Current and future implications of artificial intelligence in colonoscopy

Giulio Antonelli[a,b], Tommy Rizkala[c], Federico Iacopini[a], Cesare Hassan[c,d]

Ospedale dei Castelli Hospital, Ariccia, Rome; Sapienza University of Rome; Humanitas University, Rozzano; IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy

Abstract
Gastrointestinal endoscopy has proved to be a perfect context for the development of artificial intelligence (AI) systems that can aid endoscopists in many tasks of their daily activities. Lesion detection (computer-aided detection, CADe) and lesion characterization (computer-aided characterization, CADx) during colonoscopy are the clinical applications of AI in gastroenterology for which by far the most evidence has been published. Indeed, they are the only applications for which more than one system has been developed by different companies, is currently available on the market, and may be used in clinical practice. Both CADe and CADx, alongside hopes and hypes, come with potential drawbacks, limitations and dangers that must be known, studied and researched as much as the optimal uses of these machines, aiming to stay one step ahead of the possible misuse of what will always be an aid to the clinician and never a substitute. An AI revolution in colonoscopy is on the way, but the potential uses are infinite and only a fraction of them have currently been studied. Future applications can be designed to ensure all aspects of colonoscopy quality parameters and truly deliver a standardization of practice, regardless of the setting in which the procedure is performed. In this review, we cover the available clinical evidence on AI applications in colonoscopy and offer an overview of future directions.

Keywords Artificial intelligence, machine learning, colonoscopy, adenoma detection rate, polyp detection

Ann Gastroenterol 2023; 36 (2): 114-122

Introduction

In recent years, there are few areas of medicine that have not been touched by the advent of artificial intelligence (AI)

[a]Gastroenterology and Digestive Endoscopy Unit, Ospedale dei Castelli Hospital, Ariccia, Rome (Giulio Antonelli, Federico Iacopini); [b]Department of Anatomical, Histological, Forensic Medicine and Orthopedics Sciences, "Sapienza" University of Rome (Giulio Antonelli); [c]Department of Biomedical Sciences, Humanitas University, Rozzano, Milan (Tommy Rizkala, Cesare Hassan); [d]IRCCS Humanitas Research Hospital, Rozzano, Milan (Cesare Hassan), Italy

Conflict of Interest: None

Correspondence to: Giulio Antonelli, MD, Gastroenterology and Digestive Endoscopy Unit, Ospedale dei Castelli Hospital, Via Nettunense Km 11.5, Ariccia, Rome, Italy, e-mail: giulio.antonelli@gmail.com

systems, especially since the development of convolutional neural networks, which have empowered machines to acquire certain cognitive abilities that can be used in multiple medical fields [1,2]. Indeed, we are on the verge of a revolution for many, if not all, scenarios of modern medicine [3]. Gastrointestinal endoscopy has proved to be a perfect context for the development of AI systems that can aid endoscopists in many parts of their daily activities [3]. Indeed, AI seems to be the ideal tool to improve quality in virtually every subdomain of endoscopy [4], delivering a standardization of practice by ensuring a minimum standard under which it is virtually impossible to go.

Lesion detection (computer-aided detection, CADe) and lesion characterization (computer-aided characterization, CADx) during colonoscopy are the clinical applications of AI in gastroenterology for which by far the most evidence has been published [5] (Tables 1, 2). Indeed, they are the only applications for which more than one system has been developed by different companies, is currently available on the market, and may be used in clinical practice. This is probably due to the fact that a large body of data (i.e., images and videos) is needed to train and test reliable AI systems [6]. Especially in western countries, where colorectal cancer screening programs are now widely implemented [7], a huge number of colonoscopies are performed each day, and the overall

Table 1 Published randomized controlled trials on computer-aided detection (CAD)

Author [Ref.]	Publication year	Country	Single/Multi-center	CAD system
Repici *et al* [16]	2020	Italy	Multicenter	GI genius® (Medtronic)
Repici *et al* [17]	2021	Italy, Switzerland	Multicenter	GI genius® (Medtronic)
Gong *et al* [39]	2020	China	Single center	Endoangel
Liu *et al* [47]	2020	China	Single center	Henan Tongyu
Liu *et al* [48]	2021	China	Single center	EndoScreener
Su *et al* [40]	2020	China	Single center	Self-developed
Wang *et al* [49]	2019	China	Single center	EndoScreener
Wang *et al* [19]	2020	China	Single center	EndoScreener
Xu *et al* [50]	2021	China	Multicenter	Self-developed
Yao *et al* [41]	2021	China	Single center	Endoangel
Glissen Brown *et al* [51]	2021	US	Multicenter	EndoScreener
Kamba *et al* [52]	2021	Japan	Multicenter	YOLOv3
Wang *et al* [53]	2021	China	Single Center	EndoScreener
Wallace *et al* [18]	2022	US, UK, Italy	Multicenter	GI genius® (Medtronic)
Rondonotti *et al* [36]	2022	Italy	Multicenter	CAD-EYE® (Fujifilm)
Shaukat *et al* [54]	2022	US	Multicenter	SKOUT
Mangas-Sanjuan *et al* [55]	2022	Spain	Multicenter	GI genius® (Medtronic)

Table 2 Clinical evidence for computer-aided characterization

Author [Ref.]	Publication year	Country	Single/Multi-center	CAD system
Mori *et al* [31]	2018	Japan	Single center	Endo Brain
Hassan *et al* [34]	2022	Italy	Single center	GI genius® (Medtronic)
Rondonotti *et al* [36]	2023	Italy	Multicenter	CAD-EYE® (Fujifilm)
Barua *et al* [10]	2021	Norway, Japan, UK	Multicenter	Endo Brain

CAD, computer-aided detection

prevalence of colorectal polyps is so high that it is feasible to collect a large number of "pathological" images or videos to train a system, while it is equally relatively easy to test a system, even in a real-life scenario.

Despite recent enormous improvements in colonoscopy technologies and techniques, this procedure is still hampered by a substantial rate of missed neoplasia, representing the major cause of interval cancer [8]. In addition, an extremely high variability in adenoma detection rate (ADR), the main key quality indicator in colonoscopy, has been extensively reported [9]. For these reasons, CADe systems have gained the most attention and are supported by an ever-growing body of evidence that consistently shows that their use leads to a significant increase in ADR, the main quality parameter in colonoscopy, which, in turn, is inversely associated with post colonoscopy colorectal cancer [5,10].

After a polyp has been detected, the natural workflow is either immediate or postponed polyp resection, using a series of endoscopic techniques that vary depending on polyp characteristics: namely size, location, morphology, and optical diagnosis [11]. Polyp characterization is critical in the choice of the optimal endoscopic treatment for each lesion, and

the availability of AI systems to aid the endoscopist in this effort can be considered ground-breaking. Small, diminutive (≤5 mm) colorectal polyps of the rectosigmoid tract (DRSPs), for example, have to be categorized as either adenomatous, harboring a malignant degeneration potential over the years, or non-adenomatous (hyperplastic), which can be safely left in place as they do not carry the genetic mutations and pathways that can degenerate over time, acquiring malignant characteristics [12]. Larger polyps have to be diagnosed optically to determine the best treatment strategy. It is now possible to distinguish between a superficial submucosal invasion, which is still amenable to endoscopic resection with advanced techniques such as endoscopic submucosal dissection, and a deep submucosal invasion that, until now, has been considered an absolute indication for surgery [13,14].

Both CADe and CADx, alongside hopes and hypes, come with potential drawbacks, limitations and dangers that must be known, studied and researched as much as the optimal uses of these machines, aiming to stay one step ahead of the possible misuse of what will always be an aid to the clinician and never a substitute. The AI revolution in colonoscopy is on the way, but the potential uses are infinite and only a fraction of them have

currently been studied. Future applications can be designed to ensure all aspects of colonoscopy quality parameters and truly deliver a standardization of practice, regardless of the setting in which the procedure is performed. In this review, we cover the available clinical evidence on AI applications in colonoscopy and offer an overview of future directions.

CADe

Lesion recognition failure (i.e., failing to notice a visible lesion on the endoscopy screen) is one of the main reasons for adenoma miss rate (AMR) [8], a relevant quality parameter that correlates with ADR and in turn with post-colonoscopy colorectal cancer. Indeed, many factors contribute to recognition failure, most of which are innate in the endoscopists' human nature, such as fatigue, distraction or impatience. In addition, many lesions are subtle in their appearance and can be spotted only by the trained eye in a perfectly cleansed bowel mucosa [15].

CADe of colorectal polyps (Fig. 1) seems the perfect answer to these issues, and indeed was the first task to be mastered by multiple AI systems that have been tested in multiple randomized controlled trials (RCTs) [5,16-18] (Table 1). This was possible after extensive evidence of standalone AI performance showed the feasibility and effectiveness of the task, consisting of the detection and segmentation of any visible lesion with a specific "flag", usually a box, that can be reinforced by an audio signal.

A recent meta-analysis to be published has summarizes the data of the 17 existing RCTs on different CADe systems, reporting on more than 16,000 patients. More than half of the included studies came from China, while only 5 RCTs came from the western setting: 3 from Italy and 1 each from the US and Spain.

When looking at ADR, authors of the meta-analysis found it to be significantly higher among patients undergoing colonoscopy with CADe compared to patients in the standard group (3077/6791, 45.3% vs. 2575/6796, 37.9%; relative risk 1.28, 95% confidence interval [CI] 1.17-1.40). When using a different sub-analysis, the consistency of this result was confirmed by looking only at the first colonoscopy of tandem trials, as well as by the similar results across studies of different magnitude. No publication bias was found. This meta-analysis

is also notable because it was the first to look at AMR in the 4 tandem trials published on CADe, showing that the AMR was indeed lower in the groups where the first colonoscopy was performed using CADe, as compared to the groups where standard colonoscopy was performed first.

Among tandem studies, it is worth noting the sub-analysis carried out by Wang *et al*, which investigated the difference in the miss rate between "visible" (i.e., exposed, but not recognized by the operating endoscopist) and "invisible" (i.e., not exposed by the endoscopist) polyps [19]. They observed that if the mucosa containing a polyp is actually exposed during the examination, CADe almost never misses its detection (AMR-visible in the CADe group: 1.59%; polyp miss rate-visible in the CADe group: 2.36%), highlighting once again the importance of mucosal exposure in neoplasia detection.

Interestingly, the benefit of CADe seemed higher among studies with a low mean ADR compared to studies with a higher mean ADR, both effects being nevertheless significant. Regarding serrated polyps, the authors reported a significant, albeit slight superiority in the number of serrated lesions detected per colonoscopy in the CADe group. It must be noted that serrated polyp detection was never a primary endpoint, and we believe that this should be prioritized in future studies.

Since CADe systems autonomously learn the salient features of an image without direct supervision by humans, their output is not always predictable, resulting in false-positive activations that can alert endoscopists to areas that would not normally have attracted their attention. False-positive activations may jeopardize the effectiveness of CADe systems if they result in too much background noise, unnecessary polypectomies and increased procedure time. In a recent study [20], the authors performed a *post hoc* analysis of an RCT on CADe, and measured false-positive burden and clinical relevance, classifying false positives into 2 broad categories: artefacts from bowel wall and artefacts from bowel content. The bowel wall was found to account for nearly 90% of false-positive activations (folds, ileocecal valve, diverticula, appendicular foramen, etc.). Overall, false positives were found to impact on less than 1% of total withdrawal time. Recently, another study from the same group compared false positives from 2 different systems, with similar results [21], suggesting that the difference in the perception phase between AI and human endoscopist minimizes the negative effect of false-positive activations. That is, the human brain of a trained endoscopist effectively
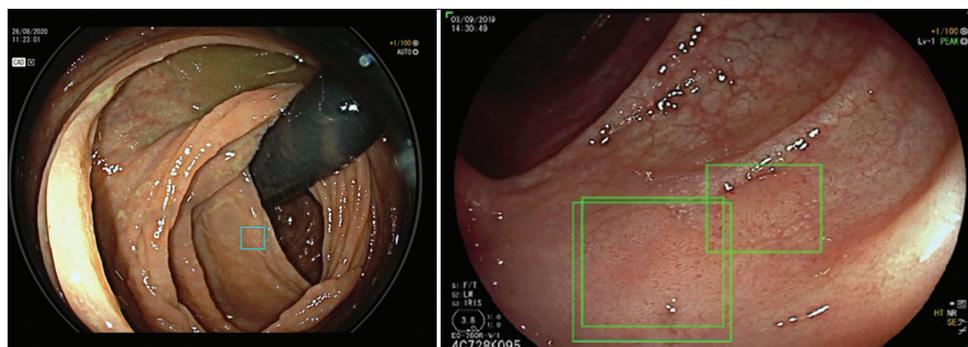

Figure 1 Examples of computer-aided detection

and quickly dismisses most false-positive activations, with no apparent negative effect on colonoscopy safety and duration.

CADx

Optical diagnosis of colorectal polyps has a critical role in determining the optimal treatment strategies for each patient. The appeal of CADx (Fig. 2) is manifold: first, AI autonomously learns features of interest that are completely different from the ones selected by the human mind, potentially increasing diagnostic accuracy; second, when applied to large polyps, it could estimate the risk of submucosal invasion better than existing classifications, known to have suboptimal performances even in expert hands [22]; third, it could indicate to non-expert endoscopists where to stop and refer the lesion to an expert center; fourth, when applied to diminutive polyps, it could efficiently permit the implementation of cost-saving strategies that up to now have failed to be widely accepted in the community; finally, it could have many applications in training and competence assessment [23].

As the foundation of training a highly performing AI system is based on the availability, quantity and quality of data, it is only natural that the first developments of CADx systems have concentrated on diminutive colorectal polyps, which represent more than 60% of all detected and resected colonic polyps [24]. In addition, the growing availability of different devices and technologies aimed at increasing ADR is dramatically increasing the number of diminutive polyps. To date, guidelines recommend resecting all detected polyps and sending them for histopathological examination [11]. Consequently, the burden of diminutive lesions on the total capacity of the entire clinical chain that starts with polyp detection is dramatic, from resection devices (forceps, snares, etc.) to vials, to the histopathological workload of technicians and physicians. These reasons have prompted the development and implementation of cost-saving strategies aimed at optical diagnosis of diminutive colorectal polyps.

These strategies were introduced by the American Society for Gastrointestinal Endoscopy (ASGE) back in 2011, with a landmark document called PIVI [25]. When applying optical diagnosis strategies, endoscopists or AI systems diagnose diminutive polyps during colonoscopy with high or low confidence. When the diagnosis of adenoma is made with high confidence, the polyps could be resected and discarded without histological evaluation (i.e., "resect-and-discard"). In addition, non-neoplastic lesions of the rectosigmoid tract could be left *in situ* and not resected, as they have no malignant potential (i.e., "leave-in-situ"). Indeed, cost effectiveness models have estimated a saving of up to 150 million dollars a year, just in Japan, for the implementation of the resect-and-discard strategy [26]. Similar figures have been demonstrated in Europe and in the US [26]. The implementation of these strategies has also been endorsed by several international societies.

The ASGE has set specific thresholds that, if met, permit the application of cost-saving strategies in clinical practice. Namely, to implement the resect-and-discard strategy, endoscopic technology (when used with high confidence) predicting the histology of polyps <5 mm in size, when combined with the histopathologic assessment of polyps >5 mm in size, should provide a >90% agreement in the assignment of post-polypectomy surveillance intervals when compared to decisions based on the histopathological assessment of all identified polyps. Furthermore, to implement the resect-and-discard strategy the technology should provide >90% negative predictive value (NPV) (when used with high confidence) for adenomatous histology [25].

Regrettably, although many years have passed since the proposal of these strategies, the uptake and implementation have been very slow and often overlooked, especially in the community setting [27,28]. Several underlying reasons have been found, the main ones being the fear of miscalculating endoscopic surveillance intervals, which ultimately depend on the number and histological characteristics of resected polyps, the fear of medico-legal implications, the lack of financial incentives to use optical diagnosis, which goes hand-in-hand with the possible loss of the incentives connected to polyp resection [29]. Apart from these, other common reasons are simply the lack of proper training in optical diagnosis and the lack of methods for competence assessment and maintenance. Indeed, the training in optical diagnosis acquisition and competence is not short, and even structured curricula proposed by scientific societies can be difficult to implement in everyday practice [23,30].

For all the above-mentioned reasons, optical diagnosis of diminutive polyps is the natural territory where the availability of a reliable AI system for CADx can be a game changer in clinical practice. All the stakeholders involved in colonoscopy
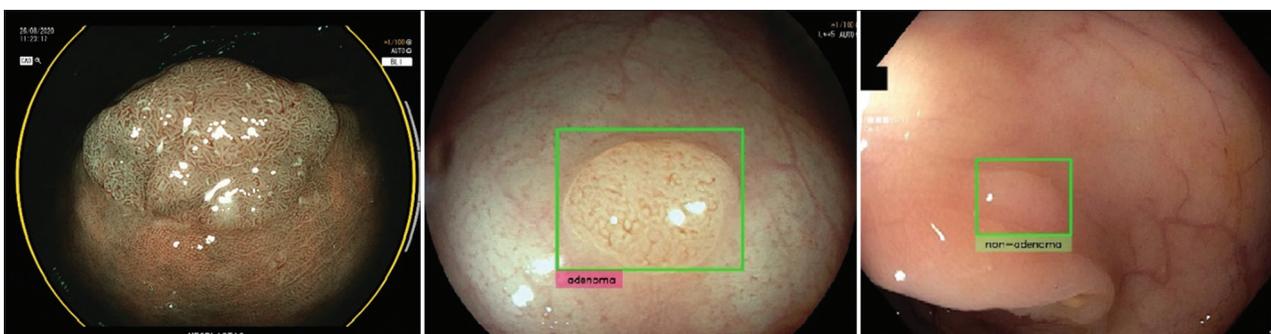
Figure 2 Examples of computer-aided characterization

could potentially benefit from the successful implementation of a CADx system. CADx can offer an unparalleled standardization of optical diagnosis performance based on the potential consistency of its prediction, which does not suffer from operator-related variables such as training, fatigue or distraction.

Clinical data

Recently, a small number of high-quality clinical trials have started to be published exploring the performance of different CADx systems (Table 2). The first clinical study was a landmark paper dating back to 2018 from Mori *et al* [31], who showed for the first time the real-time application of a CADx module in live colonoscopies. This study evaluated a CADx module paired to endocytoscopy (offering a x520 ultramagnification) plus virtual or dye-based chromoendoscopy in 791 patients undergoing colonoscopy. This system worked on still images for each polyp and the ability of CAD to differentiate neoplastic from non-neoplastic polyps was assessed, using histopathology as the gold standard.

When looking at CADx performance for diminutive polyps, CADx combined with virtual chromoendoscopy (narrow band imaging mode) showed a NPV of 96.5% (95%CI 92.1-98.9%) in identifying adenomas in the rectosigmoid tract, but a significantly lower performance (60%) if this analysis was extended to polyps proximal to the sigmoid colon.

This study was ground-breaking, because it showed for the first time the feasibility of a CADx-driven approach for diminutive rectosigmoid polyps and that a CADx system could reach recommended thresholds. It must be noted that the system tested in this study has not been approved for clinical practice in the west and is unlikely to be implemented in western everyday clinical practice because of the wide unavailability of endocytoscopy in this setting. Nevertheless, the pioneering aspect of this study cannot be sufficiently stressed, and may also understood from the fact that it took over 4 years for other clinical studies of other CADx systems to emerge.

Recently, 3 clinical trials have been published on 3 different CADx systems, starting to close the gap with standalone studies showing high performance [32-34]. The previously mentioned system was further studied in a multicenter international clinical trial [34] involving centers in Japan, Norway and the United Kingdom, and employing only non-expert (<1000 lifetime colonoscopies) endoscopists. This is the only study that selected sensitivity and specificity as primary outcomes; these, although not used as thresholds by the PIVI document, are less prone to bias given their independence from disease prevalence. In this study, no difference was found in sensitivity and specificity between human endoscopists and CADx, both showing a very high diagnostic performance. More specifically, sensitivity for the diagnosis of neoplastic polyps with standard visual inspection was 88.4% (95%CI 84.3-91.5%) compared with 90.4% (95%CI 86.8-93.1%) with CADx. Specificity was 83.1% (95%CI 79.2-86.4%) with standard visual inspection and 85.9% (95%CI 82.3-88.8%) with CADx. Most remarkably,

however, the authors showed how the proportion of polyp assessment with high confidence dramatically increased from 74.2% (95%CI 70.9-77.3%) with standard visual inspection to 92.6% (95%CI 90.6-94.3%) with CADx.

Although the lack of difference in diagnostic performance could be disappointing, a potential improvement in specificity was shown, although not significant, and more importantly still, the improvement in diagnostic confidence could potentially lead to a clinically significant reduction of unnecessary polyp resections.

The 2 remaining papers on CADx were from Italy, on 2 different systems, both approved and commercialized in Europe in clinical practice [33,34]. The first study was the CHANGE study [34], a prospective, single-arm study conducted in one open-access endoscopy center in Italy using the GI Genius CADx module (Medtronic, USA). The unique feature of this system is the capability of delivering a real-time diagnosis during white-light endoscopy, integrating the CADx system into the standard colonoscopy workflow. This study enrolled a total of 162 patients (46% male, mean age 66.6 years). A total of 544 polyps were detected and resected. Among these, 295 (54.2%) ≤5 mm rectosigmoid polyps were retrieved for histology, being adenomatous and non-adenomatous in 39/295 (13.2%) and 256/295 (86.8%) of the cases, respectively.

Of the 242 lesions predicted as non-adenomatous by CADx, 235 were confirmed as non-adenomatous at histology, corresponding to an adjusted NPV of 97.6% (95%CI 94.1-99.1%; P=0.002). More specifically, sensitivity, specificity, positive predictive value and accuracy for ≤5 mm rectosigmoid polyps were 82% (95%CI 66.5-92.5%), 93.2% (95%CI 89.4-96%), 65.3% (95%CI 50.4-78.3%), and 91.8% (95%CI 88-94.6%), respectively.

CADx predictions in the whole colon for diminutive adenomas, integrated with histological analysis for polyps >6 mm, resulted in a correct estimate of post-polypectomy endoscopic surveillance intervals of over 95%, according to both European and US guidelines [33,35]. Furthermore, the highly experienced endoscopists performing procedures and optical diagnosis with blue light imaging in the study achieved diagnostic performances that were comparable to those of CADx in white light.

This was the first time that high accuracy in the characterization of diminutive colorectal polyps was shown during real-time white light endoscopy. Diagnostic performances were sufficient to reach the thresholds set as mandatory for the clinical implementation of optical diagnosis. Currently, many endoscopists use optical diagnosis without knowing their own diagnostic performances. The availability of a second opinion that shows similar results and consistency can serve as a silent observer that can come into play whenever needed. In the study setting, the use of a leave-in-situ strategy would have resulted in an over 40% reduction in histopathological examinations, which would rise to over 80% it the resect-and-discard strategy was also applied. It is equally important to stress that for lesions proximal to the sigmoid tract this system showed a lower diagnostic accuracy, underlying the dynamic process of CADx development that is dependent on multiple training and retraining sessions.

The other Italian clinical study [36] evaluated the performance of the CAD-EYE CADx module (Fujifilm, Japan), which delivers an optical diagnosis dynamically during live colonoscopy every time virtual chromoendoscopy is activated during the procedure. This system also provides a heat map of the area most likely to harbor the provided diagnosis. This study employed a 3-step process: in the first step, the endoscopist alone characterized the polyp; in the second, the AI output was obtained and registered; in the third, the final diagnosis (adenoma vs. non-adenoma) provided by the endoscopist, combining the results of the first 2 steps, was reported. In the first and third step, the level of confidence was also expressed, while in the second, the AI diagnosis was collected only when an output was provided by the system and was considered stable during the observation time.

Looking at the primary outcome, the NPV of AI-assisted optical diagnosis (step 3) for adenomatous histology was 91.0% (95%CI 87.1-93.9%), while sensitivity, specificity and accuracy were 88.6% (95%CI 83.7-92.2%), 88.1% (95%CI 83.9-91.4%), and 88.4% (95%CI 85.3-90.9%), respectively.

Similar results, although slightly inferior to 90% regarding NPV, were found for AI alone. Agreement with the surveillance interval was over 90% for both European and the US Multisociety Task Force on Colorectal Cancer guidelines.

Interestingly, this study provided a differentiation between expert and non-expert endoscopist performance. While expert performance was stable during the study, reaching high accuracy consistently among the first and the last diminutive polyps evaluated, the performance of non-experts showed a statistically significant improvement in their own performance between the first evaluated polyps and the last evaluated polyps in the study. More specifically, the AI-assisted NPV of the last 50 DRSPs evaluated by non-experts met the PIVI threshold (NPV 95.2%, 95%CI 76.2-99.85%) and was similar to NPV calculated for the last 50 DRSPs evaluated by experts (NPV 93.9%, 95%CI 79.7-99.2%). Authors have speculated that this could be related to a positive interaction between a non-expert endoscopist and CADx, leading to a "learning effect" for optical diagnosis. This very attractive aspect of CADx should be further researched as a means to increase training opportunities.

One of the elements connecting all early CADx clinical trials is the drop in diagnostic accuracy when considering the proximal colon. This has not gone unnoticed and may be interpreted in different ways: first, it is logical and likely that in the development of the available systems the developers have prioritized the leave-in-situ strategy, namely recognizing adenomas as the first target of CADx development. Consequently, in regions where the prevalence changes (i.e., more adenomas in the right colon) the "weight" of a single wrong optical prediction can also greatly shift performance measures. Second, the superficial characteristics of proximal polyps could be harder to learn from and more variable than distal polyps, and thus need specific and more focused training. Third, the presence and increased prevalence of sessile serrated polyps in the proximal colon could have played an important role. The prevalence of rectosigmoidal serrated polyps is minimal in these studies, and a higher prevalence in the right colon might lead to diagnostic errors for CADx modules that are trained to deliver a dichotomous diagnosis. Indeed, not only have different studies considered sessile serrated lesions (SSLs) differently (neoplastic vs. non neoplastic), but no system has yet been developed to deliver a 3- or 4-way diagnosis, limiting their current use in the right colon. This point also highlights the importance, for endoscopists using a CAD system, of knowing the training data used to develop the system they are using.

Certainly, future studies and future systems will focus on this specific topic and the potential added value of CADx in the right colon and in the diagnosis of SSLs will be clarified. In addition, there is need for pragmatically designed and randomized trials, which up to now are completely lacking, to further prove the added value of CADx in colonoscopy.

Interaction between human endoscopist and CADx

The interaction between the AI machine and its user has been a subject of speculation since the first dawn of the AI concept many years ago. In our domain, and more broadly in medicine, one of the main concerns about AI implementation is the fears of negative interaction between AI and human. More specifically, when using an AI system, the human brain may fall victim to many biases, namely over-reliance (blindly accepting AI decisions as true without criticism) and under-reliance (blindly refusing AI decisions as false without considering its input). It would be redundant to state how perilous both instances may prove to be, for the doctor and the patient alike. Of course, the interaction between human and AI can also result in an improved level of performance generated from what has been called "hybrid intelligence", which in medicine can mean offering patients the highest quality of care currently available.

To explore the interaction between CADx and human endoscopists, a recent study [37] used a novel design, showing a team of both expert and non-expert endoscopists the same (reshuffled) set of colorectal diminutive lesions, first without the AI overlay and second with the AI overlay. The aim was to analyze how the decision of the endoscopist is influenced by the availability of the AI optical diagnosis output.

When looking at the results, the study found that indeed, as expected, endoscopists were influenced by the presence of the AI output. Interestingly, not only did using the AI improve the diagnostic performance overall, but the study also found that endoscopists, both experts and non-experts, were more likely to accept a correct AI opinion, even if it contradicted their own previous diagnosis, and to reject an incorrect AI opinion, if it contrasted with their diagnosis.

This study showed, for the first time in a fully scientific experimental setup, the positive interaction between a CADx module and human endoscopists. It has further explored the concepts of high and low confidence in an optical diagnosis, which was introduced many years ago but can draw new explanations from this work. Not only does this work prove that optical diagnosis is a dynamic process, but it also shows how human endoscopists are (consciously and/or unconsciously)

aware of the fluid nature of this decision making and are naturally inclined to accept changes in judgment when new elements of information are added to the equation. In this specific scenario, endoscopist were more inclined to stick with their own judgment, either when fully confident of their diagnosis, or when perceiving a low AI confidence in diagnosis. In contrast, when the endoscopist felt less confident and/or perceived a high confidence and consistency in the AI output, they were more likely to accept the AI output, even when it changed their original diagnosis.

Cost-effectiveness

The implementation of CAD in clinical practice has already begun. However, widespread implementation beyond the "usual" tertiary referral centers is a different matter and does not depend only on the credibility and solidity of scientific evidence, but also on the choices of the bodies that allocate resources. For this reason, cost-effectiveness studies, however theoretical, are vital for a 360-degree view of the implications of the use of AI. A preliminary study by Mori *et al* [26], focusing on the implementation of CADx for the leave-in-situ strategy, showed potentially dramatic cost reductions. Specifically, they estimated that the use of AI could save $119, $52, $34 and $125 per colonoscopy, and up to $149.2 million, $12.4 million, $1.1 million and $85.2 million from the annual reimbursement for colonoscopies conducted under public health insurances in Japan, England, Norway and the United States, respectively.

A recent cost-effectiveness analysis [38] was conducted on the implementation of CADe in screening colonoscopy in the US setting. The authors estimated a relative reduction in colorectal cancer incidence and mortality of 4.8% and 3.6%, respectively. The per-patient cost saving was estimated at $57 per individual. Projecting the results at a US population level, the implementation of CAD was estimated to prevent more than 7000 colorectal cancer cases and over 2000 colorectal cancer deaths, with a yearly saving of nearly $300 million.

Future implications

The potential applications of AI in colonoscopy go far beyond polyp detection and characterization. The most attractive development opportunities now fall to systems that can guarantee a standardization of the many quality parameters that have been defined for colonoscopy. Cecal intubation rate, mucosal exposure and inspection, and scope-slipping alerts have all been subjects of preliminary trials.

Gong *et al* developed an AI tool that notifies endoscopists of the withdrawal speed and blind spots. More specifically, the system was trained and then tested in real time to identify the cecum (and automatically record insertion and withdrawal time after cecum intubation) with an overall accuracy of an accuracy of 95% [39]. Su *et al* reported on a system that monitored the timing of the withdrawal phase, supervising withdrawal

stability and evaluating bowel preparation, in addition to having normal CADe functionalities [40]. A recent RCT showed that an AI system developed for real-time withdrawal speed monitoring applied to an existing CADe system improved ADR as compared to CADe or no AI alone [41].

The scoring of bowel preparation using AI was also recently investigated by 3 studies [42-44]. AI-based bowel preparation scoring is attractive: it is known that even the most studied and validated scales are prone to a huge interobserver agreement and are limited by the subjective scoring and estimation of the colonic regions of interest. AI could either give a real-time score, prompting cleaning and suctioning until a satisfactory level has been reached, or could provide a more homogeneous score that really reflects the prep in the whole colon. A real-time clinical trial involving 616 patients undergoing colonoscopy validated a previously trained system for the scoring of bowel prep using the Boston scale. This study found a significant inverse correlation between AI-based scoring and ADR, and showed potential for objective scoring of bowel preparation.

Real-time estimation of polyp size is also a task that has attracted much attention. Polyp size estimation has so far been a completely subjective task, since there is limited availability of measurement tools, essentially because of the challenge of deploying them through the endoscope service channel and since their disposal or re-use is unpractical and/ or expensive. As of today, the gold standard for *in vivo* polyp size estimation is to compare it side by side with an endoscopic tool of a known size, such as a forceps or a snare. However, in real-life clinical practice, it is very rare for this operation to be carried out systematically, because of time and cost issues. It is clear that an AI tool that could instantly and consistently provide an estimation of polyp size would be of great use in standardizing practice. Furthermore, as we have previously mentioned, the whole application of cost-saving strategies based on optical diagnosis relies first on determining that the polyp is diminutive (<5 mm). In addition, the correct assignment of post-polypectomy surveillance intervals is also partly dependent on polyp size, adding value to a tool that can homogenize polyps size estimation.

A recent proof-of-concept study showed for the first time the feasibility of an AI-based tool that uses laser technology to correctly estimate polyp size [45]. This system showed a higher accuracy for polyp size measurements than for visual size estimation (85.4% vs. 66.8%; P<0.001), using the polyp size measured after removal for reference. Of course, in these cases, the choice of the gold standard is harder than for characterization or detection, since there is no perfect methodology that can provide a reliable estimate of the *in vivo* size. The above prototype system, although extremely attractive, is limited by the need for a specially designed endoscope equipped with the laser system that is needed to function.

The last and probably most attractive quality measurement by AI is the estimation of mucosal exposure. Mucosal exposure is the most critical element in differentiating a high-quality colonoscopy, since only complete and accurate mucosal exposure can permit polyp detection, regardless of the use of a CADe system. Recently, a Chinese group developed an AI system for measuring "fold examination quality" during

withdrawal in colonoscopy [46]. They compared the system's evaluation of examination quality with the evaluation produced by expert endoscopists. Interestingly, the system showed a good correlation with experts, and the assistance in determining fold examination quality led to an increase in ADR.

It is conceivable that the combined use of CADe, CADx and quality assurance tools can increase the overall effectiveness of colonoscopy, although randomized trials are lacking in these areas. The goal of the combination of different CAD tools is the standardization of quality, providing reliable detection and characterization functions, and ultimately generating a semi-automatic report containing all the measured key performance indicators, a guarantee of mucosal exposure and all detected polyp characteristics. This can provide a guarantee for patients that they have received the highest quality of examination and could protect physicians from legal issues when they have provided the best level of care.

References

1. Gordon W. Moving past the promise of AI to real uses in health care delivery. NEJM Catal Innov Care Deliv 2022;**04**.

2. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med* 2019;**380**:1347-1358.

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

4. Messmann H, Bisschops R, Antonelli G, et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2022;**54**:1211-1231.

5. Hassan C, Spadaccini M, Iannone A, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021;**93**:77-85.e6.

6. van der Sommen F, de Groof J, Struyvenberg M, et al. Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. *Gut* 2020;**69**:2035-2045.

7. Basu P, Ponti A, Anttila A, et al. Status of implementation and organization of cancer screening in The European Union Member States-Summary results from the second European screening report. *Int J Cancer* 2018;**142**:44-56.

8. 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**:1661-1674.e11.

9. Zorzi M, Senore C, Da Re F, et al. Quality of colonoscopy in an organised colorectal cancer screening programme with immunochemical faecal occult blood test: the EQuIPE study (Evaluating Quality Indicators of the Performance of Endoscopy). *Gut* 2015;**64**:1389-1396.

10. Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* 2021;**53**:277-284.

11. Ferlitsch M, Moss A, Hassan C, et al. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2017;**49**:270-297.

12. Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000;**47**:251-255.

13. Sano Y, Tanaka S, Kudo SE, et al. Narrow-band imaging (NBI) magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI Expert Team. *Dig Endosc* 2016;**28**:526-533.

14. Pimentel-Nunes P, Libanio D, Bastiaansen BAJ, et al. Endoscopic submucosal dissection for superficial gastrointestinal lesions: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2022. *Endoscopy* 2022;**54**:591-622.

15. Radaelli F, Paggi S, Hassan C, et al. Split-dose preparation for colonoscopy increases adenoma detection rate: a randomised controlled trial in an organised screening programme. *Gut* 2017;**66**:270-277.

16. Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020;**159**:512-520.e7.

17. Repici A, Spadaccini M, Antonelli G, et al. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. *Gut* 2022;**71**:757-765.

18. Wallace MB, Sharma P, Bhandari P, et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. *Gastroenterology* 2022;**163**:295-304.e5.

19. Wang P, Liu P, Glissen Brown JR, et al. 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.

20. Hassan C, Badalamenti M, Maselli R, et al. Computer-aided detection-assisted colonoscopy: classification and relevance of false positives. *Gastrointest Endosc* 2020;**92**:900-904.e4.

21. Hassan C, Sharma P, Mori Y, et al. Comparative performance of artificial intelligence optical diagnosis systems for leaving in situ colorectal polyps. *Gastroenterology* 2022 Nov 1;S0016-5085(22)01199-4. [Online ahead of print]. doi: 10.1053/j.gastro.2022.10.021.

22. Kobayashi S, Yamada M, Takamaru H, et al. 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.

23. Dekker E, Houwen B, Puig I, et al. Curriculum for optical diagnosis training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2020;**52**:899-923.

24. Gupta N, Bansal A, Rao D, et al. Prevalence of advanced histological features in diminutive and small colon polyps. *Gastrointest Endosc* 2012;**75**:1022-1030.

25. Rex DK, Kahi C, O'Brien M, et al. 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.

26. Mori Y, Kudo SE, East JE, et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). *Gastrointest Endosc* 2020;**92**:905-911.e1.

27. Rex DK. Can we do resect and discard with artificial intelligence-assisted colon polyp "optical biopsy?". *TIGE* 2020;**22**:52-55.

28. Vu HT, Sayuk GS, Hollander TG, et al. Resect and discard approach to colon polyps: real-world applicability among academic and community gastroenterologists. *Dig Dis Sci* 2015;**60**:502-508.

29. Willems P, Djinbachian R, Ditisheim S, et al. Uptake and barriers for implementation of the resect and discard strategy: an international survey. *Endosc Int Open* 2020;**8**:E684-E692.

30. Houwen B, Hassan C, Coupe VMH, et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2022;**54**:88-99.

31. Mori Y, Kudo SE, 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**:357-366.

32. Barua I, Wieszczy P, Kudo S, et al. Real-time artificial intelligence-

based optical diagnosis of neoplastic polyps during colonoscopy. *NEJM Evid* 2022;**1**.

33. Hassan C, Antonelli G, Dumonceau JM, et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2020. *Endoscopy* 2020;**52**:687-700.

34. Hassan C, Balsamo G, Lorenzetti R, Zullo A, Antonelli G. Artificial intelligence allows leaving-in-situ colorectal polyps. *Clin Gastroenterol Hepatol* 2022;**20**:2505-2513.e4.

35. Gupta S, Lieberman D, Anderson JC, et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastrointest Endosc* 2020;**91**:463-485.e5.

36. Rondonotti E, Hassan C, Tamanini G, et al. Artificial intelligence-assisted optical diagnosis for the resect-and-discard strategy in clinical practice: the Artificial intelligence BLI Characterization (ABC) study. *Endoscopy* 2023;**55**:14-22.

37. Reverberi C, Rigon T, Solari A, et al. Experimental evidence of effective human-AI collaboration in medical decision-making. *Sci Rep* 2022;**12**:14952.

38. Areia M, Mori Y, Correale L, et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. *Lancet Digit Health* 2022;**4**:e436-e444.

39. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020;**5**:352-361.

40. Su JR, Li Z, Shao XJ, et al. 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.

41. Yao L, Zhang L, Liu J, et al. Effect of an artificial intelligence-based quality improvement system on efficacy of a computer-aided detection system in colonoscopy: a four-group parallel study. *Endoscopy* 2022;**54**:757-768.

42. Chang YY, Li PC, Chang RF, et al. Development and validation of a deep learning-based algorithm for colonoscopy quality assessment. *Surg Endosc* 2022;**36**:6446-6455.

43. Lee JY, Calderwood AH, Karnes W, Requa J, Jacobson BC, Wallace MB. Artificial intelligence for the assessment of bowel preparation. *Gastrointest Endosc* 2022;**95**:512-518.e1.

44. 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;**3**:e697-e706.

45. von Renteln D, Djinbachian R, Zarandi-Nowroozi M, Taghiakbari M. Measuring size of smaller colorectal polyps using a virtual scale function during endoscopies. *Gut* 2023;**72**:417-420.

46. Liu W, Wu Y, Yuan X, et al. Artificial intelligence-based assessments of colonoscopic withdrawal technique: a new method for measuring and enhancing the quality of fold examination. *Endoscopy* 2022;**54**:972-979.

47. Liu WN, Zhang YY, Bian XQ, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol* 2020;**26**:13-19.

48. Liu P, Wang P, Glissen Brown JR, et al. The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study. *Therap Adv Gastroenterol* 2020;**13**:1756284820979165.

49. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019;**68**:1813-1819.

50. Xu L, He X, Zhou J, et al. Artificial intelligence-assisted colonoscopy: a prospective, multicenter, randomized controlled trial of polyp detection. *Cancer Med* 2021;**10**:7184-7193.

51. Glissen Brown JR, Mansour NM, Wang P, et al. Deep learning computer-aided polyp detection reduces adenoma miss rate: a United States multi-center randomized tandem colonoscopy study (CADeT-CS Trial). *Clin Gastroenterol Hepatol* 2022;**20**:1499-1507.e4.

52. Kamba S, Tamai N, Saitoh I, et al. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. *J Gastroenterol* 2021;**56**:746-757.

53. Wang P, Liu X, Berzin TM, et al. 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.

54. Shaukat A, Lichtenstein DR, Somers SC, et al. Computer-aided detection improves adenomas per colonoscopy for screening and surveillance colonoscopy: a randomized trial. *Gastroenterology* 2022;**163**:732-741.

55. Mangas-Sanjuan C, Seoane A, Alvarez-Gonzalez MA, et al. Factors associated with lesion detection in colonoscopy among different indications. *United European Gastroenterol J* 2022;**10**:1008-1019.