111353_Auto_Edited.docx

WORD COUNT	6368	TIME SUBMITTED	02-SEP-2025 03:26PM
		PAPER ID	117986805

Name of Journal: *World Journal of Gastroenterology*
Manuscript NO: 111353
Manuscript Type: REVIEW

Artificial intelligence in inflammatory bowel disease: Current applications and future directions

Horia Minea, Ana-Maria Singeap, Manuela Minea, Stefan Chiriac, Carol Stanciu, Anca Trifan

Abstract

Inflammatory bowel disease (IBD) represents a major global health concern, significantly impacting patient quality of life and healthcare systems. Mucosal and histological healing have emerged as key therapeutic targets, offering better long-term outcomes compared with previous targets. However, accurate disease assessment remains challenging because of interobserver variability and inconsistencies between endoscopic and histological findings. Artificial intelligence (AI) is transforming IBD care by enhancing the precision and reproducibility of disease evaluation. This review provided a structured synthesis of AI applications in IBD, organized by diagnostic, histological, and therapeutic domains, and highlighted comparative model performance such as machine learning classifiers (random forest, support vector machine) and deep learning models (convolutional and recurrent neural networks) with reported accuracy between 80% and 97% and areas under the curve ranging from 0.74 to 0.99. Beyond summarizing existing tools, the review emphasized the ability of AI to reduce diagnostic variability, improve early prediction of therapeutic response, and streamline clinical workflows. These advancements support a shift toward personalized treatment strategies and more efficient care delivery. Additionally, we outlined the expanding role of AI in clinical trials in which it supports patient stratification, endpoint prediction, and automated data integration.

Key Words: Inflammatory bowel disease; Artificial intelligence; Computer-aided diagnosis; Artificial intelligence-based endoscopy; Deep learning; Machine learning; Digital pathology

Minea H, Singeap AM, Minea M, Chiriac S, Stanciu C, Trifan A. Artificial intelligence in inflammatory bowel disease: Current applications and future directions. *World J Gastroenterol* 2025; In press

Core Tip: This review focused on the integration of artificial intelligence (AI) in inflammatory bowel disease management, highlighting its applications in diagnostics, surveillance, treatment management, and clinical trial optimization. AI enables more accurate, objective, and timely evaluations of endoscopic, histological, and clinical data. By synthesizing recent advances the article emphasized the potential of AI to enhance personalized medicine and support real-time decision-making. Despite promising developments challenges remain regarding standardization, validation, and ethical concerns. This comprehensive overview outlined future directions to ensure the safe and effective implementation of AI in routine inflammatory bowel disease care.

INTRODUCTION

Inflammatory bowel diseases (IBD) represent complex conditions, characterized by heterogeneous and recurrent phenotypes with particularities due to the variable impact of underlying pathogenic mechanisms[1,2]. Due to the multifactorial etiology, the diagnosis of IBD is considered highly complex, relying on multiple investigations that involve significant variability in result interpretation closely related to the training and experience of the observers[3,4].

Artificial intelligence (AI) was first described in the 1950s with limited early technological models, which prevented their widespread adoption in medicine. With the development of deep learning (DL) techniques in the early 2000s, there have been substantial improvements[5]. Figure 1 provides an overview of the evolution of AI technologies.

Therefore, a new research concept has been encouraged, focused on the development of computer-aided diagnostic (CAD) tools that rapidly process and integrate large amounts of data, offering the possibility to improve the holistic management of endoscopic, histological, or imaging results. By eliminating inconsistencies associated with potentially subjective scoring systems and optimizing the organization of data used to accurately characterize information relationships based on clinical outcomes, these tools support the development of predictive models[6,7].

In this context the AI-based approach has become a promising reality in gastroenterology for the identification of new biomarkers with diagnostic or prognostic potential and for providing decision support that has favored the implementation of personalized interventions for the advancement of precision medicine[8-10]. Figure 2 provides a comprehensive overview of AI integration across diagnostic, prognostic, and therapeutic domains in IBD care.

Currently, AI includes a wide range of computational models that address various areas, including the extraction of information from complex data sets, natural language processing (NLP) for interpreting clinical text, machine learning (ML) with constantly improved performance through the rapid integration of specifications provided by operators, and the application of DL that involves the use of complex algorithms, such as artificial neural networks or convolutional neural network (CNN), designed to establish statistical associations between variables used to adjust behaviors and establish predictions[11-13].

The implementation of AI has influenced medical practice in recent years, allowing the storage and structured organization of data for an accurate characterization of relationships based on clinical outcomes and the development of high-performance models for predicting future events. Moreover, the emergence of AI has provided a new perspective on understanding the pathophysiological mechanisms involved in IBD[6].

This review provided a structured synthesis of AI applications in IBD, organized by diagnostic, histological, and therapeutic domains, and it outlined the expanding role of AI in clinical trials in which it supports patient stratification, endpoint prediction, and automated data integration.

MATERIALS AND METHODS

This article presented a narrative, non-systematic review of the scientific literature identified through international databases such as PubMed, CrossRef, and ScienceDirect. The review period spanned from 2016 to 2025, reflecting recent advances in AI applied to the diagnosis and monitoring of IBD. The selection criteria focused on

studies exploring the use of AI technologies, including ML, DL, and neural networks, aimed at improving endoscopic and histological assessment in the context of ulcerative colitis (UC) and Crohn's disease (CD). Additionally, the review included studies addressing disease progression prediction, therapeutic response monitoring, and the optimization of clinical trial outcomes through the use of these advanced technologies.

AI-ASSISTED ENDOSCOPIC EVALUATION IN IBD

Currently, endoscopic and histological remission represent the main targets for assessing therapeutic efficacy in IBD. However, in clinical practice a lack of concordance between the two endpoints has been frequently observed. The discrepancy is most likely due to significant interobserver variability and the subjective nature of human evaluation[7].

Several studies have questioned the reliability and consistency of colonoscopic examination, highlighting important limitations in the accurate analysis[14,15]. To support the implementation of precision medicine, the complexity of diagnosing IBD requires the involvement of professionals with up-to-date clinical knowledge and advanced endoscopic skills. When colonoscopic examinations are performed by trainees or healthcare providers with limited expertise, the risk of misclassification or incomplete assessment increases along with the potential for missed lesions. Addressing these shortcomings often necessitates repeat investigations, which generate additional costs and delay therapeutic decision-making[16].

In recent years the integration of AI in the field of endoscopic imaging has stimulated the achievement of structural and functional goals that have improved the accuracy of diagnosis and reduced the necessity of invasive biopsies by discovering microscopic and molecular features that are imperceptible to the human eye[7,17]. As a result AI has emerged as a promising tool, oriented towards achieving an objective and reproducible analysis with precise image interpretation for classifying and discriminating clinical phenotypes, risk stratification, and the prediction of complications[4].

Although endoscopic scoring systems are accessible to apply, they do not accurately reflect the degree of clinical severity within each image category and show significant interobserver variability. Recent studies that tested different DL models applying random forest and CNN for colonic video analysis reported superior performance [accuracy 86.5%-94.5%, area under the curve (AUC) 0.94-0.98] compared with expert assessment[18-21].

Subsequently, Stidham *et al*[22] confirmed the potential of a CNN model for classifying severity in UC (AUC 0.966, sensitivity 83.0%, specificity 96.0%), and the results of the Mayo Endoscopic Score (MES) assessment were comparable with the agreement of experienced raters ($\kappa = 0.840$ *vs* $\kappa = 0.860$). Despite the high performance of the model, it had limitations related to the subjectivity of the evaluators in assessing the endoscopic score, the lack of data from multiple centers, and the evaluation method using images and videos captured at high resolution, which did not reflect the technical conditions in daily practice[22].

A recent meta-analysis that included 12 clinical trials highlighted the advantage of using a CNN to identify mucosal healing in still endoscopic images (sensitivity 0.91, specificity 0.89) and videos (sensitivity 0.86, specificity 0.91) in patients with UC[23]. Furthermore, the integration of AI into the analysis of these images improved the detection of restoration of intestinal mucosal vascularization, which is a predictor of relapses in the next 12 months[3].

The construction of a computer algorithm based on pixel integration allowed the definition of a red density score that was applied to identify patterns in endoscopic imaging in patients with UC. A red density score value below 60 was associated with the recognition of histological remission (AUC 0.950, sensitivity 96.0%, specificity 80.0%) with the results being strongly correlated with Robarts Histologic Index (RHI) ($r = 0.74$, $P < 0.0001$), MES ($r = 0.76$, $P < 0.0001$), and Ulcerative Colitis Endoscopic Index Of Severity (UCEIS) ($r = 0.74$, $P < 0.0001$). However, several limitations may hinder the generalizability of the findings. First, the data were collected from a reduced number of patients all enrolled at a single tertiary center. Additionally, the progressive inclusion of

patients at different disease stages resulted in a heterogeneous cohort, which may have impacted the robustness of the statistical analysis[24].

Promoting accurate management in IBD involves an objective assessment based on the association between endoscopic and histological examination, which is particularly useful for monitoring response and predicting clinical evolution. In this context Takenaka *et al*[25] created a deep neural network (DNN) for UC, which achieved excellent results (accuracy 92.9%, sensitivity 92.4%, specificity 93.5%) to detect histological inflammation in colonic videos, excluding the need for biopsy samples. The evidence of a significant correlation between DNN for UC and examiners for UCEIS ($\kappa = 0.917$) is an argument supporting the advantage of implementing AI in clinical studies to eliminate subjective interpretations. Although the score showed an excellent performance, the classification regarding the most severe lesion may influence the results and the interpretation of the algorithm. Hence, further multicentric studies are required for external validation[25].

Standardization of image acquisition is essential for training algorithms that reflect clinical scenarios. The quality of the examination could be improved by automated analysis of video recordings, facilitating the exclusion of low-quality data[26]. Since most studies focus on static images, selection bias frequently occurs. This has led to a new video-based approach to improve the reliability of the results. As a result, Yao *et al*[27] developed a CNN applied to unedited video frames to remove low-resolution images, allowing accurate differentiation between remission (MES 0-1) and active disease (MES 2-3) with an achieved accuracy of 83.7%[27]. AI improved the quality of the procedure by reducing sampling errors, optimizing photo documentation, preventing missing blind spots, and targeting areas of maximum interest. The WISENSE system was developed using the methods of CNN and DL to significantly reduce the rate of dead spots in endoscopic videos compared with the control group (5.86% *vs* 22.46%, $P < 0.001$) with an accuracy of 94.4%[28]. Therefore, ML algorithms could alert the endoscopist in real time regarding the necessity to re-

evaluate the patient when lower quality videos are detected or when inadequate colonic coverage occurs due to rapid scope withdrawal during the examination[29].

The application of AI in clinical trials is not limited to improving the analysis and interpretation of video images. The integration of a recurrent neural network for the automatic classification and scoring of endoscopic severity in patients with UC, determined increased rates of agreement with physician expertise for MES (κ = 0.844) and UCEIS (κ = 0.855), helping to standardize the results. In addition the more accurate classification of disease progression ensures rapid detection and intervention to reduce the risk of colorectal cancer associated with IBD[30].

The definition of new scoring systems focused on the distribution and extent of intestinal inflammation detected through AI-assisted colonoscopy offers the advantage of achieving a more detailed profile of disease activity. In the UNIFI study the introduction of a cumulative disease score generated by an ML algorithm for computerized evaluation of endoscopic videos in patients with UC receiving ustekinumab as induction and maintenance therapy more accurately quantified mucosal damage compared with MES (Hedges' g = 0.743 *vs* 0.460)[31].

Capsule endoscopy represents a complex investigation in clinical practice due to the abundance and diversity of visual data contained in video recordings that may exceed the perceptual acuity of the human eye[3]. By eliminating the clinicians subjectivity, the implementation of AI represents a significant role in optimizing the interpretation of images obtained by this investigation in patients with CD, helping to achieve excellent differentiation between strictures and normal mucosa (AUC 0.989, accuracy 93.5%), complemented by an objective classification of mild (AUC 0.992), moderate (AUC 0.750), and severe ulcerative lesions (AUC 0.889). However, potential misleading information may be influenced by the large image dataset and the absence of temporal dynamics due to using still frames. Moreover, the study included only patients with confirmed bowel patency thus underrepresenting severe strictures and a lack of generalizability[32].

On the other hand the combination of capsule endoscopy with DL techniques was found to be competitive (AUC 0.860, accuracy 81.0%, sensitivity 75.0%, specificity 84.0%) for predicting the need for biologic therapy in patients newly diagnosed with CD, outperforming fecal calprotectin (AUC 0.740) and human assessments (AUC 0.700). Possible disadvantages include the reduced sample size and the heterogeneity of the decision-making process regarding biologic therapy initiation, influenced by diverse clinical factors and individual preferences. Therefore, they may compromise the robustness of the conclusions. Furthermore, the predominance of patients with mild disease limits the generalizability of the findings and robustness of the conclusions[33].

Finally, the results of a multicenter study concluded that the application of a DL algorithm for the automatic selection and processing of relevant capsule endoscopy images significantly reduced the duration of the evaluation to a median review time of 3.2 min while maintaining high interobserver concordance for the diagnosis of CD ($\kappa = 0.910$), UC ($\kappa = 0.780$), and IBD ($\kappa = 0.920$)[34].

Confocal laser endomicroscopy (pCLE) represents an emerging technique, addressed for the *in vivo* assessment of histological activity and intestinal barrier permeability by enabling real-time visualization and characterization of the extension and the inflammatory infiltrate. This approach reduces the need for invasive biopsies to discriminate between clinical phenotypes and predict progression. Although it is considered a promising tool, it requires specific training and qualification of endoscopists to accurately interpret the fluorescence-enhanced images[7,35].

Recently, Iacucci *et al*[36] developed a computerized model for the quantitative *in vivo* analysis of pCLE images based on the binding of fluorescent molecules at the time of initiation of biologic therapy. This approach allowed the validation of vascular alterations (AUC 0.930, precision 85.0%), crypt area (AUC 1.000, precision 90.0%), and crypt eccentricity (AUC 0.810, precision 80.0%) as the most important predictors of therapeutic response in patients with UC. Interpretation of the results was restricted due to the reduced sample size and the selection of patients treated primarily with anti-tumor necrosis factor (TNF), limiting applicability to other therapeutic classes. The lack

of standardized protocols for fluorescein labeling and difficulties with immunofluorescence analysis may affect data reproducibility. In addition the practical application is hampered by the high cost and the need for advanced training for the use of pCLE. However, AI-assisted analysis may standardize image interpretation and reduce subjectivity in the future[36].

Computerized analysis of images obtained through pCLE provides more reliable and real-time results regarding the assessment of deep mucosal healing without the necessity for advanced training of medical personnel[37]. On the other hand it was found that the implementation of a CAD system for *in vivo* microscopic investigation of the intestinal mucosa using endocytoscopy in patients with UC offered an increased accuracy in identifying persistent histological inflammation (Geboes score $\geq$ 3) with a perfect reproducibility (k = 1.000) compared with the assessments of experienced pathologists[38].

Because AI-assisted endoscopic imaging provides a much more precise assessment of intestinal mucosal healing, it has a considerable contribution in improving IBD management, becoming a frequently integrated tool in clinical trials due to its accelerating potential for standardization and promising results for monitoring evolution and predicting therapeutic response[39]. However, further validation of the association of endoscopy with AI in multicenter studies is necessary to obtain significant clinical outcomes for patients. Table 1 presents a summary of the developed AI models for endoscopic evaluation in IBD.

AI IN HISTOPATHOLOGICAL EVALUATION OF IBD

In recent years histological healing has emerged as a new therapeutic target associated with a reduced risk of relapse in patients with IBD. Although histopathological examination represents a key role in confirming the diagnosis, it remains invasive and time-consuming[62]. The European Crohn's and Colitis Organization has proposed over 30 scores for classifying histological activity; however, their application in clinical practice is difficult due to a subjective and inconsistent interpretation that limits the

comparison and reproducibility of results, which may impact therapeutic decisions[62,63].

AI is considered a tool with the potential to enhance the diagnostic accuracy of biopsy specimens, focusing on the objective detection and characterization of specific histological features under standardized conditions, thus improving the assessment of inflammation and healing of the intestinal tract. Traditional histological examination is laborious and requires a long time for sample processing. In contrast AI systems have been observed to significantly reduce the time for evaluation, making them ideal when handling a large number of biopsies. Furthermore, by automating early detection, quantifying histological features, and combining them with clinical, laboratory, and imaging results, these tools help physicians make rapid decisions that optimize therapeutic strategies for the implementation of personalized medicine[12].

The integration of digitalized histological slides has enabled the acquisition of high-resolution images while the automated quantification of scoring systems that characterize the evolution of these diseases considerably reduced the examination time and reliance on human observer expertise compared to traditional histopathology[64,65]. Another opportunity associated with the implementation of AI algorithms in medical practice refers to the assistance provided to pathologists in identifying areas of interest in each biopsy sample[66].

The first attempt to develop a DL algorithm for quantifying eosinophil density in colonic biopsies of active UC, confirmed by the Geboes score and RHI, led to the definition of a histological score that provided strong concordance with manual eosinophil count performed by pathologists[20]. Similarly, the implementation of a CNN model for the identification and quantification of neutrophils and plasma cells in whole-slide images achieved impressive accuracy for assessing disease severity and predicting histological remission in patients with UC. Using a random forest classifier, based on 13 features derived from tissue and cellular patterns, the Nancy histological index (NHI) values produced by the model showed strong agreement with expert pathologists ($\kappa = 0.910$, $P < 0.001$)[67]. However, the model is strictly limited to

assessing the severity of the disease and does not detect other architectural distortions. Additionally, it does not take into account variation between fragments and has a reduced performance in detecting low disease activity (NHI 0-1)[68].

Using a DL model based on a CNN architecture, Gui *et al*[69] developed the Paddington International Virtual ChromoendoScopy Score Histologic Remission Index (PHRI), defined by the presence or absence of neutrophils in the lamina propria and intestinal epithelium. The implementation of this index demonstrated remarkable potential in distinguishing active from inactive UC with strong agreement between AI and human pathologists ($\kappa = 0.840$, $P < 0.001$)[69].

Subsequently, the validation of PHRI was performed on digitized biopsies using a CNN-VGG16 model that differentiated normal from inflamed intestinal mucosa with greater accuracy than conventional histological (RHI and NHI) and endoscopic (UCEIS and PICaSSO) scores. The hazard ratio for the exacerbation of the disease according to PHRI between the biopsy samples of the groups with histological remission was better for the assessment by AI compared with the pathologist (4.64 *vs* 3.56). However, the implementation of the score in digital pathology remains limited as it does not accurately discriminate between different degrees of inflammatory severity and does not detect the presence of dysplastic changes[70].

Mucin depletion represents an important predictor of relapse risk in patients with UC in clinical and endoscopic remission. However, the quantitative assessment presented a considerable interobserver variability. This limitation was assessed by Ohara *et al*[72] by introducing a DL model capable of automatically measuring mucus areas, reporting a higher risk of relapse in patients with goblet cell depletion compared with those without (45.0% *vs* 6.5%, $P < 0.010$)[71]. Furthermore, the authors enhanced the AI system by introducing semantic segmentation and object detection models for the localization and quantification of neutrophils in whole-slide images of the epithelium and lamina propria. This advancement improved the performance of deep remission assessment. In addition the two AI-based histological scores, PICaSSO and NHI, were positively correlated with pathologists' diagnoses (Spearman's $\rho = 0.68$-0.80, $P <$

0.050), being associated with an increased risk of relapse in patients with UC (hazard ratio 3.2-5.0, $P < 0.010$)[72].

Using automatic image processing algorithms, the severity of histological activity in patients with UC was assessed based on NHI, resulting in a performance comparable with expert pathologists (intraclass correlation coefficient 87.2%) and eliminating subjectivity while also ensuring the reproducibility of the results. However, it is subject to certain limitations, including potential bias during the initial labeling phase and the requirement for a large image dataset. Furthermore, the use of partial image tiles instead of full-slide images may impact the accuracy of the scoring process[73].

The development of a DL model, based on the use of the StarDist algorithm for digitized analysis of biopsy images, achieved an automatic quantification of basal plasmacytosis, which was shown to be a key feature for discriminating between IBD and normal non-IBD mucosa (odds ratio 4.968, 95% confidence interval: 1.835-14.638) with a precision similar to human experts (90.0%) and reduced subjectivity and evaluation time. However, the decreased number of patients might limit the generalizability of the results with a possible compensation due to a large number of biopsies per patient[74].

Most AI applications have predominantly focused on classifying disease activity, assessment standardization, and clinical outcome stratification, mainly in patients with UC[68]. In contrast research efforts in CD have focused on the development of DL systems that aim at the anticipation of therapeutic response and postoperative evaluation. Although AI applications in UC have recorded a continuous upward trend, the interpretation of histological results in CD has remained insufficiently covered[75]. This is caused by segmental and transmural involvement of the intestine, complicating the standardization, training, and implementation of an AI algorithm to establish the inflammatory activity[20]. Although numerous histological scores for CD have been proposed in recent years that differ depending on the number or type of segments examined and their microscopic characteristics, they have not been accepted for use in clinical practice due to the lack of validation[64].

The first DL analysis of histological images in patients with CD identified adipocyte shrinkage and mast cell infiltration as the main predictors, achieving high accuracy (AUC 0.995) for forecasting postoperative recurrence[76]. More recently, a DL system using a CNN was validated to assess the severity of histological lesions in the myenteric plexus (AUC 0.958, sensitivity 90.6%, and specificity 85.8%) and the muscle layer (AUC 0.966, sensitivity 92%, and specificity 89%) in patients with CD. By including clinical data the construction of a multitask model provides the prediction of the risk of postoperative recurrence (AUC 0.980) for establishing potential strategies to improve long-term outcomes[77].

The quantitative approach achieved with these models surpasses traditional qualitative histology based solely on visual inspection, allowing for a much more precise assessment of the healing process and real-time adjustments of the therapeutic strategy to achieve sustained remission. Table 2 summarizes the developed AI models for histological evaluation in IBD.

AI APPLICATIONS IN THERAPEUTIC MONITORING AND OUTCOME OPTIMIZATION IN IBD

The continuous expansion of AI implementation has revolutionized IBD management, enabling a targeted treatment approach based on clinical phenotype. Therefore, it facilitates the prevention of loss of response and identifies patients requiring early intervention to avoid disease progression[8,80]. Traditional methods used to define prognoses in IBD rely on statistical regression models, which are limited in their ability to analyze complex data sets involving repeated measurements[81].

The application of AI has promoted overcoming these limitations, offering the ability to adapt interactively as new information becomes available[5]. ML and NLP represent objective and consistent applications, trained on the extraction of data from electronic medical records with the potential to stratify patients based on the dynamic identification of different patterns associated with clinical and biological variables[82].

Despite the complex architecture of the AI algorithm, achieving meaningful benefits is not a guaranteed paradigm, depending on the quality of the training data[5]. Recent advancements have increased the performance of these models for providing real-time information that ensures pharmacovigilance compliance and increases adherence among patients with IBD[3].

One of the earliest applications represented longitudinal monitoring of thiopurine administration in patients with UC. Compared with traditional testing of the metabolite 6-thioguanine nucleotide, ML algorithms built on the collection and integration of laboratory data and patient age proved superior discrimination between remission *vs* non-remission among users. The prediction of potential adverse events associated with these medications has generated clinical benefits, including a reduction in the surgical intervention rates, hospitalizations, and steroid prescription[83].

The absence or progressive decline in response to biologic therapies increases healthcare costs in IBD[84]. In this context computational analysis of multiple clinical and molecular signals represents a promising area for identifying patients who are likely to respond to biologic therapies, both in CD and UC[85].

In general vedolizumab administration leads to a slow onset of favorable outcomes in patients with UC. In contrast the implementation of an ML algorithm based on laboratory data obtained at the onset of clinical evaluation predicted the establishment of endoscopic corticosteroid-free remission within the following 12 months[86].

Another example is the Endo-Omics study, which demonstrated that CAD of *in vivo* pCLE images based on morphological vessel or crypt abnormalities and fluorescein leak detection predicted the efficacy of anti-TNF and anti-integrin α4β7 therapy in patients with UC. In addition CAD showed a strong association of fluorescein-labeled biologic therapy (infliximab and vedolizumab) on biopsy specimens with a better response in UC than in CD[36].

Recently, an eXtreme gradient boosting (XGBoost) ML model, applied to a complex set of variables of UC patients (demographic and clinical data, physiological measurements, histology, serum biomarkers), improved the predictive accuracy of

remission after induction therapy with etrolizumab. Using the SHapley Additive exPlanations (SHAP) algorithm, key factors were identified to support the optimization of future treatment strategies[87]. Similarly, the ML approach for defining the probability of non-response to ustekinumab based on the combination of demographic data with the measurement of serum biomarkers (C-reactive protein and albumin) in the first weeks of clinical monitoring has proven to be valuable for reducing costs and improving CD management[88].

Although the use of anti-TNF-α therapy is effective in IBD, a significant number of patients do not develop a favorable response and are exposed to various side effects. In recent years AI has increased the safety and cost-effectiveness of this treatment to evaluate the clinical response and predict mucosal healing. Based on the integration of specific predictors, an ML model associated with the SHAP algorithm was designed and demonstrated excellent performance for discriminating responders *vs* non-responders[89]. Novel therapeutic targets proposed in the management of IBD have gone beyond clinical remission, focusing on mucosal healing[84].

The implementation of AI to enhance endoscopic evaluation is essential for predicting long-term clinical outcomes in patients with IBD, enabling physicians to make real-time treatment decisions[90]. In this context a DNN represents a consistent application for predicting this therapeutic target and reducing the risk of hospitalization or colectomy in patients with UC[44].

AI has become a field that has transformed and redefined clinical research, accelerating the development of novel therapeutic possibilities[26]. Hence, validation standards must be defined to ensure the credibility of these models and promote individualized therapeutic strategies to reduce the financial burden on healthcare systems[67]. Table 3 summarizes the AI models developed for evaluating treatment response in IBD.

APPLICATION OF AI IN CLINICAL TRIALS FOR IBD

Currently, AI has become an increasingly integrated component in the conduct of clinical trials applied in IBD with a remarkable capacity to optimize the various stages of the research process. These include simplifying participant recruitment to generate representative cohorts, improving trial design, and adopting high-performance criteria to ensure consistent evaluation, analyzing large volumes of data for precise phenotyping to predict response and discover new therapeutic targets[3,26].

In any clinical trial patient recruitment represents a major challenge, often leading to delays and high costs. AI has accelerated the strategy of selecting eligible candidates, harnessing the potential of ML and NLP for adaptive monitoring of medical records that was previously hindered by the static nature of traditional studies[95]. As a result electronic phenotyping focuses on identifying patients with similar clinical, endoscopic, or histological characteristics to reduce the cohort heterogeneity[96]. This approach significantly shortens the time needed to identify and recruit participants eligible for a specific clinical trial[3].

AI allows for improved control of participant adherence and early detection of possible dropout risks, helping researchers maintain the integrity of the study. Using tools such as chatbots and virtual assistants, AI enhances communication between participants and study organizers. An example is the myTrialsConnect platform that fosters interactions, collects specific data, and even schedules appointments, thereby improving participant accessibility and ensure efficient workload management[96].

The implementation of synthetic control groups or digital twins has been accepted as an alternative to the traditional use of placebo participants, favoring the reduction of patient dropout rates. Therefore, the concept is attractive for future studies by accelerating timelines and lowering costs. Additionally, it has become valuable for managing scenarios in which the conventional use of placebo groups is ethically questionable[3,97].

The possibility of automated integration of complex data sets, ranging from clinical symptoms to endoscopic and histopathological findings and to gene expression and other omics results, ensures a multiparametric analysis that provides additional

information to facilitate the conduct of randomized trials in IBD[98]. The algorithms have demonstrated increased potential for patient stratification, achieving a real-time classification of disease activity through the refinement of evaluation criteria that improve the interpretation quality of endoscopic and histological images[30,99]. Furthermore, it was observed that the use of automated quality of examination indicators substantially reduced interobserver variability in the interpretation of endoscopic videos and the need for a second reader to confirm the results while also providing real-time feedback[17,100].

By eliminating dependence on human experts, the implementation of AI to automate the assessment of severity in endoscopic (cumulative disease score) and histological (PHRI index) images has contributed to reducing the costs of clinical trials[3]. The application of a ML model for the detection of missed areas during colonoscopy, based on the estimation of the depth and position of intestinal lesions, achieved segment-by-segment assessment consistent with expert evaluation[29]. In contrast a CNN model applied to endoscopic images allowed for real-time detection and processing of undefined artifacts, restoring approximately 25% of corrupted videos to increase the overall frame retention rate[101].

Using AI in clinical trials is not without challenges. The performance of these tools depends on the accuracy of the data, which could impact the validity of the conclusions. AI models trained on biased data may perpetuate inaccuracies and discriminatory outcomes. Therefore, it is recommended to implement protocols and reporting guidelines, such as Interventional Trials-Artificial Intelligence (known as SPIRIT-AI) and Reporting Trials-Artificial Intelligence (known as CONSORT-AI) to ensure transparency, standardization, and reproducibility of results[3].

CURRENT CHALLENGES AND FUTURE PERSPECTIVES

Due to its ability to integrate an impressive volume of data, AI has revolutionized the implementation of personalized medicine, enabling a systematic investigation that provides a much more detailed perspective on the complexity of IBD[62,102]. Table 4

summarizes the main findings of ML algorithms applied to IBD. The introduction of CAD systems has eliminated the subjectivity of interpreting histological specimens and optimized the performance of endoscopic techniques, offering the possibility of identification of features that could be overlooked by human operators[104-106].

The advantage of achieving a more precise evaluation of therapeutic objectives, represents a decisive role in predicting the course of the disease[75]. Considering the heterogeneity of architecture and the diversity of technical approaches used to analyze input data, AI models may exhibit different behaviors and performances that influence implementation in clinical practice[8]. Despite the fact that the application of AI is increasingly accepted due to promising results in IBD, several limitations may prevent widespread implementation into clinical practice[7].

A critical issue is represented by the algorithmic transparency, especially in the case of DNN, which function as "black boxes," with their conclusions being difficult to interpret for both clinicians and patients. This opacity affects the ability to justify medical decisions, explaining why recent research promotes the use of explainability methods, such as SHAP or LIME, designed to increase the interpretability of automated decisions[107,108]. Secondly, although the performance of AI algorithms has been confirmed in numerous studies, their application in real-world scenarios is uncertain. The lack of consistent validation on external and independent data sets in randomized clinical trials favors the vulnerability of AI models, amplifying the risk of overfitting and the occurrence of interpretation errors that limit the accuracy and comparability of results, questioning the reliability of application in different clinical settings[109,110]. Furthermore, the variability of input data in terms of quality, type, and size of the information set may compromise the extrapolation of conclusions and recommendations, making it difficult to manage multiple clinical contexts of patients[8]. Another challenge lies in algorithmic bias. If the training data underrepresents certain subpopulations, such as ethnic minorities or patients with atypical disease phenotypes, AI outputs may influence the appearance of healthcare discrepancies. Algorithmic

fairness, bias mitigation strategies, and inclusive data collection are essential to ensure equitable performance[111].

Moreover, AI systems require ongoing maintenance and recalibration to remain clinically relevant. The need for continuous model retraining to adapt to evolving data distributions and clinical practices poses logistical burdens that may impede their application[112]. Therefore, it is essential to approach standardized protocols for training, testing, and validating AI systems. This ensures transparency and reproducibility of results, making these tools trusted solutions for healthcare professionals and regulators, contributing to improved therapeutic decisions and IBD management[113].

The increased cost associated with integrating these technologies represents another obstacle. This involves significant investments in the development and maintenance of a complex infrastructure from high-performance computing capabilities to data integration platforms and qualified personnel[114]. *Furthermore, it is also noteworthy that the global impact of incompletely resolved ethical and legal concerns regarding discrimination and ensuring the confidentiality of medical data and information is significant. In this context increased awareness should be directed toward exploring patient safety and establishing clear regulations regarding the malpractice implications generated by a potential error[26].* Patients need to be clearly informed about how their data are used, the role of AI in therapeutic decisions, and the limitations of these technologies. Lack of transparency or the perception of uncontrolled automation can erode patient trust in the healthcare team. It is therefore essential that informed consent processes include explanations of AI and respect patients' rights[96].

Protection of personal data is a major concern, given the confidential nature of medical information. These measures must be harmonized with national or international ethical standards and regulations on data protection in medical research[115]. Thus, it is a major priority, given the large volume of personal information involved in training and using algorithms. Existing regulations, such as the General Data Protection Regulation (known as GDPR) in Europe or the Health

Insurance Portability and Accountability Act (known as HIPAA) in the United States, provide a legal framework but are insufficient in the face of the specific challenges of AI. In the absence of adequate transparency, it becomes increasingly difficult to establish responsibility in case of decisional error. To address these limitations the SPIRIT-AI and CONSORT-AI guidelines, developed by international consortia of experts, provide the first methodological standards dedicated to clinical trials involving AI interventions. They recommend detailed reporting of the algorithm version, the source and processing of input data, human-algorithm interaction, validation processes, and the management of algorithmic errors. The use of these guidelines contributes not only to increasing the scientific quality of studies but also to ensuring an ethical and responsible integration of AI in precision medicine[116].

It is estimated that AI would become an innovative tool that could reduce dependence on the expertise of an endoscopist or a pathologist and minimize the number of biopsies collected during monitoring procedures[75]. **In the future the integration of data from multiple sources (clinical symptoms, endoscopic scores, histology, gene expression, transcriptomics, and other relevant variables) through AI tools may allow complex multiparametric analyses with the potential to extract new meaning from clinical trials. This integrative approach could help to define disease phenotypes much more accurately, refine patient subgroups, and optimize the design of future trials[30].** ML-based analysis of multiomics data (genomic, transcriptomic, proteomic, microbiomic) opens promising prospects for the discovery of clinically relevant biomarkers that can become therapeutic targets or surrogate markers for treatment response. Integrating these data into predictive models can support personalized medicine and guide therapeutic decisions[117,118].

In this context it is essential to set a concrete direction for the integration of AI into clinical practice, including standardization of data, external validation, and explainability of models, ensuring fairness, confidentiality, and transparency of decisions. Furthermore, collaboration with regulatory authorities and validation in

randomized trials are essential steps for the safe implementation of AI in the diagnosis and monitoring of evolution in IBD[75].

Another research direction is the development of AI models to support, rather than replace, clinical reasoning. These algorithms are based on quantifiable parameters, but medical decisions often involve judgments based on non-quantifiable factors. Therefore, future research should aim to integrate AI as a complementary tool to enhance clinical reasoning[119].

Although many challenges remain to be addressed, the ability of AI to integrate multimodal data offers the advantage of a much more comprehensive picture of disease activity and the healing process that is linked with the prevention of major adverse events[8].

CONCLUSION

The implementation of AI in clinical practice requires a clear vision of its objectives to generate advances that could lead to a paradigm shift in IBD management. Future directions should focus on optimizing diagnostic accuracy and validating the performance of AI models in diverse populations to enable integration into workflows and provide real-time decision support necessary for advancing precision medicine. As these technologies evolve a long-term proactive approach to addressing current challenges would enhance the efficiency and potential for personalization of clinical trials and ensure equitable benefits for patients.

111353_Auto_Edited.docx

ORIGINALITY REPORT

2%
SIMILARITY INDEX

PRIMARY SOURCES

1	downloads.hindawi.com Internet	29 words — < 1%
2	Giovanni Santacroce, Irene Zammarchi, Olga Maria Nardone, Ivan Capobianco et al. "Rediscovering histology – the application of artificial intelligence in inflammatory bowel disease histologic assessment", Therapeutic Advances in Gastroenterology, 2025 Crossref	28 words — < 1%
3	Marietta Iacucci, Giovanni Santacroce, Maeda Yasuharu, Subrata Ghosh. "Artificial Intelligence–Driven Personalized Medicine: Transforming Clinical Practice in Inflammatory Bowel Disease", Gastroenterology, 2025 Crossref	28 words — < 1%
4	www.ncbi.nlm.nih.gov Internet	23 words — < 1%
5	www.mdpi.com Internet	22 words — < 1%
6	Pieter Sinonquel, Alessandro Schilirò, Bram Verstockt, Séverine Vermeire, Raf Bisschops. "Evaluating the potential of artificial intelligence in ulcerative colitis", Expert Review of Gastroenterology & Hepatology, 2023 Crossref	12 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS