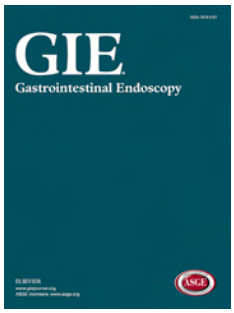


Quality Assurance of Computer-Aided Detection and Diagnosis in Colonoscopy

Daniela Guerrero Vinsard, MD, Yuichi Mori, MD, PhD, Masashi Misawa, MD, PhD, Shin-ei Kudo, MD, PhD, Amit Rastogi, MD, Ulas Bagci, PhD, Douglas K. Rex, MD, Michael B. Wallace, MD, MPH

PII: S0016-5107(19)30210-X

DOI: https://doi.org/10.1016/j.gie.2019.03.019

Reference: YMGE 11467

To appear in: *Gastrointestinal Endoscopy*

Received Date: 8 January 2019

Accepted Date: 18 March 2019

Please cite this article as: Vinsard DG, Mori Y, Misawa M, Kudo S-e, Rastogi A, Bagci U, Rex DK, Wallace MB, Quality Assurance of Computer-Aided Detection and Diagnosis in Colonoscopy, *Gastrointestinal Endoscopy* (2019), doi: https://doi.org/10.1016/j.gie.2019.03.019.

REVIEW ARTICLE

Quality Assurance of Computer-Aided Detection and Diagnosis in Colonoscopy

Running title: Standardization of computer-aided colonoscopy

Authors:

Daniela Guerrero Vinsard MD[1, 2]; Yuichi Mori MD, PhD[3]; Masashi Misawa MD, PhD[3], Shin-ei Kudo MD, PhD[3]; Amit Rastogi MD [4]; Ulas Bagci PhD[5]; Douglas K. Rex MD[6]; Michael B. Wallace MD, MPH.[7]

Authors' Affiliations:

1. Showa University International Center for Endoscopy, Showa University Northern Yokohama Hospital, Yokohama, Japan.
2. Division of Internal Medicine, University of Connecticut Health Center, Farmington, CT, USA.
3. Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan
4. Division of Gastroenterology, University of Kansas Medical Center, Kansas City, KS, USA.
5. Center for Research in Computer Vision (CRCV), University of Central Florida (UCF), Orlando, FL, USA
6. Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Indianapolis, IN, USA.
7. Division of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, FL. USA.

Keywords: artificial intelligence, optical biopsy, prediction, polyp, colon

Funding sources: None

Corresponding author information:

Yuichi Mori, MD, PhD

Digestive Disease Center, Showa University Northern Yokohama Hospital, 35-1 Chigasaki-chuo, Tsuzuki, Yokohama 224-8503, Japan

E-mail: ibusiginjp@gmail.com

Tel: +81-45-949-7000

Fax: +81-45-949-7263

ABSTRACT

Recent breakthroughs in artificial intelligence (AI), specifically via its emerging sub-field "Deep Learning," have direct implications for computer-aided detection and diagnosis (CADe/CADx) for colonoscopy. AI is expected to have at least 2 major roles in colonoscopy practice; polyp detection (CADe) and polyp characterization (CADx). CADe has the potential to decrease polyp miss rate, contributing to improving adenoma detection, whereas CADx can improve the accuracy of colorectal polyp optical diagnosis, leading to reduction of unnecessary polypectomy of non-neoplastic lesions, potential implementation of a resect and discard paradigm, and proper application of advanced resection techniques. A growing number of medical-engineering researchers are developing both, CADe and CADx systems, some of which allow real-time recognition of polyps or in vivo identification of adenomas with over 90% accuracy. However, the quality of the developed AI systems as well as that of the study designs vary significantly, hence raising some concerns regarding the generalization of the proposed AI systems. Initial studies were conducted in an exploratory or retrospective fashion using stored images and likely overestimating the results. These drawbacks potentially hinder smooth implementation of this novel technology into colonoscopy practice. The aim of this article is to review both contributions and limitations in recent machine learning based CADe/CADx colonoscopy studies and propose some principles that should underlie system development and clinical testing.

INTRODUCTION

Artificial intelligence (AI) has the potential to improve the quality of medical diagnosis and treatment. Loosely inspired by neural networks in the human brain, "Deep Learning (DL)" is capable of autonomously extracting and learning features from big data of healthcare (ie, imaging, genetics, healthcare records, and most -omics data) by the means of a multilayered system called convolutional neural networks.[1–5] It usually outperforms the traditional (non-deep learning based) machine learning methods, which extract features that scientists themselves interpreted and picked up based on experience. Such features are often called as *hand-crafted* features in machine learning literature.

In the field of gastrointestinal endoscopy, computer-aided detection and diagnosis (CADe/CADx) in colonoscopy is garnering increased attention and investigation.[6,7] AI will have 2 major initial roles in colonoscopy practice: (1) automated polyp detection (CADe) and (2) automated polyp histology characterization (CADx). CADe can minimize the probability of missing a polyp during colonoscopy, and thereby improving the adenoma detection rate (ADR) and potentially decreasing the incidence of interval cancer.[8] CADx can improve colorectal polyp optical diagnosis, leading to reduction in the resection of clinically inconsequent distal non-neoplastic lesions, potential implementation of a resect and discard paradigm and proper use of advanced resection methods such as endoscopic submucosal dissection and surgery.

More than 100 studies regarding AI in colonoscopy have been published in both engineering and medical fields.[6,7,9] However, the quality of the developed AI systems vary significantly as well as that of the study designs exploring their performance. Most of the previous studies were conducted in an experimental or retrospective fashion and the performance of AI in colonoscopy has not been sufficiently assessed in terms of its effectiveness and

reproducibility in actual clinical practice, though such pre-clinical studies are considered indispensable in the early research phase.

The goal of this review is to provide direction and facilitate appropriate research and development of CADe/CADx systems for colonoscopy. We especially highlight the following issues in this article: benefits and disadvantages of AI, published literature, current limitations, features of ideal CADe/CADx system, study design, training, and education, regulatory approval and legal issues.

BENEFITS of AI

a. CADe

A major goal of CADe in colonoscopy is to prevent missing polyps during colonoscope withdrawal, potentially increasing ADR as well as the number of adenomas per colonoscopy (APC). Lower miss rates, and thus higher ADR are strongly associated with a reduced incidence of postcolonoscopy colorectal cancers (CRC) and CRC-related mortality.[10]

b. CADx

The purpose of CADx is to predict the pathology of the detected polyps during colonoscopy. The potential benefit of CADx is to improve the accuracy of optical biopsy (eg, in vivo differentiation between neoplastic and non-neoplastic polyps using endoscopic light properties without tissue acquisition), thereby minimizing pathological assessment and unnecessary resection of distal non-neoplastic polyps leading to significant reduction in costs.[11] In addition, it would facilitate the implementation of the “resect and discard” strategy[12] into clinical practice even by inexperienced endoscopists. Future applications of CADx will include AI assessment of bowel preparation quality, lesion size measurement,

morphology description, identification of lesion features associated with deep and superficial submucosal invasion of cancer, real time guidance of therapeutic procedures, and automated report generation.

DISADVANTAGES of AI

There are potential drawbacks of AI in colonoscopy. One prospective study investigating real-time use of CADx pointed out that the time required for colonoscopy was estimated to increase by 35 to 47 seconds per polyp assessed with CADx.[13] Also, the output from CADe/CADx might distract the concentration of the endoscopists, and if inaccurate, may lead to missing/mischaracterization of polyps.[14] Reliance and/or dependence on AI may make the new generation of endoscopists less skillful and meticulous given the sense of security provided by this tool. Future prospective studies should assess the impact of these AI "pitfalls" in addition to its efficacy.

PREVIOUS CONTRIBUTIONS

In this section, we focus on clinically relevant, physician-initiated studies on AI in colonoscopy. Early research work mostly focused on technical development by computer-vision and engineering groups, and those are left outside the scope of this review.[15–20]

a) CADe

An early physician-initiated study on automated polyp detection was published by Fernandez-Esparrach et al in 2016. They used polyp boundaries information to identify polyps

effectively. They assessed their CADe model on video recordings of 31 polyps and obtained a sensitivity and specificity of >70%.[19] After this study, three additional studies on automated polyp detection were published, all of which used DL algorithms. Misawa et al[21] developed a real time CADe algorithm and assessed its performance using 50 polyp videos and 85 non-polyp videos, resulting in a sensitivity and a specificity of 90% and 63%, respectively. Urban et al[22] also developed a CADe model and reported an area under the curve of 0.991 (a measure in which values of 0.5 correspond to chance observation and 1.0 is perfect accuracy) and an accuracy of 96%. Wang et al[23] also developed a CADe model reporting over 90% values in both sensitivity and specificity. From a technical perspective, these researchers have dealt with polyp detection with already available or minimally changed DL models. Different from these retrospective analyses, Klare et al[24] conducted a prospective evaluation of a CADe model based on hand-drafted features. Their model achieved a 29.1% ADR in 55 colonoscopies, using the number of adenomas found by blinded experienced endoscopists as a reference standard.

b) CADx

Compared with CADe in which white-light endoscopy is used as the target of the image analysis, several optical technologies can be used for CADx: white light endoscopy,[25,26] magnifying narrow-band imaging (NBI),[27–32] magnifying chromoendoscopy,[33] endocytoscopy,[13,34–37] confocal laser endomicroscopy,[38,39] spectroscopy,[40,41] and autofluorescence endoscopy.[42,43,44] Among these, the most extensively studied has been magnifying NBI;[29] probably because it may have better diagnostic performance than nonmagnified NBI and does not require staining like dye-based chromoendoscopy that can be time consuming in routine clinical use.

CADx for magnifying NBI was first reported by Tischendorf et al[27] in 2010. After their work, several researchers developed CADx systems that were designed to differentiate adenomas from hyperplastic polyps based on conventional machine learning methods in early 2010s.[28,30] Their models focused on vascular patterns on the polyp surface for adenoma characterization and showed >90% sensitivities and specificities. Subsequently, Kominami et al[29] successfully evaluated their model in a prospective study, showing a 93.0% sensitivity, 93.3% specificity, 93.0% positive predictive value (PPV), and 93.3% negative predictive value (NPV). Their study also demonstrated >92.7% accuracies in predicting the surveillance interval based on optical diagnosis of diminutive polyps using CADx. Recently, 2 research teams conducted retrospective studies on newly developed CADx systems based on DL algorithms, both of which met the threshold that optical biopsy technologies require for implementation and adoption in clinical practice, namely >90% NPV for diagnosis of diminutive (≤5 mm) adenomas.[31,32]

CADx for endocytoscopy has also been investigated by a Japanese group. Endocytoscopy is performed with a colonoscope with a 520-fold ultra-magnifying function (CF-H290ECI, Olympus Corp). Although the availability of this technology is more limited than that of a magnifying colonoscope, endocytoscopy has ideal features for CADx. With endocytoscopy, endoscopists do not need to indicate the region of interest during polyp assessment given its ultra-magnification power once the tip of the device is in contact with the lesion. After several pilot studies,[34,35,37,45,46] this research group conducted a large-scale prospective study using CADx, demonstrating 91.4% sensitivity, 91.7% specificity, 88.9% PPV, and 93.7% NPV in the classification of diminutive rectosigmoid adenomas.

Laser-induced fluorescence (LIF) spectroscopy is another type of imaging modality investigated in this field. Rath et al. evaluated LIF spectroscopy CADx prospectively, reporting 100% sensitivity, 80.6% specificity, 33.3% PPV, and 100% NPV for diminutive distal adenomas.[41] However, another study by Kuiper et al[40] demonstrated less-impressive results with 83.0% sensitivity, 59.7% specificity, 71.6% PPV, and 74.2% NPV for diminutive adenomas.

CADx for white-light endoscopy, the most common endoscopic modality, has not been as extensively investigated compared with other CADx.[7] Recently, 2 research groups published preliminary results in this field; Komeda et al[25] developed a DL model, providing 75.1% accuracy with a cross-validation method. Sanchez-Montes et al[26] developed a handcrafted, predictive model based on 3 metrics (contrast, tubularity, and branching) of the polyp surface pattern, resulting in 95.0% sensitivity, 87.9% specificity, 82.6% PPV, and 96.7% NPV for diminutive rectosigmoid adenomas.

CADx has also been explored for other modalities such as confocal laser endomicroscopy[38,39] and autofluorescence endoscopy.[42,43] However, the number of publications and performance of the developed models are limited when compared with the aforementioned modalities.

CURRENT LIMITATIONS of AI

Most studies to date have developed and evaluated CADe/CADx systems using stored static and video images. These are often selected as “ideal” images of endoscopist detected lesions, and therefore the results are not truly representative of real-world effectiveness and may not be reproducible in clinical practice.

Also, pathology, which is usually used as "ground truth" for training CADx, is not always a gold standard. For example, considerable interobserver variation can be found in pathological diagnosis of sessile serrated lesions (SSLs), which creates a limitation for characterization of SSLs by means of AI. Another limitation is the lack of data on detection of inflammation and dysplasia in ulcerative and Crohn's colitis, though pilot studies in this field can be found.[47,48]

The black-box nature of the current DL algorithms can be another limitation; DL algorithms fail to reason the machine generated decision on polyp classification in CADx. Reasons causing the decision of the DL model are being investigated, and *interpretable deep learning* has already become an active area of research.

IDEAL AI SYSTEM

An ideal AI system includes at least five features: algorithm selection; ability to work real-time; appropriate output styles; smart setup of the computer; and appropriately curated data set for machine learning.

a) Algorithm selection

Before the DL era, machine learning algorithms were developed by extracting hand-crafted features (ie, features that are determined by the users) for classification of medical images.[13,28,29,34] In the DL era, algorithms learn defining features thorough exposure of images to deep neural networks.[21–23,31,32] Briefly, in the handcrafted algorithms, experts train computer systems with known features (eg, polypoid shape, surface features, vascular features) and use these features to detect and classify polyps later in test images. In contrast, DL algorithms

identify significant features from the image by a repetitive learning process. When using DL for CADe/CADx in colonoscopy, researchers should bear in mind several issues: DL algorithms usually but not necessarily outperform handcrafted algorithms[49]; although most DL algorithms are openly available (eg, Le-Net, AlexNet, VGG, GoogLeNet, ResNet) and can be installed even by "non-experts," they still require expertise and time to tune numerous parameters to achieve the best performance. This process usually entails support from industrial or engineering partners.

b) Ability to work real-time during colonoscopy

Ideally the detection and characterization of colorectal polyps should be performed by the AI system real-time during colonoscopy. For this purpose, the computer that analyzes the endoscopic images should be directly connected to the endoscopy unit. In addition, latency from capturing endoscopic image frames to outputting the analyzed results should be as short as possible, because detection of polyps with CADe later than endoscopist's detection will not be really useful.[24] To shorten the latency, it is necessary to improve the computer algorithms and use high-specification computer systems.

c) Appropriate output styles

CADe is capable of outputting 2 variables including the presence and location of polyps. Polyp presence is indicated by audible or visible alarm outside the endoscopic monitor (Figure 1-a)[21], whereas polyp location is indicated by a visible rectangle or circle that highlights the polyp (Figure 1-b).[22,23] Each type of output has pros and cons. The former method provides no information regarding polyp location, thus endoscopist has to search for them. On the other hand, its output does not distract endoscopists' attention during optical assessment of the polyp. The

latter method that makes it easier for the endoscopist to localize the polyp, but may be distracting for histology assessment.

Output of CADx also includes 2 kinds of patterns: pathology prediction[34,37,50] (eg, non-neoplastic or neoplastic) and endoscopic classification that can then be extrapolated to the histopathology of the polyp[29,31] (eg, NICE classification[51], Sano's classification[52], or Hiroshima classification[53]). Similarly, each type of output has pros and cons. The former method is considered more clinically beneficial and relevant because histopathological prediction is most useful in decision making, whereas the latter is not always a perfect indicator of pathology.[54] However, because the pathology prediction can directly influence clinical decision making, regulatory approval may be more difficult.

Regarding monitor number, dual monitor-based system (one for endoscopic image, the other for CADe/CADx) is discouraged. Taking human's visual fields into consideration, the output of AI and endoscopic image should be displayed preferably in one monitor. Several studies have demonstrated that certain visual gaze patterns on the monitor are associated with higher adenoma detection.[55] It is not yet known if CADe systems will alter visual gaze patterns and if this will improve or worsen lesion detection. Thus, visual display is an important area for research.

d) Smart setup of the computer

Assembling a stand-alone type computer in an endoscopy room is the most suitable way of implementing AI into practice smoothly. However, emerging DL technologies require high-specification in hardware setup, imposing constraints on size of the workstation and create cooling challenges. Server-based computing or cloud-computing are attractive alternatives because they may solve such installation hurdles in the endoscopy room. Nonetheless, latency

related with internet-connection speed and risk of leakage of patient's personal information are some pitfalls of cloud-computing systems.

e) Appropriately curated data set for machine learning

Whether CADe/CADx employs a conventional hand-crafted feature extraction based algorithm or a DL algorithm, both quality and quantity of the machine learning material is important to enhance its performance. Regarding the quality of the material, 3 factors should be noted: imaging modality (static images or video recordings), prevalence of positive images, and quality of annotation (ie, labeling each image frame with true data such as neoplastic/non-neoplastic or polyp/non-polyp).

Importantly, video recordings are the ideal and recommended learning material. Video contains a much larger number of image frames than static images (1second video usually includes 30 image frames). Video recordings also contain valuable low-quality images which usually cannot be found in static image collections because endoscopists tend to capture good-quality endoscopic static pictures (eg, non-blur, less stool, polyp is centered). Learning from low-quality images contributes to the robustness of the AI system.

AI systems should be trained with images that have an adequate representation of the target patient population, with a balanced proportion of polyp and non-polyp images; neoplastic and non-neoplastic images, and high- and low-quality images. It cannot be extrapolated to populations with "unnatural" disease prevalence unless the likelihood ratio is adjusted in the developed algorithm.[56]

The quality of annotations is extremely important. For fully supervised learning process, researchers should precisely annotate all the image frames (eg, non-polyp or polyp image), which will be used for machine learning. Especially when researchers use video recordings as learning material (sometimes exceed 100,000 image frames[21]), the annotation process will be likely performed by research-assistants. In that case, confirmation by expert endoscopists is mandatory to ensure the quality and accuracy of the learning material. On that premise, accurate annotations for such "big data" may require significant investment of time and resources.

Finally, a larger number of learning images contribute to a higher diagnostic accuracy, though the minimum number to reach learning plateau is still in an exploratory stage. The DL model type is another parameter that will affect the data size. Newer algorithms (such as Tiramisu[57] and SegCaps [58]) require 50% to 90% fewer parameters than conventional DL methods. To alleviate big data problems in medical imaging, DL researchers often use 2 strategies: (1) transfer learning, and (2) data augmentation. In transfer learning[59,60], the new DL model is updated from a pretrained network model which is obtained from other fields such as computer vision ImageNET where millions of natural images are made available with precise labels to train a typical neural network. In data augmentation, on the other hand, new data are artificially generated by using the available data with certain realistic manipulations such as rotating, translating, adding noise, flipping, etc. By this way, the data size can be increased considerably.[61]

STUDY DESIGN

a) Endpoints

CADe

i) Preferable endpoints

ADR is considered one of the best quality metric and endpoints to assess endoscopists' performance in clinical practice.[62] However, a limitation of ADR is that it only addresses the first adenoma found, and thus does not consider the possibility of missing subsequent adenomas. Adenoma per colonoscopy (APC) may be a more suitable endpoint to assess the ability of a CADe system to improve adenoma detection as it includes all adenomas detected per procedure.[63] In addition, polyp miss rate (PMR) is also a good option as a primary endpoint for clinical validation

ii) Definition and threshold of endpoint metrics:

ADR: defined as the proportion of screening colonoscopies performed by a physician that detect at least one histologically confirmed adenoma or adenocarcinoma.[64] The guidelines recommend ADR minimum thresholds of 25%,[65] which might be used for the performance threshold of CADe. One relevant measure of success would be to increase ADR from a lower group (e.\g, quintile as defined by Corley et al[62]) to a higher group.

APC: defined as the total number of adenomas found in all colonoscopies divided by the total number of colonoscopies. APC has a shown a good correlation with ADR in several studies.[66] A study by Kahi et al suggested that APC of 0.5 for males and 0.2 for females correspond to the current benchmarks for ADR.[67]

PMR: Defined as the total number of polyps missed from the first colonoscopy/the total number of polyps detected by both the first and the second (tandem) colonoscopy. Studies have shown

roughly 20%[68] of polyps were missed during colonoscopy, thus, a PMR of <20% may be used as a threshold for assessment of CADe. PMR can be evaluated in data sets in which 2 colonoscopies are performed on the same patients in a back-to-back manner, the second procedure may serve as a reference standard for missed lesions. In this case, missed lesions likely need to be subclassified as recognition errors (polyp on screen but not recognized) or demonstration errors (polyp hidden from view by fold, shadow, mucous, etc).

ADR, APC, and PMR should be evaluated with in vivo use of CADe during colonoscopy, but additional effect of CADe can also be independently identified in a prospective fashion if 2 rooms are prepared for the assessment (one for an endoscopist, the other for a CADe assessor)[24].

CADx

i) Preferable endpoints

The recommended thresholds proposed by the American Society of Gastrointestinal Endoscopy (ASGE) PIVI are appropriate targets to be achieved by CADx.[69]

ii) Definition and threshold of endpoint metrics

PIVI-1– resect and discard paradigm: To assess whether endoscopic optical biopsy technologies—when used with high confidence—provide ≥90% agreement in assignment of post-polypectomy surveillance compared with decisions based on histopathology. [69]

PIVI-2 – diagnose and leave paradigm: To assess whether the technology - when used with high confidence - provides a 90% or greater NPV for adenomatous histology in diminutive rectosigmoid polyps.[69]

A caveat to AI systems is that they can both detect and classify polyps[70], both of which may alter the surveillance interval. Thus, when measuring the CADx outcome, this can be assessed either separately (eg, the surveillance prediction when the CADx classifies only human-detected, and pathologically confirmed polyps) or comprehensively (eg, the surveillance prediction based on CADe plus CADx).

b. Frequently adopted study design and its issues

The vast majority of CADe/CADx systems have been evaluated in an experimental or retrospective fashion, whereas only a couple of studies were conducted prospectively with in vivo use of AI.[13,24,41,71] The steps of such experimental/retrospective studies include (1) Retrospective/prospective collection of colonoscopy images or videos. (2) Dividing the dataset into training, validation, and test data. (3) AI training with the training data and evaluated with the validation data for ensuring the correctness of the training procedure, and subsequently evaluated with test data. (4) Comparison of AI's performance with the endoscopists' performance for the same test data.

The most problematic issue of this kind of study design is the risk of selection bias; researchers tend to exclude low-quality images from the test set or omit "difficult-for-AI" cases. In addition, retrospective studies are not able to detect limitations of real-time use of AI such as additional time for the examination, endoscopist's stress burden, level of expertise and confidence, performance with low quality images and control for missing data.[14] In this regard, prospective studies on AI can provide more reliable information[13,24,29,40,41]. However, selection bias was still not eliminated in such prospective studies because they were single-arm with no controls for comparison. In the United States, the National Institutes of Health (NIH) classifies

diagnostic trials into 4 categories: phase I to phase □. A similar classification should be used for AI trials and should be accurately reported in scientific manuscripts.

c. Optimal study design

Consideration of epidemiological factors is crucial to the interpretation of the model's output. Randomized controlled clinical trials (RCTs) comparing colonoscopy with AI versus colonoscopy without AI represent the best methodology to analyze the tool's performance, safety and limitations.[72,73] This assessment should not be limited to centers with high experience in colonoscopy, otherwise, results should be adjusted to the level of expertise to preserve generalizability of recommendations. In addition, testing the model's performance in different datasets with diverse content to what was used for machine training is important to ensure external validation.[56]

Currently, there has been no RCT on AI in colonoscopy published, except for one as an abstract.[74] An example of a successfully performed RCT assessing CADx in the gynecology field was published in 2017. Pregnant women (N=47,062) were randomized to fetal heart rate machine interpretation and no machine interpretation, evaluating neonatal outcomes as the primary endpoint.[75] The study was not positive for machine interpretation, though the pilot studies provided good results.[76] This study emphasizes the need for full, real-world validation of AI systems before their routine use in clinical practice.

TRAINING AND EDUCATION

a. Required training to use AI

Because AI serves just as an adjunct to both detection and characterization of colorectal polyps (by no means an autonomous robot), basic insertion and withdrawal skill for colonoscopy is still required, though some AI software was designed to improve the quality of mucosal exposure during colonoscopy withdrawal.[77] In addition, if the AI is designed for special endoscopy such as magnifying endoscopy, endocytoscopy, or confocal laser endomicroscopy, training to capture stable endoscopic images is also required. Once endoscopists acquire these basic skills, they may be able to achieve a high diagnostic performance with the use of AI comparable with that of experts. According to a prospective study that evaluated the use of CADx for optical biopsy in-vivo, the nonexpert group provided 95.0% NPV for diminutive rectosigmoid adenomas whereas the expert group showed 91.3% NPV.[13]

b. Education

Because AI is new to most endoscopists, education programs are considered mandatory before it is adopted in clinical practice. Knowledge of strengths and weaknesses of AI can contribute to the effective use of AI and also prevent unnecessary adverse events related to its use. Especially, through an education program, endoscopists should recognize that AI sometimes outputs wrong predictions and endoscopists' final diagnosis can be strongly influenced and swayed by them.[78] For example, a study of 30 internal medicine residents showed that they exhibited a decrease in diagnostic accuracy from 57% to 48% when electrocardiograms were annotated with inaccurate CADx.[79] Importantly, endoscopists should be trained and educated about the legal responsibilities they might face before AI is implemented.

On the other hand, CADx can potentially be a valuable tool for education and training of the lesser experienced endoscopists as the endoscopist might be able to compare their thought process and diagnostic suspicion with CADx output.

REGULATORY APPROVAL AND LEGAL ISSUES

a) Regulatory approval

Because CADe/CADx for colonoscopy potentially affects the endoscopists' decision making, obtaining regulatory approval will be required for its practical use. The hurdle of obtaining approval differs according to countries and role of AI in clinical practice.[80] East et al[81] proposed three roles of CADx for colonoscopy. A second observer, a concurrent observer, or an independent decision maker. If CADx is aimed to be used independently, outstanding results from rigorously designed clinical trials will be required for its approval. In the United States, the Food and Drug Administration (FDA) recently moved to reclassify CADe software for radiology to allow an easier regulatory path to market[82,83]. For example, CADe for mammography will require class II approval, which used to require class III approval. This means that industries will no longer be required to submit a premarket approval application (PMA) for which conduct of either nonclinical or clinical trial under supervision of FDA is requested but can instead submit a less burdensome premarket notification (510(k)) before marketing their device.[82] CADe for colonoscopy may be able to follow a similar pattern in its approval and regulation process.

b) Legal issues

AI is not always beneficial for patients' care.[78] Some previous studies on CADe for mammography[82] and CADx for electrocardiography[79] demonstrated a negative effect in practice (CADe/CADx contributed to misdiagnosis). Such unintended, potentially negative effects of AI can result in legal challenges, therefore, medical malpractice insurance needs to be clear about coverage when healthcare decisions are made in part by AI.[7,84] Public guidance for the development of AI devices is now available from FDA in the United States and the Ministry of Economy, Trade and Industry of Japan; however, these documents do not establish legally enforceable responsibilities.[7]

SUMMARY

AI is expected to significantly enhance and supplement the endoscopists' performance in polyp detection and characterization. Such improvement could contribute to higher ADR (ultimately reduction of colorectal cancers) and potential implementation of a resect and discard paradigm. Although AI powered CADe/CADx systems have shown a great premise in colonoscopy, the quality of reported AI systems varies significantly. Once the efficacy and reproducibility of AI systems are validated in rigorously designed trials, they may have a significant impact on colonoscopy practice.

REFERENCES:

1. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18:500–10.
2. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare:

Past, present and future. Stroke Vasc Neurol. 2017;2:230–43.

3. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318:2199–210.
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–8.
5. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316:2402–10.
6. Liedlgruber M, Uhl A. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. IEEE Rev Biomed Eng. 2011;:73–88.
7. Mori Y, Kudo SE, Berzin TM, Misawa M, Takeda K. Computer-aided diagnosis for colonoscopy. Endoscopy. 2017;49:813–9.
8. Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, et al. Quality Indicators for Colonoscopy and the Risk of Interval Cancer. N Engl J Med. 2010;362:1795–803.
9. Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. World J Gastrointest Endosc. 2018;10:239–49.
10. Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med. 2013;369:1095–105.
11. Hassan C, Pickhardt PJ, Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. Clin Gastroenterol Hepatol. 2010;8:865-869.

12. Rex DK, Kahi C, O'Brien M, Levin TR, Pohl H, Rastogi A, 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;73:419–22.

13. Mori Y, Kudo S, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy. 2018; 169:357-366

14. Mori Y, Kudo S. Detecting colorectal polyps via machine learning. Nat Biomed Eng. 2018;2:713–4.

15. Tajbakhsh N, Gurudu SR, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans Med Imaging. 2016;35: 630-644.

16. Tajbakhsh N, Gurudu SR, Liang J. Automatic polyp detection using global geometric constraints and local intensity variation patterns. Med Image Comput Comput Assist Interv. 2014;17:179–87.

17. Karkanis SA, Iakovidis DK, Maroulis DE, Karras DA, Tzivras M. Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Trans Inf Technol Biomed. 2003;7:141–52.

18. Wang Y, Tavanapong W, Wong J, Oh JH, de Groen PC. Polyp-Alert: Near real-time feedback during colonoscopy. Comput Methods Programs Biomed. 2015 Jul;120:164–79.

19. Fernández-Esparrach G, Bernal J, López-Cerón M, Córdova H, Sánchez-Montes C, Rodríguez de Miguel C, et al. Exploring the Clinical Potencial of an Automatic Colonic Polyp Detection Method Based on Energy Maps Creation. Endoscopy. 2016;48:837–42.

20. Tajbakhsh N, Gurudu SR, Liang J. A Comprehensive Computer-Aided Polyp Detection

System for Colonoscopy Videos. Inf Process Med Imaging. 2015;24:327–38.

21. Misawa M, Kudo S ei, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. Gastroenterology. 2018;154:2027–2029.e3.

22. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep Learning Localizes and Identifies Polyps in Real Time with 96% Accuracy in Screening Colonoscopy. Gastroenterology. 2018; 155:1069–1078.

23. Wang P, Xiao X, Brown JRG, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng. 2018;2:741–8.

24. Klare P, Sander C, Prinzen M, Haller B, Nowack S, Abdelhafez M, et al. Automated polyp detection in the colorectum: a prospective study (with videos). Gastrointest Endosc. 2019;89:576–82.

25. Komeda Y, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, et al. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. Oncol. 2017;93:30–4.

26. Sanchez-Montes C, Sanchez FJ, Bernal J, Cordova H, Lopez-Ceron M, Cuatrecasas M, et al. Computer-aided prediction of polyp histology on whilte-light colonoscopy using surface pattern analysis. Endoscopy. 2019;51:261–5.

27. Tischendorf JJW, Gross S, Winograd R, Hecker H, Auer R, Behrens A, et al. Computer-aided classification of colorectal polyps based on vascular patterns: A pilot study. Endoscopy. 2010;42:203–7.

28. Gross S, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, et al. Computer-based

classification of small colorectal polyps by using narrow-band imaging with optical magnification. Gastrointest Endosc. 2011;74:1354–9.

29. Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. 2016;83:243-9

30. Takemura Y, Yoshida S, Tanaka S, Kawase R, Onji K, Oka S, et al. Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). Gastrointest Endosc. 2012;75:179–85.

31. Byrne MF, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut. 2019; 68:94-100

32. Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HHS, Tseng VS. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. Gastroenterology. 2018; 154:568-575.

33. Takemura Y, Yoshida S, Tanaka S, Onji K, Oka S, Tamaki T, et al. Quantitative analysis and development of a computer-aided system for identification of regular pit patterns of colorectal lesions. Gastrointest Endosc. 2010;72:1047–51.

34. Misawa M, Kudo SE, Mori Y, Nakamura H, Kataoka S, Maeda Y, et al. Characterization of Colorectal Lesions Using a Computer-Aided Diagnostic System for Narrow-Band Imaging Endocytoscopy. Gastroenterology. 2016;150:1531–32.e3.

35. Mori Y, Kudo S, Chiu P, Singh R, Misawa M, Wakamura K, et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-

based study. Endoscopy. 2016;48:1110–8.

36. Mori Y, Kudo SE, Mori K. Potential of artificial intelligence-assisted colonoscopy using an endocytoscope (with video). Dig Endosc. 2018;30:52–3.

37. Takeda K, Kudo SE, Mori Y, Misawa M, Kudo T, Wakamura K, et al. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. Endoscopy. 2017;49:798–802.

38. André B, Vercauteren T, Buchner AM, Krishna M, Ayache N, Wallac MB. Software for automated classification of probe-based confocal laser endomicroscopy videos of colorectal polyps. World J Gastroenterol. 2012;18:5560–9.

39. Ştefănescu D, Streba C, Cârţână ET, Săftoiu A, Gruionu G, Gruionu LG. Computer Aided Diagnosis for Confocal Laser Endomicroscopy in Advanced Colorectal Adenocarcinoma. PLoS One. 2016;11:e0154863.

40. Kuiper T, Alderlieste YA, Tytgat KM, Vlug MS, Nabuurs JA, Bastiaansen BA, et al. Automatic optical diagnosis of small colorectal lesions by laser-induced autofluorescence. Endoscopy. 2015;47:56–62.

41. Rath T, Tontini GE, Vieth M, Nägel A, Neurath MF, Neumann H. In vivo real-time assessment of colorectal polyp histology using an optical biopsy forceps system based on laser-induced fluorescence spectroscopy. Endoscopy. 2016;48:557–62.

42. Aihara H, Saito S, Inomata H, Ide D, Tamai N, Ohya TR, et al. Computer-aided diagnosis of neoplastic colorectal lesions using real-time; numerical color analysis during autofluorescence endoscopy. Eur J Gastroenterol Hepatol. 2013;25:488–94.

43. Inomata H, Tamai N, Aihara H, Sumiyama K, Saito S, Kato T, et al. Efficacy of a novel auto-fluorescence imaging system with computer-assisted color analysis for assessment of

colorectal lesions. World J Gastroenterol. 2013;19:7146–53.

44. Aihara H, Sumiyama K, Saito S, Tajiri H, Ikegami M. Numerical analysis of the autoflorescence intensity of neoplastic and non-neoplastic colorectal lesions by using a novel videoendoscopy system. Gastrointestinal Endosc. 2009;69:726-33.

45. Mori Y, Kudo SE, Wakamura K, Misawa M, Ogawa Y, Kutsukawa M, et al. Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). Gastrointest Endosc. 2015; 81:621-9.

46. Misawa M, Kudo S, Mori Y, Nakamura H, Ishida F, Wakamura K, et al. Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: comparison with experts. Int J Comput Assist Radiol Surg. 2017;12:757–66.

47. Ozawa T, Ishihara S, Fujishiro M, Saito H, Kumagai Y, Shichijo S, et al. Novel Computer-assisted Diagnosis System for Endoscopic Disease Activity in Patients with Ulcerative Colitis. Gastrointest Endosc. 2019;89:416-421.

48. Maeda Y, Kudo S-E, Mori Y, Misawa M, Ogata N, Sasanuma S, et al. 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.

49. Khan S, Yong S. A comparison of deep learning and hand crafted features in medical image modality classification. 3rd International Conference on Computer and Information Sciences (ICCOINS). 2016. IEEE. p. 633–8.

50. East JE, Vleugels JL, Roelandt P, Bhandari P, Bisschops R, Dekker E, et al. Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE)

Technology Review. 2016;48:1029–45.

51. Hewett DG, Kaltenbach T, Sano Y, Tanaka S, Saunders BP, Ponchon T, et al. Validation of a Simple Classification System for Endoscopic Diagnosis of Small Colorectal Polyps Using Narrow-Band Imaging. Gastroenterology. 2012;143:599–607.e1.

52. Tamai N, Saito Y, Sakamoto T, Nakajima T, Matsuda T, Sumiyama K, et al. Effectiveness of computer-aided diagnosis of colorectal lesions using novel software for magnifying narrow-band imaging: a pilot study. Endosc Int Open. 2017;05:E690–4.

53. Hirata M, Tanaka S, Oka S, Kaneko I, Yoshida S, Yoshihara M, et al. Evaluation of microvessels in colorectal tumors by narrow band imaging magnification. Gastrointest Endosc. 2007;66:945–52.

54. Abu Dayyeh BK, Thosani N, Konda V, Wallace MB, Rex DK, Chauhan SS, et al. 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;81:502–16.

55. Lami M, Singh H, Dilley J, Ashraf H, Edmondon M, Orihuela-Espina F, et al. Gaze patterns hold key to unlocking successful search strategies and increasing polyp detection rate in colonoscopy. Endoscopy. 2018;50:701–7.

56. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology. 2018;286:800–9.

57. Torosdagli N, Liberton DK, Verma P, Sincan M, Lee JS, Bagci U. Deep Geodesic Learning for Segmentation and Anatomical Landmarking. IEEE Trans Med Imaging. Epub 2018; Oct 12.

58. Lalonde R, Bagci U. Capsules for Object Segmentation. Med Imaging with Deep Learn. 1st Conference on Medical Imaging with Deep Learning (MIDL 2018) Amsterdam, The Netherlands. Apr 2018. Available from: https://arxiv.org/pdf/1804.04241.pdf

59. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. The MIT Press. Cambridge, Massachusetts, 2016. Available from: https://www.deeplearningbook.org

60. Hussein S, Chuquicusma MM, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and Pancreatic Tumor Characterization in Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches. IEEE Trans Med Imaging. Epub 2019; Jan 23.

61. Mortazi A, Karim R, Rhode K, Burt J, Bagci U. CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI. Springer; 2017. p. 377–85.

62. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma Detection Rate and Risk of Colorectal Cancer and Death. N Engl J Med [Internet]. 2014;370:1298–306.

63. Rex DK, Petrini J, Baron TH, Chak A, Cohen J, SE D. Quality Indicators for colonoscopy. Gastrointest Endosc. 2006;63:16–28.

64. Rex DK, Bond JH, Winawer S, Levin TR, Burt RW, Johnson DA, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: Recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol. 2002;97:1296–308.

65. Rex DK, Schoenfeld PS, Cohen J, Pike IM, Adler DG, Fennerty MB, et al. Quality indicators for colonoscopy. Am J Gastroenterol. 2015;110:72–90.

66. Hilsden RJ, Bridges R, Dube C, McGregor SE, Naugler C, Rose SM, et al. 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;111:1743–9.

67. 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;79:448–54.

68. 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;101:343–50.

69. Rex D, Kahi C, Levin T, Pohl H, Rastogi A, Burgart L, 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 diminutiv diminutive colorectal polyps. Gastrointest Endosc. 2011; 73:419-22.

70. Mori Y, Kudo S, Misawa M, Mori K. Simultaneous detection and characterization of diminutive polyps with the use of artificial intelligence during colonoscopy. Video GIE. 2019; 4:7-10.

71. Ichimasa K, Kudo SE, Mori Y, Wakamura K, Ikehara N, Kutsukawa M, et al. Double staining with crystal violet and methylene blue is appropriate for colonic endocytoscopy: An invivo prospective pilot study. Dig Endosc. 2014;26:403–8.

72. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130:515–24.

73. Leggett CL, Fasge KKW. Computer-aided diagnosis in GI endoscopy : looking into the

future. Gastrointest Endosc. 2016;84:842–4.

74. Wang P, Bharadwaj S, Berzin TM, Becq A, Li L, Liu P, et al . Assistance of a real-time automatic colon polyp detection system increases polyp and adenoma detection during colonoscopy: a prospective randomized controlled study. United Eur Gastroenterol J. 2018;87:490-491.

75. Brocklehurst P, Field D, Greene K, Juszczak E, Keith R, Kenyon S, et al. Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. Lancet. 2017;389:1719–29.

76. Keith RD, Beckley S, Garibaldi JM, Westgate JA, Ifeachor EC, Greene KR. A multicentre comparative study of 17 experts and an intelligent computer system for managing labour using the cardiotocogram. Br J Obstet Gynaecol. 1995;102:688–700.

77. Oh* J, Hwang S, Cao Y, Tavanapong W, Liu D, Wong J, et al. Measuring Objective Quality of Colonoscopy. IEEE Trans Biomed Eng. 2009;56:2190–6.

78. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. Jama. 2017;318: 517-518.

79. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: The case of electrocardiograms. J Am Med Informatics Assoc. 2003;10:478–83.

80. Chinzei K, Shimizu A, Mori K, Harada K, Takeda H, Hashizume M, et al. Regulatory Science on AI-based Medical Devices and Systems. Adv Biomed Eng. 2018;7:118–23.

81. East JE, Rees CJ. Making optical biopsy a clinical reality in colonoscopy. Lancet Gastroenterol Hepatol. 2018;3:10–2.

82. Food and Drug Administration HHS. Radiology Devices; Reclassification of Medical Image Analyzers. A proposed rule by the Food and Drug Administration. June 2018. Pp

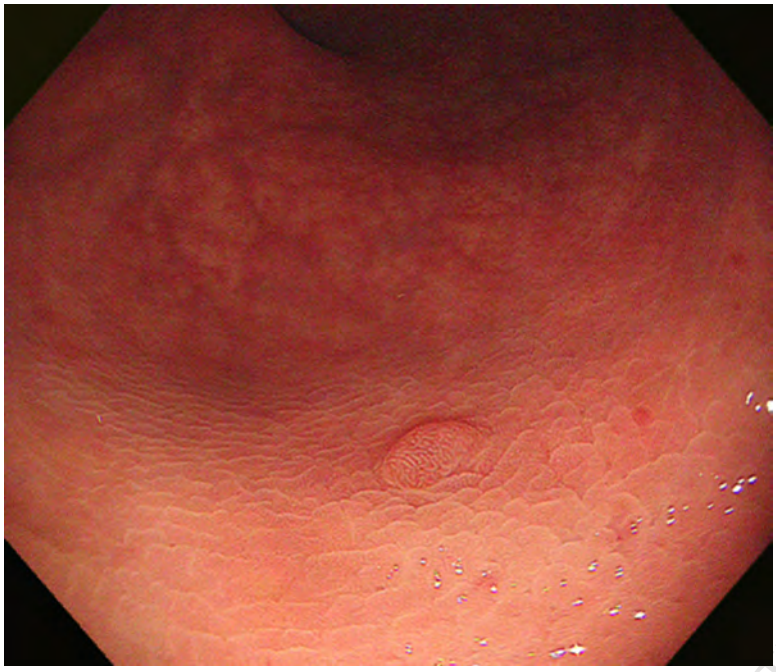
25598-25604. Available from: https://www.federalregister.gov/documents/2018/06/04/2018-11880/radiology-devices-reclassification-of-medical-image-analyzers

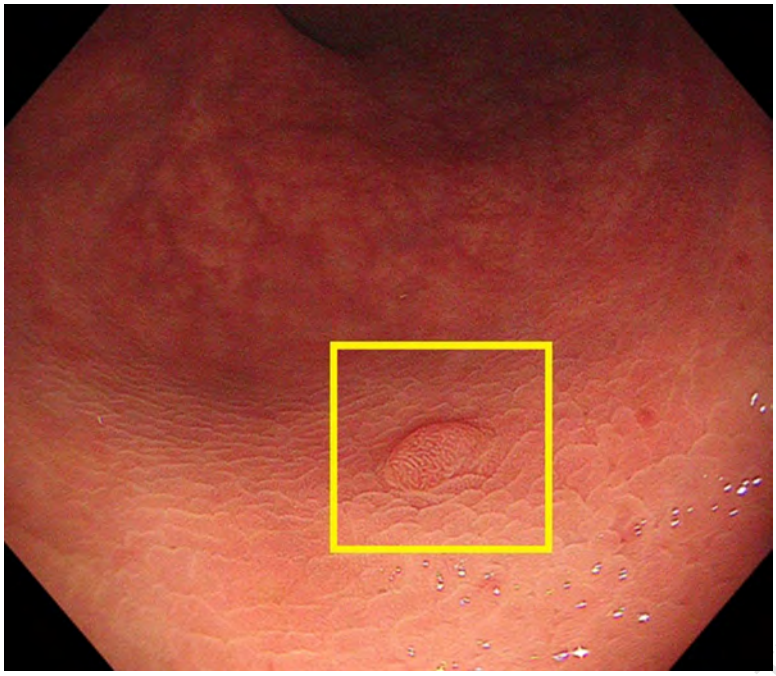
83. Park SH. Regulatory Approval versus Clinical Validation of Artificial Intelligence Diagnostic Tools. Radiology. 2018;288:910–1.

84. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2:719–31.

FIGURE LEGENDS

Figure 1. Two types of outputs for automated polyp detection. A, Presence of the polyp is indicated by a visible alarm outputting color outside the endoscopic monitor. B, Polyp location is indicated by putting a visible rectangle.





ABBREVIATIONS:

ADR: adenoma detection rate. AI: artificial intelligence. APC: adenoma per colonoscopy. CADe: computer aided detection. CADx, computer aided characterization. CRC: colorectal cancer. DL: Deep Learning. FDA: Food and Drug Administration. LIF: laser-induced fluorescence. NBI: narrow band imaging. NPV: negative predictive value. PIVI: preservation and incorporation of valuable endoscopic innovations. PMR: polyp miss rate. PPV: positive predictive value. SSLs: sessile serrated lesions.