ABOUT THE AUTHORS

Amine Zoughlami, MD, MA

Dr. Amine Zoughlami is a fifth-year Adult Gastroenterology resident at McGill University. He completed his medical training and Internal Medicine training at McGill University, and a Master's in Bioethics at the Universite de Montreal. He will be completing an Advanced Inflammatory Bowel Disease Fellowship at Western University and has an interest in medical innovation and leveraging technological initiatives in gastroenterology.

Affiliations: *Division of Gastroenterology and Hepatology, Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada*

Adel Arezki, MD

Dr. Adel Arezki is a fourth-year urology resident at McGill University. He completed his medical training at McGill University, with a research focus at the intersection of urology and technology. His work centres on artificial intelligence, clinical prediction modelling and data-driven approaches to improve diagnostic accuracy and treatment outcomes in medicine.

Affiliations: *Division of Urology, Department of Surgery, McGill University Health Centre, Montreal, Quebec, Canada*

Edgard Medawar, MD

Dr. Edgard Medawar completed his medical training at McGill University. He is currently a third-year resident in internal medicine at the University of Ottawa, and is completing a Ph.D. in Experimental Medicine at the University of Montreal. Next year, he will begin his adult gastroenterology subspecialty training at the University of Montreal. Edgard has an interest in inflammatory bowel diseases and endoscopic tissue resection.

Affiliations: *Department of Medicine, University of Ottawa*
Division of Gastroenterology, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Quebec

doi.org/10.58931/cibdt.2025.3350

Talat Bessissow, MD, MDCM, MSc, FRCPC

Dr. Talat Bessissow earned his medical degree at McGill University (2005) where he then completed post-graduate training in Internal Medicine and Gastroenterology (2005-2010). In 2012, he trained in inflammatory bowel disease and advanced endoscopic imaging at the Gasthuisberg University Hospital, Leuven, Belgium under the supervision of professor Severine Vermeire. He also competed a Master in Experimental Medicine and Epidemiology from McGill University in 2016. Since 2012, He is a full time Associate Professor in the Division of Gastroenterology and Attending Staff at the McGill University Health Center. He is member of the McGill Inflammatory Bowel Disease (IBD) group as well as the McGill small bowel program. His current research focuses on the role and outcomes of mucosal healing in IBD as well as early detection of neoplastic lesions in ulcerative colitis. His researched has led him to publish over 150 peer-reviewed full papers. He is the past president of the Canadian IBD Research Consortium. He has also served as a reviewer for multiple national and international journals.

Affiliations: *Division of Gastroenterology and Hepatology, Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada*

Artificial Intelligence in Inflammatory Bowel Disease (IBD) Diagnostics: Applications and Future Directions

Amine Zoughlami, MD, MA
Adel Arezki, MD
Edgard Medawar, MD
Talat Bessissow, MD, MDCM, MSc, FRCPC

Key Takeaways
AI can improve the accuracy, objectivity, and reproducibility of IBD disease assessments across multiple disease assessment indices.
Multiple AI models have shown expert-level performance in the assessment of endoscopic and histologic activity in IBD.
The deployment of AI models can help uniformize the quality of disease assessment across academic and community centres alike.
The next steps will involve multimodal AI models. The development of these models, and the fine-tuning of unimodal systems, will require large, diverse datasets and careful governance.

Introduction

Management of inflammatory bowel diseases (IBD) relies on clinical, endoscopic, and histologic indices to assess disease activity and guide treatment. In practice, clinicians integrate multiple datapoints to formulate a treatment plan. Advances in artificial intelligence (AI) provide a unique opportunity to integrate these inputs to deepen both our understanding and assessment of the disease.

The role of disease activity indices is pivotal to the treat-to-target strategy recommended by the STRIDE-II consensus.[1] Yet, commonly used indices often face challenges such as subjectivity, low interobserver reliability, and limited granularity in evaluating severity or phenotypic differences. AI methods can address several of these issues. This brief narrative overview introduces core AI concepts that clinicians are likely to encounter in the future, and discusses key applications spanning clinical, endoscopic, histologic, and multimodal assessments of disease.

Artificial Intelligence – What is It All About?

AI refers to the computerization of tasks that would otherwise require human cognition, such as pattern recognition, problem-solving, and decision-making. Machine learning (ML), a subset of AI, refers to models that learn directly from data rather than being explicitly programmed to do so. Deep learning (DL), a subset of ML, uses *multiple* layers of neural networks to learn complex patterns.

AI models are classically trained in a supervised or unsupervised fashion. In supervised learning, the model learns from labelled data, for example, a model would be shown an image of an ulcer which would be labelled as such in the context of IBD. In unsupervised learning, models identify patterns in the data on their own. The models are trained on one dataset and tested on another. Generalizability refers to how well a model maintains its performance when applied to new data. Key pitfalls in model performance include overfitting (when a model learns from test data but fails to perform on new data), and underfitting (when a model is not exhaustive enough to capture patterns, leading to poor performance on both training and test sets). Overfitting may occur in contexts where the training data differs radically from test data, such as differences in endoscope models, image quality, or patient case mix. To mitigate overfitting, strategies such as using diverse datasets in addition to federated learning, in which models are locally trained and centrally aggregated.

Neural networks (NN) are a class of ML algorithms inspired by the interconnected structure of neurons in the brain. They consist of multiple layers, including an input layer, one or more processing layers, and an output layer. As the NN analyzes data, the strength of the connection between nodes varies to improve the quality of the output. Among NNs, convolutional neural networks (CNNs) are more commonly used for image and video processing and are widely applied in endoscopic tasks such as polyp detection. Natural language processing (NLP) also uses NNs to enable computerized understanding and generation of human language. An application of NLP is the development of large language models (LLMs), which are trained on large data sets to predict and generate language in a conversational manner, such as ChatGPT (OpenAI, San Francisco, USA).

These concepts constitute a brief overview of core AI principles. Together, these methods underlie the IBD applications discussed in this review.

Clinical Disease Activity in IBD – Only Part of the Answer

Clinical indices, such as the Crohn's disease (CD) activity index (CDAI), the Harvey-Bradshaw Index (HBI), and the partial Mayo Score (pMS) are widely used to assess disease activity, yet each contains subjective elements. The CDAI is vulnerable to interobserver reliability, at least partly due to its reliance on subjective evaluation in key items, such as "general well-being,"[2] and it may be markedly affected by recall bias. Although simpler and easier to use, the HBI and pMS are also partly reliant on subjective items. Furthermore, several items within these indices may be confounded by conditions such as irritable bowel syndrome, which overlaps with IBD in 7–25% of patients.[3,4] These limitations highlight a potential role for AI to complement symptom assessment by integrating data from different sources, and by the use of continuous, objective measures.

Outside of a clinic appointment, patients often communicate with their treating physician through phone calls, emails, or via

a patient portal. AI may be used to identify active disease during these interactions. For example, a recent study applied NLP to an IBD online forum and identified 20 surrogate markers of clinical flare derived from patient language.[5] This study highlights the potential for NLP to analyze other patient-generated data sources, such as messages, emails, and patient portal communications. Much time is spent reviewing interim clinical interactions. LLMs have demonstrated the ability to extract patient-reported outcomes from IBD-related clinical data,[6] and AI-based chart review systems can accurately identify extraintestinal manifestations within IBD clinical notes.[7] Similar systems can be used to reduce clinical time spent on chart reviews and effectively highlight relevant between-visit changes. However, an important caveat is input quality: note forwarding, incomplete charting, or lack of quantification all contribute to misclassification and poor accuracy.

An exciting frontier in the clinical assessment of disease activity is the emerging use of wearable health sensors. In a study involving a cohort of 309 patients equipped with consumer wearables, physiological data, including heart rate (HR), resting HR, HR variability, and oxygen saturation, were paired with daily symptom surveys and biochemical markers.[8] ML models were able to predict flares (defined as symptoms *with* corroborating biochemical evidence such as fecal calprotectin, C-reactive protein, and erythrocyte sedimentation rate) up to 49 days before onset. Continuous data collection through wearables may allow early identification of patients at risk of flares, permitting earlier testing and proactive assessments. While physiologic, non-invasive data from wearable devices offers valuable information through ML, the promise for wearable-acquired biochemical data is even greater. Future developments in wearable technology may allow for real-time sensing of biochemical data. A recently developed non-invasive, perspiration-based wearable can measure sweat calprotectin, interleukin-6, and C-reactive protein levels.[9] In the study, the sensor was able to distinguish between patients with endoscopically active versus inactive ulcerative colitis (UC) based on sweat calprotectin levels. As well, perspiration-based measurements of each marker showed moderate to strong correlations with corresponding serum levels. While longitudinal validation is pending, this proof-of-concept suggests an exciting future in which real-time evidence of inflammation can facilitate rapid triage, timely assessments, and treatment modifications.

Such innovations have the potential to shift real-time disease monitoring from a periodic, timepoint based model to a proactive model where changes and discussions can occur prior to the onset of a significant clinical status change.

Endoscopic Assessment in IBD – How We Can Do Better

Endoscopic evaluation of disease activity in IBD largely relies on the Simple Endoscopic Score in CD (SES-CD) and the Mayo Endoscopic Score (MES). The SES-CD evaluates ulcers, affected areas, and stenosis across segments and has shown good inter-rater reliability among central readers in clinical trials.[10] However, its generalizability and uptake in community practice remains uncertain. In contrast, the MES offers a simpler approach but may lack precision, as it relies on subjective thresholds, such as distinguishing mild friability from friability. AI-assisted endoscopic activity assessment models can provide a systematic and reproducible endoscopic disease activity index.

In a study by Gottlieb et al., 795 endoscopic videos from a phase 2 trial of mirikizumab in UC were centrally scored using MES and Ulcerative Colitis Endoscopic Index of Severity (UCEIS) by a single reader and then analyzed with a DL model, which showed strong agreement across both indices.[11] Similarly, Fan et al. trained an AI scoring system model for UC using still images, and tested it on 20 full-length endoscopic videos divided into five segments.[12] The model achieved concordance in 83% of segments with active disease and 100% of segments with inactive disease, and generated colourized colon maps—an intuitive graphical tool representing disease severity. These findings suggest that DL algorithms can identify and distinguish active and inactive disease at an expert-level.

Granularity remains a challenge in UC scoring. In a 2023 study by Kim et al. involving a UC cohort of 492 patients who demonstrated endoscopic improvements from MES 1 to 0, the endoscopic disease activity assessments of gastroenterologists was compared to that of a DL algorithm.[13] Results show the model outperformed the consensus of a group of gastroenterology fellows, providing more accurate results and superior ability to distinguish between MES 0

and 1. Notably, the algorithm maintained its performance level on an external dataset. These findings support the use of AI as an adjunct to improve scoring in UC, and allow for subtle discrimination near clinical thresholds.

In CD, the SES-CD relies on bias-prone assessments, such as ulcer size and affected surface. Marked interobserver variability can be noted in specific subscores, even among expert gastroenterologists.[14] An AI model trained to assess ulceration in CD was shown to have a strong correlation with the total SES-CD score, a moderate but significant correlation with fecal calprotectin levels, and, importantly, superior ability to identify clinical remission compared with SES-CD.[15] For small-bowel assessment, AI applications in video capsule endoscopy have rapidly progressed, outperforming gastroenterologists in both bleeding detection and review times.[16] In IBD, computer-assisted detection of erosions and ulcers has achieved sensitivity and specificity >90%,[17] with good discrimination between superficial and severe ulcers.[18] More recently, an AI-generated score for assessing small-bowel disease severity in CD was found to be strongly correlated to the Lewis Score.[19]

Despite strong results, heterogeneity remains. A recent meta-analysis revealed marked variability in AI accuracy for assessing mucosal healing in UC across datasets,[20] highlighting the need for standardized algorithm training, and extensive external validation.

Histologic Assessment in IBD – AI as the Great Equalizer?

Histologic remission is increasingly recognized as a potential treatment target in IBD, particularly in UC. However, histologic evaluation is labour-intensive, and requires subspecialty expertise, limiting its widespread adoption. Najdawi et al. trained a series of CNNs to identify tissues and cells, generating interpretable outcome features, including cell density and affected tissue areas.[21] From these, 13 features were selected by expert consensus as most predictive of outcomes, demonstrating strong correlation with the Nancy Histological Index, and achieving 97% accuracy in detecting histologic remission. Notably, the model's agreement with gastrointestinal pathologists matched inter-pathologist agreement, indicating expert-level performance.

AI-assisted histologic assessment can also predict outcomes. Using the PICaSSO Histologic Remission Index (PHRI), an AI model was able to predict clinical relapse with similar performance to expert pathologist assessment, with the AI generated results being obtained in as little as 9.8 seconds.[22]

These results illustrate how AI can democratize histologic expertise, especially in community settings where dedicated gastroenterology pathology may be limited.

AI in IBD – Putting It Together

Decision-making in IBD is inherently multimodal, and AI is helpful in interpreting heterogeneous signals. Chen et al. developed a clinical decision support tool that used only complete blood counts to non-invasively predict the extent and severity of colonic inflammation achieving an area under the receiver operating characteristic curve as high as 0.81 when differentiating between extensive colitis and proctitis on external validation data sets.[23] Additional data points can be integrated to assess disease severity with greater certainty and granularity. The integration of multiple disease activity parameters into one index has been recognized as potentially useful, particularly for the purpose of increasing sensitivity to therapeutic response in studies with smaller sample sizes.[24] Multimodal data integration with ML has also been applied to gene expression profiles to predict clinical response to advanced therapies,[25] or to models integrating clinical history and biochemical data to predict 1-year CD-related surgical risk.[26] These efforts highlight the potential of AI in optimizing treatment selection and prognostication.

Using data from a phase 2 trial of mirikizumab in UC, an AI fusion model that combined endoscopic and histologic data inputs outperformed individual single-modality models in predicting histologic remission.[27] This study provides an important proof-of-concept for using AI to integrate multiple disease activity inputs to better predict healing. Future research should explore the application of fusion models in predicting clinical and endoscopic outcomes.

Where Do We Go From Here?

The growing role of AI in IBD holds tremendous promise, but will require collaboration and care in its implementation. First, the use of diversified and multicentric datasets are a priority to protect against overfitting and improve generalizability. *Gastronet-5M*, a publicly available endoscopy dataset compiled from eight Dutch centres using different endoscope systems (Fuji, Olympus, Pentax), illustrates how diversified training datasets can improve model performances across a variety of endoscopy-related tasks.[28]

Second, AI should augment, but not replace, clinical judgment. Recent data has shown a decrease in adenoma detection rates during standard colonoscopies following AI-assisted colonoscopies, suggesting a risk of over-reliance on AI.[29] Maintaining clinicians' skills and autonomy will remain essential.

Third, interdisciplinary collaboration will be essential as IBD research increasingly recognizes the value of transmural assessment, and explores the potential of molecular and genetic markers. Equally, the implementation of AI tools should include community centres, where expertise and patient volumes in IBD may be limited, which will help to standardize care.

Ultimately, the integration of AI into IBD care represents a paradigm shift. When implemented responsibly, these tools will provide a much-needed level of objectivity and reproducibility to disease assessment. The next step will be prospective validation, across large multicentric datasets. AI holds the potential to support gastroenterologists in delivering care that is earlier, more precise, and, importantly, equitable for all patients with IBD.

Correspondence

Talat Bessissow, MD, MDCM, MSc, FRCPC
Email: talat.bessissow@mcgill.ca

Financial Disclosures

A.Z.: None declared.
A.A.: None declared.
E.M.: None declared.
T.B.: None declared.

References

1. Turner D, Ricciuto A, Lewis A, D'Amico F, Dhaliwal J, Griffiths AM, 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. doi:10.1053/j.gastro.2020.12.031
2. Sandborn WJ, Feagan BG, Hanauer SB, Lochs H, Lofberg R, Modigliani R, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. Gastroenterology. 2002;122(2):512–530. doi:10.1053/gast.2002.31072
3. Burdine LK, Rakowsky S, Grossberg L, Rabinowitz L, Center BIR, Cheifetz AS, et al. Irritable bowel syndrome/inflammatory bowel disease overlap: less common than we think. Gastro Hep Adv. 2024;3(8):1135–1137. Published 2024 Aug 10. doi:10.1016/j.gastha.2024.08.005
4. Fairbrass KM, Costantino SJ, Gracie DJ, Ford AC. Prevalence of irritable bowel syndrome-type symptoms in patients with inflammatory bowel disease in remission: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2020;5(12):1053–1062. doi:10.1016/S2468-1253(20)30300-9
5. Ricci L, Toussaint Y, Becker J, Najjar H, Renier A, Choukour M, et al. Web-based and machine learning approaches for identification of patient-reported outcomes in inflammatory bowel disease. Dig Liver Dis. 2022;54(4):483–489. doi:10.1016/j.dld.2021.09.005
6. Patel PV, Davis C, Ralbovsky A, Tinoco D, Williams CYK, Slatter S, et al. Large language models outperform traditional natural language processing methods in extracting patient-reported outcomes in inflammatory bowel disease. Gastro Hep Adv. 2025;4(2):100563. Published 2024 Oct 10. doi:10.1016/j.gastha.2024.10.003
7. Stidham RW, Yu D, Zhao X, Bishu S, Rice M, Bourque C, et al. Identifying the Presence, Activity, and Status of Extraintestinal Manifestations of Inflammatory Bowel Disease Using Natural Language Processing of Clinical Notes. Inflamm Bowel Dis. 2023;29(4):503–510. doi:10.1093/ibd/izac109
8. Hirten RP, Danieletto M, Sanchez-Mayor M, Whang JK, Lee KW, Landell K, et al. Physiological data collected from wearable devices identify and predict inflammatory bowel disease flares. Gastroenterology. 2025;168(5):939–951.e5. doi:10.1053/j.gastro.2024.12.024
9. Shahub S, Kumar RM, Lin KC, Banga I, Choi NK, Garcia NM, et al. Continuous monitoring of CRP, IL-6, and calprotectin in inflammatory bowel disease using a perspiration-based wearable device. Inflamm Bowel Dis. 2025;31(3):647–654. doi:10.1093/ibd/izae054
10. Khanna R, Zou G, D'Haens G, Rutgeerts P, McDonald JW, Daperno M, et al. Reliability among central readers in the evaluation of endoscopic findings from patients with Crohn's disease. Gut. 2016;65(7):1119–1125. doi:10.1136/gutjnl-2014-308973

11. Gottlieb K, Requa J, Karnes W, Chandra Gudivada R, Shen J, Rael E, et al. Central reading of ulcerative colitis clinical trial videos using neural networks. Gastroenterology. 2021;160(3):710–719.e2. doi:10.1053/j.gastro.2020.10.024
12. Fan Y, Mu R, Xu H, Xie C, Zhang Y, Liu L, et al. Novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. Gastrointest Endosc. 2023;97(2):335–346. doi:10.1016/j.gie.2022.08.015
13. Kim JE, Choi YH, Lee YC, Seong G, Song JH, Kim TJ, et al. Deep learning model for distinguishing Mayo endoscopic subscore 0 and 1 in patients with ulcerative colitis. Sci Rep. 2023;13(1):11351. Published 2023 Jul 13. doi:10.1038/s41598-023-38206-6
14. Khanna R, Feagan BG, Zou G, Stitt LW, McDonald JWD, Bressler B, et al. Reliability and responsiveness of clinical and endoscopic outcome measures in Crohn's disease. Inflamm Bowel Dis. 2025;31(3):706–715. doi:10.1093/ibd/izae089
15. Cai L, Wittrup E, Minoccheri C, Hiatt TK, Rice MD, Bishu S, et al. Artificial intelligence for quantifying endoscopic mucosal ulceration in Crohn's disease. Clin Gastroenterol Hepatol. Published online August 18, 2025. doi:10.1016/j.cgh.2025.05.026
16. Spada C, Piccirelli S, Hassan C, Ferrari C, Toth E, Gonzalez-Suarez B, et al. AI-assisted capsule endoscopy reading in suspected small bowel bleeding: a multicentre prospective study. Lancet Digit Health. 2024;6(5):e345–e353. doi:10.1016/S2589-7500(24)00048-7
17. Fan S, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys Med Biol. 2018;63(16):165001. Published 2018 Aug 10. doi:10.1088/1361-6560/aad51c
18. Barash Y, Azaria L, Soffer S, Margalit Yehuda R, Shlomi O, Ben-Horin S, et al. Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. Gastrointest Endosc. 2021;93(1):187–192. doi:10.1016/j.gie.2020.05.066
19. Cardoso P, Mascarenhas M, Afonso J, Ribeiro T, Mendes F, Martins M, et al. Deep learning and minimally invasive inflammatory activity assessment: a proof-of-concept study for development and score correlation of a panendoscopy convolutional network. Therap Adv Gastroenterol. 2024;17:17562848241251569. Published 2024 May 27. doi:10.1177/17562848241251569
20. Rimondi A, Gottlieb K, Despott EJ, Iacucci M, Murino A, Tontini GE. 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. doi:10.1016/j.dld.2023.11.005
21. Najdawi F, Sucipto K, Mistry P, Hennek S, Jayson CKB, Lin M, et al. Artificial intelligence enables quantitative assessment of ulcerative colitis histology. Mod Pathol. 2023;36(6):100124. doi:10.1016/j.modpat.2023.100124
22. Iacucci M, Parigi TL, Del Amor R, Meseguer P, Mandelli G, Bozzola A, et al. Artificial intelligence enabled histological prediction of remission or activity and clinical outcomes in ulcerative colitis. Gastroenterology. 2023;164(7):1180–1188.e2. doi:10.1053/j.gastro.2023.02.031
23. Chen H, Lin X, Pan X, Xu H, Zhang X, Liang G, et al. Development and validation of a blood routine-based extent and severity clinical decision support tool for ulcerative colitis. Sci Rep. 2023;13(1):21368. Published 2023 Dec 4. doi:10.1038/s41598-023-48569-5
24. Jairath V, Jeyarajah J, Zou G, Parker CE, Olson A, Khanna R, et al. A composite disease activity index for early drug development in ulcerative colitis: development and validation of the UC-100 score. Lancet Gastroenterol Hepatol. 2019;4(1):63–70. doi:10.1016/S2468-1253(18)30306-6
25. He M, Li C, Tang W, Kang Y, Zuo Y, Wang Y. Machine learning gene expression predicting model for ustekinumab response in patients with Crohn's disease. Immun Inflamm Dis. 2021;9(4):1529–1540. doi:10.1002/iid3.506
26. Stidham RW, Liu Y, Enchakalody B, Van T, Krishnamurthy V, Su GL, et al. the use of readily available longitudinal data to predict the likelihood of surgery in Crohn disease. Inflamm Bowel Dis. 2021;27(8):1328–1334. doi:10.1093/ibd/izab035
27. Iacucci M, Santacroce G, Meseguer P, Dieguez A, Del Amor R, Kolawole BB, et al. Endo-Histo foundational fusion model: a novel artificial intelligence for assessing histologic remission and response to therapy in ulcerative colitis clinical trial. J Crohns Colitis. 2025;19(7):jjaf108. doi:10.1093/ecco-jcc/jjaf108
28. Jong MR, Boers TGW, Fockens KN, Jukema JB, Kusters CHJ, Jaspers TJM, et al. GastroNet-5M: a multicenter dataset for developing foundation models in gastrointestinal endoscopy. Gastroenterology. Published online July 30, 2025. doi:10.1053/j.gastro.2025.07.030
29. Budzyn K, Romanczyk M, Kitala D, Kolodziej P, Bugajski M, Adami HO, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol. 2025;10(10):896–903. doi:10.1016/S2468-1253(25)00133-5