Digestive Endoscopy

WILEY

ORIGINAL ARTICLE OPEN ACCESS

A Novel Switching of Artificial Intelligence to Generate Simultaneously Multimodal Images to Assess Inflammation and Predict Outcomes in Ulcerative Colitis—(With Video)

Marietta Iacucci[1,2] | Irene Zammarchi[1] | Giovanni Santacroce[1] | Bisi Bode Kolawole[3] | Ujwala Chaudhari[3] | Rocio del Amor[4] | Pablo Meseguer[4,5] | Valery Naranjo[4] | Miguel Puga-Tejada[1] | Ivan Capobianco[1] | Ilaria Ditonno[1] | Andrea Buda[6] | Brian Hayes[7] | Rory Crotty[7] | Raf Bisschops[8] | Subrata Ghosh[1] | Enrico Grisan[3] | on behalf of the PICaSSO group

[1]APC Microbiome Ireland, College of Medicine and Health, University College of Cork, Cork, Ireland | [2]University of Birmingham, Institute of Immunology and Immunotherapy, Birmingham, UK | [3]School of Engineering Computer Science and Informatics, London South Bank University, London, UK | [4]Instituto de Investigación e Innovación en Bioingeniería, HUMAN-Tech, Universitat Politècnica de València, València, Spain | [5]valgrAI—Valencian Graduate School and Research Network of Artificial Intelligence, Valencia, Spain | [6]Department of Gastrointestinal Oncological Surgery, Santa Maria del Prato Hospital, Feltre, Italy | [7]Department of Histopathology, Cork University Hospital, Cork, Ireland | [8]Division of Gastroenterology, University Hospitals Leuven, Leuven, Belgium

Correspondence: Marietta Iacucci (miacucci@ucc.ie)

Received: 24 January 2025 | **Revised:** 19 April 2025 | **Accepted:** 27 May 2025

Funding: The authors received no specific funding for this work.

Keywords: artificial intelligence | ulcerative colitis | virtual chromoendoscopy

ABSTRACT

Objectives: Virtual Chromoendoscopy (VCE) is pivotal for assessing activity and predicting outcomes in Ulcerative Colitis (UC), though interobserver variability and the need for expertise persist. Artificial intelligence (AI) offers standardized VCE-based assessment. This study introduces a novel AI model to detect and simultaneously generate various endoscopic modalities, enhancing AI-driven inflammation assessment and outcome prediction in UC.

Methods: Endoscopic videos in high-definition white-light, iScan2, iScan3, and NBI from UC patients of the international PICaSSO iScan and NBI cohort (302 and 54 patients, respectively) were used to develop a neural network to identify the acquisition modality of each frame and for inter-modality image switching. 2535 frames from 169 videos of the iScan cohort were switched to different modalities and trained a deep-learning model for inflammation assessment. Subsequently, the model was tested on a subset of the iScan and NBI cohorts (72 and 51 videos, respectively). Performance in predicting endoscopic and histological activity and outcomes was evaluated.

Results: The model efficiently classified and converted images across modalities (92% accuracy). Performance in predicting endoscopic and histological remission was excellent, especially with different modalities combined in both iScan (accuracy 81.3%

Abbreviations: AI, artificial intelligence; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; DOR, diagnostic odds ratio; HD-WLE, high-definition white-light; HR, hazard ratio; IBD, inflammatory bowel disease; IQR, interquartile range; MCC, Matthews correlation coefficient; NBI, narrow band imaging; NHI, Nancy histological index; NN, neural network; NPV, negative predictive value; PHRI, PICaSSO Histologic Remission Index; PICaSSO, Paddington International Virtual Chromoendoscopy Score; PPV, positive predictive value; RHI, Robarts Histopathology Index; SD, standard deviation; UC, ulcerative colitis; UCEIS, ulcerative colitis endoscopic index of severity; VCE, virtual chromoendoscopy.

Irene Zammarchi, Giovanni Santacroce, Bisi Bode Kolawole and Ujwala Chaudhari contributed equally to the manuscript.

The PICaSSO group members are listed in Acknowledgments section.

and 89.6%; AUROC 0.92 and 0.89 by UCEIS and PICaSSO, respectively) and the NBI cohort. Moreover, it showed a remarkable ability in predicting clinical outcomes.

Conclusions: Our multimodal "AI-switching" model innovatively detects and transitions between different endoscopic modalities, refining inflammation assessment and outcome prediction in UC by integrating model-derived images.

1 | Introduction

Endoscopic remission (ER) is the primary target in Ulcerative Colitis (UC), guiding disease management [1]. Endoscopic assessment routinely relies on white-light endoscopy (WLE), which faces challenges in differentiating between mild-patchy activity/remission and accurately assessing vascular changes [2]. Therefore, virtual chromoendoscopy (VCE) has been introduced to enhance disease assessment by highlighting subtle mucosal and vascular features suggestive of minimal inflammation [3]. VCE has proven more effective in predicting disease activity/remission and outcomes, showing a stronger correlation with histology [4]. Notably, VCE-enabled vascular assessment through Narrow-Band Imaging (NBI) has shown a remarkable ability to predict relapse risk in patients in clinical remission [5].

Different VCE platforms have been developed. The iScan platform (Pentax, Tokyo, Japan) offers various modalities: iScan1 sharpens surface vessels and enhances mucosal texture, iScan2 increases the contrast between mucosa and vessels, and iScan3 enhances vessel visibility, including those in dimly illuminated far-field regions [6]. Similarly, the NBI platform (Olympus, Tokyo, Japan) maximizes contrast between vessels and surrounding mucosa, improving assessment of microvessel architecture [7].

Efforts have been made to standardize disease assessment through VCE using appropriate scores, including the PICaSSO (Paddington International Virtual Chromoendoscopy Score), which was developed on the iScan platform [8], later reproduced on the NBI platform and validated in a multicentre study [9, 10]. Despite its better correlation with histology and effective outcome prediction compared to WLE-based scores, it is not widely adopted due to a lack of training and high inter- and intra-observer variability.

Assessing disease activity through VCE requires experience and training [11, 12]. Artificial intelligence (AI) can standardize disease assessment, offering objective evaluation of ER and outcome prediction [13, 14]. Our group pioneered an AI model applied to WLE and VCE, enabling accurate disease assessment through the Ulcerative Colitis Endoscopic Index of Severity (UCEIS) and PICaSSO score. This model effectively distinguished disease activity from remission, remarkably predicting clinical outcomes [15].

A comprehensive evaluation of disease activity, combining WLE and multiple VCE modalities, could be promising for a deeper disease assessment. Hence, this study aims to develop a new AI switching model to simultaneously detect and convert different endoscopic modalities (HD-WLE, iScan2, iScan3, and NBI), obtaining simultaneous multiple endoscopic modalities during a single procedure. Therefore, we evaluated the model's ability to predict endoscopic and histologic activity and clinical outcomes in different cohorts through different endoscopic modalities.

2 | Materials and Methods

2.1 | Ethical Approval

Videos were extracted from studies approved by the West Midlands Research Ethics Committee (17/WM/0223), Research Ethics Committee Northern Ireland (17/NI/0148) for the UK, ethics approval n.2678, and CARMS n.14392 for the UK and local competent committees for other international centers. All patients gave informed consent to participate in the study.

2.2 | Patient Selection

The AI switching model's training and first testing phase involved videos from UC patients prospectively enrolled from 11 international centres (iScan cohort) [10]. The model was further tested with videos from an external cohort of UC patients prospectively enrolled at the University of Birmingham (NBI cohort) [9].

Ulcerative Colitis patients scheduled for colonoscopy were enrolled. Exclusion criteria encompassed contraindications to procedure/biopsies, inability to provide consent, and inadequate bowel preparation.

2.3 | Endoscopic Assessment

In the iScan cohort, patients underwent assessment through HD-WLE, iScan2, and iScan3 (EPK-i7010 and EPK-i8020c processors, HiLine series colonoscopes, Pentax, Tokyo, Japan). Patients in the NBI cohort underwent assessment with HD-WLE and NBI (CV-190 processors, EVIS EXERA III 190 series colonoscopes, Olympus, Tokyo, Japan).

Experienced endoscopists graded disease activity by UCEIS for HD-WLE and PICaSSO for VCE. ER was defined by UCEIS ≤ 1 [16] and PICaSSO ≤ 3 [8].

2.4 | Histological Assessment

Biopsies were performed to grade disease activity. Inflammatory bowel disease (IBD)-experienced pathologists assessed activity through the Robarts Histopathology Index (RHI) [17], Nancy Histological Index (NHI) [18], and PICaSSO Histologic

Remission Index (PHRI) [19]. Histological remission was defined by RHI < 3 with absence of neutrophils in epithelium and lamina propria, NHI ≤ 1, and PHRI = 0.

2.5 | Follow-Up and Outcome

Clinical outcomes at 12-month follow-up were assessed (flare-ups, treatment initiation/change [steroids, immunomodulators, biological agents, oral targeted therapies], hospitalization, and colectomy).

2.6 | AI-Switching Model Development

2.6.1 | Phase 1—Detection and Conversion of VCE Modalities

A novel AI-switching model was developed to detect and convert images across different endoscopic modalities. HD-WLE, iScan, and NBI high-quality videos with endoscopic activity/remission were assessed by UCEIS and PICaSSO, respectively. We developed a neural network (NN) to identify each frame's acquisition modality, starting from 2535 extracted frames of HD-WLE, iScan, and NBI videos, and then trained a cycle-GAN model using 900 images from different modalities to allow inter-modality image switching. Subsequently, our previously developed deep-learning model [15] was trained using images obtained from the AI switching model to assess inflammation. The model was then trained to generate iScan2, iScan3, and NBI modalities starting from HD-WLE images. One expert endoscopist evaluated a set of image pairs, one original and one generated, for validation and to assess the fidelity of AI-generated images. Images were correctly classified in 55.33% of cases.

Models for both single modalities (HD-WLE, iScan2, iScan3, NBI) and multimodalities (HD-WLE + iScan2 + iScan 3; HD-WLE + NBI; HD-WLE + iScan2 + iScan 3 + NBI) were trained.

Separate training of each model was conducted to predict endoscopic or histologic scores. Appendix S1 and Figure S1 provide details on model development.

2.6.2 | Phase 2—Assessment of Inflammation/Remission and Outcome Prediction

The model was initially tested on 72 videos (1080 frames) of the iScan cohort and further tested on 51 videos (765 frames) of the NBI cohort. Diagnostic performance in assessing disease activity and predicting outcome was evaluated, as was the agreement with experts.

The model was developed from high-quality videos (absence of motion artifacts, stools, and adequate mucosal visualization). Validation and test sets were randomly selected from the entire dataset to ensure unbiased performance assessment. 5-fold cross-validation was performed to prevent overfitting.

2.7 | Study Objectives

The primary objective was to develop a novel AI switching model to simultaneously detect different VCE modalities, converting HD-WLE into iScan2, iScan3, and NBI images.

Secondary objectives were to evaluate the model's ability to predict endoscopic and histologic remission and clinical outcomes using generated unimodal and multimodal images in different UC cohorts where different endoscopic platforms were initially used. Moreover, the model's performance in predicting activity/remission against humans was assessed.

2.8 | Statistical Analysis

The CAD system developed by Iacucci et al. for detecting histological activity as PHRI > 0 [19] has 92% sensitivity, alpha, beta, and delta values of 5%, 20%, and 10%, respectively. Therefore, a sample of at least 150 cases was necessary to assess the AI switching model.

Continuous variables were described as means (standard deviation, SD) or medians (interquartile range, IQR) in agreement with the Gaussian distribution (Kolmogorov–Smirnov test). Categorical variables were described in percentages and 95% confidence intervals (CI) when corresponding.

Diagnostic performance of the model to predict remission according to endoscopic and histological scores was calculated as sensitivity (or recall), specificity, positive predictive value (PPV or precision), negative predictive value (NPV), observed agreement (accuracy), F1 score, Matthews Correlation Coefficient (MCC), area under the receiver operating characteristic curve (AUROC), and diagnostic Odds Ratio (DOR)—the equality of proportions hypothesis test and DeLong's test compared multimodal probabilities and AUROC versus each unimodal. The agreement between human and AI prediction was evaluated through Cohen's kappa. Kaplan–Meier curves for survival free from adverse outcomes were compared through the Mantel-Cox log-rank test, and corresponding hazard ratios (HR) to define effect size were computed. A p-value < 0.05 was statistically significant. Analysis was performed with R v4.3 (R Foundation for Statistical Computing; Vienna, Austria).

3 | Results

Table 1 details patients' demographic and disease characteristics.

Of 302 UC patients from the iScan cohort, 182 (59.3%) were males, averaging 48.4 years (SD 14.8). 69.2% and 72.8% of patients were in the ER by UCEIS and PICaSSO, respectively.

Of 54 UC patients from the NBI cohort, 32 (59%) were males, averaging 40 years (SD 15.3). 42.6% and 51.9% of patients were in the ER by UCEIS and PICaSSO, respectively.

TABLE 1 | Demographic characteristics of the iScan and NBI cohorts.

Characteristics	iScan cohort	NBI cohort
Patients, *N*	302	54
Age (years) mean ± SD	48.4 ± 14.8	40 ± 15.3
Gender male, *N* (%)	182 (59.3)	32 (59)
Disease duration (years) mean ± SD	15.0 ± 10.8	10.6 ± 7.9
Disease extension, *N* (%)		
Left-sided colitis	130 (43)	10 (18.5)
Sub-total or total colitis	172 (57)	44 (81.5)
Endoscopic activity		
UCEIS, *N* (%)		
Remission (≤ 1)	209 (69.2)	23 (42.6)
Activity (> 1)	93 (30.8)	31 (57.4)
PICaSSO, *N* (%)		
Remission (≤ 3)	220 (72.8)	28 (51.9)
Activity (> 3)	82 (27.2)	26 (48.1)

Abbreviations: NBI, narrow-band imaging; PICaSSO, Paddington International Virtual Chromoendoscopy Score; SD, standard deviation; UCEIS, ulcerative colitis endoscopic index of severity.

3.1 | AI Switching Model Performance

3.1.1 | Phase 1—Detection and Conversion of Different Endoscopic Modalities

The model classified and converted images across different endoscopic modalities with 92% accuracy of the NN classified on the test set.

3.1.2 | Phase 2

3.1.2.1 | Model Performance-iScan Cohort

3.1.2.1.1 | Assessment of Activity/Remission and Agreement With Human. The model showed remarkable diagnostic performance in assessing endoscopic and histological activity using single or all modalities' inputs, with the multimodal model for HD-WLE + iScan2 + iScan3 and HD-WLE + NBI outperforming corresponding unimodal models (Tables 2 and 3). Moreover, the full multimodal model (HD-WLE + iScan2 + iScan3 + NBI) showed the highest diagnostic accuracy in almost all predictions. It exhibited 81.3% (95% CI 75.8–86.0) and 89.6% (95% CI 85.1–93.2) accuracy and 0.92 and 0.89 AUROC in predicting ER by UCEIS and PICaSSO, respectively (Table 2 and Figure 1).

Similarly, the multimodal model achieved 90.5% (95% CI 86.0–93.9), 85.9% (95% CI 80.9–90.0), and 89.6% (95% CI 85.1–93.2) accuracy and 0.89, 0.92, and 0.92 AUROC in predicting histologic remission by RHI, NHI, and PHRI, respectively (Table 3 and Figure 1).

Agreement between the multimodal model and experienced endoscopists was good for UCEIS (K 0.66) and fair for PICaSSO (K 0.49). Similarly, agreement with experienced pathologists was good for RHI, NHI, and PHRI (K 0.78, 0.65, and 0.79, respectively).

3.1.2.1.2 | Outcome Prediction. 13 (18.1%) patients had an adverse outcome. The AI model outperformed human assessment in predicting clinical outcomes according to ER (hazard ratio 3.18 [95% CI 0.979–10.35, $p = 0.05$] and 1.89 [0.63–5.61, $p = 0.27$], respectively) and histological remission (HR 5.75 [95% CI 1.77–18.71, $p = 0.004$] and 2.15 [95% CI 0.66–6.97, $p = 0.22$], respectively) (Figure 2 and Table S1).

3.1.2.2 | Model Performance-NBI Cohort

3.1.2.2.1 | Assessment of Activity/Remission and Agreement With Human. The AI model showed remarkable diagnostic performance in predicting activity from unimodal and multimodal strategies (Tables S2 and S3), with the multimodal model for HD-WLE + iScan2 + iScan3 and HD-WLE + NBI outperforming corresponding unimodal models. Full multimodal assessment outperformed the unimodal one in detecting endoscopic activity through UCEIS and PICaSSO (accuracy 86.3% [95% CI 73.7.1–94.3] and 88.2% [95% CI 76.1–95.6] and AUC 0.92, respectively) (Table S2). Similarly, the multimodal model outperformed the unimodal model in detecting histological activity through RHI, NHI, and PHRI (Table S3 and Figure S2).

Agreement between the multimodal model and experienced endoscopists was excellent for UCEIS and PICaSSO (K 1). Similarly, agreement with experienced pathologists was excellent for RHI and NHI (K 0.92 and 1) and good for PHRI (K 0.74).

3.1.2.2.2 | Outcome Prediction. 21 (39%) patients had an adverse outcome. The model predicted survival-free from adverse outcomes according to ER similarly to experienced endoscopists (HR 1.7 [95% CI 0.7–4.11, $p = 0.24$]) and histological remission similarly to pathologists (HR 3.9 [95% CI 1.15–13.28, $p = 0.04$] and 5.33 [95% CI 1.24–22.94, $p = 0.02$]) (Figure S3 and Table S1).

4 | Discussion

We developed the first AI system to detect and convert images across different endoscopic modalities in IBD. Using AI-generated images, the model accurately assessed endoscopic and histologic activity/remission in the multicentre PICaSSO iScan [10] and NBI [9] cohort, achieving the best results when integrating a multimodal image approach. Furthermore, it demonstrated remarkable agreement with experts and predicted clinical outcomes in both cohorts. This model enabled the acquisition and integration of diverse images from multiple endoscopic modalities, holding potential for comprehensive IBD assessment. NBI uses narrow light bandwidths to enhance mucosal and vascular patterns, while iSCAN applies digital image post-processing to enhance structures [20]. Our model, trained

TABLE 2 | Diagnostic performance of the AI switching model for detection of endoscopic activity/remission—Iscan cohort ($n = 72$).

	Sensitivity	Specificity	PPV	NPV	Observed agreement	F1 score	MCC	AUROC	DOR
UCEIS									
HD-WLE	56/74; 75.7 (64.3–84.9)	142/167; 85.0 (78.7–90.1)	56/81; 69.1 (57.9–78.9)	142/160; 88.8 (82.8–93.2)	198/241; 82.2 (76.7–86.8)	112/155; 72.3 (64.5–79.1)	59.3	0.90	17.7 (9.0–34.9)
Iscan2	40/74; 54.1 (42.1–65.7)	135/167; 80.84 (74.0–86.5)	40/72; 55.6 (43.4–67.2)	135/169; 79.9 (73.0–85.7)	175/241; 72.61 (66.5–78.1)	80/146; 54.8 (46.4–63.0)	35.2	0.84	5.0 (2.7–9.0)
Iscan3	44/74; 59.46 (47.4–70.7)	139/167; 83.2 (76.7–88.6)	44/72; 61.1 (48.9–72.4)	139/169; 82.3 (75.6–87.7)	183/241; 75.9 (70.0–81.2)	88/146; 60.3 (51.9–68.3)	43.0	0.85	7.3 (3.9–13.5)
NBI	51/74; 68.9 (57.1–79.2)	132/167; 79.0 (72.1–85.0)	51/86; 59.3 (48.2–69.8)	132/155; 85.2 (78.6–90.4)	183/241; 75.9 (70.0–81.2)	102/160; 63.8 (55.8–71.2)	46.2	0.85	8.4 (4.5–15.5)
HD-WLE + NBI	53/74; 71.6 (60–81.5)	149/167; 89.2 (83.5–93.5)	53/71; 74.7 (62.9–84.2)	149/170; 87.7 (81.7–92.2)	202/241; 83.8 (78.6–88.2)	106/145; 73.1 (65.1–80.1)	61.6	0.90	20.9 (10.3–42.2)
HD-WLE + Iscan2 + Iscan3	57/74; 77.0 (65.8–86.0)	146/167; 87.4 (81.4–92.0)	57/78; 73.1 (61.8–82.5)	146/163; 89.6 (83.8–93.8)	203/241; 84.2 (79.0–88.6)	114/152; 75 (67.3–81.7)	63.5	0.91	23.3 (11.5–47.4)
HD-WLE + Iscan2 + Iscan3 + NBI	60/74; 81.1 (70.3–89.3)	136/167; 81.4 (74.7–87.0)	60/91; 65.9 (55.3–75.6)	136/150; 90.7 (84.8–94.8)	196/241; 81.3 (75.8–86.0)	120/165; 72.7 (65.3–79.4)	59.5	0.92	18.8 (9.3–37.9)
PICaSSO									
HD-WLE	44/56; 78.6 (65.6–88.4)	168/185; 90.8 (85.7–94.56)	44/61; 72.1 (59.2–82.9)	168/180; 93.3 (88.6–96.5)	212/241; 88 (83.2–91.8)	88/117; 75.2 (66.4–82.7)	67.4	0.85	36.2 (16.1–81.5)
Iscan2	37/56; 66.1 (52.2–78.2)	161/185; 87.03 (81.31–91.5)	37/61; 60.7 (47.3–72.9)	161/180; 89.44 (84.0–93.5)	198/241; 82.16 (76.7–86.8)	74/117; 63.3 (53.8–72)	51.6	0.75	13.1 (6.5–26.3)
Iscan3	39/56; 69.6 (56.0–81.2)	164/185; 88.7 (83.2–92.8)	39/60; 65 (51.6–76.9)	164/181; 90.6 (85.4–94.4)	203/241; 84.2 (79.0–88.6)	78/116; 67.2 (57.9–75.7)	56.9	0.80	17.9 (8.6–37.1)
NBI	35/56; 62.5 (48.6–75.1)	165/185; 89.2 (83.8–93.2)	35/55; 63.64 (49.6–76.2)	165/186; 88.7 (83.3–92.9)	200/241; 83.0 (77.6–87.5)	70/111; 63.1 (53.4–72.0)	52.0	0.83	13.8 (6.7–28.1)
HD-WLE + NBI	47/56; 83.9 (71.7–92.4)	157/185; 84.9 (78.9–89.7)	47/75; 62.7 (50.7–73.6)	157/166; 94.6 (90–97.5)	204/241; 84.7 (79.5–90)	94/131; 71.8 (63.2–79.3)	62.8	0.88	29.2 (12.9–66.4)
HD-WLE + Iscan2 + Iscan3	44/56; 78.6 (65.6–88.4)	173/185; 93.5 (88.9–96.6)	44/56; 78.6 (65.6–88.4)	173/185; 93.5 (88.9–96.6)	217/241; 90.0 (85.6–93.5)	88/112; 78.6 (69.8–85.8)	72.1	0.88	52.9 (22.2–125.7)
HD-WLE + Iscan2 + Iscan3 + NBI	46/56; 82.1 (69.6–91.1)	170/185; 91.9 (87.0–95.4)	46/61; 75.4 (62.7–85.5)	170/180; 94.4 (90.0–97.3)	216/241; 89.6 (85.1–93.2)	92/117; 78.6 (70.1–85.7)	71.9	0.89	52.1 (22.0–123.7)

Abbreviations: AUROC, area under the receiver operating characteristic; DOR, diagnostic odds ratio; HD-WLE, high-definition white-light endoscopy; MCC, Matthews correlation coefficient; NBI, narrow banding imaging; NPV, negative predictive value; PICaSSO, Paddington International Virtual Chromoendoscopy Score; PPV, positive predictive value; UCEIS, ulcerative colitis endoscopic index of severity.

TABLE 3 | Diagnostic performance of the AI switching model for detection of histological activity/remission—Iscan cohort ($n = 72$).

	Sensitivity	Specificity	PPV	NPV	Observed agreement	F1 score	MCC	AUROC	DOR
RHI									
HD-WLE	37/48; 77.1 (62.7–88.0)	171/193; 88.6 (83.3–92.7)	37/59; 62.7 (49.2–75)	171/182; 94.0 (89.4–96.9)	208/241; 86.3 (81.3–90.4)	74/107; 69.2 (59.5–77.7)	61.0	0.85	26.15 (11.7–58.6)
Iscan2	33/48; 68.8 (53.8–81.3)	177/193; 91.7 (86.9–95.2)	33/49; 67.4 (52.5–80.1)	177/192; 92.2 (87.4–95.6)	210/241; 87.1 (82.2–91.1)	66/97; 68.0 (57.8–77.2)	60	0.75	24.337 (11.0–54.0)
Iscan3	26/48; 54.2 (39.2–68.6)	172/193; 89.1 (83.9–93.1)	26/47; 55.3 (40.1–69.8)	172/194; 88.7 (83.3–92.8)	198/241; 82.2 (76.7–86.8)	52/95; 54.7 (44.2–65.0)	43.6	0.80	9.7 (4.7–20.0)
NBI	32/48; 66.7 (51.6–79.6)	173/193; 89.6 (84.5–93.6)	32/52; 61.5 (47.0–74.7)	173/189; 91.5 (86.6–95.1)	205/241; 85.1 (79.92–89.3)	64/100; 64 (53.8–73.4)	54.7	0.83	17.3 (8.1–36.9)
HD-WLE + NBI	40/48; 83.3 (69.8–92.5)	165/193; 85.5 (79.7–90.1)	40/68; 58.8 (46.2–70.6)	165/173; 95.4 (91.1–98.0)	205/241; 85.1 (79.9–89.3)	80/116; 69.0 (59.7–77.2)	61.1	0.88	29.5 (12.5–69.5)
HD-WLE + Iscan2 + Iscan3	37/48; 77.1 (62.7–88.0)	181/193; 93.8 (89.4–96.8)	37/49; 75.5 (61.1–86.7)	181/192; 94.27 (90.0–97.1)	218/241; 90.5 (86.0–93.9)	74/97; 76.3 (66.6–84.3)	70.3	0.88	50.7 (20.8–123.7)
HD-WLE + Iscan2 + Iscan3 + NBI	41/48; 85.4 (72.2–93.9)	177/193; 91.7 (86.9–95.2)	41/57; 71.9 (58.5–83.0)	177/184; 96.2 (92.3–98.5)	218/241; 90.5 (86.0–93.9)	82/105; 78.1 (69.0–85.6)	72.5	0.89	64.8 (25.0–167.7)
NHI									
HD-WLE	41/58; 70.7 (57.3–81.9)	157/183; 85.8 (79.9–90.5)	41/67; 61.2 (48. -72.9)	157/174; 90.2 (84.8–94.2)	198/241; 82.2 (76.7–86.8)	82/125; 65.6 (56.6–73.9)	53.9	0.90	14.6 (7.2–29.4)
Iscan2	37/58; 63.8 (50.1–76.0)	160/183; 87.4 (81.7–91.9)	37/60; 61.7 (48.2–73.9)	160/181; 88.4 (82.8–92.7)	197/241; 81.7 (76.3–86.4)	74/118; 62.7 (53.3–71.4)	50.6	0.84	12.3 (6.1–24.5)
Iscan3	36/58; 62.1 (48.4–74.5)	154/183; 84.2 (78.0–89.1)	36/65; 55.4 (42.5–67.7)	154/176; 87.5 (81.7–92)	190/241; 78.8 (73.1–83.8)	72/123; 58.5 (49.3–67.4)	44.5	0.85	8.7 (4.5–16.9)
NBI	36/58; 62.1 (48.4–74.5)	164/183; 89.6 (84.3–93.6)	36/55; 65.5 (51.4–77.8)	164/186; 88.2 (82.6–92.4)	200/241; 83 (77.6–87.5)	72/113; 63.7 (54.1–72.6)	52.7	0.85	14.1 (6.9–28.8)
HD-WLE + NBI	47/58; 81.0 (68.6–90.1)	150/183; 82.0 (75.6–87.3)	47/80; 58.8 (47.2–69.7)	150/161; 93.2 (88.1–96.5)	197/241; 81.7 (76.2–86.4)	94/138; 68.1 (59.7–75.8)	57.2	0.90	19.4 (9.1–41.4)
HD-WLE + Iscan2 + Iscan3	43/58; 74.1 (61.0–84.7)	160/183; 87.4 (81.7–91.9)	43/66; 65.2 (52.4–76.5)	160/175; 91.4 (86.3–95.1)	203/241; 84.2 (79.0–88.6)	86/124; 69.4 (60.4–77.3)	59.0	0.91	19.9 (9.6–41.5)
HD-WLE + Iscan2 + Iscan3 + NBI	44/58; 75.9 (62.8–86.1)	163/183; 89.1 (83.6–93.2)	44/64; 68.8 (55.9–79.8)	163/177; 92.1 (87.1–95.6)	207/241; 85.9 (80.9–90.0)	88/122; 72.1 (63.3–79.9)	62.9	0.92	25.6 (12.0–54.8)
PHRI									
HD-WLE	43/57; 75.4 (62.2–85.9)	160/184; 87.0 (81.2–91.5)	43/67; 64.2 (51.5–75.5)	160/174; 92.0 (86.9–95.5)	203/241; 84.23 (79.01–88.59)	86/124; 69.35 (60.44–77.32)	59.2	0.90	20.5 (9.8–42.9)

(Continues)

TABLE 3 | (Continued)

	Sensitivity	Specificity	PPV	NPV	Observed agreement	F1 score	MCC	AUROC	DOR
Iscan2	34/57; 59.7 (45.8–72.4)	156/184; 84.8 (78.8–89.6)	34/62; 54.8 (41.7–67.5)	156/179; 87.2 (81.4–91.7)	190/241; 78.84 (73.13–83.82)	68/119; 57.14 (47.75–66.17)	43.2	0.84	8.2 (4.2–16.0)
Iscan3	35/57; 61.4 (47.6–74)	155/184; 84.2 (78.2–89.2)	35/64; 54.7 (41.8–67.2)	155/177; 87.6 (81.8–92.0)	190/241; 78.84 (73.13–83.82)	70/121; 57.85 (48.54–66.77)	44	0.85	8.5 (4.4–16.5)
NBI	38/57; 66.7 (52.9–78.6)	158/184; 85.9 (80.0–90.6)	38/64; 59.4 (46.4–71.5)	158/177; 89.3 (83.8–93.4)	196/241; 81.33 (75.82–86.04)	76/121; 62.81 (53.56–71.42)	50.6	0.85	12.2 (6.1–24.2)
HD-WLE + NBI	50/57; 87.7 (76.3–94.9)	152/184; 82.6 (76.3–87.8)	50/82; 61.0 (49.6–71.6)	152/159; 95.6 (91.1–98.2)	202/241; 83.82 (78.55–88.23)	100/139; 71.94 (63.7–79.23)	63.1	0.90	33.9 (14.1–81.6)
HD-WLE + Iscan2 + Iscan3	44/57; 77.2 (64.2–87.3)	161/184; 87.5 (81.8–91.9)	44/67; 65.7 (53.1–76.9)	161/174; 92.5 (87.6–96)	205/241; 85.06 (79.92–89.31)	88/124; 70.97 (62.14–78.77)	61.4	0.91	23.7 (11.1–50.5)
HD-WLE + Iscan2 + Iscan3 + NBI	46/57; 80.7 (68.1–90.0)	170/184; 92.4 (87.6–95.8)	46/60; 76.7 (64.0–86.6)	170/181; 93.9 (89.4–96.9)	216/241; 89.63 (85.07–93.17)	92/117; 78.63 (70.09–85.67)	71.8	0.92	50.8 (21.6–119.3)

Abbreviations: AUROC, area under the receiver operating characteristic; DOR, diagnostic odds ratio; HD-WLE, high-definition white-light endoscopy; MCC, Matthews correlation coefficient; NBI, narrow banding imaging; NHI, Nancy histological index; NPV, negative predictive value; PHRI, PICaSSO histological remission index; PPV, positive predictive value; RHI, Robarts histopathology index.

on both modalities, learns to infer missing information by recognizing features correlating across modalities. The image generator effectively reconstructed digital enhancements needed for iScan and replicated the optical enhancement of NBI, suggesting that narrow-band lighting may emphasize features already visible with wide-band light.

Inflammatory bowel disease management heavily depends on endoscopic assessment [1]. Integrating VCE with WLE enhances mucosal and vascular features [10], identifying subtle inflammation and closely correlating with histology [2, 10]. Nonetheless, the lack of training makes disease scoring challenging even when VCE is available [11, 12]. Additionally, VCE is burdened by high inter- and intra-observer variability due to different endoscopic experiences, subjectivity, and complexity of visual assessment, potentially influencing disease management and patient outcomes. Our group developed the first VCE-based score for UC activity, the PICaSSO score, which correlates with histology better than common scores and successfully predicts clinical outcomes across platforms [8]. However, its complexity has limited its clinical use [21].

Artificial intelligence-enabled endoscopy has shown promise in objectively predicting endoscopic and histological activity [13]. Several AI systems have been developed to predict UC activity using WLE [22–24], and VCE-based scores [15]. Based on the PICaSSO score, we developed an AI system that accurately detects endoscopic inflammation/remission, predicting histological remission and outcomes [15]. However, it misses comprehensive insights from multiple modalities, particularly for vascular assessment.

Therefore, we developed a new AI model capable of switching images from one modality to another, allowing objective assessment using simultaneous images from different platforms. The AI switching model was trained using videos from iSCAN (HD-WLE, iScan2, and iScan3) and NBI platforms (HD-WLE and NBI), and accurately detected and converted images across all modalities.

Using the generated images, the model was first tested on the iScan cohort, showing remarkable diagnostic performance in predicting endoscopic and histological activity, especially when using images from all modalities simultaneously. Notably, multimodal assessment outperformed unimodal in predicting ER by UCEIS and PICaSSO and histological remission by RHI, NHI, and PHRI. To confirm the model's performance on different VCE modalities, we further tested it on an external cohort of NBI, providing consistent results.

Moreover, the AI showed good agreement with experts in predicting endoscopic and histological activity/remission, confirming its importance for disease assessment, especially in a non-expert context.

Finally, the model outperformed human assessment in predicting 12-month clinical outcomes in the iScan cohort, while showing comparable performance in the NBI cohort. This confirms the role of VCE in highlighting subtle patchy activity and vascular changes, allowing more precise disease assessment and outcome prediction to personalize UC management.

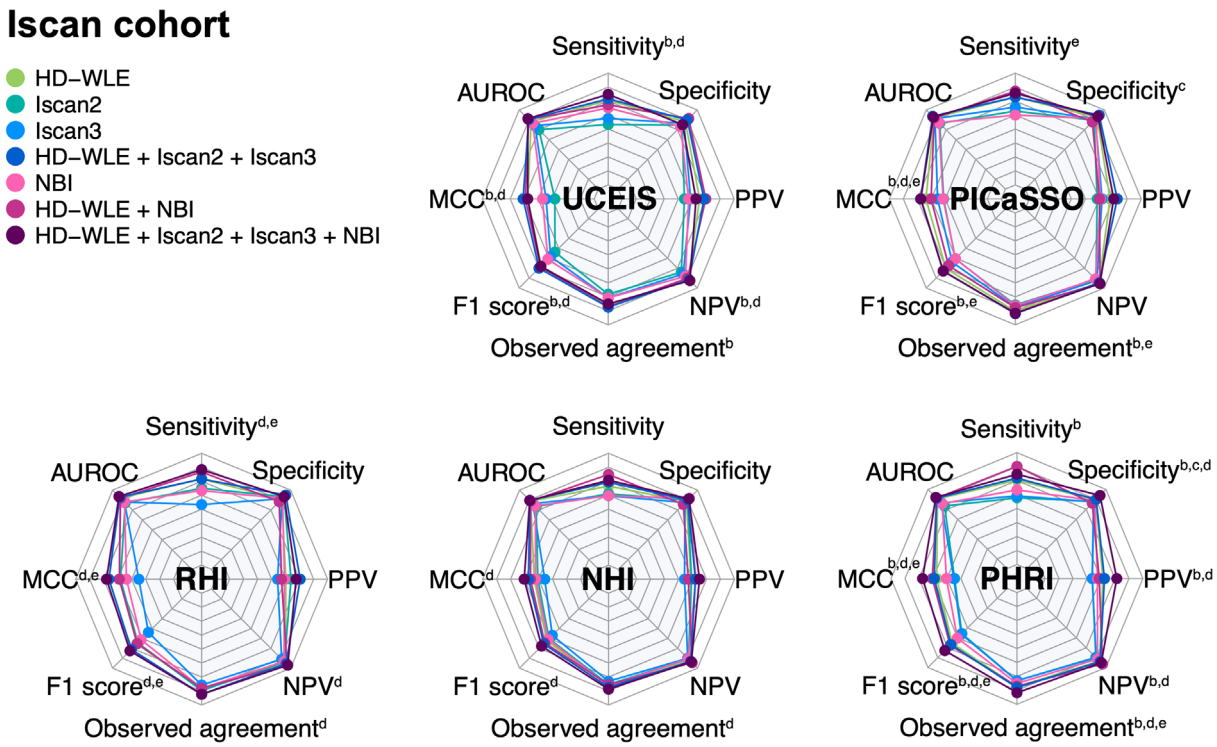

FIGURE 1 | AI switching model detection of endoscopic and histological activity in the iScan testing cohorts. The two-sample test for equality of proportions with continuity correction was statistically significant ($p < 0.05$) when comparing HD-WLE + Iscan2 + Iscan3 + NBI versus Iscan2, (b) Iscan3, (c) HD-WLE + Iscan2 + Iscan3, (d) NBI, (e) HD-WLE + NBI. Superscripts on each performance measure indicate the modalities in which a significantly difference is reached compared to the multimodal assessment. AUROC, area under the receiver operating characteristic; HD-WLE, high-definition white light endoscopy; MCC, Matthew's correlation coefficient; NBI, narrow band imaging; NHI, Nancy histological index; NPV, negative predictive value; PHRI, PICaSSO Histological Remission Index; PICaSSO, Paddington International virtual ChromoendoScopy Score; PPV, positive predictive value; RHI, Robarts histopathology index; UCEIS, ulcerative colitis endoscopic index score.

The AI switching model can simultaneously generate images starting from single modalities, achieving accurate and automated assessment of disease activity even in non-expert hands. Therefore, it can address inter-observer variability and the need for expertise, bringing endoscopy closer to histology. This is increasingly crucial as histologic remission better predicts clinical outcomes, guiding IBD management [25]. Indeed, our model was able to predict histologic activity, paving the way for targeted biopsy sampling. Moreover, studies demonstrated that AI-assisted colonoscopies utilizing VCE can objectively assess deeper healing in IBD and predict clinical outcomes [15].

Our work has several strengths. It is the first AI switching model built using the prospective PICaSSO cohort, ensuring high-quality videos, different modalities, reliable scoring, and precise outcome assessment. The novelty is the simultaneous inflammation assessment using multiple modalities and different endoscopic platforms, resulting in improved AI ability to predict activity and outcomes. The model was successfully developed and tested across iScan and NBI cohorts, offering a standardized assessment and enabling comprehensive evaluation of modality-specific features. The ability to accurately predict histological activity, especially alongside ER, can ultimately help in guiding disease management. Unlike other AI models that predict activity/remission through validated scores using single WLE images [24], our model dynamically switches and integrates different VCE images, potentially improving diagnostic accuracy. WLE focuses on gross mucosal changes in severe disease, whereas VCE enhances patchy and early inflammation. Therefore, our model can provide a more nuanced disease assessment. Moreover, previous models provided frame-level classification on endoscopic subscores, while our approach ensures video-level classification of overall scores, representing a step forward in AI implementation in daily practice.

The study also has limitations. Training GANs architecture is difficult. However, CycleGAN uses consistency and identity loss to minimize artifacts during the style transfer process. To ensure accuracy, a sample of generated images was visually inspected to contain similar details as the original without artifacts. Still, extensive validation by endoscopists is needed to confirm reliability. Second, the PICaSSO score is precise in capturing patchy inflammation, but it is difficult to implement in clinical practice due to its complexity. Therefore, the AI switching might facilitate the widespread VCE adoption.

Moreover, patients in the iScan and NBI cohorts are not equally distributed. Nonetheless, we implemented class-weighting techniques and data augmentation strategies to mitigate potential bias during training and reproducibility. The same number of frames per modality was used to train the image generator. Moreover, the only training set used for classification is the training subset of the iScan cohort, while the validation set of the iScan and NBI cohorts was used as the test set. Wider external

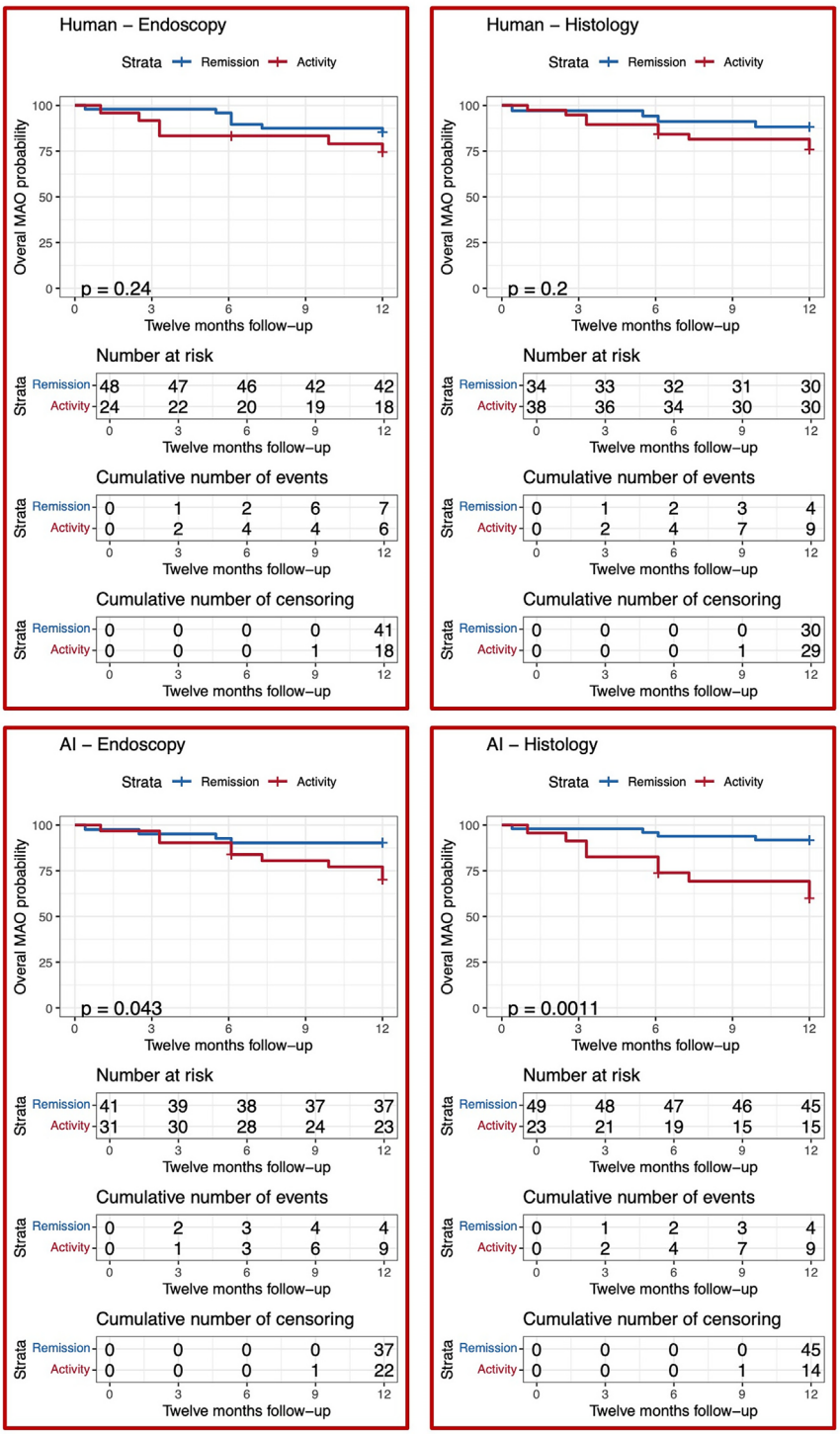

FIGURE 2 | Kaplan Meier curve describing the pooled time-to-event analysis of artificial intelligence (AI) vs. human (endoscopists or pathologists) in predicting a major adverse outcome (MAO) over a 12-month follow-up in the iScan testing cohort using endoscopic and histological scores (blue and red lines represent disease remission and activity, respectively). The pooled HR was defined in terms of disease activity at least in one endoscopic/histological score.

validation is essential, and indeed, a multicentre prospective study is ongoing to assess the AI model's robustness and generalizability beyond its original training set.

Finally, the system relies on images obtained through the iScan and the NBI mode. However, other advanced VCE platforms, including Blue Light Imaging/Linked Color Imaging (Fujifilm), Optical Enhancement-1 (Pentax), Red Dichromatic Imaging, and Texture and Color Enhancement Imaging (Olympus), are currently available. Further ongoing studies aim to demonstrate the efficacy of the algorithm on these modalities. Our ongoing international multicenter MONET Study will expand and validate the model by training and converting images across new VCE modalities.

5 | Conclusion

Our AI-switching model identifies different endoscopic images and seamlessly converts them simultaneously from one to another across multiple platforms. This algorithm can help provide different disease features and allow accurate AI-enabled disease assessment and outcomes prediction. This model represents a significant step toward making endoscopy closer to histology and can serve as a valuable tool to guide IBD management.

Author Contributions

M.I. and E.G. conceived, designed, and supervised the study, collected and analyzed the data, drafted and critically reviewed the manuscript for important intellectual content. I.Z., G.S., B.B.K., U.C. and M.P.-T. collected and analyzed the data, drafted and critically reviewed the manuscript for important intellectual content. R.A. and P.M. analyzed the data and critically reviewed the manuscript for important intellectual content. V.N. and S.G. supervised the study, analyzed the data, and critically reviewed the manuscript for important intellectual content. I.C., I.D., A.B., B.H., R.C. and R.B. collected data and critically reviewed the manuscript for important intellectual content.

Acknowledgments

The PICaSSO group includes Pradeep Bhandari, Gert de Hertogh, Jose G. Ferraz, Martin Goetz, Xianyong Gui, Bu' Hussian Hayee, Ralf Kiesslich, Mark Lazarev, Remo Panaccione, Adolfo Parra-Blanco, Luca Pastorelli, Timo Rath, Vincenzo Villanacci, Elin Synnøve Røyset, and Michael Vieth.

Ethics Statement

Videos were extracted from studies approved by the West Midlands Research Ethics Committee (17/WM/0223), the Research Ethics Committee Northern Ireland (17/NI/0148) for the UK, ethics approval n.2678, and CARMS n.14392 for the UK and the local competent committees for other international centres. All patients gave informed consent to participate in the study.

Conflicts of Interest

M.I. received research grants from Pentax, Eli Lilly, and personal fees from Pentax, Pfizer, Eli Lilly. Other authors have no conflicts of interest to declare related to this manuscript.

Data Availability Statement

Data will be made available from the corresponding author upon reasonable request.

References

1. D. Turner, A. Ricciuto, A. Lewis, 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* 160 (2021): 1570–1583.

2. M. Iacucci, R. Kiesslich, X. Gui, et al., "Beyond White Light: Optical Enhancement in Conjunction With Magnification Colonoscopy for the Assessment of Mucosal Healing in Ulcerative Colitis," *Endoscopy* 49 (2017): 553–559.

3. G. Santacroce, I. Zammarchi, C. K. Tan, et al., "Present and Future of Endoscopy Precision for Inflammatory Bowel Disease," *Digestive Endoscopy* 36 (2024): 292–304.

4. O. M. Nardone, Y. Snir, J. Hodson, et al., "Advanced Technology for Assessment of Endoscopic and Histological Activity in Ulcerative Colitis: A Systematic Review and Meta-Analysis," *Therapeutic Advances in Gastroenterology* 15 (2022): 175628482210925.

5. T. Kuroki, Y. Maeda, S. Kudo, 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)," *Gastrointestinal Endoscopy* 100 (2024): 97–108.

6. Pentax Medical, "A Leap Forward for In-Vivo Histology [Internet]," accessed December 12, 2024, https://www.pentaxmedical.com/en/products/video-processors/i-scan.

7. C. Bojarski, M. Waldner, T. Rath, et al., "Innovative Diagnostic Endoscopy in Inflammatory Bowel Diseases: From High-Definition to Molecular Endoscopy," *Frontiers in Medicine* 8 (2021): 655404.

8. M. Iacucci, M. Daperno, M. Lazarev, et al., "Development and Reliability of the New Endoscopic Virtual Chromoendoscopy Score: The PICaSSO (Paddington International Virtual ChromoendoScopy ScOre) in Ulcerative Colitis," *Gastrointestinal Endoscopy* 86 (2017): 1118–1127.e5.

9. R. Cannatelli, A. Bazarova, F. Furfaro, et al., "Reproducibility of the Electronic Chromoendoscopy PICaSSO Score (Paddington International Virtual ChromoendoScopy ScOre) in Ulcerative Colitis Using Multiple Endoscopic Platforms: A Prospective Multicenter International Study (With Video)," *Gastrointestinal Endoscopy* 96 (2022): 73–83.

10. M. Iacucci, S. C. L. Smith, A. Bazarova, et al., "An International Multicenter Real-Life Prospective Study of Electronic Chromoendoscopy Score PICaSSO in Ulcerative Colitis," *Gastroenterology* 160 (2021): 1558–1569.e8.

11. M. Daperno, M. Comberlato, F. Bossa, et al., "Inter-Observer Agreement in Endoscopic Scoring Systems: Preliminary Report of an Ongoing Study From the Italian Group for Inflammatory Bowel Disease (IG-IBD)," *Digestive and Liver Disease* 46 (2014): 969–973.

12. M. Daperno, M. Comberlato, F. Bossa, et al., "Training Programs on Endoscopic Scoring Systems for Inflammatory Bowel Disease Lead to a Significant Increase in Interobserver Agreement Among Community Gastroenterologists," *Journal of Crohn's & Colitis* 11 (2016): jjw181.

13. M. Iacucci, G. Santacroce, I. Zammarchi, et al., "Artificial Intelligence and Endo-Histo-Omics: New Dimensions of Precision Endoscopy and Histology in Inflammatory Bowel Disease," *Lancet Gastroenterology & Hepatology* 9 (2024): 758–772.

14. V. Jahagirdar, J. Bapaye, S. Chandan, 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," *Gastrointestinal Endoscopy* 98 (2023): 145–154.e8.

15. M. Iacucci, R. Cannatelli, T. L. Parigi, et al., "A Virtual Chromoendoscopy Artificial Intelligence System to Detect Endoscopic and Histologic Activity/Remission and Predict Clinical Outcomes in Ulcerative Colitis," *Endoscopy* 55 (2023): 332–341.

16. S. P. L. Travis, D. Schnell, P. Krzeski, et al., "Developing an Instrument to Assess the Endoscopic Severity of Ulcerative Colitis: The Ulcerative Colitis Endoscopic Index of Severity (UCEIS)," *Gut* 61 (2012): 535–542.

17. M. H. Mosli, B. G. Feagan, G. Zou, et al., "Development and Validation of a Histological Index for UC," *Gut* 66 (2017): 50–58.

18. A. Marchal-Bressenot, J. Salleron, C. Boulagnon-Rombi, et al., "Development and Validation of the Nancy Histological Index for UC," *Gut* 66 (2017): 43–49.

19. X. Gui, A. Bazarova, R. del Amor, et al., "PICaSSO Histologic Remission Index (PHRI) in Ulcerative Colitis: Development of a Novel Simplified Histological Score for Monitoring Mucosal Healing and Predicting Clinical Outcomes and Its Applicability in an Artificial Intelligence System," *Gut* 71 (2022): 889–898.

20. I. Zammarchi, G. Santacroce, and M. Iacucci, "Next-Generation Endoscopy in Inflammatory Bowel Disease," *Diagnostics* 13 (2023): 2547.

21. O. M. Nardone, M. Iacucci, V. Villanacci, et al., "Real-World Use of Endoscopic and Histological Indices in Ulcerative Colitis: Results of a Global Survey," *United European Gastroenterology Journal* 11 (2023): 514–519.

22. H. P. Bhambhvani and A. Zamora, "Deep Learning Enabled Classification of Mayo Endoscopic Subscore in Patients With Ulcerative Colitis," *European Journal of Gastroenterology & Hepatology* 33 (2021): 645–649.

23. M. F. Byrne, R. Panaccione, J. E. East, et al., "Application of Deep Learning Models to Improve Ulcerative Colitis Endoscopic Disease Activity Scoring Under Multiple Scoring Systems," *Journal of Crohn's & Colitis* 17 (2023): 463–471.

24. K. Takenaka, K. Ohtsuka, T. Fujii, S. Oshima, R. Okamoto, and M. Watanabe, "Deep Neural Network Accurately Predicts Prognosis of Ulcerative Colitis Using Endoscopic Images," *Gastroenterology* 160 (2021): 2175–2177.e3.

25. R. V. Bryant, D. C. Burger, J. Delo, et al., "Beyond Endoscopic Mucosal Healing in UC: Histological Remission Better Predicts Corticosteroid Use and Hospitalisation Over 6 Years of Follow-Up," *Gut* 65 (2016): 408–414.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.