Artificial Intelligence in

Gastrointestinal Endoscopy

Artificial Intelligence in Gastrointestinal Endoscopy

Contents

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastrointestinal Endoscopy*, Omer Faruk Ozkan, FEBS, MD, PhD, Professor, Surgeon, Department of General Surgery, Universtiy of Health and Science Umraniye Training and Research Hospital, Istanbul 34760, Turkey. ozkanfomer@gmail.com

AIMS AND SCOPE

The primary aim of *Artificial Intelligence in Gastrointestinal Endoscopy* (*AIGE*, *Artif Intell Gastrointest Endosc*) is to provide scholars and readers from various fields of artificial intelligence in gastrointestinal endoscopy with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

   AIGE mainly publishes articles reporting research results obtained in the field of artificial intelligence in gastrointestinal endoscopy and covering a wide range of topics, including artificial intelligence in capsule endoscopy, colonoscopy, double-balloon enteroscopy, duodenoscopy, endoscopic retrograde cholangiopancreatography, endosonography, esophagoscopy, gastrointestinal endoscopy, gastroscopy, laparoscopy, natural orifice endoscopic surgery, proctoscopy, and sigmoidoscopy.

INDEXING/ABSTRACTING

There is currently no indexing.

RESPONSIBLE EDITORS FOR THIS ISSUE

Colonoscopy and artificial intelligence: Bridging the gap or a gap needing to be bridged?

James Weiquan Li, Tiing Leong Ang

James Weiquan Li, Tiing Leong Ang, Department of Gastroenterology and Hepatology, Changi General Hospital, Singapore 529889, Singapore

Corresponding author: Tiing Leong Ang, FASGE, FRCP, MBBS, Chief Physician, Professor, Department of Gastroenterology and Hepatology, Changi General Hospital, 2 Simei Street 3, Singapore 529889, Singapore. ang.tiing.leong@singhealth.com.sg

Abstract

Research in artificial intelligence (AI) in gastroenterology has increased over the last decade. Colonoscopy represents the most widely published field with regards to its use in gastroenterology. Most studies to date center on polyp detection and characterization, as well as real-time evaluation of adequacy of mucosal exposure for inspection. This review article discusses how advances in AI has bridged certain gaps in colonoscopy. In addition, the gaps formed with the development of AI that currently prevent its routine use in colonoscopy will be explored.

Key Words: Artificial intelligence; Endoscopy; Colonoscopy; Detection; Diagnosis

Core Tip: The use of artificial intelligence (AI) for colonoscopy has been studied most extensively for polyp detection and characterization. Despite advances made in this field, AI systems studied for these purposes represent only the machine learning domain of AI, and individual machine learning algorithms used in these studies are each focused on performing a very narrow task. While they may bridge existing gaps in polyp detection and real-time optical diagnosis of colorectal polyps, the introduction of AI into colonoscopy will also mean that there are new gaps that must be bridged for AI systems to be routinely used in clinical practice.

INTRODUCTION

The use of artificial intelligence (AI) in gastroenterology has gained momentum in the past decade. This is reflected in the increasing number of publications in the field of AI in endoscopy, most of which have been centered on colonoscopy. This is understandable as the unique role of colonoscopy in the prevention and management of colorectal cancer (CRC), together with the unmet needs in this field, has created the perfect milieu for the introduction of AI into world of endoscopy.

CRC represents one of the leading causes of cancer-related morbidity and mortality worldwide[1,2]. Colonoscopy decreases CRC-related mortality[3,4], with a 1% increase in adenoma detection rate (ADR) estimated to decrease interval CRC by 3%[5]. As such, a key barrier to overcome is the adenoma miss rate (AMR), which has been estimated in a meta-analysis to be a high as 22% overall, with a higher AMR when diminutive adenomas are considered[6]. Another unmet need in colonoscopy is the need for accuracy in the optical diagnosis of colonic polyps in relation to their actual histology. Up to 90% of lesions detected on colonoscopy consist of diminutive (≤ 5 mm) and small (6-9 mm) polyps, with the progression rates to advanced adenomas or CRC postulated to be low based on evidence from available studies[7]. It is therefore no surprise that most of the literature to date has focused on computer-assisted detection (CADe)[8,9] and computer-assisted diagnosis (CADx)[10-12] applications in colonoscopy.

This review article evaluates the areas in colonoscopy where AI may be a bridge for certain gaps in clinical practice. It will also explore in detail the current limitations and pitfalls in the application of AI in colonoscopy, highlighting how despite the proliferation of literature on this topic and what it promises to offer, AI may be a new gap in endoscopy which clinicians need to work to bridge.

LITERATURE SEARCH

We performed a comprehensive literature search in the PubMed, MEDLINE and EMBASE (up to March 17, 2021) electronic databases to identify relevant clinical trials that evaluated the roles of AI systems in colonoscopy. Electronic searches were also supplemented with manual searches of the references in the included studies and review articles.

AI TERMINOLOGY IN COLONOSCOPY

What does the term AI mean in colonoscopy?

The term "artificial intelligence" was first coined by John McCarthy in 1956 at the Dartmouth Summer Research Project. In essence, it is a branch in computer science where computer systems are designed to perform tasks which would ordinarily require human intelligence. This definition is extremely broad and often confuses clinicians to what exactly the capabilities, and by inference, the limitations of AI are in their respective fields[13]. There is therefore a need to define what AI means in colonoscopy as this is a prerequisite for meaningful discussion of its role in colonoscopy.

Published and ongoing studies incorporating AI in the context of colonoscopy involve the machine learning (ML) domain of AI. ML refers to the use of algorithms, which form predictive and descriptive models based on analysis of input data provided by investigators (the training set)[14]. These algorithms undergo multiple iterations of these models with the goal of performing a specific task, the aim of which is to come to a specified classification output (*e.g.*, polyp or no polyp) when the algorithms are tested on an unseen set of data (the test set). In practical terms and in the context of colonoscopy, this is achieved using either handcrafted models or deep learning (DL).

A useful mental model in understanding the scope of and roles which AI plays in colonoscopy is to regard the progress made in this field as "waves"[15]. It is crucial to understand that the methods, technologies, and results from earlier AI studies are not obsolete the moment a "better" or "faster" computer system is available based on results we as clinicians are familiar with such as the ADR and adenoma per colonoscopy (APC), or technical matrices that we may gravitate towards such as the processing speed of an algorithm. Rather, these "waves" are continuously interacting

and building on top of each other, and as a result, have a strong influence on the development of later technologies. The earlier "waves" remain relevant and may sometimes harbor solutions to certain issues faced with CADe and CADx support tools, which will be discussed later in this article. Having this mental model also helps us better understand the intrinsic biases present in all forms of ML regardless of advancements made in AI, which is essential for critical appraisal of literature surrounding AI in clinical practice.

AI terminology relevant to colonoscopy

Commonly used terms in AI which are relevant to this review article will be discussed here. This list is not meant to be exhaustive and is meant instead to highlight terms which will help the reader understand the later critiques and solutions offered in this paper.

AI can be categorized very broadly into weak (or narrow) AI and strong AI. The former refers to systems built to solve a specific problem or performing a single task extremely well, without an emphasis on elucidating how human reasoning works. This type of AI operates within significant constraints and a limited context. The latter term, also referred to as artificial general intelligence, aims to build systems which think like humans.

Features in ML refer to the set of numbers which quantitatively summarize and represent in a compact fashion the input data. For example, differences in morphology of polyps as defined in the Paris classification[16] and pit patterns[17] can be converted into different arrays of numbers which an ML algorithm can use to generate a prediction such as "polyp" or "no polyp" in a CADe application. Conventional learning by the ML algorithm may be supervised, where training takes place on labeled data sets, or unsupervised, where commonalities are used to identify groups within data. Supervised learning occurs on pre-established input and output pairs, enabling the ML algorithm to learn predictive mathematical models which can then map the input from unseen data into an outcome of interest (*e.g.*, neoplastic, or hyperplastic). In contrast, unsupervised learning predicts similarities between data points through looking at the underlying structure of the data provided, with no prior knowledge of its significance.

Handcrafted knowledge represents the first "wave" of AI. This consisted of knowledge-based methods where manual extraction and selection of characteristics of an object such as polyp shape and texture, are used to create mathematical models which can achieve a class or numerical output. This is labor-intensive and as a result, are usually implemented on small sets of data. These systems do not have the ability to learn and were of limited clinical use. DL is another form of ML where an artificial neural network (ANN) is used to perform the same task. ANNs are supervised ML models where interconnected artificial neurons form layered networks. Signals travel *via* weighted inputs from artificial neurons in the previous layer to the next layer, which then propagate the signal when a predefined threshold is reached, like how biological neurons work. Classification can be optimized, and the system enhanced by adjustment of the weights given to these inter-neuron connections.

Deep convolutional neural networks (DCNNs) have enabled more hidden layers to be added to the input and output layers of ANN, a development which has been facilitated by advancements made in other areas of computer science as this is computationally expansive. In addition, convolutional layers apply filters (a set of weights) in a systematic fashion to each overlapping part of the input data. In this manner, large numbers of filters can be applied to the training set of data in parallel under the constraints of the intended task, for example classification of an image as having a polyp or not in colonoscopy, allowing information to be extracted directly from images training data to form a feature map. DCNN usually require large amounts of labelled training data, which are derived wither from public databases or private collections in individual institutions.

Hyperparameters in ML refer to all parameters that have been arbitrarily set by the investigator and are used to configure the model for optimal performance at a specific task or on a specific dataset. As opposed to model parameters, which are learned automatically during training of the model, hyperparameters are manually set and affects the learning process and ultimately, the behavior of the model. This is useful in understanding the roles (and potential biases resulting from) the optimization and training process of AI models used in colonoscopy. The training set refers to the initial dataset used to determine optimal parameters after multiple rounds or iterations of adjustments. The validation set is mostly (but not always) a different dataset where these parameters are tested and adjusted. It is also used to optimize the hyperpara-meters in the model. Lastly, the test set refers to a new set of unseen data which is used

to test the model and its generalizability.

AI: BRIDGING THE GAP IN COLONOSCOPY

AI in the field of colonoscopy has been studied primarily for polyp detection, polyp characterization in terms of predicted histology, and for quality assurance in the performance of colonoscopy.

Polyp detection

The rate of missed polyps was mentioned earlier in the introduction. The AMR in influenced by different factors, among which the endoscopist is considered one of the major determinants[18-21]. These human biases may be due to distraction during colonoscopy, fatigue, or the inability to maintain a sustained level of alertness during withdrawal. These lead to errors in perception where the endoscopist may miss polyps which are visible on the monitor. The role of "second readers" in colonoscopy in increasing ADR[22,23] lends support to the hypothesis that CADe may help increase APC and ADR, and decrease AMR, during colonoscopy.

At the time of writing, there are six randomized controlled trials (RCTs)[24-29] to date that have evaluated the role of CADe in colonoscopy. Hassan *et al*[9] recently performed a systematic review and meta-analysis of five of these studies[24,25,27-29], which consisted of 4354 participants. The pooled ADR was significantly higher in the CADe group compared with the control group (36.6% *vs* 25.2%; relative risk [RR] 1.44; 95% confidence interval [CI]: 1.27-1.62; $P < 0.1$), with all of the included RCTs reporting a significant increase in ADR individually. APC, which is defined as the total number of adenomas found divided by the total number of colonoscopies and has good correlation with ADR[30,31], was also significantly higher in the CADe compared to the control group (0.58 *vs* 0.36; RR 1.70; 95%CI: 1.53-1.89; $P < 0.01$). The mean withdrawal time in the CADe and control groups was shown to be statistically different in this meta-analysis.

An interesting prospective study conducted by Wang *et al*[32] showed that the AMR was decreased with CADe. This study differed from the RCT mentioned above in that tandem colonoscopies were performed. Patients in this study were randomly assigned to colonoscopy with CADe or colonoscopy without CADe by an endoscopist, followed immediately by the other procedure. The study showed that the AMR and polyp miss rates were significantly lower in the CADe colonoscopy group compared to the routine colonoscopy group (13.89% *vs* 40.00%, $P < 0.0001$ and 12.98% *vs* 45.90%; $P < 0.0001$, respectively). These results were also consistent regardless of colonic segments, *i.e.* the AMR was significantly lower in the CADe group in the ascending, transverse, and descending colon.

Polyp characterization (optical prediction of polyp histology)

In contrast to CADe for polyp detection, CADx deals with the interpretation of polyp appearance during colonoscopy to determine the predicted histology. Polyp classification systems such as the Kudo pit pattern[17], Sano *et al*[33], NBI International Colorectal Endoscopic (NICE)[34], and Japan NBI Expert Team (JNET)[35] classifications were developed with the purpose of predicting polyp histology and severity of neoplasia to guide therapy. The use of these classification systems for optical prediction of colorectal polyp histology requires the proper equipment, structured training, and experience in clinical application. Studies have shown wide variation in the sensitivity and specificity of NICE and JNET classifications, with most studies reporting a moderate interobserver agreement at best[36-39].

With the clinical use of CADe, the detection of diminutive polyps is likely to increase exponentially, as demonstrated in the CADe RCT mentioned[24,25,27-29]. Most diminutive polyps tend to be hyperplastic in nature with low malignant potential. The "resect and discard" and "detect and leave" strategies for such polyps were previously studied to address these issues before the emergence of AI but have failed to gain traction due to the need for better quality training and quality assurance in the accurate optical diagnosis of colon polyps[40-42]. The threshold for optical biopsy technologies in high confidence predictions established by the American Society for Gastrointestinal Endoscopy (ASGE) Preservation and Incorporation of Valuable Endoscopic Innovations (PIVI)[43] are deemed appropriate targets for CADx support tools[44]. A systematic review and meta-analysis by ASGE[45] showed that these thresholds were met using NBI only among NBI experts, illustrating the difficulty and practical limitations of replying on the use of these forms of imaging by

endoscopists in general to achieve accurate optical diagnoses of colorectal polyps. Hence, this represents a significant clinical gap which AI has the potential to bridge in colonoscopy.

CADx is postulated to aid in this field of colorectal polyp management by using DL models to increase the accuracy of prediction of polyp histology during colonoscopy[46]. At the time of writing, there are currently no RCT evaluating CADx in colonoscopy. In a study by Jin *et al*[10], a DCNN was trained to differentiate between adenomatous and hyperplastic diminutive colorectal polyps with an overall accuracy of 86.7% using polyp histology as the gold standard. The system was tested on 22 endoscopists with varying expertise such as novice endoscopists, colonoscopy experts with differing levels of expertise in NBI, and NBI-trained experts. The use of CADx markedly improved the accuracy of novice endoscopists in differentiating adenomatous and hyperplastic polyps from 73.8% to 85.6% ($P < 0.05$), which was comparable to the baseline accuracy of NBI-trained experts (87.6%). However, in the colonoscopy expert and NBI-trained expert groups, this increase in accuracy was less impressive (83.8% to 89.0% and 87.6% to 90.0, respectively). The overall time to diagnosis per polyp was also decreased from 3.92 s to 3.37 s; $P = 0.42$).

A review of CADx predictions[47] for diminutive polyp histology which included 9 studies[48-56] showed a pooled sensitivity of 93.5% (95%CI: 90.7%-95.6%) and specificity of 90.8% (95%CI: 86.3%-95.9%), with a pooled area under the curve of 0.98. This pooled analysis of diminutive polyps had a negative predictive value (NPV) of 0.91 (95%CI: 0.89-0.94). This meets the 90% or greater threshold for NPV in adenomatous histology in rectosigmoid diminutive polyps recommended by the ASGE PIVI[43] and thus would in theory support a "diagnose and leave" strategy if these applications are validated in clinical use. However, most of these studies are retrospective in nature or, when conducted prospectively, involved the use of *ex vivo* video or still images.

Few prospective studies on CADx in real-time colonoscopy are currently available in the literature. In a single-center, open-label, prospective study of 791 consecutive patients undergoing colonoscopy in a university hospital, Mori *et al*[54] evaluated the performance of CADx in a clinical setting using endocytoscopy (CF-H290ECI; Olympus Corp, Tokyo, Japan). NBI was applied to visualize the microvascular pattern and methylene blue staining for cellular structure under these ultra-magnifying colonoscopes with 520X optical zoom capability. Of the 466 diminutive polyps found in this study, 250 polyps were in the rectosigmoid colon. The CADx system using endocytoscopy had an NPV for diminutive rectosigmoid adenomas ranging from 93.7% to 96.4% with methylene blue staining and 95.2% to 96.5% with NBI. This is well above the "diagnose and leave" threshold of 90% recommended by the ASGE PIVI[43] described. This prospective study also provides evidence for utilization of CADx for prediction of polyp histology in a clinical setting which may have an impact on decisions on polyp management real-time.

In an earlier study with a similar design by Horiuchi *et al*[56], CADx was evaluated with the use of autofluorescence imaging (AFI) to differentiate diminutive rectosigmoid polyps in real-time colonoscopies. The CADx system used software-based automatic color intensity analysis, which utilized AFI's ability to differentiate polyps based on the ratio of green to red tone intensities and was tested on 258 rectosigmoid polyps in 95 patients undergoing colonoscopy. The CAD-AFI system achieved an NPV for adenomatous polyps of 93.4% (95%CI: 89.0%-96.4%), which again exceeds the 90% "diagnose and leave" threshold[43]. In addition, the NPV using CAD-AFI was comparable to that of diagnoses made by endoscopists using AFI in the study (94.9%; 95%CI: 90.8%-97.5%).

Quality assurance in colonoscopy

Quality indices such as a high cecal intubation rate and adequate withdrawal time have been studied extensively[57,58]. However, these quality indices in colonoscopy performance and reporting are not always adhered to for a variety of factors such as training, lack of real-time feedback and failure of enforcement[59-61]. In an RCT of 704 patients by Gong *et al*[26], which used an AI system called ENDOANGEL, the withdrawal speed and time, as well as the adequacy of mucosal exposure, was monitored in real-time and in an automated fashion. The resulted in a significantly longer withdrawal time in the ENDOANGEL[62] *vs* the control group (mean 6.38 min *vs* 4.76 min, respectively; $P < 0.0001$). This translated into an increased ADR in the ENDOANGEL group and, more significantly, is the only RCT to date which demonstrates an AI system which can increase the rate of detection of adenomas 10 mm or larger in size (10/355 *vs* 1/349, respectively; odds ratio [OR] 9.50, 95%CI: 1.19-75.75; $P = 0.034$). Su *et al*[28] used both a CADe tool together with an automatic quality

control system (AQCS) to increase ADR and APC. The AQUS consisted of a timer on the monitor and audio prompts for the Endoscopist to slow down withdrawal speed when unstable and blurry frames were displayed or when the Boston Bowel Preparation Scale (BPPS) in a colonic segment was < 2. This study showed an improved withdrawal time (7.03 min *vs* 5.68 min; $P < 0.001$) and rate of adequate bowel preparation (87.34% *vs* 80.63%; $P = 0.023$) in the AQCS group in addition to the mentioned significant increase in ADR and APC.

AI: A GAP NEEDING TO BE BRIDGED IN COLONOSCOPY?

While AI has emerged in the world of endoscopy with much promise, there are several significant gaps which need to be bridged before it can be routinely applied in colonoscopy in a clinical setting.

Undefined and unspecified role in clinical environment

A major bridge which needs to be bridged before AI systems can be applied in routine environments is its generalizability. Three of the five CADe RCT[25,27,28] available involved senior endoscopists with extensive experience in colonoscopy. ADR is dependent on several factors, one of which includes experience. A more experienced endoscopist is not only skilled in recognition, but also in scope handling and consequent mucosal exposure during withdrawal. The role of a "second reader" in previous studies[22,23] in increasing small adenoma detection rates suggests that trainees and Nurses, who by inference have less "experience" than the senior endoscopist, have no issues recognizing a polyp visible on screen. In addition, as discussed in the ENDOANGEL study, one of the largest increments in ADR and the only increase in detection of adenomas larger then 10 mm was seen in the RCT by Gong *et al*[26], where real-time feedback on adequacy of mucosal exposure was studied. An obvious but less often mentioned fact is that any CADe algorithm is still completely dependent on the endoscopist to present optimal images with adequately exposed colonic mucosa in each real-time colonoscopy performed in a busy clinical setting. A polyp not visible on the screen will not be detected by a CADe tool, no matter how powerful the algorithm is[33]. This has implications on how generalizable available data is for clinical use, as more studies involving both "high detectors" and "low detectors" are required[25,63].

Most RCT in CADe to date were conducted in single centers. Moreover, except for the study by Wang *et al*[27] where a second monitor was used and visible only to an observer who reported the alerts, the rest of the RCT were non-blinded studies[24-26,28-29]. It is not known what the impact of the latter factor may be in actual clinical practice, as non-blinded endoscopists in these studies may put in more effort in exposing colonic mucosa for inspection when they are under observation. This Hawthorne Effect, together with the single-center experiences of most of these RCT, also limit their generalizability to routine clinical practice. While single monitors are encouraged[44] due to presumed gaze limitations of endoscopists and the need to reduce distractions, it is the opinion of the authors that a dual monitor setting in clinical trials plays a crucial role in achieving a double-blind and objective environment for assessment of the performance of the AI system and to bridge this gap. Furthermore, it resembles tandem colonoscopy in that the performance of the AI system can be compared directly against endoscopists of varying skill levels and experience. Useful information such as the AMR can be determined accurately without the patient having to go through an additional colonoscopy like in a traditional tandem study with this methodology.

Another limitation to the generalizability of the published results of AI systems for polyp detection and characterization is the differences in operational environments of different endoscopy suites and centers. These can vary greatly between institutions, even those located in the same country[64]. Unlike a new endoscopic method or classification system which can be taught or standardized in training or with major society guidelines, different AI algorithms have unique hardware and software requirements which must be fulfilled for technical integration into the operational environment. For instance, some may be fully integrated into the processing unit[65] while others may be web-based applications or require an additional laptop to be linked to the endoscopy stack to function. The latter may require cloud integration support, which in turn is likely to be vendor-specific and has implications in procurement and cybersecurity. This technical integration into the operational environment is key, as the development environment from which these AI systems are derived may be vastly

different[66]. Most clinical trials understandably focus on the clinical aspects like the ADR and APC and the outcomes will inevitably be based on these primary objectives. However, few studies have reported the technical specifications and limitations of the AI systems they are investigating. The rare studies that do report them, do so in varying details, most of which are insufficient for interpretation and contextualization into the operational environment. Moreover, most of the published trials have been conducted in academic or expert centers and in several instances, in the same institutions where the AI algorithm was developed, *i.e.* the development and operational environment are the same[3,47]. Individual institutions may have difficulty integrating these systems due to budgeting constraints, existence of legacy systems which are incompatible with the software and hardware requirements of the AI systems, logistical limitations such as space, and established workflows in endoscopy which does not cater to the introduction of an AI system.

The current scope of AI applications in colonoscopy in the literature is also largely skewed towards to polyp detection, characterization, and assessment of adequacy of mucosal exposure, which is ultimately linked to ADR. When translated to clinical practice, this effectively confines the indications for which AI should be used in colonoscopy to CRC screening or indications where one might expect to find colorectal polyps in the process of performing a colonoscopy. All systems developed in the field of AI in colonoscopy, from handcrafted models to the most complex DCNN, are fundamentally "weak AI." This is a term used to describe AI systems designed to solve a single problem or narrow task[15]. In a clinical setting, indications for colonoscopy are widely variable and the pre-test probability of finding of a polyp may be low. An endoscopist will be able to process the demographic data, clinical course, medical history, clinical condition, laboratory investigations and concerns of the patient and use this information during the colonoscopy. For example, an 85-year-old patient who is troubled by per rectal bleeding has a hugely different indication and clinical index of suspicion than a 50-year-old male with a family history of early CRC. In the former case, the endoscopist's focus may be on looking for angiodysplasia, diverticular disease or hemorrhoids as the etiology. A "strong AI" system would be able to think and adapt like a human and calibrate the weights in its layers to perform the task at hand, determine the appropriate classification output and achieve the correct alarm settings. However, current AI systems will continue looking for polyps and may present a distraction to the Endoscopist if used in this clinical example, prolonging the time taken for colonoscopy in an elderly patient, who may have multiple co-morbidities and for whom resection of small or diminutive adenomas may not have clinical relevance, much less answer the clinical question at hand. A trainee endoscopist or an experienced nurse, on the other hand, would be able to immediately recognize an unusual finding, such as multiple angiodysplasia or extensive diverticular disease, even if they were not formally trained to recognize these abnormalities.

It should be noted that AI has also been studied in colonoscopy outside the context of polyp detection, characterization, and quality assurance. Endocytoscopy has been used with AI to accurately detect persistent histologic inflammation in patients with ulcerative colitis (UC) which was reproducible based on static images[67]. A separate group used a deep neural network to predict endoscopic and histologic remission in UC patients based on evaluation of static images obtained from colonoscopy with high accuracy[68]. However, studies looking at indications other than polyp detection and characterization are few and far between.

Technical biases and lack of technical knowledge among clinicians
There is significant variability and a lack of standardization in reporting of the technical aspects of AI algorithms in clinical trials[69]. In addition, clinicians may not have the technical knowledge to critically appraise AI literature given that this has not been a formal part of training or an emphasis in clinical practice until relatively recently. A "minimum reporting standard" and practical knowledge of terms and potential biases on the part of investigators and clinicians, respectively, is required to bridge these gaps[70-72].

A practical knowledge of commonly used terms and how AI systems are derived is necessary for the clinician to appreciate the technical biases inherent to these algorithms. While the inclusion criteria of patients in clinical trials is clearly defined, the criteria for inclusion of the input data for the AI system during training and validation may not always be included in the methodology. This is crucial as most AI systems for CADe were tested in the same centers where they were developed[73]. This is often due to the ease with which large amounts of data are readily available for training and validation. Although the training, validation, and test datasets may be

different, they could be derived from the same database in a single, often expert, center, which is then split to form these datasets. The nature of the images used could be highly similar in terms of quality (*e.g.*, no confounding fecal material and bubbles and polyps always centered in the image) and labelling (*e.g.*, experts from different centers may mark out the most obvious abnormal area or delineate even the most minute detail which does not look like normal colonic mucosa for sessile serrated polyps depending on their level of skill and the training received, while experts from the same center are more likely to label lesions similarly). Prevalence and variability in presentations of disease may also differ depending on the populations studied, but the sample of images used in training and validating the AI algorithm may not necessarily reflect this natural variability of disease if data from a single center is used in the development of the AI system. This is a form of selection bias, as input data is not selected at random and hence is not fully representative of the study population in which the AI system is meant to function. This could impact the hyperparameters chosen during validation, and lead to overfitting, which occurs when the mathematical model derived is optimized to work on the training data and fits this data too tightly. This would limit its generalizability when new data is presented to the same AI algorithm.

Moreover, the proportion of "positive" to "normal" images used for training is not often mentioned in the published literature. For example, in a CADe application, polyps of various shapes, sizes and colors may be included in the training dataset to expose the AI algorithm to all possible eventualities when presented with an image with even the subtlest polyp. However, the "normal" images used may be disproportionately lower when compared to the natural prevalence of adenomas in the population. In addition, there may not be the same rigor in the selection of "normal" images for training. Variations in degrees of bowel preparation, bubbles, and artefacts due to the light source reflecting off normal colonic mucosa may thus not be reflected in images supplied to the AI algorithm for training. Positive and negative predictive values are determined by the prevalence of disease, and this may result in a higher proportion of false positives per true positive detected in clinical practice, depending on how the ratio of "positive" to "normal" images used in training compares with the true prevalence of the lesion of interest (*e.g.*, polyps) in the study population. This is a factor which needs to be adjusted for in the AI algorithm[74].

A certain form of publication bias may also exist as clinicians who wish to publish on the topic of AI will search for references almost exclusively from medical journals. For example, meta-analysis and systematic reviews on the use of AI in colonoscopy may take a very clinical slant, while publications in computer science and engineering journals which may add technical dept to the chosen topic on AI being discussed will not be included. Even if a search were performed for these articles, the inclusion criteria for the literature search will inevitably involve clinical-based endpoints like ADR and APC, and almost always exclude publications from computer science and engineering journals as a result. The barrier to entry in medical journals for these studies is high, as editors and reviewers, who themselves are clinicians, may not have enough technical knowledge to feel comfortable about accepting these articles for publication, and may also be compounded by fear of a lack of interest or understanding in the readership. On the other hand, AI and ML experts will not be familiar with the clinical aspects or relevance of their research and would not be able to pitch it at a level that would be acceptable to a Medical journal and its readership. This may result in a "reinforcement bias" of sorts, where only certain types of publications from a few expert centers and which revolve around common themes are published repeatedly and in different forms in Medical journals, whereas significant developments in AI and ML which may have the potential for changing clinical practice are missed out. The same technical terms specific to these publications will also be mentioned repeatedly, while novel approaches and new technical terms unfamiliar to clinicians may never see publication in a medical journal. The endoscopy readership may already have been "overfitted" towards polyp detection and characterization in the endoscopy literature[75], while neglecting the fact that, as mentioned, the use of AI in colonoscopy to date has utilized only an extremely limited aspect of AI and in a very narrow clinical context. Including computer science experts in the editorship and as reviewers for Medical journals may help to bridge the gap in these technical and publication biases.

Physician sentiment towards AI
Physician sentiment is a significant determinant on how quickly technologies and recommendations are deployed in a clinical setting. A recently conducted online survey among Gastroenterologists in the United States showed high overall interest in

CADe and perception that it would increase ADR (85.5% and 75.8%, respectively)[76]. However, the same survey also showed that majority of the respondents felt that CADe will prolong the time taken per colonoscopy, despite evidence to the contrary[9,24,25,27-29].

Concerns about operator dependence, or "deskilling", of the Endoscopist due to reliance on CADe and CADx for detection and characterization of polyps, respectively, are also mentioned in this survey[76] and other reviews[44,73]. Another major concern shown in the survey by Wadhwa *et al*[76] was the perceived increase in cost per procedure (75.2%). While concerns such as withdrawal time have been addressed independently in several RCT, others such as operator dependance and cost-effectiveness have not studied. Hence, physician sentiment may be another significant gap in AI which needs to be bridged in the field of colonoscopy.

Medicolegal challenges and future directions

AI algorithms which utilize DL are considered "black box" models, meaning that it is almost impossible to trace the decision-making process which led to the output determined by the algorithm when faced with a specific task (*e.g.*, polyp or no polyp in the image, hyperplastic or adenomatous). One of the major gaps in clinical use of AI systems in colonoscopy is medicolegal liability when a misdiagnosis or missed diagnosis occurs. While a clinician's account of events and the accompanying documentation can be helped up to scrutiny, the black box nature of DL algorithms means that the root cause and mitigating factors surrounding such a case may never be elucidated or even discovered. This has ethical implications in the event of harm to a patient[77], particularly if no clear protocol exists to define how an AI system should interface with its user and what its limits are, as the error may be due to deviation from safe use of the system or from an error of the AI system itself[78].

As AI systems, like other healthcare interventions, may have unpredictable errors, this inability to explain the errors or to detect them as they occur due to their black box nature may result in a perpetuation of systemic errors with unknown clinical implications if they are scaled up rapidly for routine clinical use in all colonoscopies. It is also unknown if the liability rests with the manufacturer, the regulatory body approving its use, or the clinician interfacing with the AI system. Having a reliable and accountable post-deployment surveillance plan is perhaps one of the strategies to minimize this risk.

Lastly, while AI systems have been shown to improve various quality indices associated with colonoscopy, one should remember that they are still limited most of all by our current expertise in this field. A useful example to illustrate this is the fact that there is currently no AI system capable of detecting dysplasia in UC. The availability of DCNN with high computing power and hardware to support the required processing speeds would have made this a rather simple task from an ML point of view. However, the optimal method of surveillance for dysplasia in UC and its optical features do not have the same clinical certainty as colorectal polyps in CRC screening, with resultant discrepancies in surveillance and biopsy practices[79,80]. Moreover, there is wide interobserver variability in the histological diagnosis of dysplasia in UC[81] and an inadequate understanding of its pathogenesis[82]. It is therefore understandable that there would be a paucity of expertly labelled data for "dysplasia" and "non-dysplasia" controls in UC patients for the training of an ML algorithm. Similarly, other potential AI applications in colonoscopy could include localization of diverticular bleeding and an automated scoring system for adequacy of bowel preparation which includes the BPPS[83] and the newly validated Colon Endoscopic Bubble Scale[84]. The clinical expertise and research in these fields must progress sufficiently for an accompanying increase in standardized and labelled data to be available for such future AI systems to be trained on and to materialize.

CONCLUSION

Despite the advances made in the field of AI, most notably for polyp detection and characterization in colonoscopy, there remain significant gaps which need to be bridged before its routine clinical use in colonoscopy.

REFERENCES

1　**Bray F**, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *Lancet Oncol* 2012; **13**: 790-801 [PMID: 22658655 DOI: 10.1016/S1470-2045(12)70211-5]

2　**Araghi M**, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, Arnold M. Global trends in colorectal cancer mortality: projections to the year 2035. *Int J Cancer* 2019; **144**: 2992-3000 [PMID: 30536395 DOI: 10.1002/ijc.32055]

3　**Kaminski MF**, Thomas-Gibson S, Bugajski M, Bretthauer M, Rees CJ, Dekker E, Hoff G, Jover R, Suchanek S, Ferlitsch M, Anderson J, Roesch T, Hultcranz R, Racz I, Kuipers EJ, Garborg K, East JE, Rupinski M, Seip B, Bennett C, Senore C, Minozzi S, Bisschops R, Domagk D, Valori R, Spada C, Hassan C, Dinis-Ribeiro M, Rutter MD. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *Endoscopy* 2017; **49**: 378-397 [PMID: 28268235 DOI: 10.1055/s-0043-103411]

4　**Rex DK**, Schoenfeld PS, Cohen J, Pike IM, Adler DG, Fennerty MB, Lieb JG 2nd, Park WG, Rizk MK, Sawhney MS, Shaheen NJ, Wani S, Weinberg DS. Quality indicators for colonoscopy. *Gastrointest Endosc* 2015; **81**: 31-53 [PMID: 25480100 DOI: 10.1016/j.gie.2014.07.058]

5　**Corley DA**, Levin TR, Doubeni CA. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014; **370**: 2541 [PMID: 24963577 DOI: 10.1056/NEJMc1405329]

6　**van Rijn JC**, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006; **101**: 343-350 [PMID: 16454841 DOI: 10.1111/j.1572-0241.2006.00390.x]

7　**Vleugels JLA**, Hazewinkel Y, Fockens P, Dekker E. Natural history of diminutive and small colorectal polyps: a systematic literature review. *Gastrointest Endosc* 2017; **85**: 1169-1176. e1 [PMID: 28024986 DOI: 10.1016/j.gie.2016.12.014]

8　**Barua I**, Vinsard DG, Jodal HC, Løberg M, Kalager M, Holme Ø, Misawa M, Bretthauer M, Mori Y. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* 2021; **53**: 277-284 [PMID: 32557490 DOI: 10.1055/a-1201-7165]

9　**Hassan C**, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, Antonelli G, Yu H, Areia M, Dinis-Ribeiro M, Bhandari P, Sharma P, Rex DK, Rösch T, Wallace M, Repici A. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021; **93**: 77-85. e6 [PMID: 32598963 DOI: 10.1016/j.gie.2020.06.059]

10　**Jin EH**, Lee D, Bae JH, Kang HY, Kwak MS, Seo JY, Yang JI, Yang SY, Lim SH, Yim JY, Lim JH, Chung GE, Chung SJ, Choi JM, Han YM, Kang SJ, Lee J, Chan Kim H, Kim JS. Improved Accuracy in Optical Diagnosis of Colorectal Polyps Using Convolutional Neural Networks with Visual Explanations. *Gastroenterology* 2020; **158**: 2169-2179. e8 [PMID: 32119927 DOI: 10.1053/j.gastro.2020.02.036]

11　**Renner J**, Phlipsen H, Haller B, Navarro-Avila F, Saint-Hill-Febles Y, Mateus D, Ponchon T, Poszler A, Abdelhafez M, Schmid RM, von Delius S, Klare P. Optical classification of neoplastic colorectal polyps - a computer-assisted approach (the COACH study). *Scand J Gastroenterol* 2018; **53**: 1100-1106 [PMID: 30270677 DOI: 10.1080/00365521.2018.1501092]

12　**Kudo SE**, Mori Y, Misawa M, Takeda K, Kudo T, Itoh H, Oda M, Mori K. Artificial intelligence and colonoscopy: Current status and future perspectives. *Dig Endosc* 2019; **31**: 363-371 [PMID: 30624835 DOI: 10.1111/den.13340]

13　**Tang A**, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, Chepelev L, Cairns R, Mitchell JR, Cicero MD, Poudrette MG, Jaremko JL, Reinhold C, Gallix B, Gray B, Geis R; Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Can Assoc Radiol J* 2018; **69**: 120-135 [PMID: 29655580 DOI: 10.1016/j.carj.2018.02.002]

14　**Shalev-Shwartz S**, Ben-David S.　Understanding Machine Learning: from theory to algorithms. New York: Cambridge University Press, 2014

15　**Wittenberg T**, Raithel M. Artificial Intelligence-Based Polyp Detection in Colonoscopy: Where Have We Been, Where Do We Stand, and Where Are We Headed? *Visc Med* 2020; **36**: 428-438 [PMID: 33447598 DOI: 10.1159/000512438]

16　Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003; **58**: S3-43 [PMID: 14652541 DOI: 10.1016/s0016-5107(03)02159-x]

17　**Kudo S**, Tamura S, Nakajima T, Yamano H, Kusaka H, Watanabe H. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointest Endosc* 1996; **44**: 8-14 [PMID: 8836710 DOI: 10.1016/s0016-5107(96)70222-5]

18　**Leufkens AM**, van Oijen MG, Vleggaar FP, Siersema PD. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. *Endoscopy* 2012; **44**: 470-475 [PMID: 22441756 DOI: 10.1055/s-0031-1291666]

19　**Rex DK**, Cutler CS, Lemmel GT, Rahmani EY, Clark DW, Helper DJ, Lehman GA, Mark DG. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* 1997; **112**: 24-28 [PMID: 8978338 DOI: 10.1016/s0016-5085(97)70214-2]

20　**Zhao S**, Wang S, Pan P, Xia T, Chang X, Yang X, Guo L, Meng Q, Yang F, Qian W, Xu Z, Wang Y, Wang Z, Gu L, Wang R, Jia F, Yao J, Li Z, Bai Y. Magnitude, Risk Factors, and Factors Associated

With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. *Gastroenterology* 2019; **156**: 1661-1674. e11 [PMID: 30738046 DOI: 10.1053/j.gastro.2019.01.260]

21   **Anderson R**, Burr NE, Valori R. Causes of Post-Colonoscopy Colorectal Cancers Based on World Endoscopy Organization System of Analysis. *Gastroenterology* 2020; **158**: 1287-1299. e2 [PMID: 31926170 DOI: 10.1053/j.gastro.2019.12.031]

22   **Buchner AM**, Shahid MW, Heckman MG, Diehl NN, McNeil RB, Cleveland P, Gill KR, Schore A, Ghabril M, Raimondo M, Gross SA, Wallace MB. Trainee participation is associated with increased small adenoma detection. *Gastrointest Endosc* 2011; **73**: 1223-1231 [PMID: 21481861 DOI: 10.1016/j.gie.2011.01.060]

23   **Lee CK**, Park DI, Lee SH, Hwangbo Y, Eun CS, Han DS, Cha JM, Lee BI, Shin JE. 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**: 1094-1102 [PMID: 21889137 DOI: 10.1016/j.gie.2011.06.033]

24   **Wang P**, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, Liu P, Li L, Song Y, Zhang D, Li Y, Xu G, Tu M, Liu X. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; **68**: 1813-1819 [PMID: 30814121 DOI: 10.1136/gutjnl-2018-317500]

25   **Repici A**, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, Ferrara E, Spadaccini M, Alkandari A, Fugazza A, Anderloni A, Galtieri PA, Pellegatta G, Carrara S, Di Leo M, Craviotto V, Lamonaca L, Lorenzetti R, Andrealli A, Antonelli G, Wallace M, Sharma P, Rosch T, Hassan C. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* 2020; **159**: 512-520. e7 [PMID: 32371116 DOI: 10.1053/j.gastro.2020.04.062]

26   **Gong D**, Wu L, Zhang J, Mu G, Shen L, Liu J, Wang Z, Zhou W, An P, Huang X, Jiang X, Li Y, Wan X, Hu S, Chen Y, Hu X, Xu Y, Zhu X, Li S, Yao L, He X, Chen D, Huang L, Wei X, Wang X, Yu H. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020; **5**: 352-361 [PMID: 31981518 DOI: 10.1016/S2468-1253(19)30413-3]

27   **Wang P**, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, Lei L, Li L, Guo Z, Lei S, Xiong F, Wang H, Song Y, Pan Y, Zhou G. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol* 2020; **5**: 343-351 [PMID: 31981517 DOI: 10.1016/S2468-1253(19)30411-X]

28   **Su JR**, Li Z, Shao XJ, Ji CR, Ji R, Zhou RC, Li GC, Liu GQ, He YS, Zuo XL, Li YQ. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointest Endosc* 2020; **91**: 415-424. e4 [PMID: 31454493 DOI: 10.1016/j.gie.2019.08.026]

29   **Liu WN**, Zhang YY, Bian XQ, Wang LJ, Yang Q, Zhang XD, Huang J. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol* 2020; **26**: 13-19 [PMID: 31898644 DOI: 10.4103/sjg.SJG_377_19]

30   **Kahi CJ**, Vemulapalli KC, Johnson CS, Rex DK. Improving measurement of the adenoma detection rate and adenoma per colonoscopy quality metric: the Indiana University experience. *Gastrointest Endosc* 2014; **79**: 448-454 [PMID: 24246797 DOI: 10.1016/j.gie.2013.10.013]

31   **Hilsden RJ**, Bridges R, Dube C, McGregor SE, Naugler C, Rose SM, Rostom A, Heitman SJ. Defining Benchmarks for Adenoma Detection Rate and Adenomas Per Colonoscopy in Patients Undergoing Colonoscopy Due to a Positive Fecal Immunochemical Test. *Am J Gastroenterol* 2016; **111**: 1743-1749 [PMID: 27725649 DOI: 10.1038/ajg.2016.449]

32   **Wang P**, Liu P, Glissen Brown JR, Berzin TM, Zhou G, Lei S, Liu X, Li L, Xiao X. Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study. *Gastroenterology* 2020; **159**: 1252-1261. e5 [PMID: 32562721 DOI: 10.1053/j.gastro.2020.06.023]

33   **Sano Y**, Ikematsu H, Fu KI, Emura F, Katagiri A, Horimatsu T, Kaneko K, Soetikno R, Yoshida S. Meshed capillary vessels by use of narrow-band imaging for differential diagnosis of small colorectal polyps. *Gastrointest Endosc* 2009; **69**: 278-283 [PMID: 18951131 DOI: 10.1016/j.gie.2008.04.066]

34   **Rex DK**. Narrow-band imaging without optical magnification for histologic analysis of colorectal polyps. *Gastroenterology* 2009; **136**: 1174-1181 [PMID: 19187781 DOI: 10.1053/j.gastro.2008.12.009]

35   **Sano Y**, Tanaka S, Kudo SE, Saito S, Matsuda T, Wada Y, Fujii T, Ikematsu H, Uraoka T, Kobayashi N, Nakamura H, Hotta K, Horimatsu T, Sakamoto N, Fu KI, Tsuruta O, Kawano H, Kashida H, Takeuchi Y, Machida H, Kusaka T, Yoshida N, Hirata I, Terai T, Yamano HO, Kaneko K, Nakajima T, Sakamoto T, Yamaguchi Y, Tamai N, Nakano N, Hayashi N, Oka S, Iwatate M, Ishikawa H, Murakami Y, Yoshida S, Saito Y. Narrow-band imaging (NBI) magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI Expert Team. *Dig Endosc* 2016; **28**: 526-533 [PMID: 26927367 DOI: 10.1111/den.12644]

36   **Hewett DG**, Kaltenbach T, Sano Y, Tanaka S, Saunders BP, Ponchon T, Soetikno R, Rex DK. Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. *Gastroenterology* 2012; **143**: 599-607. e1 [PMID: 22609383 DOI: 10.1053/j.gastro.2012.05.006]

37   **Repici A**, Ciscato C, Correale L, Bisschops R, Bhandari P, Dekker E, Pech O, Radaelli F, Hassan C. Narrow-band Imaging International Colorectal Endoscopic Classification to predict polyp histology: REDEFINE study (with videos). *Gastrointest Endosc* 2016; **84**: 479-486. e3 [PMID: 26928372 DOI:

10.1016/j.gie.2016.02.020]

38   **Komeda Y**, Kashida H, Sakurai T, Asakuma Y, Tribonias G, Nagai T, Kono M, Minaga K, Takenaka M, Arizumi T, Hagiwara S, Matsui S, Watanabe T, Nishida N, Chikugo T, Chiba Y, Kudo M. Magnifying Narrow Band Imaging (NBI) for the Diagnosis of Localized Colorectal Lesions Using the Japan NBI Expert Team (JNET) Classification. *Oncology* 2017; **93** Suppl 1: 49-54 [PMID: 29258091 DOI: 10.1159/000481230]

39   **Kobayashi S**, Yamada M, Takamaru H, Sakamoto T, Matsuda T, Sekine S, Igarashi Y, Saito Y. Diagnostic yield of the Japan NBI Expert Team (JNET) classification for endoscopic diagnosis of superficial colorectal neoplasms in a large-scale clinical practice database. *United European Gastroenterol J* 2019; **7**: 914-923 [PMID: 31428416 DOI: 10.1177/2050640619845987]

40   **Kandel P**, Wallace MB. Should We Resect and Discard Low Risk Diminutive Colon Polyps. *Clin Endosc* 2019; **52**: 239-246 [PMID: 30661337 DOI: 10.5946/ce.2018.136]

41   **Neumann H**, Neumann Sen H, Vieth M, Bisschops R, Thieringer F, Rahman KF, Gamstätter T, Tontini GE, Galle PR. Leaving colorectal polyps in place can be achieved with high accuracy using blue light imaging (BLI). *United European Gastroenterol J* 2018; **6**: 1099-1105 [PMID: 30228899 DOI: 10.1177/2050640618769731]

42   **von Renteln D**, Kaltenbach T, Rastogi A, Anderson JC, Rösch T, Soetikno R, Pohl H. Simplifying Resect and Discard Strategies for Real-Time Assessment of Diminutive Colorectal Polyps. *Clin Gastroenterol Hepatol* 2018; **16**: 706-714 [PMID: 29174789 DOI: 10.1016/j.cgh.2017.11.036]

43   **Rex DK**, Kahi C, O'Brien M, Levin TR, Pohl H, Rastogi A, Burgart L, Imperiale T, Ladabaum U, Cohen J, Lieberman DA. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2011; **73**: 419-422 [PMID: 21353837 DOI: 10.1016/j.gie.2011.01.023]

44   **Vinsard DG**, Mori Y, Misawa M, Kudo SE, Rastogi A, Bagci U, Rex DK, Wallace MB. Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointest Endosc* 2019; **90**: 55-63 [PMID: 30926431 DOI: 10.1016/j.gie.2019.03.019]

45   **ASGE Technology Committee**. , Abu Dayyeh BK, Thosani N, Konda V, Wallace MB, Rex DK, Chauhan SS, Hwang JH, Komanduri S, Manfredi M, Maple JT, Murad FM, Siddiqui UD, Banerjee S. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015; **81**: 502.e1-502. e16 [PMID: 25597420 DOI: 10.1016/j.gie.2014.12.022]

46   **Song EM**, Park B, Ha CA, Hwang SW, Park SH, Yang DH, Ye BD, Myung SJ, Yang SK, Kim N, Byeon JS. Endoscopic diagnosis and treatment planning for colorectal polyps using a deep-learning model. *Sci Rep* 2020; **10**: 30 [PMID: 31913337 DOI: 10.1038/s41598-019-56697-0]

47   **Lui TKL**, Guo CG, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointest Endosc* 2020; **92**: 11-22. e6 [PMID: 32119938 DOI: 10.1016/j.gie.2020.02.033]

48   **Mori Y**, Kudo SE, Chiu PW, Singh R, Misawa M, Wakamura K, Kudo T, Hayashi T, Katagiri A, Miyachi H, Ishida F, Maeda Y, Inoue H, Nimura Y, Oda M, Mori K. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study. *Endoscopy* 2016; **48**: 1110-1118 [PMID: 27494455 DOI: 10.1055/s-0042-113609]

49   **Kominami Y**, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, Tamaki T, Koide T, Kaneda K, Chayama K. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016; **83**: 643-649 [PMID: 26264431 DOI: 10.1016/j.gie.2015.08.004]

50   **Byrne MF**, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, Iqbal N, Chandelier F, Rex DK. 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 [PMID: 29066576 DOI: 10.1136/gutjnl-2017-314547]

51   **Gross S**, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, Schirin-Sokhan R, Hecker H, Aach T, Tischendorf JJ. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. *Gastrointest Endosc* 2011; **74**: 1354-1359 [PMID: 22000791 DOI: 10.1016/j.gie.2011.08.001]

52   **Lui TKL**, To WP, Ko KK. Superiority of the artificial intelligence image classifier for histological prediction of diminutive colorectal polyps based on non-magnifying endoscopic images. *Hong Kong Med J* 2019; **25**: 31

53   **Chen PJ**, Lin MC, Lai MJ, Lin JC, Lu HH, Tseng VS. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology* 2018; **154**: 568-575 [PMID: 29042219 DOI: 10.1053/j.gastro.2017.10.010]

54   **Mori Y**, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, Ohtsuka K, Urushibara F, Kataoka S, Ogawa Y, Maeda Y, Takeda K, Nakamura H, Ichimasa K, Kudo T, Hayashi T, Wakamura K, Ishida F, Inoue H, Itoh H, Oda M, Mori K. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Ann Intern Med* 2018; **169**: 357-366 [PMID: 30105375 DOI: 10.7326/M18-0249]

55   **Kudo SE**, Misawa M, Mori Y, Hotta K, Ohtsuka K, Ikematsu H, Saito Y, Takeda K, Nakamura H, Ichimasa K, Ishigaki T, Toyoshima N, Kudo T, Hayashi T, Wakamura K, Baba T, Ishida F, Inoue H, Itoh H, Oda M, Mori K. Artificial Intelligence-assisted System Improves Endoscopic Identification of

Colorectal Neoplasms. *Clin Gastroenterol Hepatol* 2020; **18**: 1874-1881. e2 [PMID: 31525512 DOI: 10.1016/j.cgh.2019.09.009]

56   **Horiuchi H**, Tamai N, Kamba S, Inomata H, Ohya TR, Sumiyama K. Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software. *Scand J Gastroenterol* 2019; **54**: 800-805 [PMID: 31195905 DOI: 10.1080/00365521.2019.1627407]

57   **Mathews SC**, Zhao N, Holub JL, Lieberman D. Improvement in colonoscopy quality metrics in clinical practice from 2000 to 2014. *Gastrointest Endosc* 2019; **90**: 651-655. e3 [PMID: 31207221 DOI: 10.1016/j.gie.2019.06.004]

58   **Clark BT**, Rustagi T, Laine L. What level of bowel prep quality requires early repeat colonoscopy: systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate. *Am J Gastroenterol* 2014; **109**: 1714-23; quiz 1724 [PMID: 25135006 DOI: 10.1038/ajg.2014.232]

59   **de Jonge V**, Sint Nicolaas J, Cahen DL, Moolenaar W, Ouwendijk RJ, Tang TJ, van Tilburg AJ, Kuipers EJ, van Leerdam ME; SCoPE Consortium. Quality evaluation of colonoscopy reporting and colonoscopy performance in daily clinical practice. *Gastrointest Endosc* 2012; **75**: 98-106 [PMID: 21907986 DOI: 10.1016/j.gie.2011.06.032]

60   **Leyden JE**, Doherty GA, Hanley A, McNamara DA, Shields C, Leader M, Murray FE, Patchett SE, Harewood GC. Quality of colonoscopy performance among gastroenterology and surgical trainees: a need for common training standards for all trainees? *Endoscopy* 2011; **43**: 935-940 [PMID: 21997723 DOI: 10.1055/s-0030-1256633]

61   **Coe SG**, Panjala C, Heckman MG, Patel M, Qumseya BJ, Wang YR, Dalton B, Tran P, Palmer W, Diehl N, Wallace MB, Raimondo M. Quality in colonoscopy reporting: an assessment of compliance and performance improvement. *Dig Liver Dis* 2012; **44**: 660-664 [PMID: 22579446 DOI: 10.1016/j.dld.2012.03.022]

62   **Zhou J**, Wu L, Wan X, Shen L, Liu J, Zhang J, Jiang X, Wang Z, Yu S, Kang J, Li M, Hu S, Hu X, Gong D, Chen D, Yao L, Zhu Y, Yu H. A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020; **91**: 428-435. e2 [PMID: 31783029 DOI: 10.1016/j.gie.2019.11.026]

63   **Hassan C**, Wallace MB, Sharma P, Maselli R, Craviotto V, Spadaccini M, Repici A. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. *Gut* 2020; **69**: 799-800 [PMID: 31615835 DOI: 10.1136/gutjnl-2019-319914]

64   **Day LW**, Bhuket T, Inadomi JM, Yee HF. Diversity of endoscopy center operations and practice variation across California's safety-net hospital system: a statewide survey. *BMC Res Notes* 2013; **6**: 233 [PMID: 23767938 DOI: 10.1186/1756-0500-6-233]

65   **Weigt J**, Repici A, Antonelli G, Afifi A, Kliegis L, Correale L, Hassan C, Neumann H. Performance of a new integrated CADe/CADx system for detection and characterization of colorectal neoplasia. *Endoscopy* 2021 [PMID: 33494106 DOI: 10.1055/a-1372-0419]

66   **Kelly CJ**, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; **17**: 195 [PMID: 31665002 DOI: 10.1186/s12916-019-1426-2]

67   **Maeda Y**, Kudo SE, Mori Y, Misawa M, Ogata N, Sasanuma S, Wakamura K, Oda M, Mori K, Ohtsuka K. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019; **89**: 408-415 [PMID: 30268542 DOI: 10.1016/j.gie.2018.09.024]

68   **Takenaka K**, Ohtsuka K, Fujii T, Negi M, Suzuki K, Shimizu H, Oshima S, Akiyama S, Motobayashi M, Nagahori M, Saito E, Matsuoka K, Watanabe M. Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis. *Gastroenterology* 2020; **158**: 2150-2157 [PMID: 32060000 DOI: 10.1053/j.gastro.2020.02.012]

69   **Nagendran M**, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, Topol EJ, Ioannidis JPA, Collins GS, Maruthappu M. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020; **368**: m689 [PMID: 32213531 DOI: 10.1136/bmj.m689]

70   **Collins GS**, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019; **393**: 1577-1579 [PMID: 31007185 DOI: 10.1016/S0140-6736(19)30037-6]

71   **CONSORT-AI and SPIRIT-AI Steering Group**. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med* 2019; **25**: 1467-1468 [PMID: 31551578 DOI: 10.1038/s41591-019-0603-3]

72   **Rivera SC**, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ* 2020; **370**: m3210 [PMID: 32907797 DOI: 10.1136/bmj.m3210]

73   **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 [PMID: 33179322 DOI: 10.1111/jgh.15339]

74   **Park SH**, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology* 2018; **286**: 800-809 [PMID: 29309734 DOI: 10.1148/radiol.2017171920]

75   **Ang TL**, Carneiro G. Artificial intelligence in gastrointestinal endoscopy. *J Gastroenterol Hepatol* 2021; **36**: 5-6 [PMID: 33448513 DOI: 10.1111/jgh.15344]

76   **Wadhwa V**, Alagappan M, Gonzalez A, Gupta K, Brown JRG, Cohen J, Sawhney M, Pleskow D, Berzin TM. Physician sentiment toward artificial intelligence (AI) in colonoscopic practice: a survey of US gastroenterologists. *Endosc Int Open* 2020; **8**: E1379-E1384 [PMID: 33015341 DOI: 10.1055/a-1223-1926]

77   **Habli I**, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. *Bull World Health Organ* 2020; **98**: 251-256 [PMID: 32284648 DOI: 10.2471/BLT.19.237487]

78   **Wiens J**, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, Jung K, Heller K, Kale D, Saeed M, Ossorio PN, Thadaney-Israni S, Goldenberg A. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019; **25**: 1337-1340 [PMID: 31427808 DOI: 10.1038/s41591-019-0548-6]

79   **Singh K**, Al Khoury A, Kurti Z, Gonczi L, Reinglas J, Verdon C, Kohen R, Bessissow T, Afif W, Wild G, Seidman E, Bitton A, Lakatos PL. High Adherence to Surveillance Guidelines in Inflammatory Bowel Disease Patients Results in Low Colorectal Cancer and Dysplasia Rates, While Rates of Dysplasia are Low Before the Suggested Onset of Surveillance. *J Crohns Colitis* 2019; **13**: 1343-1350 [PMID: 30918959 DOI: 10.1093/ecco-jcc/jjz066]

80   **Watanabe T**, Ajioka Y, Mitsuyama K, Watanabe K, Hanai H, Nakase H, Kunisaki R, Matsuda K, Iwakiri R, Hida N, Tanaka S, Takeuchi Y, Ohtsuka K, Murakami K, Kobayashi K, Iwao Y, Nagahori M, Iizuka B, Hata K, Igarashi M, Hirata I, Kudo SE, Matsumoto T, Ueno F, Watanabe G, Ikegami M, Ito Y, Oba K, Inoue E, Tomotsugu N, Takebayashi T, Sugihara K, Suzuki Y, Watanabe M, Hibi T. Comparison of Targeted vs Random Biopsies for Surveillance of Ulcerative Colitis-Associated Colorectal Cancer. *Gastroenterology* 2016; **151**: 1122-1130 [PMID: 27523980 DOI: 10.1053/j.gastro.2016.08.002]

81   **Allende D**, Elmessiry M, Hao W, DaSilva G, Wexner SD, Bejarano P, Berho M, Al-Qadasi M. Inter-observer and intra-observer variability in the diagnosis of dysplasia in patients with inflammatory bowel disease: correlation of pathological and endoscopic findings. *Colorectal Dis* 2014; **16**: 710-8; discussion 718 [PMID: 24836541 DOI: 10.1111/codi.12667]

82   **Beaugerie L**, Itzkowitz SH. Cancers complicating inflammatory bowel disease. *N Engl J Med* 2015; **372**: 1441-1452 [PMID: 25853748 DOI: 10.1056/NEJMra1403718]

83   **Lai EJ**, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc* 2009; **69**: 620-625 [PMID: 19136102 DOI: 10.1016/j.gie.2008.05.057]

84   **Taveira F**, Hassan C, Kaminski MF, Ponchon T, Benamouzig R, Bugajski M, de Castelbajac F, Cesaro P, Chergui H, Goran L, Minelli Grazioli L, Janičko M, Januszewicz W, Lamonaca L, Lenz J, Negreanu L, Repici A, Spada C, Spadaccini M, State M, Szlak J, Veseliny E, Dinis-Ribeiro M, Areia M. The Colon Endoscopic Bubble Scale (CEBuS): a two-phase evaluation study. *Endoscopy* 2020 [PMID: 33285583 DOI: 10.1055/a-1331-4325]