

## Open Access

# Artificial Intelligence in Lower Gastrointestinal Endoscopy: The Current Status and Future Perspective

Sebastian Manuel Milluzzo<sup>1,2</sup>, Paola Cesaro<sup>1</sup>, Leonardo Minelli Grazioli<sup>1</sup>, Nicola Olivari<sup>1</sup> and Cristiano Spada<sup>1,2</sup>

<sup>1</sup>Digestive Endoscopy Unit and Gastroenterology, Fondazione Poliambulanza, Brescia, <sup>2</sup>Department of Gastroenterology, Fondazione Policlinico Universitario A. Gemelli IRCCS -Università Cattolica del Sacro Cuore, Roma, Italy

The present manuscript aims to review the history, recent advances, evidence, and challenges of artificial intelligence (AI) in colonoscopy. Although it is mainly focused on polyp detection and characterization, it also considers other potential applications (i.e., inflammatory bowel disease) and future perspectives. Some of the most recent algorithms show promising results that are similar to human expert performance. The integration of AI in routine clinical practice will be challenging, with significant issues to overcome (i.e., regulatory, reimbursement). Medico-legal issues will also need to be addressed. With the exception of an AI system that is already available in selected countries (GI Genius; Medtronic, Minneapolis, MN, USA), the majority of the technology is still in its infancy and has not yet been proven to reach a sufficient diagnostic performance to be adopted in the clinical practice. However, larger players will enter the arena of AI in the next few months. **Clin Endosc 2021;54:329-339**

**Key Words:** Artificial intelligence; Colon capsule endoscopy; Colonoscopy; Endoscopy

## INTRODUCTION

Colorectal cancer (CRC) still represents a major cause of death,<sup>1</sup> although its incidence and mortality declined over the last decades as a consequence of screening programs and polypectomy of the adenomatous polyps.<sup>2,3</sup> Nevertheless, the endoscopic detection of CRC precursor lesions remains a challenge. A recent systematic review and meta-analysis<sup>4</sup> showed an adenoma miss rate of 25% for any adenoma, 9% for advanced adenomas, and 27% for serrated polyps in tandem colonoscopy studies. This can partially explain the occurrence of interval cancer. Adenoma detection rate (ADR) is the current most reliable measure of a colonoscopist's capability to

detect adenomas. Higher ADR is associated with lower interval CRCs and lower CRC mortality.<sup>5,6</sup> ADR varies dramatically among different endoscopists (7%–53%)<sup>7</sup> and also for each operator during the course of the day (approximately 7%), probably due to the intrinsic imprecision of the human eye brought about by fatigue and rush activities.<sup>8</sup>

New technologies have been reported in the literature to improve ADR, including enhanced optics, distal attachments, cap-assisted techniques, and balloon-assisted devices, with the goal of improving mucosa visualization and eventually diagnosing hardly detectable polyps.<sup>9</sup>

By aiding polyp detection on colonoscopy images (Figs. 1 and 2), artificial intelligence (AI) can reduce performance variability. AI is an information technology evolution with software and hardware development to create computer machines with human-like characteristics, such as visual-spatial perceptions and decisional algorithms that are able to solve relatively new problems never seen before (machine learning). In the late 1970s, a specific software was able to teach and train a newly developed neural network (NN) evolving from an expert system (system that emulates the decision-making ability of a human expert) to the concept of AI. NNs are mathematical computing systems that analyze a large variety of inputs

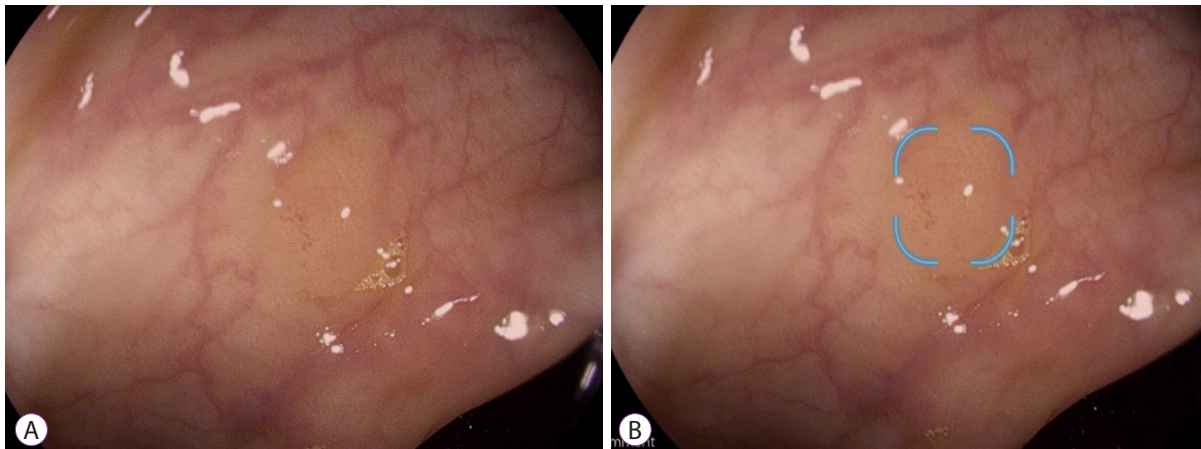
Received: March 31, 2020 Revised: June 16, 2020

Accepted: July 30, 2020

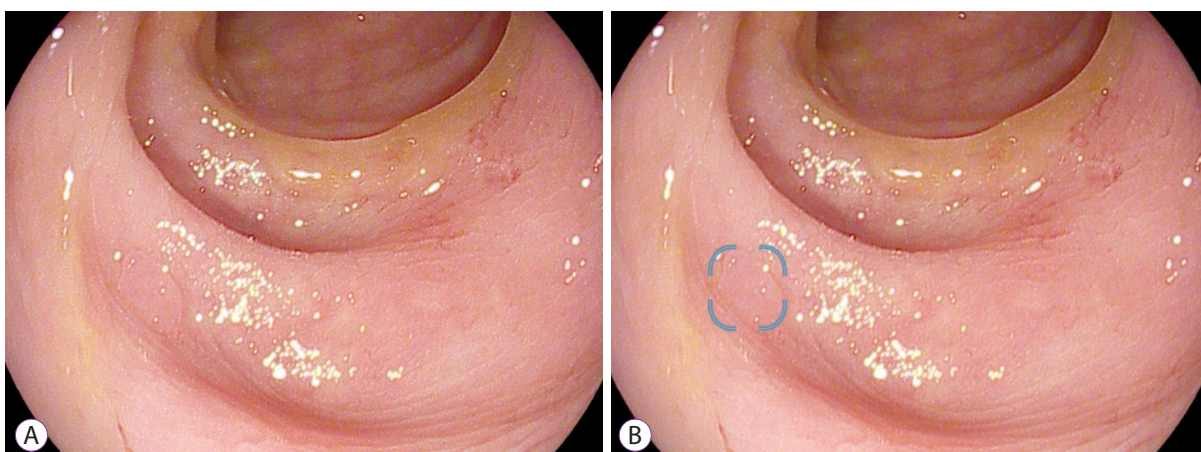
**Correspondence:** Cristiano Spada  
Digestive Endoscopy Unit and Gastroenterology, Fondazione Poliambulanza, Via L. Bissolati 57, Brescia 25125, Italy  
**Tel:** +39-030-3515-370, **Fax:** +39-030-3515-372, **E-mail:** cristiano.spada@poliambulanza.it

**ORCID:** <https://orcid.org/0000-0002-5692-0960>

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Fig. 1.** A 5-mm polyp is visualized during colonoscopy (A) and with the support of DISCOVERY (PENTAX Medical, Tokyo, Japan) artificial intelligence system (B) which generates a small box on each frame where a polyp is detected.



**Fig. 2.** A 3-mm polyp is visualized during colonoscopy (A) and with the support of DISCOVERY (PENTAX Medical, Tokyo, Japan) artificial intelligence system (B) which generates a small box on each frame where a polyp is detected.

to obtain an output through a middle (or hidden) layer that works like natural brain cells in terms of adaptive connections. The strength and type of connections within the hidden layer in NN are obtained through the training phase with supervised learning (declared inputs and outputs), semi-supervised learning (declared inputs and some of the outputs), or unsupervised learning (declared inputs without known correct outputs) with a correction system of “reward and punishment”.

When applied to lower gastrointestinal (GI) endoscopy, AI plays as additional or virtual eyes, that is, an additional endoscopist who continuously stands behind the endoscopist to help detect polyps, with the final goal to improve the quality of diagnosis, reduce operators’ variability, and improve the outcomes.

This review aimed to show the current status of AI in lower GI endoscopy and underline future potentials and limitations of the said technology.

### ARTIFICIAL INTELLIGENCE IN LOWER GASTROINTESTINAL ENDOSCOPY

Computer-aided diagnosis (CAD) potentially promises the reduction in colonoscopy performance variation. The available evidence reveals that endoscopists are flawed in detecting colonic lesions. A recent meta-analysis showed that adenoma miss rates might be as high as approximately 25%,<sup>4</sup> and a post-colonoscopy cancer may occur in approximately 9% of

cases within 3 years of an apparently negative colonoscopy.<sup>10</sup> CAD using AI and deep learning techniques was designed to reduce human variability. In this sense, AI is going to be rapidly embedded in routine endoscopy and in CRC screening, initially, with two main goals: (1) detection and (2) histological characterization.

To pass the “test”, AI systems must show that they are accurate, sensitive, specific, and have a fast latency time. Ultimately, the performance of AI should be evaluated in trials with long-term follow-up with the incidence and mortality from CRC as outcomes. A summary of the most relevant literature is presented in Tables 1, 2, and 3.

**Table 1.** Studies that Evaluated the Role of Artificial Intelligence for Polyp Detection

Study	Study design	Number of patients or polyps	Sensitivity (%)	Specificity (%)	Accuracy (%)	ADR (%) (AI vs. standard)
Wang et al. (2015) <sup>11</sup>	Retrospective	43 polyps	-	-	97.7	-
Fernández-Esparrach et al. (2016) <sup>12</sup>	Retrospective	31 polyps	70.4	72.4	-	-
Misawa et al. (2018) <sup>13</sup>	Retrospective	155 polyps	90	63	76.5	-
Urban et al. (2018) <sup>14</sup>	Prospective	4,088 polyps	96.9	88.1	96.4	-
Wang et al. (2019) <sup>15</sup>	RCT	1,058 patients	94.4	95.2	-	29.1 vs. 20.3 <i>p</i> <0.001
Wang et al. (2020) <sup>16</sup>	RCT	1,046 patients	-	-	-	34 vs. 28 <i>p</i> =0.03
Liu et al. (2020) <sup>17</sup>	RCT	1,026 patients	-	-	-	39 vs. 23 <i>p</i> <0.001
Gong et al. (2020) <sup>18</sup>	RCT	704 patients	-	-	-	16% vs. 8% <i>p</i> =0.0010
Hassan et al. (2020) <sup>19</sup>	Prospective	105 patients	99.7	99	-	-

ADR, adenoma detection rate; AI, artificial intelligence; RCT, randomized controlled trial.

**Table 2.** Studies that Evaluated the Role of Artificial Intelligence for Polyp Characterization

Study	Study design	Number of patients or polyps	Sensitivity (%)	Specificity (%)	Accuracy (%)
Tischendorf et al. (2010) <sup>22</sup>	Prospective	128 patients	90	70	-
Kudo et al. (2020) <sup>25</sup>	Retrospective	100 polyps	96.6	94.3	96
Chen et al. (2018) <sup>26</sup>	Prospective	284 polyps	96.3	78.1	90.1
Byrne et al. (2019) <sup>27</sup>	Retrospective	125 polyps	98	83	94
Shahidi et al. (2020) <sup>28</sup>	Prospective	644 polyps	90.3	90.9	89.6
Zachariah et al. (2020) <sup>29</sup>	Prospective	634 polyps	95.7	89.9	93.6

**Table 3.** Studies that Evaluated the Role of Artificial Intelligence in Ulcerative Colitis

Study	Study design	Number of patients or polyps	Sensitivity (%)	Specificity (%)	Accuracy (%)	Evaluation
Maeda et al. (2019) <sup>30</sup>	Retrospective	87 patients	74	97	91	Histologic inflammation
Ozawa et al. (2019) <sup>31</sup>	Retrospective	114 patients	AUROCs = 0.86 and 0.98 to identify Mayo 0 and 0–1			Mucosal disease activity (Mayo score)
Stidham et al. (2019) <sup>32</sup>	Retrospective	3,082	83	96	-	Endoscopic severity
Takenaka et al. (2020) <sup>33</sup>	Prospective	875	93.3	87.8	90.1	Endoscopic remission

AUROC, areas under the receiver operating characteristic curves.

## COMPUTER-AIDED POLYP DETECTION IN COLONOSCOPY

The preliminary AI-based systems were designed to assist the endoscopists in improving polyp detection. Early projects focused on techniques developed by differentiating polyp features (i.e., color, shape, texture) from the surrounding normal mucosa. For example, Wang et al.<sup>11</sup> developed a Polyp-Alert software system, which used a color–texture analysis method (local binary pattern and opponent color local binary pattern) to continuously analyze image streams at 10 frames/sec, to identify polyps using their edges. Despite the high accuracy (97.7%) and short latency (0.02 sec), this computer-aided detection (CADE) was burdened by a high number of false-positives, which was caused by artifacts owing to inadequate bowel preparation or normal findings (i.e., the folds, appendicular orifice, ileocecal valve). Other attempts were performed by Fernández-Esparrach et al.<sup>12</sup> who assessed the potentiality of the Window Median Depth of Valleys Accumulation (WMDOVA) energy maps system that defined polyps as protrusions in the mucosa and their boundaries as intensity valleys. Polyp detection was achieved with 70.4% sensitivity and 72.4% specificity.

Handcrafted AI algorithms have been surpassed and are going to be replaced by convolutional neural networks (CNN) with the ability to perform real-time polyp detection. CNN is a deep learning algorithm, which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and differentiate one from the other. CNN, for instance, separates the various characteristics of the image (red, green, and blue colors, narrow-band imaging [NBI], chromoendoscopy, depression or elevation, etc.) in multiple separate algorithms in the hidden layers of the NN that works in parallel to enforcing or weakening the probability of a certain output.

Several studies reported the results of different systems. For example, Misawa et al.<sup>13</sup> developed an original CADE system based on three-dimensional CNN, which was trained on white light endoscopy (WLE) images. They used a dataset consisting of 73 colonoscopy video sequences, including 155 polyps. Moreover, flat lesions populated the dataset. The system showed 90% sensitivity and 63% specificity, with 76.5% accuracy. In a prospective study, Urban et al.<sup>14</sup> trained the AI on a database of images without polyps and then retrained it on images with polyps, using both WLE and NBI. This CADE system detected the polyps with a processing time of 10 ms/frame. The accuracy was 96.4%, and area under the curve (AUC) of 0.991. Interestingly, the system was also tested on 20 colonoscopy videos. Three colonoscopists (ADR > 50%) were

asked to identify all polyps in 9 colonoscopy videos, without benefit of the CNN. Their “polyp encounters” were then combined by consensus. CNN-overlaid videos were generated by superimposing a small green box on each frame where a polyp was detected. A senior expert (ADR  $\geq$  50%, > 20,000 colonoscopies), who was used as reference, was asked to review the CNN-overlaid videos and judge the true polyp presence. In the first dataset of 9 videos, 28 polyps were removed by the colonoscopist. The four experts identified 8 additional polyps without CNN assistance that had not been removed, and identified an additional 17 polyps with CNN assistance. The CNN false-positive rate was 7%. The authors concluded that CNN assistance is able to affect ADR, highlighting polyps that could potentially be overlooked.

Similarly, promising results were confirmed in three recent randomized controlled trials (RCTs) by Wang et al. and Liu et al.<sup>15-17</sup> In a study published by Wang et al.,<sup>15</sup> 1,058 patients were randomized to standard colonoscopy ( $n=536$ ) and colonoscopy with CAD ( $n=522$ ). The primary outcome was ADR. The real-time AI system (Shanghai Wision AI Co., Ltd., Shanghai, China) was designed on a deep learning architecture and showed a per-image sensitivity and specificity of 94.38% and 95.92%, respectively. The AUC was 0.984. In this study, the authors showed that CAD-assisted colonoscopy significantly improved ADR (29.1% vs. 20.3%,  $p<0.001$ ) and the mean number of adenomas per patient (0.53 vs. 0.31,  $p<0.001$ ). It should be noted that the higher ADR was mainly due to a higher number of diminutive adenomas (185 vs. 102,  $p<0.001$ ), although there was no statistical difference for bigger adenomas (77 vs. 58,  $p=0.075$ ). Correspondingly, the number of hyperplastic polyps was significantly increased (114 vs. 52,  $p<0.001$ ). The same system was validated by Wang et al. in another double-blind randomized trial.<sup>16</sup> Patients were randomly allocated to colonoscopy with either the CADE system or the sham system. The primary outcome was the ADR that was 34% for the CADE group and 28% for the sham system (odds ratio [OR], 1.36; 95% confidence interval [CI], 1.03–1.79;  $p=0.030$ ). Polyps initially missed by the endoscopist, but identified by the CADE system were small, isochromatic, flat, had unclear borders, were hidden by folds, and were on the edge of the visual field.

Similarly, in the study by Liu et al.,<sup>17</sup> 1,026 patients were prospectively randomized for colonoscopy with the CADE or without (control group). The polyp detection rate (PDR) was the primary outcome. The CADE system (Henan Xuanweitang Medical Information Technology Co., Ltd., Zhengzhou, China) was developed on the basis of indepth learning architecture that was trained on 535 videos with and without polyps. The PDR in the control and CAD groups were 0.28 and



0.44, respectively (OR = 1.57; 95% CI, 1.586–2.483;  $p < 0.001$ ). The ADR in the control and CADe groups was 0.23 and 0.39, respectively (OR = 1.64; 95% CI, 1.201–2.220;  $p < 0.001$ ). Of note, as shown by Wang et al.,<sup>15</sup> although the average number of adenomas and the number of small adenomas and hyperplastic polyps significantly increased ( $p < 0.01$ ), the number of larger adenomas (> 10 mm) was comparable between the groups ( $p > 0.05$ ). This may give rise to the issue regarding the clinical relevance of the adjunctive polyps detected by the AI-based systems as well as the cost–benefit ratio (i.e., AI may result in the useless removal of hyperplastic polyps) in which the PDR improvement needs to be further evaluated.

The ENDOANGEL system<sup>18</sup> was developed using deep neural networks (DNN). It was designed not only to increase ADR, but also to actually assist the endoscopist in improving the colonoscopy quality, monitoring withdrawal speed, timing of colonoscopy intubation and withdrawal, and reminding endoscopists of blind spots caused by endoscope slipping. Standard endoscopes from Olympus Optical (Tokyo, Japan) and Fujifilm (Kanagawa, Japan) were used for training. The system was tested in a single-center, prospective, randomized trial by Gong et al.<sup>18</sup> Patients were randomized to either colonoscopy with the ENDOANGEL system or unassisted colonoscopy (control). The primary endpoint was ADR. Sixteen percent of patients allocated in the ENDOANGEL-assisted colonoscopy were diagnosed as having one or more adenomas compared with 8% allocated in the control colonoscopy (OR, 2.30; 95%

CI, 1.40–3.77;  $p = 0.0010$ ).

The vast majority of CNN-based systems are still in a proof-of-concept stage and they need to be further evaluated for their integration in the endoscopic towers. GI Genius (Medtronic, Minneapolis, MN, USA) is the only available system for routine clinical practice in selected countries. GI Genius was recently validated by Hassan et al.<sup>19</sup> on 338 polyps (168 adenomas or serrated polyps) from 105 patients. The algorithm was previously trained on WLE videos of 2,684 histologically confirmed polyps from 840 patients who underwent high-definition white-light colonoscopy. Patients were randomized between the validation and training groups. For the validation phase, 338 polyps (168/338 adenomas or sessile serrated adenomas, 49.7%) from 105 patients were used. The overall per lesion sensitivity was 99.7%, and the false-positive rate was less than 1%. The AI system anticipated the polyp detection against the average of the five reference endoscopists in 82% of cases. The difference in reaction time was  $-1.27 \pm 3.81$  sec. The study confirms that GI Genius is able to virtually detect all the lesions diagnosed by expert endoscopists with an anticipation of the diagnosis compared with the human reader in the vast majority of cases, with a negligible rate of false-positive cases. Similarly, another system has been designed and is going to be released in the next few months: DISCOVERY (PENTAX Medical, Tokyo, Japan).<sup>20</sup> DISCOVERY is intended to support endoscopists in polyp or lesion detection. It incorporates the AI based on a DNN in a panel



**Fig. 3.** DISCOVERY (PENTAX Medical, Tokyo, Japan) incorporates the artificial intelligence based on a deep neural network in a panel PC with a 32 inch LCD display. This panel PC can be connected with a signal cable (DVI/ HD-SDI) to each PENTAX HD+ video processor for integration and is intended to be used as a secondary monitor.

PC with a 32 inch LCD display (Fig. 3). This panel PC can be connected to each PENTAX HD+ video processor for integration and is intended to be used as a secondary monitor. The system was trained using 10,467 colonoscopy images from 504 polyps. To evaluate DNN's real-time capability to detect polyps, a set of 45 videos has been used. To ensure a realistic evaluation, 5 out of the 45 videos did not include any polyp to estimate the system's false-positive rate. Polyps that populated the videos were representative of all the morphologic patterns, including flat morphology, and either diminutive ( $\leq 5$  mm) or small (6–10 mm) polyps. The DNN's sensitivity and specificity for polyp detection was 90% and 94.6%, respectively. AUC for classification was 97%. DISCOVERY is going to be released in the coming months, and a multicenter study will be conducted to further investigate the system.

Recently, a meta-analysis that evaluated the role of AI for polyp detection was published. Overall, the AUC of AI for polyp detection was 0.90 (95% CI, 0.67–1.00). The AI sensitivity for polyp detection was 95.0% (95% CI, 91.0%–97.0%) and specificity was 88.0% (95% CI, 58.0%–99.0%). When limiting the analysis to studies that used a deep learning model (i.e., CNN models), the AUC, sensitivity, and specificity of AI for polyp detection was 0.91 (95% CI, 0.73–1.00), 94.4% (95% CI, 89.9%–97.0%), and 91.9% (95% CI, 44.3%–99.4%), respectively.<sup>21</sup>

## COMPUTER-AIDED POLYP CHARACTERIZATION IN COLONOSCOPY

In addition to detection, AI has been designed for automated polyp characterization (CADx). The AI-assisted classification of colorectal polyps using NBI and magnification was initially evaluated by Tischendorf et al.,<sup>22</sup> who analyzed 209 polyps. The evaluation was based on three features: mean vessel length, mean vessel circumference, and mean brightness at detected blood vessels as well as the combination of all three features. The primary outcome was to distinguish non-adenomatous from adenomatous polyps. Histology was the gold standard and the comparison was between CADx and expert endoscopists who were blinded to the histology. CADx showed a sensitivity of 90% and specificity of 70% in differentiating neoplastic from non-neoplastic lesions. However, human observers performed better, with a sensitivity and specificity of 93.8% and 85.7%, respectively. Several subsequent studies, mainly from Japan, were developed using the endoscopic magnification. Although some of these studies reported promising results, the generalizability of these systems

is limited since magnification endoscopy is available only in few, highly specialized centers.

Endocytoscopy permits cellular visualization *in vivo* providing an ultra-magnification ( $\times 450$ ), which allows visualization of the nuclei. Several CAD systems for endocytoscopy (EC-CAD) were evaluated. The sensitivity, specificity, and accuracy for the identification of neoplastic colonic lesions ranged between 89.0%–92.0%, 79.5%–88.0%, and 81.0%–89.0%, respectively.<sup>23</sup> None of the systems presented in literature were able to show any significant difference when compared with expert evaluation. EC-CAD systems were also developed to predict invasive cancer from adenomatous lesions. A proper evaluation in this sense is crucial in selecting lesions where the endoscopic treatment is not appropriate.

The combination of EC-CAD with NBI was evaluated by Mori et al.<sup>24</sup> They reported a preliminary experience in 4 patients using a CADx-based (EndoBRAIN; Cybernet Systems Co., Tokyo, Japan) on the microvascular aspect at NBI images. These results were confirmed in a larger study by Kudo et al.<sup>25</sup> on 100 polyps from 89 patients. The authors performed a retrospective study wherein they compared the diagnostic performance of the EndoBRAIN to the diagnostic performance of 30 endoscopists (20 trainees and 10 experts). The endoscopists were asked to evaluate the images using white-light microscopy, endocytoscopy with methylene blue staining, and endocytoscopy with NBI. The EndoBRAIN was used to assess endocytoscopic images with methylene blue staining and NBI, but not with white-light. The accuracy of the EndoBRAIN and endoscopists in distinguishing neoplasms from non-neoplasms was the primary outcome. Pathology analysis was the gold standard. All the accuracy parameters (sensitivity, specificity, accuracy, positive predictive value [PPV], and negative predictive value [NPV]) of the EndoBRAIN with methylene blue were significantly greater than those of the endoscopy trainees and experts. The accuracy parameters of the NBI-EndoBRAIN were all significantly higher than those of the endoscopy trainees. However, when compared with the performance of the experts, only the sensitivity and NPV were significantly higher, while the other values were comparable. As previously mentioned for magnifying endoscopy, also the generalizability of these results is limited since endocytoscopy is available only in highly specialized centers.

More recently, optimized technologies were reported in the literature, all having promising accuracy parameters. For example, Chen et al.<sup>26</sup> designed a DNN CAD system to characterize diminutive polyps using NBI with optical magnification. Histology was the reference standard. They compared the polyp's characterization between the NBI-based CADx and novel and expert endoscopists. AI was faster

( $0.45 \pm 0.07$  sec) than experts ( $1.54 \pm 1.30$  sec) and novel endoscopists ( $1.77 \pm 1.37$  sec). It correctly classified the neoplastic histology with 96.3% sensitivity and 78.1% specificity. The accuracy was 90.1%. The system was able to better characterize the polyps than novel endoscopists and was comparable to the experts. Deep learning was also developed for standard colonoscopies (i.e., without magnification), with or without NBI. This represents a huge step forward since the algorithm may enrich instruments that are already used in endoscopic services to date. Diminutive polyps represent one of the most attractive areas of interest in the field of AI. The approach in case of small lesions ( $< 5$  mm) is still under discussion, and a unique management is far from standardization. AI may play a pivotal role in the identification of lesions that may benefit from “resect and discard” and/or “diagnose and leave” strategies. Byrne et al.<sup>27</sup> described a DNN model for the real-time assessment of colorectal polyps. The model was developed on videos containing colorectal polyps captured with 190 series Olympus (Olympus Co., Tokyo, Japan) colonoscopies using NBI. Video sequences were populated by polyps of varying sizes. Videos of the colonic mucosa without polyps were also used to train the model. The NICE classification was used to classify polyps and for training the deep learning machine. For testing, 158 consecutive diminutive polyps were included. Pathological examination was the gold standard. The accuracy, sensitivity, specificity, NPV, and PPV for the identification of adenomas was 94.0% (95% CI, 86.0%–97.0%), 98.0% (95% CI, 92.0%–100%), 83.0% (95% CI, 67.0%–93.0%), 97.0% and 90.0%, respectively. The system operates in real-time, with an unremarkable delay of 50 ms/frame. The same system was also used in the study by Shahidi et al.,<sup>28</sup> who aimed to estimate the discrepancy between endoscopic and pathologic diagnoses of lesions  $\leq 3$  mm. A total of 644 lesions  $\leq 3$  mm, diagnosed during optical evaluation as adenomatous, were included. Discrepancy between endoscopic and pathologic diagnoses occurred in 28.9% of the lesions. Overall, there was agreement between the model and the endoscopic diagnosis in 89.6% of the lesions. In those cases, where the endoscopic and pathologic evaluations were in disagreement, the model agreed with the endoscopic diagnosis in 90.3% of the lesions. In approximately 91.0% of the lesions identified on pathology as normal mucosa, the algorithm agreed with the endoscopic diagnosis (i.e., adenomatous lesion). The results of this study are relevant. In fact, although AI needs to be optimized, if these results will be confirmed by larger studies, AI may be used to arbitrate between endoscopic and pathologic diagnoses and offer an appropriate surveillance colonoscopy interval.

Other studies evaluated the role of CNN in distinguishing real-time diminutive polyps. Zachariah et al.<sup>29</sup> designed a

CNN developed to classify adenomatous polyps versus non adenomatous polyps (i.e., hyperplastic and sessile serrated polyps). The training data set was populated by 5,278 diminutive polyp images (3,310 adenomatous polyps; 1,968 serrated polyps). The CNN model was tested using NBI vs. WLE. Pathology evaluation was the gold standard. The overall accuracy, sensitivity, specificity, PPV, and NPV of the model for adenomatous polyps was 93.6% (95% CI, 92.9%–94.2%), 95.7% (95% CI, 95.1%–96.4%), 89.9% (95% CI, 88.6%–91.3%), 94.1% (95% CI, 93.3%–94.9), and 92.6% (95% CI, 91.5%–93.8%), respectively. The model showed a comparable accuracy when used with WLE and NBI: 91.9% (95% CI, 90.2%–93.6%) and 94.0% (95% CI, 93.2%–94.7%), respectively.

To summarize the results, a meta-analysis was recently published with the aim to evaluate the accuracy of AI on histology prediction and detection of colorectal polyps.<sup>21</sup> Overall, the pooled sensitivity and specificity on polyp histology prediction is 92.3% (95% CI, 88.8%–94.9%) and 89.8% (95% CI, 85.3%–93.0%), respectively. The AUC of the AI in the prediction of polyp histology was 0.96 (95% CI, 0.95–0.98). The sensitivity and specificity when dealing with diminutive polyps were 93.5% (95% CI, 90.7%–95.6%) and 90.8% (95% CI, 86.3%–95.9%), respectively. The NPV was 0.91 (95% CI, 0.89–0.94).

## ROLE OF ARTIFICIAL INTELLIGENCE IN INFLAMMATORY BOWEL DISEASE

Objective evaluations of ulcerative colitis (UC) based on a combination of endoscopic and histologic assessments are crucial in selecting the type of treatment and in monitoring the therapeutic response. Several studies showed that endoscopic patterns are predictive of the clinical outcomes of UC. However, endoscopic evaluations often differ between endoscopists where inter- and intra-observer variability is high, and biopsies are often performed for histologic evaluation of the disease activity. Recently, several studies have reported on an NN system to assess endoscopic severity in UC.<sup>30–33</sup> For instance, the potential role of AI for the evaluation of the inflammation in UC was investigated by Maeda et al.<sup>30</sup> The AI adopted in this study was first trained on 525 sets of 525 segments from 100 patients who underwent colonoscopy using endocytoscopy with biopsy. The system was then tested on 87 patients to predict histologic inflammation. It provided a diagnosis for 100% of the validation images. The processing time was 40 ms/frame. The overall sensitivity, specificity, and accuracy was 74.0% (95% CI, 65.0%–81.0%), 97.0% (95% CI, 95.0%–99.0%), and 91.0% (95% CI, 83.0%–95.0%), respectively. The CAD for segments

with a Mayo endoscopic score of 0 or 1 showed an overall diagnostic sensitivity, specificity, accuracy, PPV, and NPV of 65.0% (95% CI, 54.0%–75.0%), 98% (95% CI, 94.0%–99.0%), 91.0% (95% CI, 88.0%–94.0%), 87.0% (95% CI, 76.0%–94.0%), and 92.0% (95% CI, 89.0%–95.0%), respectively. Studies were also performed using algorithms applied to standard colonoscopies. For example, Ozawa et al.<sup>31</sup> performed a study aimed to evaluate if a CNN-based algorithm is reliable in classifying the endoscopic disease activity in patients with UC. The model was previously trained using 26,304 images from 841 patients with UC with different disease activities. The validation phase was performed on 114 patients with UC. The processing time was 20 ms/frame. Seventy-three percent of the Mayo 0 images, 70% of the Mayo 1 images, and 63% of the Mayo 2–3 images were correctly classified with appropriate Mayo scores by the CNN. The AUC was 0.86 (95% CI, 0.84–87) when differentiating Mayo 0 from 1–3 and 0.98 (95% CI, 0.97–98) when identifying Mayo 0–1 versus 2–3. The issue of grading the endoscopic severity of UC was also evaluated by Stidham et al.,<sup>32</sup> who aimed to define whether deep learning systems are able to grade the endoscopic severity of UC to the same accuracy as experienced endoscopists. The deep learning model was developed to categorize images into two clinically relevant groups: remission (Mayo 0 or 1) and moderate/severe disease (Mayo 2 or 3). The AI system performed similarly with the experienced endoscopists in grading the endoscopic severity of UC, being able to distinguish endoscopic remission from moderate-to-severe disease with an AUC of 0.966 (95% CI, 0.967–0.972), PPV of 0.87 (95% CI, 0.85–0.88), sensitivity of 83.0% (95% CI, 80.8%–85.4%), specificity of 96.0% (95% CI, 95.1%–97.1%), and NPV of 0.94 (95% CI, 0.93–0.95). Takenaka et al.<sup>33</sup> developed an AI system aimed to predict histologic remission from endoscopic images in UC patients. A DNN was trained with 40,758 images of UC colonoscopies. Targeted biopsies for surveillance were obtained, and the histologic data were linked to three consecutive images from biopsy site during colonoscopy. All this information was used to train the DNN, which learned the endoscopic images and their corresponding scores. For the validation phase, a total of 875 UC patients were prospectively enrolled. The accuracy of the system was comparable to that of endoscopists in evaluating mucosal inflammation in patients with UC. In fact, the model identified patients with endoscopic remission with 90.1% accuracy (95% CI, 89.2%–90.9%), using findings reported by endoscopists as the reference standard. In addition, the system was able to predict histologic remission without the need for biopsy. In fact, the model identified patients in histologic remission with 92.9% accuracy (95% CI, 92.1%–93.7%), using pathologic report as the reference standard. The Kappa coeffi-

cient between the model and the pathologic results was 0.859 (95% CI, 0.841–0.875).

## FUTURE PERSPECTIVE

Early attempts of CAD for colonoscopy have been proposed since 1990, but AI became feasible in real-life activities only with the most recent development of deep learning and CNN. Deep learning comprises complex AI systems composed of multiple processing layers designed to learn representations of data with multiple levels of abstraction. CNN uses a complex deep learning architecture composed of hierarchical feature representation of several layers (called “feature maps”) and “neurons” that connect with the adjacent neurons from the previous layers (receptive field). This system is able to perform filtering and pooling operations, learning automatically from the data to classify images and detect objects.<sup>34–36</sup> For a large-scale clinical application, an automatic polyp detection system should have high sensitivity, high specificity (i.e., low false-positive rate), near real-time response, and on-screen alerting system. A system with suboptimal specificity would have the potential consequences of being unreliable and disturbing for the endoscopist, while a low AI-related false-positive rate is likely to exclude any relevant detrimental effect of AI on withdrawal time. Conversely, an inadequate sensitivity would have an impact on the PDR. Moreover, real-time detection should be efficient and the time of analysis is preferably fast, with no perceptible delay to the endoscopist.

The clinical implications of AI in endoscopy and colonoscopy, in particular, are relevant when assuming that most of the adenoma miss rates at colonoscopy as well as variability in ADR between endoscopists are related to perceptual errors, owing to the fact that individual endoscopists may fail to recognize polyps. Factors that may explain such limitations include an inaccurate human visual perception, fatigue, distraction, and alertness during colonoscopy. AI appears as the best way of mitigating a suboptimal performance of colonoscopy, provided that polyps are actually visible in the monitor (Figs. 1 and 2). AI will not be able to compensate the risk of a suboptimal colonoscopy quality in case of a suboptimal exploration of the colorectal mucosa, inadequate level of cleansing, short withdrawal time, and/or inadequate colonoscopic technique. All these factors remain prerequisites to maximize the performance of AI. Several studies were recently published, all addressing the performance of different systems in polyp detection. Interestingly, overall, the number of diminutive adenomas detected by the AI was higher than that of the endoscopists. In the most recent RCTs, AI did improve the ADR and



PDR; however, it did not significantly increase the number of larger adenomas.<sup>15,17</sup> Indeed, it increased the number of diminutive polyps detected, including hyperplastic polyps, which have no clinical relevance. For example, Wang et al. confirmed that CADe was associated with an increase in ADR, PDR, and the mean number of polyps and adenomas per colonoscopy compared with the control group.<sup>15</sup> However, such increase was mainly due to higher detection of diminutive adenomas, suggesting that small polyps are more likely to be missed during colonoscopy rather than larger polyps. The authors also showed a higher detection of small hyperplastic polyps, which may lead to additional unnecessary polypectomies. Although diminutive adenomas have a low risk of neoplastic progression, the increase in overall ADR may result in: (1) eventually contribute to achieve a “clean” colon, (2) contribute to a decreased risk of interval cancer, (3) play consequences on the “protective” role of colonoscopy, and (4) have an impact on surveillance intervals.

In addition, when dealing with diminutive polyps, in the future, systems developed for polyp detection may be combined and integrated with a CADx system to complete all the processes, that is, to detect, diagnose, and disregard strategy to avoid unnecessary polypectomies. CADx can potentially allow a resect-and-discard strategy for diminutive/hyperplastic polyps and/or precluding unnecessary polypectomies, which has been estimated by Hassan et al.,<sup>19</sup> promising an increase of \$33 million dollars in savings per year in the United States alone. CADx can also help endoscopists by discouraging them from performing inappropriate endoscopic resections and suggesting to perform only targeted biopsies in case of advanced neoplastic lesions. In this sense, Byrne et al.<sup>27</sup> reported an accuracy and sensitivity for polyp characterization of 94% and 98%, respectively. Detection and characterization are considered a priority and represent the major fields of interest of AI in lower GI endoscopy.

As previously mentioned, the quality and technique of colonoscopy remain the prerequisites to maximize the performance of AI. AI is not intended to replace quality. Nevertheless, AI might be useful to help the endoscopists to improve the colonoscopy quality. AI has been initially designed for real-time histological classification and detection of polyps, with the aim to improve ADR and allow real-time polyp diagnosis. Limited research has been performed to evaluate the role of deep learning in monitoring or standardizing the endoscopists' technical skills. However, AI may be of paramount importance to monitor and guide the endoscopists in real-time to increase the overall colonoscopy quality. The ENDOANGEL was designed to guide the endoscopists in improving the colonoscopy quality and mitigate skill variations.

The final goal is to homogenize technical variations of endoscopists and alert them in case of suboptimal quality. Although preliminary, the study by Gong et al.<sup>18</sup> suggests how wide the field of application of AI could be, which will not be limited to polyp detection and characterization alone, but may interface with other issues, including training and quality monitoring. Small increments in the colonoscopy quality may have significant effects in any CRC screening program. In this regard, and looking at the positive effect on the ADR described in several trials, the use of AI might be an important quality assurance for screening colonoscopy.

The integration of AI in routine clinical practice will be challenging. For its widespread use, systems should not depend on specific light source, but they should be available across multiple scope manufacturers without requiring specialized scopes, apart from the processing unit connected to the video source and monitor. There will be regulatory and reimbursement difficulties to overcome. Medico-legal issues will also need to be addressed. There are only few recent studies on automatic polyp detection and characterization, and they have small sample sizes. They came from proof-of-concept projects, and systems rarely were tested in real-life situations. Performances of the systems achieved on still images or pre-recorded videos in a lab setting may not directly translate to live testing. Real-life data are needed. Evaluation of AI systems in a real-time colonoscopy requires rapid processing time that may not always be feasible or, at least, may not be feasible for all the systems described in literature. Trials have used heterogeneous study designs and outcomes. This is a consequence and reflects the figures involved in the technology development, where both the engineers, who design the software, and the clinicians, who tailor it for clinical use, play a complementary role. Preliminary studies frequently focused on per-frame detection rate. However, studies focusing on a clinical approach should emphasize the sensitivity, specificity, ADR, and PDR.

Only few studies are randomized. The interest is rapidly increasing and probably with the emergence of deep learning, we will assist important advances in the years ahead. If prospective clinical trials will confirm the preliminary data already published, the approach to diminutive colorectal polyps will be revolutionized by enabling the “resect and discard” and “leave distal” strategies for distal colon hyperplastic polyps. On the other hand, we should admit that the increased ADR observed in most of the trials might not be generalizable in settings with high baseline ADR, since evidence has shown that a second observer strategy is controversial when the ADR is high.<sup>37</sup> Standardization still remains an issue. The experts need to define thresholds. For example, the American Society for Gastrointestinal Endoscopy (ASGE) published a Preser-

vation and Incorporation of Valuable endoscopic Innovations (PIVI) statement that recommends technologies that perform real-time polyp classification to have at least 90% NPV for adenoma to be used in a “diagnose and leave” strategy.<sup>38</sup> Research should focus on the most clinically meaningful questions and should standardize outcomes. To adopt AI in the clinical practice, more clinical trial data coming from multicenter trial based on live colonoscopies will be needed. To our knowledge, with the exception of the GI Genius system that is already available in some countries, the majority of technologies are still in their infancy and have not yet proven to reach a sufficient diagnostic performance to be adopted in the clinical practice. Nevertheless, large players will enter the arena of AI in the coming months. The field will be demanding, but exciting.

## CONCLUSIONS

AI will hit considerably in medical applications. In the field of lower GI endoscopy, the evidence is mainly related to the detection and characterization of colonic lesions. Preliminary results are encouraging; however, larger, multicenter, randomized trials are needed to understand the real impact of AI in a real-life setting. In the coming months, new systems will be available and integrated into endoscopic towers. New areas of interest will emerge. The different technologies that are now separately developed to cover specific areas will hopefully be integrated into a single system.

### Conflicts of Interest

Cristiano Spada is a consultant for Medtronic, Norgine, and AlfaSigma, and received grants from Olympus and Pentax. The other authors have no potential conflicts of interest.

### Funding

None.

### Author Contributions

Conceptualization: Sebastian Manuel Milluzzo

Data curation: SMM

Formal analysis: SMM

Investigation: SMM

Methodology: Nicola Olivari

Project administration: NO

Resources: Paola Cesaro, Leonardo Minelli Grazioli

Software: NO

Supervision: PC, LMG

Validation: Cristiano Spada

Visualization: PC, LMG

Writing-original draft: CS

Writing-review&editing: SMM, CS

### ORCID

Sebastian Manuel Milluzzo:	<a href="https://orcid.org/0000-0002-6786-1783">https://orcid.org/0000-0002-6786-1783</a>
Paola Cesaro:	<a href="https://orcid.org/0000-0003-0899-9034">https://orcid.org/0000-0003-0899-9034</a>
Leonardo Minelli Grazioli:	<a href="https://orcid.org/0000-0002-3400-7820">https://orcid.org/0000-0002-3400-7820</a>
Nicola Olivari:	<a href="https://orcid.org/0000-0001-5004-3621">https://orcid.org/0000-0001-5004-3621</a>
Cristiano Spada:	<a href="https://orcid.org/0000-0002-5692-0960">https://orcid.org/0000-0002-5692-0960</a>

## REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
2. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993;329:1977-1981.
3. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. *Gastroenterology* 2014;146:709-717.
4. 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.
5. Burt RW, Cannon JA, David DS, et al. Colorectal cancer screening. *J Natl Compr Canc Netw* 2013;11:1538-1575.
6. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010;362:1795-1803.
7. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014;370:1298-1306.
8. Marcondes FO, Gourevitch RA, Schoen RE, Crockett SD, Morris M, Mehrotra A. Adenoma detection rate falls at the end of the day in a large multi-site sample. *Dig Dis Sci* 2018;63:856-859.
9. Gkolfakis P, Tziatzios G, Facciorusso A, Muscatiello N, Triantafyllou K. Meta-analysis indicates that add-on devices and new endoscopes reduce colonoscopy adenoma miss rate. *Eur J Gastroenterol Hepatol* 2018;30:1482-1490.
10. Morris EJ, Rutter MD, Finan PJ, Thomas JD, Valori R. Post-colonoscopy colorectal cancer (PCCRC) rates vary considerably depending on the method used to calculate them: a retrospective observational population-based study of PCCRC in the English National Health Service. *Gut* 2015;64:1248-1256.
11. Wang Y, Tavanapong W, Wong J, Oh JH, de Groen PC. Polyp-Alert: near real-time feedback during colonoscopy. *Comput Methods Programs Biomed* 2015;120:164-179.
12. Fernández-Esparrach G, Bernal J, López-Cerón M, et al. Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. *Endoscopy* 2016;48:837-842.
13. Misawa M, Kudo SE, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018;154:2027-2029.e3.
14. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018;155:1069-1078.e8.
15. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019;68:1813-1819.
16. 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.
17. Liu WN, Zhang YY, Bian XQ, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroen-*

- terol 2020;26:13-19.
18. 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.
  19. Hassan C, Wallace MB, Sharma P, et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. *Gut* 2020;69:799-800.
  20. Seibt H, Beyer A, Häfner M, Eggert C, Huber H, Rath T. Evaluation of a real-time artificial intelligence system using a deep neural network for polyp detection and localization in the lower gastrointestinal tract. *Gastrointest Endosc* 2020;91(6 Suppl):AB249.
  21. Lui TKL, Guo CG, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointest Endosc* 2020;92:11-22.e6.
  22. Tischendorf JJ, Gross S, Winograd R, et al. Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study. *Endoscopy* 2010;42:203-207.
  23. Ahmad OF, Soares AS, Mazomenos E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *Lancet Gastroenterol Hepatol* 2019;4:71-80.
  24. Mori Y, Kudo SE, Misawa M, Mori K. Simultaneous detection and characterization of diminutive polyps with the use of artificial intelligence during colonoscopy. *VideoGIE* 2019;4:7-10.
  25. Kudo SE, Misawa M, Mori Y, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. *Clin Gastroenterol Hepatol* 2020;18:1874-1881.e2.
  26. Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HH, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018;154:568-575.
  27. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94-100.
  28. Shahidi N, Rex DK, Kaltenbach T, Rastogi A, Ghalehjegh SH, Byrne MF. Use of endoscopic impression, artificial intelligence, and pathologist interpretation to resolve discrepancies between endoscopy and pathology analyses of diminutive colorectal polyps. *Gastroenterology* 2020;158:783-785.e1.
  29. Zachariah R, Samarasekera J, Luba D, et al. Prediction of polyp pathology using convolutional neural networks achieves "resect and discard" thresholds. *Am J Gastroenterol* 2020;115:138-144.
  30. Maeda Y, Kudo SE, Mori Y, 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.
  31. Ozawa T, Ishihara S, Fujishiro M, et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019;89:416-421.e1.
  32. Stidham RW, Liu W, Bishu S, et al. Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open* 2019;2:e193963.
  33. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology* 2020;158:2150-2157.
  34. Krishnan SM, Tan CS, Chan KL. Closed-boundary extraction of large intestinal lumen. In: *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; 1994 Nov 3-6; Baltimore (MD), USA. Piscataway (NJ): IEEE; 1994. p. 610-611.
  35. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-444.
  36. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90.
  37. Tziatzios G, Gkolfakis P, Triantafyllou K. Effect of fellow involvement on colonoscopy outcomes: a systematic review and meta-analysis. *Dig Liver Dis* 2019;51:1079-1085.
  38. 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.