ADVANCES IN IBD

Current Developments in the Treatment of Inflammatory Bowel Disease

Section Editor: Stephen B. Hanauer, MD

Artificial Intelligence in Inflammatory Bowel Disease

Ryan W. Stidham, MD, MS
Associate Professor
Inflammatory Bowel Disease Center
Department of Internal Medicine
Department of Computational Medicine and Bioinformatics
University of Michigan
Ann Arbor, Michigan

G&H How can artificial intelligence be defined, and why is it being used in inflammatory bowel disease?

RS The colloquial concept of artificial intelligence (AI) represents a collection of analytic methods that are able to ingest data of any type—numerals, text, sounds, images, or videos—and find patterns that can predict an output (eg, a judgment, clinical outcome, or therapeutic decision). These analytic methods encompass types of machine learning, including support vector machines, ensemble methods such as random forest, and neural networks, with each tool suited for different types of data and situations. AI is neither artificial nor necessarily intelligent, although it can appear so by replicating complex human tasks and interpretations. The availability of data has fueled opportunities that show what AI is capable of achieving. It is very important to remember that AI is just another tool in a doctor's tool belt; it just happens to be exceptionally powerful when used for the right problem.

AI is appealing because treating inflammatory bowel disease (IBD) is difficult and requires many laborious, tedious, and time-consuming tasks. I would love to hand off tasks such as reviewing laboratory results for evidence of medication toxicity or efficacy, watching capsule endoscopy videos to look for ulcers, and grading endoscopic severity as a central reviewer, and these tasks need to be done competently, reliably, and fast, which AI can do. In addition, doctors often do not have the luxury of opening textbooks to figure out what to do or have complete confidence that a course of therapy will be helpful. They

also do not see every detail when interpreting images from computed tomography (CT), magnetic resonance (MR) imaging, and endoscopy despite understanding the nature and severity of disease. AI provides opportunities to quantify medical imaging in new ways that are impossible for humans to achieve. AI's capabilities for self-learning and recognizing patterns have the potential to improve decision aids and support tools. AI may even be able to write clinic and procedural notes. It can help remove much of the labor needed for great care and provide guidance for difficult decisions in IBD.

G&H How well can AI replicate IBD experts?

RS It depends on what AI is trying to replicate, but in certain cases, AI's performance is excellent. This is particularly true for replicating expert interpretation of imaging. Endoscopy is critical for determining disease severity and therapeutic response. Like art, doctors know disease when they see it, but reproducibly grading severity and quantifying exactly what constitutes a Mayo score of 1 vs 0 in ulcerative colitis (UC) is not very clear. The central reading approach essentially trained humans to act like machines and provide less biased, more experienced, and more rigorous reporting of disease activity in IBD. This was much better but still imperfect, as bias remained, disagreements could not completely be avoided even among experts, the approach was time-consuming, and only so many readers were available.

Neural networks excel at learning the specific features of images and their complex interactions to enable

> … AI neural networks can grade IBD endoscopic severity for still images and full-motion videos with accuracy indistinguishable from that of experienced reviewers but with high speed, near-perfect reproducibility, and minimal cost.

biometric identification such as fingerprint and facial recognition. Dozens of groups have shown that AI neural networks can grade IBD endoscopic severity for still images and full-motion videos with accuracy indistinguishable from that of experienced reviewers but with high speed, near-perfect reproducibility, and minimal cost. These AI systems can be deployed in clinical trials, academic research settings, and day-to-day clinical practice. The PICASSO study has shown that AI can review pathology and both categorize and grade histologic remission in UC with accuracy again indistinguishable from that of an expert pathologist. Taking it further, the PICASSO team also showed that using AI-pathology image analysis could predict clinical outcomes in UC at 1 year. It is also starting to become possible to interpret CT enterography and MR enterography, and soon ultrasound, automatically with AI image recognition systems. These cases highlight how AI can be trained to encode the knowledge and experience of experts to perform very specific image interpretation tasks with excellent accuracy.

G&H What is the current status of IBD decision support tools that use AI?

RS Clinical decision support tools (CDSTs) that are reliable and functional are still far away, although there are a few examples of useful CDSTs, usually for very specific decisions. In work pioneered by my colleagues Dr Peter Higgins and Dr Akbar Waljee, a machine learning laboratory data analysis known as THIOmon has been shown to provide very good prediction of optimized thiopurine dosing. THIOmon is enabled in the University of Michigan's electronic medical record (EMR) and, importantly, provides an explanation of why the model made its prediction. As use of thiopurines is becoming less

common, AI methods are being applied to help optimize biologic medication dosing by combining drug levels and laboratory and clinical data.

However, most CDSTs currently available are certainly artificial and not very intelligent. Many researchers in the clinical data science space are working hard to use AI to build prediction models to help with difficult treatment decisions. Despite encouraging performance for predicting outcomes, most AI predictions do not impact management. Consider a patient whose small bowel Crohn's disease has failed to respond to 2 biologics and whose doctor is considering a third biologic vs proceeding with surgery. I made an AI prediction model with an 80% accuracy for predicting therapeutic success or ultimate surgery at 1 year. However, I have never used it in practice because patients have more to lose with surgery than medical therapy, patients usually do not want surgery, and 80% accuracy is not good enough for this decision. What if the accuracy for medical failure was 90%, 95%, or 99%? Decision curve analysis and other methods have been used to determine this cut point, but 95% to 99% accuracy is often needed for caregivers to act on divergent CDST recommendations.

The question is how to get CDSTs to sufficient accuracy to make a difference in decision-making. AI models do not have all of the information that is available to a seasoned clinician. For example, perhaps the AI model has knowledge of prior medications and fecal calprotectin levels and has even automatically interpreted endoscopy and cross-sectional imaging. However, the AI model does not know the influential disease details not captured by the endoscopic score that an endoscopist would qualitatively appreciate. The AI model would not know that a patient's prior biologic use was very effective but was stopped not because of failure but because of an adverse event or financial coverage change. AI models are powerful, but they need the privileged information that is not easy to put into a data set without a lot of work. Natural language processing (NLP) and other sophisticated machine learning methods are powering better ways to comprehensively collect information about patients to give AI more information and likely better predictions.

G&H Can you expand on the use of NLP in AI?

RS NLP uses multiple machine learning techniques to mimic the language comprehension of a human and teaches machines to "read." Think about all of the detailed information in all of the notes and documents reviewed when meeting a new patient (often hundreds of pages). The notes are digitized in the patient's EMR, but must still be read by a human to obtain the information.

NLP can be trained to extract this information, structure it, and make it accessible to AI CDSTs. Reading is very complicated, and NLP is more than just finding keywords. NLP systems must identify and understand disease concepts, synonyms, grammar, syntax, temporal references, and modifiers. NLP must also have methods to adjudicate conflicts, assign priority to references, and handle inconsistent and varied documentation styles of the author's prose. NLP applications are just beginning to

> Rather than aiming for perfect correctness, it may be more important to emphasize the explainability of AI so that doctors can clearly evaluate the evidence and rationale of AI decisions …

be used in IBD, and promising results are being seen for specific and targeted applications, including extraction of medication use behavior, symptom severity, and historical interventions. These tasks are simple for a human but very challenging for a machine. However, once trained, NLP can be scaled to interpret numbers of documents that are not feasible by human reviewers. In the near future, NLP will find and extract specific information from text documents. Not much later, NLP will be providing a longitudinal synopsis of an IBD patient's medical history, making chart reviews history.

G&H What are the limitations of AI in IBD?

RS Currently, AI is only as good as its teachers. Neural networks used for endoscopic image analysis replicate an expert's knowledge and judgment, as well as that person's bias and potential fatigue captured in training. Overcoming these limitations is difficult. Using increasingly large image data sets or many human experts is appealing, but large numbers do not solve all of the problems. Regression to the mean includes error and bias that is difficult to dilute by sample size. Another vital aspect of developing effective AI is ensuring that a diverse population of patients is included in the training set. Socioeconomic, racial, and ethnic considerations too often remain an afterthought in study design, and assuming that all groups behave the same has been repeatedly proven to be fraught with problems.

Another challenge is the concept that a single correct answer may not uniformly apply to the qualitative judgments relied on in practice and research. In my own work, in cases where AI disagrees with the human expert, a blinded adjudicator is often equally as likely to pick the AI response as the human response for being the most correct. Rather than aiming for perfect correctness, it may be more important to emphasize the explainability of AI so that doctors can clearly evaluate the evidence and rationale of AI decisions, whether they involve the endoscopic severity of UC or the escalation of a biologic.

Finally, patients have a difficult time trusting AI in medicine. In a survey, patients were asked whether they would want an operation performed autonomously by an AI robot with a 2% rate of complication or a human surgeon with a 15% rate of complication. Patients overwhelmingly selected the human surgeon, which is illogical. The trepidation may stem from human errors being understandable, whereas AI errors can be erratic and unexplainable. Familiar examples are autonomous car accidents owing to vehicles suddenly braking for no clear reason or misinterpreting their own reflection for another vehicle. Explainable AI, known as XAI, is vital for the adoption of future AI tools.

G&H What concerns exist regarding the use of AI in IBD care?

RS I have no concern about losing my job to AI. I am hopeful that AI will enable management of patient populations by helping to expand delivery of expert care using AI for monitoring, useful alerting, and select automated decision-making. I do have concerns that the efficiency, speed, and unloading of traditional IBD clinical labor provided by AI could further increase the productivity demands, volume, and ultimate scope of responsibilities for clinicians. It will soon be known whether AI reduces, increases, or has no net effect on clinician labor while hopefully treating more patients better and faster.

Malpractice and liability should also be considered in the era of AI in IBD. For example, who is responsible if a clinician disagrees with the AI recommendation and the patient has a bad outcome, an endoscopist decides not to use AI-enhanced polyp detection during colonoscopy and a patient develops colon cancer the following year, or a physician follows the AI recommendation and a bad outcome ensues? In these scenarios, is the clinician, AI manufacturer, model developer, or hospital liable for undesired or bad outcomes? This is a complex and fluid legal landscape.

Finally, my greatest concern is privacy. The entire health care industry has ever-increasing access to rich, specific, detailed medical data. Patients are rightfully and understandably interested in what is done with their data. Sensitized by the transgressions of some social media companies, patients want to know how their own data are used and whether the data being generated in routine care are being monetized for commercial products. Patients also want to know whether the data they gave permission to be used for one purpose are later being used for a different purpose, commercial interest, or specific company (eg, the Cambridge Analytica scandal). Collecting background data with insufficient or, frankly, no consent for use in AI product development at scale presents real danger for breaching the trust of patients. Current concepts of secondary data use and 1-time consent authorizing indefinite sharing or repurposing of anonymized data to any party need to be updated for the 21st century and the new expectations of patients. The health care industry must be extremely clear in what will be done with an individual's data, why the data are needed, and who could potentially use the data. Patients should be empowered to direct how their information is and could be used and be given the tools to be in reasonable control of access to that information. The overwhelming majority of patients enthusiastically support medical innovations and often ask how they can help. It is the responsibility of the entire health care industry to foster complete transparency of intent, purpose, and possible use of AI in IBD using simple and proactive language and consent.

G&H What does the future hold for AI in IBD?

RS In the near term, AI tools for image recognition in endoscopy, pathology, and cross-sectional imaging will be deployed in IBD applications. Central reviewers in clinical trials will initially be augmented by an AI assistant to provide a preliminary interpretation or corroborate a human's interpretation. However, it will not be long before AI demonstrates it can function independently. AI image analysis will move into the clinical space; AI will be part of endoscopy processors, CT scanners, and slide scanners. Using image recognition, endoscopy notes will be automatically written by software detailing all of a doctor's findings and interventions.

Shortly thereafter, AI will exit the replication phase and move into the information-generation phase. Image analysis will extract new measures of disease that better capture activity and prognosis than conventional concepts. NLP will reliably collect detailed granular information from text records. Together, these more comprehensive information sources will power future CDSTs and make them more useful for real-world decision-making. AI will likely be used to manage less-complex IBD patients, hopefully giving clinicians more time with IBD challenges that still require a human touch. There will also be an explosion of pathophysiologic insights relying on AI for detailed measurements of transcriptional and cellular behavior that offer new directions for therapeutic developments.

Eventually, AI will reach an independent phase, where it can teach itself about IBD. This will be the era of gastroenterology general intelligence, where self-learning machines need very little prompting to understand the surrounding world. I do not know what will happen with intuitively, context-aware AI, but this phase will come about.

Disclosures

Dr Stidham holds intellectual property related to IBD image analysis for endoscopy and cross-sectional imaging that has been licensed by the University of Michigan to AMI, EIQ, and Prenovo. He has received research funding related to AI development from Janssen Research and Development as well as AbbVie.

Suggested Reading

Stidham RW, Takenaka K. Artificial intelligence for disease assessment in inflammatory bowel disease: how will it change our practice? *Gastroenterology*. 2022;162(5):1493-1506.

Stidham RW, Yu D, Zhao X, et al. Identifying the presence, activity, and status of extraintestinal manifestations of inflammatory bowel disease using natural language processing of clinical notes [published online June 3, 2022]. *Inflamm Bowel Dis*. doi:10.1093/ibd/izac109.

Waljee AK, Sauder K, Patel A, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J Crohns Colitis*. 2017;11(7):801-810.

Zand A, Stokes Z, Sharma A, van Deen WK, Hommes D. Artificial intelligence for inflammatory bowel diseases (IBD); accurately predicting adverse outcomes using machine learning [published online April 27, 2022]. *Dig Dis Sci*. doi:10.1007/s10620-022-07506-8.