ADVANCES IN GERD

Current Developments in the Management of Acid-Related GI Disorders

Section Editor: Prateek Sharma, MD

Artificial Intelligence in Esophageal Diseases

Cadman L. Leggett, MD
Assistant Professor of Medicine
Department of Gastroenterology and Hepatology
Mayo Clinic
Rochester, Minnesota

G&H How may artificial intelligence be used to help improve clinical or endoscopic evaluation of esophageal diseases?

CL There are several aspects of how artificial intelligence (AI) is having and will continue to have an impact in the field of esophagology. Most algorithms to date have focused on enhancing endoscopic surveillance by improving detection of neoplastic lesions with computer-aided detection (CADe). For example, the Barrett's

> … the use of AI is not meant to substitute the cognitive skills employed during endoscopy; on the contrary, AI is meant to enhance these skills by providing a real-time second opinion.

Oesophagus Imaging for Artificial Intelligence consortium developed a CADe algorithm that significantly improved the sensitivity for neoplasia detection among endoscopists from 74% to 88% when tested in patients with Barrett esophagus. Similarly, a recent large multicenter tandem randomized controlled trial conducted in China demonstrated the benefit of a CADe system in detecting superficial esophageal squamous cell carcinoma. However, it is important to keep in mind that the performance of these algorithms relies on the quality of the endoscopic examination. There are quality algorithms capable of monitoring blind spots in real time during endoscopy. I feel that the combination of quality monitoring and CADe algorithms will yield the highest diagnostic performance.

The clinical care of patients with esophageal diseases relies on the histologic evaluation of biopsies and resection specimens. In this regard, the application of AI to digital endoscopy will play a key role in improving and standardizing clinical care. An example is a deep learning model developed by Mayo Clinic investigators that is capable of performing histologic evaluation of dysplasia in Barrett esophagus. The model, which evaluated digitized histology slides, showed an F1 score (measure of precision and recall) of greater than 80% each for the diagnosis of nondysplastic low-grade dysplasia and high-grade dysplasia. This result is particularly important in the effort to standardize a diagnosis of low-grade dysplasia, which currently has a high interobserver variability among pathologists.

G&H What types of tasks can AI perform, and can they enhance technical and cognitive skills and endoscopic quality?

CL AI can perform several of the tasks involved in diagnostic endoscopy. These include detection of a lesion, characterization of a lesion (benign vs malignant), and in cases of a neoplastic lesion, estimating the depth of invasion and performing delineation for resection. In this context, the use of AI is not meant to substitute the cognitive skills employed during endoscopy; on the contrary, AI is meant to enhance these skills by providing a real-time second opinion. Most endoscopy CADe systems use an alarm, such as a bounding box (a geometric shape surrounding one or more objects in an image), to alert the endoscopist to the presence of a lesion. A challenge with this approach is that it currently is not tailored to the endoscopist's level of expertise. For an expert, an alarm by an AI system might be seen as intrusive, ie, a cognitive burden that distracts instead of a tool that enhances detection. On the contrary, for a nonexpert, an alarm may represent an opportunity to detect a lesion that would have otherwise been missed. Researchers are beginning to learn how the use of AI changes endoscopist behavior during endoscopy and whether an AI algorithm can adapt to the user's level of experience.

Regarding endoscopic quality, the endoscopist performing an examination must always strive to perform the highest quality examination. It is important to remember that an AI image-processing algorithm relies on the quality of the images it analyses when providing a prediction. Endoscopists should be conscious of their skill set and ability to perform a good-quality examination so that they can obtain the best performance of the algorithm they are using.

G&H Can AI predict or detect early esophageal cancer? If so, how?

CL There are two important aspects of how AI can enhance our ability to detect early esophageal cancer. The first involves improving the ability to identify individuals at risk for early esophageal cancer. AI-powered risk prediction models analyze large amounts of data from the electronic medical record (EMR) system. These models rely on natural language processing, a branch of AI that can extract meaning from textual information. As an example, Mayo Clinic investigators used the Clinical Data Analytics Platform, a deidentified EMR database of 6 million patients, to develop a machine-learning natural language processing algorithm capable of predicting incident Barrett esophagus and esophageal cancer with an area under the curve of 0.84. This machine learning model appears to outperform conventional risk factor–based risk scores and has the advantage that it can be incorporated into the EMR to generate an automated

flag in the patient's file to prompt screening.

The second way in which AI can enhance the detection of early esophageal cancer is by improving user diagnostic performance during surveillance endoscopy. As previously mentioned, CADe systems have been developed and tested for early esophageal cancer in both Barrett esophagus and esophageal squamous cell carcinoma. What is lacking is pragmatic data on the real-world utilization of these algorithms; however, more data should become available in the near future once CADe systems receive clearance from the US Food and Drug Administration.

> The promise of AI is that it will have a direct positive impact on the clinical outcomes of patients while simultaneously relieving clinicians from the clerical burden of the EMR system.

AI is also being used to predict progression to cancer in patients with Barrett esophagus. The tissue systems pathology-9 test (TissueCypher, Castle Biosciences), which is commercially available, is performed on endoscopic biopsies and provides a risk score (low, intermediate, or high) for progression to high-grade dysplasia/esophageal adenocarcinoma in 5 years.

G&H How well does AI perform in the diagnosis of malignant and benign esophageal disease?

CL A wide array of AI algorithms have been developed for the diagnosis of both malignant and benign esophageal disorders using high-definition white light endoscopy in addition to advanced imaging modalities. The understanding among AI developers of how best to test the performance of these algorithms has evolved over time. For instance, one way to test the performance of an algorithm is to allocate part of the data for testing (ie, internal validation). Researchers are now aware of several types of biases that can be intrinsic to the data that were used to train these algorithms. For this reason, it is important to make sure that algorithms are vali-

dated with datasets that are independent of those used for training (ie, external validation). Researchers are also becoming keenly aware that stand-alone performance metrics are not necessarily representative of real-world performance. This is because there exists an interdependence between the AI algorithm and the user that can significantly impact behavior during endoscopy. When an AI algorithm provides a prediction, the user is required to agree or disagree with the prediction. In the current state, the AI algorithm is not capable of providing an explanation as to why it reached a prediction. As such, users are asked to *trust* the algorithm, but this trust should never be blind; it should always involve clinical judgment.

G&H What effect might AI have on clinician labor and care of patients with esophageal disease?

CL The promise of AI is that it will have a direct positive impact on the clinical outcomes of patients while simultaneously relieving clinicians from the clerical burden of the EMR system. To date, the application of AI in esophageal diseases has primarily focused on improving

> Current algorithms are fine-tuned to provide a prediction but lack the capability to explain their decision-making process, a conundrum sometimes referred to as the black box of AI.

clinical outcomes. I have provided examples of how AI is capable of identifying patients at risk of esophageal cancer and how CADe systems can improve detection of early neoplasia. I suspect that the next wave of AI will focus on how to optimize the overall management of patients with esophageal disease. With the advent of generative AI tools, it will not be too long before automated clinical and procedural documentation can be incorporated into daily clinical practice. AI tools will likely also be available to help optimize patient scheduling and procedural workflow.

G&H How is AI limited in the evaluation of esophageal diseases? What is the black box of AI?

CL Current algorithms are fine-tuned to provide a prediction but lack the capability to explain their decision-making process, a conundrum sometimes referred to as the black box of AI. Some users may see this as a limitation because they are unable to interrogate the AI or understand how the algorithm performs a prediction. In medicine, this concept naturally makes clinicians uncomfortable, partially because they often rely on a set of established criteria in the diagnostic process. I personally do not see the black box of AI as a limitation, as long as clinicians measure the prediction of the AI algorithm against their own prediction. For example, if during an endoscopy an AI algorithm provides a bounding box to alert of early neoplasia in a patient with Barrett esophagus, the endoscopist would evaluate this area for established features associated with early neoplasia such as vascular and/or mucosal irregularities (ie, Barrett's International Narrow-band Imaging Group classification). If such features exist, the endoscopist can confidently rely on the algorithm's prediction. If such features are lacking, then the endoscopist would need to decide whether or not to trust the algorithm's prediction. As I mentioned, the AI algorithm is essentially serving as a second pair of eyes, or a second opinion in real time. Techniques such as gradient-weighted class activation mapping have been used to superimpose a heat map over an image to highlight the areas of importance in making a prediction. Although having this information does not fully explain the algorithm's prediction, it may help the endoscopist in recognizing the image features associated with a given prediction.

G&H What steps are being taken to ensure responsible, safe implementation of AI algorithms?

CL The American Society for Gastrointestinal Endoscopy (ASGE) has established an AI Task Force with the primary mission to set clinical and research priorities for AI applications in gastrointestinal endoscopy. Regulation and implementation of AI into clinical practice are topics that have evolved over time and are discussed at the ASGE AI Task Force yearly Gastroenterology and Artificial Intelligence Summit. A summary of recommendations on the clinical use and implementation of AI can be found in a recent publication by the AI Task Force. Broadly, there are efforts across the AI community for responsible use of AI in health care. The Coalition for Health AI is leading some of these efforts and is

primarily focused on providing guidelines and standards that will drive high-quality AI systems in health care that are free of potential biases.

G&H What are current and future applications of AI for esophageal diseases?

CL Commercially available AI algorithms in the United States are primarily in the space of colorectal polyp detection. As more algorithms are approved by the US Food and Drug Administration, I look forward to seeing the impact that they will have in the field of esophagology, in particular for the detection of early esophageal cancer. Whether these algorithms will remain as stand-alone systems or integrated into established endoscopy video processors remains to be determined. As previously mentioned, the combination of quality and detection algorithms is a key step in improving implementation into clinical practice. Finally, as researchers begin to understand the impact of AI algorithms on human behavior, I envision that second-generation algorithms will have capabilities that adapt and learn from their users, improving trust while decreasing cognitive burden.

Disclosures
Dr Leggett has no relevant conflicts of interest to disclose.

Suggested Reading

Faghani S, Codipilly DC, Vogelsang D, et al. Development of a deep learning model for the histologic diagnosis of dysplasia in Barrett's esophagus. *Gastrointest Endosc.* 2022;96(6):918-925.e3.

Fockens KN, Jong MR, Jukema JB, et al; Barrett's Oesophagus Imaging for Artificial Intelligence (BONS-AI) consortium. A deep learning system for detection of early Barrett's neoplasia: a model development and validation study. *Lancet Digit Health.* 2023;5(12):e905-e916.

Iyer PG, Sachdeva K, Leggett CL, et al. Development of electronic health record-based machine learning models to predict Barrett's esophagus and esophageal adenocarcinoma risk. *Clin Transl Gastroenterol.* 2023;14(10):e00637.

Kahn A, Leggett CL. Artificial intelligence in the age of cognitive endoscopy. *Gastrointest Endosc.* 2020;91(6):1251-1252.

Leggett CL. Endoscopic screening for oesophageal cancer: empowering artificial intelligence with a high-quality examination. *Lancet Gastroenterol Hepatol.* 2024;9(1):4-5.

Parasa S, Repici A, Berzin T, Leggett C, Gross SA, Sharma P. Framework and metrics for the clinical use and implementation of artificial intelligence algorithms into endoscopy practice: recommendations from the American Society for Gastrointestinal Endoscopy Artificial Intelligence Task Force. *Gastrointest Endosc.* 2023;97(5):815-824.e1.

Sharma P, Hassan C. Artificial intelligence and deep learning for upper gastrointestinal neoplasia. *Gastroenterology.* 2022;162(4):1056-1066.

Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut.* 2019;68(12):2161-2169.

Yuan XL, Liu W, Lin YX, et al. Effect of an artificial intelligence-assisted system on endoscopic diagnosis of superficial oesophageal squamous cell carcinoma and precancerous lesions: a multicentre, tandem, double-blind, randomised controlled trial. *Lancet Gastroenterol Hepatol.* 2024;9(1):34-44.