Deep learning for automated bowel preparation assessment during colonoscopy: time to embrace a new approach?

Published Online
September 16, 2021
https://doi.org/10.1016/S2589-7500(21)00143-6

See Articles page e697

Some of the most translationally mature artificial intelligence (AI) applications in health care belong to colonoscopy, where algorithms can now be used in clinical practice to assist colorectal polyp detection and characterisation.[1] The rapid pace of translation reflects the desperate need to identify solutions to drive up quality in colonoscopy and overcome the operator dependence associated with variable colorectal cancer protection. It is not surprising that AI developers are expanding research to further use cases applied to colonoscopy, particularly those that relate to performance measures and quality metrics.

Suboptimal bowel preparation is a major barrier to effective colonoscopy as it is associated with missed polyps, incomplete procedures, unsatisfactory patient experience, shorter surveillance intervals, and increased health-care costs.[2] Endoscopic professional societies have defined minimum standards for the adequacy of bowel preparation.[3] There are multiple scoring systems available, including the Boston bowel preparation scale (BBPS), which is the most thoroughly validated scale.[4] This system has some potential limitations in real-world clinical practice, such as the requirement of segmental scoring of the colon, possible variable reporting or even omission from reports, and many endoscopists score based on recall following completion of the procedure which could lead to inaccuracies. The concept of an automated, more objective, and reliable scoring process could address these limitations.

In their study,[5] Wei Zhou and colleagues developed an automated deep learning-based bowel preparation scoring system (automatic BBPS [e-BBPS]), which consisted of two deep convolutional neural network models, one to filter out unqualified colonoscopy image or video frames and a second one to classify frames according to BBPS categories. The e-BBPS score was then calculated using proportions of frames classified as BPPS scores 0–1 (suboptimal bowel preparation). The two deep convolutional neural network (DCNN) models were trained using 43 001 images for DCCN1 and 24 410 images for DCNN2. Five endoscopists provided reference annotations following a BBPS training module. After performing initial internal and external validation using retrospective data including both still images and video frames, the model was evaluated in a prospective observational study using consecutive video recordings from 616 patients undergoing screening colonoscopies. The e-BPPS score was significantly inversely correlated with the adenoma detection rate (ADR; Spearman's rank –0·976, p<0·010). Furthermore, based on the 25% ADR standard for screening colonoscopy, a threshold score of 3 was determined for e-BPPS that could guarantee an ADR of more than 25% and be used to define adequate bowel preparation in practice.

The results of Zhou and colleagues suggest that the concept of automated AI based bowel preparation assessment has a promising future.[5] However, there are some important limitations that should be considered before clinical implementation can occur. The prospective component of the study was limited to a single-centre; the system should be externally validated in larger multicentre studies to show generalisability, with robust comparisons to standard scoring methods. The study by Zhou and colleagues mandated that endoscopists performed thorough washing and suctioning during insertion before the model scored the withdrawal. In real-world clinical settings, it is not uncommon for endoscopists to wash during the withdrawal, and, therefore, this could pose as a challenge for the algorithm. Moreover, it is unclear how the algorithm would be best integrated into the existing clinical pathway. The model could output the score for reporting purposes at the end of the procedure to replace or accompany existing traditional endoscopist-derived scoring methods. It is not inconceivable that future AI models will also output other important procedure metrics by producing auto-generated endoscopy reports, to be verified by endoscopists, providing the precious gift of time to physicians.[6] However, by limiting the algorithm output to the post-procedure phase, the potential for feedback and improvement of bowel cleansing by the endoscopist that could be achieved with a real-time display is arguably diminished.

Overall, the endoscopic community needs to prepare for a dramatic increase of AI applications that are being developed for use cases across the spectrum of colonoscopy practice. It is even possible that novel computer vision metrics will replace and indeed overcome limitations of existing human-based quality performance measures. For instance, deep learning-based depth estimation has been used to accurately detect deficient colonoscopy coverage during withdrawals.[7] At the moment, a crude measure of withdrawal time is used as a surrogate for inspection quality. However, it is important that we remain cognisant of the potential pitfalls and limitations of AI-based software and solutions for colonoscopy by developing a robust framework for prospective evaluation of models within the intended clinical pathway, accounting for real-world endoscopist–AI interaction, and reporting endpoints that ultimately reflect patient outcomes.[8,9]

I declare no competing interests.

Omer F Ahmad
o.ahmad@ucl.ac.uk

Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, W1W 7TS, UK

1 Mori Y, Neumann H, Misawa M, Kudo SE, Bretthauer M. Artificial intelligence in colonoscopy—now on the market. What's next? *J Gastroenterol Hepatol* 2021; **36:** 7–11.

2 Hassan C, East J, Radaelli F, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline—update 2019. *Endoscopy* 2019; **51:** 775–94.

3 Kaminski MF, Thomas-Gibson S, Bugajski M, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. *Endoscopy* 2017; **49:** 378–97.

4 Parmar R, Martel M, Rostom A, Barkun AN. Validated scales for colon cleansing: a systematic review. *Am J Gastroenterol* 2016; **111:** 197–204.

5 Zhou W, Yao L, Wu H, et al. Multi-step validation of a deep learning-based system for the quantification of bowel preparation: a prospective, observational study. *Lancet Digit Health* 2021; published online Sept 16. https://doi.org/10.1016/S2589-7500(21)00109-6.

6 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25:** 44–56.

7 Freedman D, Blau Y, Katzir L, et al. Detecting deficient coverage in colonoscopies. *IEEE Trans Med Imaging* 2020; **39:** 3451–62.

8 Ahmad OF, Mori Y, Misawa M, et al. Establishing key research questions for the implementation of artificial intelligence in colonoscopy: a modified Delphi method. *Endoscopy* 2020; published online Nov 9. https://doi.org/10.1055/a-1306-7590.

9 Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019; **25:** 1337–40.