Guns, germs, and steel…and artificial intelligence

When we think back on the most significant turning points that have affected humanity, most will agree with the 1998 Pulitzer Prize–winning author of *Guns, Germs, and Steel* that the invention of guns, the discovery of microorganisms, and the invention of steel are among the key turning points.[1] For recent years we might add the invention of electricity and the birth of the internet to this list. In years to come, however, we will look back and likely state that the introduction of artificial intelligence (AI) into our society was another big turning point for humanity. Here and now, AI permeates all aspects of life, including transportation, construction, business, military, healthcare, social media, and even art and music.

In 2020, we are embarking on an unchartered journey of medicine integrating with AI—and how exciting it is. AI algorithms are used to help pathologists diagnose cancer with less error, aid patients with monitoring symptoms, and even assist administrators to predict clinical, financial, and operational risk.[2] What makes gastroenterology especially primed for the implementation of AI is our most coveted tool, endoscopy. Crudely speaking, endoscopy is nothing more than a silent stream of still images shot by a fast-clicking camera captured by a highly trained cameraperson. These still frames are ripe for a specific brand of AI called computer vision, a branch of computer science that leans on AI for augmenting our perception and interpretation of things we perceive visually.

How can AI help us endoscopists? I see endoscopy as the sum of 3 main skills: (1) understanding the biologic, sociodemographic, and clinical factors surrounding the case; (2) technical expertise in manipulation of the endoscope and its adjunctive instruments; and (3) real-time interpretation of the visual matter displayed on the screen. The interpretation of visual matter on the screen is a challenging skill to develop and is an area that often falls short during training programs. Examination of Barrett's esophagus is a prime example. Despite high-definition endoscopes in the hands of gastroenterologists, the sensitivity for detecting neoplastic lesions in this disease continues to be suboptimal. Why is this? Perhaps it is because our training programs lack ample case experience with neoplastic lesions, and hence the opportunity to learn the nuances of detailed inspection are limited. Another may be, as Struyvenberg and colleagues[3] point out in their article, some of the visual classification systems are too vague, whereas others are overly complicated. Whatever the reason, there remains significant room for improvement, and the argument to use AI to augment an endoscopist's image interpretation skills (or lack thereof) is certainly valid.

In the high-quality examination of Barrett's esophagus, there should be 2 discrete phases. The first is a "red flagging" phase whose goal is to identify areas of suspicion for neoplasia. The second is a "characterization" phase in which these areas of suspicion are interrogated and a mental calculation of neoplasia probability is made. Harnessing the power of AI through computer vision can help with both phases of the Barrett's examination procedure. In fact, several groups, including our own group, have focused AI projects around the red flagging phase of the Barrett's examination often referred to as computer-aided detection (CADe).[4,5] In this article, Struyvenberg et al train and test an algorithm that specializes in the characterization phase of the examination, increasingly being referred to as computeraided diagnosis (CADx).[3] Characterization is often performed with the use of virtual chromoendoscopy, and in this study narrow-band imaging (NBI) with zoom magnification was used. The CADx system demonstrated an accuracy, sensitivity, and specificity for detection of BE neoplasia using NBI-zoom images of 84%, 88%, and 78%, respectively. When video frames were analyzed by the CADx system, accuracy, sensitivity, and specificity were 83%, 85%, and 83%, respectively. The authors admit that these results are fairly good but are not ready for real-time use yet. They nicely include a clear course for the future of how their algorithm will be improved.

> Ideally, a CADe or CADx algorithm would be endoscope agnostic, meaning that the algorithm performs optimally across multiple endoscope manufacturers. In reality, this is difficult to achieve.

Copyright © 2021 by the American Society for Gastrointestinal Endoscopy
0016-5107/$36.00
https://doi.org/10.1016/j.gie.2020.07.021

I believe that several pitfalls and weaknesses should be highlighted when we read and evaluate studies on AI systems that perform these types of image interpretation tasks. In this regard, Struyvenberg and colleagues should be congratulated for their attention to these issues and clever execution to avoid and address them.

Pitfall No. 1: Overfitting and data leakage

Overfitting occurs when a machine learning model has become too attuned to the data on which it was trained and therefore loses its applicability to another dataset. When overfitting occurs, prediction accuracy is falsely elevated and overly optimistic. A common mistake for investigators is not carefully separating the images in the training set from images in the test set. The authors went a step further, in this study, where data from specific patients included in the training set pool were not included in the test set pool, preventing an error called data leakage.

Pitfall No. 2: Lack of diversity in the dataset

The authors trained the algorithm based on an extensive dataset of images of >800 neoplastic images and 600 nonneoplastic images in addition to still frames obtained from videos. Often more important than the number of images is the diversity of the image set. The authors notably point out that each image in their training set was quite unique to the other. For characterization algorithms, a high specificity is key, given that potentially risky therapeutic interventions may be carried out if a lesion is deemed neoplastic. The ideal therefore would be to have a training dataset that encompasses all lesions and all possible imaging features that one would encounter in clinical practice. Obviously, this is impractical to achieve, but the greater the diversity in the training set, the stronger the algorithm likely will be. How many more images will be needed to achieve optimum results is to be determined, but the authors point out that in the future their training set will expand.

Pitfall No. 3: Prediction speeds too slow for real time

As mentioned earlier, video is a stream of still images played in rapid sequence. Thirty frames per second is typical for native endoscopy. For an algorithm to be most helpful to the modern endoscopist, the prediction speed ideally would match or better this frame rate. This is most important for a CADe algorithm, where an endoscopist would like to see a suggestive lesion flagged instantly without pausing or freezing the image. The authors are sure to report that this new CADx algorithm was able to generate predictions at a very fast 38 frames per second.

Pitfall No. 4: Cross-compatibility and external validation

Ideally, a CADe or CADx algorithm would be endoscope agnostic, meaning that the algorithm performs optimally across multiple endoscope manufacturers. In reality, this is difficult to achieve. Most institutions use a single endoscope manufacturer, and hence the images populating the database are all from the same brand of endoscope. Each brand has slightly different imaging attributes. As a result, a high accuracy result of an algorithm trained on a single brand may not necessarily translate to another. Another important factor is the nature of the algorithm, inasmuch as some algorithms seem to be more cross compatible than others. Colon polyp detection, for example, appears to be an "easier" algorithm for cross compatibility because the visual differences between a polyp and normal colonic mucosa are quite distinct. The characterization of Barrett's neoplasia is inclined to be a more difficult algorithm for cross compatibility because the visual differences between neoplastic and nonneoplastic tissue is more subtle. To make a call on something more visually challenging like Barrett's neoplasia, the algorithm will understandably prefer to work with material with attributes most similar to its training material.

External validation is a hugely important aspect of this research, which means testing the algorithm on data that are completely independent of the training set and possibly even independent of the research group or country. This ensures that the algorithm can be universally applied to patients from outside the institution. The authors acknowledge that an external validation phase was not performed on this algorithm but will be the focus of future study.

In closing, the ongoing work of Struyvenberg and colleagues[3] is a great step toward a real-time CADx system for Barrett's esophagus, and the authors should be commended for their high-quality and responsible approach to achieve this goal. Algorithms such as this can and will continuously undergo iterative enhancements. As has been demonstrated, real-time processing speeds are now possible with off-the-shelf computing hardware. There persists a groundswell of excitement from researchers, industry, government and regulatory firms to implement AI technology into medical devices. I assure you as technology such as this is refined, CADe and CADx algorithm integration into the endoscope is inevitable. Artificial intelligence is here and it is changing our world…even endoscopy! Let's embrace it and do it right.

DISCLOSURE

Dr Samarasena has ownership in Docbot and is a consultant and lecturer for for Medtronic, Olympus, *and Conmed; the recipient of an educational grant from Cook Medical; a consultant for Neptune Medical, Steris, and Microtech; and a lecturer for Mauna Kea Technologies.*

Jason B. Samarasena, MD
H. H. Chao Comprehensive Digestive Disease Center
Division of Gastroenterology and Hepatology
University of California Irvine
Orange, California, USA

Abbreviations: AI, artificial intelligence; CADe, computer-assisted detection; CADx, computer-assisted diagnosis; NBI, narrow-band imaging.

REFERENCES

1. Diamond JM. *Guns, Germs, And Steel: The Fates of Human Societies*. New York: W.W. Norton & Co; 1998.
2. Daley S. Surgical robots, new medicines and better care: 32 examples of AI in healthcare 2020. Available at: https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare. Accessed June 15, 2020.
3. Struyvenberg MR, de Groof AJ, van der Putten J, et al. A computer-assisted algorithm for narrow-band-imaging-based tissue characterization in Barrett's esophagus. Gastrointest Endosc 2021;93:89-98.
4. Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). Gastrointest Endosc 2020;91:1264-71 e1.
5. de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. Gastroenterology 2020;158:915-29 e4.

Online Audio and Podcasting

Audio and Podcasts of article abstracts published in *Gastrointestinal Endoscopy* are now available online. Recordings are performed by Deborah Bowman, MFA, ELS, Senior Managing Editor of Clinical Publications.

Scan the QR code or log on to **www.giejournal.org** to listen to recordings from the current issue.