Journal of Gastroenterology and Hepatology JGHF

doi:10.1111/jgh.15341

SPECIAL SECTION ARTICLE

Artificial intelligence in small bowel capsule endoscopy - current status, challenges and future promise

Xavier Dray,[*,†] Dimitris Iakovidis,[‡] Charles Houdeville,[*] Rodrigo Jover,[§] Dimitris Diamantis,[‡] Aymeric Histace[†] and Anastasios Koulaouzidis[¶]

[*]Sorbonne Université, Centre d'Endoscopie Digestive, Hôpital Saint-Antoine, APHP, Paris, France [†]ETIS UMR 8051 (CY Paris Cergy University, ENSEA, CNRS), Cergy, France [‡]Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece [§]Servicio de Medicina Digestiva, Hospital General Universitario de Alicante, Instituto de Investigación Biomédica ISABIAL, Alicante, Spain [¶]Endoscopy Unit, The Royal Infirmary of Edinburgh, Edinburgh, UK

Key words
algorithms, artificial intelligence, capsule endoscopy, deep learning, neural networks, small bowel.

Accepted for publication 5 November 2020.

Correspondence
Prof Xavier Dray, Sorbonne Université, Centre d'Endoscopie Digestive, Hôpital Saint-Antoine, APHP, 184 rue du Faubourg Saint-Antoine, 75012 Paris, France.
Email: xavier.dray@aphp.fr

Declaration of conflict of interest: Xavier Dray and Aymeric Histace are cofounders and shareholders of the Augmented Endoscopy. Xavier Dray has received lecture fees from Bouchara Recordati, Fujifilm, Medtronic, MSD, and Pfizer and has acted as a consultant for Alfasigma, Boston Scientific, Norgine, and Pentax. Rodrigo Jover has acted as a consultant for Alfasigma, CPP Pharmaceuticals, GI Supply, MSD, and Norgine. Anastasios Koulaouzidis has received research support from SynMed/IntroMedic and travel support from Aquilant/Jinshan and has been in advisory board meeting in the United Kingdom for Ankon.

Abstract

Neural network-based solutions are under development to alleviate physicians from the tedious task of small-bowel capsule endoscopy reviewing. Computer-assisted detection is a critical step, aiming to reduce reading times while maintaining accuracy. Weakly supervised solutions have shown promising results; however, video-level evaluations are scarce, and no prospective studies have been conducted yet. Automated characterization (in terms of diagnosis and pertinence) by supervised machine learning solutions is the next step. It relies on large, thoroughly labeled databases, for which preliminary "ground truth" definitions by experts are of tremendous importance. Other developments are under ways, to assist physicians in localizing anatomical landmarks and findings in the small bowel, in measuring lesions, and in rating bowel cleanliness. It is still questioned whether artificial intelligence will enter the market with proprietary, built-in or plug-in software, or with a universal cloud-based service, and how it will be accepted by physicians and patients.

Introduction

Capsule endoscopy (CE) is used worldwide mainly to investigate the small bowel (SB), and its use is endorsed by national and international guidelines for obscure gastrointestinal (GI) bleeding (OGIB), iron deficiency anemia, suspected Crohn's disease (CD), polyposis, and refractory celiac disease. Colon capsule has also become an acceptable option in patients for whom conventional colonoscopy is contraindicated, incomplete, or refused.[1]

Gastroenterologists spend 30–120 min to review and interpret full-length small-bowel capsule endoscopy (SBCE) recordings encompassing tens of thousands of images.[2] This task is tedious, often monotonous and demanding, as reading requires the physician to find a dedicated time slot without unwanted distractions.[3] Artificial intelligence (AI), on the other hand, affects every aspect of our life, including multiple health-care domains. It has therefore become obvious to caregivers, academic researchers, medical, and information technology (IT) companies that computer-aided diagnosis solutions are needed for medical imaging, for example, GI endoscopy and more specifically SBCE.[4] Among seven randomized trials of AI deep neural networks (DNNs) in medicine, six studies had been conducted in the field of GI endoscopy.[5] AI has entered the market with several commercially available solutions in colonoscopy and even one in CE.

This review looks into current and future perspectives of AI in CE, as well as potential barriers for their full implementation.

A summary of artificial intelligence methods for capsule endoscopy

During the last decade, there has been a significant increase in computational power, which has provoked a wave of advancements in machine learning (ML), especially with the introduction of deep learning (DL) (Fig. 1). ML is a subdomain of AI, where mathematical models are used to construct systems that can automatically learn and improve their performance based on data. Learning is performed in a supervised or unsupervised way (Fig. 2), using algorithms that iteratively try optimally adjusting the parameters of the models. Supervised learning is based on a set of ground truth data (called training data) that can be used as a gold standard to train ML systems. Once an ML system is trained, it can be used to infer decisions.[6]

The term DL has been devised to characterize ML for a particular class of artificial neural network (ANN) systems. Unlike the

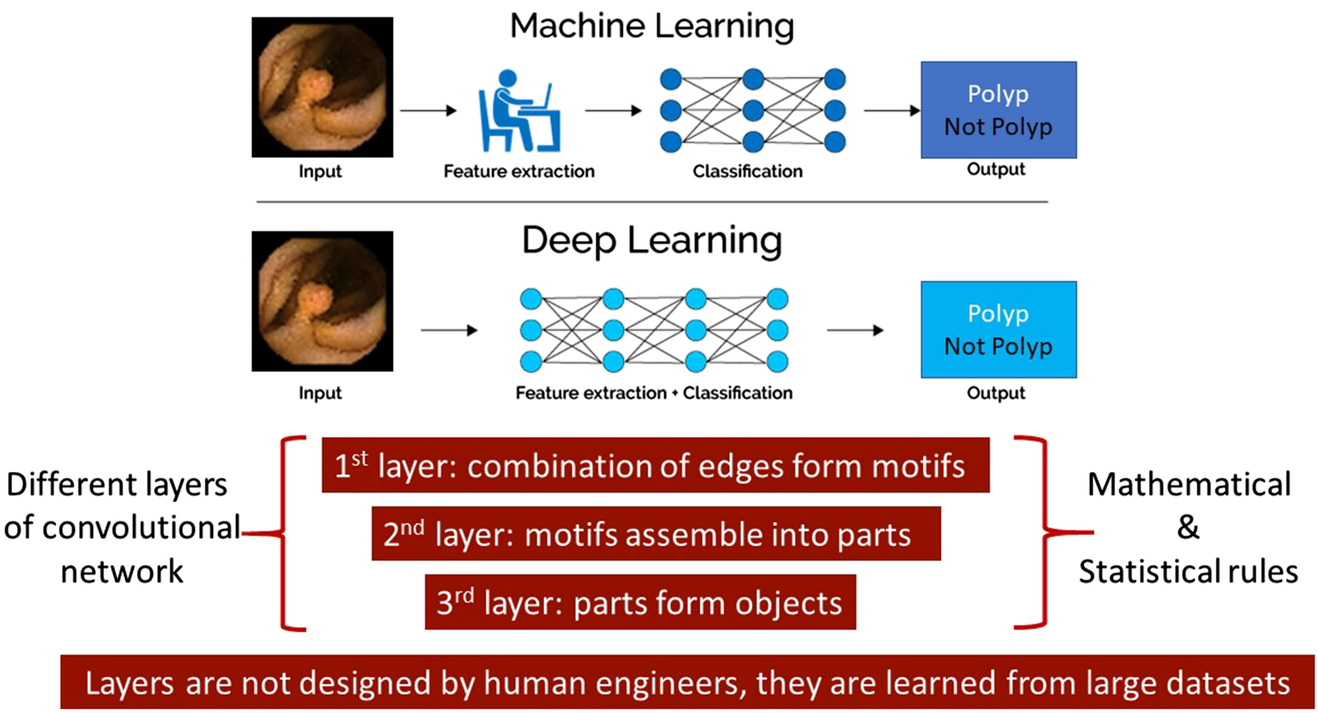

Figure 1 Schematic representation of machine learning and deep learning. Machine learning uses predefined engineered features, and experts collaborate making the extraction of data. In deep learning, mathematical algorithms permit the software to train itself to perform tasks and extract the data. In convolutional networks, different layers are working in defining the characteristics of an image; this network is trained with datasets from the desired image and, through mathematical and statistical rules, is finally able to identify these images by themselves. [Color figure can be viewed at wileyonlinelibrary.com]

conventional ANNs, this class consists of several neuronal layers, forming DNNs. The most popular DNN structure with a significant impact on medical image analysis is the convolutional neural network (CNN), which has a bioinspired neuron connection arrangement that resembles the organization of the animal visual cortex.[7] The main advantage of these networks over conventional ANNs is their capability to automatically extract image features, so that they can be used by computers to characterize their content. In previous approaches, feature extraction was based on mathematical formulas devised by humans (called "handcrafted features").

A drawback of supervised learning is that the training data should be provided by domain experts. In the context of CE, a standard annotation process constructing a training set includes the examination of each CE image, the delineation of regions of interest within the image using a graphic annotation tool, and labeling, preferably based on a standard terminology.[8] Considering that this process may be repeated for thousands of images, it becomes evident that it is a time-consuming and costly task.

To reduce the annotation effort, weakly-supervised ML algorithms have been proposed. This approach requires only image-level labels, indicating the categories in which their content belong to, without the need to manually delineate each regions of interest within each image.[9] Another approach to reduce the annotation effort with promising results in CE is transfer learning. This type of learning includes the application of ML systems that are pretrained on large datasets of nonendoscopic images for lesion detection in endoscopy images.[10]

The specific features of artificial intelligence in capsule endoscopy: lower time constraint, but higher sensitivity required

In conventional endoscopy, the operator is maneuvering the scope, adjusting it according to what is observed on the screen. In this setting, synchronous to the "human" look, state-of-the-art AI solutions provide further assistance for efficient detection (possibly characterization and/or therapeutic decisions) of lesions. Therefore, the main challenge for AI in this setting is to accurately display findings synchronously, that is, as fast as the frame capture rate (24–60 images per second).

Artificial intelligence in CE faces an almost reverse challenge. Although nowadays, a race for speed in every aspect of our life is evident, time constraint is not the main issue with asynchronous CE reading. The capsule's journey along the GI tract remains out the control of the operator; in a future, when AI will select a limited number of frames to be analyzed by the endoscopist (let us say 1%), the remainder 99% of frames will never get a human check. Overall, the primary aim of AI in CE is currently to reduce reading times while maintaining a high sensitivity for the detection of abnormalities.

Detection of lesions and abnormalities

This task has been among the first addressed in the development of AI in CE. Recent, most advanced evaluations are reported in Table 1.

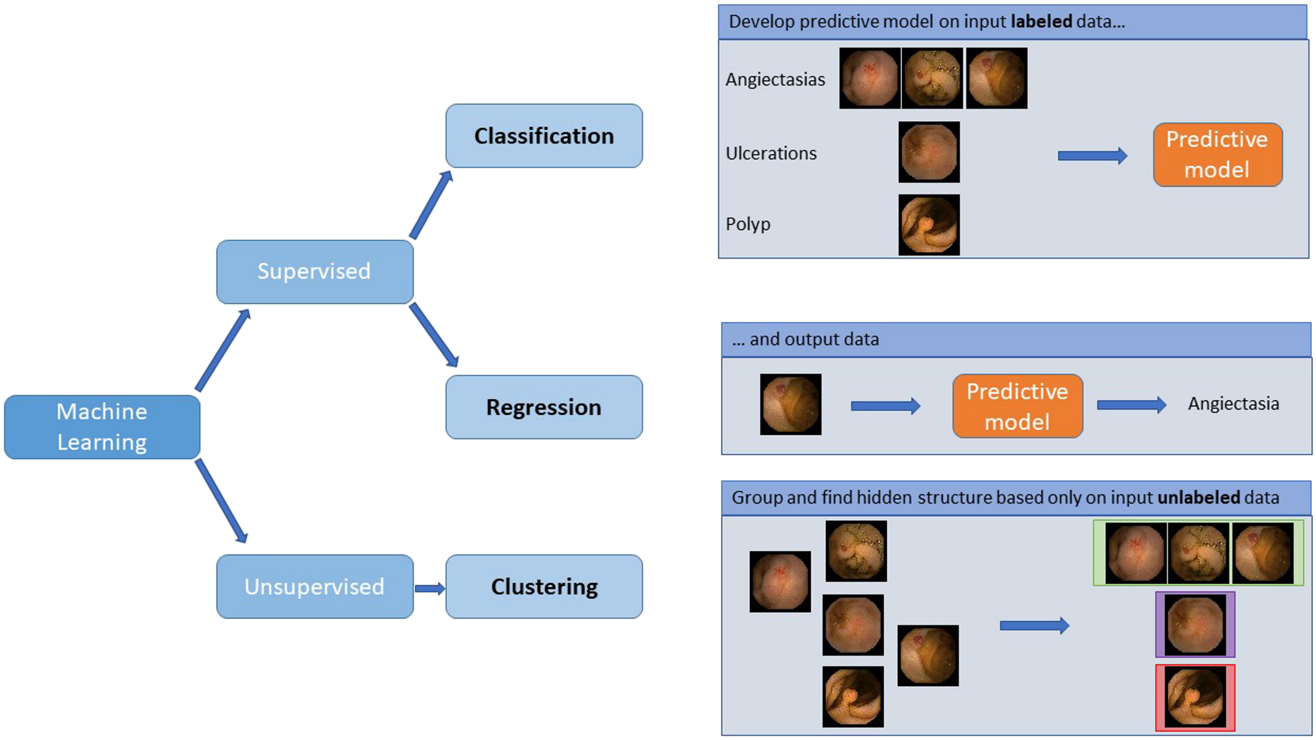

Figure 2 Principles of supervised and unsupervised machine learning. [Color figure can be viewed at wileyonlinelibrary.com]

Blood. In the pre-AI era, a proprietary Suspected Blood Indicator (SBI) performance had high sensitivity (96%) but poor specificity (variable 17–65%).[11] Since then, Xing *et al.*[12] proposed a strongly supervised DL method based on the analysis of superpixel-color histograms for classification of 500 frames with blood content and 500 normal frames. It outperformed earlier classifiers with sensitivity, specificity, and accuracy more than 98% and displayed a segmentation of bleeding zones within abnormal frames. Furthermore, Aoki *et al.*[13] developed a less supervised neural network-based algorithm based on PillCam SBCE still images. This AI solution was trained on 6503 images with blood and 21 334 without blood and then assessed on 208 images with blood and 10 000 without blood. It demonstrated a sensitivity of 96.6% (outperforming the SBI), a specificity of 99.9%, and an accuracy of 99.9%.

These results seem promising; there is though no evaluation at the video level. Although blood detection seems intuitive for human readers, AI does not seem ready for the task. The clinical approach to CE reading in bleeders is specific (and relies on video reading rather than still frames review). Therefore, clinical (and now AI) strategy in bleeders should be not only to detect blood but also to identify the location of bleeding and the lesion. Building databases for machine training (and testing) in this setting is challenging due to diversity (fresh blood *vs* melena, red blood *vs* dark clots, and traces *vs* massive bleeding) and need for videos (instead of a huge amount of still frames from a few patients, with a risk of overfitting).

Angiectasias. Gastrointestinal angiectasias (GIAs)—or angiodysplasias—are the most common lesions in SBCE. These well-demarcated lesions allow either a "strong" (pixel-level segmentation) or a "weak" (bounding boxes) manual labeling for AI development. Two proof-of-concept studies on still images paved the way to computer-assisted detection of GIA. In 2017, Noya *et al.*[14] used a database of 799 frames with GIA and 849 normal frames extracted from 36 different PillCam SBCE videos. They developed a deep feature extraction, CNN-based approach. Their system achieved a sensitivity of 89.5% and a specificity of 96.8%. Leenhardt *et al.*[15] further evaluated a similar approach in a larger population. A dataset of 1200 PillCam SBCE still frames (half with GIA from 208 patients and half without GIA from 200 patients) was split into two subsets (one for ML and one for testing). The algorithm yielded a sensitivity of 100% and a specificity of 96% for GIA detection. In 2020, Tsuboi *et al.*[16] worked on a dataset of 189 SBCE examinations. First, 2237 GIA frames captured from a training subset of 141 videos were put into a single-shot multibox detector architecture through a Caffe DL framework. Then, 488 GIA images from a different subset of 28 videos, plus 10 000 frames from 20 normal videos, were used for testing. This algorithm yielded a 98.8% sensitivity and 98.4% specificity. These promising results call for large prospective clinical trials at the video level.

Erosions, ulcerations, and ulcers. Erosions, ulcerations, and ulcers are seen in several SB conditions, including CD, celiac disease, drug toxicity, infectious, and neoplastic diseases, and are often surrounded by ill-limited, inflamed mucosa. Their diversity in depth, size, shapes, and etiology is challenging the research on AI in CE. In the study by Fan *et al.*,[17] a DL framework was trained and assessed on a large database (4910 still frames with

Table 1 Main recent studies on artificial intelligence for detection and/or characterization of lesions and abnormalities in capsule endoscopy

First author, (reference) year	Application	Number of patients/images	Design	Training & validation	Source of images	Performance measures
Xing X,[12] 2018	Blood	30 patients 1000 images	Retrospective	Yes	Still frames	Sensitivity: 98.5% Specificity: 99.5%
Aoki T,[13] 2020	Blood	41 patients 49 191 images	Retrospective	Yes	Still frames	Sensitivity: 96.6% Specificity: 99.9%
Noya F,[14] 2017	Angiectasias	36 patients 1648 images	Retrospective	No	Still frames	Sensitivity: 89.5% Specificity: 96.8%
Leenhardt R,[15] 2019	Angiectasias	408 patients 1200 images	Retrospective	No	Still frames	Sensitivity: 100% Specificity: 96%
Tsuboi A,[16] 2020	Angiectasias	48 patients 10 488 images	Retrospective	Yes	Still frames	Sensitivity: 98.8% Specificity: 98.4%
Fan S,[17] 2018	Ulcerations	144 patients 21 160 images	Retrospective	Yes	Still frames	Sensitivity: 96.8% Specificity: 94.8%
Aoki T,[18] 2019	Ulcerations	180 patients 15 800 images	Retrospective	Yes	Still frames	Sensitivity: 88.2% Specificity: 90.9%
		16 patients	Retrospective	No	Videos	3-min reading†
Wang S,[20] 2019	Ulcerations	1504 patients 47 202 images	Retrospective	Yes	Still frames	Sensitivity: 89.7% Specificity: 90.5%
Klang E,[21] 2020	Ulcerations	49 patients 17 640 images	Retrospective	Yes	Still frames	Sensitivity: 96.8% Specificity: 96.6%
Saito H,[22] 2020	Protruding lesions	385 patients 48 091 images	Retrospective	Yes	Still frames	Sensitivity: 90.7% Specificity: 79.8%
Iakovidis D,[23] 2014	Multiclass	251 patients 1370 images	Retrospective	Yes	Still frames	Sensitivity: 94.0% Specificity: 95.4%
Ding Z,[24] 2019	Multiclass	6970 patients 113 million images	Retrospective	Yes	Videos	Sensitivity: 99.9% Specificity: 97.0% 6-min reading†
Otani K,[27] 2020	Multiclass detection + characterization	167 patients 39 963 images + 288 patients	Retrospective	Yes	Still frames Videos	Mean AUROC: Ulceration = 0.93 Angiectasias = 0.88 Protrusion = 0.90

†Mean reading time.
AUROC, area under receiver-operating characteristic curve.

erosions and 3250 with ulcers). The proposed computer-assisted detection solution showed excellent performance, with sensitivity 96.8% and 93.7% and specificity 94.8% and 96.0%, for ulcer and erosion detection, respectively. Aoki *et al*.[18] further used 5800 still images with erosions and ulcers, from 180 PillCam videos. Their architecture demonstrated a sensitivity of 88.2% and a specificity of 90.9%, with lower detection rates related to the smaller size and to poor preparation. The same authors evaluated their system in a pilot series of 16 videos with 37 mucosal breaks and 4 videos without abnormal findings. The experts' reading times significantly decreased from 12 to 3 min, whereas detection rates did not change (87% and 84%, respectively).[19] To overcome the difficulties related to small-sized or ill-limited ulcerated lesions, Wang *et al*.[20] proposed a refined, two-stage detection approach. They used a dataset of 22 738 still images extracted from 1504 Ankon SBCE videos and demonstrated an overall sensitivity of 89.7% and a specificity of 90.5%. For CD lesions, Klang *et al*.[21] developed and assessed a CNN-based detection solution. The dataset comprised 7391 ulcerated still frames with ulcerated lesions and 10 249 normal ones, extracted from 49 PillCam videos. Sensitivity ranged between 92.5% and 96.8%, and specificity between 96.6% and 98.1%. Overall, detection solutions for ulcerated lesions still need to be refined and to be better evaluated at the video level.

Protruding lesions. Detecting protruding lesions is an important challenge for AI in CE. After preliminary studies from IT teams suggested encouraging results, Saito *et al*.[22] collected 30 584 images of such lesions from 292 patients to train a CNN-based solution. The computer-aided diagnosis system was then tested on 7507 images with protruding lesions from 73 patients and on 10 000 normal frames from 20 patients. Sensitivity was good (90.7%), and specificity was fair (79.8%). Although the proof-of-concept is now close to be achieved, there is no study available at the video level on the critical matter of automated detection of protruding lesions. Characterizing these lesions as bulges, polyps, nodules, mass, venous, or lymphatic structures, relative to various benign or malignant conditions, will be even more challenging.

Multiclass detection. In 2014, Iakovidis *et al*.[23] proposed a supervised methodology capable of learning from normal and abnormal frames and therefore capable of detecting several different types of lesions. The mean average performance, as the area under the receiver-operating characteristic curve, reached 0.892. The best average performance was obtained for angiectasias (0.975) and nodular lymphangiectasias (0.963). Few years later, Ding *et al*.[24] using the DL model and an impressive number of CE images (> 113 million) collected from 77 Chinese SBCE centers achieved gastroenterologist-level identification of SB diseases and normal variants in CE. The algorithm used in this study had an error rate of only 3% in discriminating abnormal findings from normal frames. The most notable fact though was that the mean reading time per video was 97 min by conventional reading and 6 min by CNN-based auxiliary reading.

Characterization

Once a SB lesion is detected within a video, it is important to categorize it in terms of diagnosis and clinical relevance. Whatever for human readers of today or for AI solutions tomorrow, any decision for further treatment will rely on these factors.

At the video level, only very few contributions have been reported in the last 3-year literature. Recently, Leenhardt *et al*.[15] proposed to use a pretrained U-Net architecture to segment GIAs. They showed that a precise segmentation can be obtained for the learning stage and could lead to pixel-wise multiple labeling of CE images paving the way for feature extractions like size, type, and depth, which could be then related to clinical parameters. Major limitations of this kind of approach are in their very demanding computational resources for both the learning and the segmentation tasks.

A higher level of supervision of neural network should allow current technology to face the challenge of classifying lesions in terms of clinical relevance. However, the main barrier for this development today is more on the clinicians' side: Any ML is illusive before the reproducibility of the interpretation of SB findings is high. Unfortunately, both the interobserver concordance rates regarding the diagnosis and the pertinence of SB lesions are around 60%.[25] This is probably because numerous types of findings are seen in the SB (including many "normal variants"). Furthermore, most SB findings found in CE are rarely sampled for pathological examination. Another issue is that vocabulary varies a lot and even more so within each specific category (an ulceration can be "aphthoid," "superficial," "deep,"...).[8] Lastly, because the interpretation depends on the clinical setting, GIA are relevant in OGIB, but not for suspected CD. Overall, although essential for ML and for evaluation of AI solutions, any "ground truth" regarding categories in SB lesions remains debatable. A European group addressed this challenge by proposing experts' consensus on the nomenclature, description, and pertinence of vascular, ulcerated and protruding lesions, according to the main indications of SBCE.[26] Future databases and AI solutions should include these categories for better characterization and reporting.

Combined multiclass detection and characterization

Combining a "weakly-supervised" approach for detection—as proposed by others[23,24]—with different attached subnetworks for classification detailed in their prior works,[16,18,19,22] Otani *et al*.[27] have recently reported on a solution with mean area under the receiver-operating characteristic curve of 0.996 for ulcerations, 0.950 for vascular abnormalities, and 0.950 for protruding lesions, based on still images. Performances were slightly lower (0.928, 0.884, and 0.902, respectively), but still outstanding, on an independent external validation dataset. Again, prospective evaluations at the video level are expected before translation in clinical practice.

Localization and size measurement

Localizing abnormalities in the GI tract is essential to guide clinical decisions.[28] Commercially available methods include wearable external sensor arrays to communicate with the capsule. Recent solutions, promising more comfortable CE localization, are based on the analysis of CE image sequences using ML. Initial techniques include video segmentation capable of automatically dividing the CE image sequence into consecutive segments with coherent topographic content corresponding to the different parts of the GI tract. These methods are based on the automated recognition of the respective tissues and anatomical landmarks, using color and texture handcrafted image features.[29] Such approaches can be combined with AI-enabled methods for CE motion estimation, to achieve more accurate CE localization. These can track the motion of the CE by tracking image features frame by frame. In recent works, ANNs have been used to accurately estimate the motion of the CE in physical units, based solely on image features.[30,31] The main challenges for this approach include the motility of the GI tract and the presence of intestinal content, which can affect the localization performance. Such methods have also been used as a basis to develop other useful methods in GI endoscopy, including the reconstruction of the GI tract for visualization,[32] and the precise measurement of size of a lesion *in vivo* during a CE examination.[33]

Assessment of bowel cleanliness

Bowel cleanliness is a crucial point and different scales are now adopted for colonoscopy. In the context of SBCE, two most recent AI-based algorithms have been proposed.[34,35] Both showed that AI can lead to a robust and reliable cleanliness metric at image level. A recent contribution shows that the principle of DL approaches can be extended to the video level.[35] In the context of SBCE, two most recent AI-based algorithms have been proposed (Table 2).

Limitations

Several important limitations of the current studies on AI in CE should be acknowledged. First and foremost, these studies are retrospective. Second, most datasets for AI training comprise selected still images, with an inherent risk of overfitting. Third, validation studies are mostly performed on datasets and rarely comprise randomly extracted SB video sequences.[18,24] Fourth, external validation studies are scarce (if none) possibly because research teams use different devices, networks architecture, and images (with various contrast, resolution, and labels).

Needs

It was noted early in the clinical implementation history of CE that there is a moderate degree of interobserver discrepancies on the interpretation of videos, and it has been suggested that a second

Table 2 Main recent studies on artificial intelligence for assessment of the cleanliness of small bowel during capsule endoscopy

First author,(reference) year	Ground truth/outcome	Device	Training set	Testing/validation sets	Performance measures in validation set
Noorda R,[34] 2020	Nonvalidated, self-developed, 2-item to 4-item scale, scored by two independent experts (no adjudication)	Pillcam SB3, Medtronic, USA	563 images (augmented to 55 293) from 35 videos	854 images from 30 videos	Agreement-based intraclass coefficient: 0.82
Leenhardt R,[35] 2020	Validated, 2-time to 10-time scales, scored by three independent experts (with adjudication when needed)	Pillcam SB3, Medtronic, USA	600 images (augmented to 3000) from 30 patients	156 full-length videos	Accuracy: 89.7% Sensitivity: 90.3% Specificity: 83.3%

reading by an experienced viewer might improve the diagnostic accuracy of the modality.[36] A multicenter study a few years later confirmed that there is substantial agreement between experts and only moderate agreement between trainees.[25] To achieve higher accuracies and better interobserver agreement, we need not only more experience with CE but also consensus regarding CE terminology. However, in 2012, Rondonotti *et al*.[37] showed that both the interobserver agreement and the detection rate of significant findings are low, regardless of the readers' experience. Furthermore, the training scheme tested in this study did not significantly increase the performance of readers with different experience. It should be noted though that the aforementioned studies were all of small-size and there was never a standard and diverse database. Recently, Beg *et al*.[38] showed that CE reader's accuracy declines after reading just one capsule study. This has potential implications for clinical practice, with the experienced readers in this group reporting their normal practice is on average of 3.4 capsule readings per clinical session. Furthermore, neither subjective nor objective measure(s) of fatigue were sufficient to predict the onset of the effects of fatigue. The perspective here is obvious: There is simply no favorable solution that involves humans, either in the form of prereaders or CE validators that could beat applications of AI in CE. There is currently immense activity from all CE manufacturers to provide self-reporting capsules. The first leap forward was done with in 2019, when a large multicenter Chinese study confirmed that a CNN-based algorithm was able to accurately identify abnormalities in CE images and able to drastically reduce the CE reading times to around 6 min per video.[24] Recently, we alluded to the fact that CE is open to a high level of scrutiny and medicolegal challenge, as the recorded video data can be readily accessible for further on-demand review.[3] AI solutions are not immune from such audits.

Nevertheless, computer-assisted analysis of CEs is faced with the curse of clinical databases, that is, not easy to collect and/or annotate. Further real challenges for the close future are (i) how to collect more and more relevant annotated data?, (ii) how to deal with currently accessible databases to develop robust and reliable AI algorithms?, and (iii) can we consider very recent DL approaches coming from the deep metric learning field to propose new methods that only need limited extra work to reach very good classification performance? For instance, active learning is introduced for colonoscopy as a strong alternative to manual annotation by proposing automatic generation of hard-negative examples[39]; classic data augmentation can be used but few/one-shot learning strategies introduced in deep metric learning context are very promising strategy to tackle classification problems only by using one or two examples of the lesions with no need for huge numbers of annotated data.[40]

Perspectives

Although initial comparison of a panoramic capsule (CapsoCam SV1; CapsoVision) with the traditional axial viewing CE (PillCam SB3; Medtronic) showed that panoramic CE detected more lesions, relevant bleeding sources were visualized by both types of capsules.[41] A subsequent, single-center, British study confirmed that SB diagnostic yield was comparable in the panoramic and axial viewing groups in patients with overt OGIB, but axial viewing offered a better gastric diagnostic yield than panoramic.[42] Recently, we reported the results of multicenter double-headed study, in 204 patients. We found that the use of double-headed CE provides more information, which has the potential to change clinical diagnosis and therefore patient management.[43] Afterall, the capsule being a nonsteerable device can only record what the bowel allows it to do; hence, more coverage will eventually provide better results if the study is properly powered. However, a major downside in this would be the longer time required to go through the recorded images per head (here, the use of dual mode can be confusing).

Unilateral development of AI systems ignores the needs and expectations of patients who are perhaps the most important stakeholders. AI systems need to fulfill certain preconditions for this technology to be embraced by society. Going beyond the efficiency of AI in detecting and characterizing lesions in CE, patients' perspectives focus on other items such as accountability, data protection, opaque decision-making (also known as "AI blackbox"), personal interaction, procedural knowledge, or the way patients are informed of the results obtained by means of AI. As mentioned above, a specificity of AI in CE is that, at the end of the line, there will be no human synchronous check of supposedly normal images (as proposed for colonoscopy). In a questionnaire developed with the aim of knowing which are the patients' perspectives in the use of AI for radiology (which, in this matter, is quite similar), patients mention they prefer humans when research shows that humans and computers are equally skilled in performing their job. However, when scientific research shows that computers are indeed superior to humans, most patients indicate that they would rather let the computer do the work than the radiologist.[44] Another specific issue is the potential for bias by the AI program towards certain population subgroups if inappropriate sampling and training of the algorithms has occurred.[45] In the setting of CE, how an AI solution developed and evaluated in Europe or America would translate in Asia, and the other ways around? Finally, it is also important to consider data protection issues, when computers are directly involved in diagnosis.

Conclusion. Twenty years after its revolutionary implementation in GI endoscopy, CE is about to live its golden age thanks to the development of AI. Automated detection solutions are now commercially available. Prospective controlled trials by independent teams are still needed before physicians can fully rely on such solutions and significantly reduce their reading times while maintaining (and possibly improving) detection accuracy. When this first critical step is taken, automated characterization (in terms of diagnosis and relevance) will be the second. It is still questioned how AI will enter the market, either introduced by a capsule company (with a proprietary, built-in or plug-in software, for example) or by an IT company (with a universal cloud-based service, for instance). The journey is much longer however, as the technology of CE is following its own path, dealing with panenteric examinations, virtual chromoendoscopy, active locomotion, and therapeutic options, numerous topics where AI may also play a role.

References

1 Rondonotti E, Spada C, Adler S *et al*. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) technical review. *Endoscopy* 2018; **50**: 423–46.

2 ASGE Technology Committee, Wang A, Banerjee S *et al*. Wireless capsule endoscopy. *Gastrointest. Endosc.* 2013; **78**: 805–15.

3 Rondonotti E, Pennazio M, Toth E, Koulaouzidis A. How to read small bowel capsule endoscopy: a practical guide for everyday use. *Endosc. Int. Open* 2020; **8**: 1220–4.

4 Soffer S, Klang E, Shimon O *et al*. Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. *Gastrointest. Endosc.* 2020; **92**: 831–9.

5 Topol EJ. Welcoming new guidelines for AI clinical research. *Nat. Med.* 2020; **26**: 1318–20.

6 Theodoridis S, Koutroumbas K. *Pattern Recognition*, 4th edn. Academic Press, 2009.

7 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.

8 Korman LY, Delvaux M, Gay G *et al*. Capsule endoscopy structured terminology (CEST): proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. *Endoscopy* 2005; **37**: 951–9.

9 Vasilakakis MD, Diamantis D, Spyrou E, Koulaouzidis A, Iakovidis DK. Weakly supervised multilabel classification for semantic interpretation of endoscopy video frames. *Evolv. Syst.* 2020; **11**: 409–21.

10 Diamantis DE, Iakovidis DK, Koulaouzidis A. Look-behind fully convolutional neural network for computer-aided endoscopy. *Biomed. Signal Process. Control.* 2019; **49**: 192–201.

11 Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. *Expert Rev. Gastroenterol. Hepatol.* 2017; **11**: 43–51.

12 Xing X, Jia X, Meng MQ. Bleeding detection in wireless capsule endoscopy image video using superpixel-color histogram and a subspace KNN classifier. *Annu Int Conf IEEE Eng Med Biol Soc* 2018: 1–4.

13 Aoki T, Yamada A, Kato Y *et al*. Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network. *J. Gastroenterol. Hepatol.* 2020; **35**: 1196–200.

14 Noya F, Alvarez-Gonzalez MA, Benitez R. Automated angiodysplasia detection from wireless capsule endoscopy. *Annu Int Conf IEEE Eng Med Biol Soc* 2017: 3158–61.

15 Leenhardt R, Vasseur P, Li C *et al*. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest. Endosc.* 2019; **89**: 189–94.

16 Tsuboi A, Oka S, Aoyama K *et al*. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. *Dig. Endosc.* 2020; **32**: 382–90.

17 Fan S, Xu L, Fan Y *et al*. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys. Med. Biol.* 2018; **63**: 165001.

18 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.

19 Aoki T, Yamada A, Aoyama K *et al*. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. *Dig. Endosc.* 2020; **32**: 585–91.

20 Wang S, Xing Y, Zhang L *et al*. A systematic evaluation and optimization of automatic detection of ulcers in wireless capsule endoscopy on a large dataset using deep convolutional neural networks. *Phys. Med. Biol.* 2019; **64**: 235014.

21 Klang E, Barash Y, Margalit RY *et al*. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest. Endosc.* 2020; **91**: 606–13.

22 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**: 144–51.

23 Iakovidis DK, Koulaouzidis A. Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. *Gastrointest. Endosc.* 2014; **80**: 877–83.

24 Ding Z, Shi H, Zhang H *et al*. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019; **157**: 1044–54.

25 Jang BI, Lee SH, Moon JS *et al*. Inter-observer agreement on the interpretation of capsule endoscopy findings based on capsule endoscopy structured terminology: a multicenter study by the Korean Gut Image Study Group. *Scand. J. Gastroenterol.* 2010; **45**: 370–4.

26 Leenhardt R, Koulaouzidis A, McNamara D. A guide for assessing the clinical relevance of findings in small bowel capsule endoscopy: analysis of 8064 answers of international experts to an illustrated script questionnaire. *Clin. Res. Hepatol. Gastroenterol.* (in press).

27 Otani K, Nakada A, Kurose Y et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* 2020 (Pubmed ahead of print)

28 Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nat. Rev. Gastroenterol. Hepatol.* 2015; **12**: 172–86.

29 Zhou S, Yang H, Siddique MA, Xu J, Zhou P. A novel method for automatically locating the pylorus in the wireless capsule endoscopy. *Biomed Tech (Berl)* 2017; **62**: 1–12.

30 Dimas G, Spyrou E, Iakovidis DK, Koulaouzidis A. Intelligent visual localization of wireless capsule endoscopes enhanced by color information. *Comput. Biol. Med.* 2017; **89**: 429–40.

31 Iakovidis DK, Dimas G, Karargyris A, Bianchi F, Ciuti G, Koulaouzidis A. Deep endoscopic visual measurements. *IEEE J. Biomed. Health Inform.* 2019; **23**: 2211–9.
32 Koulaouzidis A, Iakovidis DK, Yung DE *et al*. Novel experimental and software methods for image reconstruction and localization in capsule endoscopy. *Endosc Int Open.* 2018; **6**: E205–10.
33 Dimas G, Bianchi F, Iakovidis DK, Karargyris A, Ciuti G, Koulaouzidis A. Endoscopic single-image size measurements. *Meas. Sci. Technol.* 2020; **31**: 074010.
34 Noorda R, Nevárez A, Colomer A, Beltrán VP, Naranjo V. Automatic evaluation of degree of cleanliness in capsule endoscopy based on a novel CNN architecture. *Sci. Rep.* 2020; **10**: 17706.
35 Leenhardt R, Souchaud M, Houist G *et al*. A neural network-based algorithm for assessing the cleanliness of small bowel during capsule endoscopy. *Endoscopy* 2020 (in press).
36 Lai LH, Wong GLH, Chow DKL, Lau JYW, Sung JJY, Leung WK. Inter-observer variations on interpretation of capsule endoscopies. *Eur. J. Gastroenterol. Hepatol.* 2006; **18**: 283–6.
37 Rondonotti E, Soncini M, Girelli CM *et al*. Can we improve the detection rate and interobserver agreement in capsule endoscopy? *Dig. Liver Dis.* 2012; **44**: 1006–11.
38 Beg S, Card T, Sidhu R, Wronska E, Ragunath K. How many capsule endoscopy cases can be read before accuracy is affected? *Gut* 2018; **67**: A1–30.
39 Angermann Q, Bernal J, Sánchez-Montes C, et al. Towards real-time polyp detection in colonoscopy videos: adapting still frame-based methodologies for video sequences analysis. 4th International Workshop, CARE 2017, and 6th International Workshop, CLIP 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceeding 2017:29–41
40 Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015. *J Machine Learning Res* 2015; **37**: 1–8.
41 Zwinger LL, Siegmund B, Stroux A *et al*. CapsoCam SV-1 versus PillCam SB 3 in the detection of obscure gastrointestinal bleeding: results of a prospective randomized comparative multicenter study. *J. Clin. Gastroenterol.* 2019; **53**: e101–6.
42 Chetcuti Zammit S, McAlindon ME, Sidhu R. Panoramic versus axial small bowel capsule endoscopy in overt obscure gastrointestinal bleeding. *Gut* 2019; **68**: A1–26.
43 Yung D, Brzeszczynska J, Rahman I *et al*. Double-headed capsule endoscopy: real-world experience from a multicentre British study. *Gut* 2019; **68**: A1–26.
44 Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. *J. R. Soc. Med.* 2019; **112**: 22–8.
45 Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science* 2017; **356**: 183–6.