



Artificial intelligence for colorectal polyp detection: are we ready for prime time?

Omer F. Ahmad¹, Laurence B. Lovat^{1,2}

¹Wellcome/EPSCRC Centre for Interventional & Surgical Sciences, ²Division of Surgery & Interventional Science, University College London, London, UK

Contributions: (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Omer F. Ahmad. Wellcome/EPSCRC Centre for Interventional & Surgical Sciences, University College London, London, W1W 7TS, UK. Email: o.ahmad@doctors.org.uk.

Abstract: Colorectal cancer (CRC) is a leading cause of cancer-related mortality worldwide. Colonoscopy is protective against CRC through the detection and removal of neoplastic polyps. Unfortunately, the procedure is highly operator dependent with significant miss rates for polyps. Artificial intelligence (AI) and computer-aided detection software offers a promising solution by providing real-time assistance to highlight lesions that may otherwise be overlooked. Rapid advances have occurred in the field with recent prospective clinical trials demonstrating an improved adenoma detection rate (ADR) with AI assistance. Deployment in routine clinical practice is possible in the near future although further robust clinical trials are necessary and important practical challenges relating to real-world implementation must be addressed.

Keywords: Colonoscopy; artificial intelligence (AI); machine learning; computer-aided detection; endoscopy

Received: 19 August 2019; Accepted: 30 August 2019; Published: 17 September 2019.

doi: 10.21037/jmai.2019.09.02

View this article at: <http://dx.doi.org/10.21037/jmai.2019.09.02>

Introduction

Colorectal cancer (CRC) is a leading cause of cancer-related mortality worldwide, and its burden is expected to increase significantly over the coming decade (1). Colonoscopy is effective at preventing CRC through the removal of neoplastic polyps (e.g., adenomas) (2). The adenoma detection rate (ADR) during colonoscopy is independently associated with the incidence of interval CRC (defined as cancer diagnosed between the screening and post-screening surveillance examinations) (3). However, colonoscopy is imperfect and highly operator dependent. Wide variability still exists between endoscopists when considering ADRs (4). Moreover, a recent meta-analysis of more than 15,000 tandem colonoscopies (two same day colonoscopies performed on a patient) estimated that miss rates for adenomas were as high as 26% (5).

Clearly the quality of bowel preparation and colonic

inspection technique must be optimised. Beyond these, an important contributory factor is that polyps can be missed even when in the endoscopic field of view. Whilst human fatigue and poor concentration are obvious reasons, there is now a greater appreciation that many lesions are also visually subtle and as a result may be overlooked by the endoscopist. This is particularly true for flat or depressed type polyps, which are also more likely to harbour advanced histopathology with an associated increased risk of cancer progression. A number of different solutions have been developed in attempt to overcome this issue. These include enhanced imaging technologies, such as virtual chromoendoscopy, to increase the contrast between the background normal tissue and abnormal appearances of lesions, although studies have been inconclusive in demonstrating improved ADRs in average-risk patients (6). Meanwhile, educational quality improvement initiatives

to modify endoscopist behaviours, including pattern recognition of subtle polyps, have demonstrated an improved ADR but are challenging to sustain and implement at scale (7). A number of studies have shown that the presence of an additional observer of the colonoscopy video screen during procedures, such as by an experienced nurse, can lead to an increased ADR (8).

Artificial intelligence (AI) for polyp detection

An AI based computer-aided polyp detection system could act as a 'second observer' of the screen in real-time, potentially providing a performance level similar to that of an expert endoscopist. This concept has been the subject of research particularly in the computer science and engineering fields for over a decade (9). Early work focused on classical computer vision techniques, requiring human researchers to design meaningful image features, which could then be used to develop a prediction algorithm to detect polyps. Such techniques were guided by features such as shapes or colours of polyps to distinguish them from background normal mucosal appearances (10). These studies were often based on small image datasets with limitations in wider application due to the significant variation in polyp features observed during colonoscopy and associated high false positive rates.

An important initiative, known as the 'Automatic Polyp Detection Challenge' was led by a group of computer scientists, as part of the international Medical Image Computing and Computer Assisted Intervention (MICCAI) conference in 2015 (11). Such competitions allow for comparisons of different computer vision methods submitted by international groups using standardised datasets and performance metrics. Results from this competition were published and revealed that deep-learning methods using convolutional neural networks (CNNs) offered the best performance.

There has since been a dramatic increase in the number of publications related to the application of deep learning techniques for colorectal polyp detection. This rise is due to a combination of factors, including advances in algorithm development, enhanced computational power and specifically for colonoscopy, the availability of large annotated endoscopic imaging datasets which has been facilitated primarily by increasing clinician interest in the technology.

Most early clinician initiated pilot studies were developed and evaluated on retrospectively collected colonoscopy

datasets. Misawa *et al.* developed a CNN using 73 colonoscopy videos containing 155 polyps, of which 64.5% were flat shaped, which are typically difficult to detect (12). In addition, 391 polyp negative short videos were created from the colonoscopy procedures. Two expert endoscopists provided annotations for polyp presence in each frame which acted as the gold standard. The dataset was divided into random short, polyp-positive and polyp-negative videos for the purposes of training and testing. Based on a receiver operating characteristic (ROC) analysis, a cut-off value for the probability of detecting a polyp was set at 15%. Using a frame-based analysis, the algorithm achieved a sensitivity of 90.0%, specificity of 63.3% and accuracy of 76.5% on a 135 short video test dataset.

Urban *et al.* initially created a dataset of 8,641 colonoscopic images from 2,000 patients containing 4,088 unique polyps and 4,553 non-polyp images (13). Polyps were annotated by a team of colonoscopists using bounding boxes which represented the ground truth. A CNN was developed which was able to detect polyps with a cross-validation accuracy of 96.4% and area under the receiver operating characteristic curve of 0.991. The CNN was also evaluated further on small colonoscopy video datasets, most importantly one dataset consisting of 11 videos containing 73 polyps, where 'missed polyp' scenarios were simulated deliberately by the recording colonoscopist. The system was able to identify 67 of 73 polyps with a low frame-by-frame false positive rate of 5%. Although the video dataset was small, this provided preliminary evidence to support the hypothesis that an AI based polyp detection system could reduce the number of missed polyps in clinical practice.

More recently, a number of prospective clinical studies using automated colorectal polyp detection technologies have been published. Klare *et al.* performed a prospective observational cohort study to evaluate a prototype automated polyp detection software (APDS) during 55 routine colonoscopy procedures (mean patient age 67.4 years) performed by six colonoscopists (14). The APDS software analysed a weighted combination of colours, structure, textures and motion information to detect images containing possible polyps. A region of interest was marked with small green rings as an alert on an additional high definition video monitor. For the purposes of this study, the outcome of the APDS was not visible to the endoscopists. Instead, the additional monitor with the APDS output was available to an independent investigator who was out of sight of the endoscopist. Endoscopists were asked to give a verbal signal once a polyp had been detected. The independent investigator recorded whether

the APDS had detected the same polyp as the endoscopist and whether the APDS detection occurred before human detection. The APDS detected 55 of 73 polyps (75.3%). In the study, the ADR of the colonoscopists was 30.9% and the APDS was 29.1%. The APDS produced a mean number of 6 false positive alerts per procedure. Smaller polyp size and flat morphology were correlated with insufficient polyp detection by the APDS. Crucially, no polyp was detected by the APDS before the endoscopist. Whilst promising, the study highlighted that the system was not ready for clinical application.

Randomised clinical trials

Wang *et al.* conducted the first prospective randomised controlled trial examining the use of an automated polyp detection system during colonoscopy (15). The algorithm was a CNN that had previously been validated achieving a per-image sensitivity of 91.6% on 138 polyp positive videos and per-image-specificity of 95.4% on 54 polyp negative video procedures (16). In this single-centre, open, non-blinded trial, consecutive patients were randomised to undergo colonoscopy with or without assistance of the real-time software. The output of the system was displayed on an adjacent monitor providing a simultaneous visual and sound alarm. A total of 1,058 patients were included (536 standard colonoscopy and 522 AI assisted). There was a statistically significant higher ADR in the AI assisted arm versus the control group (29% and 20% respectively). Of all the detected polyps in the AI assisted arm, none were missed by the system. There was a total of 39 false alarms in the AI assistance group giving an average of 0.075 false alarms per colonoscopy, equivalent to an average of one false alarm for every 13 colonoscopies. It should be noted that the increase in ADR was predominantly due to an increase in diminutive (<5 mm) adenomas. There was also a significant increase in hyperplastic polyps.

The study by Wang *et al.* now provides high quality evidence that AI assistance for colorectal polyp detection can improve ADR. We will undoubtedly see numerous other prospective clinical trials in the near future as multiple developers look to evaluate their software. The optimal trial design and clinical end-points used to evaluate AI polyp detection software are unclear and become part of wider debate on methodologies for evaluating innovations in diagnostic colonoscopy. Beyond the relative strengths and weaknesses of randomised parallel or tandem study designs, there are important specific issues relating to AI

evaluation. Firstly, the inability to blind the endoscopist makes it difficult to assess the actual beneficial contribution of AI and account for potential observational bias. Some have proposed directly comparing AI detections and human operators in real-time within trials. However, claims relating to the potential benefits of AI such as ‘earlier detection’ or categorising a lesion as otherwise ‘missed’ by the endoscopist and revealed only by the addition of AI, can be challenging to define and record in a standardised manner. A randomised, double-blind study, has been presented only in abstract form to date, using a ‘sham’ or ‘false detection’ AI system in the control group versus a previously validated AI system in the research group (17). The output of either system was shown on a second monitor which was only visible to an observer who reported an area flagged by the system that was not seen by the endoscopist. The ‘false detection’ AI system was designed specifically to detect with a similar false positive rate as the genuine AI detection system. The trial demonstrated a statistically significant increase in ADR in the research group using the genuine AI polyp detection system.

Future perspectives

The existing prospective studies also highlight some initial limitations of the systems developed to date. For example, in the study by Wang *et al.*, the increase in detection rate is predominantly due to diminutive (<5 mm) adenomas, which arguably could have less impact on interval CRC rates. In addition, there was no difference in advanced adenoma or sessile serrated lesion detection rates in the two study arms. This suggests that further technological development should focus on the detection of more subtle, advanced lesions. Moreover, the unintended consequences of using AI assistance may include the increased detection of benign lesions, such as hyperplastic polyps, which could lead to the additional unnecessary removal of these polyps with associated risks and costs. In the future, this could be addressed by accurate AI assisted polyp characterisation or ‘optical biopsy’ diagnostic systems that allow for these to be disregarded. Promising results have already been reported for polyp characterisation AI systems (18,19). Ideally, future polyp detection and characterisation algorithms could be incorporated into one system to complement the normal endoscopist workflow.

The need for multicentre validation is crucial for AI, particularly in the context of deep-learning, where generalisability of results should be demonstrated

beyond the population in which the training data was used for algorithm development for effective widespread deployment. In addition, independent assessment of an algorithm outside of the setting where it was developed can be important to overcome any potential perceived conflict of interest.

Current AI polyp detection systems only address the issue of polyps that might be missed within the endoscopic field of view. Inadequate exposure of the colonic mucosa, whether due to poor bowel preparation or differences in operator technique, can lead to polyps remaining completely hidden out of view and undetectable by AI systems. Real-time feedback on colonoscopy withdrawal technique provided by computer software has shown promise in preliminary studies. Stanek *et al.* developed an image analysis system that included assessment of video frame quality, stool detection and withdrawal spiral motions of the colonoscope (20). Endoscopists were provided with real-time feedback in the form of a green marker that was displayed when each quadrant of the image was inspected along with a score. The software resulted in an improvement in the quality of colonoscopy inspection performed by third year gastroenterology trainees, based on objective assessments by two blinded investigators reviewing video recordings. More recently, deep learning approaches are being developed to predict depth and produce a 3D map of the colon, which could potentially provide a real-time quantitative measure of colonic mucosal inspection (21). However, this work is in its infancy and therefore in the more immediate future it is likely that AI polyp detection systems will be combined with other technologies aimed at addressing inadequate mucosal exposure such as mechanical add-on devices to colonoscopes.

Conclusions

Despite these limitations, we are now witnessing a watershed moment for AI assisted colonoscopy. It is encouraging that the focus of discussion is now moving rapidly towards the practical challenges of real-world implementation. Many of the issues relate broadly to the general deployment of AI technology in healthcare for example, data sharing and privacy, patient safety, accountability, transparency, cost-effectiveness and regulatory issues. Much can be learnt from other imaging-based specialties which are more advanced down the AI translational pathway, such as diagnostic radiology. However, it should be noted that there are

particular challenges related to colonoscopy, especially integration into the clinical workflow (22). Colonoscopy is a highly dynamic, video-based procedure where decisions need to be made in real-time, in contrast to the relatively controlled environments in which AI can be deployed in diagnostic radiology for instance. The ability of AI to augment endoscopic practice without causing unnecessary distractions or increasing procedure time will be vital for clinical adoption. The future success of AI assisted endoscopy will depend largely on the initial results of AI assisted polyp detection systems, which is the clinical application closest to widespread routine deployment. It is important that we take careful steps in this early phase, with wide engagement of all stakeholders, to harness the full potential of a future human-machine collaboration that will revolutionize endoscopic practice.

Acknowledgments

Funding: This work was supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. This work was also supported by the CRUK Experimental Cancer Medicine Centre at UCL and the Wellcome/EPSCRC Centre for Interventional and Surgical Sciences (WEISS) at UCL (203145Z/16/Z).

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Michael F. Byrne and Brandon J. Teng) for the series “Artificial Intelligence and Gastrointestinal Cancer Column” published in *Journal of Medical Artificial Intelligence*. The article has undergone external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jmai.2019.09.02>). The series “Artificial Intelligence and Gastrointestinal Cancer Column” was commissioned by the editorial office without any funding or sponsorship. LB Lovat: Minor shareholder in Odin Vision & Dynamx Medical. Research grants from Medtronic, Pentax Medical, DynamX Medical. Scientific Advisory Boards: Dynamx Medical, Odin Vision, Ninepoint Medical. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all

aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Arnold M, Sierra MS, Laversanne M, et al. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017;66:683-91.
2. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of Colorectal Cancer by Colonoscopic Polypectomy. *N Engl J Med* 1993;329:1977-81.
3. Kaminski MF, Regula J, Kraszewska E, et al. Quality Indicators for Colonoscopy and the Risk of Interval Cancer. *N Engl J Med* 2010;362:1795-803.
4. Barclay RL, Vicari JJ, Doughty AS, et al. Colonoscopic Withdrawal Times and Adenoma Detection during Screening Colonoscopy. *N Engl J Med* 2006;355:2533-41.
5. Zhao S, Wang S, Pan P, et al. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: a Systematic Review and Meta-analysis. *Gastroenterology* 2019;156:1661-74.e11.
6. Kamiński MF, Hassan C, Bisschops R, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2014;46:435-49.
7. Coe SG, Crook JE, Diehl NN, et al. An Endoscopic Quality Improvement Program Improves Detection of Colorectal Adenomas. *Am J Gastroenterol* 2013;108:219-26.
8. Lee CK, Park DI, Lee SH, et al. Participation by experienced endoscopy nurses increases the detection rate of colon polyps during a screening colonoscopy: a multicenter, prospective, randomized study. *Gastrointest Endosc* 2011;74:1094-102.
9. Ahmad OF, Soares AS, Mazomenos E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *Lancet Gastroenterol Hepatol* 2019;4:71-80.
10. Karkanis SA, Iakovidis DK, Maroulis DE, et al. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans Inf Technol Biomed* 2003;7:141-52.
11. Bernal J, Tajkbaksh N, Sanchez FJ, et al. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results from the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Trans Med Imaging* 2017;36:1231-49.
12. Misawa M, Kudo S, Mori Y, et al. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* 2018;154:2027-2029.e3.
13. 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.e8.
14. Klare P, Sander C, Prinzen M, et al. Automated polyp detection in the colorectum: a prospective study (with videos). *Gastrointest Endosc* 2019;89:576-82.e1.
15. 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. [Epub ahead of print].
16. 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.
17. Zhou G, Liu X, Berzin T, et al. A Real-Time Automatic Deep Learning Polyp Detection System Increases Polyp and Adenoma Detection During Colonoscopy: a Prospective Double-Blind Randomized Study. *Gastroenterology* 2019;156:S1511.
18. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94-100.
19. Mori Y, Kudo S, Misawa M, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: a Prospective Study. *Ann Intern Med* 2018;169:357-66.
20. Stanek SR, Tavanapong W, Wong J, et al. SAPPHERE: a toolkit for building efficient stream programs for medical video analysis. *Comput Methods Programs Biomed* 2013;112:407-21.
21. Rau A, Edwards PJE, Ahmad OF, et al. Implicit domain

adaptation with conditional generative adversarial networks for depth prediction in endoscopy. *Int J Comput Assist Radiol Surg* 2019;14:1167-76.

- 22 Ahmad OF, Stoyanov D, Lovat LB. Human-machine collaboration: bringing artificial intelligence into colonoscopy. *Frontline Gastroenterol* 2019;10:198 LP-199.

doi: 10.21037/jmai.2019.09.02

Cite this article as: Ahmad OF, Lovat LB. Artificial intelligence for colorectal polyp detection: are we ready for prime time? *J Med Artif Intell* 2019;2:16.