Recent developments in small bowel endoscopy: the “black box” is now open!

Luigina Vanessa Alemanni1,2,*, Stefano Fabbri1,2,*, Emanuele Rondonotti3, Alessandro Mussetto1

1Gastroenterology Unit, Santa Maria delle Croci Hospital, Ravenna; 2Department of Medical and Surgical Sciences, S. Orsola-Malpighi Hospital, Bologna; 3Gastroenterology Unit, Valduce Hospital, Como, Italy

Over the last few years, capsule endoscopy has been established as a fundamental device in the practicing gastroenterologist's toolbox. Its utilization in diagnostic algorithms for suspected small bowel bleeding, Crohn’s disease, and small bowel tumors has been approved by several guidelines. The advent of double-balloon enteroscopy has significantly increased the therapeutic possibilities and release of multiple devices (single-balloon enteroscopy and spiral enteroscopy) aimed at improving the performance of small bowel enteroscopy. Recently, some important innovations have appeared in the small bowel endoscopy scene, providing further improvement to its evolution. Artificial intelligence in capsule endoscopy should increase diagnostic accuracy and reading efficiency, and the introduction of motorized spiral enteroscopy into clinical practice could also improve the therapeutic yield. This review focuses on the most recent studies on artificial-intelligence-assisted capsule endoscopy and motorized spiral enteroscopy.

Keywords: Artificial intelligence; Capsule endoscopy; Enteroscopy; Motorized spiral enteroscopy; Small bowel endoscopy

INTRODUCTION

The 21st century has marked the onset of a revolution in the treatment of small bowel diseases. In fact, evaluation of the small bowel is traditionally challenging for gastroenterologists because of its length and tortuous anatomy in the abdominal cavity. In particular, the introduction of capsule endoscopy (CE) and double-balloon enteroscopy has enabled evaluation of the entire gastrointestinal (GI) tract.1,2 CE has revolutionized small bowel imaging by providing a non-invasive method of examination of the mucosal surface. Given the increased detection of small bowel diseases by CE, the release of multiple device-assisted enteroscopes in the market has been crucial for histopathological sampling and endoscopic therapy in selected cases, thus avoiding the need for angiography and surgery.

Further technological advancements have occurred in recent years concerning small bowel endoscopy. In particular, some CE software now employs artificial intelligence (AI), and device-assisted enteroscopy (DAE) has revved up with power spiral motorization. This narrative review highlights the most recent studies of these two new techniques.

ARTIFICIAL INTELLIGENCE, CONVOLUTIONAL NEURAL NETWORK-MODELS, AND CAPSULE ENDOSCOPY

Since its introduction in 2000, the clinical use of CE has increased, and the indications for small bowel CE (SBCE) include suspected small bowel bleeding, assessment of Crohn’s disease,
screening for polyposis syndromes, celiac disease, and small bowel tumor investigation. However, because it captures and transmits approximately 60,000 images/person through the entire GI tract, SBCE results are time-consuming for physicians, requiring intense focus for an average time estimated between 30 and 120 minutes. It has been described that after just one capsule study examination, the accuracy of SBCE readers declines. To overcome such limitations, AI has recently been proposed for SBCE readers’ assistance, mainly as a first reader, with the aim of reducing workload while improving accuracy.

AI allows computerized approaches to the resolution of cognitive problems. Owing to recent improvements in AI, particularly the implementation of machine learning (ML) technologies, in the last few decades, research on models for computer-aided diagnosis (CAD) and computer-assisted image analysis have become major areas of interest in medical imaging. ML is a subdomain of AI that provides systems that can independently learn and infer decisions. In supervised ML, learning is achieved through a training data-set used as reference according to human experts indications, for example a large database of normal and abnormal GI images, after which the model extracts image features and translates them to mathematical values, which are statistically elaborated with predictive models. In medical imaging, ML systems are able to recognize disease patterns and estimate the probability that one specific image belongs to different classes (e.g., normal vs. abnormal, malignant vs. benign). The performance of these methods can be affected by the reliability of experts’ first inputs, effectiveness of mathematical formulation translation, and difficulty of translating a large complex of normal and abnormal variations in medical imaging. Most recent technologies have led to a further subdomain of ML, namely deep learning (DL). DL includes ML models based on particular artificial neural network systems with complex multi-layer architecture. The most important type of DL technology is based on convolutional neural network (CNN) systems (Fig. 1), which are extremely useful in medical image analysis because they are structured to resemble the organization of the animal visual cortex. When properly trained with a large dataset of images, CNN systems are able to independently discover image features for further classification without requiring manually designed inputs; in fact, CNN systems are expected to be superior to hand-engineered features input.

Recently, an increasing number of CNN-based CAD models have been proposed and validated for medical imaging. Many studies, both bio-computational and clinical, have shown encouraging results upon their application. CNN-based CAD systems were first employed for single abnormality detection, but new implementations are being conducted for SBCE multiclass lesion detection, localization, characterization, and even some attempts regarding entirely self-reporting capsules. Even if they are still in the experimental phase and have not yet been introduced in clinical practice, CNN-based CAD models will probably revolutionize SBCE readings.

The initial performance was promising, as many CNN-based systems have been trained for blood and angioectasia detection, always with excellent accuracy results, with some algorithms achieving a sensitivity of 98.8% to 100% and a specificity of 96% to 98.4%. Recently, other CNN-based systems have been developed for the detection of ulcers and erosions in both inflammatory bowel disease (IBD) and non-IBD patients. A recent meta-analysis by Soffer et al. evaluated CNN-based system accuracy, pooling together five studies on bleeding source detection and five studies on ulcer detection, providing
a bleeding content/source pooled sensitivity of 98% (95% confidence interval [CI], 96%–99%) and pooled specificity of 99% (95% CI, 97%–99%). For ulcer detection, the pooled sensitivity was 95% (95% CI, 89%–98%), while the pooled specificity was 94% (95% CI, 90%–96%). Similar performance was described in another meta-analysis by Mohan et al., including nine studies and aimed to assess CNN-based system performance in hemorrhage and/or ulceration detection, establishing an overall pooled accuracy of 95.4% (sensitivity 95.5%, specificity 95.8%, positive predictive value 95.8%, and negative predictive value 96.8%).

Thus, CNN-based CAD could be useful in IBD diagnosis and assessment, especially in Crohn’s disease. Some algorithms have already shown good results in this context, proving their value in grading the severity of ulcerations and in differentiating strictures from normal mucosa or other ulcerative lesions. The first results have also been described for celiac small bowel changes and protruding lesions, such as tumors and polyps, suggesting a potentially increasing role for SBCE in celiac disease diagnosis and follow-up, and in familial polyposis screening.

Today, the main challenge for AI-assisted CE is the development of CNN-based systems capable of detecting, classifying, and characterizing multiple types of lesions. Accuracy is proven to be higher when systems are trained independently for different types of abnormalities; thus, very large datasets with high-quality annotations are required for these systems.

In 2019, Ding et al. conducted a large study, collecting data from 6,970 patients who underwent SBCE in 77 medical centers for two years; a CNN-based model was trained with a 158,235 image dataset from 1,970 patients with the aim of classifying them into normal, inflammation, polyps, bleeding, and other pathological findings. The CNN model was further validated in 5,000 patients and showed excellent performance compared with conventional gastroenterologist reading, with high sensitivity (99.8%–99.9%) and an error rate of only 3% in distinguishing abnormal from normal images.

Similarly, the RetinaNet AI system trained by Otani et al. provided high area under the receiver operative curve (AUC) values for detecting ulcers/erosions (0.966), vascular lesions (0.95), and tumors (0.95), with slightly inferior AUC values after external validation (0.928, 0.884, and 0.9, respectively). Some studies have also reported algorithms able to recognize small bowel segments and/or localize lesions. The model trained in a study by Dimas et al. showed an average localization error of 2.70±1.62 cm. In addition, the first attempts at bowel cleanliness assessment and for entirely self-reporting SBCE have also been described. In particular, regarding small bowel cleanliness, Nam et al. developed a DL-based automation software for calculating its score; a scoring system based on mucosal visibility was trained for DL in the training set, and the performance of the trained software was subsequently evaluated in the validation set.

Despite these promising results, many relevant concerns about medico-legal issues are emerging: in clinical practice, a fully developed and potentially self-reporting CNN-based system entrusts a machine with the entire diagnostic procedure. Even if just used as first reader assistance, this would result in a huge number of images never being reviewed by any human expert. Such concerns can only be overcome with adequate evidence on the usefulness of AI in SBCE in terms of cost-effectiveness, time saved, and improved accuracy in lesion detection.

A multicenter retrospective study conducted by Aoki et al. compared the CE reading time of physicians, alone or after CNN-aided first screening, in 20 videos of the entire small bowel, each of which included 0 to 5 lesions of small bowel mucosal breaks. The mean reading time was significantly shorter when the capsule study was performed after CNN assessment, both in the expert readers group (3.1 vs. 12.2 min, p<0.001) and in the trainee readers group (5.2 vs. 20.7 min, p<0.001), without affecting the accuracy in abnormalities detection (i.e., erosions or ulcerations). In the previously mentioned meta-analysis, Mohan et al. established in nine studies an average SBCE reading time of 4.5 minutes using CNN-based systems.

A large study published in 2019 by Ding et al. compared the results of conventional analysis and multiclass-trained CNN-assisted analysis performed by 20 gastroenterologists. The model outperformed conventional analysis, providing not only a significantly shorter mean reading time (5.9±2.2 min vs. 96.6±22.5 min, p<0.001), but also higher sensitivity (99.9% vs. 74.6%) and negative predictive value (99.8% vs. 67.4%), where the total detection rate increased by 16.3%.

Therefore, CNN-based CAD is promising for being potentially more valuable than conventional physician SBCE readings. However, before being fully introduced in clinical practice, there are still other challenges to solve: most studies are conducted under experimental conditions and have a retrospective single-center study design; in addition, models are mainly trained on a large number of selected still images (with risks of overfitting) and are also usually validated on image datasets,
rather than using full-length videos. External validation of all CNN models is limited by the variability of devices, networks, and images, but also by the possibility of discordance in lesion characterization or in medical terminology between expert groups.9

**MOTORIZED SPIRAL ENTEROSCOPY**

The motorized spiral enteroscope (PowerSpiral; Olympus Medical, Tokyo, Japan) is a novel advancement in the field of enteroscopy. This enteroscope operates on the same principle as the manual spiral enteroscope, with a spiral overtube mounted on the insertion tube portion, converting the rotational energy into a linear force to pull the bowel onto the scope. This innovation is powered by the integrated electric motor located in the endoscope handle, which allows rotation of the spiral overtube. Owing to the automated rotation, the procedure may be performed faster by a single operator.

PowerSpiral enteroscope (PSE) is composed of three elements (Fig. 2). (1) A reusable endoscope, similar to a standard pediatric colonoscope, with a working length of 168 cm, an outer diameter of 12.8 mm, and a 3.2 mm caliber accessory channel with a separate water irrigating channel. Therefore, routine colonoscopic accessories can be used for therapeutic procedures. In the latest endoscopes being released, PSE is equipped with image enhancement technology, such as high-definition imaging and narrow-band imaging. (2) A 24 cm disposable and atraumatic spiral overtube, made of soft spiral fins with an outer diameter of 31 mm, was attached to the rotation coupler on the tip of the endoscope. (3) A control unit with a foot pedal and visual force gauge. The operator controlled the integrated electric motor for rotating the spiral overtube through a foot pedal. During the procedure, a visual force gauge was used to monitor the direction of overtube rotation and to control the amount of torque applied to the small bowel. When excessive resistance is detected, spiral rotation is stopped automatically to avoid perforation, as reported in 0.27% of cases with conventional spiral enteroscopy.26

The first in-human procedure using PSE was performed in November 2015 by Neuhaus et al.27 (Dusseldorf, Germany) on a 48-year-old patient with iron deficiency anemia. PSE was inserted approximately 250 cm distal to the ligament of Treitz within 20 minutes, and jejunal angioectasia was detected and treated with argon plasma coagulation. Iatrogenic mucosal trauma or delayed adverse events (AEs) were registered.

The first prospective study of PSE was published by Beyna et al. in 2021.28 A total of 132 patients with suspected or confirmed small bowel diseases underwent antegrade enteroscopy with PSE at two European tertiary referral endoscopic centers (Dusseldorf, Brussels). The technical success rate of PSE, de-

![Fig. 2. Motorized spiral enteroscopy system.](image-url)
defined as reaching at least the ligament of Treitz, was 97%. The overall diagnostic yield of PSE, which was the primary endpoint of this study, was 74.2%. Therapeutic procedures were performed in 68.2% of cases. The median depth of maximum insertion (DMI) was 450 cm beyond the ligament of Treitz, with a median insertion time of 25 minutes. The median duration of PSE enteroscopy was 54 minutes. Total antegrade enteroscopy of the cecum was performed in 14 patients (10.6%). The overall AE rate was 14.4%, with two patients experiencing major AEs (1.5%): a perforation in the terminal ileum treated with laparoscopic suturing and upper GI bleeding from a Mallory-Weiss lesion of the gastric cardia that required endoscopic treatment and hemotransfusions.

A case series of 14 novel motorized spiral enteroscopies was recently published by Prasad et al. in India. The indications for the procedure were suspected small bowel thickening or stricture in 10 patients and obscure GI bleeding in the four remaining patients. The diagnostic yield obtained was 92.8% (13/14), which was defined as reaching the target lesion or performing a total enteroscopy. The average duration was 61.1 minutes for anterograde enteroscopy and 90 minutes for retrograde enteroscopy. Panenteroscopy, confirmed on entering the cecum from the antegrade approach, was achieved in two patients, with procedure times of 90 and 105 minutes. The therapeutic procedures performed were argon plasma coagulation (two patients) and balloon dilation of the stricture (one patient). Three patients experienced mild postprocedural odynophagia due to cricopharynx abrasions, and one patient developed pancreatitis and was treated conservatively.

Beyna et al. performed a second prospective trial of 30 patients at two tertiary referral centers, in which attention was focused on the ability of motorized spiral enteroscopy to perform panenteroscopy using the antegrade approach along with a retrograde approach, if needed. Indeed, the primary outcome of this study was the total endoscopic small bowel visualization rate, achieved in 72.4% of the patients, of which only five (16.6%) procedures were completed using an antegrade-alone approach. The median procedure times required for the antegrade and retrograde approaches were 51 and 40 minutes, respectively. The overall diagnostic yield of PSE in this study was 80%.

A further retrospective study by Ramchandani et al., including 61 patients from an Indian population, showed similar results, wherein the overall total enteroscopy rate was 60.6%, of which 31.1% was achieved using an antegrade-alone approach. In both case series, no major AEs were reported.

However, all of the aforementioned studies included patients without a history of abdominal surgery. A very recent paper from the Netherlands showed good results in a population of 170 patients, 34 of whom (20%) had a history of surgically altered anatomy (mostly partial small bowel resection/strictureplasty or bariatric Roux-en-Y gastric bypass). The overall diagnostic yield was 64.1%, the panenteroscopy rate (when intended) was 70.3%, and most importantly, only minor AEs were registered in 15.8% of procedures, without differences between patients with or without altered anatomy. Moreover, the results of a large prospective observational multicenter trial in a real-life setting were recently published, which included 54 out of 251 patients (enrolled for the “core” phase of the study) with previous abdominal surgery (21.5%), resulting in surgically altered GI anatomy in 25 (10%). The overall serious AE rate (core and “training” phase patients) was 2.3% (7/298), which did not increase in the subgroup of post-surgical patients (1.9%). Diagnostic and therapeutic yields were 83% and 60.2%, respectively. Panenteroscopy was performed in 51% of the patients who were initially planned for total enteroscopy. All the published studies mentioned above are summarized in Table 1.

In addition, motorized spiral enteroscopy showed promising outcomes in the first clinical case of enteroscopy-assisted endoscopic retrograde cholangiopancreatography (ERCP) in a patient with altered anatomy. DAE-ERCP with balloon dilation of bilioenteric anastomotic strictures was successfully performed using a novel motorized spiral enteroscope in a 78-year-old man who underwent duodenum-preserving pancreatic head resection and Roux-en-Y reconstruction. No AEs were observed during or after the procedure.

The results of these previous studies suggest that the motorized spiral enteroscope is a promising and reliable technique, in line with other DAE techniques in terms of diagnostic yield, DMI, and AE rate. Motorized spiral enteroscopy potentially provides many advantages in performing deep enteroscopy, including decreased procedure time, increased total enteroscopy rates, and the possibility of performing a single-operator procedure. However, at present, only indirect comparisons with other DAE are possible because of the lack of randomized controlled trials.

**CONCLUSIONS**

The introduction of AI in SBCE software and the release in the
market of power spiral enteroscopy have revolutionized the approach to small bowel pathologies. The small bowel, once considered the black box for gastroenterologists, can now be studied and approached endoscopically more easily and with more possibilities. A definitive introduction in clinical practice of CNN-based CE and PSE will require further evidence prospectively developed and validated in real-life scenarios.

Conflicts of Interest
The authors have no potential conflicts of interest.

Funding
None.

Author Contributions
Conceptualization: AM; Data curation: LVA, SF, AM; Formal analysis: AM; Supervision: ER; Validation: AM, ER; Writing-original draft: LVA, SF, AM; Writing-review & editing: LVA, SF, ER, AM.

ORCID
Luigina Vanessa Alemanni https://orcid.org/0000-0003-3013-7772
Stefano Fabbri https://orcid.org/0000-0003-2125-7588
Emanuele Rondonotti https://orcid.org/0000-0001-5013-0514
Alessandro Mussetto https://orcid.org/0000-0001-8140-4223

REFERENCES

### Table 1. Summary of studies evaluating PowerSpiral enteroscopy

<table>
<thead>
<tr>
<th>Study</th>
<th>Study design</th>
<th>No. of patients</th>
<th>Diagnostic yield (%)</th>
<th>Therapeutic yield (%)</th>
<th>Panenteroscopy rate (%)</th>
<th>Major adverse events (%)</th>
<th>Minor adverse events (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beyna et al., 2020</td>
<td>Prospective</td>
<td>132</td>
<td>74.2</td>
<td>68.2</td>
<td>10.6</td>
<td>1.5</td>
<td>14.4</td>
</tr>
<tr>
<td>Prasad et al., 2020</td>
<td>Case series</td>
<td>14</td>
<td>92.8</td>
<td>35.7</td>
<td>35.7</td>
<td>28.5</td>
<td>42.8</td>
</tr>
<tr>
<td>Ramchandani et al., 2021</td>
<td>Retrospective</td>
<td>61</td>
<td>65.5</td>
<td>23.0</td>
<td>60.6</td>
<td>0</td>
<td>24.5</td>
</tr>
<tr>
<td>Beyna et al., 2021</td>
<td>Prospective</td>
<td>30</td>
<td>80.0</td>
<td>86.7</td>
<td>70.0</td>
<td>0</td>
<td>16.7</td>
</tr>
<tr>
<td>Al-Toma et al., 2022</td>
<td>Prospective</td>
<td>170</td>
<td>64.1</td>
<td>53.5</td>
<td>70.3</td>
<td>0</td>
<td>15.8</td>
</tr>
<tr>
<td>Beyna et al., 2022</td>
<td>Prospective</td>
<td>298</td>
<td>83.0</td>
<td>60.2</td>
<td>51.9</td>
<td>2.3</td>
<td>8.7</td>
</tr>
</tbody>
</table>


