

Artificial intelligence in inflammatory bowel disease: innovations in diagnosis, monitoring, and personalized care

Raseen Tariq  and Anita Afzali

Ther Adv Gastroenterol

2025, Vol. 18: 1–16

DOI: 10.1177/
17562848251357407

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

Abstract: Artificial intelligence (AI) is redefining the management of inflammatory bowel diseases (IBD) by enhancing diagnostic accuracy, refining disease classification, and optimizing disease monitoring. This review highlights AI's potential to transform IBD management by streamlining clinical workflows, improving diagnostic precision, and supporting personalized treatment strategies. By addressing the limitations of traditional clinical assessments including variability, subjectivity, and resource intensity, AI offers unbiased, consistent, and efficient solutions. Concluding with a forward-looking perspective, this paper emphasizes how integrating AI into clinical practice could lead to more precise, proactive, and patient-centric approaches to IBD care, ultimately enhancing clinical outcomes and quality of life for these patients.

Plain language summary

Role of artificial intelligence in diagnosis and management of inflammatory bowel diseases

Inflammatory Bowel Diseases (IBD), like Crohn's disease and ulcerative colitis, are chronic conditions that affect the digestive system. Diagnosing and managing IBD can be complex, often requiring multiple tests and expert interpretation. Artificial Intelligence (AI) offers new tools to improve IBD care. AI systems can quickly and accurately analyze medical images, predict treatment responses, and support doctors in making better decisions. This review explains how AI can make diagnosing IBD more accurate, treatment plans more personalized, and overall care more efficient. By using AI, healthcare providers could help patients with IBD get faster, more reliable, and more tailored treatments, improving their quality of life.

Keywords: artificial intelligence, computer-aided diagnosis, Crohn's disease, deep learning, disease monitoring, dysplasia detection, inflammatory bowel disease, machine learning, natural language processing, prognosis prediction, radiomics, ulcerative colitis

Received: 20 February 2025; revised manuscript accepted: 24 June 2025.

Introduction

The management of inflammatory bowel diseases (IBD) requires specialized expertise for making accurate diagnosis and nuanced decision-making to optimize patient outcomes. Accurate diagnosis, disease classification, and proactive disease monitoring are fundamental to effective care,

relying on endoscopy, imaging, and histopathology tools to evaluate treatment targets and management of disease.¹ However, interpreting these tools is often subjective, prone to bias, costly, and time-intensive. While standardized clinical and endoscopic scoring systems exist to bring consistency to these assessments, their adoption in

Correspondence to:
Raseen Tariq
Division of
Gastroenterology and
Hepatology, Virginia
Commonwealth University,
1200 E Broad Street, P.O.
Box 980341, Richmond, VA
23298, USA
raseen.tariq@vcuhealth.org
Anita Afzali
Division of
Gastroenterology and
Hepatology, University
of Cincinnati, Cincinnati,
OH, USA

routine practice has been limited due to their complexity and the fact that they are often cumbersome and difficult to integrate efficiently into clinical workflows. Consequently, disease activity, phenotype, and therapeutic response evaluations can differ, introducing variability in care.

Furthermore, crafting effective treatment strategies for IBD goes beyond understanding disease activity alone. A patient's coexisting conditions, prior treatments, extraintestinal manifestations, and personal preferences often shape management decisions, resulting in highly individualized care plans. While clinical guidelines provide a general roadmap, the intricate and dynamic nature of IBD care demands advanced expertise and experience to navigate these complexities.

Artificial intelligence (AI) refers to the simulation of human intelligence in machines programmed to think, learn, and solve problems. In healthcare, AI systems are primarily powered by machine learning (ML), which enables computers to identify patterns and make predictions based on data. One common approach is supervised learning, where models are trained on labeled datasets to forecast outcomes in new cases.²

Deep learning is a subfield of ML that uses neural networks with multiple layers that learn complex relationships directly from raw inputs. Convolutional neural networks (CNNs), in particular, are highly effective at analyzing visual data, making them well-suited for tasks like endoscopic image interpretation, radiologic pattern detection, and histological feature recognition.^{3,4} Natural language processing (NLP), another key technique, allows AI systems to extract meaningful information from unstructured clinical text, such as physician notes or patient reports.⁵

AI is poised to transform how IBD is diagnosed, classified, and monitored, addressing many challenges faced in current clinical workflows.⁶ By leveraging digitized medical data such as imaging, pathology slides, clinical notes, and patient-reported outcomes alongside advanced analytical methods including ML and neural networks, AI can identify patterns that may be difficult for clinicians to detect. For instance, AI-powered tools can provide rapid, unbiased, and consistent interpretation of endoscopic and radiological images, minimizing variability and ensuring uniformity across providers and settings.^{7,8}

In addition, AI applications are advancing beyond imaging to include NLP capabilities that can analyze clinical documentation and patient narratives, extracting actionable insights to support diagnostic and prognostic decisions. ML models have shown promise in categorizing patients into distinct disease phenotypes, predicting complications, and tracking therapeutic responses over time, offering a personalized and proactive approach to care.⁹

This review explores the expanding role of AI in IBD care, focusing on its applications in disease diagnosis, classification, and monitoring (Figure 1). The integration of AI into clinical practice has the potential to streamline workflows, improve diagnostic precision, reduce variability, and enhance patient outcomes, marking a transformative step forward in the management of IBD.

Search strategy and review methodology

This article is a narrative review designed to synthesize key advances in AI applications in the diagnosis, classification, and monitoring of IBD. To identify relevant studies, we searched PubMed, Scopus, and Web of Science databases for English-language articles published between January 2012 to December 2024 using combinations of the following terms: “inflammatory bowel disease,” “ulcerative colitis,” “Crohn’s disease,” “artificial intelligence,” “machine learning,” “deep learning,” “natural language processing,” “radiomics,” “endoscopy,” “histology,” and “diagnostic imaging.”

We included peer-reviewed original studies, systematic reviews, and meta-analyses that applied AI to clinical or imaging data in the context of IBD. Exclusion criteria were non-English publications, conference abstracts, non-human studies, and papers focused purely on algorithm development without clinical relevance. As this was a narrative review rather than a systematic one, study selection was guided by relevance to clinical practice and recent impact in the field.

AI in disease diagnosis and classification

Differentiating UC and CD

Distinguishing ulcerative colitis (UC) from Crohn’s disease (CD) presents several challenges due to overlapping clinical, endoscopic, and

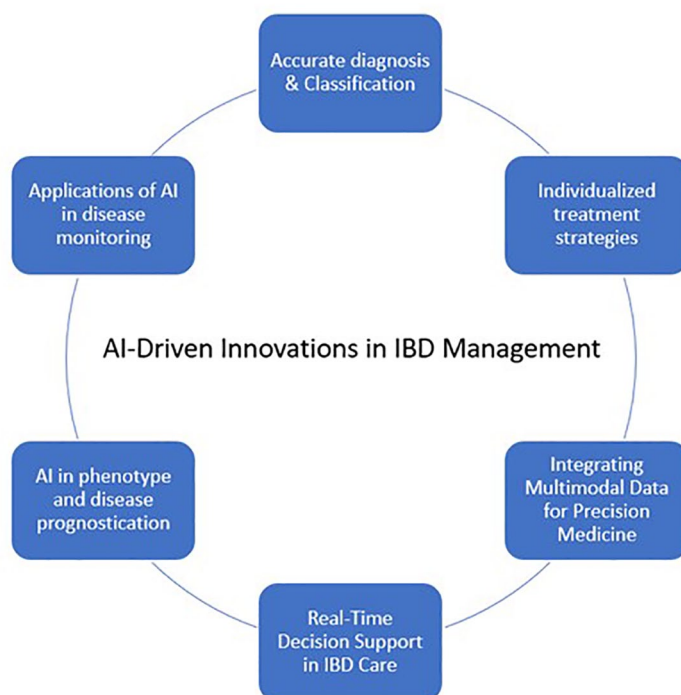


Figure 1. AI applications in IBD management, showcasing innovations in diagnosis, monitoring, and personalized care.

AI, artificial intelligence; IBD, inflammatory bowel diseases.

histopathological characteristics. Endoscopically, UC is typically characterized by continuous colonic involvement, loss of the normal vascular pattern, and superficial ulcerations. While CD often features cobble stoning, deep ulcers, and strictures,¹⁰ however, these findings are not pathognomonic, and overlap can occur. Atypical presentations, such as rectal sparing and patchy inflammation in UC, may resemble CD, further complicating the diagnostic process.¹¹ The use of immunohistochemical markers, (e.g., Das-1 and CG-3) and quantification of CD30+ lymphocytes and eosinophils in biopsy samples, has been proposed to enhance the differentiation between UC and CD. However, the use of immunohistochemical marker is limited by variability in marker expression, lack of absolute specificity, and the requirement for specialized expertise and standardized protocol.^{12,13} Several studies have highlighted the potential of AI in overcoming these challenges in distinguishing UC from CD through various advanced methodologies, thereby enhancing diagnostic accuracy.

Chierici et al.¹⁴ developed a deep learning framework using endoscopic images to classify IBD subtypes and distinguish healthy controls. The model exhibited moderate performance in differentiating UC from CD (Matthews Correlation Coefficient (MCC)=0.688). The model's lower performance may reflect intrinsic difficulty in distinguishing IBD subtypes. Conversely, the model demonstrated excellent performance in differentiating UC from healthy controls with an MCC of 0.931.¹⁴

Another study employed whole transcriptome analysis on endoscopic samples to identify differentially expressed genes, including *PI3*, *ANXA1*, and *VDR*, which exhibited significant performance in discriminating CD from UC with an area under the curve (AUC) of 0.84. This system leverages the differential expression of specific genes to accurately classify the two conditions.¹⁵

Furthermore, Manandhar et al.¹⁶ utilized supervised ML on gut microbiome data to identify differential bacterial taxa and operational taxonomic

units that distinguish IBD patients from healthy controls and CD from UC. The model achieved an AUC of >0.90 for differentiating CD from UC.¹⁶

Another AI system based on Raman spectroscopy, in conjunction with support vector machines (SVMs), demonstrated the ability to differentiate between CD and UC with 98.9% accuracy. This method utilizes the unique molecular signatures captured by Raman spectroscopy to distinguish between the two types of IBD, providing a highly accurate diagnostic tool.¹⁷

These studies suggest that AI-based approaches show great promise in accurately distinguishing UC from CD, addressing critical gaps in clinical practice where misdiagnosis leads to suboptimal outcomes. Additionally, by integrating microbiome data and unique molecular signatures, this represents a significant leap in diagnostic precision. However, their clinical application requires further refinement, validation, and standardized protocols.

Endoscopy. AI technologies, particularly deep learning algorithms, have demonstrated remarkable accuracy in detecting and grading mucosal inflammation and lesions. Rimondi et al.¹⁸ conducted a comprehensive systematic review and meta-analysis of studies evaluating the diagnostic accuracy of AI systems assessing mucosal healing in patients with UC using endoscopic images and videos. The AI systems exhibited sensitivity and specificity of 91% and 89%, respectively, and a diagnostic odds ratio (OR) of 92.42 for fixed images. For videos, sensitivity was 86%, specificity was 91%, and OR of 70.86. The AUC was 0.957 for fixed images and 0.941 for videos, indicating exceptional diagnostic performance. Despite the impressive diagnostic performance, the findings were subject to moderate to high heterogeneity due to variations in training algorithms, datasets, and mucosal healing definitions.¹⁸

Xie et al.¹⁹ evaluated the performance of a deep learning model in detecting and grading small-bowel CD ulcers from double balloon images. The model demonstrated high accuracies for detecting ulcers (96.3%), non-inflammatory stenosis (95.7%), and inflammatory stenosis (96.7%), with AUC values exceeding 0.98 for all categories. In ulcer grading, it exhibited

accuracies of 87.3% (95% CI, 84.6%–89.6%) for ulcerated surface, 87.8% (95% CI, 85.0%–90.2%) for ulcer size, and 85.2% (95% CI, 83.2%–87.0%) for ulcer depth. Compared to endoscopists, the AI model outperformed junior and intermediate levels and performed similarly to senior endoscopists, demonstrating expert-level accuracy in lesion detection and objective severity grading. These findings highlight the potential of AI to enhance the accuracy, efficiency, and objectivity of small-bowel CD evaluations.¹⁹

Capsule endoscopy (CE) is a crucial modality in diagnostic evaluation and assessing the extent of disease in patients with CD. Ferriera et al. developed and validated a CNN-based model for automatically detecting ulcers and erosions in the gastrointestinal tract using images from the PillCam™ Crohn's Capsule (PCC). Images from 59 PCC examinations conducted at two centers were divided into a training dataset (80%) and a validation dataset (20%). The model achieved an overall accuracy of 98.8%, sensitivity of 98%, specificity of 99%, and an AUC of 1 for detecting ulcers and erosions in the PCC images. Notably, the CNN demonstrated comparable diagnostic accuracy to expert gastroenterologists.²⁰

Similarly, Klang et al.²¹ developed a deep learning algorithm for the automated detection of small-bowel ulcers in CD CE images. The study included 17,640 images from 49 patients, AUC values of 0.94–0.99, and accuracy ranging from 95.4% to 96.7%, demonstrated high diagnostic accuracy.²¹ In 2019, Aoki et al.²² developed a CNN-based model for automatically detecting ulcers and erosions in computed CE images. The model exhibited overall performance with an AUC of 0.95, sensitivity of 88%, and specificity of 91%.²²

The transformative role of AI in improving diagnostic workflows, particularly in areas requiring precise lesion detection and grading is quite evident from these studies (Table 1). By automating image analysis with high accuracy and efficiency, AI-based models hold promise for enhancing the diagnostic process, standardizing assessments, and alleviating the clinical workload. Continued validation in diverse clinical settings and integration into practice are essential steps to realize the full potential of AI in IBD diagnostics.

Table 1. Summary of AI applications in endoscopy, histology, and imaging for inflammatory bowel disease.

Study	AI methodology	Dataset	Key findings	Clinical impact	Limitations	Clinical readiness
Endoscopy						
Rimondi et al. ¹⁸	AI for mucosal healing assessment	Endoscopic images/videos	AUC=0.95 (fixed), 0.94 (videos)	High accuracy for mucosal healing scoring	Heterogeneous definitions, dataset variability	Research-only
Xie et al. ¹⁹	Deep learning for ulcer detection (CD)	Double-balloon endoscopy	AUC >0.98; Accuracy ~96%	Outperformed junior/intermediate endoscopists	Single center; lacks external validation	Pilot phase
Klang et al. ²¹	DL for CE lesion detection	CE	AUC 0.94–0.99 for ulcer detection	Improves workflow and automation	Retrospective design; narrow scope	Research-only
Aoki et al. ²²	CNN for ulcer/erosion detection in CE	Computed CE images	AUC=0.95; sensitivity 88%, specificity 91%	Supports ulcer detection in CE evaluations	Limited generalizability; single center	Pilot phase
Histology						
Rymarczyk et al. ²⁴	AI for histology severity classification	6431 trial biopsies	Accuracy 76%–94%	Standardizes histology grading	Class imbalance; moderate accuracy in CD ileum	Pilot phase
Furlanello et al. ²³	AI for plasma cell quantification	4981 histology images	OR=4.97 for IBD vs normal	Detects basal plasmacytosis accurately	Lacks external validation	Research-only
Noguchi et al. ²⁹	CNN for p53 mutation prediction (dysplasia)	IHC-stained slides	Accuracy 86%–91%	Enhances dysplasia recognition	Pathology-specific; limited validation	Pilot phase
Lopez-Serrano et al. ³⁰	CADe system (Discovery™) vs virtual chromoendoscopy	Prospective surveillance colonoscopy in UC	Comparable dysplasia detection rates	Validates real-world use of AI for dysplasia in IBD	No significant superiority; tool-dependent	Real-world
Imaging						
Stidham et al. ²⁷	ML for cumulative ileal damage scoring	8242 ileal CTE segments	AUC=0.76 for surgery prediction	Aids surgical decision-making in CD	Retrospective: clinical utility not confirmed	Pilot phase
Carter et al. ²⁵	CNN for bowel wall thickening detection	IUS images	Accuracy 90.1%	Standardizes IUS-based inflammation detection	Operator variability; no external dataset	Research-only
Naziroglu et al. ²⁶	Semi-automated bowel wall thickness (MRI)	53 patients with CD	ICC 0.88 (AI) vs 0.45 (manual)	Improves reproducibility of MRI measurements	Small sample size; limited scope	Research-only
Zeng et al. ⁴³	Radiomics for fibrosis risk (CT)	MSCT of 218 IBD patients	AUC=0.971 (train), 0.865 (test)	Predicts fibrosis and supports stratification	Retrospective; no external validation	Pilot phase

AI, artificial intelligence; AUC, area under the curve; CADe, computer-aided detection; CD, Crohn's disease; CE, capsule endoscopy; CNN, convolutional neural network; CTE, computed tomography enterography; IBD, inflammatory bowel diseases; ICC Intraclass Correlation Coefficient; IHC, Immunohistochemistry; IUS, intestinal ultrasound; ML, machine learning; MSCT, multi-slice computed tomography; OR, odds ratio; UC, ulcerative colitis.

Histology. Several studies have evaluated the potential of AI for histological diagnosis of IBD. Furlanello et al.²³ utilized 4981 annotated histological images to develop an AI system for semi-automated detection and quantification of plasma cells, specifically focusing on basal plasmacytosis, a key histological feature in IBD. The model was validated using 356 biopsies from CD, UC, and control samples. The AI system demonstrated reliable detection of plasma cells with high sensitivity, with these cells being more prevalent in colonic regions compared to the ileum, aligning well with human assessments. UC cases exhibited significantly higher plasma cell counts compared to CD cases, reflecting established histological patterns. The OR for IBD diagnosis versus normal tissues was 4.968, highlighting the AI system's accuracy.²³

The study by Rymarczyk et al.²⁴ utilized 6431 biopsies from 1189 patients enrolled in six phase II and III clinical trials for CD and UC. Biopsies were collected from specific anatomical sites, including the terminal ileum and colon for CD and the rectum and sigmoid colon for UC. The study evaluated three multi-instance learning methods and selected the model with the best overall performance for detailed analysis. Model predictions were compared against scores assigned by a central pathologist (gold standard) and an independent panel of five experienced pathologists. For colonic biopsies (CD and UC), the model achieved 87%–94% accuracy. For CD ileum biopsies, accuracy ranged from 76% to 83%. The authors acknowledged data imbalances, with fewer CD ileum biopsies and an overrepresentation of normal or mild disease severity in some subgroups, potentially limiting the model's generalizability.²⁴

These studies highlight the ability of AI to complement pathologists by streamlining and standardizing histological assessments, particularly for large datasets (Table 1). However, challenges such as data diversity, representation of disease severity, and anatomical site variability must be addressed to enhance generalizability and clinical applicability.

Cross-sectional and diagnostic imaging. AI has been utilized in cross-sectional imaging modalities such as intestinal ultrasound (IUS) to identify bowel wall thickening, a surrogate marker for bowel inflammation in CD. For instance, Carter

et al.²⁵ developed an AI-based system that achieved an overall accuracy of 90.1%, sensitivity of 86.4%, and specificity of 94% in detecting bowel wall thickening exceeding 3 mm on IUS images. This AI module facilitates the utilization of IUS by less experienced operators, potentially standardizing the interpretation of IUS imaging and enhancing diagnostic consistency.²⁵

Naziroglu et al.²⁶ employed magnetic resonance enterography (MRE) to evaluate a semiautomatic method for measuring bowel wall thickness (BWT) in patients with CD. The study analyzed the MRE dataset of 53 patients. The algorithm-generated measurements of BWT exhibited superior interobserver agreement compared to the human-assessed measurements (intraclass correlation coefficient 0.88 vs 0.45, $p=0.005$).²⁶

Computed tomography enterography (CTE) is a valuable diagnostic tool for identifying IBD. Stidham et al.²⁷ employed ML to automate the assessment of cumulative ileal injury utilizing 8242 ileal mini segments extracted from 229 CTE scans of patients diagnosed with ileal CD. The ML-predicted injury grades exhibited substantial concordance with the radiologists' assessments ($\kappa=0.80$), which is comparable to the inter-radiologist agreement ($\kappa=0.87$). Notably, the ML method demonstrated high accuracy (74.8% exact match with radiologists) and exhibited particularly strong performance in distinguishing mild-moderate from severe disease (88.6% accuracy²⁷; Table 1).

AI in IBD dysplasia

AI systems, particularly those utilizing deep learning algorithms, have accurately identified and classified neoplastic lesions in IBD patients.

Yamamoto et al.²⁸ evaluated the diagnostic accuracy of an AI system against four expert and three non-expert endoscopists using 862 non-magnified endoscopic images from 99 IBD-associated neoplasia. The AI system was designed to differentiate high-grade dysplasia and adenocarcinoma from low-grade dysplasia, sporadic adenomas, and non-neoplastic mucosa. The image-based diagnostic ability of the system yielded sensitivity, specificity, and accuracy of 64.5%, 89.5%, and 80.6%, respectively. The lesion-based diagnostic ability of the system yielded sensitivity, specificity, and accuracy of 74.4%, 85%, and

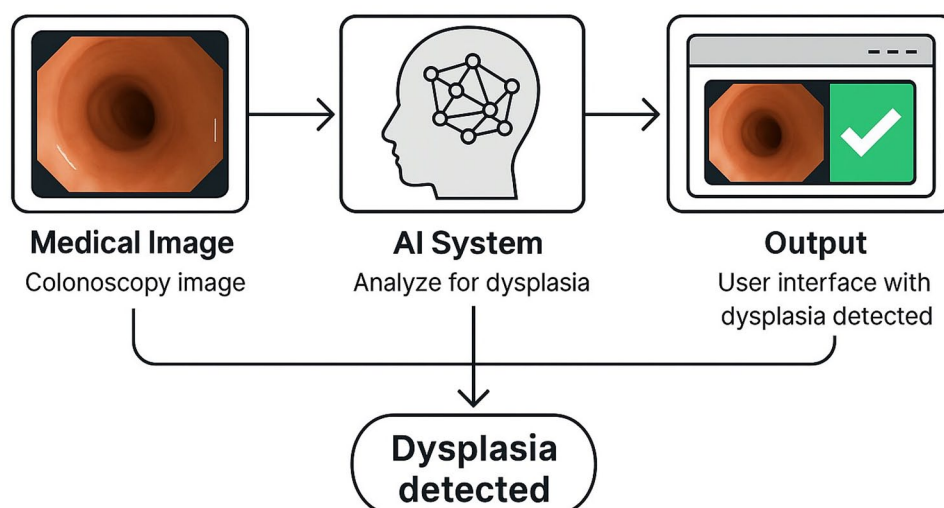


Figure 2. AI-assisted workflow for dysplasia detection in colonoscopy images: a colonoscopy image is analyzed by an AI system to detect dysplasia. The result is displayed through a user interface, indicating whether dysplasia is present to support clinical decision-making. AI, artificial intelligence.

80.8%, respectively. The AI system demonstrated a higher accuracy of 79% compared to both experts (77.8%) and non-experts (75.8%). While the AI system outperformed experts in sensitivity (72.5% vs 60.5%), it had slightly lower specificity (82.9% vs 88.0%). Additionally, the study results may not be generalizable due to the smaller sample size.²⁸

Histologic evaluation is a vital component in diagnosing UC-associated cancer or dysplasia. Immunohistochemical analysis of p53 mutation in the biopsy samples is a key study to detect dysplasia and colitis-associated carcinoma. Noguchi et al.²⁹ developed a CNN model based on p53 positivity with an average precision of 0.71–0.754. The model predicted the p53 immunohistochemistry staining with an accuracy of 86%–91%.²⁹

A recent prospective, cross-sectional, non-inferiority study by López-Serrano et al.³⁰ compared a computer-aided detection (CADE) system (Discovery™) with virtual chromoendoscopy (VCE using iSCAN) during surveillance colonoscopy in patients with UC at risk for colorectal cancer. The CADE system demonstrated comparable diagnostic performance to VCE, identifying dysplasia in a similar proportion of lesions and patients. These findings offer important real-world validation for AI-assisted endoscopy in the IBD surveillance setting, while also highlighting

practical considerations for integrating AI tools into routine clinical workflows.³⁰

AI has the potential to improve dysplasia detection in IBD through high-accuracy imaging, histology, and advanced optical tools, enhancing precision and consistency. Future work should focus on integrating AI into clinical practice for better outcomes. To further support reader understanding of AI-based imaging workflows in IBD, we include a schematic overview (Figure 2) illustrating how medical images such as those from colonoscopy or cross-sectional imaging are processed by AI systems to generate standardized outputs. These outputs can aid in disease activity scoring, dysplasia detection, or structural classification, ultimately enhancing real-time interpretation and clinical decision support.

Disease monitoring

Traditionally, scoring systems are employed to monitor disease activity in patients with IBD. Commonly utilized scoring systems for CD include the Crohn's Disease Activity Index (CDAI), the Simple Endoscopic Score for Crohn's Disease, the Crohn's Disease Endoscopic Index of Severity (CDEIS),³¹ and for UC the Mayo Endoscopic score and the Ulcerative Colitis Endoscopic Index of Severity (UCEIS).³² Several imaging-based scoring systems have

been developed to assess disease activity for CD, such as the Magnetic Resonance Index of Activity (MaRIA), the Clermont and London score.³³ However, these scoring systems encounter significant limitations, including their time-consuming nature, variable sensitivity and specificity, and interobserver variability. These limitations restrict their clinical applicability and present an opportunity for AI to enhance patient care. Studies have demonstrated significant potential in addressing these limitations³⁴ (Table 2).

The study conducted by Fan et al.³⁵ evaluated the efficacy of a deep learning-based system in assessing inflammatory activity in UC using 5875 endoscopic images and 20 full-length videos from 332 UC patients. The system demonstrated an accuracy of 86.54% for Mayo Score classification and accuracies of 90.7%, 84.6%, and 77.7% for UCEIS sub-scores for vascular pattern, erosions/ulcers, and bleeding, respectively, with kappa coefficients exceeding 0.7. These findings exhibited a high level of agreement with the expert endoscopist. Additionally, the system generated two-dimensional images that provided visual representations of inflammation severity and distribution, facilitating the identification of disease activity. The study also reported the system's ability to track changes in inflammation before and after treatment, which correlated with clinical outcomes. However, it was noted that the system exhibited lower performance for certain disease categories, particularly in cases of mild inflammation (Mayo 1) and UCEIS bleeding scores.³⁵ While the model demonstrated robust performance overall and strong agreement with expert endoscopists, its reduced accuracy in detecting mild inflammation and bleeding suggests the need for further refinement before routine clinical adoption.

Cai et al.³⁶ evaluated ML models for predicting disease activity based on non-invasive, routinely collected clinical and laboratory data from 876 individuals with IBD. The study included 601 patients with CD and 275 patients with UC. Disease activity was assessed using the CDAI for CD and the Mayo score for UC. Out of the seven algorithms tested, the SVM exhibited superior performance for predicting disease activity in CD patients, while the Adaptive Boosting (AdaBoost) algorithm demonstrated the best performance for UC patients. For disease activity prediction, the ML models achieved an accuracy of 93%,

sensitivity of 94.7%, specificity of 92%, and an AUC of 0.975 for SVM. Similarly, AdaBoost achieved an accuracy of 85.5%, sensitivity of 84.4%, specificity of 87.5%, and AUC of 0.911 for UC patients.³⁶ These findings demonstrate ML models' high accuracy and reliability in predicting disease activity, suggesting potential applications in clinical decision-making. However, the study's retrospective design, potential for selection bias, and lack of external validation limit its generalizability and emphasize the need for prospective, multicenter validation before clinical integration.

Puylaert et al.³⁷ developed a semiautomatic scoring system to evaluate the disease activity for CD using the MRI. Virtual gastrointestinal tract (VIGOR) score was derived from semiautomatic measurements of BWT, excess volume, and dynamic contrast enhancement, combined with radiologist-assessed features like mural T2 signal. Scores were compared to established MRI activity indices (MaRIA, London score, and CD MRI index) and the CDEIS as the reference standard. The VIGOR score exhibited moderate correlation with the CDEIS ($r=0.58$ for observer 1 and $r=0.59$ for observer 2), comparable to other MRI activity scores. Notably, the VIGOR score demonstrated superior interobserver agreement (ICC=0.81 vs 0.44–0.59). Furthermore, the VIGOR score achieved a diagnostic accuracy of 80%–81% for detecting active disease, comparable to other scores.³⁷

The study by Rymarczyk et al.²⁴ developed AI-based severity scoring system both for CD and UC utilizing histological dataset. The model's performance was evaluated by comparing its predicted disease severity with the assessments of central readers. The global histology activity score for CD demonstrated accuracies ranging from 65% to 89% with kappa values of 0.46–0.67. Similarly, the simplified Geboes score for UC achieved accuracies ranging from 65% to 85% with kappa values of 0.44–0.68. The model exhibited optimal performance at the extremes of severity (grades 0 and 3) but exhibited a decline in accuracy for intermediate grades. The model's performance was comparable to that of independent pathologists for most features and subgrades, with minor discrepancies in accuracy for specific categories (detecting neutrophils in CD ileum biopsies). Furthermore, the model's predictions of histological improvement were significantly

Table 2. Summary of AI applications in disease monitoring and prognosis for inflammatory bowel disease.

Study	AI methodology	Dataset	Key findings	Clinical impact	Limitations	Clinical readiness
Disease monitoring						
Fan et al. ³⁵	AI-based endoscopic scoring (Mayo/UCEIS)	5875 images + 20 videos	Accuracy 86.5%; subscore AUCs 0.77–0.91	Tracks endoscopic disease activity	Lower accuracy for mild inflammation; no external validation	Pilot phase
Cai et al. ³⁶	ML for CDAI/Mayo activity prediction	876 patients (clinical/lab data)	AUC 0.975 (CD), 0.911 (UC)	Predicts activity from routine data	Retrospective; no external validation	Research-only
Puylaert et al. ³⁷	MRI-based VIGOR activity scoring	MRI dataset	$r=0.58$ – 0.59 vs CDEIS; ICC=0.81	Improves reproducibility of MRI activity scoring	Internal validation only; semiautomated method	Pilot phase
Rymarczyk et al. ²⁴	Histology-based disease severity scoring	Biopsies from 6 trials	Accuracy 65%–89%; kappa 0.44–0.68	Enables histologic disease tracking	Class imbalance; less accurate for intermediate severity	Pilot phase
Prognosis						
Iacucci et al. ³⁹	PHRI-based flare prediction (UC)	UC histologic biopsies	HR=4.64 for relapse within 1 year	Predicts clinical outcomes from histology	Not integrated into clinical workflow	Pilot phase
Ohara et al. ⁴⁰	GCR relapse prediction	114 UC patients in remission	Relapse 45% vs 6.5% if GCR $\leq 12\%$	Identifies at-risk patients despite endoscopic remission	Small sample; moderate interobserver variation	Research-only
Klein et al. ⁴¹	Histology-based fibrosis and penetrating CD risk	CD histologic slides	AUC 0.74–0.78 for fibro stenosis and penetration	Long-term risk prediction from baseline histology	Early study; not validated in real-time settings	Research-only
Zhu et al. ⁴⁴	Radiomics for fibrosis and biologic response (UC)	CT from 119 UC patients	AUC 0.86 (fibrosis), 0.80 (response prediction)	Supports treatment selection in chronic UC	Retrospective; small cohort	Pilot phase
Stidham et al. ²⁷	ML-based CTE severity scoring for surgery prediction	8242 ileal CTE segments	AUC=0.76 for predicting surgery	Identifies CD patients needing surgery	Internal-only score; not prospectively validated	Pilot phase
AI, artificial intelligence; AUC, area under the curve; CD, Crohn's disease; CDAI, Crohn's Disease Activity Index; CDEIS, Crohn's Disease Endoscopic Index of Severity; CTE, computed tomography enterography; GCR, Goblet Cell Ratio; HR, hazard ratio; ML, machine learning; PHRI, PICaSSO Histologic Remission Index; UC, ulcerative colitis; UCEIS, Ulcerative Colitis Endoscopic Index of Severity; VIGOR, Virtual gastrointestinal tract.						

correlated with clinical remission and endoscopic improvement.²⁴

AI is transforming disease monitoring in IBD by enhancing the accuracy and efficiency of activity

assessments. It supports scoring systems, such as Mayo and UCEIS, through endoscopic image analysis, predicts disease activity using clinical data, and refines imaging evaluations with MRI and CTE-based tools.

In addition to transforming traditional scoring methods, AI aligns with the STRIDE-II framework, a cornerstone in IBD management that emphasizes a treat-to-target approach. STRIDE-II defines therapeutic goals, progressing from clinical remission and biomarker normalization to endoscopic and histological healing.³⁸ AI tools enhance this framework by enabling real-time, objective assessments of disease activity and treatment response. For example, endoscopic AI systems refine mucosal healing evaluations, while ML models predict treatment outcomes and track biomarker normalization, supporting intermediate goals. Advanced imaging-based AI tools provide detailed assessments of transmural and mucosal healing, aligning with long-term targets. By integrating AI-driven insights with STRIDE-II, clinicians can implement timely interventions and personalized care strategies, ensuring optimal outcomes for patients with IBD.

Prognosis

A substantial amount of research has been conducted to identify the prognostic factors that can reduce the risk of complications and prompt timely intervention. Nevertheless, these factors have limited predictive value due to their inability to distinguish between disease remission and activity, lack of standardization, and variable predictive accuracy. Studies have demonstrated that AI-based models generally exhibit superior prognostic accuracy compared to conventional methods in IBD (Table 2).

Iacucci et al.³⁹ evaluated an AI-based computer-aided diagnosis system for assessing histological disease activity and predicting clinical outcomes in UC. The model was developed utilizing the PICaSSO Histologic Remission Index (PHRI). The model demonstrated the ability to differentiate between histologic remission and activity with a sensitivity of 89%, specificity of 85%, and accuracy of 87%. Furthermore, the model-predicted PHRI exhibited a strong correlation with flare-ups within the subsequent year, with a hazard ratio of 4.64 (95% confidence interval (CI): 2.76–7.80), comparable to or surpassing human assessments.³⁹

Ohara et al.⁴⁰ developed a deep learning-based model to quantify goblet cell mucus (GCM) in colonic biopsies from 114 UC patients with clinical and endoscopic remission (Mayo Endoscopic

Subscore ≤ 1). The model demonstrated high accuracy in identifying GCM areas in histologic images. Patients who experienced relapse (Mayo score ≥ 3) within 12 months exhibited significantly lower Goblet Cell Ratio (GCR; defined as the ratio of GCM area to epithelial cell) in the rectum, cecum, and ascending colon compared to the relapse-free group. A GCR threshold of $\leq 12\%$ in rectal specimens was strongly associated with relapse (45% vs 6.5%, $p < 0.01$). Interobserver agreement for pathologists assessing mucin depletion was moderate (Cohen's kappa = 0.59), while the AI model exhibited excellent reproducibility.⁴⁰ Likewise, Klein et al.⁴¹ designed a system to analyze baseline histological images from individuals with CD. This system demonstrated the ability to predict the likelihood of developing fibro stenosis (AUC 0.74) and internal penetrating disease behavior (AUC 0.78) within 5 years.⁴¹

A recent systematic review by Maeda et al.⁴² of AI-assisted colonoscopy in identifying histologic remission and predicting clinical outcomes in patients with UC reported that AI systems demonstrated performance comparable to or exceeding experienced endoscopists in detecting the histologic remission. The sensitivity ranged from 65% to 98%, and the specificity ranged from 80% to 97%.⁴² Moreover, these models demonstrated the ability to identify patients at risk of relapse based on both endoscopic and histological features.

Imaging-based studies offer a non-invasive approach to monitor disease activity and predict outcomes. Several studies have demonstrated the utility of radiomics in predicting IBD outcomes. Zeng et al.⁴³ developed and validated a radiomics nomogram for IBD patients using multi-slice computed tomography (MSCT) and clinical data to stratify the risk of intestinal fibrosis. The study included data from 218 IBD patients (113 with CD and 105 with UC) who underwent MSCT imaging and endoscopic or histological evaluations. A clinical-radiomics nomogram was constructed by integrating selected radiomics features with clinical factors (e.g., lesion location, engorged vasa recta, and computed tomography (CT) value of arterial phase enhancement). In the training set, the nomogram demonstrated AUC of 0.971 and in the test set, it had an AUC of 0.865 (95% CI: 0.738–0.992) and an accuracy of 79%, reflecting excellent predictive performance. This study indicates that the integration of

radiomics and clinical data can provide superior predictive accuracy compared to models using single data sources.⁴³

The study by Zhu et al.⁴⁴ evaluated radiomics models based on CT imaging of the bowel wall and mesenteric adipose tissue to identify the severity of colonic fibrosis and predict clinical response to biologics in UC patients. Radiomics features were extracted from the CT images of 119 UC patients (patients undergoing proctocolectomy, 72 and patients starting biologics, 47). Two radiomic models were developed: bowel wall radiomic model (BW-RM) focused on bowel wall features, and the mesenteric adipose tissue radiomic model (MAT-RM) focused on mesenteric fat characteristics. Regarding predicting colonic fibrosis, BW-RM had an AUC of 0.86, and MAT-RM performed with an AUC of 0.83. Both models significantly outperformed visual assessment by radiologists (AUC ~0.60). In predicting response to biologics, MAT-RM showed superior performance compared to BW-RM (AUC; 0.71–0.80 vs 0.61–0.72).⁴⁴

Stidham et al.²⁷ studied the CTE image-based ML model to quantify cumulative ileal damage and to predict surgical outcomes. They compared ML-derived scores with traditional imaging features to predict bowel resection within 3 years. Patients who underwent surgery had significantly higher Simple Cumulative Ileal Damage Severity scores (S-CIDSS (46.6 vs 30.4, $p=0.0007$)) and mean severity grades (1.80 vs 1.42, $p<0.0001$). ML models combining S-CIDSS and mean severity grade achieved an AUC of 0.76 for predicting surgery, outperforming traditional imaging measures (AUC 0.62).²⁷ These studies support the integration of radiomics into clinical workflows for personalized management of IBD patients.

Another study based on data from the OptumLabs® Data Warehouse assessed the feasibility of ML in predicting adverse outcomes in IBD patients. The study included 72,178 patients in the training set and 69,165 in the validation set, and 108 predictive variables were incorporated in model evaluation. Random forest had the best overall performance for predicting IBD-related hospitalization (AUC 0.73), biologic initiation (AUC 0.92), steroid use (AUC 0.81), and surgery (AUC 0.71). Common predictors of adverse outcomes included prior hospitalizations,

use of steroids, antibiotics, and biologics, frequency of office visits, and IBD-related procedures.⁴⁵ The study highlights the potential of ML models to predict adverse outcomes in IBD and paves the way for implementing data-driven, preemptive care in IBD management.

AI demonstrates significant potential in improving prognostication for IBD, offering superior accuracy in predicting disease relapse, fibrosis, and adverse outcomes. By leveraging histological, imaging, and clinical data, AI-based models enable personalized, data-driven care. Future efforts should focus on integrating these models into routine practice to enhance predictive precision and patient management.

Personalized treatment strategy

Managing IBD requires personalized strategies due to its diverse phenotypes, variable disease progression, and outcomes. AI has the potential to revolutionize IBD care by processing patient data, including medical history and treatment outcomes, to generate tailored treatment recommendations that align with individual needs and preferences.⁴⁶ This not only enhances patient adherence and clinical outcomes but also optimizes direct and indirect costs including payers' financial returns. By integrating omics data, historical records, and predictive treatment responses, AI can uncover patterns and biomarkers that guide the selection of the most effective therapies for each patient.^{47,48} Additionally, as the medicine cabinet for IBD expands, offering novel therapeutic options such as Janus kinase inhibitors, S1P receptor modulators, and anti-integrin therapies. AI can integrate emerging data on these agents to refine their positioning within treatment algorithms and ensure their appropriate use in clinical practice.

It has been reported that single therapeutic agents often reach a plateau with limited remission rates⁴⁹ recognized as the “therapeutic ceiling.” This could be due to multiple pathological pathways driving the inflammatory process in the IBD. Treatment strategies involving combination of established single agents could be an alternative approach to address the complex inflammatory pathways.⁴⁹ The existing biologic therapies for IBD are supported by predictive models and clinical decision-support tools, yet there is room to improve their evaluation and application.^{47,50–52}

AI's capacity to refine these tools and assess their predictive performance can lead to more precise treatment strategies, resulting in better patient outcomes and more effective disease management.⁵³

AI and ML algorithms can forecast disease trajectories from diagnosis, enabling clinicians to determine the most suitable treatment pathways. These algorithms excel at detecting patterns within structured data while MLP tools provide insights from unstructured sources such as clinical notes and patient-reported outcomes, supporting the creation of highly individualized management and treatment care plans.

Real-world implementation and ongoing clinical trials

In addition to retrospective model development, recent efforts have also focused on operationalizing AI tools for both clinical practice and within IBD clinical trials. A real-time deep learning system for Mayo Endoscopic Score classification has been evaluated in endoscopy units to assist in disease activity grading and reduce interobserver variability.⁵⁴ In parallel, AI-designed therapeutics such as ISM5411 have progressed through phase I trials, with phase II studies in UC expected to begin in late 2025, marking early translation of AI into therapeutic discovery.⁵⁵

Complementing these efforts, emerging translational frameworks are shaping how AI may be integrated into IBD trials and care pathways. Sedano et al.⁵⁶ proposed a comprehensive AI roadmap for IBD clinical trials, emphasizing its role in digital enrichment, automated eligibility screening, and dynamic outcome assessment. Similarly, Ahmad et al.⁵⁷ highlighted the utility of AI-assisted endoscopy for reducing variability in mucosal healing endpoints, a major challenge in trial reproducibility. These frameworks underscore the increasing recognition of AI as both a diagnostic and trial optimization tool, even as formal implementation studies remain limited.

Current limitations

While the integration of artificial AI into IBD care shows considerable promise, several limitations must be addressed before widespread clinical adoption is feasible. First, data heterogeneity across institutions and platforms presents a

significant challenge. Most AI models are trained on retrospective, single-center datasets using varying imaging protocols, histologic scoring systems, and clinical documentation standards, limiting cross-institutional reproducibility. Additionally, the lack of external validation for many models restricts their generalizability, with few undergoing prospective or real-world testing. Generalizability concerns are further compounded by narrow patient populations, particularly when training cohorts lack diversity in age, ethnicity, or disease phenotype. These limitations increase the risk of model performance degradation in broader clinical populations. Another major concern is algorithmic bias, which can arise from training datasets that reflect existing disparities in healthcare delivery or documentation. Without careful oversight, AI models may inadvertently perpetuate health inequities, particularly for underserved populations. From a practical standpoint, integration into clinical workflows remains challenging. Many AI tools are not designed with interoperability or end-user efficiency in mind, resulting in poor adoption or redundancy with existing systems. This highlights the importance of co-designing tools with input from clinicians, informaticians, and patients. Lastly, regulatory and ethical considerations including patient privacy, data governance, and the explainability of AI models pose ongoing hurdles. Technical pitfalls and overreliance on automation warrant caution. Many models, especially deep learning systems, are susceptible to overfitting, wherein performance on training data is strong but fails to generalize to unseen patient populations. In addition, the black-box nature of most AI algorithms can obscure how decisions are made, limiting interpretability and raising concerns about trust, transparency, and accountability in clinical care. As AI becomes more embedded in diagnostic workflows, there is also a risk of overreliance, where clinicians may defer too heavily to algorithmic outputs. This highlights the need to maintain clinical oversight and emphasizes that AI should augment—not replace—human judgment. Future development must prioritize explainability, validation in diverse settings, and continuous feedback mechanisms to ensure safe and effective deployment. Additionally, the impact of AI on healthcare disparities, particularly its potential to either exacerbate or reduce inequities in under-resourced settings, remains a critical area for ongoing evaluation.

Future directions and conclusion

Looking ahead, the future of AI in IBD lies in collaborative, cross-disciplinary innovation. Key priorities include building multi-institutional and diverse datasets to enhance generalizability, developing explainable AI models to enhance trust and improve interpretability; and embedding these tools into clinical decision support systems that are both intuitive and actionable for healthcare providers. The incorporation of multi-modal data including genomics, microbiome profiles, imaging, and patient-reported outcomes offers a unique opportunity to develop robust predictive models for disease trajectory, treatment response, and complications. Federated learning and other privacy-preserving AI frameworks may help overcome data-sharing barriers while enabling for broader validation efforts. AI has the potential to support treat-to-target strategies, streamline disease monitoring, and individualize therapy based through real-time data. As regulatory frameworks evolve and health systems embrace digital transformation, AI is poised to become an indispensable component of precision medicine in IBD.

In conclusion, while substantial challenges remain, the current trajectory of AI in IBD is promising. With continued research, rigorous validation, and ethical implementation, AI has the potential to revolutionize IBD care by enhancing diagnostic accuracy, reducing clinical variability, and enabling personalized treatment pathways.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author contributions

Raseen Tariq: Conceptualization; Data curation; Methodology; Writing – original draft; Writing – review & editing.

Anita Afzali: Conceptualization; Methodology; Supervision; Writing – review & editing.

Acknowledgements

None.

Funding

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

Competing interests

R.T.: None. A.A.: Consulting and/or speaking: AbbVie, Bristol Myers Squibb, Eli Lilly, Johnson & Johnson, Pfizer, Takeda; Grants and/or Research: AbbVie, Bristol Myers Squibb, Eli Lilly, Johnson & Johnson, Pfizer, Takeda; Other: Co-Founder and Board Member of IBD Horizons and Scrubs & Heels Foundation.

Availability of data and materials

Not applicable.

ORCID iD

Raseen Tariq  <https://orcid.org/0000-0001-7586-6691>

References

1. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019; 13(2): 144–164.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25(1): 44–56.
3. LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521(7553): 436–444.
4. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019; 25(1): 30–36.
5. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017; 2(4): 230–243.
6. Cohen-Mekelburg S, Berry S, Stidham RW, et al. Clinical applications of artificial intelligence and machine learning-based methods in inflammatory bowel disease. *J Gastroenterol Hepatol* 2021; 36(2): 279–285.
7. Lamash Y, Kurugol S and Warfield SK. Semi-automated extraction of Crohn's disease MR imaging markers using a 3D residual CNN with distance prior. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018)*. 2018; 11045: 218–226.

8. Puylaert CAJ, Nolthenius CJT, Tielbeek JAW, et al. Comparison of MRI activity scoring systems and features for the terminal ileum in patients with Crohn disease. *Am J Roentgenol* 2019; 212(2): W25–W31.
9. Ahmad HA, East JE, Panaccione R, et al. Artificial intelligence in inflammatory bowel disease: implications for clinical practice and future directions. *Intest Res* 2023; 21(3): 283–294.
10. American Society for Gastrointestinal Endoscopy Standards of Practice Committee; Shergill AK, Lightdale JR, Bruining DH, et al. The role of endoscopy in inflammatory bowel disease. *Gastrointest Endosc* 2015; 81(5): 1101–1121. e1–13.
11. Yantiss RK and Odze RD. Pitfalls in the interpretation of nonneoplastic mucosal biopsies in inflammatory bowel disease. *Am J Gastroenterol* 2007; 102(4): 890–904.
12. Flores C, Francesconi CF and Meurer L. Quantitative assessment of CD30+ lymphocytes and eosinophils for the histopathological differential diagnosis of inflammatory bowel disease. *J Crohns Colitis* 2015; 9(9): 763–768.
13. Yantiss RK, Das KM, Farraye FA, et al. Alterations in the immunohistochemical expression of Das-1 and CG-3 in colonic mucosal biopsy specimens helps distinguish ulcerative colitis from Crohn disease and from other forms of colitis. *Am J Surg Pathol* 2008; 32(6): 844–850.
14. Chierici M, Puica N, Pozzi M, et al. Automatically detecting Crohn's disease and ulcerative colitis from endoscopic imaging. *BMC Med Inform Decis Mak* 2022; 22(Suppl. 6): 300.
15. James JP, Nielsen BS, Christensen IJ, et al. Mucosal expression of PI3, ANXA1, and VDR discriminates Crohn's disease from ulcerative colitis. *Sci Rep* 2023; 13(1): 18421.
16. Manandhar I, Alimadadi A, Aryal S, et al. Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases. *Am J Physiol Gastrointest Liver Physiol* 2021; 320(3): G328–G337.
17. Bielecki C, Bocklitz TW, Schmitt M, et al. Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells. *J Biomed Opt* 2012; 17(7): 076030.
18. Rimondi A, Gottlieb K, Despott EJ, et al. Can artificial intelligence replace endoscopists when assessing mucosal healing in ulcerative colitis? A systematic review and diagnostic test accuracy meta-analysis. *Dig Liver Dis* 2024; 56(7): 1164–1172.
19. Xie W, Hu J, Liang P, et al. Deep learning-based lesion detection and severity grading of small-bowel Crohn's disease ulcers on double-balloon endoscopy images. *Gastrointest Endosc* 2024; 99(5): 767–777.e5.
20. Ferreira JPS, de Mascarenhas Saraiva M, Afonso JPL, et al. Identification of ulcers and erosions by the novel Pillcam™ Crohn's capsule using a convolutional neural network: a multicentre pilot study. *J Crohns Colitis* 2022; 16(1): 169–172.
21. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020; 91(3): 606–613.e2.
22. 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(2): 357–363.e2.
23. Furlanello C, Bussola N, Merzi N, et al. The development of artificial intelligence in the histological diagnosis of inflammatory bowel disease (IBD-AI). *Dig Liver Dis* 2025; 57(1): 184–189.
24. Rymarczyk D, Schultz W, Borowa A, et al. Deep learning models capture histological disease activity in Crohn's disease and ulcerative colitis with high fidelity. *J Crohns Colitis* 2024; 18(4): 604–614.
25. Carter D, Albshesh A, Shimon C, et al. Automatized detection of Crohn's disease in intestinal ultrasound using convolutional neural network. *Inflamm Bowel Dis* 2023; 29(12): 1901–1906.
26. Naziroglu RE, Puylaert CAJ, Tielbeek JAW, et al. Semi-automatic bowel wall thickness measurements on MR enterography in patients with Crohn's disease. *Br J Radiol* 2017; 90(1074): 20160654.
27. Stidham RW, Enchakalody B, Wang SC, et al. Artificial intelligence for quantifying cumulative small bowel disease severity on CT-enterography in Crohn's disease. *Am J Gastroenterol* 2024; 119(9): 1885–1893.
28. Yamamoto S, Kinugasa H, Hamada K, et al. The diagnostic ability to classify neoplasias occurring in inflammatory bowel disease by artificial intelligence and endoscopists: a pilot study. *J Gastroenterol Hepatol* 2022; 37(8): 1610–1616.

29. Noguchi T, Ando T, Emoto S, et al. Artificial intelligence program to predict p53 mutations in ulcerative colitis-associated cancer or dysplasia. *Inflamm Bowel Dis* 2022; 28(7): 1072–1080.
30. Lopez-Serrano A, Voces A, Lorente JR, et al. Artificial intelligence for dysplasia detection during surveillance colonoscopy in patients with ulcerative colitis: a cross-sectional, non-inferiority, diagnostic test comparison study. *Gastroenterol Hepatol* 2025; 48(2): 502210.
31. Lichtenstein GR, Loftus EV, Isaacs KL, et al. ACG clinical guideline: management of Crohn's disease in adults. *Am J Gastroenterol* 2018; 113(4): 481–517.
32. Rubin DT, Ananthakrishnan AN, Siegel CA, et al. ACG clinical guideline: ulcerative colitis in adults. *Am J Gastroenterol* 2019; 114(3): 384–413.
33. Buisson A, Pereira B, Goutte M, et al. Magnetic resonance index of activity (MaRIA) and Clermont score are highly and equally effective MRI indices in detecting mucosal healing in Crohn's disease. *Dig Liver Dis* 2017; 49(11): 1211–1217.
34. Gu P, Mendonca O, Carter D, et al. AI-luminating artificial intelligence in inflammatory bowel diseases: a narrative review on the role of AI in endoscopy, histology, and imaging for IBD. *Inflamm Bowel Dis* 2024; 30(12): 2467–2485.
35. Fan Y, Mu R, Xu H, et al. Novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. *Gastrointest Endosc* 2023; 97(2): 335–346.
36. Cai W, Xu J, Chen Y, et al. Performance of machine learning algorithms for predicting disease activity in inflammatory bowel disease. *Inflammation* 2023; 46(4): 1561–1574.
37. Puylaert CAJ, Schuffler PJ, Naziroglu RE, et al. Semiautomatic assessment of the terminal ileum and colon in patients with Crohn disease using MRI (the VIGOR++ Project). *Acad Radiol* 2018; 25(8): 1038–1045.
38. 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(5): 1570–1583.
39. Iacucci M, Parigi TL, Del Amor R, et al. Artificial intelligence enabled histological prediction of remission or activity and clinical outcomes in ulcerative colitis. *Gastroenterology* 2023; 164(7): 1180–1188.e2.
40. Ohara J, Nemoto T, Maeda Y, et al. Deep learning-based automated quantification of goblet cell mucus using histological images as a predictor of clinical relapse of ulcerative colitis with endoscopic remission. *J Gastroenterol* 2022; 57(12): 962–970.
41. Klein A, Mazor Y, Karban A, et al. Early histological findings may predict the clinical phenotype in Crohn's colitis. *United European Gastroenterol J* 2017; 5(5): 694–701.
42. 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(7): 1119–1125.
43. Zeng X, Jiang H, Dai Y, et al. A radiomics nomogram based on MSCT and clinical factors can stratify fibrosis in inflammatory bowel disease. *Sci Rep* 2024; 14(1): 1176.
44. Zhu F, Dong T, Tang C, et al. A mesenteric fat-derived radiomic model to identify colonic fibrosis and predict treatment response to biologics in chronic ulcerative colitis. *Dis Colon Rectum* 2024; 67(12): 1544–1554.
45. Zand A, Stokes Z, Sharma A, et al. Artificial intelligence for inflammatory bowel diseases (IBD); accurately predicting adverse outcomes using machine learning. *Dig Dis Sci* 2022; 67(10): 4874–4885.
46. Reddy BK, Delen D and Agrawal RK. Predicting and explaining inflammation in Crohn's disease patients using predictive analytics methods and electronic medical record data. *Health Informatics J* 2019; 25(4): 1201–1218.
47. Park SK, Kim YB, Kim S, et al. Development of a machine learning model to predict non-durable response to anti-TNF therapy in Crohn's disease using transcriptome imputed from genotypes. *J Pers Med* 2022; 12(6): 947.
48. Venkatapurapu SP, Iwakiri R, Udagawa E, et al. A computational platform integrating a mechanistic model of Crohn's disease for predicting temporal progression of mucosal damage and healing. *Adv Ther* 2022; 39(7): 3225–3247.
49. Solitano V, Ma C, Hanzel J, et al. Advanced combination treatment with biologic agents and novel small molecule drugs for inflammatory bowel disease. *Gastroenterol Hepatol (NY)* 2023; 19(5): 251–263.

50. Alric H, Amiot A, Kirchgesner J, et al. Vedolizumab clinical decision support tool predicts efficacy of vedolizumab but not ustekinumab in refractory Crohn's disease. *Inflamm Bowel Dis* 2022; 28(2): 218–225.
51. He M, Li C, Tang W, et al. Machine learning gene expression predicting model for ustekinumab response in patients with Crohn's disease. *Immun Inflamm Dis* 2021; 9(4): 1529–1540.
52. Park J, Chun J, Yoon H, et al. Feasibility of a clinical decision support tool for ustekinumab to predict clinical remission and relapse in patients with Crohn's disease: a multicenter observational study. *Inflamm Bowel Dis* 2023; 29(4): 548–554.
53. Vickers AJ, Van Calster B and Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016; 352: i6.
54. Møller BL, Lo BZS, Burisch J, et al. Building an AI support tool for real-time ulcerative colitis diagnosis. *Künstl Intell* 2024; 38: 1–8.
55. <https://www.prnewswire.com/news-releases/insilico-received-positive-topline-results-from-two-phase-1-trials-of-ism5411-new-drug-designed-using-generative-ai-for-the-treatment-of-inflammatory-bowel-disease-302344176.html>
56. Sedano R, Solitano V, Vuyyuru SK, et al. Artificial intelligence to revolutionize IBD clinical trials: a comprehensive review. *Therap Adv Gastroenterol* 2025; 18: 17562848251321915.
57. Ahmad HA, East JE, Panaccione R, et al. Artificial intelligence in inflammatory bowel disease endoscopy: implications for clinical trials. *J Crohns Colitis* 2023; 17(8): 1342–1353.

Visit Sage journals online
[journals.sagepub.com/
home/tag](https://journals.sagepub.com/home/tag)

 Sage journals