Artificial intelligence and capsule endoscopy: unravelling the future

Miguel Mascarenhas, João Afonso, Patrícia Andrade, Hélder Cardoso, Guilherme Macedo
Hospital de São João, Porto, Portugal

Abstract
The applicability of artificial intelligence (AI) in gastroenterology is a hot topic because of its disruptive nature. Capsule endoscopy plays an important role in several areas of digestive pathology, namely in the investigation of obscure hemorrhagic lesions and the management of inflammatory bowel disease. Therefore, there is growing interest in the use of AI in capsule endoscopy. Several studies have demonstrated the enormous potential of using convolutional neural networks in various areas of capsule endoscopy. The exponential development of the usefulness of AI in capsule endoscopy requires consideration of its medium- and long-term impact on clinical practice. Indeed, the advent of deep learning in the field of capsule endoscopy, with its evolutionary character, could lead to a paradigm shift in clinical activity in this setting. In this review, we aim to illustrate the state of the art of AI in the field of capsule endoscopy.

Keywords Capsule endoscopy, artificial intelligence, deep learning, machine learning, gastroenterology

Introduction
Artificial intelligence (AI) has played an increasing role in the technological development of clinical practice and biomedical academic activity [1]. The potential of AI has applications over a range of different medical specialties, while specialties with a strong imaging and diagnostic component have assumed a leading position in the implementation of this technology [2]. Indeed, there is a growing awareness and perception of the innumerable opportunities and disruptive nature of AI in clinical practice [3].

AI is defined as the use of computers and technology to simulate intelligent behavior and critical thinking comparable to that of a human being [4]. The ever growing need to provide high-quality and cost-efficient global healthcare has resulted in a corresponding expansion in the development of computer-based and robotic healthcare tools that rely on artificially intelligent technologies [5]. In 2016, healthcare was the most funded sector regarding AI research, and investment continues to pour into this sector [6]. AI, machine learning (ML), and deep learning are overlapping disciplines [7], with many current applications in the various fields of the healthcare sector. With the advent of the big data era, the accumulation of a gigantic number of digital images and medical records created an unparalleled set of resources for ML [8]. The relationship between AI, ML, and deep learning is summarized in Fig. 1. ML is based on the recognition of patterns that can be applied to medical images [9], laboratory medicine [10], drug discovery [11], and even clinical practice [12]. ML is based on the introduction of algorithms that ingest input data, apply computer analysis to predict output values within an acceptable range of accuracy, identify patterns and trends within the data, and finally learn from previous experience [13]. ML can be either supervised or unsupervised.

A supervised ML algorithm uses the available training data (images from capsule endoscopy for example) to learn a function by mapping certain input variables/features from the training data onto a qualitative or quantitative output/target (e.g., identifying protuberant lesions in the small bowel) [14]. A frequently used example is training a model to differentiate between apples, oranges and lemons. The "label" of each type of fruit is supplied to the algorithm, along with features such as color, size, weight and shape, and by referring to a set of learning data the algorithm determines the combinations of features that differentiate the fruits [15]. In medical applications, once a model has been developed and perfected, it is tested on novel patients whose data were not included in the training set, to determine its external validity and subsequent applicability to other patients [13].

On the other hand, unsupervised ML methods rely on the arbitrary aggregation of unlabeled data sets to yield groups...
or clusters of entities with shared similarities that may be unknown to the user prior to the analysis [14]. Unsupervised ML algorithms are data-driven techniques that automatically learn from the relationships between elementary bits of information associated with each variable of a dataset [16]. The combination of and potential synergy between supervised and unsupervised methods of ML holds great promise in the field of gastroenterological endoscopy.

Deep learning is a subset of ML. The structure of neural networks, organized in multiple layers, allows them to address complex tasks [17]. Deep neural networks use the compositional hierarchy of signals, in which higher-level features are obtained by combining lower-level ones [18]. A convolutional neural network (CNN, or ConvNet) is a class of deep neural networks tailored to visual imagery analysis. CNNs resemble neurobiological processes, emulating the connectivity pattern between neurons [19]. In Fig. 2 we can see the similarities between a human neural network and a deep learning algorithm. CNNs are a type of feed-forward artificial neural network inspired by the organization of the animal visual cortex, whose individual neurons are arranged in such a way that they respond to overlapping regions tiling the visual field [20]. Therefore, CNNs require less preprocessing and are also less dependent on prior knowledge and human effort. CNNs exhibit superior performance when compared to other deep learning architectures, namely in terms of object detection and recognition [21]. The fields of application of CNNs vary from abnormality detection and disease classification to computer-aided diagnosis [22]. Deep learning and CNNs are disruptive and have excelled in the detection of a range of diseases in capsule endoscopy [23].

**Application in capsule endoscopy**

Capsule endoscopy is one of the branches of gastroenterology that can benefit the most from the application of this type of technology. Indeed, the use of AI in this field shows great promise and capsule endoscopy can serve as a stepping stone for the broader application of AI in endoscopy and gastroenterology. Below, we summarize the state of the art regarding the use of AI in capsule endoscopy.

**AI and bleeding lesions**

One of the fields in which the automation of videocapsule diagnostics has undergone enormous advances is in the detection of gastrointestinal (GI) hemorrhage, namely from ulcers and vascular lesions. In 2007, Lau et al developed a model capable of detecting the presence of hemorrhage with a sensitivity of 88.3%, using simple color coding. However, this model was limited by the very low quality of the analyzed video images [24]. In the following year, Giritharan et al analyzed 400 frames of GI hemorrhage using a support-vector
machine (SVM) model and obtained results similar to Lau, with a sensitivity greater than 80% in the detection of positive bleeding cases [25].

In 2009, Li et al took a database of 200 hemorrhage images from 10 patients and, using a multilayer perceptron (MLP) model, developed an ML algorithm capable of detecting areas of bleeding with a sensitivity, specificity and accuracy greater than 90%. This study was particularly important because it was able to surpass the detection rate of the state-of-the-art methods at that time [26]. In the same year, Pan et al developed a CNN by analyzing the color and texture of the images. The algorithm was tested using 150 full videos of wireless capsule endoscopy (WCE), consisting of 3172 hemorrhage images and 11,458 of normal mucosa. This model achieved a sensitivity of 93% with a specificity of 96% for the detection of cases. The large number of images analyzed contributed to the robustness of this experiment [27].

In 2010, Charisis et al developed an SVM using a dataset of 40 images of normal mucosa and 40 images of ulcers. This model was able to detect positive cases with a sensitivity and specificity greater than 95%. However, it was only able to detect cases of medium or higher severity, which reduces its applicability in real clinical practice [28].

In 2014, Fu et al developed a computer-aided design (CAD) method based on SVM, capable of detecting hemorrhage with a sensitivity, specificity and accuracy of 99%, 94% and 95%, respectively. This method was particularly interesting because it introduced a new form of image analysis. The developed model analyzed super pixels—grouped sets of pixels of similar characteristics in each frame—which made it possible to reduce the computation costs compared to the analysis of each isolated pixel, while improving the detection capacity compared to the overall analysis of a frame [29]. In the same year, Gosh et al used 30 videos of WCE and, using 50 images of hemorrhage and 200 of normal mucosa for training the model, developed an SVM classifier applied to 2000 test images, achieving a sensitivity of 93% and specificity of 95% [30].

In December 2015, Hassan et al used 1200 training frames and 1720 testing frames to develop a new local texture descriptor that was capable of obtaining sensitivities and specificities above 98.9%, significantly higher than what had been done to date. In addition, this method had a low computational cost, making it suitable for real-time implementation [31].

In 2018, Fan et al developed a method for simultaneous detection of ulcers and mucosal erosions, with a high accuracy of 95.2% and 95.3%, sensitivity of 96.8% and 93.7%, and specificity of 94.8% and 96.0% in detecting ulcers and erosions, respectively. This study was relevant since it did not evaluate an isolated lesion, but instead a set of pathological entities [32].

In January 2019, Leenhardt et al developed a CNN method capable of detecting small-bowel angiectasias, using 6360 still frames from 4166 different videocapsule videos. This study, given the large number of patients covered, proved to be extremely robust and presented excellent results, with a sensitivity of 100% and specificity of 96%, an excellent starting point for future automated diagnostic software [33]. In fact, angiectasias are the most common lesions diagnosed in patients with medium GI bleeding undergoing video capsule endoscopy.

In August of 2019, Pokorelov et al developed a combined color and texture algorithm with excellent computational cost and efficiency. Using 300 bleeding frames and 200 nonbleeding or normal frames for the training dataset (500 frames) and 500 bleeding and 200 nonbleeding frames (700 frames) for the testing dataset, they were able to obtain a sensitivity, specificity and accuracy of 97.6%, 95.9% and 97.6%, respectively [34]. Also, in August of the same year, Aoki et al developed a CNN method that compared the time and effectiveness of videocapsule reading by 2 processes: (A) endoscopist-alone readings; and (B) endoscopist readings after a first screening by the proposed CNN. Mean reading time of small-bowel sections by endoscopists was significantly shorter during process B (expert, 3.1 min; trainee, 5.2 min) compared to process A (expert, 12.2 min; trainee, 20.7 min) (P<0.001). For 37 mucosal breaks, the detection rate by endoscopists was not significantly lower in process B (expert, 87%; trainee, 55%) compared to process A (expert, 84%; trainee, 47%). This study was extremely important because it demonstrates the applicability of these auxiliary diagnostic methods in daily clinical practice, enabling a significant reduction in the video capsule reading time [35].

More recently, in March 2020, Tsuboi et al used 2237 images of WCE and created a CNN system capable of detecting small-bowel angiectasias with a sensitivity of 98.8% and specificity of 98.4% [36]. In July 2020, Aoki et al developed a CNN using images from 41 patients, with a total of 27,847 images, capable of detecting blood in the intestinal lumen with a sensitivity of 96.6%, specificity of 99.9% and accuracy of 99.9%. The performance of the network was compared with a conventional tool (suspected blood indicator) and proved able to outperform this tool [37].

**AI and protuberant lesions**

One of the most profitable areas of investigation in this context is the detection and classification of protruding structures of the small intestine mucosa, since its analysis by other methods is extremely difficult. However, using videocapsule images it is also possible to detect abnormal structures present elsewhere in the GI tract.

In 2008, Barbosa et al, based on 100 images of normal mucosa and 92 images of tumor lesions and using an MLP method, developed an algorithm capable of being applied to real data, with a sensitivity of 98.7% and specificity of 96.6% in the detection of tumors of the small intestine [38]. The following year, using the same AI method, Li et al analyzed 300 video images from 2 WCE exams and developed a model with an accuracy of 86.1% (sensitivity and specificity of 89.8%). The fact that they only used data from 2 patients limits the applicability of this model in other settings, such as real-life medical practice [39]. The same author, in April 2011, applying a model based on the color difference between tumor lesions and normal mucosa, used a dataset of 1200 images from 10 different patients to develop a CAD system that demonstrated a sensitivity of 82.3%
and specificity of 84.7% in the detection of GI tumors through the analysis of WCE exams [40].

Barbosa et al also carried out a further study in 2011, with a more comprehensive dataset (700 tumor frames and 2300 normal frames). Through the analysis of mucosal textural information, they developed a method with sensitivity and specificity greater than 93% for the detection of tumors of the small intestine [41].

Zhao et al, in the same year, created a dataset of 1120 images (560 of polyps and 560 of normal mucosa), with the particularity of including a group of consecutive frames of injury and normal mucosa, to verify whether the simultaneous analysis of 5 frames of the same lesion is superior to the analysis of 1 isolated frame. This study was an important innovation, since until then most of the methods developed were based on the analysis of only one image of each lesion. Zhao demonstrated that a polyp sequence can have apparently normal frames and that a normal mucosa sequence can have apparently abnormal frames. By analyzing several consecutive frames, the number of false negatives and false positives in the model can be reduced. In this case, with the analysis of consecutive images, they managed to improve the specificity and sensitivity of single frame evaluation, from 91% and 83% to 95% and 92%, respectively [42].

In August 2015, Vieira et al compared a method of SVM and one of MLP in the automatic detection of small intestine tumors, through the analysis of 700 abnormal frames from 14 patients and 2500 normal frames from 19 individuals, concluding that the MLP method is superior to the older AI method in sensitivity, specificity and accuracy [43]. In 2017, Yuan et al developed a CAD method capable of identifying polyps and also distinguishing other structures, such as bubbles and the presence of cloudy luminal material, with an accuracy greater than 95%. This method is particularly important, since it allows the removal of luminal content that makes it difficult to evaluate images [44].

In March 2019, Blanes-Vidal et al managed to establish a correlation in 225 patients between the detection of colorectal polyps in colonoscopy and those detected in WCE in the same patients, with a sensitivity of 97.1% and specificity of 93.3%. This study represents an important advance in the applicability of this technique as a possible method of screening for colorectal cancer in the future [45].

In February 2020, Saito et al, through the analysis of a robust database of 30,584 images of protruding small intestine lesions, developed a CNN method capable of not only identifying lesions but also classifying them as polyps, nodules, epithelial tumors, submucosal tumors and venous structures, with sensitivities of 86.5%, 92.0%, 95.8%, 77.0%, and 94.4%, respectively. This study was a pioneer in the use of several types of lesions in a single model, and allowed these methods to approach more closely to real clinical practice, where several pathological changes can occur simultaneously and require proper distinction [46].

Al and inflammatory bowel disease

Another medical field where the videocapsule has a well-established role is in the evaluation of patients with inflammatory bowel disease, particularly those with Crohn's disease (CD), since it allows an assessment of all the small intestine mucosa. In addition to being able to assist in the confirmation of the CD diagnosis, it also allows the extent of disease activity and response to therapy to be assessed, through the application of scores such as the Lewis score [47].

In 2010, Seshamani et al, using an SVM-based similarity learning method, used videocapsule images of 47 exams of CD patients, to manually extract 724 images of injury areas. In this way, they developed a model capable of detecting suggestive areas of injury with an accuracy of 88%, which allowed to drastically reduce the training time of the model, without compromising its effectiveness [48].

In March 2020, Klang et al developed a deep-learning algorithm, using the analysis of 17,640 endoscopic capsule images from 49 patients with CD and healthy individuals, that achieved an accuracy greater than 95%, revealing the potential of this technology in the prediction of small-bowel findings based on videocapsule endoscopy in CD patients [49]. Also, in March 2020, Freitas et al assessed the correlation between classic videocapsule reading and the use of a new software tool of the RAPID Reader®, TOP100, in the application of the Lewis score in CD patients. They examined 115 patients and showed a strong agreement (89.6% of the cases) between the 2 methods of capsule reading. This study is particularly important because it demonstrates the clinical applicability of this type of diagnostic aid [50].

More recently, in June 2020, Y. Barash, in collaboration with the aforementioned E. Klang, developed a neural network capable of classifying the severity of ulcers in patients with CD. To achieve that, they classified 2598 images containing ulcers on a numerical scale of 1-3. They divided the experiment into 2 parts. In the first part, they evaluated the interobserver agreement between 2 different evaluators, and in the second they used a CNN to automatically classify the ulcers. They obtained a global human interobserver agreement of 31% (76% for grade I-III ulcers) vs. a global neural network agreement of 67% (91% for grade I-III) [51].

Al and celiac disease

Celiac disease affects around 1% of the world population, with an increasing prevalence in recent years. This chronic autoimmune disorder, characterized by an immune attack on the small intestine mucosa, is triggered by the ingestion of gluten in genetically susceptible individuals [52]. The gold standard for diagnosis is the presence of duodenal villous atrophy in endoscopic biopsies. However, this is an invasive and expensive procedure. Therefore, capsule endoscopy appears as a more practical approach in some settings and an alternative with fewer associated risks [52]. With the increasing use of this diagnostic method, computer models have been developed to assist doctors in diagnosing this disorder using a videocapsule enteroscopy video.

In 2010 Ciacco et al developed a threshold classifier to classify images of patients with celiac disease. Using images
from 21 exams (11 of patients with celiac disease and 10 controls) and through the analysis of 9 different characteristics of each frame, they developed a model capable of predicting the occurrence of the disease with a sensitivity of 80% and a specificity of 96%. Later, in 2014, the same investigation team developed a new model capable of predicting the occurrence of the disease with a sensitivity of 84.6% and specificity of 92.3%, using base images from patients and controls [54].

In 2017, Zou et al., using data from 6 patients with celiac disease and 5 controls, developed a CNN to quantitatively measure the presence and degree of intestinal mucosa damage. Its model, using the latest technology in the field of AI, obtained a sensitivity and specificity of 100% in the small group tested. In addition, they were also able to classify the degree of mucosal injury, opening doors for the future analysis of a correlation between the videocapsule images and the histological evaluation [55].

More recently, Koh et al developed a computer-aided detection system, decomposing the video images of 13 control tests and 13 patients. This system, with an accuracy of 86.5% and a sensitivity and specificity of 88.4% and 84.6%, respectively, demonstrates the potential to effectively identify patients with celiac disease [56].

In April 2020, Wang et al., using a deep learning method, developed a CNN system, based on data from 52 patients and 55 healthy controls, that demonstrated a remarkable accuracy (accuracy, sensitivity and specificity 95.9%, 97.2% and 95.6%, respectively). This study was particularly robust given the large number of images collected, as well as the type of analysis used [57].

### AI and luminal content

AI may also play a key role in locating the capsule in the GI tract, as well as in the detection and elimination of artifacts that may compromise the mucosal evaluation, thus reducing the required examination reading time and also reducing bias and interpretation errors. In 2012, Seguí et al developed a model capable of detecting, isolating and classifying luminal content, to remove it from image view. For this, he resorted to images of clean mucosa and images of luminal content, which they divided into turbid liquid and bubbles. The proposed system was then evaluated using a large dataset. The statistical analysis of the performance showed an accuracy above 90%, far superior to that of previously existing models. In addition, this was the first work to distinguish between the different artifacts detected throughout the video capsule examination [58].

In 2013, Ionescu et al analyzed more than 10,000 frames from 10 different patients to detect images with artifacts and thus reduce the number of images that would have to be analyzed by the clinician, thus making the reading process faster and more effective. Through a CNN method, they managed to develop an algorithm with an accuracy of 88.2% in the detection of bubbles and food debris [59].

In 2018, Wang et al proposed to develop a model capable of automatically detecting the location of the boundaries between the stomach and the duodenum–pylorus. For this, they analyzed 42,000 images and randomly selected 3801 images from the pyloric region, 1822 pre-pyloric and 1979 post-pyloric. Using an SVM method, the investigators were able to detect the location of the pylorus in 30 real WCE videos, with an accuracy of 97.1% and a specificity of 95.4% [60]. All these types of analysis can contribute greatly to the optimization of the evaluation of videocapsule images, to make the reading process less time consuming and considerably more effective.

### AI and hookworm

Parasitic infections represent another type of pathological entity that can be detected by this diagnostic method. Of all the parasites that reach the GI tract, hookworm infection is one of the most common and serious, affecting about 600 million people worldwide. The hookworm is a helminth that presents itself as a tubular structure, with grayish, white or pinkish semi-transparent body [61].

In March 2016, Xiao et al used 440,000 images from 11 patients to develop a mechanism capable of automatically detecting these helminths in videocapsule images. This was one of the first studies to address this topic. Its model showed a sensitivity and specificity close to 78%. The low effectiveness of this model is mainly due to the difficulty in correctly distinguishing the parasite's structure from some bubbles and intestinal folds. As a way of correcting this low performance, they raise the possibility of considering the temporal and spatial relationship between consecutive images in future works [62]. In May 2018, He et al used 1500 images to create a CNN model capable of detecting the presence of hookworms with a sensitivity of 84.6% and specificity of 88.6%; these results were superior to those previously obtained in this area [63].

The automatic detection of this type of parasite remains a very challenging task, since the wide variety of aspects that they can present is a huge obstacle to the development of effective methods for their detection. Thus, it will be necessary to develop more research in order to improve the accuracy of these methods in the detection of intestinal helminths. Although there are alternative tests available, such as the parasitological examination of the stools, this task remains an important proof of concept for AI in video capsule endoscopy. A summary of all the studies discussed can be found in Table 1

### AI: promises and pitfalls

In several studies AI was able to compensate for the limited experience of novice endoscopists and some errors by even the most experienced endoscopists. Human nature makes their performance variable, and diagnostic performance may certainly be impaired by a decrease of awareness and attention, or forgetfulness due to fatigue, anxiety, or any other physical or emotional stress [64]. The scarcity of human resources and the increasing workload can be alleviated with the implementation of AI systems. AI may also have a particularly important role in the emergency department, where less time
### Table 1 Summary of studies using AI methods to aid videocapsule video analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Field of application</th>
<th>Year of publication</th>
<th>Proposed goals</th>
<th>Number of subjects</th>
<th>AI type</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ciaccio et al [54]</td>
<td>Celiac disease</td>
<td>2010</td>
<td>Evaluate if quantitative markers could assist in the screening for celiac disease</td>
<td>11 patients and 10 controls</td>
<td>Threshold classifier (quantitative analysis of 9 different frame characteristics)</td>
<td>Sensitivity of 80% and a specificity of 96% in predicting the occurrence of the disease</td>
</tr>
<tr>
<td>Ciaccio et al [54]</td>
<td>Celiac disease</td>
<td>2014</td>
<td>Improve the image-based detection of villous atrophy and other abnormality in videocapsule endoscopy, using the grayscale brightness of each frame</td>
<td>13 patients and 13 controls</td>
<td>Threshold classifier (quantitative analysis of grayscale brightness)</td>
<td>Sensitivity of 84.6% and specificity of 92.3%</td>
</tr>
<tr>
<td>Zou et al [55]</td>
<td>Celiac disease</td>
<td>2017</td>
<td>Measure the presence and degree of intestinal mucosa damage</td>
<td>6 patients and 6 controls</td>
<td>CNN</td>
<td>Sensitivity and specificity of 100% in the cases tested. They were also capable to classify the degree of mucosal damage</td>
</tr>
<tr>
<td>Koh et al [56]</td>
<td>Celiac disease</td>
<td>2019</td>
<td>Identify patients with celiac disease</td>
<td>13 patients and 13 controls</td>
<td>SVM</td>
<td>Accuracy of 86.5% and a sensitivity and specificity of 88.4% and 84.6% respectively</td>
</tr>
<tr>
<td>Wang et al [57]</td>
<td>Celiac disease</td>
<td>2020</td>
<td>Identify patients with celiac disease</td>
<td>52 patients and 55 controls</td>
<td>CNN</td>
<td>Accuracy of 95.9%, sensitivity of 97.2% and specificity of 95.6%</td>
</tr>
<tr>
<td>Seshamani et al [48]</td>
<td>Inflammatory bowel disease</td>
<td>2010</td>
<td>Detecting suggestive areas of Crohn's disease injury</td>
<td>47 exams</td>
<td>SVM</td>
<td>Accuracy of 88%</td>
</tr>
<tr>
<td>Klang et al [49]</td>
<td>Inflammatory bowel disease</td>
<td>2020</td>
<td>Detecting suggestive areas of Crohn's disease injury</td>
<td>49 exams</td>
<td>Deep learning algorithm</td>
<td>Accuracy greater than 95%</td>
</tr>
<tr>
<td>Freitas et al [50]</td>
<td>Inflammatory bowel disease</td>
<td>2020</td>
<td>Assess the correlation between classic videocapsule reading and the use of a new software tool in patients with CD</td>
<td>115 patients</td>
<td>NA</td>
<td>Agreement on 89.6% of the cases between the 2 methods</td>
</tr>
<tr>
<td>Barash et al [51]</td>
<td>Inflammatory bowel disease</td>
<td>2020</td>
<td>Access the interobserver agreement between 2 human observers and a neural network</td>
<td>49 exams</td>
<td>Ordinary neural network</td>
<td>Global human interobserver agreement of 31% (76% between grade I-III ulcers) vs. a global neural network agreement of 67% (91% between grade I-III)</td>
</tr>
<tr>
<td>Xiao et al [62]</td>
<td>Hookworm infection</td>
<td>2016</td>
<td>Automatically detect hookworms in WCE videos</td>
<td>11 patients</td>
<td>Rusboost method</td>
<td>Sensitivity and specificity close to 78%</td>
</tr>
<tr>
<td>He et al [63]</td>
<td>Hookworm infection</td>
<td>2018</td>
<td>Detect hookworm presence</td>
<td>1500 images</td>
<td>CNN</td>
<td>Sensitivity of 84.6% and specificity of 88.6%</td>
</tr>
<tr>
<td>Segui et al [58]</td>
<td>Luminal content</td>
<td>2012</td>
<td>Detect video artefacts</td>
<td>Large dataset</td>
<td>NA</td>
<td>Accuracy above 90% - first work to distinguish between the different artifacts</td>
</tr>
<tr>
<td>Wang et al [59]</td>
<td>Capsule location</td>
<td>2018</td>
<td>Detect pylorus location</td>
<td>3801 images from the pyloric region</td>
<td>SVM</td>
<td>Accuracy of 97.1% and specificity of 95.4%</td>
</tr>
</tbody>
</table>

(Contd...)
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Field of application</th>
<th>Year of publication</th>
<th>Proposed goals</th>
<th>Number of subjects</th>
<th>AI type</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbosa et al</td>
<td>Protruding structures</td>
<td>2008</td>
<td>Small intestine tumor detection</td>
<td>100 images of normal mucosa and 92 images of tumor lesions</td>
<td>MLP</td>
<td>Sensitivity of 98.7% and specificity of 96.6%</td>
</tr>
<tr>
<td>Li et al</td>
<td>Protruding structures</td>
<td>2009</td>
<td>Small intestine tumor detection</td>
<td>WCE videos from 2 patients</td>
<td>MLP</td>
<td>Accuracy of 86.1%, sensitivity of 89.8% and specificity of 89.8%</td>
</tr>
<tr>
<td>Li et al</td>
<td>Protruding structures</td>
<td>2011</td>
<td>GI tumor detection</td>
<td>10 patients</td>
<td>SVM</td>
<td>Sensitivity of 82.3% and specificity of 84.7%</td>
</tr>
<tr>
<td>Barbosa et al</td>
<td>Protruding structures</td>
<td>2012</td>
<td>Small intestine tumor detection</td>
<td>700 tumor frames and 2300 normal frames</td>
<td>MLP</td>
<td>Sensitivity and specificity greater than 93%</td>
</tr>
<tr>
<td>Zhao et al</td>
<td>Protruding structures</td>
<td>2012</td>
<td>Verify if using consecutive frames from the same lesion was more effective than single frame analysis</td>
<td>560 polyp frames and 560 normal mucosa frames</td>
<td>NA</td>
<td>Managed to increase the specificity and sensitivity of single frame evaluation from 91% and 83% to 95% and 92% with the analysis of consecutive images</td>
</tr>
<tr>
<td>Vieira et al</td>
<td>Protruding structures</td>
<td>2015</td>
<td>Compare a method of SVM and one of MLP in the automatic detection of small intestine tumors</td>
<td>700 abnormal and 2500 normal frames from 19 patients</td>
<td>MLP vs. SVM</td>
<td>The MLP method was superior to the older AI method in sensitivity, specificity and accuracy</td>
</tr>
<tr>
<td>Yuan et al</td>
<td>Protruding structures</td>
<td>2017</td>
<td>Detect polyps and distinguish it from other structures</td>
<td>3000 normal mucosa frames and 1000 polyps</td>
<td>Stacked Sparse Autoencoder with Image Manifold Constraint (SSAEIM)</td>
<td>Accuracy greater than 95%</td>
</tr>
<tr>
<td>Blaines-Vidal et al</td>
<td>Protruding structures</td>
<td>2019</td>
<td>Establish a match between the colorectal polyps detected in colonoscopy and those detected in WCE</td>
<td>255 patients</td>
<td>CNN</td>
<td>Accuracy of 96.4%, sensitivity of 97.1%) and specificity of 93.3%</td>
</tr>
<tr>
<td>Saito et al</td>
<td>Protruding structures</td>
<td>2020</td>
<td>Identify lesions but also classify them as polyps, nodules, Epithelial tumors, submucosal tumors and venous structures</td>
<td>30,584 images of protruding small intestine lesions</td>
<td>CNN</td>
<td>Sensitivities of 86.5%, 92.0%, 95.8%, 77.0%, and 94.4% for the different types of structures</td>
</tr>
<tr>
<td>Lau et al</td>
<td>GI hemorrhage</td>
<td>2007</td>
<td>Detect the presence of GI hemorrhage using a simple color coding</td>
<td>577 abnormal images</td>
<td>NA</td>
<td>Sensitivity of 88.3%</td>
</tr>
<tr>
<td>Girtharan et al</td>
<td>GI hemorrhage</td>
<td>2008</td>
<td>Detect the presence of GI hemorrhage</td>
<td>400 GI bleeding frames</td>
<td>SVM</td>
<td>Sensitivity greater than 80%</td>
</tr>
<tr>
<td>Li et al</td>
<td>GI hemorrhage</td>
<td>2009</td>
<td>Detect the presence of GI hemorrhage</td>
<td>10 patients (200 bleeding frames)</td>
<td>MLP</td>
<td>Sensitivity, specificity and accuracy greater than 90%</td>
</tr>
<tr>
<td>Pan et al</td>
<td>GI hemorrhage</td>
<td>2009</td>
<td>Detect the presence of GI hemorrhage analyzing the color and texture</td>
<td>150 full WCE videos</td>
<td>CNN</td>
<td>Sensitivity of 93% with a specificity of 96% for the detection of cases</td>
</tr>
<tr>
<td>Charisis et al</td>
<td>GI hemorrhage</td>
<td>2010</td>
<td>Detect GI ulcers</td>
<td>40 normal mucosa and 40 ulcer images</td>
<td>SVM</td>
<td>Sensitivity and specificity greater than 95% (only cases of medium or higher severity)</td>
</tr>
</tbody>
</table>

(Contd...)
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Field of application</th>
<th>Year of publication</th>
<th>Proposed goals</th>
<th>Number of subjects</th>
<th>AI type</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fu et al [29]</td>
<td>GI hemorrhage</td>
<td>2014</td>
<td>Detect GI bleeding using super pixel analysis</td>
<td>20 different WCE videos</td>
<td>SVM</td>
<td>Sensitivity, specificity and accuracy of 99%, 94% and 95% respectively</td>
</tr>
<tr>
<td>Gosh et al [30]</td>
<td>GI hemorrhage</td>
<td>2014</td>
<td>Detect GI bleeding</td>
<td>30 WCE videos</td>
<td>SVM</td>
<td>Sensitivity of 93%, specificity 94.9%</td>
</tr>
<tr>
<td>Hassan et al [31]</td>
<td>GI hemorrhage</td>
<td>2015</td>
<td>Detect GI bleeding</td>
<td>1200 training frames and 1720 testing frames</td>
<td>SVM</td>
<td>Sensitivity and specificity above 98.9%</td>
</tr>
<tr>
<td>Fan et al [32]</td>
<td>GI hemorrhage</td>
<td>2018</td>
<td>Simultaneous detection of ulcers and mucosal erosions</td>
<td>144 full WCE videos</td>
<td>CAD method based on deep learning framework</td>
<td>Accuracy of 95.2% and 95.3%, sensitivity of 96.8% and 93.7%, and specificity of 94.8% and 95.9%, in detecting ulcers and erosions respectively</td>
</tr>
<tr>
<td>Leenhardt et al [33]</td>
<td>GI hemorrhage</td>
<td>2019</td>
<td>Detecting small bowel angiectasias</td>
<td>6360 still frames from 4166 different videocapsule videos</td>
<td>CNN</td>
<td>Sensitivity of 100% and specificity of 96%</td>
</tr>
<tr>
<td>Pokorelov et al [34]</td>
<td>GI hemorrhage</td>
<td>2019</td>
<td>Develop a color and texture algorithm with excellent computational costs</td>
<td>500 frames for the training dataset and 700 frames for the testing dataset</td>
<td>SVM</td>
<td>Sensitivity, specificity and accuracy of 97.6%, 95.9% and 97.6%</td>
</tr>
<tr>
<td>Aoki et al [35]</td>
<td>GI hemorrhage</td>
<td>2019</td>
<td>Compare the time and effectiveness of videocapsule reading by endoscopist-alone and endoscopist readings after the first screening by a proposed CNN</td>
<td>NA</td>
<td>CNN</td>
<td>Mean reading time was significantly shorter during the second process without significantly decreasing in the detection rate</td>
</tr>
<tr>
<td>Tsuboi et al [36]</td>
<td>GI hemorrhage</td>
<td>2020</td>
<td>Detect small bowel angiectasis</td>
<td>2237 WCE images</td>
<td>CNN</td>
<td>Sensitivity of 98.8% and specificity of 98.4%</td>
</tr>
<tr>
<td>Aoki et al [37]</td>
<td>GI hemorrhage</td>
<td>2020</td>
<td>Detect GI bleeding</td>
<td>41 patients, with a total of 27447 images</td>
<td>CNN</td>
<td>Sensitivity of 96.6%, specificity of 99.9% and accuracy of 99.9%</td>
</tr>
</tbody>
</table>

AI, artificial intelligence; CAD, computer aided design; CNN, convolutional neural network; GI, gastrointestinal; ML, machine learning; MLP, multilayer perceptron; NA, not applicable; SVM, support-vector machine; WCE, wireless capsule endoscopy.

is available for full capsule visualization and a faster view time is often necessary.

Despite convincing results and growing evidence of the central role of AI in technological evolution in the area of digestive endoscopy, the overwhelming majority of studies were designed in a retrospective manner. Furthermore, inherent bias, such as selection bias, cannot be excluded in this situation and real-life clinical application should be carefully tested and taken into consideration before validation of the AI solution.

Spectrum bias is another pitfall of the current AI application to capsule endoscopy. Spectrum bias occurs when a diagnostic test is studied in a range of individuals who are different from the intended population for the test. AI systems are tailor-made, designed to fit the training dataset, and the risk of overfitting should not be ignored. As a matter of fact, the efficiency and validity of an AI learning model may not be completely applicable to a new dataset, and AI learning models are still vulnerable to overfitting issues despite recent mitigation efforts.

On the other hand, the efficiency and accuracy of ML increases as the amount of data increases. Capsule endoscopy produces a considerable quantity of data to feed the growth of the ML systems. Additionally, the advent of the big data era will inexorably propel the exponential development of AI in capsule endoscopy. Despite the many challenges, the fast development of AI will ensure a relevant role for AI in capsule endoscopy in clinical practice.
Concluding remarks

The exponential development of the computational capacity of new computers, coupled with a greater understanding and accessibility of deep learning technologies, has made it possible to develop algorithms that are increasingly effective and applicable in the most diverse areas. Health, and particularly gastroenterology, are no exception. Undoubtedly, the future of the analysis of capsule endoscopy videos involves the use of auxiliary computerized methods that will not only facilitate the analysis of these images, but also improve the accuracy of diagnosis.

However, there is a pressing need for more research studies proving the usefulness of this technology in a clinical context, taking into account the computational costs, efficiency and accuracy of the technology. Indeed, there is still a long way to go before AI takes its place as an integral part of the daily clinical practice of the gastroenterologist.

References


