=0.84$) and was similar to the agreement between experienced reviewers ($\kappa=0.86$).

Endoscopic AI diagnosis has also been used in two later randomized controlled trials to overcome the inter-observer variability affecting the off-site reading of pre and post-treatment endoscopic videos. In a branch study of a phase II multicenter, randomized, double-blind, parallel placebo controlled study of mirikizumab, Gottlieb *et al.*¹⁷ developed a CNN system to assess mucosal activity according to MES and UCEIS. A total of 795 prospectively recorded full-length endoscopy procedure videos were adopted for the training set (80% of total video frames) and a hold-out test (20% of total video frames). The trained model was then employed to make a final prediction on the 20% hold-out test set and results were compared with a quadratic weighted kappa inter-observer statistic (QWK) with blinded human off-site expert readers. Agreement was almost perfect for both the MES and the UCEIS (QWK 0.844 and 0.855, respectively). Notably, the model's performance was better for MES scores 0

and 3 and worse for MES scores 1 and 2, in which inter-observer variability among human readers is higher.

Another CNN system was trained by Yao *et al.*¹⁸ with a first cohort of 51 UC patients undergoing high-resolution colonoscopy. The system was later tested on 264 videos from 157 patients taking part in the LYC-30937-EC study, an international phase II randomized clinical trial of an investigational oral therapy for moderate to severe UC. Interestingly, a deep learning CNN was employed here for the first time automatically to exclude non-informative video frames. This CNN performed well in the automated scoring of local high-resolution video ($\kappa=0.84$) but performed less well in the unadjusted analysis of the external cohort of patients ($\kappa=0.59$). However, in this subset of patients a higher concordance level was found if a distinction between Mayo 0, 1 and Mayo 2, 3 was made (83.7%, 221 of 264).

Beyond standard luminal endoscopic imaging, several studies have also endorsed the use of confocal laser endomicroscopy (CLE) to assess deep mucosal healing and predict the disease course by evaluating microscopic healing (up to 1250-fold) throughout the colon and rectum^{19,20} or assessing intestinal barrier dysfunction in the terminal ileum^{21–23} of IBD patients. Briefly, CLE is based on the emission of a low power blue laser after topical (acriflavine hydrochloride, cresyl violet) or systemic (fluorescein sodium) administration of contrast agents.²⁴ CLE studies in IBD patients were conducted in referral centers and mostly rely on an expert post hoc revision of CLE images. AI-driven CLE could solve this issue by providing real-time CLE results without the need for advanced training. Recently, Quénéhervé *et al.*²⁵ were the first to explore the potential of AI-driven CLE diagnosis in a retrospective analysis of colorectal mucosal architecture from endomicroscopy images in IBD patients (23 CD and 27 UC) and healthy subjects ($n=9$). Excellent accuracy for IBD diagnosis (100% sensitivity and specificity) or UC *versus* CD differentiation (92% sensitivity, 91% specificity) was obtained in real time, confirming the high technical performance of CLE previously validated by expert revision of recorded CLE images.²⁶

Another advanced endoscopic imaging technique enabling *in vivo* microscopic imaging of the gastrointestinal mucosa (up to 1390-fold magnification)

is endocytoscopy. This technique is based on the principle of contact light microscopy, utilizes a fixed-focus, high-power objective lens, and requires mucolysis (e.g. N-acetyl-cysteine) and mucosal staining with an absorptive agent (e.g. methylene blue, toluidine blue or cresyl violet) or narrow band imaging modality.²⁴ As for CLE, this technique is also operator dependent and requires specific advanced skills. In 2019, Maeda *et al.*²⁷ developed the first CAD system predicting *in vivo* microscopic inflammation using 12,900 ultra-magnified endocytoscopic images (520-fold, CF Y-0058-I prototype and H290ECI; Olympus Corporation, Tokyo, Japan) from 87 UC patients. This system was found to provide an accuracy of 91% in predicting the presence of inflammation at histology defined by a Geboes score ≥ 3.1 as assessed by experienced and blinded pathologists. A prospective study is currently ongoing to compare long-term clinical prognoses with fully automated endocytoscopic diagnoses in UC.

More recently, Bossuyt *et al.*²⁸ have described a new CAD technique to assess images obtained with a new prototype endoscope with single short wave-length monochromatic LED light illumination (Fujifilm, Tokyo, Japan). This novel advanced endoscopic imaging technique allows the visualization of mucosal architectural features (e.g. crypts, pericryptal capillaries) up to a depth of approximately 50–200 μm in real time and without the need for contrast agents. In this prospective study, 58 UC subjects were assessed with an automated feature extraction technique that provides the number of pixels with bleeding and computes the vascular pattern by calculating the density of mucosal vessels per pixel. These two automated features were combined to optimize correlation with the histological Geboes score. The resulting CAD automated algorithm successfully predicted UC histological remission with high accuracy (86%) as compared with the use of standard endoscopic scoring systems (MES 74%, UCEIS 79%). Larger studies and validation in independent cohorts of patients are on the way.

New horizons. Several AI systems have already shown very promising results in evaluating both endoscopic and microscopic disease activity in UC, and many more will be implemented in the coming years. After proper validation in large, prospective multicenter trials and real-life clinical settings, the fully automated assessment of

endomicroscopic (or ‘deep’) mucosal activity in IBD patients will gradually become a new diagnostic standard to be integrated with the expertise of the IBD endoscopist and pathologist. In a more remote future, AI will probably reduce the need for dedicated IBD expert teams and multiple biopsies during routine follow-up endoscopy, facilitating the widespread implementation of precision medicine in IBD care management.

Colitis-associated neoplasia

Patients with long-standing IBD colitis have a higher risk of colorectal cancer, and colitis-associated neoplasia is often challenging to detect at an early stage as it can have a flat appearance, an unclear boundary and an atypical pit pattern. Despite technical advances in endoscopic imaging,²⁹ personalized strategies based on risk stratification³⁰ and the implementation of key quality measures in every-day practice,³¹ surveillance colonoscopy remains, from many points of view, the Achilles heel of IBD endoscopy. To date, there is no AI application for colorectal surveillance in patients with long-standing IBD colitis. Very recently, Maeda *et al.*³² have described the first ever reported case of AI-assisted detection of colitis-associated neoplasia in a 72-year-old man with an 18-year history of pancolitis. In another study, during surveillance colonoscopy, two flat lesions with low-grade dysplasia were clearly highlighted by EndoBRAIN-EYE (Cybernet Systems, Tokyo, Japan), an AI-based polyp detection system successfully adopted in previous trials to identify colorectal lesions in non-IBD patients.³³ Pending future software implementation supported by studies in IBD cohorts, this single experience suggests that this AI-based polyp detection system can already help non-expert endoscopists in detecting colitis-associated dysplasia in long-standing IBD colitis.

New horizons. The implementation of currently available AI-based polyp detection and characterization systems requires large IBD initiatives and big data analyses to assist the discovery of straightforward patterns enabling both the AI-assisted detection and the characterization of colitis-associated neoplasia despite active and post-inflammatory changes.²

Capsule endoscopy

Recent IBD guidelines advocate the increased use of capsule endoscopy (CE) to assess location and

activity in patients with suspected or established CD.^{3,5} However, CE reading requires dedicated training³⁴ and is a time-consuming activity that requires full concentration by the endoscopist for at least 1 h. In addition, inter-observer variability remains a major limitation to achieving reproducible assessment of the small bowel in IBD.³⁵ To overcome these limitations, an increasing number of studies have already addressed the potential of AI in the field of CE. Recent advances in deep learning algorithms in CE have been summarized in a 2021 systematic review with meta-analysis by Mohan *et al.*,³⁶ who evaluated the overall ability of these newly developed systems in diagnosing small bowel ulcers and/or bleeding in IBD and non-IBD patients. This systematic review included all studies that employed CNN models for AI training. All small bowel pathologies and associated lesions were included. A total of 4245 studies were retrieved and 88 full-length articles were assessed. Nine studies were then included in the final statistical analysis. For each paper, a contingency table recording accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) was created. The overall pooled accuracy of CE in detecting ulcers or bleeding was 95.4% (sensitivity 95.5%, specificity 95.8%, PPV 95.8% and NPV 96.8%), thus highlighting the potential of this innovation in CE reading.³⁶

As regards the specific IBD focus of our systematic review, a full application of an AI system based on a hybrid adaptive filtering and differential lacunarity (HFA DLac) architecture was explored in 2016 by Charisis and Hadjileontiadis for describing and detecting CD-associated lesions in CE.³⁷ The system was trained with a CE image database containing 400 images depicting CD-related lesions with the lowest possible similarity and 400 lesion-free frames which were judged by two independent endoscopists. The efficacy was later validated with two open CE databases containing six normal and 22 CD-related lesions images and 60 normal and 14 CD-related lesions, respectively. The accuracy of this HFA DLac system was then calculated based on the severity of the lesions, and ranged from 81.2% in mild lesions to 93.8% in severe lesions (total 90.5%). Although promising, the small sample size and the use of still images suggest the need for further confirmatory trials.³⁷ More recently, an Israeli group led by Klang and Barash extensively explored the use of AI systems based

on a CNN in the automated detection and grading of ulcers and strictures in CE for CD.³⁸ In a first retrospective single-center study aimed at automatically detecting CD mucosal ulcers, CNN-based software was trained with CE images taken from 49 patients (36 with CD and ulcerated small bowel mucosa, two with CD and normal mucosa, and 11 healthy controls) for a total of 17,640 images (7391 with mucosal ulcers and 10,249 with normal mucosa). The database was split into 80% of images that were employed as training modules and 20% of images that were used in the validation process. This AI system was then challenged in two distinct phases: in the first phase, endoscopic images from the same patient appeared in both the training and validation datasets. In the second phase, the CNN was trained with images taken from $n-1$ patients and then challenged on the unseen patient. The authors found that in phase I the CNN showed high accuracy in retrieving ulcers on randomly split images [area under the curve (AUC) 0.99, accuracy 95.4–96.7%] and in phase II the CNN was able to detect ulcerations in consecutive images from individual patients with high AUC levels (AUCs 0.94–0.99, accuracy 73.7–98.2%).³⁸ In a similar retrospective single-center study, a deep learning system based on EfficientNetB5 (CNN architecture; Google) was trained and challenged with 27,892 CE images consisting of 14,266 normal mucosa images, 1942 stricture images and 11,684 ulcer images categorized as mild ulcers (7075), moderate ulcers (2386) and severe ulcers (2233). The database of images was reviewed by capsule experts. The aim was to train the AI system to detect intestinal strictures and ulcers in CD capsule enteroscopy. The system was challenged by creating 10 patient-level experiments and it achieved an AUC of 0.942 in differentiating between strictures and all ulcers, an AUC of 0.989 in differentiating between strictures and normal mucosa, and AUCs of 0.992, 0.975 and 0.889 in differentiating strictures from mild, moderate and severe ulcers, respectively.³⁹ Another 2020 study from the same working group ascertained the accuracy of a CNN system in grading the severity of ulcers detected on CE. The experiment was divided into a first part in which inter-observer variability between two human experts was tested and a second part in which the CNN was trained and tested against the consensus of three expert capsule readers. A total of 17,640 CE images taken from CD patients (7391 with mucosal ulcers and 10,249 normal

images) were used to train and test the CNN. Of 1108 CE images, 488 were graded as mild ulcerations, 436 as moderate ulcerations and 184 as severe ulcerations. The inter-observer agreement between human readers was 76% when evaluating the difference between mild and severe ulcers, 40% between mild and intermediate ulcers, and 36% between intermediate and severe ulcers. Regarding agreement between AI and consensus reading, there was an overall agreement of 67% that reached 91% when discriminating between mild and severe ulcers (AUC 0.958, specificity 0.91%, sensitivity 0.91%), whereas agreement reduced when distinguishing mild from intermediate ulcers (65%; AUC 0.565, specificity 0.71%, sensitivity 0.34%) and intermediate from severe ulcers (79%; AUC 0.939, specificity 0.91%, sensitivity 0.73%).

New horizons. Capsule endoscopy is the most reproducible examination in digestive endoscopy as the storage of recorded images allows review by another endoscopist. Integrated algorithms highlighting relevant details and images are already offered by manufacturers to facilitate and shorten reading time.⁴⁰ The nature of CE makes it suitable for AI applications. A semi-automated CE reading system based on cloud technology will soon become a new standard owing to ongoing integration with AI systems for automated lesion detection, characterization and localization.

Other AI applications for IBD endoscopy

AI in detecting non-informative frames in colonoscopy. In a recent paper presented to the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Yao *et al.*¹⁸ explained how different artificial learning system architectures could improve AI automated performance of a simple yet significant task such as removing non-informative frames from IBD endoscopy-captured videos. To achieve this, the authors extracted 16,659 video frames from 10 colonoscopies in 10 patients with IBD. These video frames were reviewed manually by a single endoscopist and four types of frames were defined as non-informative: those with motion blur, those with poor prep, those that were too close to colonic mucosa, and those with overexposure due to excessive lightning. A total of 3829 frames were classified as informative, while 12,830 frames were annotated as non-informative. Then, four different AI methods were applied,

resulting in different AUCs: 0.909 ($p=0.02$) for hand-crafted features plus random forest (RF); 0.924 ($p=0.02$) for deep learning; 0.928 ($p=0.01$) for bottleneck features plus RF; and 0.939 ($p<0.01$) for feature fusion plus RF. While external validation in perspective cohorts from a real-life clinical scenario is awaited, this initial result suggests that feature fusion plus RF has the potential to improve the performance of AI learning systems that are now mainly based on deep learning features.

Endoscopic reporting. An underestimated yet important point concerns endoscopic reporting and IBD classification. The endoscopic report often has a remarkable impact on the clinical management of IBD patients, especially its concluding section. A retrospective study of 6399 Chinese patients who received a diagnosis of UC (5128), CD (875) or intestinal tuberculosis (ITB) (396) after endoscopic evaluation, was conducted by Tong *et al.*⁴¹ The aim of this study was to aid the endoscopic diagnosis of these three illnesses with the help of natural language processing and machine learning in the analysis of the endoscopic report. For this purpose, RF and CNN were employed to create an algorithm capable of automatically classifying UC, CD and ITB based on endoscopic results on a form or in free text. The overall performance of the algorithm in differentiating between UC and CD in terms of sensitivity, specificity and AUC was 0.89%, 0.84% and 0.93 when using RF; in differentiating between UC and ITB it was 0.83%, 0.82% and 0.89 when using RF; and in differentiating between CD and ITB it was 0.72%, 0.77% and 0.82 when using RF and 0.90%, 0.77% and 0.91 when using CNN. This preliminary study underlines the importance of standardized endoscopic wording and descriptors when reporting on IBD and suggests a successful application of a novel machine learning technique to improve endoscopic reporting.⁴¹

AI in correlating clinical data and colonoscopy findings. Finally, we report a retrospective single-center cohort study from Popa *et al.*⁴² who aimed to use clinical and endoscopic data in an AI algorithm to predict clinical remission in UC at 1 year. Patients with an established UC diagnosis, in clinical remission and in maintenance therapy with an anti-TNF drug (infliximab or adalimumab) who underwent colonoscopy for disease evaluation and had a follow-up of 1 year after

initial evaluation were retrospectively included. A neural network model was trained to identify patients who would clinically relapse at 1 year based on baseline endoscopic activity, neutrophil count, platelet distribution width, mean platelet volume, platelet large cell ratio, C-reactive protein and alpha 1 globulins. The initial dataset of 50 patients was randomly divided into a training group of 40 patients and a test group of 10 patients. A validation set of five patients was added independently. This newly developed neural network system had a well-performing receiver operating characteristic (ROC) curve (PPV 100%, NPV 100%; $p < 0.001$) in a small subset of patients, accurately differentiating those who will achieve clinical remission from those who will have active disease. Large and prospective external validation cohorts of unselected patients are now awaited to confirm the promising performance of this model mixing clinical and endoscopic features for predicting clinical relapse or remission in UC.

Discussion

In the era of personalized medicine and treat-to-target strategies, IBD endoscopy has become highly specialized and requires dedicated professionals with up-to-date skills and heterogeneous clinical and endoscopic competences.⁴³ In non-expert hands, IBD endoscopy can result in incomplete examination or reporting and missed or misclassified lesions. In clinical practice, this often means that repeat endoscopic examinations are required, leading to additional costs and decision-making delay. Also, in experimental settings, there can be variability between observers when subjective judgment is part of the observation, potentially leading to observer bias⁴⁴ and low result reliability. Several interventions and strategies have been implemented to address these problems. Endoscopic training and skill appraisal is now recognized as key to modern endoscopic training and several endoscopic societies are investing substantial resources in training programs.⁴⁵ The European Crohn's and Colitis Organisation has recently conducted a systematic review focused on the standardization of reporting in IBD endoscopy to improve quality and reproducibility in clinical practice and facilitate comparison of research data.⁹ In the past few years, off-site revision of endoscopic videos performed by blinded experts has been recognized as an essential evaluation criterion for clinical trials,

as it has been shown the additional costs are accompanied by improved result reliability and statistical efficiency.⁴⁶

Within this context, AI is widely expected to improve everyday diagnostic accuracy and reproducibility by enhancing the detection of subtle mucosal changes and standardizing lesion characterization and classification. Furthermore, synergic integration of AI-driven endoscopy with novel advanced endoscopic imaging techniques such as red density, confocal laser endomicroscopy, short wave-length monochromatic LED light illumination, and molecular imaging, is on the horizon. Combined with advanced endoscopic imaging, AI will reveal microscopic and molecular details invisible to the human eye, thereby heralding the achievement of ultrastructural, molecular and functional endpoints in IBD endoscopy as previously theorized by Neurath and Travis in a pioneering paper on mucosal healing.⁴⁷

In this systematic review we have focused on the early experimental implementation of AI in the field of IBD endoscopy. Selected records allowed identification of the main research areas and the technical capabilities and clinical potential of different AI applications for IBD endoscopy. Potential limitations of this work should be acknowledged. Our research was performed during the early stage of AI development and so included a relatively small number of studies, mostly with a cross-sectional design and conducted in a preclinical setting. Included studies had very heterogeneous aims, designs and endpoints, hampering direct comparison, aggregated evaluation or meta-analysis. In addition, as often happens when a promising research field starts to develop, all records showed positive results, raising the risk of potential publication bias across studies. Several questions and unmet needs have not yet been properly addressed, including the impact of mucosal visibility and bowel preparation or of other factors (e.g. scars, post-inflammatory polyposis) on CAD performance and reliability. Also, the role of high-definition imaging and chromoendoscopy combined with AI systems has not yet been assessed.

Therefore, our systematic review cannot provide a conclusive account of the use of AI in IBD endoscopy but only describe the current state of

the art and future perspectives. However, this systematic review, the first addressing AI development in the field of IBD endoscopy, has pointed out some major findings and provided some reflections. First, this research field is receiving a lot of attention, which is expected to increase further in the next few years. Second, although AI-assisted endoscopy has only recently been introduced in the field of IBD, all published studies reported very promising results and continuous technological progress. Within the next few years, we expect the implementation of AI-assisted IBD endoscopy in large and unselected prospective cohorts of patients better reflecting the real-life situation, to test the extra accuracy and additional benefit of employing AI compared with human endoscopists.

Author contributions

GET: study concept, search strategy, acquisition of data, interpretation of data, drafting of the manuscript, critical revision of the manuscript, study coordination.

AR, HN, MauV: interpretation of data, drafting of the manuscript, critical revision of the manuscript.

FC, CB, MarV: acquisition of data, interpretation of data, critical revision of the manuscript.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Funding

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

ORCID iD

Gian Eugenio Tontini  <https://orcid.org/0000-0002-8964-5686>

Supplemental material

Supplemental material for this article is available online.

References

1. Sinonquel P, Eelbode T, Bossuyt P, *et al.* Artificial intelligence and its impact on quality improvement in upper and lower gastrointestinal endoscopy. *Dig Endosc* 2021; 33: 242–253.
2. Tontini GE and Neumann H. Artificial intelligence: thinking outside the box. *Best Pract Res Clin Gastroenterol*. Epub ahead of print 5 December 2020. DOI: 10.1016/j.bpg.2020.101720
3. 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.
4. Leighton JA, Shen B, Baron TH, *et al.* ASGE guideline: endoscopy in the diagnosis and treatment of inflammatory bowel disease. *Gastrointest Endosc* 2006; 63: 558–565.
5. Lamb CA, Kennedy NA, Raine T, *et al.* British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* 2019; 68: s1–s106.
6. Aromataris E and Munn Z (eds). *Joanna Briggs Institute reviewer's manual*. Adelaide, Australia: The Joanna Briggs Institute, 2014.
7. Moher D, Liberati A, Tetzlaff J, *et al.*; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097.
8. Turner D, Ricciuto A, Lewis A, *et al.*; International Organization for the Study of IBD. 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.
9. Adamina M, Winterthur CH, Feakins R, *et al.* ECCO topical review optimising reporting in surgery, endoscopy, and histopathology. *J Crohns Colitis*. Epub ahead of print 11 January 2021. DOI: 10.1093/ecco-jcc/jjab011
10. Tontini GE, Bisschops R and Neumann H. Endoscopic scoring systems for inflammatory bowel disease: pros and cons. *Expert Rev Gastroenterol Hepatol* 2014; 8: 543–554.
11. Sasaki Y, Hada R and Munakata K. Computer-aided grading system for endoscopic severity in patients with ulcerative colitis. *Gastroenterol Endosc* 2004; 46: 2319–2324.
12. 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: 416–421.e1.
13. 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.
14. Takenaka K, Ohtsuka K, Fujii T, *et al.* Deep neural network accurately predicts prognosis of ulcerative colitis using endoscopic images. *Gastroenterology*. Epub ahead of print 21 January 2021. DOI: 10.1053/j.gastro.2021.01.210
 15. 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.
 16. Stidham R, 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.
 17. Gottlieb K, Requa J, Karnes W, *et al.* Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology* 2021; 160: 710–719.e2.
 18. Yao H, Stidham RW, Soroushmehr R, *et al.* Automated detection of non-informative frames for colonoscopy through a combination of deep learning and feature extraction. *Annu Int Conf IEEE Eng Med Biol Soc* 2019; 2019: 2402–2406.
 19. Tontini GE, Mudter J, Vieth M, *et al.* Prediction of clinical outcomes in Crohn's disease by using confocal laser endomicroscopy: results from a prospective multicenter study. *Gastrointest Endosc* 2018; 87: 1505–1514.e3.
 20. Buda A, Hatem G, Neumann H, *et al.* Confocal laser endomicroscopy for prediction of disease relapse in ulcerative colitis: a pilot study. *J Crohns Colitis* 2014; 8: 304–311.
 21. Liu JJ, Wong K, Thiesen AL, *et al.* Increased epithelial gaps in the small intestines of patients with inflammatory bowel disease: density matters. *Gastrointest Endosc* 2011; 73: 1174–1180.
 22. 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.
 23. Karstensen J, Săftoiu A, Brynskov J, *et al.* Confocal laser endomicroscopy: a novel method for prediction of relapse in Crohn's disease. *Endoscopy* 2016; 48: 364–372.
 24. Rath T, Tontini GE, Neurath MF, *et al.* From the surface to the single cell: novel endoscopic approaches in inflammatory bowel disease. *World J Gastroenterol* 2015; 21: 11260–11272.
 25. Quénéhervé L, David G, Boureille A, *et al.* Quantitative assessment of mucosal architecture using computer-based analysis of confocal laser endomicroscopy in inflammatory bowel disease. *Gastrointest Endosc* 2019; 89: 626–636.
 26. Tontini G, Mudter J, Vieth M, *et al.* Confocal laser endomicroscopy for the differential diagnosis of ulcerative colitis and Crohn's disease: a pilot study. *Endoscopy* 2015; 47: 437–443.
 27. Maeda Y, Kudo S-E, 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.
 28. Bossuyt P, De Hertogh G, Eelbode T, *et al.* Computer-aided diagnosis with monochromatic light endoscopy for scoring histologic remission in ulcerative colitis. *Gastroenterology* 2021; 160: 23–25.
 29. van der Laan JJH, van der Waaij AM, Gabriëls RY, *et al.* Endoscopic imaging in inflammatory bowel disease: current developments and emerging strategies. *Expert Rev Gastroenterol Hepatol* 2021; 15: 115–126.
 30. Bisschops R, East JE, Hassan C, *et al.* Correction: advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) guideline – update 2019. *Endoscopy* 2019; 51: C6.
 31. Iacucci M, Cannatelli R, Tontini G, *et al.* Improving the quality of surveillance colonoscopy in inflammatory bowel disease. *Lancet Gastroenterol Hepatol* 2019; 4: 971–983.
 32. Maeda Y, Kudo SE, Ogata N, *et al.* Can artificial intelligence help to detect dysplasia in patients with ulcerative colitis? *Endoscopy*. Epub ahead of print 1 October 2020. DOI: 10.1055/a-1261-2944
 33. Misawa M, Kudo S-E, Mori Y, *et al.* Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointest Endosc* 2021; 93: 960–967.e3.
 34. Sidhu R, Chetcuti Zammit S, Baltes P, *et al.* Curriculum for small-bowel capsule endoscopy and device-assisted enteroscopy training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) position statement. *Endoscopy* 2020; 52: 669–686.
 35. Leenhardt R, Buisson A, Bourreille A, *et al.* Nomenclature and semantic descriptions of ulcerative and inflammatory lesions seen

- in Crohn's disease in small bowel capsule endoscopy: an international Delphi consensus statement. *United Eur Gastroenterol J* 2020; 8: 99–107.
36. Mohan BP, Khan SR, Kassab LL, *et al.* High pooled performance of convolutional neural networks in computer-aided diagnosis of gastrointestinal ulcers and/or hemorrhage on wireless capsule endoscopy images: a systematic review and meta-analysis. *Gastrointest Endosc* 2020; 93: 356–364.e4.
37. Charisis VS and Hadjileontiadis LJ. Potential of hybrid adaptive filtering in inflammatory lesion detection from capsule endoscopy images. *World J Gastroenterol* 2016; 22: 8641–8657.
38. Klang E, Barash Y, Margalit RY, *et al.* Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020; 91: 606–613.e2.
39. Klang E, Grinman A, Soffer S, *et al.* Automated detection of Crohn's disease intestinal strictures on capsule endoscopy images using deep neural networks. *J Crohns Colitis*. Epub ahead of print 20 November 2020. DOI: 10.1093/ecco-jcc/jjaa234
40. Omori T, Hara T, Sakasai S, *et al.* Does the PillCam SB3 capsule endoscopy system improve image reading efficiency irrespective of experience? A pilot study. *Endosc Int Open* 2018; 6: E669–E675.
41. 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: 1–9.
42. Popa IV, Burlacu A, Mihai C, *et al.* 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. Shen B, Kochhar G and Hull TL. Bridging medical and surgical treatment of inflammatory bowel disease: the role of interventional IBD. *Am J Gastroenterol* 2019; 114: 539–540.
44. Bankhead C, Spencer E and Nunan D. Information bias. In: *Sackett catalogue of biases*. <https://catalogofbias.org/biases/information-bias/> (2019, accessed 10 January 2021).
45. Daperno M, Comberlato M, Bossa F, *et al.* Training programs on endoscopic scoring systems for inflammatory bowel disease lead to a significant increase in interobserver agreement among community gastroenterologists. *J Crohns Colitis* 2017; 11: 556–561.
46. 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: 149–157.e2.
47. Neurath MF and Travis SPL. Mucosal healing in inflammatory bowel diseases: a systematic review. *Gut* 2012; 61: 1619–1635.