

Title	AI-based quantification of inflammatory extent for relapse prediction in ulcerative colitis: a prospective cohort study
Authors	Maeda, Yasuharu;Kudo, Shin Ei;Ogata, Noriyuki;Takenaka, Kento;Takabayashi, Kaoru;Kuroki, Takanori;Kawabata, Yurie;Okumura, Taishi;Sakurai, Tatsuya;Kouyama, Yuta;Ichimasa, Katsuro;Hayashi, Takemasa;Baba, Toshiyuki;Ogata, Haruhiko;Ohtsuka, Kazuo;Mori, Yuichi;Iacucci, Marietta;Misawa, Masashi
Publication date	2026-07-30
Original Citation	Maeda, Y, Kudo, S E, Ogata, N, Takenaka, K, Takabayashi, K, Kuroki, T, Kawabata, Y, Okumura, T, Sakurai, T, Kouyama, Y, Ichimasa, K, Hayashi, T, Baba, T, Ogata, H, Ohtsuka, K, Mori, Y, Iacucci, M & Misawa, M 2026, 'AI-based quantification of inflammatory extent for relapse prediction in ulcerative colitis: a prospective cohort study', Journal of Crohn's and Colitis, vol. 20, no. 7, jjag115, pp. 1-10. https://doi.org/10.1093/ecco-jcc/jjag115
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.scopus.com/pages/publications/105046204346 - 10.1093/ecco-jcc/jjag115
Rights	© 2026, the Author(s) . Published by Oxford University Press on behalf of European Crohn's and Colitis Organisation. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-12 19:39:53
Item downloaded from	https://hdl.handle.net/10468/19168



University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

AI-based quantification of inflammatory extent for relapse prediction in ulcerative colitis: a prospective cohort study

Yasuharu Maeda^{*1, }, Shin-Ei Kudo¹, Noriyuki Ogata¹, Kento Takenaka^{2, }, Kaoru Takabayashi³, Takanori Kuroki¹, Yurie Kawabata¹, Taishi Okumura¹, Tatsuya Sakurai¹, Yuta Kouyama¹, Katsuro Ichimasa¹, Takemasa Hayashi¹, Toshiyuki Baba¹, Haruhiko Ogata^{4,5}, Kazuo Ohtsuka⁶, Yuichi Mori⁷, Marietta Iacucci⁸, Masashi Misawa¹

¹Digestive Disease Center, Showa Medical University Northern Yokohama Hospital, Yokohama, Kanagawa, 224-8503, Japan

²Department of Gastroenterology and Hepatology, Institute of Science Tokyo, Tokyo, 113-8519, Japan

³Center for Diagnostic and Therapeutic Endoscopy, Keio University School of Medicine, Tokyo, 160-8582, Japan

⁴Center for Preventive Medicine, Keio University, Tokyo, 106-0041, Japan

⁵Fujita Medical Innovation Center Tokyo, Tokyo, 144-0041, Japan

⁶Endoscopy Unit, Institute of Science Tokyo, Tokyo, 113-8519, Japan

⁷Clinical Effectiveness Research Group, Institute of Health and Society, University of Oslo, Oslo, 0317 Norway

⁸APC Microbiome Ireland, College of Medicine and Health, University College Cork, Cork, T12 YT20, Ireland

*Corresponding author: Yasuharu Maeda, Digestive Disease Center, Showa Medical University Northern Yokohama Hospital, 35-1 Chigasaki-chuo, Tsuzuki, Yokohama 224-8503, Japan (yasuharu@med.showa-u.ac.jp).

Author Contributions: Y. Maeda and N. Ogata contributed equally.

Abstract

Background and Aims: Endoscopic remission is a key therapeutic goal in ulcerative colitis (UC); however, conventional indices focus on peak severity without accounting for the spatial extent of inflammation. We developed a deep learning-based score, Quantitative Ulcerative Colitis Assessment using Deep Learning (QUAD), and evaluated whether incorporating inflammatory extent improves relapse prediction in patients with UC in clinical remission. To our knowledge, no studies have examined whether artificial intelligence (AI)-derived assessment of inflammatory extent predicts clinical relapse.

Methods: This prospective cohort study evaluated relapse prediction over 24 months in patients with UC in clinical remission. The QUAD model assigns a score of 0-3 to each image quadrant, yielding a total score of 0-12. The model was trained on 84743 images from 998 patients with UC across three centers. Still image-based validation and automated full-length video analysis were conducted to assess the impact of inflammatory extent on relapse prediction.

Results: Clinical relapse occurred in 19.4% of patients with QUAD ≥ 4 compared with 5.2% of those with QUAD < 4 ($P = .01$), with an area under the curve (AUC) of 0.67 (95% confidence interval [CI]: 0.57-0.76). Notably, automated video-based analysis showed that distal inflammatory burden yielded the highest predictive performance, with analysis of the distal 10% segment achieving an AUC of 0.73 (95% CI: 0.62-0.85), compared with whole-colon assessment (AUC, 0.62; 95% CI: 0.48-0.76) and still image-based evaluation.

Conclusion: AI-augmented assessment integrating inflammatory severity and extent may provide complementary prognostic information beyond severity-based evaluation.

Key words: deep learning; endoscopic remission; relapse prediction

1. Introduction

Ulcerative colitis (UC) is a chronic inflammatory bowel disease marked by alternating periods of remission and relapse.^{1,2} Achieving and maintaining endoscopic remission through appropriate pharmacological treatment is essential because it supports long-term clinical remission, reduces the likelihood of surgery for management of recurrent inflammation, and lowers the risk of colitis-associated colorectal cancer.^{3,4}

With treat-to-target strategies now standard in UC management, endoscopic remission is widely recognized as a

key long-term goal.^{5,6} Conventional endoscopic indices, such as the Mayo Endoscopic Subscore (MES),⁷ focus primarily on the most severely affected mucosal areas.⁷ However, recent evidence indicates that both the severity and extent of mucosal inflammation have important prognostic value.^{8,9}

Advances in artificial intelligence (AI) have enabled automated endoscopic assessment of UC, including approaches that incorporate not only image-level severity^{10,11} but also classification of inflammatory extent.^{12,13} Despite these methodological advances, the clinical utility of incorporating inflammatory

extent into relapse prediction remains unclear.¹⁴ To address this limitation, we developed the Quantitative Ulcerative Colitis Assessment using Deep Learning (QUAD), a quadrant-based AI score designed to capture both the severity and extent of inflammation.

The aim of this prospective cohort study was to evaluate whether QUAD predicts clinical relapse in patients with UC in clinical remission and to explore whether still image-based and video-based assessments provide prognostic information.

2. Methods

This study comprised two main phases: multicenter AI model development (Phase I) and prospective validation (Phase II). Phase II included two analytic components: still image-based validation and automated full-length video analysis.

We developed and prospectively validated a quantitative, AI-driven colonoscopic scoring system, ranging from 0 to 12, called QUAD (Figure 1). The QUAD algorithm was trained using endoscopic data from three centers: Keio University Hospital, the Institute of Science Tokyo Hospital, and Showa Medical University Northern Yokohama Hospital. The validation study comprised three sequential components. First, we assessed the feasibility of the QUAD algorithm by examining its correlation with endoscopic disease activity in patients with UC. Second, we performed a single-group, open-label, prospective study to evaluate whether a QUAD score derived from still images could predict clinical relapse in patients with UC. Third, we applied automated video analysis to full-length withdrawal videos from the same cohort to quantify inflammatory extent. We then evaluated whether the AI-quantified inflammatory extent was associated with relapse. The primary endpoint was clinical relapse within 24 months. Secondary analyses included time-to-relapse assessment and exploratory video-based evaluation of inflammatory extent.

2.1. System description

The development of our QUAD system utilized the capabilities of a convolutional neural network, using the 16-layer Visual Geometry Group (VGG16) architecture. The algorithm was trained on 84743 colonoscopic images from 998 patients with UC, collected between October 2016 and June 2022 across Keio University Hospital, the Institute of Science Tokyo Hospital, and Showa Medical University Northern Yokohama Hospital. Each colonoscopic image was divided vertically and horizontally to produce four quadrant ($\frac{1}{4}$ -frame) images. Two board-certified gastrointestinal endoscopists—each with experience in more than 20000 colonoscopies—independently scored every quadrant according to a modified MES. Quadrants showing normal mucosa or scar tissue were assigned a score of 0, and those with erythema, yellow-white spots, or decreased vascular pattern received a score of 1. When erosion or mucosal granularity occupied $\geq 50\%$ of the quadrant, a score of 2 was given; when an ulcer or spontaneous bleeding occupied $\geq 50\%$ of the quadrant, a score of 3 was assigned. The set of quadrants labeled in this manner formed the training dataset for development of the QUAD-score algorithm. A QUAD score, ranging from 0 to 12, was calculated by summing the scores from the four quadrants in each image. Each quadrant was classified independently by the AI model, and the final QUAD score was calculated as the sum of the four quadrant-level predictions. A

QUAD score of ≥ 4 was determined as the threshold for relapse prediction based on receiver operating characteristic curve analysis of data from a previous study.¹¹ Low-quality frame exclusion was performed using a previously validated quality-control algorithm¹¹ before QUAD scoring. The development dataset consisted of colonoscopic images collected between October 2016 and June 2022 from three institutions and was used exclusively for model training. In Phase I, an additional prospective dataset collected in July 2022 was used to assess the correlation between QUAD scores and MES but was not used for model training or threshold optimization. Prospective validation was subsequently conducted in an independent cohort enrolled between August 2022 and February 2023. No patient overlap was permitted among the development, Phase I assessment, and prospective validation cohorts.

This model offers two complementary workflows:

Real-time mode. During examination, every captured still image is automatically assigned a 13-class QUAD score, with the highest score serving as the representative value for relapse-risk stratification. The 13-class output corresponds to possible summed quadrant scores ranging from 0 to 12, representing all combinations of quadrant-level inflammation severity.

Post-colonoscopy mode. After the procedure, the full withdrawal video is uploaded to the system. To characterize the spatial distribution of inflammation, the software samples the video at one frame per second, automatically discarding poor-quality frames. Frame sampling at one frame per second was selected a priori to balance computational feasibility and representative coverage of mucosal evaluation. Each remaining frame is assigned a QUAD score, and the patient's mean QUAD score is calculated and recorded.

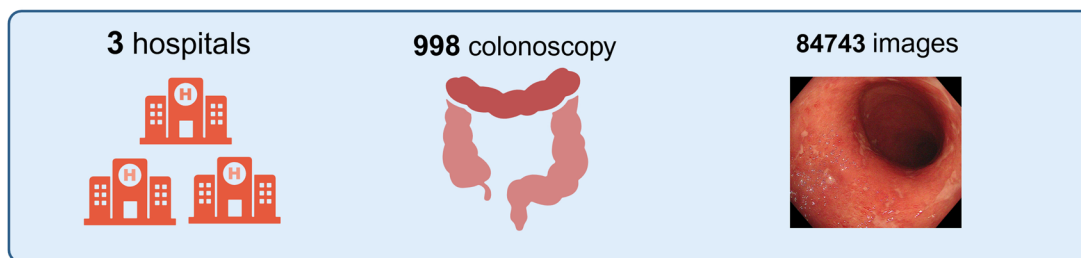
Key features include:

- Automatic exclusion of low-quality images ineligible for analysis because of halation, inadequate air insufflation, close proximity to the mucosa, or the use of chromoendoscopy or virtual chromoendoscopy.
- Fully automated 13-class scoring based on frame-by-frame analysis. The model visualizes AI-detected inflammatory activity by displaying color-coded indicators (red, yellow, blue) in the four corners of the screen. This intuitive visual aid helps endoscopists locate specific regions of active inflammation identified by the AI.
- Processing of full-length withdrawal videos to quantify inflammatory extent. Frames are extracted at one frame per second, screened for quality, and scored according to QUAD. The average QUAD score is then calculated and recorded. In addition to generating scores for the entire colon, the system can output scores for user-specified segments. Simultaneously, visual representations of inflammation severity are displayed as annotated images on the output screen (Figure 1).

2.2. Colonoscopy procedure for validation study

Commercially available colonoscopes (CF-HQ290ZI, PCF-H290ZI, PCF-PQ260I/L, CF-XZ1200I, and CF-EZ1500DI; Olympus Medical Systems, Tokyo, Japan) were used in this study. For bowel preparation, patients ingested 2–4 L of polyethylene glycol solution on the morning of the examination. Clearance of intestinal fluid was verified before initiating the

Learning Samples



Model Algorithm

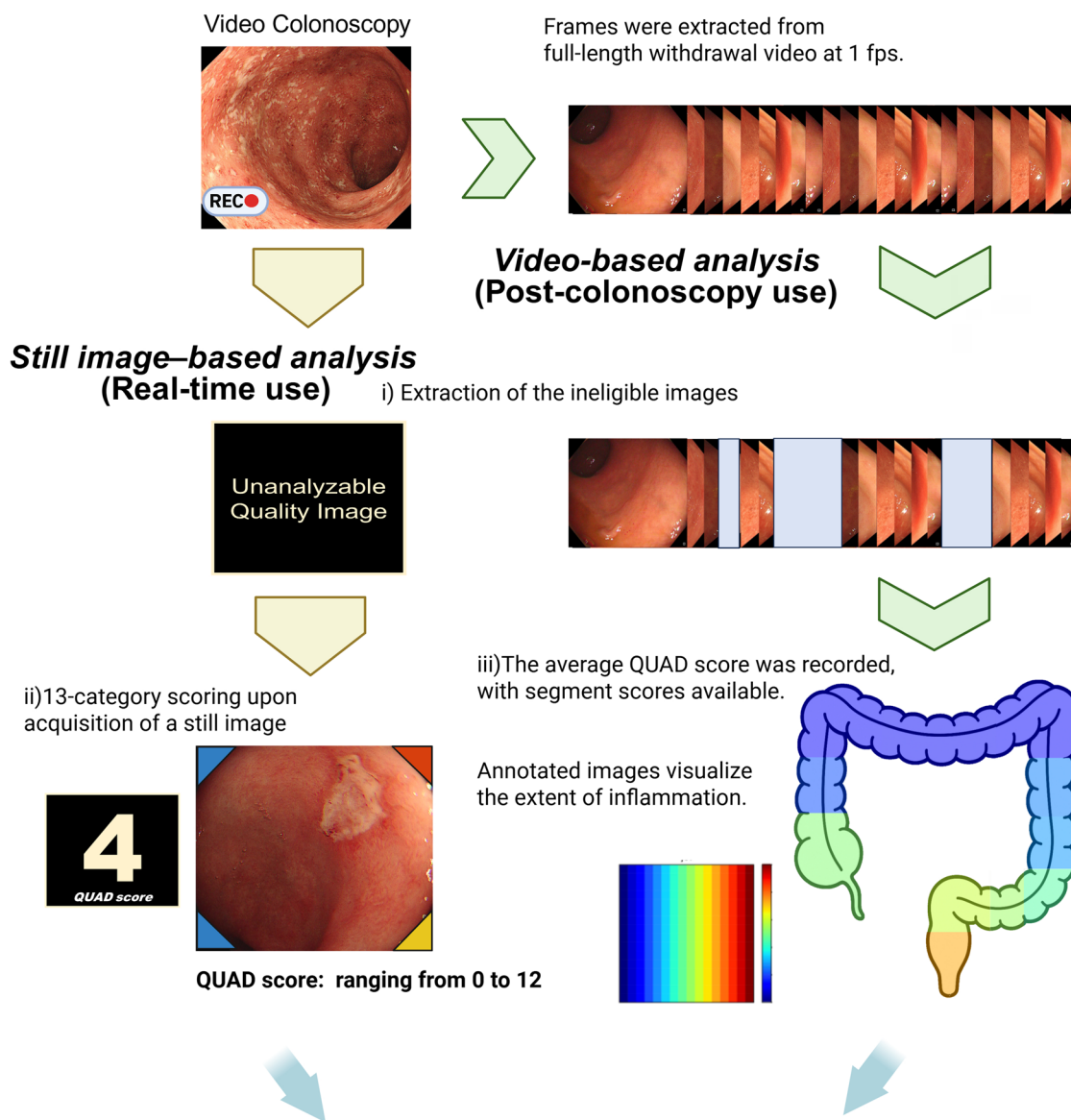


Figure 1. Overview of the Quantitative Ulcerative Colitis Assessment using Deep Learning (QUAD) framework and its clinical application. The QUAD system can be used in two complementary ways. During colonoscopy, still image-based analysis provides a real-time QUAD score reflecting inflammatory severity. After colonoscopy, automated video analysis quantifies inflammatory extent throughout the examined bowel. Both outputs may be used for relapse-risk stratification in patients with ulcerative colitis. Created with <https://BioRender.com>

colonoscopy, which was performed under conscious sedation with intravenous diazepam (5–10 mg) or midazolam (2–10 mg).

2.2.1. First phase

In July 2022, we prospectively collected endoscopic still images. The study included all patients with UC at our institution who underwent colonoscopy for clinical purposes and consented to participate. During withdrawal from the cecum, the endoscopists recorded still images according to clinical requirements, without observing the QUAD algorithm outputs. An expert endoscopist (N.O.) reviewed all recorded still images and assigned an MES. We then examined the correlations between the QUAD score and the MES.

2.2.2. Second phase

2.2.2.1. Still image-based analysis

A single-group, open-label, prospective study was conducted to evaluate whether a still image-based QUAD score could predict clinical relapse in patients with UC.

Study population. Eligible patients with UC undergoing colonoscopy at Showa Medical University Northern Yokohama Hospital between August 2022 and February 2023 were prospectively enrolled. A flowchart of patient inclusion is shown in Figure 2.

Inclusion criteria. Patients were eligible if they had a confirmed diagnosis of UC and were in clinical remission at the time of colonoscopy.

Exclusion criteria. Patients were excluded if they had active UC (partial Mayo score of >2), a history of colon surgery, or systemic prednisone use within 6 months before colonoscopy.

Follow-up protocol. Patients were followed in the outpatient clinic every 8–12 weeks for up to 24 months or until clinical relapse.

Outcome definition. The primary outcome measure was the rate of clinical relapse within 24 months after colonoscopy. Clinical relapse was defined as a partial Mayo score of >2 .

During withdrawal of the colonoscope from the cecum, the endoscopist captured at least one still image of the most severely inflamed area in each of the five colorectal segments (cecum–ascending colon, transverse colon, descending colon, sigmoid colon, and rectum) to generate diagnostic AI output. After each colonoscopy, the QUAD score for each image, as output by the AI, was recorded in a comma-separated file. One endoscopist (M.M.), blinded to the patient's clinical course and endoscopic information, aggregated the AI outputs from the file and registered the AI-based QUAD assignments. Both patients and their outpatient clinicians were blinded to the AI output. Expert endoscopists (Y.M. and N.O.) entered each patient's characteristics and clinical course data into the database. Patients were classified into two groups based on the QUAD score (<4 vs. ≥ 4) of the worst segment. The cutoff value of $\text{QUAD} \geq 4$ was determined a priori using an independent dataset from our previous study¹¹ before initiation of the present analysis and was not modified based on the current validation cohort. For segment scoring, when multiple images were available, the AI prediction was determined using the most frequently occurring QUAD score (eg, three outputs of QUAD score 0, two of score 1, and five of score 2 were classified as QUAD score 2). In cases of ties, the higher score was selected. The most frequent score was used to minimize the influence of outlier frames and to better represent the dominant inflammatory pattern within each segment.

2.2.2.2. Video-based analysis

Because still image-based analysis relied on endoscopist-selected representative images, we also performed a separate automated

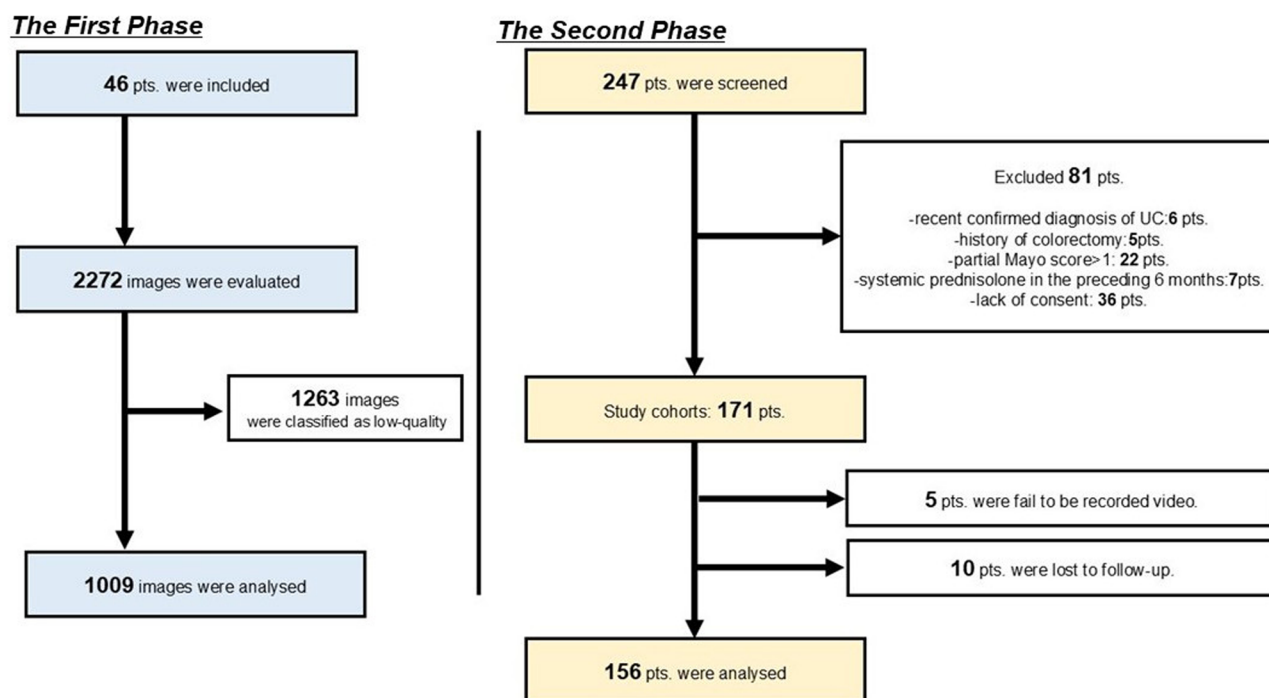


Figure 2. Study flow diagram across the three study components. In the first component, 46 patients provided 2272 images, of which 1009 high-quality images were analyzed. In the second and third components, 247 patients were screened; after exclusions and loss to follow-up, 156 patients were included in the final analysis.

video-based analysis to reduce frame-selection bias. Withdrawal videos were analyzed from cecal withdrawal to completion of the examination. Frames were extracted at 1 frame per second, and images classified as low quality by the predefined quality-control algorithm were excluded. For segmental analyses, the distal 50%, 40%, 30%, 20%, 10%, and 5% portions were defined according to withdrawal time and used as surrogate markers of progressively distal colonic segments. A schematic overview of the video segmentation workflow, including frame extraction, low-quality frame exclusion, segment definition, and mean QUAD score calculation, is provided in Figure S1.

We compared the area under the curve (AUC) of the score derived from the full video (100%-QUAD score) with scores based on only the distal 50%, 40%, 30%, 20%, 10%, and 5% of the video (50%-QUAD score through 5%-QUAD score, respectively) in predicting clinical relapse within 24 months, using the cohort from the second phase.

2.3. Statistical analysis

Statistical analysis of the primary endpoint was performed using a per-protocol set that excluded patients with missing data. All statistical analyses were performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria). Quantitative data were expressed as median (interquartile range) for continuous variables, and statistical significance was assessed using Fisher's exact test, Student's t-test, or the Mann-Whitney U test, as appropriate. A two-sided *P*-value of <.05 was considered statistically significant. The required sample size was estimated with reference to our previous study using AI-assisted colonoscopy. We assumed 24-month clinical relapse rates of 5% for patients with a QUAD score of <4 and 30% for those with a QUAD score of ≥4, with a two-sided significance level of 0.05 and a power of 80%. The enrollment ratio of 1:2 for a QUAD score of <4 versus ≥4 was based on an unpublished preliminary validation study using consecutive cases from our hospital. This calculation indicated that 45 and 90 patients would be required in the <4 and ≥4 groups, respectively, at a 1:2 ratio. Allowing for patient dropouts, we determined that 160 participants would be needed.

2.4. Research enrollment and ethics

All patients provided informed consent for both the procedures and participation in the study. The Ethics Committee of Showa Medical University Northern Yokohama Hospital approved the study protocol (No. 21H018). The study was registered in the clinical trial registry of the University Hospital Medical Information Network (UMIN000049559) and conducted in accordance with the guidelines of the Declaration of Helsinki. All authors had full access to the study data and reviewed and approved the final manuscript.

3. Results

3.1. Model performance and correlation with endoscopic severity

In the first phase, 2272 images from 46 patients were prospectively collected. Image quality assessment and exclusion of non-analyzable frames were performed automatically by the predefined AI pipeline. Of these, 1009 images met the

predefined quality criteria and were included in the analysis. QUAD scores showed a strong correlation with MES (Spearman $r=0.809$, $P<.01$) (Figure 3).

3.2. Prospective cohort characteristics

Among 252 screened patients, 156 (median age, 51 years; 49% male) were included in the prospective validation cohort (Figure 2). Baseline characteristics are shown in Table 1. During follow-up, clinical relapse occurred in 22 patients.

3.3. Primary analysis: still image-based relapse prediction

Still image-based QUAD scoring predicted clinical relapse within 24 months. Relapse occurred in 19.4% (19/98) of patients with QUAD ≥4 compared with 5.2% (3/58) of those with QUAD <4 ($P=.01$). The AUC for relapse prediction was 0.67 (95% confidence interval [CI]: 0.57-0.76), with a sensitivity of 86.4% and specificity of 41.0%. Kaplan-Meier analysis demonstrated significant differences in relapse-free survival between groups (log-rank test, $P=.01$) (Figure 4a).

Histological assessment was available for all 156 patients in the prospective validation cohort. Histological activity, defined as a Geboes histological index ≥3.1, was significantly more frequent among patients with a worst still image-based QUAD score ≥4 than among those with a QUAD score <4 (31.6% [31/98] vs 15.5% [9/58]; $P=.04$).

To evaluate the prognostic value of QUAD in relation to established relapse predictors, we performed a multivariable Cox proportional hazards analysis including MES, histological activity, maximum disease extent, use of biologics or Janus kinase (JAK) inhibitors, and the still image-based QUAD score. In this model, a worst MES ≥1 showed a trend toward association with clinical relapse (hazard ratio [HR], 3.22; 95% CI, 0.93-11.19; $P=.07$). Similarly, a worst still image-based QUAD score ≥4 showed a comparable trend toward association with relapse (HR, 3.16; 95% CI, 0.91-10.97; $P=.07$). Histological activity, maximum disease extent, and use of biologics or JAK inhibitors were not significantly associated with relapse in this model. (Table 2).

3.4. Secondary exploratory analysis: video-based assessment of inflammatory extent

Still image-based and video-based analyses were evaluated separately because the former depended on endoscopist-selected images, whereas the latter used automated frame extraction from withdrawal videos. Automated full-length video analysis was performed to evaluate whether incorporating inflammatory extent improved relapse prediction. Full-length (100%) video analysis yielded an AUC of 0.62 (95% CI: 0.48-0.76), with a sensitivity of 63.6% and specificity of 61.2%. Segmental analyses revealed progressively improved predictive performance with increasing distal focus. Analyses restricted to the distal 50%, 40%, 30%, and 20% segments showed stepwise increases in discrimination (AUC range, 0.63-0.69) (Table 3). Predictive performance peaked at the distal 10% segment, which achieved the highest accuracy for relapse prediction (AUC, 0.73; 95% CI: 0.62-0.85; sensitivity, 72.7%; specificity, 70.9%). Further restriction to the distal 5% segment resulted in a slight decrease in performance (AUC, 0.71; 95% CI: 0.59-0.82), suggesting the presence of an optimal distal assessment window rather than a simple linear improvement with

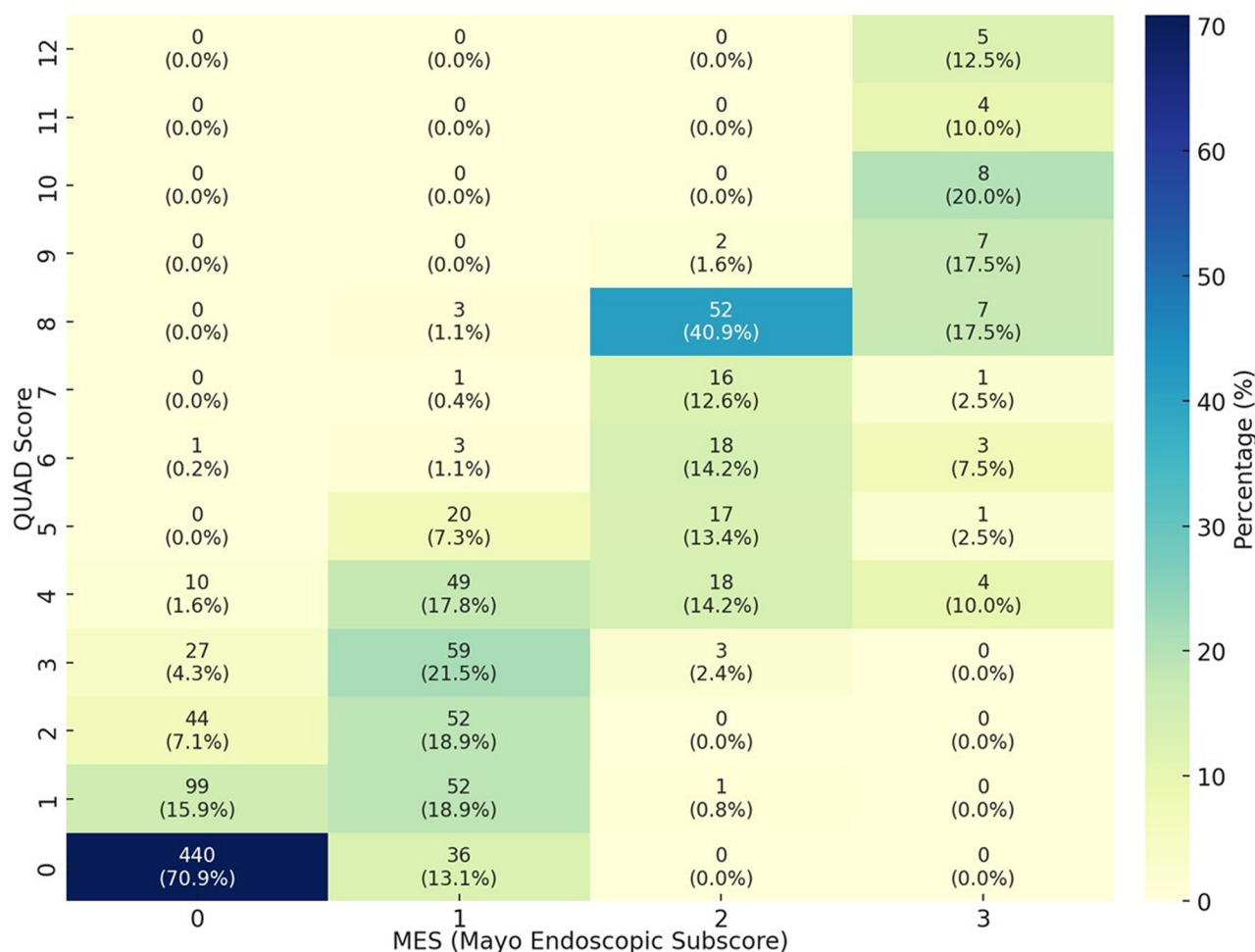


Figure 3. Heatmap of QUAD score distribution across MES categories. Each cell displays the number of cases and the corresponding percentage within each MES column. The vertical axis represents the QUAD score (0-12), with increasing severity from bottom to top. The color gradient reflects the relative proportion of each QUAD score within the respective MES group. MES, Mayo Endoscopic Subscore; QUAD, Quantitative Ulcerative Colitis Assessment using Deep Learning.

increasing distal restriction. Overall, incorporation of distal inflammatory extent improved relapse discrimination beyond both whole-colon video assessment and still image-based evaluation.

4. Discussion

In this prospective study, we demonstrated that incorporating inflammatory extent into AI-based endoscopic assessment improves relapse risk stratification in UC. Compared with conventional severity-focused evaluation, analysis incorporating inflammatory extent provided higher predictive performance. Notably, distal inflammatory burden showed the strongest predictive performance in this cohort, suggesting that quantification of inflammatory extent may provide prognostic information beyond peak severity alone.

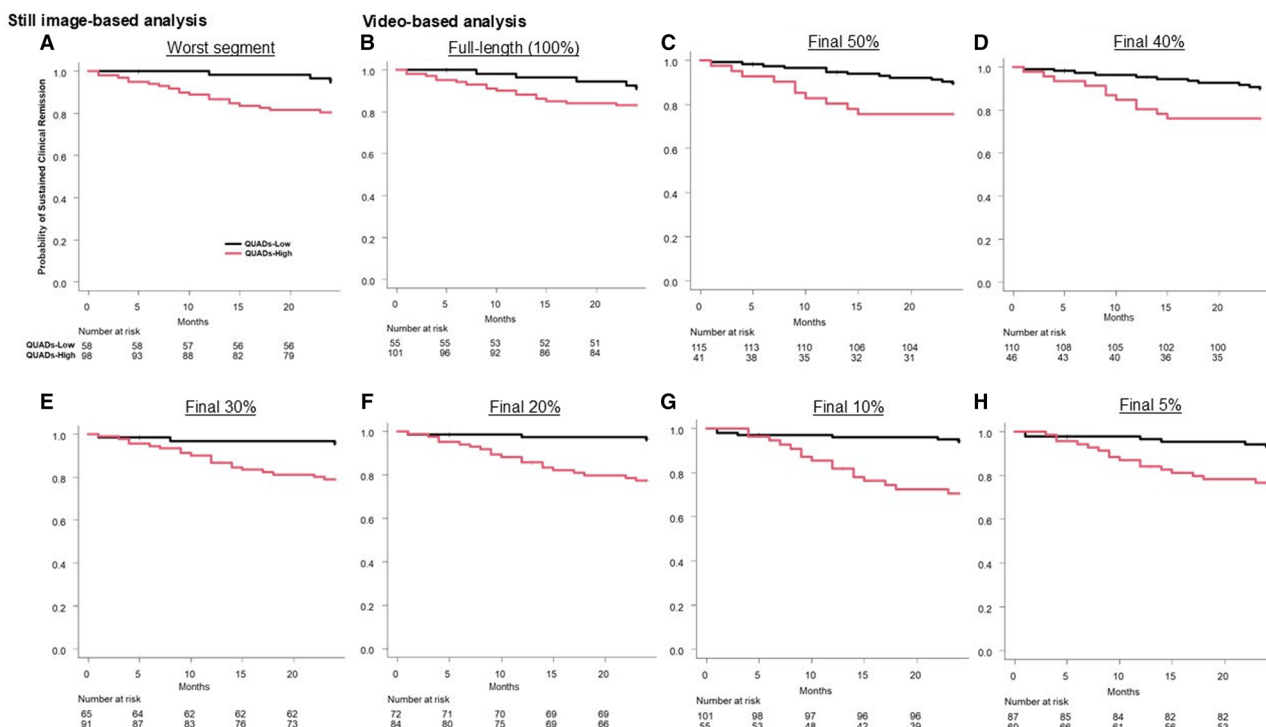
Our study showed that relapse prediction accuracy was higher when focusing on the distal colon rather than the entire colon. Specifically, analysis of the final 10% of the full-length colonoscopic video—corresponding to the distal sigmoid colon and rectum—outperformed conventional evaluations based on severity alone and whole-colon analysis. The AUC for this distal segment analysis was 0.73, compared with 0.68 for the single

highest still-image score and 0.62 for whole-colon analysis. Although AI-based models have been recognized for enhancing objectivity in image interpretation,^{15–18} the subjectivity of image acquisition by endoscopists remains a critical challenge.^{19,20} Full-length video analysis may reduce, but does not eliminate, the frame-selection bias associated with endoscopist-selected still images because withdrawal technique, inspection time, mucosal cleaning, and image acquisition practices were not fully standardized.²¹ The cohort was not restricted to left-sided disease because 60.9% of patients had extensive colitis, suggesting that the superior predictive value of distal assessment was not driven solely by disease-distribution bias; however, the biological and technical explanations underlying this observation remain uncertain and should be interpreted cautiously. One possible explanation is that residual inflammation in the distal colon, particularly the distal sigmoid colon and rectum, may represent a sensitive marker of incomplete mucosal healing. By contrast, averaging inflammatory activity across the entire colon may dilute localized signals associated with subsequent relapse. Additionally, rectosigmoid segments are generally visualized more consistently during colonoscopy, which may improve the robustness of AI-based assessment. These hypotheses require further investigation and validation in independent cohorts.

Table 1. Patient background characteristics.

	Overall (n=156)	Non-relapse (n=134)	Relapse (n=22)	P-values
Male, n (%)	76 (48.7)	67 (50.0)	9 (40.9)	.50
Age, years, median (IQR)	51 (44-63)	52 (44-64)	50 (41-55)	.67
Duration of disease, years, median (IQR)	14 (7-19)	14 (8-20)	10 (7-19)	.12
Extent of disease, n (%)				.07
Extensive colitis	95 (60.9)	78 (58.2)	17 (77.3)	
Left-sided colitis	39 (25.0)	34 (25.4)	5 (22.7)	
Proctitis	22 (14.1)	22 (16.4)	0	
Mayo Endoscopic Subscore, n (%)				<.01
0	72 (46.2)	69 (51.4)	3 (13.6)	
1	65 (41.7)	53 (39.6)	12 (54.6)	
2-3	19 (12.1)	12 (9.0)	7 (31.8)	
Geboes histological index				.03
<3.1	116 (74.4)	104 (77.6)	12 (54.6)	
≥3.1	40 (25.6)	30 (22.4)	10 (45.6)	
Concomitant therapy, n (%)				>.95
Oral 5-aminosalicylic acid	131 (84.0)	110 (82.1)	18 (81.8)	
Topical 5-aminosalicylic acid	22 (14.1)	16 (11.9)	6 (27.2)	.09
Immunomodulator	24 (15.4)	18 (13.4)	6 (27.3)	.11
Biological agent	42 (26.9)	34 (25.4)	8 (36.4)	.31
Serum evaluation, median (IQR)				
White blood cell count, /μL	5630 (4750-6770)	5680 (4785-6770)	5445 (4760-6455)	.74
Hemoglobin, g/dL	14.1 (12.7-15.3)	14.2 (12.7-15.4)	13.8 (12.6-14.6)	.62
Albumin, g/dL	4.5 (4.2-4.7)	4.5 (4.2-4.7)	4.3 (4.1-4.7)	.13
C-reactive protein, mg/dL	0.06 (0.04-0.16)	0.06 (0.04-0.16)	0.07(0.05-0.30)	.33

Abbreviation: IQR, interquartile range.



This observation is biologically plausible because distal residual inflammation may represent a sensitive marker of persistent disease activity and has been linked to subsequent relapse in prior studies.²²⁻²⁵ These findings extend previous work evaluating

cumulative inflammatory burden and AI-assisted endoscopic scoring systems in UC. Previous studies have focused primarily on lesion-level classification or treatment-response assessment,^{12,13} whereas our study addresses relapse prediction in patients already

in clinical remission, representing a clinically relevant unmet need in treat-to-target management strategies. Within this evolving framework of AI-assisted endoscopic assessment, the aim of this study was not to replace conventional MES scoring but to explore whether incorporating inflammatory extent provides complementary prognostic information beyond severity-focused assessment. However, the predictive performance observed in this study was moderate (AUC, 0.67-0.73), suggesting that QUAD should be considered a complementary prognostic marker rather than a standalone clinical decision-making tool. Further integration with clinical, biochemical, and histologic markers, along with validation in independent cohorts, will be necessary before widespread clinical implementation.

Stidham et al.¹² recently developed a video-based cumulative score that quantifies mucosal disease activity from the descending colon to the rectum, enabling a more efficient assessment of therapeutic response than the conventional MES. A major

methodological distinction is that unlike the previous study, which focused on treatment response in patients experiencing clinical relapse, the still image-based component of the present study was designed to investigate clinical relapse exclusively in patients who were in clinical remission. Gutierrez-Becker et al.¹³ subsequently used full-colon datasets to create an AI scoring system that provides a spatially resolved evaluation of disease severity. To our knowledge, the present study is the first to investigate whether these novel indices are useful for predicting relapse during routine monitoring.

These findings highlight practical considerations for the clinical implementation of AI-based scoring. Reliable AI performance depends on high-quality colonoscopic imaging,^{14,26,27} and our system may further support image-quality standardization through real-time feedback mechanisms, although this requires further evaluation in future studies.

Complementing the concept of inflammatory extent across the colon, another distinctive feature of our model is its ability to capture patchy inflammatory patterns within a single endoscopic view, which may be particularly relevant in treated UC, where residual inflammation often becomes heterogeneous. Figure 5 illustrates images from the first component of this study that were interpreted by endoscopists as MES 3, with corresponding QUAD scores of 5, 8, and 11. However, when MES and QUAD were included in the same multivariable model, neither variable reached statistical significance, although both showed similar effect sizes. This finding suggests that QUAD and MES may capture partly overlapping aspects of residual endoscopic inflammation. In addition, the limited number of relapse events may have reduced the statistical

Table 2. Multivariable Cox proportional hazards analysis for clinical relapse within 24 months after colonoscopy.

Variable	Hazard ratio	95% CI	P-value
MES≥1	3.22	0.93-11.20	.07
Still image-based QUAD score≥4	3.16	0.91-10.97	.07
Use of biologics or JAK inhibitors	1.05	0.43-2.55	.92
Geboes histological index≥3.1	2.29	0.84-6.23	.11
Disease extent (extensive colitis)	1.59	0.66-3.83	.30

Abbreviations: CI, confidence interval; JAK, Janus kinase; MES, Mayo Endoscopic Subscore; QUAD, Quantitative Ulcerative Colitis Assessment using Deep Learning.

Table 3. Diagnostic performance for predicting clinical relapse within 24 months after colonoscopy.

	Area under the curve	95% confidence interval	QUADs cutoff value	Sensitivity (%)	Specificity (%)
Still image-based analysis	0.67	0.57-0.76	4	86.4	41.0
Video-based analysis					
Full-length (100%)	0.62	0.48-0.76	1.56	63.6	61.2
Final 50%	0.63	0.50-0.76	2.35	50.0	73.9
Final 40%	0.65	0.53-0.78	2.66	50.0	76.9
Final 30%	0.67	0.55-0.78	1.70	86.4	46.3
Final 20%	0.69	0.58-0.81	2.11	86.4	51.5
Final 10%	0.73	0.62-0.85	3.36	72.7	70.9
Final 5%	0.71	0.59-0.82	3.20	72.7	60.4

Abbreviation: QUADs, Quantitative Ulcerative Colitis Assessment using Deep Learning score.

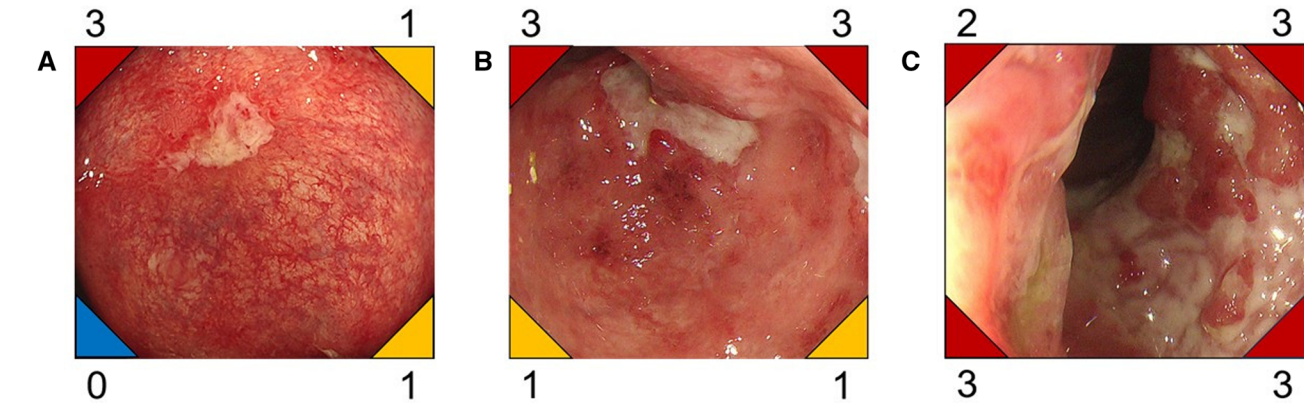


Figure 5. Representative images generated by the Quantitative Ulcerative Colitis Assessment using Deep Learning (QUAD) algorithm. (A) QUAD score 5, (B) QUAD score 8, and (C) QUAD score 11.

power to distinguish the independent contribution of each endoscopic predictor. Therefore, the incremental value of QUAD beyond conventional endoscopic indices should be further evaluated in larger independent cohorts.

Several limitations should be acknowledged. Although multicenter data were used for model development, prospective validation was performed at a single specialized center, which may limit generalizability. The relatively small number of relapse events may limit the precision and stability of effect estimates and increase uncertainty around the observed predictive performance. Nevertheless, the use of standard white-light colonoscopy suggests potential applicability to routine clinical practice. Therapeutic strategies were not standardized and may have influenced relapse risk. Fecal calprotectin was not systematically available and could not be included in the multivariable model; therefore, the added value of QUAD in combination with biochemical markers requires further investigation. In addition, reliance on image quality may introduce bias despite automated filtering. Therefore, external validation in independent cohorts from different institutions and healthcare systems will be necessary to establish generalizability. Future prospective studies, including randomized comparisons of AI-assisted and standard endoscopic assessment strategies, are warranted to determine whether AI-guided risk stratification improves long-term clinical outcomes.

In conclusion, incorporating inflammatory extent into endoscopic assessment may enhance relapse risk stratification beyond conventional severity-focused evaluation. Our findings suggest that distal inflammatory burden represents a clinically relevant target for AI-assisted analysis during routine colonoscopy. However, given the moderate predictive performance observed, integration with clinical and non-endoscopic factors, along with validation in larger independent cohorts, will be essential before widespread clinical implementation.

Acknowledgments

We thank Angela Morben, DVM, ELS, from Edanz (<https://jp.edanz.com/ac>), for editing a draft of the manuscript.

Author contributions

Conceptualization: Y.Ma, M.M.; data collection: Y.Ma, N.O., Ke.T., Ka.T., Y.Ka., T.Ku., T.O., T.S., Y.Ko., K.I., T.H., T.B., H.O., K.O., M.M.; writing—original draft preparation: Y.Ma; writing—review and editing: S.K., N.O., Ke.T., Ka.T., H.O., K.O., Y.Mo., M.I., M.M.; supervision: S.K., H.O., K.O., M.I. All authors have read and agreed to the published version of the manuscript.

Supplementary material

Supplementary material is available at ECCO-JCC online.

Funding

This work was supported by a grant from JSPS KAKENHI [grant number JP23K09537], the Takeda Science Foundation, and The Japanese Foundation for Research and Promotion of Endoscopy Grant.

Conflicts of interest

S.K., Yu.M., and M.M.: consultant and speaker fees from Olympus Corp.; ownership interest in Cybernet Systems. K.O.: consultant and speaker fees from Olympus.

Patient and public involvement

Patients and/or the public were involved in the design, conduct, reporting, or dissemination plans of this research.

Patient consent for publication

Not applicable.

Ethical approval

The Ethics Committee of Showa Medical University Northern Yokohama Hospital approved the study protocol (No. 21H0003).

Data availability

The data supporting the findings of this study are not publicly available because of ethical and privacy restrictions, as participants did not consent to public data sharing.

References

1. Gros B, Kaplan GG. Ulcerative colitis in adults: a review. *JAMA*. 2023;330:951-965.
2. Rubin DT, Ananthakrishnan AN, Siegel CA, et al. ACG clinical guideline update: ulcerative colitis in adults. *Am J Gastroenterol*. 2025;120:1187-1224.
3. Manginot C, Baumann C, Peyrin-Biroulet L. An endoscopic Mayo score of 0 is associated with a lower risk of colectomy than a score of 1 in ulcerative colitis. *Gut*. 2015;64:1181.2-1182.
4. Levesque BG, Sandborn WJ, Ruel J, et al. Converging goals of treatment of inflammatory bowel disease from clinical trials and practice. *Gastroenterology*. 2015;148:37-51 e1.
5. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: an update on the selecting therapeutic targets in inflammatory bowel disease (STRIDE) initiative of the International Organization for the Study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. *Gastroenterology*. 2021;160:1570-1583.
6. Neurath MF, Vieth M. Different levels of healing in inflammatory bowel diseases: mucosal, histological, transmural, barrier and complete healing. *Gut*. 2023;72:2164-2183.
7. Schroeder KW, Tremaine WJ, Ilstrup DM, et al. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med*. 1987;317:1625-1629.
8. Kaneshiro M, Takenaka K, Suzuki K, et al. Pancolonic endoscopic and histologic evaluation for relapse prediction in patients with ulcerative colitis in clinical remission. *Aliment Pharmacol Ther*. 2021;53:900-907.
9. Rowan CR, Cullen G, Mulcahy HE, et al. DUBLIN [degree of ulcerative colitis burden of luminal inflammation] score, a simple method to quantify inflammatory burden in ulcerative colitis. *J Crohns Colitis*. 2019;13:1365-1371.
10. 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.

11. Ogata N, Maeda Y, Misawa M, et al. Artificial intelligence-assisted video colonoscopy for disease monitoring of ulcerative colitis: a prospective study. *J Crohns Colitis*. 2025;19:jjae080.
12. Stridham RW, Cai L, Cheng S, et al. Using computer vision to improve endoscopic disease quantification in therapeutic clinical trials of ulcerative colitis. *Gastroenterology*. 2024;166:155-167.e2.
13. Gutierrez-Becker B, Fraessle S, Yao H, et al. Ulcerative Colitis Severity Classification And Localized Extent (UC-SCALE): an artificial intelligence scoring system for a spatial assessment of disease severity in ulcerative colitis. *J Crohns Colitis*. 2025;19:jjae187.
14. Maeda Y, Kudo SE, Santacroce G, et al. Artificial intelligence-assisted colonoscopy to identify histologic remission and predict the outcomes of patients with ulcerative colitis: a systematic review. *Dig Liver Dis*. 2024;56:1119-1125.
15. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc*. 2019;89:408-415.
16. 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;7:230-237.
17. 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.
18. Lo B, Liu Z, Bendtsen F, Igel C, Vind I, Burisch J. High accuracy in classifying endoscopic severity in ulcerative colitis using convolutional neural network. *Am J Gastroenterol*. 2022;117:1648-1654.
19. Gregorio V, Maeda Y, Kudo SE, et al. Evolving role of artificial intelligence in endoscopic management of inflammatory bowel disease: diagnosis, surveillance, and assessment. *Dig Endosc*. 2025;37:1148-1161.
20. Iacucci M, Santacroce G, Maeda Y, et al. Artificial intelligence-driven personalized medicine: transforming clinical practice in inflammatory bowel disease. *Gastroenterology*. 2025;169:416-431.
21. Maeda Y, Kudo S-e, Kuroki T, et al. A scoping review of endoscopic assessment considering disease extent in ulcerative colitis: insights for the artificial intelligence era. *Gut Liver*. 2026;20:511-523.
22. Bitton A, Peppercorn MA, Antonioli DA, et al. Clinical, biological, and histologic parameters as predictors of relapse in ulcerative colitis. *Gastroenterology*. 2001;120:13-20.
23. Nishio Y, Ando T, Maeda O, et al. Pit patterns in rectal mucosa assessed by magnifying colonoscope are predictive of relapse in patients with quiescent ulcerative colitis. *Gut*. 2006;55:1768-1773.
24. Ohara J, Maeda Y, Ogata N, et al. Automated neutrophil quantification and histological score estimation in ulcerative colitis. *Clin Gastroenterol Hepatol*. 2025;23:846-854.e7.
25. Kuroki T, Maeda Y, Kudo SE, 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;100:97-108.
26. Kuroki T, Maeda Y, Kudo SE, et al. Combination of white-light imaging based and narrow-band imaging based artificial intelligence models during colonoscopy in patients with ulcerative colitis. *J Crohns Colitis*. 2025;19:jjaf014.
27. 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;9:758-772.