Review

Artificial intelligence and its impact on quality improvement in upper and lower gastrointestinal endoscopy

Pieter Sinonquel,1,3 Tom Eelbode,2,4 Peter Bossuyt,1,5 Frederik Maes2,4 and Raf Bisschops1,3

1Department of Gastroenterology and Hepatology, 2Medical Imaging Research Center (MIRC), University Hospitals Leuven, 3Departments of Translational Research in Gastrointestinal Diseases (TARGID), 4Electrical Engineering (ESAT/PSI), KU Leuven, Leuven and 5Department of Gastroenterology and Hepatology, Imelda Hospital, Bonheiden, Belgium

Artificial intelligence (AI) and its application in medicine has grown large interest. Within gastrointestinal (GI) endoscopy, the field of colonoscopy and polyp detection is the most investigated, however, upper GI follows the lead. Since endoscopy is performed by humans, it is inherently an imperfect procedure. Computer-aided diagnosis may improve its quality by helping prevent missing lesions and supporting optical diagnosis for those detected. An entire evolution in AI systems has been established in the last decades, resulting in optimization of the diagnostic performance with lower variability and matching or even outperformance of expert endoscopists. This shows a great potential for future quality improvement of endoscopy, given the outstanding diagnostic features of AI. With this narrative review, we highlight the potential benefit of AI to improve overall quality in daily endoscopy and describe the most recent developments for characterization and diagnosis as well as the recent conditions for regulatory approval.

Key words: artificial intelligence, endoscopy, lower gastrointestinal tract, quality, upper gastrointestinal tract

INTRODUCTION

IN THE LAST decades, there is a growing attention for artificial intelligence (AI) in biomedical sciences. Several medical AI applications are currently being investigated and clinically tested with the ultimate goal of improving quality of diagnosis in clinical practice.1–3 Image-based medicine and diagnosis like gastrointestinal (GI) endoscopy is prone to go hand in hand with AI applications. Within the field of GI endoscopy, AI is expected to have the potential to overcome certain clinical-human hurdles that are difficult to solve with the current available imaging techniques, hereby increasing the quality. AI seems to be applicable in many different fields within GI endoscopy, especially in detection of and differentiation between neoplastic and non-neoplastic lesions in both the upper and lower GI tract.4,5 For instance, optical diagnosis of (early) dysplasia in Barrett’s esophagus (BE) remains a struggle mainly for non-expert endoscopists. To overcome this hurdle several endoscopic enhancement techniques have been developed and validated with significant improvement when used by experts. However, in non-expert hands many BE lesions are missed.6 In colonoscopy, a widely used quality metric is the adenoma detection rate (ADR). The risk of interval colorectal cancer (CRC) and cancer mortality is inversely correlated with the ADR.7 Several factors influence the ADR of an endoscopist, so AI might be an additional tool to improve the ADR.8 To improve this scattered quality landscape, AI is expected to support all endoscopists and improve overall daily performances and quality.

In this physician–engineer co-authored narrative review article we summarize the literature of AI in both upper and lower GI endoscopy and how AI can improve quality of daily endoscopy.

THE UPPER GI TRACT (OVERVIEW, TABLE 1)

Barrett’s esophagus and esophageal adenocarcinoma

WITH A RISING incidence and prevalence of esophageal adenocarcinoma (EAC), the need of high-quality surveillance to identify early malignant lesions
is more prominent than ever. The Seattle protocol and the introduction of enhanced imaging techniques have facilitated diagnosis but with a remaining low sensitivity, sampling error and need of expert hands.9−11 The American Society of GI Endoscopy endorsed the use of these advanced imaging techniques to shift from random towards targeted biopsies only under specific circumstances; performance thresholds for optical diagnosis were set at a per-patient sensitivity of ≥90%, a negative predictive value (NPV) of ≥98% and a specificity of ≥80% for detection of HGD or EAC.12 These high demands in combination with the long learning curve makes that this only feasible for very experienced endoscopists. It is well known that 50% of endoscopists in community centers do not adhere to the guideline of following the Seattle protocol and that it is difficult to identify early lesions.6,13 Indeed, in up to 75% of cases expert endoscopists can identify visible lesions that are not found before referral to an expert center.6

Computer-aided detection (CADe) has the potential to support all endoscopists regardless of their experience in the detection of early neoplasia. Van der Sommen et al.14 reported a CADe for detection of early neoplasia in BE in comparison to four expert endoscopists resulting in a reasonable performance. The ARGOS project used a CADe with hand-crafted features based on color and texture showing a sensitivity 95%,14 Hashimoto et al.15 trained an AI system (different types of AI, see Appendix S1) based on 916 images of histologically proven BE with high grade dysplasia or a T1 stage EAC and 916 non-dysplastic BE images with good performance to binary outcome (dysplastic, non-dysplastic) with an accuracy of 95.4% per image. The binary classifier ran at a speed of 72 frames per second, allowing a real-time application. Other bench studies of similar CADe systems, studied in a variety of study designs, report similar performances as shown by De Souza et al.16 Recently, de Groof et al.17 validated a multistep trained CADe system (pretraining stage followed by a training phase and an internal validation phase) for external validation on two different datasets of both 80 images, showing a classifying accuracy of 89% (dysplastic vs non-dysplastic). Additionally, their system outperformed every one of the 53 expert endoscopists and indicated the optimal site for targeted biopsy in 97% and 92% of the cases in external validation datasets 1 and 2, respectively.

A major limitation of the previous mentioned studies is the offline, often retrospective, use of the CADe. Therefore, the research has pivoted more and more in the direction of prospective real-time or live studies of CADe in the last 2 years. In a pilot trial of de Groof et al.18 a CADe system for detection of early neoplasia in BE was used during endoscopy in 10 patients with non-dysplastic BE and 10 patients with confirmed dysplastic BE. During endoscopy three white light (WL) images were taken at every 2 cm level of the BE segment and were analyzed by the CADe system at the same time. This resulted in a per-level accuracy of the CADe system of 90%. In 75% of all levels, the system produced three concordant predictions. Ebigbo et al.19 validated their CADe system on 62 images with 36 images of early EAC and 26 normal BE images from 14 patients concurrently assessed by an expert endoscopist and with histopathological confirmation. The system showed a sensitivity of 83.7% and a specificity of 100%. In parallel the same group tested a CADe system for detection of EAC showing a sensitivity and specificity of 97% and 88%, respectively, for WL images and 94% and 80% for narrow band imaging (NBI) images.20 Furthermore, these performances significantly outperformed 11 of the 13 endoscopists for either WL, NBI or both images.

Esophageal squamous cell carcinoma

Pharyngeal and esophageal squamous cell carcinoma are frequent and often diagnosed in an advanced stage. Since early esophageal squamous cell carcinoma (ESCC) is often subtle or impossible to visualize with WL endoscopy (WLE), Lugol’s chromoendoscopy is the gold standard screening method in high risk patients. Endoscopic enhancement techniques as NBI facilitate the detection of ESCC without the use of iodine staining21,22; however in a recent randomized controlled trial, the specificity was rather low (42.1%) and resulted in a sensitivity of only 53% in inexperienced hands.23,24 So again, like in BE, detection appears to depend a lot on the endoscopist’s experience, a caveat that could potentially be resolved by CAD.

Several benchmark studies with a CADe system have been published in the last decade. Recently, CADe application for pharyngeal cancers has been emerging with promising results. Two Japanese research groups retrospectively showed good sensitivities (both >85%) of the CNN-based CADe detection of pharyngeal cancer on NBI images.25,26 Horie et al.27 developed a CADe for detection of ESCC based on WL images and compared it with histopathology. The CADe showed an excellent sensitivity of 98% and the ability of recognizing all seven lesions smaller than 10 mm. Cai et al.28 compared their CADe system (trained on 2428 WL images from 746 patients) with 16 endoscopists of all levels after interpretation of a validation set of 187 WL images from 52 patients. The CADe system showed a sensitivity and accuracy of 97.8% and 91.3%, respectively. The diagnostic accuracy of the high-level endoscopists (four endoscopists with an endoscopic experience of >15 years) was 88.8%, the mid-level endoscopists (six endoscopists with an endoscopic experience of 5–15 years) was 81.6% and the junior-level endoscopists (six endoscopists with an endoscopic experience...
Table 1  Overview of diagnostic challenges in gastrointestinal endoscopy and the potential interest of artificial intelligence within those fields and the corresponding main references

| Clinical indication | Clinical challenge | Opportunities for artificial intelligence | Main References |
|---------------------|--------------------|------------------------------------------|-----------------|
| **Upper gastrointestinal tract** | | | |
| Barrett esophagus (BE) | Differentiation of dysplastic vs non-dysplastic lesions | Classification dysplastic vs nondysplastic BE | de Souza et al.  
De Groof et al.  
Ebigbo et al. |
| | High burden of multiple sampling in with low yield Missed diagnosis of esophageal adenocarcinoma | Indication of targeted biopsy location Detection of neoplasia esophageal adenocarcinoma detection | Ishihara et al.  
Van der Sommen et al.  
Ebigbo et al.  
de Groof et al. |
| **Esophageal Squamous Cell Carcinoma (ESCC)** | High burden of Lugol staining for ESCC detection | ESCC detection | Horie et al.  
Cai et al.  
Zhao et al.  
Guo et al. |
| **Gastric cancer (GC)** | Differentiation cancerous vs non-cancerous lesion | Classification of cancerous vs non-cancerous lesion | Miyaki et al.  
Kanesaka et al.  
Hirasawa et al.  
Li et al. |
| | Difficult interpretation of invasion depth and impact on planning therapeutic endoscopy | Invasion depth determination | Nakagawa et al.  
Tokai et al. |
| **Helicobacter pylori (HP) Infection** | Optical detection and diagnosis of HP infection without random/targeted biopsies | Prediction of HP infection | Huang et al.  
Shichijo et al.  
Itoh et al.  
Nakashima et al.  
Zheng et al. |
| **Gastric ulcer** | Difficulty in differentiation of subtle gastric lesions (ulcers, tumors, polyps) | Recognition (vs GC) Prediction of recurrence | Namikawa et al.  
Wong et al. |
| **Lower gastrointestinal tract** | High miss-rates of colorectal polyps Difficult optical diagnosis of colorectal polyps | Polyp/adenoma detection Polyp characterization | Urban et al.  
Klare et al.  
Wang et al.  
Wang et al.  
Hassan et al.  
Repici et al.  
Takemura et al.  
Kominami et al.  
Byrne et al.  
Komeda et al.  
Sanchez-Montes et al.  
Kudo et al. |
| Colorectal cancer screening | High proportion of polyps’ size overestimation | Polyp size estimation | Suykens et al. |
of <5 years) was 77.2%. After referring to the CADx outcome, the average diagnostic performance of all endoscopists improved, predominantly in terms of sensitivity (74.2% vs 89.2%, \( P = 0.000 \)), accuracy (81.7 vs 91.1%, \( P = 0.000 \)).

Guo et al.\(^{29}\) developed a computer-aided diagnostic tool (CADx) trained on 6473 NBI images including precancerous, early ESCC and normal images and video data. A heat map for every input image was generated and in case the CADx system identified a suspicious lesion. The per-image sensitivity for all 1480 malignant images in 59 patients (dataset A) was 98.04% and the per-image specificity for all 5191 non-cancerous images from 2004 patients (dataset B) was 95.03%. With an analysis processing speed of 25 frames per second and a latency of less than 100 milliseconds (ms), the per-frame sensitivity (dataset C, 27 non-magnifying videoclips and 20 magnifying videoclips) and specificity (dataset D, 30 non-magnifying and three magnifying videoclips) was 91.5% and 99.9%, respectively. The group of Zhao et al.\(^{30}\) developed a CADe system for classification of intrapapillary capillary loops for detection and classification of ESCC on NBI magnifying images with a mean diagnostic accuracy of 89.2% and 93.0% at the lesion and pixel levels, respectively, which was significantly better than the endoscopists.

Quality of ESCC treatment can also be improved by better predicting submucosal invasion, which is a decisive parameter for patient selection. Two Japanese research groups recently published their data of a CADx system developed and trained to differentiate invasion depth of ESCC. Nakagawa et al.\(^{31}\) showed a sensitivity and accuracy of 90.1%, and 91.0%, respectively, for differentiation of pathologic mucosal and submucosal microinvasive from deep invasive cancers. These results were comparable with the outcome of the 16 experienced endoscopists. Tokai et al.\(^{32}\) retrospectively collected 1751 WL images of ESCC for training the CADx system and subsequently collected 291 WL images for independent validation for invasion depth of the ESCC by both the CADx system and 13 experienced endoscopists resulting in a CADx accuracy of 80.9% outperforming 12/13 endoscopists’ accuracy (accuracy of 73.5%; median range 58.4–82.4%).

### Gastric cancer

Cardia and noncardia gastric cancer combined are the fifth most frequently diagnosed cancer.\(^{33}\) Endoscopic diagnosis is difficult due to two major problems. First, early gastric cancer mostly shows a subtle depression or elevation with faint redness which hinders endoscopic recognition. Second, the prediction of the invasion depth in the gastric wall is hard but important in the treatment selection. Optical diagnosis using image enhancement techniques as NBI, flexible spectral imaging color (FICE) or blue light imaging (BLI) has shown to be useful but demand again a substantial expertise. CADx might help to overcome these problems.

The group of Miyaki et al.\(^{34}\) developed a CADx system for differentiation between cancerous and non-cancerous gastric lesions on magnifying FICE images, yielding a detection accuracy of 86% and a sensitivity of 85%. In 2015, the same group applied the same CADx to magnifying BLI images for a quantitative validation, generating a significantly greater average output of 0.846 ± 0.220 for cancerous lesions than that for reddened lesions or surrounding tissue.\(^{35}\) Further advances were made by Kanesaka et al.\(^{36}\) developing a CADe tool not only for detection but also delineation of the border between cancerous and non-cancerous gastric lesions with a sensitivity of 97% and a specificity of 95%. Hirasawa et al.\(^{37}\) were the first to report detection of early gastric cancer using a deep learning (DL)
model trained with 13,584 WL images instead of BLI, NBI or FICE images. CADe-analysis of a test set with 77 cancer lesions was achieved in only 47 s and resulted in a correct diagnosis of 72/77 lesions (sensitivity 92%) but a misdiagnosis of cancer in 161 non-cancerous cases (positive predictive value 31%). The same sensitivity and specificity results have also been recently shown in CADe models for detection based on NBI images by two other groups.\textsuperscript{38,39} CADx models for the diagnosis of the invasion depth of early gastric cancer have been developed by Kubota et al. and improved by Zhu et al., resulting higher sensitivity and specificity than those achieved with endoscopists’ visual inspection.\textsuperscript{40,41} Furthermore, three research groups recently developed and validated their DL models for recognition of gastric deformities (ulcers, cancer, polyps, erosions).\textsuperscript{42–44}

**Non-cancerous gastric disorders**

*Helicobacter pylori* infection

Huang et al. trained a CADe model with WL images of 30 patients and used parameters obtained from 74 patients to predict *Helicobacter pylori* (HP)\textsuperscript{45} with a sensitivity and specificity of 85% and 95%, respectively. Shichijo et al. constructed a convolutional neural network (CNN) for HP detection by using images that were classified according to their location in the stomach.\textsuperscript{46,47} Due to the locational information, the sensitivity, specificity and accuracy increased from 81.9% to 88.9%, 83.4% to 87.4% and from 83.1% to 87.7%, respectively, outperforming immunohistochemistry tests not histology, so this system only improves patient and location selection for biopsy taking not the diagnosis itself.\textsuperscript{46,48} A group of 23 endoscopists (six certified, nine relatively experienced, eight beginners) independently evaluated retrospectively the same images with an overall significant lower sensitivity, specificity and accuracy of 79.0%, 83.2% and 82.4%, respectively.

Recently, Itoh et al. showed their CADe system being able to detect and diagnose HP infection on WLE images with both a sensitivity and specificity of 86.7%.\textsuperscript{49} Consecutively Nakashima et al. improved sensitivity to both 96.7% using BLI and linked color imaging, respectively.\textsuperscript{50,51} Most recently the pilot study of Zheng et al.\textsuperscript{52} evaluated a on WL images trained CNN for HP detection showing a sensitivity, specificity, and accuracy for a single gastric image of 81.4%, 90.1%, and 84.5%, respectively. If multiple gastric images were used per patient, the diagnostic accuracy on a per-patient basis was very high with a sensitivity, specificity, and accuracy of 91.6%, 98.6%, and 93.8%, respectively, comparable to sensitivity and specificity of histological testing.\textsuperscript{38}

Artificial intelligence for patient management of gastric ulcer bleeding. Wong et al. used machine learning to derive a predictive score to identify patients with a high risk of recurring idiopathic HP negative gastric bleeding.\textsuperscript{53} The machine learning model predicted the 1-year risk of gastric ulcer bleeding recurrence with an accuracy of 84.3%.

**THE LOWER GI TRACT (OVERVIEW, TABLE 1)**

With the introduction of local and national screening programs, CRC mortality has declined by 32% among individuals aged $\geq$50 years between 2000 and 2014.\textsuperscript{54,55} However, tandem colonoscopy studies show colonoscopy to be an imperfect test with miss-rates up to 27% for diminutive polyps.\textsuperscript{56,57} ADR is a widely accepted and validated quality metric for screening colonoscopy and is defined as the percentage of colonoscopies performed by an endoscopist in which at least one adenoma is detected. ADR is inversely correlated with the risk of interval CRC, in a way that every increase of 1% in ADR is associated with a 3% decrease in risk of CRC and a 5% decrease in risk of fatal CRC.\textsuperscript{5} Many applications, both mechanical and optical, have been developed to increase ADR, all with shared shortcomings such as the need of specialized hardware and/or a long learning curve.\textsuperscript{58–61} This explains the keen interest for AI in colonoscopy as an adjunctive detection and diagnosis.\textsuperscript{52–64}

**Polyp detection**

In the last two decades a tremendous evolution in CADe for colorectal polyp detection has occurred, with a shift from hand-crafted algorithms to DL CNNs with or without transfer learning.\textsuperscript{65–69} Not only the technical aspect like processing time and data processing became better, also the diagnostic performances improved tremendously over the past decade. However, all of these systems were only tested on the bench.

Real-time use of CNN was first introduced in 2018 by Urban et al.,\textsuperscript{5} using a CNN algorithm pre-trained with images from ImageNet and subsequently trained and tested on multiple datasets of colonoscopy images and additionally tested the algorithm on a set of 11 quick withdrawal videos resulting in an accuracy of 96%. They also assessed the efficacy of their algorithm by comparing assessments of nine videos with CADe vs without CADe and showed that the assessment with CADe identified nine more polyps than the assessment without CADe (45 vs 36 identified polyps, respectively).

The most recent step in the AI evolution is CADe application in vivo during live colonoscopy. First Klare et al.\textsuperscript{70} prospectively studied CADe for polyp detection during live colonoscopy performed by a trained endoscopist
while a second observer monitored the CADe output. The system analyzed with an average delay of only 50 ms and achieved a polyp detection rate (PDR) of 51% and ADR of 29%, comparable to the endoscopist’s PDR of 56% and ADR of 31%. No additional detections were made by the system, hereby questioning the real-world applicability and additional clinical value of such a system. The first randomized controlled trial (RCT) comparing a CADe system to standard colonoscopy for polyp detection has been conducted by Wang et al. 1058 patients were randomized 1:1 to either standard colonoscopy or colonoscopy with CADe. The CADe system significantly increased the ADR (29.1% vs 20.3%, \( P < 0.001 \)) and the mean number of adenomas per patient (0.53 vs 0.31, \( P < 0.001 \)). The increase in ADR was due to a higher number of diminutive adenomas, while there was no statistical difference in adenomas larger than 10 mm. In addition, the number of hyperplastic polyps was also significantly increased (43.6% vs 34.9%, \( P < 0.001 \)) without a difference in advanced adenomas or sessile serrated polyps detected. The CADe system gave only 39 false-positive results (i.e. a continuous trace of the CADe system deemed by the endoscopist not to be a polyp) in the CADe group, which is on average 0.075 false positives per colonoscopy. Similar results were recently published by Wang et al. who conducted a RCT sham vs CADe, resulting in an increase of ADR of 23% (28% vs 34%) and only 0.1 consistent false-positive per colonoscopy.

The first commercially available CADe system for polyp detection (GI Genius, Medtronic, Table 2) was recently studied in a retrospective validation trial by Hassan et al., showing an excellent performance with a sensitivity per lesion of 99.7%. The reaction time of the CADe system was faster in 82% of the cases when compared to expert endoscopists. Moreover, Repici et al. evaluated the same CADe system in a prospective RCT with 685 patients showing an ADR of the CADe of 54.8% vs an ADR in the control group (without CADe) of 40.4% (relative risk 1.30, 95% CI 1.14–1.45) and a higher mean number of adenomas detected in the CADe group (1.07 ± 1.54 vs 0.71 ± 1.20). Adenomas of ≤5 and 6–9 mm were detected in a significant higher proportion in the CADe group.

These studies and performances highlight the full potential of AI in colorectal polyp detection as well as its possible impact on the future quality of colorectal endoscopy.

### Polyp characterization and diagnosis

Many research teams developed a CADx system designed for differentiation between adenomas and hyperplastic polyps based on vascular and surface patterns, like in the NICE classification. Kominami et al. tested prospectively in a pilot trial a CADx system showing accuracies of >92.7% in predicting a surveillance interval based on CADx diagnosis of diminutive polyps. In 2012 Takemura et al. and more recently (2019) Byrne et al. retrospectively tested both a DL CADx system for magnified-NBI polyp characterization that both met the PIVI-threshold for optical diagnosis (NPV >90% for diagnosis of diminutive polyps).76-77

White light endoscopy and CADx has not been so extensively investigated. Preliminary results have been published by Komeda et al., who developed a DL model providing an accuracy of 75.1%. Sanchez-Montes et al. developed a hand-crafted CADx based on contrast, tubularity and branching of the polyp’s surface resulting in a sensitivity of 95% for diminutive rectosigmoid adenomas. With a significant increase of diagnostic performances when compared to expert endoscopists in both stained endoscopic and NBI images a new CADx system, EndoBRAIN, has recently shown the potential to improve low identification rates of neoplasms without interobserver variability (Table 2).

For effective diagnosis, treatment and determining surveillance intervals, polyp size measurement is important. Often polyp size is overestimated, Suykens et al. reported an AI system that can objectively infer polyp size based on a reference tool (i.e. biopsy forceps) in the endoscopic image. They used two separate DL algorithms: (i) delineation of the polyp and (ii) detection of two landmarks on the forceps, resulting in a size estimation that can detect the polyp and forceps in 71% of the 35 test images. The trimmed average difference is +0.52 mm [Standard Deviation (SD) 1.78 mm] and +1.40 mm (SD 1.82 mm) between the ground truth and predicted size by the algorithm or endoscopist, respectively. This leads to a decrease in overestimation bias by 63% (\( P \)-value < 0.1).

### Inflammatory bowel disease and the small bowel

Assessment of endoscopic disease activity in inflammatory bowel disease (IBD) is important for patient management. Three different research groups focused on wireless capsule endoscopy for detection of ulcerative lesions using support vector machine or DL-CNN among Crohn’s disease (CD) patients, with the accuracy of these algorithms ranging from 89.3% to 96.7%. Endoscopic remission is a therapeutic target in ulcerative colitis (UC). In 2019, a CCN-model performed with high AUCs of 0.93–0.99 between Mayo 0–1 and Mayo 2–3 lesions on 3981 images of 114 UC patients. Another CNN by Takenaka et al. achieved an accuracy of 90.1% and 92.9% for the determination of endoscopic and
histological remission, respectively. Since a disconnection between endoscopic and histological disease activity exists, Bossuyt et al.87 constructed an algorithm (Red Density, RD) based on the integration of pixel color data from the redness color map along with vascular pattern detection with histological remission as gold standard. The RD score correlated well with both histological and endoscopic scores, providing an objective assessment of disease activity in UC. Maeda et al.88 retrospectively evaluated automated assessment of in vivo intramucosal capillary network and crypt architecture patterns on endocytoscopic images for the stratification of the risk of clinical relapse in UC patients with an MES of 1. By automation of endoscopic and histological scoring in UC the quality of clinical care can improve since it avoids variability in the interpretation of the endoscopic image.

Determination of disease activity is important to tailor IBD therapy, it remains to be seen how automated endoscopic scores can be incorporated in algorithms including biomarkers to finetune therapy,89–92 but at least the objectivity of these systems will improve endoscopic scoring of disease activity.

Small bowel capsule endoscopy (SB-CE) gained special interest in the field of AI since it remains a time-consuming procedure. Several research groups investigated their CNN-based SB-CE device, showing good results in terms of recognition and reducing time of analysis.93–95 Recently, Otani et al.96 developed and validated their SB-CE system for multiclass detection of erosions and ulcers, vascular lesions, and tumors, showing mean AUCs of 0.996 (95% CI, 0.992–0.999), 0.950 (95% CI, 0.923–0.978), and 0.950 (95% CI, 0.913–0.988), respectively.

### Automated assessment of quality in endoscopy

Although many of the abovementioned features in AI seem promising for future application in daily endoscopy, the performance of the best AI system will still largely depend on the technical skills of the endoscopists. Indeed, if a polyp is not presented to the AI system due to low quality colonoscopy technique, it will not be detected. Therefore, it is conceivable that AI systems need to incorporate quality control mechanisms (retraction time, inspection time stomach, 3D reconstruction, automated reporting photo documentation). First steps in this field have been made by Wu et al.97 showing a significant reduction in blind spots during esophagastroduodenoscopy (EGD) with the WISENSE system. Consecutively, this research group conducted a trial with ENDOANGEL, evaluating the difference of blind spot rate between sedated or unsedated patients and the use of a standard or unsedated ultrathin endoscopy, resulting in the lowest blind spot rate during conventional EGD in sedated patients.98

### DISCUSSION

In this review, we highlighted the potential benefit of AI to improve overall quality in daily endoscopy. AI is without any doubt an attractive option for standardizing

| Name                          | Approval | Manufacturer                                              | Date of release | Function                                      |
|-------------------------------|----------|-----------------------------------------------------------|-----------------|-----------------------------------------------|
| Upper gastrointestinal tract  |          |                                                           |                 |                                               |
| IRIS™                         | FDA      | NinePoint Medical Inc. (Bedford, MA, USA)                  | November 2018   | Esophageal image feature segmentation        |
| Lower gastrointestinal tract  |          |                                                           |                 |                                               |
| ai4gi®                        | CE-label | Olympus America – Satis Operations and Imagia (Center Valley, PA, USA) | June 2018       | Colorectal polyp detection and differentiation |
| EndoBrain®                    | PMDA     | Olympus and Cybernet System Co. (Tokyo, Japan)            | December 2018   | Optical diagnosis of colorectal polyps        |
| GI Genius®                    | CE-label | Medtronic (Minneapolis, MN, USA)                          | October 2019    | Colorectal polyp detection                   |
| DISCOVERY™                    | CE-label | Pentax Medical (Tokyo Japan)                              | January 2020    | Colorectal polyp detection                   |
| CADEYE®                       | CE-label | FUJIFILM Europe GmbH (Tokyo, Japan)                       | February 2020   | Colorectal polyp detection and characterization |

CE, Conformité Européenne; Co, Corporation; FDA, Food and Drug Administration, United States of America; GmbH, Gesellschaft mit beschränkter Haftung; Inc, incorporation; IRIS, Intelligent Real-time Image Segmentation™; PMDA, Pharmaceuticals and Medical Devices Agency, Japan.

© 2020 Japan Gastroenterological Endoscopy Society
endoscopy, which is inherently imperfect due to human error. Even though the abovementioned studies report promising results in terms of diagnostic performance, supporting evidence to actually proof the beneficial effect of CADx on the quality in upper or lower GI endoscopy is currently very sparse. We need more robust data from future trials with (i) an optimal study design with a prospective and comparative (CADx vs no-CADx) set-up, in either randomized controlled or single-arm setting, (ii) real-time use (in vivo), (iii) robust endpoints like ADR or miss-rates rather than technical AUCs and (iv) a multicenter non-expert setting to enable reproduction of the results. Additionally, questions about what to report, how to unify metrics, which datasets to use should be addressed by international societies to enable a uniform reporting and basis for comparison between all of the systems.

At this point, it is however reasonable to accept that at least in the colon for polyp detection, the current technical data on AI suggest that it could certainly serve as an adjunct to both detection and diagnosis of GI lesions. However, basic endoscopic skills (i.e. insertion, withdrawal, inspection, mucosal exposure) will remain and will probably become even more important to have an acceptable performance of AI. Therefore, maybe even more important than the detection “skills” of AI tools, is that this is associated with software that actually monitors the technical performance of the endoscopists and mucosal visualization. Recently Wu et al. developed AI-software to improve quality of mucosal exposure in the stomach, as did Oh et al. for colonoscopy, and this is most likely the way forward.97,99 The use of advanced imaging techniques for characterization, which are often the ground truth of AI systems, also requires thorough training to capture a stable endoscopic image. Thus, combination of the good (benchmark) diagnostic performances of the current AI systems for both upper and lower GI endoscopy and well-trained endoscopists may improve the quality of daily endoscopy. With this respect, AI itself can be a valid tool for future education and training of less experienced endoscopists or trainees, to give feedback on quality during the procedure. In the end, the AI tool may also serve to generate an automated endoscopy report with relevant photo documentation, standardized terminology and inclusion of all performance measures like defined by for instance ESGE.100

In conclusion, due to recent breakthroughs in AI the interest in CAD is gaining attraction as a novel approach to improve the quality of upper and lower GI endoscopy. The advantage of CAD in GI endoscopy can be anticipated given the high performances provided by the state-of-the-art technology of DL. With larger and more well-designed prospective trials, this novel technology for GI endoscopy will be implemented in clinical practice in the near future.

CONFLICT OF INTEREST

PB HAS RECEIVED financial support for research from AbbVie, Mundipharma, Pfizer, Janssen and Mylan; lecture fees from AbbVie, Takeda, Pfizer and Janssen; advisory board fees from AbbVie, BMS, Takeda, Hospira, Janssen, MSD, Mundipharma, Roche, Pfizer, Sandoz and Pentax. RB received speaker’s fees, consultancy and research support from Pentax, Fujifilm and Medtronic. Other authors have no COI to disclose.

FUNDING INFORMATION

RB AND TE are supported by a grant of Research Foundation Flanders. PS is supported by an unrestricted research grant from Pentax. FM is supported by KU Leuven internal grant C24/18/047 and by the Flemish Government (AI Research Program).

REFERENCES

1 Esteva A, Kuprel B, Novoa RA et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542: 115–8.
2 Baker JA, Rosen EL, Lo JY et al. Computer-aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detection architectural distortion. Am J Roentgenol 2003; 181: 1083–8.
3 Gulshan V, Peng L, Coram M et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316: 2402–10.
4 Van Der Sommen F, Zinger S, Curvers WL et al. Computer-aided detection of early neoplastic lesions in Barrett’s esophagus. Endoscopy 2016; 48: 617–24.
5 Urban G, Tripathi P, Alkalayi T et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy of the HHS public access. Gastroenterology 2018; 155: 1069–78.
6 Schölvinck DW, Van Der Meulen K, Bergman JIGHM et al. Detection of lesions in dysplastic Barrett’s esophagus by community and expert endoscopists. Endoscopy 2017; 49: 113–20.
7 Kaminski MF, Regula J, Kraszewska E et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010; 362: 1795–803.
8 Kaminski MF, Thomas-Gibson S, Bugajski M et al. Performance measures for lower gastrointestinal endoscopy: A European Society of Gastrointestinal Endoscopy (ESGE)
quality improvement initiative. *United Eur Gastroenterol J* 2017; 5: 309–34.
9 Smyth EC, Lagergren J, Fitzgerald RC, et al. Oesophageal Cancer Europe PMC Funders Group. *Nat Rev Dis Prim* 2018; 3: 1–44.
10 Reid BJ, Blount PL, Feng Z, et al. Optimizing endoscopic biopsy detection of early cancers in Barrett’s high-grade dysplasia. *Am J Gastroenterol* 2000; 95: 3089–96.
11 Falk GW, Rice TW, Goldblum JR, et al. Jumbo biopsy forceps protocol still misses unsuspected cancer in Barrett’s esophagus with high-grade dysplasia. *Gastrointest Endoscopy* 1999; 49: 170–6.
12 Sharma P, Savides TJ, Canto MI, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on imaging in Barrett’s Esophagus. *Gastrointest Endosc* 2012; 76: 252–4.
13 Abrams JA, Kapel RC, Lindberg GM, et al. Surveillance in the community setting in the United States. *Clin Gastroenterol Hepatol* 2009; 7: 710–36.
14 De Groof J, Van Der Sommen F, Van Der Putten J, et al. The Argos project: The development of a computer-aided detection system to improve detection of Barrett’s neoplasia on white light endoscopy. *Unit 2019; 7: 538–47.
15 Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett’s esophagus (with video). *Gastrointest Endosc* 2020; 91: 1264–71.e1.
16 de Souza LA, Palm C, Mendel R, et al. A survey on Barrett’s esophagus analysis using machine learning. *Comput Biol Med* 2018; 96: 203–13.
17 de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-learning system detects neoplasia in patients with Barrett’s esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology* 2020; 158: 915–29.e4.
18 de Groof AJ, Struyvenberg MR, Fockens KN, et al. Deep learning algorithm detection of Barrett’s neoplasia with high accuracy during live endoscopic procedures: A pilot study (with video). *Gastrointest Endosc* 2020; 91: 1242–50.
19 Ebigbo A, Mendel R, Probst A, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett’s oesophagus. *Gut* 2020; 69: 615–6.
20 Ebigbo A, Mendel R, Probst A, et al. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut* 2019; 68: 1143–5.
21 Nagami Y, Tominaga K, Machida H, et al. Usefulness of non-magnifying narrow-band imaging in screening of early esophageal squamous cell carcinoma: A prospective comparative study using propensity score matching. *Am J Gastroenterol* 2014; 109: 845–54.
22 Lee Y-C, Wang C-P, Chen C-C, et al. Transnasal endoscopy with narrow-band imaging and Lugol staining to screen patients with head and neck cancer whose condition limits oral intubation with standard endoscope (with video). *Gastrointest Endosc* 2009; 69: 408–17.
23 Muto M, Minashi K, Yano T, et al. Early detection of superficial squamous cell carcinoma in the head and neck region and esophagus by narrow band imaging: A multicenter randomized controlled trial. *J Clin Oncol* 2010; 28: 1566–72.
24 Ishihara R, Takeuchi Y, Chatani R, et al. Prospective evaluation of narrow-band imaging endoscopy for screening of esophageal squamous mucosal high-grade neoplasia in experienced and less experienced endoscopists. *Dis Esophagus* 2010; 23: 480–6.
25 Tamashiro A, Yoshio T, Ishiyama A, et al. Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks. *Dig Endosc* 2020; 32: 1057–65.
26 Kono M, Ishihara R, Kato Y, et al. Diagnosis of pharyngeal cancer on endoscopic video images by Mask region-based convolutional neural network. *Dig Endosc* Published online: 26 Jul 2020; https://doi.org/10.1111/den.13800
27 Horie Y, Yoshio T, Aoyama K, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019; 89: 25–32.
28 Cai S-L, Li B, Tan W-M, et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest Endosc* 2019; 90: 745–53.e2.
29 Guo L, Xiao X, Wu C, et al. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest Endosc* 2020; 91: 41–51.
30 Zhao Y-Y, Xue D-X, Wang Y-L, et al. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy. *Endoscopy* 2019; 51: 333–41.
31 Nakagawa K, Ishihara R, Aoyama K, et al. Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists. *Gastrointest Endosc* 2019; 90: 407–14.
32 Tokai Y, Yoshio T, Aoyama K, et al. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. *Esophagus* 2020; 17: 250–6.
33 Bray F, Ferlay J, Soerjomataram I, et al. 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.
34 Miyaki R, Yoshida S, Tanaka S, et al. Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement. *J Gastroenterol Hepatol* 2013; 28: 841–7.
35 Miyaki R, Yoshida S, Tanaka S, et al. A computer system to be used with laser-based endoscopy for quantitative diagnosis of early gastric cancer. *J Clin Gastroenterol* 2015; 49: 108–15.
36 Kanesaka T, Lee TC, Uedo N, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. *Gastrointest Endosc* 2018; 87: 1339–44.
Convolutional neural networks in the diagnosis of Helicobacter pylori infection based on endoscopic images: Preliminary experience. Clin Transl Gastroenterol 2019; 10: 1–6.

Wong GL-H, Ma AJ, Deng H et al. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. Aliment Pharmacol Ther 2019; 49: 912–8.

Cronin KA, Lake AJ, Scott S et al. Annual report to the nation on the status of cancer. Part I: National Cancer Statistics. Cancer 2018; 124: 2801–15.

Siegel RL, Miller KD, Fedewa SA et al. Colorectal cancer statistics, 2017. CA Cancer J Clin 2017; 67: 177–93.

Rex DK, Cutler CS, Lemmel GT et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. Gastroenterology 1997; 112: 24–8.

Bisschops R, Tejpar S, Willekens H et al. Virtual chromoendoscopy (I-SCAN) detects more polyps in patients with Lynch syndrome: A randomized controlled crossover trial. Endoscopy 2017; 49: 342–50.

Omata F, Ohde S, Deshpande GA et al. Image-enhanced, chromo, and cap-assisted colonoscopy for improving adenoma/neoplasia detection rate: A systematic review and meta-analysis. Scand J Gastroenterol 2014; 49: 222–37.

Wada Y, Fukuda M, Ohtsuka K et al. Efficacy of Endocuff-assisted colonoscopy in the detection of colorectal polyps. Endosc Int Open 2018; 06: E425–31.

Atkinson NSS, Ket S, Bassett P et al. Narrow-band imaging for detection of neoplasia at colonoscopy: A meta-analysis of data from individual patients in randomized controlled trials. Gastroenterology 2019; 157: 462–71.

Hassan C, Senore C, Radaelli F et al. Full-spectrum (FUSE) versus standard forward-viewing colonoscopy in an organised colorectal cancer screening programme. Gastroenterology 2019; 157: 462–71.

Mori Y, Kudo S, Berzin TM et al. Computer-aided diagnosis for colonoscopy. Endoscopy 2017; 49: 813–9.

Alagappan M, Glissen Brown JR, Mori Y et al. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. World J Gastrointest Endosc 2018; 10: 239–49.

Liedlgrubner M, Uhl A. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. IEEE Rev Biomed Eng 2011; 4: 73–88.

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

Wang Y, Tavanapong W, Wong J et al. Polyp-alert: Near real-time feedback during colonoscopy. Comput Methods Programs Biomed 2015; 120: 164–79.

Angermann Q, Berna J, Sánchez-Montes C et al. Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. In: MJ Cardoso, T Arbel, X Luo, eds. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Heidelberg, Berlin: Springer Verlag, 2017; 29–41.
68 Billah M, Waheed S, Rahman MM. An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. Int J Biomed Imaging 2017; 2017: 1–17.
69 Zhang R, Zheng Y, Mak TWC et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. IEEE J Biomed Heal Informatics 2017; 21: 41–7.
70 Klare P, Sander C, Prinzen M et al. Automated polyp detection in the colorectum: A prospective study (with videos). Gastrointest Endosc 2019; 89: 576–82.e1.
71 Zhang R, Zheng Y, Mak TWC et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. IEEE J Biomed Heal Informatics 2017; 21: 41–7.
72 Wang P, Berzin TM, Romek Glissen Brown J et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. Gut 2019; 68: 1813–9.
73 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.
74 Repici A, Badalamenti M, Maselli R et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020; 159: 512–20.e7.
75 Kominami Y, Yoshida S, Tanaka S et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointest Endosc 2016; 83: 643–9.
76 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.
77 Takemura Y, Yoshida S, Tanaka 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.
78 Komeda Y, Handa H, Watanabe T et al. Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: Preliminary experience. Oncology 2017; 93: 30–4.
79 Sanchez-Montes C, Sanchez FJ, Bernal J et al. Computer-aided prediction of polyp histology on white light using surface pattern analysis. Endoscopy 2019; 51: 261–5.
80 Kudo S, Misawa M, Mori Y et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol 2020; 18: 1874–81.
81 Suykens J, Eelbode T, Daenen J et al. Sa2012 automated polyp size estimation with deep learning reduces interobserver variability. Gastrointest Endosc 2020; 91: AB241–2.
82 Saito H, Aoki T, Aoyama K et al. Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc 2020; 92: 44–51.e1.
83 Charisis VS, Hadjileontiadis Vasileios S, Charisis LJ, Hadjileontiadis LJ. Potential of hybrid adaptive filtering in inflammatory lesion detection from capsule endoscopy images. World J Gastroenterol 2016; 22: 8641–57.
84 Klang E, Barash Y, Yehuda Margalit R et al. Deep learning algorithms for automated detection of Crohn’s disease ulcers by video capsule endoscopy. Gastrointest Endosc 2020; 91: 606–13.
85 Ozawa T, Ishihara S, Fujishiro M et al. Novel computer-aided diagnosis system for endoscopic disease activity in patients with ulcerative colitis. Gastrointest Endosc 2019; 89: 416–21.e1.
86 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: 2150–7.
87 Bossuyt P, Nakase H, Vermeire S et al. Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density. Gut Published online: 8 Jan 2020; https://doi.org/10.1136/gutjnl-2019-320056.
88 Maeda Y, Kudo S, Ogata N et al. Endocytoscopic intramuscular capillary network changes and crypt architecture abnormalities can predict relapse in patients with an ulcerative colitis Mayo endoscopic score of 1. Dig Endosc 2020; 32: 1082–91.
89 Waljee AK, Sauder K, Patel A et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. J Crohn’s Colitis 2017; 11: 801–10.
90 Waljee AK, Liu B, Sauder K et al. Predicting corticosteroid-free biologic remission with vedolizumab in Crohn’s disease. Inflamm Bowel Dis 2018; 24: 1185–92.
91 Dong Y, Xu L, Fan Y et al. A novel surgical predictive model for Chinese Crohn’s disease patients. Medicine 2019; 98: 1–10.
92 Jain S, Kedia S, Sethi T et al. Predictors of long-term outcomes in patients with acute severe colitis: A northern Indian cohort study. J Gastroenterol Hepatol 2018; 33: 615–22.
93 Ding Z, Shi H, Zhang H et al. Gastroenterologist-level identification of small-bowel diseases&nbv;and normal variants by capsule endoscopy using a deep-learning model. Gastroenterology 2019; 157: 1044–54.
94 Aoki T, Yamada A, Aoyama K et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc 2019; 89: 357–63.
95 Leenhardt R, Li C, Le Mouel J-P et al. CAD-CAP: A 25,000-image database serving the development of artificial intelligence for capsule endoscopy. Endosc Int Open 2020; 8: E415–20.
96 Otani K, Nakada A, Kurose Y et al. Automatic detection of small-bowel lesions on capsule endoscopy images using a
newly developed deep convolutional neural network. *Endoscopy* 2020; 52: 786–91.

97. Wu L, Zhang J, Zhou W et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagastroduodenoscopy. *Gut* 2019; 68: 2161–9.

98. Chen D, Wu L, Li Y et al. Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: A prospective, single-blind, 3-parallel-group, randomized, single-center trial. *Gastrointest Endoscopy* 2020; 91: 332–9.

99. Oh JH, Hwang S, Cao Y et al. Measuring objective quality of colonoscopy. *IEEE Trans Biomed Eng* 2009; 56: 2190–6.

100. Rey JF, Lambert R, Axon A et al. ESGE recommendations for quality control in gastrointestinal endoscopy: Guidelines for image documentation in upper and lower GI endoscopy. *Endoscopy* 2001; 33: 901–3.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of this article at the publisher’s web site.

Appendix S1 Overview of technical aspects.