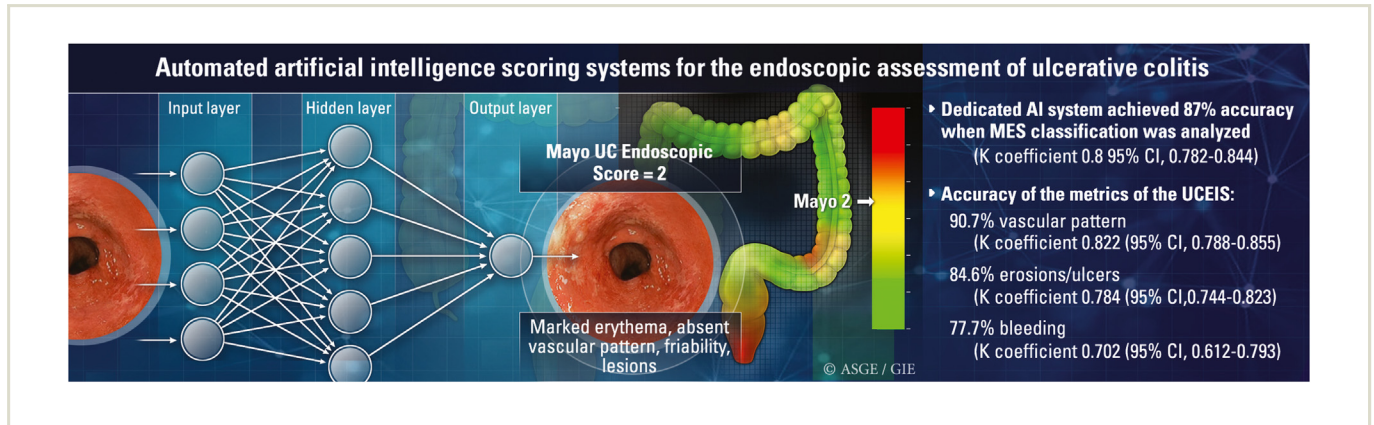




## Automated artificial intelligence scoring systems for the endoscopic assessment of ulcerative colitis: How far are we from clinical application?

### GRAPHICAL ABSTRACT



Artificial intelligence (AI) is going to drastically change our approach to diagnostic endoscopy. In contrast to its human counterpart, AI can manage an exceptional amount of data simultaneously, does not get fatigued, and can be highly effective and efficient. In the past couple of years, we have witnessed a literal blossom of AI systems applied to digestive endoscopy. Industries have been leading this first part of AI application, with the launch of real-time automated polyp detection and characterization systems to screening colonoscopy.<sup>1,2</sup>

In addition, a growing body of literature on AI systems for digestive endoscopy is being developed in academic settings with the combined efforts of medical, physics, and mathematics departments. These AI systems are mostly developed by the use of in-house databases, and they are usually not available as real-time software but rather are applied on fixed images or on recorded endoscopic videos. However, their field of application is vast, and they can possibly solve many frustrating issues that still linger unresolved. For instance, in the field of inflammatory bowel diseases, some successful attempts have already been made to apply AI to endocytomicroscopy so as to spare biopsies, or in video capsule endoscopy to automatically deliver a diagnosis of small-bowel Crohn's disease.<sup>3-6</sup>

However, reliable automated scoring for ulcerative colitis (UC) inflammatory activity was still missing; the most widespread endoscopic score for grading inflammation, the Mayo endoscopic score (MES), was never validated

Artificial intelligence (AI) is going to drastically change our approach to diagnostic endoscopy. In contrast to its human counterpart, AI can manage an exceptional amount of data simultaneously, does not get fatigued, and can be highly effective and efficient.

and suffers from moderate interobserver agreement for scores 1 and 2, being particularly relevant for official study inclusion criteria. To overcome this issue, endoscopic videos for including patients in clinical studies are still submitted to centralized readers to decrease this variability.<sup>7,8</sup>

In fact, although excellent overall diagnostic accuracy is obtained when AI performs on fixed images, it tends to underperform when an entire video is processed, because some disturbing factors (eg, blurred images, bleeding from biopsy sites) may at times hamper the quality of the endoscopic view, requiring amendments and adjusting factors.

In addition, the reproducibility of AI is dependent on the training dataset, and the correspondence between

the videos/images and the grade of inflammation is determined by human readers. When humans judge a training set of images, they introduce a bias that could not be eliminated and bears a degree of uncertainty. Another potential source of bias is represented by the training database itself, which could be “unbalanced” toward a lesser or higher degree of inflammation.<sup>9</sup>

In this issue of *Gastrointestinal Endoscopy*, Fan et al<sup>10</sup> report the development of an interactive AI scoring system dedicated to the endoscopic assessment of the colonic inflammation caused by UC. This AI system was initially trained by the use of 5875 white-light images and then applied to score 20 videos of full-length colonoscopy (ie, from the terminal ileum to the rectum), performed in 18 UC patients. The large bowel was divided into 5 macro areas (cecum and ascending colon, transverse colon, descending colon, sigmoid colon, and rectum), which were then broken down into multiple subsegments.

As a result, this novel, dedicated AI system achieved 87% accuracy when the MES classification was analyzed, with almost perfect agreement ( $\kappa$  coefficient 0.8; 95% CI, 0.782-0.844). The accuracies of the metrics of the Ulcerative Colitis Endoscopic Index of Severity were partially encouraging, being approximately 90.7%, 84.6%, and 77.7%, respectively, for vascular pattern, erosions/ulcers, and bleeding, with associated  $\kappa$  coefficients of 0.822 (95% CI, 0.788-0.855), 0.784 (95% CI, 0.744-0.823), and 0.702 (95% CI, 0.612-0.793). Finally, the AI scoring system evaluated and scored the 5 bowel sections and outlined colonic subsegmental inflammation activity, which was reproduced with a 2-dimensional image and graded by using different colors reflecting the severity of the inflammation.

To achieve these results, 4 endoscopists with different lengths of experience, ranging from 6 to 30 years, were asked to assess the 5875 white-light images used for the training phase, before the validation phase. When interobserver agreement was reached, the images were named “clean labels.” When agreement was not reached, the images were defined as “noisy labels.” Only clean labels were used for the validation set. It should be noted that the majority of the existing papers, related to the use of AI systems for grading UC, usually lacked a systematic explanation of how an agreement among the experts had been reached. In this sense, the authors tried to solve this problem before the validation set, weighing the so-called noisy labels, which certainly reduced a potential bias of selection.

Another important contribution of this study by Fan et al<sup>10</sup> is that, in contrast to previous studies, this novel AI system allowed a detailed assessment of the entire large bowel, using videos rather than static images. Also, as mentioned before, the authors provided a 2-dimensional map of inflammation, introducing a parameter of length to the endoscopic score, as for the newly proposed Modified Endoscopic Mayo Score. In this sense, the authors’ effort to propose AI software with integrated spatial evaluation is remarkable, inasmuch as studies have demon-

strated that the MES highly correlates with important biologic variables and outcomes, which could potentially ease the patient’s follow-up care. As a result, the AI system was able to independently calculate the MES and the Ulcerative Colitis Endoscopic Index of Severity, providing a detailed “map” of the severity of the colonic inflammation.

The authors, as stated in the introduction section, aimed to determine whether the AI system could provide a similar evaluation to both novice and experienced endoscopists. However, these 2 different levels of experience were not defined in the study, and this specific comparison was not reported.

The authors also found that in some cases of severe inflammation, the AI system overestimated the severity of the inflammation and failed to eliminate a repeated image; in other words, it was not able to recognize an already assessed image, which was repeatedly scored as a new image, increasing the overall severity of inflammation. Although we agree with the authors that this represents a limitation, we are also very optimistic that this bias will be addressed and eliminated soon, because of the implicit nature of the AI systems, which continue to progress and improve as new and more data are processed.

It is instead quite interesting that the lowest diagnostic performance related to the Ulcerative Colitis Endoscopic Index of Severity was bleeding, being approximately 78% in accuracy. This indeed is in keeping with the fact that, as previously mentioned, blood and other factors such as stools, debris, and bubbles may confuse the AI system, leading to lower specificity and sensibility. It is difficult to predict whether further development of the AI systems in endoscopy will overcome this limitation, which is also present in other fields of endoscopy, such as polyp detection. As it stands, human supervision is still needed, and it is indeed too early to completely rely on this newly developed AI system.

To conclude, this system has the potential to become a future real-life application, leading to significant changes in the endoscopic surveillance of UC. However, we are indeed still far from this scenario because this AI system has been validated only in 20 recorded videos and not in real-life endoscopy. Therefore, future prospective real-time studies are required to confirm these encouraging results.

## DISCLOSURE

*Both authors disclosed no financial relationships.*

**Alberto Murino, MD**

*Royal Free Unit for Endoscopy*

*The Royal Free Hospital and University College London*

*Institute for Liver and Digestive Health*

*Hampstead*

*Department of Gastroenterology*

*Cleveland Clinic London*

*London, United Kingdom*

**Alessandro Rimondi, MD***Royal Free Unit for Endoscopy**The Royal Free Hospital and University College London**Institute for Liver and Digestive Health**Hampstead**Department of Pathophysiology and Transplantation**University of Milan, Italy*

Abbreviations: AI, artificial intelligence; MES, Mayo endoscopic score; UC, ulcerative colitis.

**REFERENCES**

1. Rondonotti E, Hassan C, Tamanini G, et al. Artificial intelligence assisted optical diagnosis for resect and discard strategy in clinical practice (Artificial intelligence BLI Characterization; ABC study). *Endoscopy* 2023;55:14-22.
2. Kudo S-E, Misawa M, Mori Y, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. *Clin Gastroenterol Hepatol* 2020;18:1874-81.e2.
3. Tontini GE, Rimondi A, Venero M, et al. Artificial intelligence in gastrointestinal endoscopy for inflammatory bowel disease: a systematic review and new horizons. *Therap Adv Gastroenterol* 2021;14:17562848211917730.
4. 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;160:23-5.
5. Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2019;89:357-63.e2.
6. Ferreira JPS, de Mascarenhas Saraiva MJ da QEC, 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;16:16972.
7. 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* 2020;160:1570-83.
8. Sharara AI, Malaeb M, Lenfant M, et al. Assessment of endoscopic disease activity in ulcerative colitis: is simplicity the ultimate sophistication? *Inflamm Intest Dis* 2021;7:7-12.
9. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2020;2:e200029.
10. Fan Y, Mu R, Xu H, et al. A novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. *Gastrointest Endosc* 2023;97:335-46.

**Read Articles in Press Online Today!  
Scan the QR code or visit [www.giejournal.org](http://www.giejournal.org)**

*Gastrointestinal Endoscopy* now posts in-press articles online in advance of their appearance in the print edition of the Journal. These articles are available at the *Gastrointestinal Endoscopy* website, [www.giejournal.org](http://www.giejournal.org), by clicking on the "Articles in Press" link, as well as at Elsevier's ScienceDirect website, [www.sciencedirect.com](http://www.sciencedirect.com). Articles in Press represent the final edited text of articles that are accepted for publication but not yet scheduled to appear in the print journal. They are considered officially published as of the date of web publication, which means readers can access the information and authors can cite the research months prior to its availability in print. To cite Articles in Press, include the journal title, year, and the article's Digital Object Identifier (DOI), located in the article footnote. Please visit *Gastrointestinal Endoscopy* online today to read Articles in Press and stay current on the latest research in the field of gastrointestinal endoscopy.

