Artificial Intelligence for Understanding Imaging, Text, and Data in Gastroenterology

Ryan W. Stidham, MD, MS[1,2,3]

[1]Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan

[2]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan

[3]Morphomics Analysis Program, University of Michigan, Ann Arbor, Michigan

Corresponding author:
Dr Ryan W. Stidham
University of Michigan
1500 East Medical Center Drive
3912 Taubman Center
Ann Arbor, MI 48109
Tel: (734) 936-9377
E-mail: ryanstid@med.umich.edu

Abstract: Artificial intelligence (AI) could change the practice of gastroenterology through its ability to both acquire and analyze information with speed, reproducibility, and, potentially, insight that may exceed that of human medical specialists. AI is powered by computational methods that allow machines to replicate clinical pattern recognition used by gastroenterology specialists to interpret endoscopic or cross-sectional images; understand the meaning and intent of medical documents; and merge different types of data to infer a diagnosis, prognosis, or expected outcome. Ongoing research is studying the use of AI for automated interpretation of text from colonoscopy and clinical documents for improved quality and patient phenotyping as well as enhanced detection and descriptions of polyps and other endoscopic lesions, and for predicting the probability of future therapeutic response early in a treatment course. This article introduces emerging technologies of natural language processing, machine vision, and machine learning for data analytics, and describes current and future applications in gastroenterology.

Artificial intelligence (AI) has arrived, touching industries from entertainment and education to manufacturing and medicine. AI is a concept of machine capabilities to independently collect information and then make measurements, judgments, and predictions in the context of prior knowledge. AI is powered by machine learning, a collection of computational methods used to learn from patterns and relationships in training data to predict outcomes or events. Packaged in both the hope and hype of AI are concepts of accuracy, speed, reduced costs, and improved insights beyond what humans can perceive. These potential facets of AI make it an attractive tool for helping to realize the promises of precision, value, and innovation in health care through technical innovations. Supporting the emergence of AI in medicine is the maturation of 3 pillars. First,

high volumes of instantaneously available information and outcomes are now available as a result of the digitization of clinical, laboratory, and imaging medical data into electronic health records (EHRs). Second, machine learning can now extract meaningful information from unstructured data sources, including medical images and office notes, at speed and scale. Finally, modern machine learning analytics are better suited to aggregate and process the diverse types of data encountered in medicine to predict outcomes and provide new physiologic insights. These advances in medical information acquisition and analytics have resulted in computational techniques being more useful than ever in medicine.

As a specialty reliant upon imaging, gastroenterology is well-suited to leverage the opportunities afforded by AI. Advances in machine vision have led to improvements in automated detection, description, and quantification of disease features on endoscopy. Familiar applications of AI in gastroenterology include replicating expert-level detection of colonic polyps, distinguishing benign from malignant tissue without histology, and grading disease severity. Other technologies beyond image recognition will power AI in the coming years for use in research, population health, and day-to-day clinical care. This article introduces AI technologies used in gastroenterology, including natural language processing (NLP) for extracting knowledge from text, neural networks for image analysis, and machine learning for predictive analytics, and includes a nonexhaustive list of example applications.

## Automated Text Analysis Using Natural Language Processing

Text documents, including specialist office notes, pathology and radiology results, and even patient telephone and e-mail notes, are rich sources of clinical information. These documents offer far more detail on phenotype, symptom severity, and patient behavior than what is captured by diagnostic codes, laboratory results, and medication orders alone.[1] However, unlike administrative billing codes and laboratory data, which are structured and readily available, text information is trapped within documents. Conventional methods of chart review for extracting and organizing information from documents are time-consuming, expensive, and error-prone. The ability to intelligently and automatically collect information from text documents, known as NLP, is an emerging AI technology that will be relevant to gastroenterologists.

NLP is a collection of computational approaches to automate the extraction and transformation of information from text documents into usable structured datasets better suited for analysis. Understanding language is a complex task, and simple keyword or phrase searches are

of limited utility. Comprehension of written language requires knowledge of concepts and their synonyms, grammar, the relationships between phrases, and temporal references. For example, consider using office notes to identify patients with active fistulizing Crohn's disease and encountering this sentence: "We had concerns that periianal disease was present, but T3 MR–pelvis was negative." Concluding the absence of fistulizing disease is intuitive for a person but complex for machines (Figure 1). NLP information extraction and inference of document meaning begins with text preprocessing, including spelling and punctuation cleaning or stripping. Then, concepts or their synonyms are identified, including anatomy, symptoms, diagnostic tests, medications, or procedural topics, typically through a process called named entity recognition using public or commercial concept reference libraries. Sentences or phrases are then broken down into grammatical elements using a part-of-speech analysis to determine if the word or phrase is a past participle, preposition, noun, or interrogative; this information is used to link concepts. Using these elements, analytic techniques such as hidden Markov models or support vector machines can be employed to infer meaning of the phrase for information extraction.

NLP uses complex rules-based methods and machine learning strategies, in conjunction with publicly available analytic resources and reference libraries, to aid in AI understanding of medical documents. The National Library of Medicine supports a continually updated medical metathesaurus called the Unified Medical Language System, which contains a hierarchical listing of medical concepts, including anatomy-, symptomatology-, pathology-, and medication-related terms that serve as a concept reference. Several open-source toolkits provide software packages to help with grammatical analysis for relationship linkage, negation and affirmation detection, and logic inference. A commonly used coding resource is the Natural Language Toolkit (Python). However, other open-source software packages requiring less coding knowledge are available, including the clinical Text Analysis and Knowledge Extraction System (Apache).[2] The Clinical Language Annotation, Modeling, and Processing Toolkit developed at the University of Texas Health Science Center at Houston offers a graphical user interface to design NLP tasks for medical information retrieval.[3] Successive iterations of NLP technology are increasingly user-friendly and designed for noncomputer scientists (eg, physicians) to build NLP tools to suit their needs.

Early applications of NLP in gastroenterology include aiding clinical workflows, facilitating quality assurance practices, and improving clinical phenotyping. Imler and colleagues developed NLP systems to extract information regarding colonic adenomas from
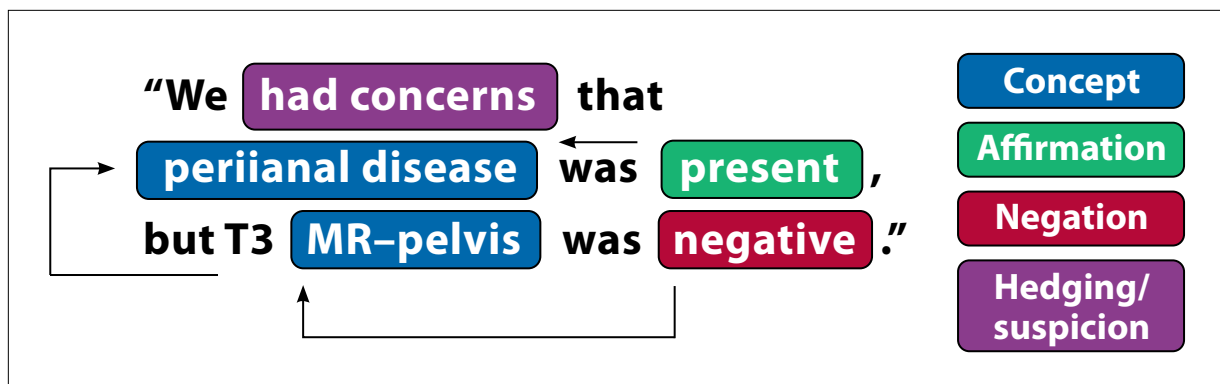
Figure 1. A conceptual general description of natural language processing (NLP). In this example, the NLP task is to determine whether a document indicates the presence or absence of fistulizing Crohn's disease. First, relevant concepts and their synonyms—here, it is perianal disease (disease complication) and diagnostic tests (MR imaging of the pelvis)—must be identified, and misspellings (periianal) and conceptual synonyms (perianal disease's relationship to fistulizing disease) must be detected. Second, affirmation and negation concepts (the MR test was negative) must be identified. Third, temporal references, including *prior*, *present*, *was*, and *had*, must be identified. Finally, grammatical understanding based on part-of-speech analysis is applied to determine the relationships between concepts, negation and/or affirmation phrases, and temporal references. Complex NLP analytics attempt to detect hedging or suspicion statements to identify and quantify an author's documented uncertainty regarding related concepts.

MR, magnetic resonance.

colonoscopy and pathology reports for the purpose of automating both adenoma detection rate calculations and guideline-based cancer surveillance recommendations.[4] In 750 paired colonoscopy and pathology text reports from 13 different centers, NLP performance for automated extraction of adenoma data was excellent, with an accuracy of 94.6% to 99.6% for classifying the histologic lesion type, 87.0% to 99.8% for lesion localization, and 92.0% for correct adenoma count. Groups have also used NLP to automatically detect endoscopic quality measures using report documents. NLP automatically extracted 19 different quality measures of colonoscopy (eg, cecal intubation, withdrawal time, bowel preparation), with an overall accuracy of 89% compared to experts across all quality measures.[5] A similar study of 13 quality measures from 24,674 documents on endoscopic retrograde cholangioscopy reported NLP accuracy of 84% to 100%.[6]

NLP has been shown to improve the accuracy and validation of diagnoses assigned to patients. In inflammatory bowel disease (IBD), diagnostic International Classification of Diseases (ICD) codes have been shown to have an accuracy of only 69% compared to manual expert document review.[7] Automated NLP systems increase the accuracy of IBD diagnosis in large datasets to 97%, representing a 12% improvement in classification accuracy over optimized administrative data.[8] Similar results for improving diagnostic accuracy have been reported for

liver diseases, where NLP methods outperformed both ICD codes and free text search for correctly distinguishing nonalcoholic fatty liver disease from other liver diseases, as assessed by F2 scores (NLP, 0.92; ICD, 0.34; free text search, 0.81).[1] Finally, NLP may automatically capture the character and severity of symptoms. In 4108 IBD patients, NLP models identified not only the presence of extraintestinal manifestations but their degree of activity, with an overall sensitivity, specificity, and accuracy of 81.8%, 92.9%, and 91.2%, respectively.[9]

NLP may provide a new information source to power future predictive analytics and decision-making tools, although significant limitations remain. Perfect NLP is a herculean effort considering the complexity of language, variation in documentation style between physicians, and the simple fact that clinical notes may be incomplete or incorrect. Research examining extraintestinal manifestation activity classification in IBD found that between 15.4% and 54.2% of physician notes were determined to be ambiguous,[9] and the presence of quality measure documentation in colonoscopy reports varied from 14.6% to 86.1% between hospitals.[10] Hypothetically perfect automated document review using NLP is still subject to the quality and completeness of the source documents. Despite these barriers, NLP implementation will continue to expand in research, clinical care, and education, with speculative expectations of what the

future holds. Automated outside document review, or scanning through thousands of pages of records to quickly find and organize information relevant to gastroenterologists, is likely to be built into EHRs. Groups are also exploring the application of NLP AI chatbots to collect patient symptom and history information as well as its application to student training, with aspirational goals of ultimately providing therapeutic advice.[11] Finally, NLP methods are being used for sentiment analysis to discern a patient's or physician's emotional tone in e-mails, text messages, transcribed telephone notes, and social media posts.[12]

Automated Image Analysis and Recognition Using Machine Vision

Similar to the wealth of information trapped within text, abundant information is also found within clinical imaging, but extracting it at scale is tedious or impossible. Machine vision, also known as computer vision, is a series of computational methods used for automated image analysis. Machine vision is garnering attention for its ability to reproduce expert-level interpretation of medical imaging with excellent performance and high speed. In gastroenterology, examples of machine vision include detecting colonic, gastric, and esophageal findings; distinguishing dysplastic from benign lesions; and grading the severity of mucosal damage from endoscopic images with comparable performance to experts.

The mechanics of machine vision involve using large sets of images that have been labeled or annotated by expert reviewers for the presence, absence, or location of a finding of interest to train and then test computational models aimed at replicating expert performance. Modern machine vision methods utilize artificial neural networks, which analyze data and identify patterns similar to the interconnectivity of natural biologic neural networks. Images are subsampled into smaller groups of pixels, which are then transformed or convolved by filters analyzing specific image characteristics called layers. Layer filters used early in the neural network detect simple image characteristics, such as color intensity or high-contrast edges and boundaries, but deeper layers become increasingly more complex, analyzing abstract features. Increasing filter complexity in deeper layers is what is referred to as deep learning. The interactions between the output of the convolutions of each layer are aggregated and analyzed to determine quantitative patterns that are associated with the expert label, such as the presence of a polyp; this is called a convolutional neural network (CNN; Figure 2). The CNN is then tested on images that were unseen in training to evaluate its performance compared to that of experts. While neural networks can be used with many types of data, they have particular advantages in handling images.

Endoscopy can be revolutionized by machine vision, and mounting research is providing example applications, principally in replicating expert-level interpretation and judgment. Numerous examples of automated polyp detection using endoscopic still images and real-time video are now available, with accuracy improving from approximately 70% over the last 5 to 10 years to more than 90% in more recent studies using modern CNNs.[13-15] In a prospective study of 1000 patients undergoing colonoscopy, machine vision assistance resulted in a small but significant increase in the adenoma detection rate (34% vs 28%; P=.03), although most additionally detected polyps were diminutive.[16,17] Beyond detecting polyps, machine vision can analyze sophisticated endoscopic image features that are difficult to standardize and challenging for human operators to learn. Using narrow-band imaging, researchers trained a CNN classification system to distinguish small adenomas from hyperplastic polyps with an accuracy of 94% (95% CI, 86%-97%).[18] Similarly, endocytoscopic systems, which provide up to 520-fold magnification of the mucosal surface, offer rich histologic information that is hindered by the challenges of interpretation by nonexperts. CNN models have matched endocytoscopic experts in image interpretation accuracy for distinguishing polyp histology (96.0% vs 94.6%; P=.141), but notably could substantially outperform gastroenterology trainees (96.0% vs 70.4%; P<.0001).[19] Finally, despite the performance of AI image recognition, some gastroenterologists have questioned the incremental value added by adopting AI technologies in endoscopy.[20] Highlighting a study assessing the impact of AI on clinical decision-making, Jin and colleagues demonstrated that a machine vision assistance system improved the accuracy of novice endoscopists' discrimination of hyperplastic from adenomatous polyps from 73.8% to 85.6% (P<.05), which is similar to the accuracy of experts (89.0%; P=.102).[21]

When tasked with other clinical image recognition tasks beyond the detection of polyps, machine vision has demonstrated similar approximation of expert interpretation for the detection of dysplasia in Barrett esophagus,[22] as well as both small bowel angioectasias and ulcerations on video capsule endoscopy with more than 95% accuracy.[23-25] Proof-of-concept studies support the potential for automated CNN models to replicate endoscopy grading of ulcerative colitis with similar agreement compared to expert reviewers (κ=0.86 vs κ=0.84) for exact Mayo endoscopic score, translating to an area under the curve (AUC) of 0.97 for distinguishing remission from active disease.[26] Another group reported similar results using the Ulcerative Colitis Endoscopic Index of Severity score, but
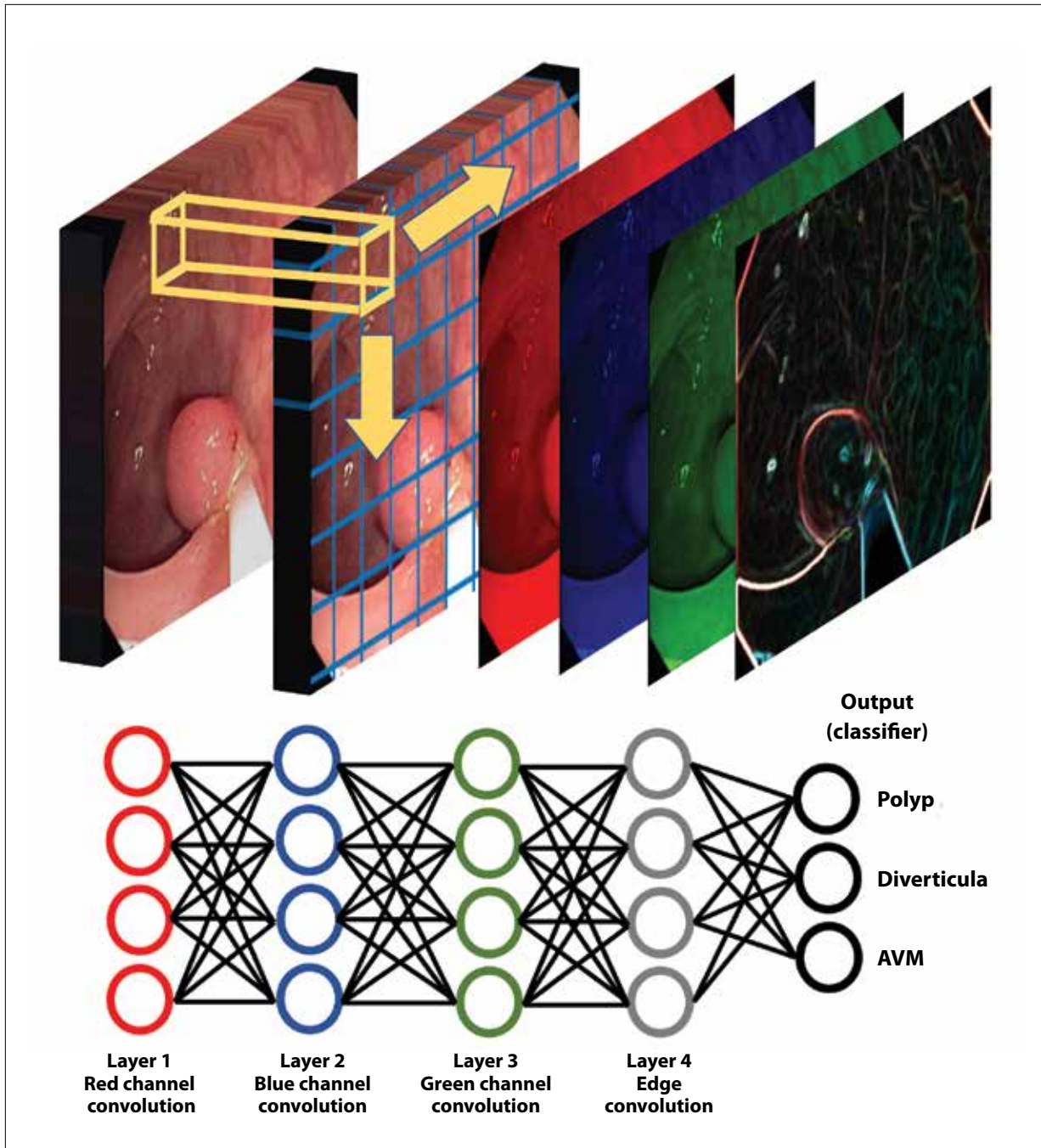
Figure 2. Basic concepts of convolutional neural networks for image classification. An image is subsampled into groups of pixels to facilitate analysis, as demonstrated by the yellow rectangle scanning the image to generate smaller image partitions of equal size. Subsampled images are then passed through analysis layers. Layers contain image filters designed to detect image properties. There can be dozens or hundreds of layers with filter properties, including color intensity, vertical edges, horizontal edges, and textures, with more complex filters in deeper layers. In each subsample of the image, the degree of closeness to the filter properties is calculated; this is the convolution. The network of convolutions can then be analyzed for patterns of interconnected features associated with an overall image classification provided by an expert. For further technical detail, see the article by Liu and colleagues[46] and the accompanying commentary.

AVM, arteriovenous malformation.

added that CNN-based deep learning analysis of endoscopic still images could predict histologic remission with 92.9% accuracy, highlighting the potential for inferring pathologic activity from endoscopic image analysis.[27]

Automated image analysis applications are equally relevant in gastrointestinal radiology and pathology. Outside of classifying images (eg, dog vs cat; adenoma vs hyperplastic polyp), machine vision can also segment images into their component parts using CNN methods. In Crohn's disease, researchers piloted automated segmentation of enterography studies to collect tedious, but clinically important, bowel measurements, including wall thickness, dilation, and lumen diameter with indistinguishable agreement compared to paired expert radiologists.[28] Using automatically extracted measurements, this group also predicted radiologist judgment on classifying disease as stricturing vs nonstricturing with an accuracy of 84.4% to 87.6%. Similar segmentation concepts have been applied to gastrointestinal pathology and interpretation of digitized images from histologic slides. In a prospective study of 102 children from 3 countries, CNN models automatically differentiated celiac disease, environmental enteropathy, and normal cases using duodenal biopsy images with a case detection of 93.4% and a false-negative rate of 2.4%.[29] Interestingly, the CNN model was able to indicate the regions of the pathology slide that resulted in the prediction. The capability to display regions of clinical or medical interest on an image affords opportunities in biologic mechanism and drug discovery.

Expected near-term roles for machine vision in gastroenterology surround themes of lesion detection assistance, quality assurance, and standardization of severity grading to improve reliability, interobserver agreement, and time efficiency. Computer-assisted polyp detection systems are likely to be directly incorporated into next-generation endoscopy hardware, potentially with histology inference systems to distinguish benign from dysplastic lesions, making high-confidence resect-and-discard practices a reality. Colonoscopy quality assurance for features such as bowel preparation and provider-level adenoma detection rate are both likely to be automated using the discussed technologies.[30] Video capsule endoscopy review should be expected to be augmented by AI lesion detection, with 1 study reporting reduction in review time to only 3 to 4 minutes compared to 40 to 50 minutes without AI assistance.[23]

Machine vision applications to date have impressive performance in replicating expert interpretations, but many important limitations remain. Expert opinion is not identical. Further, all current AI machine vision systems will incorporate any bias, subjectivity, and variation that is contained in the ground truth. Great care

in standardizing and qualifying the reference sets used as the gold standard for training AI will be critical and should involve regulatory, professional society, and practicing physician stakeholders. Work from the AI4GI academic-commercial collaborative provides an example of the degree of discrepancy between expert endoscopic and histologic opinion in colonic polyp interpretation, as well as opportunities for AI methods to aid and adjudicate disagreements.[31] In 644 biopsied colonic polyps, disagreement of endoscopy and pathology diagnosis occurred in 28.9% of cases. The authors highlight that pathology is hindered due to sampling or histology processing artifacts and may not provide a superior ground truth compared to endoscopy due to these limitations. Analytic methods may be able to merge the strengths of expert assessments in histology, endoscopy, and molecular science for improved disease assessments as an ensemble opinion rather than a single reference truth source, changing our concepts of reliance on a singular gold standard.

Despite real-world images and video having great variation in collection and digitization methods and image quality, as well as the potential for physiologic noise including debris, excessive stool, iatrogenic bleeding, and anatomic variation, most training and testing datasets are carefully crafted and are of uniform high quality. Next-generation machine vision methods will need to account for real-world variation and noise without the need for painstaking cleaning and labeling to advance current AI capabilities. Early examples include work in Barrett esophagus developing point-of-care visual analysis systems to identify dysplastic tissue.[32] Rather than using optimal exemplar still images or simply providing classification alone, an automated esophageal analysis system used in a 14-patient pilot study distinguished nondysplastic Barrett esophagus from early adenocarcinoma using real-time endoscopic streams and also localized the lesions to aid resection efforts with an accuracy of 89.9%.[32] These advancements in development may power the next wave of image-driven technologies in gastroenterology, including improved histologic inference, therapeutic monitoring, comprehensive disease assessments, and, eventually, robotic endoscopic procedures.[33]

Modeling Data to Predict Outcomes Using Artificial Intelligence Analytics

Improvements in computational analysis of large, high-quality datasets are powering predictive models for clinical outcomes, physician behavior, and patient behavior in gastroenterology. Most predictive models rely on a dataset in which an outcome, event, impression, or judgment is known for a given observation. Multiple machine learning methods can be used, with common

examples including artificial neural networks, support vector machines, and random forest techniques; their technical functions are summarized in several reviews.[34,35] Ultimately, these methods optimize the classification of an event, which in clinical settings is often a future outcome (eg, survival or therapeutic response). While machine learning models do not always outperform traditional statistical methods such as logistic regression, they are better suited to manage real-world data, where information may be missing, be imbalanced, or contain many variables.[35]

Predicting therapeutic response is challenging and often imprecise, but machine learning models of IBD outcomes provide examples of AI capabilities to aid gastroenterologist decision-making. Using data from a phase 3 clinical trial of vedolizumab (Entyvio, Takeda) in ulcerative colitis, random forest ensemble models combining baseline and week 6 clinical and laboratory data were able to predict corticosteroid-free endoscopic remission at week 52 with an AUC of 0.73 (95% CI, 0.65-0.82).[36] Similarly in Crohn's disease, machine learning models using baseline and week 8 data from phase 3 trial data for ustekinumab (Stelara, Janssen) predicted biologic responders beyond week 42 with an AUC of 0.78.[37] AI may also aid with personalizing medication dosing. Among thiopurine users with IBD, machine learning models predicted future clinical and biologic response in IBD better than thiopurine metabolite levels (AUC, 0.79 vs 0.49) using only serial complete blood counts and comprehensive metabolic panels.[38,39] Clinical outcomes in other domains also show utility for improving existing clinical predictions. Machine learning models can predict the 1-year survival rate in cirrhotic patients with 90% accuracy, as well as the probability of rapid hepatitis C virus progression to advanced fibrosis.[40] Neural networks have predicted survival duration following liver transplantation with improved results compared to traditional regression models (86.4% vs 80.7%), although the margin was minimal.[41] Similar applications of AI are being studied for a broad range of conditions, including predicting future response to neoadjuvant chemotherapy in rectal cancers[42] and recurrent bleeding from peptic ulcers,[43] with substantial improvement over traditional statistical methods.

In the near future, gastroenterologists can expect a multitude of AI-powered prediction models and clinical decision support tools, likely built into many EHRs. Some may be very useful, others may be suspect, and many will predict outcomes that are already known to be true. As EHR adoption increases and automated extraction of information from text and imaging improves, so too will the quality, utility, and value of AI predictive analytics for virtually any condition where a dataset can be generated. Further, AI models will have the capability to continually update both predictions and the models themselves in real time. Key issues facing the implementation of AI-decision support tools include qualification processes, audit of predictive performance, and the evolving degree and types of oversight recommended by regulatory agencies.

Challenges to Artificial Intelligence Implementation in Gastroenterology

Anticipation for AI solutions in gastroenterology is palpable. However, numerous issues need resolution before implementing AI into clinical decision-making. First, incorporating AI fundamentals into medical training and continuing professional education will be essential. Technologic advances in automation of AI system design will soon allow physicians to design custom AI solutions without the assistance of computer engineers.[44] Second, safety is of paramount importance and, although machine learning assistance aims to improve the quality and consistency of care, AI-derived clinical actions, diagnoses, or prognoses have the potential to harm patients. Regulatory agencies should be involved in supervising both the validation of AI technologies as well as providing oversight for implementation. The US Food and Drug Administration has published its Digital Health Innovation Action Plan detailing the handling of AI through the lens of software as a medical device, which will be distinct from regulatory guidance for traditional mechanical devices or pharmaceuticals.[45] Additional practical considerations are verification of AI reliability in different health care settings as well as intersystem interoperability, both of which will be important performance metrics for automation systems. AI will require access to large amounts of data, likely to be interchanged between systems, highlighting the need for new cyber security practices to prevent exposure of personally identifiable health information and tracking the use of an individual's information. Finally, legal considerations of who is responsible for the quality of AI systems, as well as the consequences of following, or perhaps not following, AI predictions have yet to be determined. Liability for untoward patient events may be held by physicians, technology vendors, or both.

Conclusion

Both well-qualified and potentially insufficiently validated varieties of AI may arrive in the coming years to the gastroenterology space. Most near-term AI image analysis technology will be tasked with replicating expert interpretation of endoscopic features, pathology slides, and cross-sectional imaging. Potential early introductions of AI in gastroenterology may be in the form of rapid preliminary interpretations, automated second opinions

for consensus of specialist judgment, visual assistance systems for endoscopy, and standardization of disease activity grading in clinical trials. Similar to the teleradiology and digital pathology movements of the last decade, AI applications in telehealth may be of greatest value in low-resource areas and are anticipated to focus on digital imaging. AI may increase efficiency, reduce the volume of tedious tasks, improve clinical outcomes, and, if implemented correctly, could reduce burnout and enhance the time shared between physicians and patients. The promise of AI will require new collaborations between medical specialists and engineers, rigor in validation and testing, and an open mind for changing the practice of gastroenterology for the better.

Disclosures

Dr Stidham serves as a consultant for AbbVie, Janssen, Merck, Takeda, and Corrona, LLC. He has received investigator-initiated research support from AbbVie. The University of Michigan has filed patents on his behalf related to the use of computer vision for imaging applications in gastroenterology, with technology elements licensed to AMI, Inc.

References

1. Van Vleck TT, Chan L, Coca SG, et al. Augmented intelligence with natural language processing applied to electronic health records for identifying patients with non-alcoholic fatty liver disease at risk for disease progression. *Int J Med Inform*. 2019;129:334-341.

2. Masanz J, Pakhomov SV, Xu H, Wu ST, Chute CG, Liu H. Open source clinical NLP—more than any single system. *AMIA Jt Summits Transl Sci Proc*. 2014;2014:76-82.

3. Soysal E, Wang J, Jiang M, et al. CLAMP—a toolkit for efficiently building customized clinical natural language processing pipelines. *J Am Med Inform Assoc*. 2018;25(3):331-336.

4. Imler TD, Morea J, Kahi C, et al. Multi-center colonoscopy quality measurement utilizing natural language processing. *Am J Gastroenterol*. 2015;110(4):543-552.

5. Harkema H, Chapman WW, Saul M, Dellon ES, Schoen RE, Mehrotra A. Developing a natural language processing application for measuring the quality of colonoscopy procedures. *J Am Med Inform Assoc*. 2011;18(suppl 1):i150-i156.

6. Imler TD, Sherman S, Imperiale TF, et al. Provider-specific quality measurement for ERCP using natural language processing. *Gastrointest Endosc*. 2018;87(1):164-173.e2.

7. Hou JK, Tan M, Stidham RW, et al. Accuracy of diagnostic codes for identifying patients with ulcerative colitis and Crohn's disease in the Veterans Affairs Health Care System. *Dig Dis Sci*. 2014;59(10):2406-2410.

8. Ananthakrishnan AN, Cai T, Savova G, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis*. 2013;19(7):1411-1420.

9. Stidham R, Yu D, Lahiri S, Vydiswaran V. P311 Detection and characterisation of extra-intestinal manifestations of IBD in clinical office notes using natural language processing. *J Crohns Colitis*. 2020;14(suppl 1):S309-S310.

10. Mehrotra A, Dellon ES, Schoen RE, et al. Applying a natural language processing tool to electronic health records to assess performance on colonoscopy quality measures. *Gastrointest Endosc*. 2012;75(6):1233-1239.e14.

11. Reiswich A, Haag M. Evaluation of chatbot prototypes for taking the virtual patient's history. *Stud Health Technol Inform*. 2019;260:73-80.

12. Carrillo-de-Albornoz J, Rodríguez Vidal J, Plaza L. Feature engineering for sentiment analysis in e-health forums. *PLoS One*. 2018;13(11):e0207996.

13. Fernández-Esparrach G, Bernal J, López-Cerón M, et al. Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. *Endoscopy*. 2016;48(9):837-842.

14. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*. 2018;155(4):1069-1078.e8.

15. Wang P, Xiao X, Glissen Brown JR, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng*. 2018;2(10):741-748.

16. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68(10):1813-1819.

17. Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol*. 2020;5(4):343-351.

18. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*. 2019;68(1):94-100.

19. Kudo S-E, Misawa M, Mori Y, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. *Clin Gastroenterol Hepatol*. 2020;18(8):1874-1881.e2.

20. Byrne MF. Hype or reality? Will artificial intelligence actually make us better at performing optical biopsy of colon polyps? *Gastroenterology*. 2020;158(8):2049-2051.

21. Jin EH, Lee D, Bae JH, et al. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. *Gastroenterology*. 2020;158(8):2169-2179.e8.

22. van der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy*. 2016;48(7):617-624.

23. Leenhardt R, Vasseur P, Li C, et al; CAD-CAP Database Working Group. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc*. 2019;89(1):189-194.

24. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc*. 2020;91(3):606-613.e2.

25. Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc*. 2019;89(2):357-363.e2.

26. Stidham RW, Liu W, Bishu S, et al. Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open*. 2019;2(5):e193963.

27. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology*. 2020;158(8):2150-2157.

28. Stidham RW, Enchakalody B, Waljee AK, et al. Assessing small bowel stricturing and morphology in Crohn's disease using semi-automated image analysis. *Inflamm Bowel Dis*. 2020;26(5):734-742.

29. Syed S, Al-Boni M, Khan MN, et al. Assessment of machine learning detection of environmental enteropathy and celiac disease in children. *JAMA Netw Open*. 2019;2(6):e195822.

30. Zhou J, Wu L, Wan X, et al. A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc*. 2020;91(2):428-435.e2.

31. Shahidi N, Rex DK, Kaltenbach T, Rastogi A, Ghalehjegh SH, Byrne MF. Use of endoscopic impression, artificial intelligence, and pathologist interpretation to resolve discrepancies between endoscopy and pathology analyses of diminutive colorectal polyps. *Gastroenterology*. 2020;158(3):783-785.e1.

32. Ebigbo A, Mendel R, Probst A, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut*. 2020;69(4):615-616.

33. Yeung C-K, Cheung JL, Sreedhar B. Emerging next-generation robotic colonoscopy systems towards painless colonoscopy. *J Dig Dis*. 2019;20(4):196-205.

34. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. *Nat Mater*. 2019;18(5):410-414.

35. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233-234.

36. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther*. 2018;47(6):763-772.

37. Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and validation of machine learning models in prediction of remission in patients with moderate to severe Crohn disease. *JAMA Netw Open*. 2019;2(5):e193721.

38. Waljee AK, Joyce JC, Wang S, et al. Algorithms outperform metabolite tests

in predicting response of patients with inflammatory bowel disease to thiopurines. *Clin Gastroenterol Hepatol*. 2010;8(2):143-150.

39. 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.

40. Konerman MA, Lu D, Zhang Y, et al. Assessing risk of fibrosis progression and liver-related clinical outcomes among patients with both early stage and advanced chronic hepatitis C. *PLoS One*. 2017;12(11):e0187344.

41. Khosravi B, Pourahmad S, Bahreini A, Nikeghbalian S, Mehrdad G. Five years survival of patients after liver transplantation and its effective factors by neural network and Cox poroportional hazard regression models. *Hepat Mon*. 2015;15(9):e25164.

42. Bibault J-E, Giraud P, Housset M, et al. Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep*. 2018;8(1):12611-12618.

43. Wong GL-H, Ma AJ, Deng H, et al. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. *Aliment Pharmacol Ther*. 2019;49(7):912-918.

44. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health*. 2019;1(5):e232-e242.

45. US Department of Health and Human Services; Food and Drug Administration; Center for Devices and Radiological Health. Software as a medical device (SAMD): clinical evaluation—guidance for industry and Food and Drug Administrative staff. https://www.fda.gov/media/100714/download. Published December 8, 2017. Accessed June 17, 2020.

46. Liu Y, Chen PC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. *JAMA*. 2019;322(18):1806-1816.