

Artificial intelligence in inflammatory bowel disease endoscopy – a review of current evidence and a critical perspective on future challenges

Ther Adv Gastroenterol

2025, Vol. 18: 1–16

DOI: 10.1177/
17562848251350896

© The Author(s), 2025.
Article reuse guidelines:
[sagepub.com/journals-](https://sagepub.com/journals-permissions)
permissions

Ilaria Lodola, Ferdinando D'Amico, Silvio Danese and Tommaso Lorenzo Parigi 

Abstract: Inflammatory bowel disease (IBD) is a chronic and relapsing immune-mediated condition with a rising global prevalence. Endoscopic diagnosis, monitoring and surveillance currently depend on individual endoscopists, introducing subjectivity, variability, delays and potential diagnostic discrepancies. Artificial intelligence (AI) is poised to transform these processes. To date, most AI applications have focused on ulcerative colitis (UC) severity assessment, demonstrating promising results in replicating human evaluation, standardizing severity evaluation and facilitating the application of more complex scoring systems. Research into AI for Crohn's disease (CD) has lagged behind UC, due to challenges such as disease heterogeneity and transmural extension; nevertheless, significant progress has been made to automate capsule endoscopy readings for CD. Beyond the grading of disease severity, AI is also being explored for tasks such as identifying dysplastic lesions, differentiating IBD from other conditions, assessing intestinal barrier permeability, guiding treatment decisions and integrating data from multiple omics, though studies in these areas remain exploratory. This review examines the current landscape of AI applications in IBD endoscopy, summarizes key studies in the field and explores the future potential of AI in IBD care.

Plain language summary

The role of artificial intelligence in inflammatory bowel disease endoscopy: current evidence and future challenges

Inflammatory bowel disease (IBD), including both ulcerative colitis (UC) and Crohn's disease (CD), is a long-term condition affecting the digestive system and its prevalence is increasing worldwide. Currently, the diagnosis and monitoring of IBD by endoscopy relies heavily on the expertise of individual endoscopists, which can lead to differences in interpretation and delays in making the correct diagnosis. Artificial intelligence (AI) has the potential to improve this process by making it more accurate and consistent. Most AI research to date has focused on ulcerative colitis. AI systems have shown promise in assessing disease severity, providing more consistent assessments and allowing for more accurate scoring methods. Research into AI for Crohn's disease has been slower due to the disease's complexity and the deeper tissue involvement, although significant progress has been made in automating capsule endoscopy reading for CD. Beyond grading disease severity, AI is also being explored for detecting precancerous changes, differentiating IBD from other diseases and even guiding treatment decisions. However, these applications are still at an early stage. This review looks at the current role of AI in IBD endoscopy, highlights key studies and discusses how AI could shape the future of IBD care.

Correspondence to:
Tommaso Lorenzo Parigi
Department of
Gastroenterology and
Endoscopy, IRCCS
Ospedale San Raffaele, Via
Olgettina 58, Milan 20132,
Italy

Division of Immunology,
Transplantation and
Infectious Disease,
University Vita-Salute San
Raffaele, Milan, Italy

Department of Pathology,
Case Western Reserve
University, Cleveland,
OH, USA
parigi.tommaso@hsr.it
tommaso.parigi@gmail.com

Ilaria Lodola
Ferdinando D'Amico
Silvio Danese
Department of
Gastroenterology and
Endoscopy, IRCCS
Ospedale San Raffaele,
Milan, Italy

Division of Immunology,
Transplantation and
Infectious Disease,
University Vita-Salute San
Raffaele, Milan, Italy

Keywords: artificial intelligence, Crohn's disease, deep learning, endoscopy, inflammatory bowel disease, Mayo endoscopic subscore, neural networks, ulcerative colitis

Received: 2 February 2025; revised manuscript accepted: 2 June 2025.

Introduction

Artificial intelligence (AI) refers to the ability of machines to perform tasks such as learning and problem-solving, typically associated with human cognition. Through continuous learning and processing of large datasets, AI systems could, at least in theory, 1 day surpass humans in detecting patterns opening the door to personalized medicine.

Deep learning (DL), a prominent branch of AI, leverages multi-layered neural networks to learn complex data representations. Among these, convolutional neural networks (CNNs) are particularly effective for image and video analysis, making them a cornerstone of many DL applications in medical imaging.^{1,2}

Gastrointestinal (GI) endoscopy is a particularly fertile ground for AI applications due to the wealth of data generated and stored in the

recording of endoscopic procedures. Indeed, computer-aided detection (CADe) systems for polyps were among the first applications of AI to be approved for use in medicine.^{3,4}

The heterogeneity of inflammatory bowel disease (IBD), its unpredictable course and the difficulty in measuring its severity have sparked interest in AI-based tools to solve these issues. To date, most applications of AI in IBD have focused on detecting and assessing disease activity to reduce subjectivity and improve the prediction of disease course. However, other tasks are being explored such as identifying dysplastic lesions, differentiating IBD from other conditions and guiding treatment decisions (Figure 1).

In this manuscript, we will review the most relevant studies using AI applications to interpret and improve endoscopy, including advanced endoscopic tools and capsule endoscopy (CE), in the

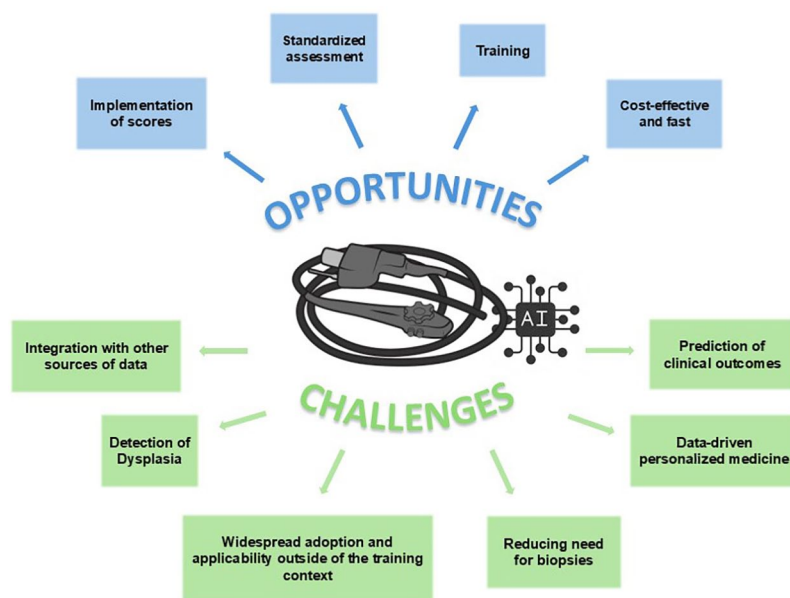


Figure 1. Opportunities and challenges of AI in IBD endoscopy. AI, artificial intelligence; IBD, inflammatory bowel disease.

setting of IBD. Finally, we will discuss future developments and limitations.

Disease monitoring

Ulcerative colitis

Endoscopy is the cornerstone of the treat-to-target management of UC and 'mucosal healing' is associated with reduced steroid use, hospitalizations, colectomies and improved quality of life.

UC endoscopy activity is assessed using various scoring systems including the Mayo endoscopic subscore (MES) and the ulcerative colitis endoscopic index of severity (UCEIS).

However, the high interobserver variability challenges the reproducibility of the measures, which is especially problematic in clinical trials.

Agreement among physicians tends to be higher for worse disease severity and lower for moderate or mild activity. For example, a study from Travis *et al.*⁵ found an inter-observer agreement of 76% for severe disease, but only of 37% for moderate disease and 27% for remission. Such inconsistencies are problematic in clinical trials where accurate outcome measures are essential. To overcome this, most clinical trials now rely on central reading by blinded reviewers to reduce bias, though this process is expensive and imperfect. AI offers a fast, cost-effective and more standardized solution (Table 1).

The first proof-of-concept application of AI to IBD endoscopy was reported in 2019 by Stidham *et al.* This CNN model separated endoscopic images of endoscopic remission (MES 0 or 1) from those of moderate-to-severe disease (MES 2 or 3), with an accuracy of 92.8% and a strong agreement with human reviewers (k -value 0.84), which were nearly equivalent to the agreement of human endoscopists between themselves (k -value 0.86).⁶

Soon after, Takenaka *et al.* published a neural network-based system designed to classify endoscopic images of UC in remission, based on the UCEIS of 0, which is a more strict definition compared to MES 0 or 1. The study was based on a larger dataset (40,758 images vs 16,514) and a more

balanced partition between training and testing, possibly accounting for the slightly lower accuracy (90.1%) and reviewer agreement ($k=0.798$), but a higher potential generalizability.

The same study went a step further training the model to predict also the corresponding histological activity/remission, for which the model achieved 92.9% accuracy and 0.859 κ agreement score compared to pathologists. The high agreement with biopsy suggests that AI could potentially reduce reliance on invasive biopsy procedures, streamline diagnostic processes or guide the endoscopist on where to take the biopsies. More importantly, this demonstrated the ability of AI models to make prediction on endpoints (histology) different from the inputs (endoscopy).⁷

While both studies showed that a neural network-based model could reach high diagnostic performance, they relied on still images introducing a potential selection bias. In fact, a real colonoscopy includes moments of suboptimal vision, due to poor lighting, bubbles, debris or other obstacles. Images captured in these settings are instead removed from the dataset.

Gottlieb *et al.* advanced AI applications by incorporating real-time video analysis of colonoscopies. Their machine model achieved impressive accuracy rates of 97% for UCEIS defined remission and 95.5% for the MES. Agreement metrics were equally robust, with weighted K values of 0.844 for MES (95% confidence interval (CI), 0.787–0.901) and 0.855 (95% CI, 0.80–0.91) for UCEIS, when compared to central reader performance.⁸

Similarly, Takenaka *et al.* presented an updated model to process full video colonoscopy in a prospective multicentre study. Their system demonstrated high accuracy in identifying endoscopic remission (UCEIS ≤ 1) with 82% sensitivity and 95% specificity and a strong correlation between the model's and the central readers, with an intra-class correlation coefficient of 0.93. These models represented a significant innovation as the first assessments of UC severity in full-length colonoscopy videos. Avoiding the selection bias inherent in fixed image and enhances clinical applicability through real-time functionality.⁹

Table 1. Main studies on AI-based tools for UC monitoring.

Reference	UC scoring system	Study design	Sample	Outcome	Results (accuracy/sensitivity/specificity)
Stidham et al. ⁶	MES	Retrospective single centre	14,862 training images (2778 pts). 1652 test images (304 pts).	MES 0–1 vs MES 2–3	92.8%/83.0%/96.0%
Takenaka et al. ⁷	UCEIS	Prospective single centre	40,758 training images (2012 pts). 4187 test images (875 pts).	Endoscopic remission (UCEIS=0)	93.3%/87.8%/90.1%
Gottlieb et al. ⁸	MES and UCEIS	Prospective multicentre	636 training videos. 159 test videos. (249 pts)	MES 0 vs 1–3 UCEIS 0 vs 2–8	95.5%/–/ 97.0%/–/–
Takenaka et al. ⁹	UCEIS	Prospective multicentre	770 patients evaluated using previously developed model.	Endoscopic remission (UCEIS ≤ 1)	–/81.5%/94.7%
Byrne et al. ¹¹	MES and UCEIS	Prospective single centre	134 training videos 1,550,030 images. 100 test videos.	MES 0–1 vs 2–3 UCEIS ≤ 3 vs > 3	94%/96%/91% 94.0%/93.4%/93.9%
Fan et al. ¹⁹	MES and UCEIS	Retrospective Single centre	5875 training images 20 test videos. (332 pts)	Grading of MES UCEIS items (vascular pattern, erosions and ulcers, bleeding)	86.54%/–/ 90.7%, 84.6%, 77.7%/–/–
Iacucci et al. ¹³	UCEIS and PICASSO	Prospective multicentre	1090 training videos. 67,280 training images (283 pts). 242 test videos.	UCEIS ≤ 1 PICASSO ≤ 3 (using VCE)	–/72%/87% –/79%/95%
Luo et al. ¹⁵	MES	Retrospective single centre	8583 training images. 2861 validation and test images. (1317 pts)	MES 0 vs 1–3 Grading of MES (MES 0/1/2/3)	98.9%/98.6%/– 91.6%/85.9%/–

AI, artificial intelligence; MES, mayo endoscopic score; UC, ulcerative colitis; UCEIS, ulcerative colitis endoscopic index of severity; VCE, virtual chromoendoscopy.

These studies also raise the question of whether the reference score used to train the model impacts its performance. Indeed, Gottlieb's findings suggest that the UCEIS-based model performed better than the MES-based one, perhaps as UCEIS allowed the system to anchor classification on more specific features. This is a common observation in clinical practice where more detailed scoring systems provide more accurate assessments at the price of less practicality as in the case of UCEIS that is more accurate than MES though less used.^{10,11}

Supporting these findings a meta-analysis showed that AI models trained with UCEIS had significantly

superior performances, with sensitivities of 93.6% (95% CI, 87.5–96.8%) for UCEIS compared to 82% (95% CI, 75.6–87%) for MES ($p=0.003$), and higher positive predictive value (PPV) 93.6% versus 83.6% for UCEIS and MES, respectively ($p=0.007$).¹²

Another study presenting two models trained using UCEIS and the PICaSSO score (Paddington International virtual ChromoendoScopy ScOre), which is even more detailed than UCEIS, found the model trained with the PICaSSO score to be slightly more accurate than the one based on UCEIS.¹³ Taken together, these studies support the idea that AI could facilitate the use of more

sophisticated scores, which are known to be more accurate but are less used in practice because of being time-consuming.¹⁴

While most early models focused on binary classification (remission vs active disease), recent work has aimed at building multiclass models to distinguish between degrees of severity, a task that remains clinically challenging. One notable example is the UC-Dense-Net model, which combines convolutional and recurrent neural networks. By integrating temporal information across video frames, the model achieved up to 4% higher accuracy and 3.5% better precision compared to previous methods, facilitating better detection of subtle inflammation.^{15,16}

Another major limitation of traditional endoscopic scoring systems for UC is the rigid classification of severity based on the most severely affected segment, overlooking activity in other bowel regions. For example, a patient with severe pancolitis may achieve a significant remission through most of the colon but still have a single persistent ulcer in the rectum and nevertheless the MES would remain unchanged, leading to a miscalculation of the true disease burden.¹⁷ AI has the potential to solve this, and several approaches have already been proposed such as having the AI score on a continuous scale from 0 to 10¹⁸ or dividing the colon into pre-defined sections, allowing the system to autonomously assess the inflammation at multiple points within each video.¹⁹

However, the most significant progress has been the cumulative disease score (CDS) proposed by Stidham *et al.*²⁰ This model analyses endoscopic videos and calculates a cumulative score by summing the contributions from all colonic segments, providing a more complete representation of disease activity. While endoscopists can manually sum the scores for each segment, this process is tedious and can only be done for a finite number of segments. Automating such task with AI significantly improves the efficiency and accuracy of clinical workflows.

The CDS showed a strong correlation with MES^{21,22} and increased the sensitivity to improvements in disease activity, a feature known as responsiveness. By capturing cumulative recovery across all segments of the colon, it better reflects treatment effects and, as a result, it increases the

statistical power and reduces the sample size needed to observe a statistical difference in clinical trials. In fact, when applying the CDS to the endoscopies of the UNIFI trial, the differences between ustekinumab and placebo increased as compared to MES, resulting in a 50% drop of sample size needed to detect a difference between the two groups. This makes the AI quantification of severity a valuable tool for improving the efficiency of clinical trials.²⁰

More recently, the UC-SCALE has built on these advances. This AI-based algorithm processes colonoscopy videos, assigns a MES to each readable frame and maps it to specific anatomical locations. UC-SCALE was trained on a large and heterogeneous dataset of 4326 sigmoidoscopy videos from 1953 UC patients collected at 554 clinical sites as part of 5 clinical trials evaluating Etrolizumab. To date, this is the largest dataset used to develop an automated scoring system in IBD, and more than four times larger than datasets used in previously published models. The diversity of this dataset, both in terms of clinical sites and patient demographics, significantly reduces the risk of overfitting and improves the generalisability of the model to new data. To ensure quality, an automated algorithm pre-selected readable frames from the videos.

UC-SCALE showed good inter-rater agreement (quadratic weighted kappa=0.73) with expert readings provided by a central group of five gastroenterologists. While this level of agreement is not perfect, it is consistent with the heterogeneity of the data. The model also showed significant correlations with clinical markers, including faecal calprotectin ($r_s=0.50$), C-reactive protein (CRP; $r_s=0.45$) and Physician Global Assessment ($r_s=0.45$), all of which were highly significant ($p<0.0001$), supporting the clinical validity of UC-SCALE.²³

Aside of CNNs, other computational methods have been employed to detect mucosal inflammation. One such method, the red density (RD) score, isolates colours from endoscopic images and correlates the density of red with inflammation, thus quantitatively measuring mucosal hypervascularization in real time. This approach is a promising alternative to assessing inflammation based on the detection of disease features such as erosions or ulcers, and does not require a human input for training.²⁴

The RD score showed strong correlations with established indices, including the MES ($r=0.76$, $p<0.0001$), the UCEIS ($r=0.74$, $p<0.0001$) as well as with histological activity measured with the Robarts Histopathology Index (RHI; $r=0.74$, $p<0.0001$) and effectively differentiated active inflammation from histological remission (RHI of ≤ 6), with 96% sensitivity and 80% specificity. However, the study did not report detailed metrics of diagnostic performance for endoscopy, which remains to be determined.²⁵

In a pilot study of 39 patients, RDS was also tested as a potential predictor of sustained clinical remission over 5 years resulting in a disappointing AUC of 0.68. This modest performance is likely imputed to the difficulty of predicting outcome, even for human physicians especially on the long term, when multiple factors come into play and reduce the relevance of the initial endoscopy. A larger study assessing RD in UC, the PROCEED trial, is ongoing (NCT04408703).²⁶

The same group investigated another CADx system using short-wavelength monochromatic LED light to assess mucosal architecture, including crypts and peri-cryptal capillaries, at depths of 50–200 μm . In active UC, increased inter-cryptal distance and crypt wall thickness, associated with inflammatory cell infiltration, are currently difficult to quantify in vivo. Although not included in histological scores, changes in peri-cryptal mucosal vascularization are associated with the degree of inflammation. The algorithm outperformed MES and UCEIS in detecting histological remission, achieving 86% accuracy compared to 74% and 79%, respectively, and demonstrated a PPV of 0.83 for histological remission compared to 0.65 for UCEIS and 0.59 for MES.²⁷

Another innovative technique in the assessment of ulcerative colitis (UC) is endocytoscopy (EC), an optical contact-based endoscopic system with up to 520-fold magnification. EC, coupled with the vital stains such as methylene blue and crystal violet, allows in vivo cellular imaging during GI endoscopy.

Maeda et al. developed a CAD system using ultra-magnification to detect histological inflammation (Geboes score >3), achieving a sensitivity of 74% (95% CI: 65%–81%), specificity of 97% (95% CI: 95%–99%) and diagnostic accuracy of 91% (95% CI: 93%–95%) and, even more

impressively, a perfect reproducibility with a k -value of 1.

The system was further developed in the EndoBRAIN, which classifies patients into healing or histologically active.²⁸ A prospective study showed that EndoBRAIN can predict relapse in UC patients in clinical remission, with significantly higher relapse rates in the AI-active group (28.4%) compared to the AI-healing group (4.9%) over a 12-month follow-up. To date, ENDOBRAIN is the only commercially available AI model for IBD endoscopy, although its use is limited by the need for specialised equipment and expertise, and it is only available in Japan.²⁹

Another system leveraging the subtle changes in vasculature, but compatible with a wider range of endoscopes, has been proposed as well. The AI provides an objective binary diagnosis of ‘AI-based vascular healing’ or ‘AI-based vascular activity’ showing significantly higher recurrence rate in patients classified as vascularly active group (23.9%) compared in the vascularly healed group (3.0%), although the AUC for outcome prediction remained modest, suggesting that even with the highest quality scopes the prediction of flares remains elusive.³⁰

In summary, by automating the analysis of endoscopic images and videos, AI offers a promising solution for achieving a more standardized, accurate and cost-effective assessment of disease activity both in daily practice and clinical trials. In addition, the integration of high-resolution imaging and novel scoring systems might improve our ability to predict histological remission and therefore outcomes, although accurate flare prediction is still far from being achieved.

Crohn's disease

While AI has been widely applied to conventional endoscopy for UC, in CD AI applications have focused primarily on CE and, more recently, intestinal ultrasound (IUS) rather than conventional endoscopy. This difference is due to the discontinuous and transmural nature of inflammation in CD and its frequent involvement of the proximal small bowel, all of which pose significant challenges to traditional endoscopy. CE is particularly useful for the detection of proximal small bowel lesions, which are associated with poorer long-term outcomes and can effectively

guide a treat-to-target strategy.^{31,32} In a recent randomised controlled trial patients with CD in clinical remission but high Lewis score (>350 ; a CE measure of disease activity) benefited from an optimised treatment approach and showed a reduced risk of clinical relapse compared to those who continued with standard care.³³

Nevertheless, CE adoption remains limited by long reading times, interobserver variability, difficulty in interpreting findings and cost. These limitations present a compelling case for AI-based solutions to improve both efficiency and diagnostic accuracy.

DL models, particularly CNNs, have demonstrated high performance in identifying small bowel erosions, ulcers and strictures with AUC values exceeding 0.94.^{34,35} Building on these results, research has progressively shifted towards investigating the role of pan-enteric CE – allowing simultaneous assessment of both the small bowel and colon – to improve the clinical utility of CE in CD management. Promising results have been achieved, with a reported sensitivity of 95.7% and specificity of 99.8%, for the detection of CD-related lesions. Notably, this model not only detected lesions but also classified disease severity, a critical factor in predicting disease progression and guiding therapy.^{36–39}

However, these studies did not validate results across different capsule devices and clinical settings. This issue was later addressed in another multicentre study that validated an AI model using data from two different CE platforms across multiple centres in Europe and the USA. Although the diagnostic sensitivity (94.6%) and accuracy (86.1%) were slightly lower than those reported in previous studies, the real-world performance underscores the robustness and interoperability of the model required for clinical translation.⁴⁰

Additionally, CE is emerging as a potential rapid rule-out tool for patients suspected of CD. The AXARO framework has demonstrated a negative predictive value of 97% for IBD and to a mean review time of less than 4 min per patients while maintaining an almost perfect agreement with human readers. This approach aligns with the increasing need to optimize the diagnostic workflow in response to growing workload and suggests that AI-assisted CE could serve as non-invasive option to rule out CD.^{41–43}

Another valuable non-invasive tool for CD monitoring is IUS. Adoption of IUS is increasing though its operator-dependence and image interpretation still limit widespread use. AI-driven solutions, including emerging vision transformers (ViTs), have shown potential in automating the detection of inflamed bowel regions in IUS, further expanding the role of AI in CD management.^{44,45} Significant progress has also been made in applying AI to therapeutic decision-making. Waljee *et al.*⁴⁶ have shown that, in patients with active Crohn's disease (CD), machine learning models using the week-6 albumin to CRP ratio can predict long-term responders to ustekinumab by week 8, potentially reducing both costs and delays in remission. Although these topics are beyond the scope of our review.

Surveillance

Risk of colorectal cancer (CRC) is increased in patients with IBD proportionally to the extent, severity and duration of inflammation. Current guidelines recommend periodic surveillance colonoscopy starting 8–10 years after initial diagnosis to ensure early detection of dysplastic changes in the colonic mucosa.^{47,48} Endoscopic surveillance of IBD is considered to be among the most challenging settings of diagnostic endoscopy due to the similarity of inflammatory and dysplastic changes.

Various surveillance strategies are possible including high-definition endoscopy with either white light, virtual chromoendoscopy or dye-based chromoendoscopy^{49–51} and advanced endoscopic imaging technologies, including laser confocal endomicroscopy and EC, are being investigated to improve detection. However, the early detection of IBD-associated dysplasia remains a challenge.⁵²

While, numerous CAD systems have been commercialized for the detection of colorectal lesions in the general population, none has been validated specifically for use in patients with IBD.^{53,54}

For instance, in a multicentre study conducted by Kudo *et al.*⁵⁵ to evaluate the accuracy of EndoBRAIN, a system trained on endocytoscopic images, in differentiating neoplastic from non-neoplastic colorectal lesions, patients with IBD were excluded due to the lack of sufficient data for adequate machine learning training. Similarly, the same exclusion criteria for IBD patients were

applied in the EndoBRAIN EYE, a CAD system with over 90% sensitivity and specificity for detecting colorectal polypoid lesions.⁵⁶

Nevertheless, the use of CAD systems designed for sporadic polyps has been occasionally reported in IBD. Using EndoBRAIN, Fukunaga et al.⁵⁷ identified a flat lesion with high-grade dysplasia in a patient with long-standing UC, which was subsequently removed by endoscopic submucosal dissection. Similarly, Maeda et al.⁵⁸ reported the detection of two low-grade dysplastic lesions in a patient with long-standing UC using EndoBRAIN EYE. Although anecdotal, these cases support the potential utility of the AI system in the detection of colitis-associated dysplasia and CRC.

However, the diagnosis of colitis-associated dysplasia remains particularly challenging due to the limited visibility of dysplastic lesions – often flat with poorly defined borders – and the high risk of false positives due to mucosal changes induced by chronic inflammation.⁵⁹

To address these issues, several CAD systems specifically designed for IBD surveillance are currently under investigation, although not yet clinically validated for routine use.

Yamamoto et al. conducted a pilot study in which CNNs were trained on 862 images to classify IBD-associated neoplasia into two categories: ‘adenocarcinoma/high-grade dysplasia’ and ‘low-grade dysplasia/sporadic adenoma/normal mucosa’. The AI achieved a sensitivity of 72.5%, specificity of 82.9% and overall accuracy of 79.0%, outperforming expert endoscopists (sensitivity: 60.5%, specificity: 88.0%, accuracy: 77.8%) and non-experts (sensitivity: 70.5%, specificity: 78.8%, accuracy: 75.8%). Interestingly, the non-experts showed higher sensitivity than the experts, which may reflect a less conservative diagnostic approach, aimed at avoiding missed lesions even at the expense of specificity.

Although preliminary, this study highlights the potential of AI to improve the accuracy of IBD surveillance by outperforming both expert and non-expert endoscopists in the diagnostic classification of IBD-associated neoplasia.⁶⁰

In another study, Vinsard et al. first applied a model trained on non-IBD lesions to detect IBD

dysplasia observing a poor sensitivity (50%) that testifies the difficulty in identifying IBD dysplasia and need for specifically trained systems. The authors then retrained the model on IBD images achieving promising sensitivity and specificity particularly with images in dye-chromoendoscopy (95.1% and 98.8%, respectively), albeit selected.⁶¹

Finally, Abdelrahim et al. developed a DL AI model for the detection and characterisation of IBD-associated neoplasia. The system was trained on over 18,000 endoscopic images, combining data from both IBD and non-IBD mucosa to improve generalizability and minimize overfitting. Specifically, the training set included images of both flat lesions and inflamed background mucosa with varying degrees of inflammation to address the two major challenges in IBD surveillance. The model was then validated on a separate dataset of 478 images from 30 IBD patients, achieving 93.5% sensitivity and 80.6% specificity for lesion detection, and 87.5% sensitivity and 80.6% specificity for lesion characterization. This approach also took into account regenerative and inflammatory lesions that may macroscopically resemble neoplasia but are histologically benign – such as pseudopolyps – and trained the system to detect and characterize them appropriately.⁶²

Overall, AI-based surveillance systems for IBD still face significant challenges, though progress is being made and a combination of CADE models with high sensitivity, for detection of lesions, and others with high specificity, for confirmation of dysplastic findings, is a realistic possibility for the near future.

Limitations

Despite the advances in the development of AI applications for IBD endoscopy, there are still several limitations.

Most of the models available have been developed and validated in patients with UC. While their performance is promising, these tools are more realistically suited for centralized reading in clinical trials, rather than immediate integration into routine clinical practice. Wider adoption will require external validation across different clinical settings, endoscopic devices and patient populations. This need was recently emphasized in a systematic review which highlighted the lack of external validation in many published studies, limiting the generalizability and real-world

Table 2. Main studies on AI-based tools for IBD diagnosis and differentiation from mimics.

Reference	Study design	Sample	Outcome	Results
Kim et al. ⁶⁵	Retrospective single centre	1635 training frames, 161 test frames.	Differentiating IBD from Behçet's disease and GI tuberculosis. (using WLE)	65.15% accuracy, 0.78 AUROC
Guimarães et al. ⁶⁶	Retrospective single centre	6617 training frames, 683 test frames.	Differentiating IBD from ischaemic colitis and infectious colitis. (using WLE)	70.9% accuracy, 0.727 AUROC
Wang et al. ⁶⁸	Retrospective multicentre	57,597 training frames, 1458 test frames.	Differentiation between CD and UC. (using WLE)	92.04% accuracy
Brodersen et al. ⁴¹	Prospective multicentre	131 CE of pts with suspected CD.	Identification of CD and IBD. (using CE)	92%–96% sensitivity 83%–90% specificity for CD; 97% sensitivity 90%–91% specificity for IBD.
Quénéhervé et al. ⁶⁹	Retrospective multicentre	23 CD pts videos 27 UC pts videos 9 controls pts videos	Identification of IBD (using CLE) Differentiation between UC and CD. (using CLE)	100% accuracy, sensitivity and specificity 92% sensitivity 91% specificity
AI, artificial intelligence; CD, Crohn's disease; CE, capsule endoscopy; CLE, confocal laser endomicroscopy; GI, gastrointestinal; IBD, inflammatory bowel disease; UC, ulcerative colitis; WLE, white-light endoscopy.				

applicability of AI tools. External validation should be considered a standard part of the development of robust AI models.⁶³ Another common limitation is the reliance on binary classification systems – active versus inactive disease – without capturing the nuanced spectrum of inflammation severity. Although recent approaches have introduced multiclass classification and continuous scoring systems, these are still at an early stage.⁶⁴

In CD, the challenges are compounded by the limited focus of AI research on conventional ileo-colonoscopy. Most AI models have focused on CE, leaving a significant gap in the development of models applicable to conventional ileo-colonoscopy, which remains the standard diagnostic and monitoring tool in clinical practice.

A further challenge lies in developing reliable AI models for the detection of colitis-associated dysplasia and cancer. These lesions are relatively rare, subtle and heterogeneous, often appearing as flat or poorly demarcated areas within an inflamed mucosal background. Their rarity limits the availability of large training datasets; their

subtle morphological features make them difficult to distinguish – similar to how AI performance drops with modest inflammatory changes. Moreover, their heterogeneity and frequent association with pseudopolyps and regenerative changes introduce additional noise, further complicating AI-based detection.

Although recent studies have attempted to address these challenges by including diverse datasets with inflammatory changes and flat lesions, further research and large-scale, multi-centre validation are needed.

Another critical issue is the differentiation of IBD from its mimics (Table 2). Although AI has made strides, it still falls short in reliably differentiating between the different forms of colitis that resemble IBD, as evidenced by the available data. Endoscopic image-based algorithms, while promising, do not outperform experienced endoscopists.^{65,66} This limitation is primarily due to the variability in clinical presentations, which complicate effective algorithm training without large and diverse datasets that adequately

represent different conditions. One option could be combining clinical parameters with image-based data to enhance diagnostic accuracy, reflecting the multifaceted nature of human clinical decision-making.^{66,67} The main diagnostic challenge in IBD diagnosis does not lie in distinguishing between UC and CD^{41,68,69} – where diagnostic accuracy exceeds 90% – but rather in differentiating IBD from non-IBD conditions, for which accuracy is lower and the clinical cost of misdiagnosis higher.

The gap between the technological potential of AI and its clinical implementation remains a significant challenge. In uncontrolled real-world settings, AI models struggle due to flaws in training data, such as selection bias, overfitting due to small datasets, data leakage, underrepresentation of certain subgroups (e.g. minorities or anatomical variants) and the lack of a universally reliable ‘ground truth’. However, advanced approaches such as self-supervised models and ViTs (Table 3) offer promising solutions to some of these challenges by enabling models to learn from large-scale unlabelled data and capture complex patterns with improved generalizability.

Addressing these limitations requires a multifaceted strategy. The integration of open data repositories can help mitigate overfitting, while publicly accessible algorithms promote transparency and reproducibility in AI research. A particularly promising approach is federated learning, which allows AI models to be trained across multiple institutions without sharing sensitive patient data. This method can help overcome data-sharing barriers, foster collaboration and preserve patient confidentiality.^{71–74}

Beyond technical and practical considerations, there are also ethical implications of AI in healthcare, particularly regarding the responsibility for AI-assisted medical decisions. AI algorithms are often considered ‘black boxes’ meaning their decision-making processes are not always transparent or understandable to healthcare professionals. If an AI system contributes to a medical error: who is responsible for it? The developer of the algorithm, the healthcare institution or the medical professional who relied on it. To address this issue, it is crucial to establish clear frameworks for legal responsibility in AI-assisted medical decisions and to improve the transparency of AI

models through the application of explainable AI so that clinicians can interpret the results of the models and provide the patient with clear information on how a decision was made.^{1,75}

Explainability is also important to build trust in AI models, which are necessary for successful adoption in clinical practice. Equally important is ensuring that healthcare professionals become familiar with AI tools. This may require integrating AI education into medical training and offering continuous professional development programmes to keep clinicians informed about emerging technologies. Moreover, embracing AI in healthcare will demand a cultural shift, addressing scepticism and helping to overcome ‘technophobia’ among practitioners.⁷⁶

In addition, developing globally accepted safety and efficacy standards for AI applications in healthcare could improve the consistency of AI-driven solutions and ensure the protection of patients’ rights across borders. Currently, there are significant regulatory differences between jurisdictions. The European Union, through the Medical Device Regulation and the Artificial Intelligence Act, has introduced stricter requirements to ensure shared standards on quality and data protection, but this may result in additional obstacles for small and medium-sized companies and leading some to reconsider their market strategies. In contrast, the US FDA is adopting a more flexible regulatory framework for AI technologies. These regulatory discrepancies can lead to inequalities in the development and deployment of AI solutions. International harmonization of standards could help overcome these challenges, ensure safety and promote innovation.⁷⁷

An additional consideration is the economic feasibility of AI deployment. Currently, the costs of AI models for IBD are unknown, as no tool is commercially available. However, extrapolating from existing economic models for AI-assisted CRC screening and polyp detection, after the initial investment, AI model could be cost-efficient in the long term.⁷⁸

Only by addressing these challenges with collaborative approaches, appropriate regulations and technological innovations can the full potential of AI in IBD endoscopy be realized.

Table 3. AI terminology.

Term	Description
Neural network	Computational model composed of multiple layers of interconnected nodes (neurons), designed to mimic the structure and function of biological neural networks.
CNN	A type of neural network with a particular type of architecture useful for deep learning, CNN learns hierarchical features through back-propagation making it, useful for detection and recognition tasks in images (e.g. lesion identification in endoscopic imaging).
ViT ⁷⁰	A transformer-based deep learning model that utilizes self-attention mechanisms instead of convolutional layers, enabling it to capture global contextual information across image patches – useful in medical image analysis.
Supervised learning	A learning method that enables an algorithm to associate an input to an output based on labelled example data.
Unsupervised learning	A learning method that enables an algorithm to identify patterns within unlabelled example data.
CNN, convolutional neural network; ViT, vision transformer.	

Future perspectives

Recent advances in AI-assisted endoscopy have seen a shift from traditional CNNs to ViTs, which use self-attention mechanisms to analyse images globally and capture complex relationships across an image (Table 3).⁷⁹ Unlike CNNs which focus on local feature extraction, ViTs provide a more comprehensive understanding of complex datasets, enhancing their robustness to image imperfections such as bubbles, debris or suboptimal lighting.

A significant breakthrough in ViT applications is their integration with self-supervised learning, a paradigm that allows models to learn from unlabelled data, overcoming the limitations of traditional CNNs that rely on manually annotated datasets, which often introduce biases and inconsistencies. A notable real-world application of this approach is the Certai model for disease assessment of UC. By integrating ViTs with the self-supervised DINOv2 framework, Certai analyses large endoscopic datasets without requiring manually labelled ground truth. The model is pre-trained using DINOv2 and then refined through expert input, following a hybrid AI-human approach that enhances accuracy and clinical reliability.^{80,81} Despite these advances, hybrid AI models remain essential, as human oversight continues to play a critical role in validating AI-driven decisions. Furthermore, challenges such as class imbalance in medical datasets

persist, requiring advanced techniques like high-frequency balancing, including up-sampling (e.g. by rotating or modifying underrepresented images), and generative adversarial networks to generate synthetic data, thereby ensuring stable model performance across disease severities.^{79,82}

Moreover, AI is increasingly moving towards multimodal approaches, integrating multiple data sources for a more comprehensive understanding of disease, in line with the goals of precision medicine.⁸³ A 'Fusion Model' that combines endoscopic and histological data to predict patient outcomes and identify early responders to therapy has recently been presented, although the clear benefits over individual assessment are yet to be proved.⁸⁴

The studies discussed highlight the trajectory of AI in endoscopy, underscoring the growing importance of self-supervised learning, multimodal approaches and hybrid AI-human models (Figure 2).

Conclusion

AI holds significant promise for transforming IBD endoscopy by enabling more accurate, standardized and accessible care. In the short term, its most realistic application lies in enhancing the consistency and reproducibility of clinical trials. Looking ahead, as more sophisticated and

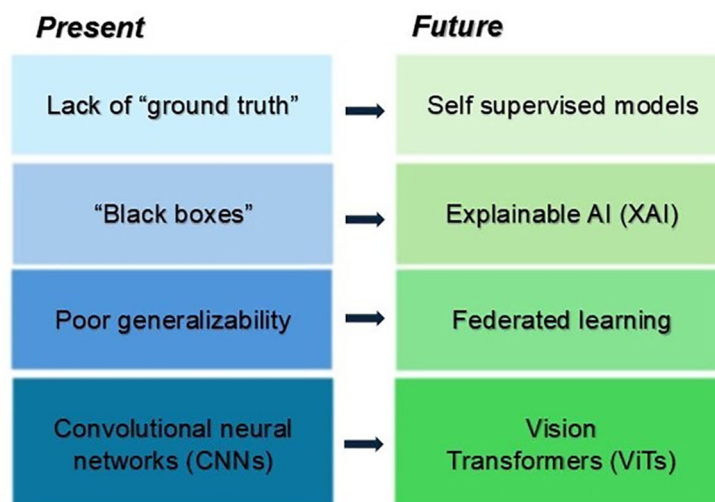


Figure 2. Present and future of AI.
AI, artificial intelligence.

robust models emerge, AI may streamline diagnostic workflows, reduce unnecessary interventions and ultimately improve patient outcomes. However, key challenges remain, including the development of generalizable and transparent algorithms, the navigation of complex regulatory frameworks and the need for strong interdisciplinary collaboration to ensure the design of safe, effective and patient-centred human–AI systems.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author contributions

Ilaria Lodola: Conceptualization; Writing – original draft.

Ferdinando D’Amico: Conceptualization; Supervision; Writing – review & editing.

Silvio Danese: Conceptualization; Supervision; Writing – review & editing.

Tommaso Lorenzo Parigi: Conceptualization; Writing – original draft; Writing – review & editing.

Acknowledgements

None.

Funding

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

Competing interests

F.D. served as a speaker for Sandoz, Janssen, Abbvie, Tillots, Takeda, Omega Pharma and Galapagos and served as an advisory board member for Abbvie, Galapagos and Nestlé. S.D. served as a speaker, consultant and advisory board member for Schering-Plow, Abbott (AbbVie) Laboratories, Merck, UCB Pharma, Ferring, Cellerix, Millennium Takeda, Nycomed, Pharmacosmos, Actelion, Alfa Wassermann, Genentech, Grunenthal, Pfizer, AstraZeneca, Novo Nordisk, Cosmo Pharmaceuticals, Vifor and Johnson and Johnson. T.L.P. has received speaker fees from Abbvie, Takeda, Tillots, Fresenius Kabi, Ferring and Janssen; Travel grants from Abbvie, Alfasigma, Janssen, CADI group, Pfizer, Takeda. I.L. declares no conflict of interest.

Availability of data and materials

Not applicable. This manuscript is a review of published articles all referenced in the draft.

ORCID iD

Tommaso Lorenzo Parigi  <https://orcid.org/0000-0002-3398-0231>

References

1. Iacucci M, Santacroce G, Zammarchi I, et al. Artificial intelligence and endo-histo-omics: new dimensions of precision endoscopy and histology in inflammatory bowel disease. *Lancet Gastroenterol Hepatol* 2024; 9(8): 758–772.
2. Da Rio L, Spadaccini M, Parigi TL, et al. Artificial intelligence and inflammatory bowel disease: where are we going? *World J Gastroenterol* 2023; 29(3): 508–520.
3. Ahmad HA, East JE, Panaccione R, et al. Artificial intelligence in inflammatory bowel disease endoscopy: implications for clinical trials. *J Crohns Colitis* 2023; 17(8): 1342–1353.
4. Pessarelli T, Tontini GE and Neumann H. Advanced endoscopic imaging for assessing mucosal healing and histologic remission in inflammatory bowel diseases. *Gastrointest Endosc Clin N Am* 2025; 35(1): 159–177.
5. Travis SPL, 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(4): 535–542.
6. 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(5): e193963.
7. 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(8): 2150–2157.
8. Gottlieb K, Requa J, Karnes W, et al. Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology* 2021; 160(3): 710–719.e2.
9. Takenaka K, Fujii T, Kawamoto A, et al. Deep neural network for video colonoscopy of ulcerative colitis: a cross-sectional study. *Lancet Gastroenterol Hepatol* 2022; 7(3): 230–237.
10. Nardone OM, Iacucci M, Villanacci V, et al. Real-world use of endoscopic and histological indices in ulcerative colitis: results of a global survey. *United Eur Gastroenterol J* 2023; 11(6): 514–519.
11. Byrne MF, Panaccione R, East JE, et al. Application of deep learning models to improve ulcerative colitis endoscopic disease activity scoring under multiple scoring systems. *J Crohns Colitis* 2023; 17(4): 463–471.
12. Jahagirdar V, Bapaye J, Chandan S, et al. Diagnostic accuracy of convolutional neural network-based machine learning algorithms in endoscopic severity prediction of ulcerative colitis: a systematic review and meta-analysis. *Gastrointest Endosc* 2023; 98(2): 145–154.e8.
13. Iacucci M, Smith SCL, Bazarova A, et al. An international multicenter real-life prospective study of electronic chromoendoscopy score PICA^{SSO} in ulcerative colitis. *Gastroenterology* 2021; 160(5): 1558–1569.e8.
14. Iacucci M, Cannatelli R, Parigi TL, et al. A virtual chromoendoscopy artificial intelligence system to detect endoscopic and histologic activity/remission and predict clinical outcomes in ulcerative colitis. *Endoscopy* 2022; 55(4): 332–341.
15. Luo X, Zhang J, Li Z, et al. Diagnosis of ulcerative colitis from endoscopic images based on deep learning. *Biomed Signal Process Control* 2022; 73: 103443.
16. Qi J, Ruan G, Liu J, et al. PHF3 technique: a pyramid hybrid feature fusion framework for severity classification of ulcerative colitis using endoscopic images. *Bioengineering* 2022; 9(11): 632.
17. Maeda Y, Kudo S-E, Kuroki T, et al. Automated endoscopic diagnosis in IBD: the emerging role of artificial intelligence. *Gastrointest Endosc Clin N Am* 2025; 35(1): 213–233.
18. Takabayashi K, Kobayashi T, Matsuoka K, et al. Artificial intelligence quantifying endoscopic severity of ulcerative colitis in gradation scale. *Dig Endosc* 2024; 36(5): 582–590.
19. Fan Y, Mu R, Xu H, et al. Novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. *Gastrointest Endosc* 2023; 97(2): 335–346.
20. Stidham RW, Cai L, Cheng S, et al. Using computer vision to improve endoscopic disease quantification in therapeutic clinical trials of ulcerative colitis. *Gastroenterology* 2024; 166(1): 155–167.e2.
21. Gu P, Mendonca O, Carter D, et al. AI-luminating artificial intelligence in inflammatory bowel diseases: a narrative review on the role of AI in endoscopy, histology, and imaging for IBD. *Inflamm Bowel Dis* 2024; 30(12): 2467–2485.

22. Ahmed M, Stone ML and Stidham RW. Artificial intelligence and IBD: where are we now and where will we be in the future? *Curr Gastroenterol Rep* 2024; 26(5): 137–144.
23. Gutierrez Becker B, Fraessle S, Yao H, et al. Ulcerative colitis severity classification and localised extent (UC-SCALE): an artificial intelligence scoring system for a spatial assessment of disease severity in ulcerative colitis. *J Crohns Colitis* 2025; 19(2): jjaf031.
24. Biamonte P, D'Amico F, Fasulo E, et al. New technologies in digestive endoscopy for ulcerative colitis patients. *Biomedicines* 2023; 11(8): 2139.
25. 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(10): 1778–1786.
26. Sinonquel P, Bossuyt P, Sabino JPG, et al. Long-term follow-up of the red density pilot trial: a basis for long-term prediction of sustained clinical remission in ulcerative colitis? *Endosc Int Open* 2023; 11: E880–E884.
27. Bossuyt P, Hertogh GD, Eelbode T, et al. Computer-aided diagnosis with monochromatic light endoscopy for scoring histologic remission in ulcerative colitis. *Gastroenterology* 2021; 160(1): 23–25.
28. Santacroce G, Zammarchi I, Tan CK, et al. Present and future of endoscopy precision for inflammatory bowel disease. *Dig Endosc* 2024; 36(3): 292–304.
29. 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(2): 408–415.
30. Kuroki T, Maeda Y, Kudo S-E, et al. A novel artificial intelligence-assisted 'vascular healing' diagnosis for prediction of future clinical relapse in patients with ulcerative colitis: a prospective cohort study (with video). *Gastrointest Endosc* 2024; 100(1): 97–108.
31. Lazarev M, Huang C, Bitton A, et al. Relationship between proximal Crohn's disease location and disease behavior and surgery: a cross-sectional study of the IBD genetics consortium. *Am J Gastroenterol* 2013; 108(1): 106.
32. Ben-Horin S, Lahat A, Amitai MM, et al. Assessment of small bowel mucosal healing by video capsule endoscopy for the prediction of short-term and long-term risk of Crohn's disease flare: a prospective cohort study. *Lancet Gastroenterol Hepatol* 2019; 4(7): 519–528.
33. Ben-Horin S, Lahat A, Ungar B, et al. Capsule endoscopy-guided proactive treat-to-target versus continued standard care in patients with quiescent Crohn's disease: a randomized controlled trial. *Gastroenterology*. Epub ahead of print March 2025. DOI: 10.1053/j.gastro.2025.02.031.
34. 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(3): 606–613.e2.
35. O'Hara FJ and Namara DM. Capsule endoscopy with artificial intelligence-assisted technology: real-world usage of a validated AI model for capsule image review. *Endosc Int Open* 2023; 11(10): E970–E975.
36. Majtner T, Brodersen JB, Herp J, et al. A deep learning framework for autonomous detection and classification of Crohn's disease lesions in the small bowel and colon with capsule endoscopy. *Endosc Int Open* 2021; 09: E1361–E1370.
37. Oliva S, Veraldi S, Russo G, et al. Pan-enteric capsule endoscopy to characterize Crohn's disease phenotypes and predict clinical outcomes in children and adults: the Bomiro study. *Inflamm Bowel Dis* 2025; 31(3): 636–646.
38. Ferreira JPS, de Mascarenhas Saraiva MJdaQEC, Afonso JPL, et al. Identification of ulcers and erosions by the novel Pillcam™ Crohn's capsule using a convolutional neural network: a multicentre pilot study. *J Crohns Colitis* 2022; 16(1): 169–172.
39. Ukashi O, Soffer S, Klang E, et al. Capsule Endoscopy in inflammatory bowel disease: panenteric capsule endoscopy and application of artificial intelligence. *Gut Liver* 2023; 17(4): 516–528.
40. Cardoso P, Mascarenhas M, Mendes F, et al. DOP089 shaping the future of IBD diagnostics: a multicentric AI-driven capsule endoscopy study. *J Crohns Colitis* 2025; 19(Suppl._1): i247.
41. Brodersen JB, Jensen MD, Leenhardt R, et al. Artificial intelligence-assisted analysis of Pan-enteric capsule endoscopy in patients with suspected Crohn's disease: a study on diagnostic performance. *J Crohns Colitis* 2024; 18(1): 75–81.

42. Oh DJ, Hwang Y, Kim SH, et al. Reading of small bowel capsule endoscopy after frame reduction using an artificial intelligence algorithm. *BMC Gastroenterol* 2024; 24(1): 80.
43. Eidler P, Kopylov U and Ukashi O. Capsule endoscopy in inflammatory bowel disease: evolving role and recent advances. *Gastrointest Endosc Clin N Am* 2025; 35(1): 73–102.
44. Delaire J, Mohtaram N, Yzet C, et al. P0400 Automated real-time detection of intestinal tract and inflammatory bowel disease in intestinal ultrasound with the vision transformer (ViT). *J Crohns Colitis* 2025; 19(Suppl_1): i896.
45. Carter D, Albshesh A, Shimon C, et al. Automatized detection of Crohn's disease in intestinal ultrasound using convolutional neural network. *Inflamm Bowel Dis* 2023; 29(12): 1901–1906.
46. Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and validation of machine learning models in prediction of remission in patients with moderate to severe Crohn disease. *JAMA Netw Open* 2019; 2(5): e193721.
47. Magro F, Gionchetti P, Eliakim R, et al. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. Part 1: definitions, diagnosis, extra-intestinal manifestations, pregnancy, cancer surveillance, surgery, and ileo-anal pouch disorders. *J Crohns Colitis* 2017; 11(6): 649–670.
48. Marques KF, Marques AF, Lopes MA, et al. Artificial intelligence in colorectal cancer screening in patients with inflammatory bowel disease. *Artif Intell Gastrointest Endosc* 2022; 3(1): 1–8.
49. Shukla R, Salem M and Hou JK. Use and barriers to chromoendoscopy for dysplasia surveillance in inflammatory bowel disease. *World J Gastrointest Endosc* 2017; 9(8): 359–367.
50. Marion JF, Wayne JD, Present DH, et al. Chromoendoscopy-targeted biopsies are superior to standard colonoscopic surveillance for detecting dysplasia in inflammatory bowel disease patients: a prospective endoscopic trial. *Am J Gastroenterol* 2008; 103(9): 2342–2349.
51. Iacucci M, Bonovas S, Bazarova A, et al. Validation of a new optical diagnosis training module to improve dysplasia characterization in inflammatory bowel disease: a multicenter international study. *Gastrointest Endosc* 2024; 99(5): 756–766.e4.
52. Diaconu C, State M, Birligea M, et al. The role of artificial intelligence in monitoring inflammatory bowel disease – the future is now. *Diagnostics (Basel)* 2023; 13(4): 735.
53. Misawa M, Kudo S-E, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018; 154(8): 2027–2029.e3.
54. Wallace MB, Sharma P, Bhandari P, et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. *Gastroenterology* 2022; 163(1): 295–304.e5.
55. Kudo S-E, Misawa M, Mori Y, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. *Clin Gastroenterol Hepatol* 2020; 18(8): 1874–1881.e2.
56. 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(4): 960–967.e3.
57. Fukunaga S, Kusaba Y, Ohuchi A, et al. Is artificial intelligence a superior diagnostician in ulcerative colitis? *Endoscopy* 2020; 53: E75–E76.
58. Maeda Y, Kudo S-E, Ogata N, et al. Can artificial intelligence help to detect dysplasia in patients with ulcerative colitis? *Endoscopy* 2020; 53: E273–E274.
59. Solitano V, Zilli A, Franchellucci G, et al. Artificial endoscopy and inflammatory bowel disease: welcome to the future. *J Clin Med* 2022; 11(3): 569.
60. Yamamoto S, Kinugasa H, Hamada K, et al. The diagnostic ability to classify neoplasias occurring in inflammatory bowel disease by artificial intelligence and endoscopists: a pilot study. *J Gastroenterol Hepatol* 2022; 37(8): 1610–1616.
61. Vinsard DG, Fetzer JR, Agrawal U, et al. Development of an artificial intelligence tool for detecting colorectal lesions in inflammatory bowel disease. *iGIE* 2023; 2(2): P91–P101.e6.
62. Abdelrahim M, Siggins K, Iwadata Y, et al. New AI model for neoplasia detection and characterisation in inflammatory bowel disease. *Gut* 2024; 73(5): 725–728.
63. Lee MCM, Farahvash A and Zazos P. Artificial intelligence for classification of endoscopic severity of inflammatory bowel disease: a systematic review and critical appraisal. *Inflamm Bowel Dis*. March 2025. DOI: 10.1093/ibd/izaf050.
64. Kusumam K, Mohankumar G, Favory X, et al. DOP077 EndoUC: an AI-assisted endoscopic ulcerative colitis activity grading application for

- deployment in clinical trials. *J Crohns Colitis* 2025; 19(Suppl_1): i226–i228.
65. Kim JM, Kang JG, Kim S, et al. Deep-learning system for real-time differentiation between Crohn's disease, intestinal Behçet's disease, and intestinal tuberculosis. *J Gastroenterol Hepatol* 2021; 36(8): 2141–2148.
66. Guimarães P, Finkler H, Reichert MC, et al. Artificial-intelligence-based decision support tools for the differential diagnosis of colitis. *Eur J Clin Invest* 2023; 53(6): e13960.
67. Jung Y, Hwangbo Y, Yoon SM, et al. Predictive factors for differentiating between Crohn's disease and intestinal tuberculosis in Koreans. *Am J Gastroenterol* 2016; 111(8): 1156.
68. Wang L, Chen L, Wang X, et al. Development of a convolutional neural network-based colonoscopy image assessment model for differentiating Crohn's disease and ulcerative colitis. *Front Med (Lausanne)* 2022; 9: 789862.
69. Quénéhervé L, David G, Bourreille A, et al. Quantitative assessment of mucosal architecture using computer-based analysis of confocal laser endomicroscopy in inflammatory bowel diseases. *Gastrointest Endosc* 2019; 89(3): 626–636.
70. Parvaiz A, Khalid MA, Zafar R, et al. Vision transformers in medical computer vision – a contemplative retrospection. *Eng Appl Artif Intell* 2023; 122: 106126.
71. Hassan C, Mori Y and Sharma P. The pros and cons of artificial intelligence in endoscopy. *Am J Gastroenterol* 2023; 118(10): 1720.
72. Aristidou A, Jena R and Topol EJ. Bridging the chasm between AI and clinical implementation. *Lancet* 2022; 399(10325): 620.
73. Zammarchi I, Santacroce G and Iacucci M. Next-generation endoscopy in inflammatory bowel disease. *Diagnostics (Basel)* 2023; 13(15): 2547.
74. Riva M, Parigi TL, Ungaro F, et al. Hugging face's impact on medical applications of artificial intelligence. *Comput Struct Biotechnol Rep* 2024; 1: 100003.
75. van der Velden BHM, Kuijff HJ, Gilhuijs KGA, et al. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal* 2022; 79: 102470.
76. Ahmad HA, East JE, Panaccione R, et al. Artificial intelligence in inflammatory bowel disease: implications for clinical practice and future directions. *Intest Res* 2023; 21(3): 283–294.
77. El-Sayed A, Lovat LB and Ahmad OF. Clinical implementation of artificial intelligence in gastroenterology: current landscape, regulatory challenges, and ethical issues. *Gastroenterology*. Epub ahead of print March 2025. DOI: 10.1053/j.gastro.2025.01.254.
78. Areia M, Mori Y, Correale L, et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. *Lancet Digit Health* 2022; 4(6): e436–e444.
79. Shah SA, Taj I, Usman SM, et al. A hybrid approach of vision transformers and CNNs for detection of ulcerative colitis. *Sci Rep* 2024; 14(1): 24771.
80. Byrne M, Requa J, Panés J, et al. P0371 building a robust artificial intelligence solution for use in ulcerative colitis clinical trials. *J Crohns Colitis* 2025; 19(Suppl_1): i852.
81. Okegunna Y, Utriainen M, Cloots MJJ, et al. P0956 Predicting clinical remission in Crohn's disease: a comparative study of expert-generated and computer-generated Bayesian networks. *J Crohns Colitis* 2025; 19(Suppl_1): i1783.
82. Turan M and Durmus F. UC-NfNet: deep learning-enabled assessment of ulcerative colitis from colonoscopy images. *Med Image Anal* 2022; 82: 102587.
83. Iacucci M, Zammarchi I, Santacroce G, et al. P0556 A novel switching of artificial intelligence to generate simultaneously multimodal images to assess inflammation and predict outcomes in ulcerative colitis. *J Crohns Colitis* 2025; 19(Suppl_1): i1122–i1123.
84. P0340 Endo-histo foundational fusion model: a novel artificial intelligence approach for predicting histologic remission and early response to therapy in a phase 2 ulcerative colitis clinical trial. *J Crohns Colitis* 2025; 19(Suppl_1): i806–i807. Table 1. Main studies on AI-based tools for UC monitoring.