Computer-Aided Polyp Detection

Contributor: Antonio Z. Gimeno-García, Anjara Hernández-Pérez, David Nicolás-Pérez, Manuel Hernández-Guerra

Artificial intelligence (AI) applied to medicine is here to stay. In gastroenterology, AI computer vision applications have been stated as a research priority. The two main AI system categories are computer-aided polyp detection (CADe) and computer-assisted diagnosis (CADx). CADe can detect previously unrecognized lesions and serve as "a second set of eyes" by continuously monitoring mechanisms. Furthermore, these systems may help to reduce in-day variations in ADRs due to operator fatigue.

---

1. Computed Aid Quality Improvement (CAQ)

Computer-aided quality improvement (CAQ) for real-time monitoring of withdrawal speed has been tested to increase ADR, with good results. Since the benefit of CADe systems depends on the amount of exposed mucosa, it would seem logical to assume that the combination of CADe and CAQ could increase the ADR. This hypothesis was tested in a recent randomized controlled study including 1076 patients assigned to a control group (regular practice), CADe, CAQ and the combination of CADe and CAQ. ADR was significantly higher in the combination group than in the CADe and control groups. Although the ADR was also higher in the combination group than in the CAQ group, the difference was not statistically significant. However, to improve the ADR, it makes sense that this type of technology should cover all dimensions of colonoscopy quality. Some studies have shown that tiredness can decrease the ADR in such a way that endoscopists working the full day detect fewer adenomas in afternoon colonoscopies. A recent study aimed to assess whether the use of AI systems (CADe, CAQ or the combination of both) may overcome the tiredness effect. A total of 1780 patients enrolled in two randomized controlled studies that compared AI systems with conventional colonoscopy were included. While in the conventional colonoscopy group, the ADR declined with each hourly interval (13.73% vs. 5.70%; p = 0.005; OR, 2.42; 95% CI, 1.31–4.47), it was not a significant factor in the AI system group (22.95% vs. 22.06%, p = 0.78; OR, 0.96; 95% CI; 0.71–1.29).

2. Computer-Aided Polyp Diagnosis (CADx)

Management of polyps is based on a precise optical diagnosis to choose the proper resection technique, either by endoscopy or surgery, and then carry out the subsequent histological analysis. The Preservation and Incorporation of Valuable endoscopic Innovations (PIVI) initiative was developed to assist the colonoscopic management of diminutive polyps located in the rectosigmoid colon based on the limited clinical importance of these polyps and with the main purpose of decreasing costs. However, even with the assistance of enhanced imaging techniques such as narrow band imaging (NBI), PIVI thresholds for resecting and discarding that prevent histological analysis of tiny hyperplastic lesions (90% agreement in the assignment of postpolypectomy surveillance intervals between histology and optical diagnosis) and leave in situ strategy and that give a recommendation for postpolypectomy endoscopic surveillance just after the endoscopic procedure (90% negative predictive value for adenomatous histology) are only met in expert centers.

One of the potential applications of AI in colonoscopy is to provide an accurate, almost instantaneous prediction of whether a polyp is neoplastic. With a high rate of histological prediction through AI, endoscopists could adopt a resect and discard strategy, obviating the need for complementary histopathological studies in tiny polyps. Some AI systems already exceed these thresholds in the experimental environment, combining CADx with NBI or endocytoscopy. Employed AI systems combining them with other advanced techniques, such as narrow band imaging, magnification, near-focus, blue light imaging (BLI), endocytoscopy, confocal endomicroscopy or laser-induced autofluorescence. These AI systems were trained with different types of information sources (still images and/or video images). They showed a negative predictive value for the detection of neoplastic lesions higher than the threshold value of 90% established for the diagnosis-and-leave criterion established by the ASGE.

However, caution should be exerted with the application of these AI systems in clinical practice given a lower diagnostic accuracy in the proximal colon compared with the sigmoid and rectum. In addition, internal and external validation of algorithms should be mandatory. It is also important to know how the AI system (training set quality) has been trained, concerning the number and type of lesions included according to their morphology and size. The number and qualification of the observers who classify the images is not trivial either. Unfortunately, this information is not always provided in the reported studies or by the companies trading these systems. Despite the complex and scarcely spreading endoscopic diagnosis techniques used in some studies (i.e., endocytoscopy, blue light imaging, autofluorescence), others using white-light endoscopy (WLE) also showed values higher than the threshold established by ASGE.

The theoretical benefits of this strategy would be cost savings, a shorter procedure duration and fewer adverse effects by avoiding unnecessary polypectomies. It has been suggested that these systems could decrease the learning curve in non-expert endoscopists. One study assessed whether a CNN trained for the characterization of colorectal polyps could increase the skills of endoscopists with different experience. The study found that the use of CNN led to an improvement in non-expert endoscopists (73.8% to 85.6%, $p < 0.05$) for colorectal polyp characterization, who almost reached the accuracy of experts (89.0%, $p = 0.10$).

The implementation of hybrid AI systems, packed into an "all-in-one" solution that includes both detection of lesions (CADe) and their characterization (CADx), should enable a higher detection rate of adenomas and serrated adenomas, compensated by a precise selection of candidate lesions for resection. This would result in a reduction in the incidence of PCCRC at reasonable costs. A cost-effectiveness analysis using a Markov model with microsimulation, which compared colonoscopy with and without AI for colorectal cancer screening for individuals at average risk, estimated an incremental gain of 4.8% in the reduction of CRC incidence when colonoscopy vs. colonoscopy plus AI were compared (44.2 vs. 48.9%), with a savings per individual of $57. Hybrid AI systems may be especially relevant in contexts such as colonoscopies by nonexpert endoscopists with low ADR or in fecal occult blood test-based screening programs where the prevalence of neoplastic lesions is higher. However, there are concerns regarding the usefulness of certain endoscopic technological innovations in CRC screening programs since they do not always increase the ADR when patients undergo high-quality colonoscopy by endoscopists with a high ADR.

3. Prediction of the Depth of Submucosal Invasion

Enhanced imaging techniques have been used for predicting submucosal invasion. The international classification NICE (Narrow band imaging International Colorectal Endoscopic classification) and JNET (Japanese Expert Team NBI classification) rely on the examination of surface and vascular patterns for predicting deep submucosal invasion. Unfortunately, training for this prediction is not easy; the prediction is endoscopist dependent and requires experience, and, even in expert hands, the sensitivity for some patterns is suboptimal. In two recent studies, the Type 2B pattern of the JNET classification only reached a sensitivity of 43–44% for differentiating high-grade dysplasia and shallow submucosal invasive carcinoma from deep submucosal invasion. Overall, the accuracy for deep submucosal invasion ranged from 59% to 84% across the studies. An inadequate diagnosis of deep submucosal invasion is relevant, leading to improper and harmful treatments, either due to overtreatment with the indication of surgery in patients with shallow neoplastic tumors (i.e., high-grade dysplastic tumors or with superficial submucosal invasion) or undertreatment, such as noncurative endoscopic resections with potential severe complications. AI may have the potential to overcome these issues and decrease interobserver variability.

There are no real-time trials in this setting thus far, and the published studies have aimed to develop and validate a convolutional neural network (CNN) to predict deep submucosal invasion. One limitation of these systems is that the prediction is usually based on still endoscopic images. Although some studies used images obtained with advanced technology such as NBI with magnification or endocytoscopy for training the CNN, this technology is not available in all endoscopy units. In general, they are single-center studies, and the composition of training sets is variable. A higher performance is achieved when the studies include advanced CRCs; however, the performance is lower for differentiating noninvasive or shallow invasive early CRCs from deep invasive early CRCs. The studies showed that AI is superior to trainees for predicting deep submucosal invasion but not to experts' assessment but seems to be faster. In a recent multicenter study, the authors designed and validated a CNN that included a multimodal data analysis (clinical information plus WLE endoscopy images plus enhanced endoscopy images) along with prior knowledge (distal location, large size, Paris type, surface morphology and the Narrow Band Laser Imaging International Colorectal Endoscopic classification). The accuracy of this system for predicting deep submucosal invasion was 90.4%. This type of combined system seems to work better than the sole deep learning model.

4. Assessment of the Colon Preparation

Several studies have been carried out to train and validate CNNs for detecting bowel cleansing during colonoscopy based on validated cleansing scales. These systems can also overcome the limitations of interobserver variability in rating colon cleansing during colonoscopy. ENDOANGEL is the only commercially available system for real-time assessment of colon cleansing. This system was trained with still images and colonoscopy videos, achieving an accuracy for bowel preparation assessment between 93.3% and 89.4% and better than the performance of endoscopists. ENDOANGEL was able to detect inadequate bowel cleansing in 100% of cases.

A recent study assessed in a randomized fashion the prediction of bowel cleansing before colonoscopy by a CNN trained with rectal effluents during bowel preparation. This approach is especially interesting since it has the potential to guide rescuing strategies before the colonoscopy (i.e., recommending additional bowel preparation in those patients with a CNN prediction of inadequate cleansing). In this study, patients were randomized to real-time CNN assessment of the rectal effluent after bowel cleansing intake and the prediction reported by patients following a set of images resembling different qualities of rectal effluents. There was no significant difference between the two predictions. CNN prediction was compared with the Boston Bowel Preparation Scale (BBPS) during the colonoscopy; however, the capability of discrimination between adequate (BBPS ≥ 6 points) or inadequate (BBPS < 6 points) bowel preparation was poor, detecting only 8.5% of the patients who finally had inadequate bowel preparation during the colonoscopy. Despite this, this type of tool lays the foundation for research on additional cleansing strategies carried out before colonoscopy.

5. Other Issues

The estimation of polyp size is an important issue since it can impact the therapeutic approach as well as the recommended surveillance intervals. Endoscopic estimation cannot rely on visual estimation, and usually different tools, such as open snares or forceps, are used. However, this process is time-consuming and is not always exact. Measures after formalin fixation are also inaccurate because of shrinking and fragmentation of the sample. A recent study aimed to assess the accuracy of an AI tool designed to measure polyp size. This measure relies on the distance between the main vessel branches. This approach proved to be a more accurate and reliable method of measurement than visual estimation or estimation assisted by open biopsy forceps during colonoscopy.

Another potential application of AI systems in colonoscopy is to accurately evaluate the location of the tip of the colonoscope during the examination. It is known that endoscopist assessment is unreliable. A recent study trained a CNN with images of a magnetic endoscope imaging positioning device. Although the predicted location dividing the colon into five segments was suboptimal (overall accuracy of 63%, sensitivity of 63%, specificity of 89%), this study set the foundation for further investigation in this field.