EDITORIAL

Artificial intelligence: finding the intersection of predictive modeling and clinical utility

Artificial intelligence (AI) refers to the ability of computers to perform tasks normally reserved for human intelligence. AI is a broad concept and encompasses machine learning in which computers use data to create a binary predictive algorithm: deep learning that enhances machine learning by creating algorithms that select and weigh different variables to best predict an outcome. Convoluted neural networks are also created with interconnected "neurons" for pattern recognition, thereby weighing predictive features and learning from the data to predict outcomes. One strength of AI is its ability to process data accurately and in ways not grasped by the human mind. This is particularly evident when applied to computer systems to process images and videos to gain novel information. For example, AI in endoscopy has been investigated largely to improve the detection of malignant and premalignant lesions of the GI tract.[1,2]

Benign lesions of the GI tract are often encountered in clinical endoscopic practice. Indeed, significant amounts of resources are spent in surveillance and management of benign lesions that cannot be reliably differentiated by malignant potential. Esophageal lesions perhaps serve as a clear example of this conundrum. Overall, benign esophageal lesions are fairly rare, with a reported prevalence of approximately 1% in autopsy series.[3] Unlike mucosal lesions, which can be diagnosed with endoscopic biopsy during standard white-light EGD, benign esophageal neoplasms are commonly intramural. Esophageal leiomyomas represent two-thirds of all benign esophageal lesions. Esophageal leiomyomas are largely indolent, asymptomatic, and, when <2 cm, carry almost no risk of malignant transformation.[4] Consequently, the American Society for Gastrointestinal Endoscopy guidelines do not recommend resection of small, asymptomatic esophageal leiomyomas.[5] However, the diagnosis of leiomyomas can be challenging on the basis of EUS features, which may overlap with rarer esophageal tumors including leiomyosarcomas and GI stromal tumors (GISTs), which carry greater malignant potential. Immunostaining with CD117, DOG1, smooth muscle actin, and desmin can distinguish these lesions, but up to one-third of EUS-guided biopsies will not provide adequate tissue for immunohistochemical staining, leading to diagnostic uncertainty.[6] Consequently, these same guidelines recommend EUS surveillance in such patients with esophageal leiomyomas.

In this issue of *Gastrointestinal Endoscopy*, Zhang et al[7] present a retrospective single-institution study using 3-deep convolution neural network (CNN) models in the attempt to supplement this diagnostic gap for benign esophageal lesions. A total of 1217 patients who underwent white-light EGD and EUS at a single institution between January 2015 and April 2020 were identified. Three different deep CNN models were created. In the first model, the goal was to discriminate benign esophageal lesions from normal EGD findings. This consisted of 17,279 white-light EGD images from 200 patients with normal findings (5536 images), 200 with esophagitis (6280 images), and 198 patients with protruding esophageal lesions consisting of esophageal papillomas, cysts, and leiomyomas (5463 images). Model 2 sought to distinguish subtypes of protruding esophageal lesions on white-light EGD and included 3226 white-light EGD images from 161 patients with esophageal papillomas (829 images), 119 patients with esophageal cysts (814 images), and 163 patients with esophageal leiomyomas (1583 images). Finally, model 3 sought to discriminate esophageal leiomyoma from esophageal cysts on EUS and included 3411 EUS images from 85 patients with esophageal cysts (1197 images) and 163 patients with esophageal leiomyomas (2214 images). In addition, 2 expert endoscopists, 2 senior endoscopists, and 2 novice endoscopists reviewed all images in models 2 and 3 in a blinded fashion and classified each lesion. In model 1, the area under the curve (AUC) was fair, at 0.751 (95% CI, 0.65-0.85) for identifying an esophageal

> As physicians we spend considerable effort to mitigate risk. As a result, what predictive accuracy would a deep CNN model need to achieve in order to substantially change the management and diagnostic evaluation of a potentially malignant lesion in a very small patient population?

0016-5107/$36.00
https://doi.org/10.1016/j.gie.2020.12.008

protruding lesion. In model 2, the CNN model performed well, with an AUC of 0.907 (95% CI, 0.835-0.979) and a predictive accuracy of 87.6% for esophageal papilloma; AUC of 0.897 (95% CI, 0.84-0.953) with a predictive accuracy of 86.7% for esophageal leiomyoma; and AUC of 0.868 (95% CI, 0.769-0.968) with a predictive accuracy of 89.26% for an esophageal cyst. This CNN model outperformed endoscopist readers for leiomyoma and cysts but was inferior to the endoscopists' interpretation for esophageal papillomas. In the third model, the CNN model performance was fair, with an AUC of 0.739 (95% CI, 0.6-0.878) for leiomyoma and AUC of 0.724 (95% CI, 0.567-0.881) for esophageal cysts on EUS images. The model had a higher sensitivity and specificity for these lesions than did the endoscopists, but notably, accuracy was fairly comparable between the model and the endoscopist readers for esophageal leiomyoma (predictive accuracy of 79.6% for CNN model vs 63.3% to 81.6% for endoscopists) and esophageal cysts (predictive accuracy of 79.6% for CNN model vs 63.2% to 81.6% for endoscopists). Subsequent analysis found that the combination of the CNN model with the endoscopists' interpretation of images resulted in an improved predictive accuracy, with an increase of sensitivity and specificity of 5.8% and 9.3%, respectively, for esophageal leiomyoma on white-light EGD. Similarly, the combination of the CNN model with endoscopists' interpretations significantly improved diagnostic performance, with an increase in predictive accuracy from 80.2% to 90.9%. The authors conclude that these preliminary CNN models highlight the potential ability of AI to assist endoscopists during EGD to reliably differentiate and identify benign esophageal protruding lesions.

Although intriguing, this study also highlights potential limitations of the application of AI in clinical practice. The relevance of differentiating esophageal protruding lesions is the exclusion or prediction of developing malignancy. In the vast majority of cases, this diagnostic uncertainty is most germane to submucosal esophageal lesions, specifically differentiating esophageal leiomyomas from GISTs. This assessment is dependent on several factors, including patient demographics, prevalence, endoscopic features, and immunohistochemical staining. In the majority of cases, current tools are reliable to rule out a GIST or a leiomyosarcoma. In a study of 84 patients with leiomyoma, 13 GISTs, 5 leiomyosarcomas, and 4 granular cell tumors, EUS with FNA had a negative predictive value of 100% for leiomyosacroma and GIST. However, the same study found that EUS with FNA had a modest positive predictive value for leiomyoma of 68%.[8] Consequently, an inconclusive diagnostic evaluation or an erroneous diagnosis of a leiomyosarcoma or GIST is possible in the presence of a benign leiomyoma. In such cases, other supporting features such as lesion size, the rarity of esophageal GISTs and leiomyosarcomas, and patient age are used to council patients regarding the next steps in management. However, in the study by Zhang et al,[7] only small benign esophageal lesions with a mean size of 0.87 ± 0.68 cm for esophageal leiomyomas were considered. Consequently, although the developed deep CNN model showed additive benefit in differentiating small and known benign esophageal lesions, its performance in differentiating such lesions from larger potentially malignant lesions is unclear. Further, if and to what degree it provides additive benefit in comparison with other clinical factors is unknown. Additionally, the development of a deep CNN model from a dataset created at a single institution is subject to undetected inherent bias, which may limit its generalizability. This point is of particular importance when the results from a deep CNN model are considered, in which the logic and rationale used is unclear. Furthermore, the clinical impact of an effective deep CNN model in the assessment of protruding esophageal lesions is itself uncertain. As physicians we spend considerable effort to mitigate risk. As a result, what predictive accuracy would a deep CNN model need to achieve in order to substantially change the management and diagnostic evaluation of a potentially malignant lesion in a very small patient population?

On the other hand, the model on which this study is based is a good one. It used large numbers of images required to perform such a study. Furthermore, it emphasized partnership rather than exclusivity of AI over the human eye; experienced operator visualization in conjunction with other patient factors will always remain an essential part of endoscopic recognition, as outlined by these authors. A future state of AI may very well determine what the computer is seeing that endoscopists are not, allowing improved personal endoscopy skills and making the teaching in AI a "2-way street."

In conclusion, the results of this well-performed study using deep CNN models represent promise in the ability of AI to reliably differentiate benign esophageal lesions and obviate the need for further diagnostic evaluation or surveillance. Furthermore, the application of deep CNN models in concert with physicians' image interpretation during real-time endoscopy is an exciting advancement to improve the quality of patient care. However, careful development and application of AI models in a clinically relevant manner is paramount as the role of and application of AI in the field of gastroenterology continues to expand.

DISCLOSURE

The author disclosed no financial relationships.

Karthik Ravi, MD
Department of Medicine
Mayo Clinic College of Medicine
Rochester, Minnesota, USA

Abbreviations: AI, artificial intelligence; CNN, convolution neural network; GIST, GI stromal tumor.

REFERENCES

1. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology: promises and pitfalls. Am J Gastroenterol 2019;114:4228.
2. La Berre C, Sandborn WJ, Aridhi S, et al. Application of artificial intelligence to gastroenterology and hepatology. Gastroenterology 2020;158:76-94.
3. Ha C, Regan J, Bulent Cetindag I, et al. Benign esophageal tumors. Surg Clin N Am 2015;95:491-514.
4. Lee LS, Singhal S, Brinster CJ, et al. Current management of esophageal leiomyoma. J Am Coll Surg 2004;198:136-46.
5. Faulx AL, Kathari S, Acosta RD, et al. The role of endoscopy in subepithelial lesions of the GI tract. Gastrointest Endosc 2017;85:1117-32.
6. Hoda KM, Rodriguez SA, Faigel DO. EUS-guided sampling of suspected GI stromal tumors. Gastrointest Endosc 2009;69:1218-23.
7. Zhang M, Zhu C, Wang Y, et al. Differential diagnosis for esophageal protruded lesions using a deep convolution neural network in endoscopic images. Gastrointest Endosc 2021;93:1261-72.
8. Codipilly DC, Fang H, Alexander JA, et al. Subepithelial esophageal tumors: a single-center review of resected and surveilled leisons. Gastrointest Endosc 2018;87:370-7.