



EndoBRAIN-EYE and the SUN database: important steps forward for computer-aided polyp detection

Colonoscopy is a durable cancer screening and prevention strategy in the United States and worldwide. Over the past several years, there has been increased attention toward the development and study of artificial intelligence (AI)-based computer-aided detection (CADe) systems for colonoscopy to augment polyp detection by the endoscopist during screening and surveillance colonoscopy.

Over this period of time, there have been a wide variety of publications focused on training and validation of various computer vision systems intended to detect polyps in the colon. Early work used traditional machine-learning techniques with explicit feature extraction methods, validated on still image frames from colonoscopic video.^{1,2} The advent of deep learning—a subset of machine learning that involves the extraction of many feature layers and that harnesses neural networks that have been compared with the human nervous system to produce complicated predictive outputs—led to significant improvements in the diagnostic performance of CADe technology. At least 4 previous studies have described the development and validation of CADe systems for polyps in the colon based on deep learning.³⁻⁶ These systems showed a per-frame sensitivity of 90.0% to 97.3% and per-frame specificity of 63.3% to 95.4%, with a high area under the receiver operating characteristic curve. There have since been several trials examining these systems prospectively during live colonoscopy.^{7,8}

In this issue of *Gastrointestinal Endoscopy*, Misawa et al⁹ report training and validation data for a new CADe system. The authors developed the system using image data collected at 5 endoscopy centers across Japan. A total of 56,668 frames were extracted from 3106 colonoscopy cases and divided into training (51,899 images) and validation/tuning (4769 images) data. Data were annotated by 3 research assistants and 2 expert endoscopists. The algorithm was developed using YoloV3, a freely available and widely used object detection algorithm that consists of a deep convolutional neural network, which can be trained to detect any category of image.⁹ The resultant AI algorithm (EndoBRAIN-EYE) had a per-frame sensitivity of 90.5% (95% confidence interval [CI], 90.2%-90.7%), per-

frame specificity of 93.7% (95% CI, 93.5%-93.8%) and a per-polyp sensitivity of 98% (95% CI, 93%-99.8%). The authors also reported per-frame and per-polyp measures on diminutive, protruding, and flat lesions.

As the authors point out, one obvious strength of the current study is that they were able to achieve modest improvements in performance with the EndoBRAIN-EYE compared with other similar deep learning algorithms. This was likely due to several factors, including an increase in the number of images in the training dataset (56,668 compared with 5545-8641 images in previous studies),

The pace of study of computer vision applied to endoscopy in general and colonoscopy in particular is blistering, and as more data are published on the development, validation, and prospective testing of computer-aided detection systems, it will become increasingly important to compare these systems directly.

the use of YoloV3, the use of an advanced object adaptive learning rate optimization algorithm (Adam), and the use of still images from colonoscopic videos, which may provide a more accurate representation of images/videos than still images captured by the endoscopist. To this last point, as the authors also describe, the use of stills captured by the physician may lead to a quality bias if used as the ground truth for algorithm development and cause the algorithm to underperform when tested on external data. The authors used many best practices in machine learning to develop and validate their new algorithm, including the use of an advanced object adaptive learning rate optimization algorithm, the use of still images of videos for their ground truth, and the use of a large database of outside images for their test set for external validation.

What is perhaps the most exciting aspect of this study, however, especially to clinicians who are less focused on the granular details of machine learning and AI algorithm development, is the creation of a public database of colonoscopic videos, which is available to use by request (<http://sundatabase.org>).

The Showa University and Nagoya University (SUN) database, contains 49,799 polyp frames (representing 100 randomly selected polyps) annotated with bounding boxes and 102,761 frames without polyps. Videos were recorded using high-definition colonoscopes. Images were annotated by 3 research assistants and then underwent several additional steps of expert adjudication for quality assurance.

The resultant database contains a diverse spread of polyp types (7 hyperplastic polyps, 4 sessile serrate lesions, 82 low-grade adenomas, 2 traditional serrated adenomas, 4 high-grade adenomas, and 1 invasive cancer). This database was created for the dual purposes of testing the EndoBrain-EYE CAde system and serving as a benchmark dataset for the direct comparison of CAde algorithms. Regarding the distribution of lesion-positive images in the database, although this is not the exact prevalence of these lesions in clinical practice, having a diverse variety of lesions emulates the real-world clinical setting, and most clinically important lesions are represented in the SUN database. One slight limitation in the resultant data is the paucity of flat lesions, such as sessile serrated adenomas and traditional serrated adenomas, which are a particularly important target for CAde systems given their subtle surface features and potential to be missed by the human eye.

The pace of study of computer vision applied to endoscopy in general and colonoscopy in particular is blistering, and as more data are published on the development, validation, and prospective testing of CAde systems, it will become increasingly important to compare these systems directly. As can be seen from the current study, the way that algorithms are developed, from the input data and ground truth, to data augmentation, to steps taken to avoid overfitting and optimistic estimates of performance, matters. As more algorithms reach the level of prospective study, premarket submission, and regulatory approval, we will need methods of direct comparison for quality assurance. One conventional method would be to compare CAde systems head to head in multi-arm, parallel, randomized clinical trials. However, given that most systems are built in a similar fashion, with somewhat similar input data and deep learning methods, and, more importantly, that each system has the potential for iterative improvement, such a method of comparison could quickly become impractical. An alternative method of comparison, as Misawa et al⁹ have laid out in this article, is the direct comparison of CAde systems using a standardized database of images and video. The SUN database is an incredibly important step toward this goal.

Although it does not appear to be the intention of the authors, creating publicly available data is also crucial to the development of future deep learning systems. Most CAde systems are built using private data sources,

harvested and annotated from a single hospital system or network. As the AI revolution in clinical medicine and in GI endoscopy continues, it will be incredibly important to have high-quality input data to build new systems and improve existing ones. Several groups are already laying the groundwork for this from colonoscopic video to video capsule endoscopy.¹⁰

In summary, Dr Misawa and colleagues⁹ have published a novel study in this issue of *Gastrointestinal Endoscopy*, summarizing the development and validation data of a new CAde system, the EndoBRAIN-EYE, designed to detect polyps in the colon during colonoscopy. They have also created an annotated database of images, the SUN database, designed to serve as a benchmark dataset for the direct comparison of CAde systems going forward. It is clear that the new CAde system described here, along with a benchmark dataset for future studies, represent equally important contributions to the AI revolution in GI endoscopy.

DISCLOSURE

Dr Berzin has served as a consultant for Wision AI, Fujifilm, and Medtronic. Dr Glissen Brown disclosed no financial relationships.

Jeremy R. Glissen Brown, MD

Tyler M. Berzin, MD, MS

Center for Advanced Endoscopy

Division of Gastroenterology

Beth Israel Deaconess Medical Center and Harvard Medical

School

Boston, Massachusetts, USA

Reprint requests: Jeremy R. Glissen Brown, MD, Center for Advanced Endoscopy, Division of Gastroenterology and Hepatology, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Boston, MA 02130.

Abbreviations: AI, artificial intelligence; CAde, computer-aided detection; CI, confidence interval.

REFERENCES

1. Fernandez-Esparrach G, Bernal J, Lopez-Ceron M, et al. Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. *Endoscopy* 2016;48:837-42.
2. Wang Y, Tavanapong W, Wong J, et al. Polyp-Alert: near real-time feedback during colonoscopy. *Comput Methods Programs Biomed* 2015;120:164-79.
3. 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;2:741-8.
4. 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;155:1069-78.e1068.

5. Yamada M, Saito Y, Imaoka H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep* 2019;9:14465.
6. Misawa M, Kudo SE, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018;154:2027-9.e2023.
7. Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020;159:512-20.e7.
8. 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;68:1813-9.
9. Misawa M, Kudo SE, Mori Y, et al. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). *Gastrointest Endosc* 2021;93:960-7.
10. Jha D, Smedsrud PH, Riegler MA, et al. Kvasir-SEG: a segmented polyp dataset. *arXiv* 2019;1911.07069.

GIE on LinkedIn

Find *GIE* on LinkedIn. Followers will receive news updates and links to author interviews, podcasts, articles, and tables of contents. Use the QR code or search on LinkedIn for “GIE: Gastrointestinal Endoscopy with Editor Michael B. Wallace” and follow us today.

