REVISTA ESPAÑOLA DE
ENFERMEDADES DIGESTIVAS
The Spanish Journal of Gastroenterology

Title:
Artificial intelligence in gastrointestinal endoscopy: evolution to a new era

Authors:
Oswaldo Ortiz Zúñiga, María Glòria Fernández Esparrach, María Daca, María Pellisé

DOI: 10.17235/reed.2022.8961/2022
Link: PubMed (Epub ahead of print)

Please cite this article as:
Ortiz Zúñiga Oswaldo, Fernández Esparrach María Glòria, Daca María, Pellisé María. Artificial intelligence in gastrointestinal endoscopy: evolution to a new era. Rev Esp Enferm Dig 2022. doi: 10.17235/reed.2022.8961/2022.

REVISTA ESPAÑOLA DE
ENFERMEDADES DIGESTIVAS
The Spanish Journal of Gastroenterology

Artificial intelligence in gastrointestinal endoscopy: evolution to a new era

Oswaldo Ortiz[1], Glòria Fernández-Esparrach[1], Maria Daca[1], Maria Pellisé[1]

Department of Gastroenterology, Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain.

Correspondence:

Dr. Maria Pellisé, MD, PhD; Department of Gastroenterology, Hospital Clínic of Barcelona; Villarroel 170; 08036 Barcelona, Spain. Phone number: +34 93 227 5418; Fax number: +34 93 227 55-89: email: mpellise@clinic.cat

ABSTRACT

Artificial intelligence (AI) systems based on machine learning have evolved in the last few years with an increasing applicability in gastrointestinal endoscopy. Thanks to AI, an image (input) can be transformed into a clinical decision (output). Although AI systems have been studied mainly to improve detection (CADe) and characterization of colorectal polyps (CADx), other indications are being currently investigated as detection of blind spots, scope guidance, or delineation/measurement of lesions.

The objective of these review is to summarize the current evidence on applicability of AI systems in gastrointestinal endoscopy, highlight strengths and limitations of the technology and review regulatory and ethical aspects for its general implementation in gastrointestinal endoscopy.

Key words: Artificial intelligence. Gastrointestinal endoscopy.

REVISTA ESPAÑOLA DE
ENFERMEDADES DIGESTIVAS
The Spanish Journal of Gastroenterology

GENERAL OVERVIEW

Over the last years, AI technologies have appeared to revolutionize gastrointestinal endoscopy. The term **"artificial intelligence" (AI)** refers to the ability of computer algorithms to mimic human cognitive decision-making process. AI algorithms are designed to compensate human limitations and weakness during the procedures such as fatigability, stress, or lack of experience and to help clinicians to make decisions. Some of the potential advantages of AI in gastrointestinal endoscopy, but not the only ones, are to short the endoscopists' learning curve or increase their inter-observer agreement.

AI systems designed for gastrointestinal endoscopy are based on **machine learning (ML),** an automatic learning algorithm that transforms inputs (a large amount of information from real word) into outputs (decision making process) without being previously programmed for this function[1]. **Deep learning (DL),** a subset of ML, use convolutional neural networks (**CNNs**) that are composed of several hierarchic layers to recognize a combination of visual characteristics that tends to appear together to differentiate one image from another. The input is fed with raw data and pass through different interconnected layers (simulating neuronal connections) with an increasingly level of complexity until it results as an output[1]. **Figure 1.** With DL, the program adjusts itself and has a continuous improving as more data are included until it gets a high level of precision**.**

To develop a DL model, three types of data sets are usually used: Training data introduced for algorithm initial development, validation data that is used to make that algorithm reach an acceptable error margin and test data set for evaluation of the algorithm in real practice.

In gastrointestinal endoscopy, AI has a wide spectrum of possible indications but the first systems developed are focused on computer-aided detection (CADe) and characterization/diagnosis (CADx) of lesions (mainly colorectal polyps). There are some other systems developed and trained to assess other aspects of the endoscopic procedure such as: detection of blind spots, delineation and measurement of lesions,

scope guidance.

The aim of this article is to give a general and updated overview of the applications of AI systems in gastrointestinal endoscopy based on current evidence.

APPLICATIONS IN GASTROINTESTINAL ENDOSCOPY

Esophagus

Esophageal adenocarcinoma (EAC) incidence is increasing in the last years. Since the tumor stage at diagnosis have an important impact in the prognosis in EAC the importance of surveillance in patients with premalignant conditions such as Barrett´s esophagus (BE) is essential to detect early cancers and/or associated focal dysplasia which can be treated by minimally invasive procedures.

Patients with known BE have regular endoscopic surveillance according to guidelines[2]. However, there is a high level of complexity for optical diagnosis of dysplasia in BE due to focal and subtle morphological changes requiring high-definition scopes and expert endoscopists[3]. Moreover, guidelines recommend taking biopsies according to Seattle protocol [2] resulting in most cases in an inefficient, time consuming and low diagnostic rate procedure. For these reasons many technologies have been developed to overcome this limitation, but until now none of them have significantly increased the diagnostic yield of dysplasia in Barret esophagus which it's still a problem in a non-expert context[4]. To overcome these limitations studies have been designed to assess these topics and are summarized in **Table 1.**

De Groof et al[5] developed a DL based CAD system using 5 data sets. In stage 1 (pre-training) they used 494364 images of all intestinal segments, training data (set 2 and 3) with 1704 esophageal images of early neoplastic BE and non-dysplastic Barret esophagus from 669 patients. System performance was assessed using data sets 4 and 5. The CAD system classified between neoplastic and non-neoplastic BE with an 89% accuracy, 90% sensitivity and 88% specificity in dataset 4. In dataset 5 (80 patients and images) the CAD system performance was compared to 53 general endoscopists with 88% vs 73% accuracy, 93% vs 72% sensitivity, and 83% vs 74% specificity respectively. **Swager et al** [6] developed a ML algorithm for detection of dysplasia in BE based on volumetric laser endomicroscopy features (higher VLE surface signal and lack of

layering). The algorithm was developed using 60 VLE images (30 non-dyplastic and 30 HGD/early carcinoma) from ex vivo VLE histologic correlations. The feature "layering and signal decay statistics" showed the optimal performance, with an area under the receiver operating characteristic curve (AUC) of .95. Corresponding sensitivity and specificity were 90% and 93%, respectively.

Van Der Sommen et al [7]in a study performed in a tertiary center used 100 white-light endoscopic images from 44 patients with Barret´s esophagus of which 21 showed histologically proven early neoplasia (60 images). They evaluate the performance of a computer algorithm that employed texture, color filters and ML for detection of early neoplastic lesions in Barret´s esophagus. The system identified on per-patient analysis early neoplastic lesions with a sensitivity and specificity of 0.86 and 0.87 respectively[7].

Gastric cancer and pre-neoplastic lesions

Gastric cancer is the third cause of cancer related death worldwide[8]. Early detection and treatment are the best measure to improve survival in these patients. Gastroscopy is useful for detecting gastric cancer, however most of the cancers are diagnosed at advance stage, due to flat-depressed subtle morphology of early gastric cancer (EGC).

Several techniques and tools such as dye based and virtual chromoendoscopy with magnifying endoscopy have been demonstrated to improve detection of these lesions. However, despite of these advantages the diagnostic accuracy of EGC in non-expert setting is still low[9].

AI application in gastric cancer has been evaluated in different scenarios as to identify blind spots during the procedure and improving endoscopy quality[10], to improve detection of EGC[11–13], to delineate lesions and predict deep submucosal invasion[14,15], to recognize pre-neoplastic lesions such as atrophy or intestinal metaplasia and to identify H. pylori infection based on some endoscopic changes such as mucosal redness, swelling and nodular change[16–18].

Regarding endoscopy quality **Wu et al**[10] developed ENDOANGEL, a CNN and deep reinforcement algorithm, and compared with endoscopists in a prospective real

time in vivo RCT with 1050 patients. ENDOANGEL group had fewer blind spots (mean 5.38 [SD] 4.32 vs. 9.82 [SD 4.98]; p<0.001) and longer inspection time (5.40 [SD 3.82] vs. 4.38 [SD 3.91] minutes; P<0.001) compared to white-light endoscopy with an accuracy of nearly 85% and sensitivity of 100% for detecting gastric cancers[10]. **Hu et al**[13] tested an algorithm developed for detection of EGC with NBI magnifying endoscopy. They compared the performance of AI system versus endoscopists evaluating magnifying endoscopy images in a training and validation cohort. There was similar performance between senior endoscopists and the AI (accuracy: .770 vs .755, P Z .355; sensitivity: .792 vs .767, P Z .183; specificity: .745 vs .742, P Z .931) but better than the junior endoscopists (accuracy: .770 vs .728, P < .05). **Ikenoyama et al**[11] developed a CNN system using 13,584 endoscopic images from 2639 patients and subsequently they compare and independent data set of 2940 images from 140 cases for diagnosis of EGC with 67 endoscopists. The developed CNN system detected more early gastric cancer in a shorter time than endoscopists. Sensitivity was significantly higher for the CNN (by 26.5%, 95% CI 14.9–32.5%), and the specificity and PPV were significantly higher for the endoscopists (specificity: by 9.9%, 95% CI 8.7–11.1%; PPV: by 20.2%, 95% CI 16.6–23.8%). Their NPV were comparable (96.5% vs. 94.9%, by 1.6%, 95% CI 1.0–2.1%).

For diagnosis of H. pylori infection **Nakashima et al**[16] developed a DL based CAD system based on Linked Color Imaging and white light endoscopy images (6639 WLE and 6248 LCI images from 395 subjects) to classify H. pylori status into three categories: uninfected, currently infected, and post-eradication. The diagnostic accuracy was higher for LCI-CAD system (84.2% for uninfected, 82.5% for currently infected and 79% for post-eradication) than with WLE-CAD system (75% for uninfected, 77.5% for currently infected and 74% for post-eradication). **Zheng et al**[17] retrospectively evaluated a CCN system based on endoscopic white light images of 1,959 patients, 1,507 (847 with H. pylori infection) were assigned to the derivation cohort, and 452 (including 310 (69%) with H. pylori infection) were assigned to the validation cohort. The area under the curve for a single gastric image was 0.93 (95% [CI] 0.92-0.94) with sensitivity, specificity, and accuracy of 81.4% (95% CI 79.8%-82.9%), 90.1% (95% CI 88.4%-91.7%), and 84.5% (95% CI 83.3%-85.7%),

respectively. Area under the curve for multiple gastric images (8.3 ± 3.3) per patient was 0.97 (95% CI 0.96-0.99) with sensitivity, specificity, and accuracy of 91.6% (95% CI 88.0%-94.4%), 98.6% (95% CI 95.0%-99.8%), and 93.8% (95% CI 91.2%-95.8%), respectively, using an optimal cutoff value of 0.4.

Other possible indications of AI in upper gastrointestinal endoscopy currently investigated are differentiation between the two most common subepithelial lesions: GIST and leiomyomas during endoscopic ultrasonography. The AI systems was trained using 10439 EUS images from 752 patients with GIST or leiomyomas. AI was then evaluated through retrospective and prospective tests. In the prospective evaluation 132 patients (36 GISTs, 44 leiomyomas and 52 other types of sub-epithelial lesions) were histologically diagnosed among 508 subjects. The total accuracy of endoscopists increased from 69.7% (95% confidence interval [CI] 61.4%–76.9%) to 78.8% (95%CI 71.0%–84.9%; P=0.01).[19]

The main studies that analyze the impact of AI in these settings are summarized in **table 2**.

Small bowel pathology

Video capsule endoscopy is a widely endoscopic procedure for diagnosis of small bowel pathology. Thousands of images are obtained in a typical small bowel study, leading to a no negligible percentage of miss rate of lesions. The main causes of this miss rate are related to human attention and the fact that some lesions appeared only in a few frames makes a real possibility of overlooking lesions during the reading. To overcome these limitations AI systems have been studied in the following scenarios: detection GI ulcers and bleeding lesions which represents one of the main indications of video capsule, identification of protruding lesions/tumors that are more difficult to detect during CE because they usually are isolated lesions and with a color similar to mucosa in comparison with vascular lesions and/or ulcers and optimal localization and size of capsule lesion(s)[20] to determine the best therapeutic approach and evaluating endoscopic features of celiac disease[21]

Mascarenhas M. et al also develop a AI based on CCN algorithm with VCE 18625 images (2830 showing protruding lesions and 15795 normal mucosa) to detect

protruding lesions in small bowel during VCE. Training and validation datasets were built in a 80%/20% distribution. The accuracy of the system was 92.5% with a sensitivity of 93.8 and specificity of 96.5%. The system analyzed the images with a rate of 70 frames per second.

The main studies assessing these scenarios are summarized in **Table 3**

Inflammatory bowel disease

There are many endoscopic indexes to assess and grade the severity of disease activity in IBD. However, these scores may have high intra and inter-observer heterogeneity1919. AI systems could improve these imprecisions leading to increase in accuracy for assessing the endoscopic severity. There are some studies evaluating the performance of AI in these scenarios showing promising results, nevertheless, they are retrospective. This reflects the difficulty for assessing good high-quality and representative images with adequate correlation with histology in this pathology[23–25].

Another potential scenario in IBD patients that AI system must be evaluated is on the detection of Colitis associated neoplasia due to the difficulties in detection and demarcation of these lesions.

Colorectal lesions

Colorectal cancer (CRC) is the third most common cancer worldwide and the second leading cause of cancer-related death66. Colonoscopy has already demonstrated to reduce colon cancer incidence and mortality. However, colonoscopy it is not a perfect procedure and post-colonoscopy CRC is still a major issue[26].To overcome these limitations several strategies have been developed for improving quality of bowel preparation and improving adenoma detection rate. Some of these strategies include dye based or virtual chromoendoscopy, distal attachments, although previous studies have failed in demonstrate a substantial benefit in detection of lesions in average-risk populations[27].

Polyp detection (CADe) is the most studied field of applications of AI systems in gastrointestinal endoscopy with several high-quality studies assessing the usefulness of AI systems already published. In a recent meta-analysis, **Hassan et al.** including 4350

patients from 5 randomized controlled trials demonstrate a significant increase in adenoma detection rate in the CADe group relative risk 1.44 95% [CI], 1.27-1.62 with moderate heterogeneity[28]. Similar results were found in another meta-analysis made by **Barua and colleagues**[29] with an increase in ADR in CADe group RR 1.48. However, there was no significant differences in detection of advance adenomas, probably due to lack of statistical power to demonstrate these differences. Thus, studies addressing this subtype of lesions are needed. In view of the increasingly evidence, the European Society of Gastrointestinal Endoscopy in its advance imaging for detection and differentiation of colorectal neoplasia have dedicated a recommendation regarding the role of AI system in detection of characterization of lesions suggesting the possible incorporation of CAD systems in these setting if acceptable evidence is demonstrated in in vivo studies[30].

Other possible methods to improve polyp detection besides deep learning methods are also being evaluated. **Fernandez-Esparrach et al**[31] developed a model that defines with more accuracy polyp boundaries and integrated and highlight the lesions into energy maps that represents the likelihood of the presence of a polyp. The system (WM-DOVA). In 24 videos containing polyps the mean of maximum values on the energy maps were higher in frames where polyps were presents than without ($p<0.001$). The performance in high-quality frames were high (AUC=0.79 [95%CI 0.70-0.87] VS. 0.75 [95%CI 0.66-0.83]. **Sánchez-Montes** [32]et al developed a CAD system for in-vivo characterization of colorectal polyps using WLE images**.** The system development was inspired in pit patterns of Kudo classification and recognizes surface textural patterns (textons) which can help to detect dysplasia. Images of 225 polyps (142 dysplastic and 83 non-dysplastic). The CAD system correctly classified 205 polyps 91.1% 92.3% dysplastic and 89.2% non-dysplastic. There were no significant differences between CAD and endoscopists performance.

Accurate in vivo **polyp characterization** (CADx) is crucial to differentiate between neoplastic or non-neoplastic lesions enabling the implementation of “resect and discard” and “diagnose and leave in situ” strategies proposed by The American Society for Gastrointestinal Endoscopy PIVI [3] (Preservation and Incorporation of

Valuable Endoscopic Innovations) in order to avoid unnecessary polypectomies and histopathology analysis with a subsequent improve in cost-effectiveness of the procedure. However, these criteria could be difficult to achieve in clinical practice. To simplify these inconvenient and made a more realistic optical diagnosis competence standards SODA criteria were recent developed through an expert Delphi consensus for implementation of these strategies in diminutive colorectal polyps[33]. In order to implement the "leave in situ" strategy is acceptable if the endoscopist achieve at least 90% of sensitivity and 80% of specificity and for "resect and discard" strategy at least 80% sensitivity and specificity in high confidence characterization. Based on this, AI systems developed and trained for optical diagnosis of colonic lesions should provide similar values in order to achieve sufficient agreement in assignment of post-polypectomy surveillance intervals when compared to decisions based on pathology according to guidelines.

For characterization of polyps there is a retrospective validation study made by Zachariah and colleagues evaluating a CCN system trained with over 6000 colorectal polyps images that demonstrated to achieve PIVI *thresholds*[34]. Clinical in-vivo studies are expected to confirm the efficacy of CADx systems in optical diagnosis of colorectal lesions. Studies regarding detection and characterization of colorectal lesions are summarized in **Table 4.**

Other applications of AI in colonoscopy could be the evaluation of colorectal lesions include Blind spots and guidance to scope insertion, polyp size measurement, optical diagnosis for prediction of invasion and for making prediction models of lymphatic invasion in patients with T1 colorectal cancers[35].

LIMITATIONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES AND IMPLEMENTATION ISSUES

The applicability and external validation of the algorithms depends on the quality of the training datasets, that is an aspect that clinicians must be careful when evaluating AI technologies.

Training data set could lead to selection bias and overfitting. Selection bias occurs when the data set used for training the model is not large enough to represent the real clinical scenario. “Overfitting” happens when too many details are introduced in the deep learning algorithm making too much heterogeneity and imprecision in categorizing the data. Selection bias and overfitting errors could arise mistakes in clinical practice.

Also, despite the potential high accuracy that AI systems could reach, deep learning models could be inscrutable for the clinicians turn them into black boxes and making impossible to infer/discuss causal relationships in decision making in clinical practice[36].

There is also concern about the effects of general implementation of artificial intelligence systems on learning curve of the endoscopist with the possibility of deskilling and on the influence of the device on endoscopist attention during the procedure and possible increasing of procedural time.

Another important aspect that needs to be addressed is potential “overdiagnosis” of lesions in context of screening programs that could led to overtreatment of lesions with negligible potential to develop cancer. This could have implications (patient anxiety, prone to complications related to the procedure) in patients and health-care system (higher burden and health-care system costs).

ETHICAL ASPECTS, REGULATIONS AND IMPLEMENTATION

AI technologies are regulated in U.S. Food and Drug Administration consider AI tools for clinical support as medical devices. Such regulatory enable the FDA and manufacturers to evaluate and monitor a software product from its premarket development to post-commercialization performance.

To regulate the report of clinical trials evaluating any AI component the CONSORT-AI (Consolidated Standards of Reporting Trials-Artificial Intelligence) and SPIRIT AI (Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence) report guidelines were designed in parallel through an international consensus and literature review of expert through a Delphi survey. CONSORT-AI its adapted from CONSORT 2010 guidelines adding specific issues regarding implementation and evaluation of AI interventions[37,38].

Ethical aspects regarding informed consent and reliable of the diagnosis and miss or incorrect diagnosis and responsibilities in decisions making process depends on the endoscopists.

As AI technologies are validated and introduced in clinical practice, cost-effective studies to assess the real impact in costs of AI in gastrointestinal endoscopy daily practice are going to be needed.

CONCLUSION

Artificial intelligence systems are being evaluated in a wide spectrum of indications in gastrointestinal endoscopy. Different algorithms, training methods, aided technologies and indications for diagnosis, prognosis and management are addressed with promising outcomes that could led to a substantial change of gastrointestinal endoscopy daily practice.

Although there are many aspects that need to be addressed in order to standardize the use of these technologies and more prospective, randomized in-vivo studies are needed to confirm the results in some indications, we expect than in the next few years AI will be a reality as a valuable tool for making decisions and will work side to side with physicians rather than replace them in gastrointestinal endoscopy practice.

REFERENCES

1. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nature Medicine*. 2019;25(1). doi:10.1038/s41591-018-0316-z
2. Weusten BLAM, Bisschops R, Coron E, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) position statement. *Endoscopy*. 2017;49(2):191-198. doi:10.1055/s-0042-122140
3. Sharma P, Savides TJ, Canto MI, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on imaging in Barrett's Esophagus. *Gastrointestinal Endoscopy*. 2012;76(2):252-254. doi:10.1016/j.gie.2012.05.007
4. Harrison M, Allen JE, Gorrepati VS, López-Jamar JME, Sharma P. Management of Barrett's esophagus with low-grade dysplasia. *Diseases of the Esophagus*. 2018;31(4). doi:10.1093/dote/doy004
5. de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology*. 2020;158(4):915-929.e4. doi:10.1053/j.gastro.2019.11.030
6. Swager AF, van der Sommen F, Klomp SR, et al. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointestinal Endoscopy*. 2017;86(5):839-846. doi:10.1016/j.gie.2017.03.011
7. Van Der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy*. 2016;48(7):617-624. doi:10.1055/s-0042-105284
8. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6). doi:10.3322/caac.21492
9. East JE, Vleugels JL, Roelandt P, et al. Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review. *Endoscopy*. 2016;48(11):1029-1045. doi:10.1055/s-0042-118087
10. Wu L, He X, Liu M, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric

cancer: A randomized controlled trial. *Endoscopy*. Published online 2021. doi:10.1055/a-1350-5583

11. Ikenoyama Y, Hirasawa T, Ishioka M, et al. Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists. *Digestive Endoscopy*. 2021;33(1):141-150. doi:10.1111/den.13688
12. Li L, Chen Y, Shen Z, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer*. 2020;23(1):126-132. doi:10.1007/s10120-019-00992-2
13. Hu H, Gong L, Dong D, et al. Identifying early gastric cancer under magnifying narrow-band images with deep learning: a multicenter study. *Gastrointestinal Endoscopy*. 2021;93(6):1333-1341.e3. doi:10.1016/j.gie.2020.11.014
14. Yoon HJ, Kim S, Kim JH, et al. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. *Journal of Clinical Medicine*. 2019;8(9):1310. doi:10.3390/jcm8091310
15. Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointestinal Endoscopy*. 2019;89(4):806-815.e1. doi:10.1016/j.gie.2018.11.011
16. Nakashima H, Kawahira H, Kawachi H, Sakaki N. Endoscopic three-categorical diagnosis of Helicobacter pylori infection using linked color imaging and deep learning: a single-center prospective study (with video). *Gastric Cancer*. 2020;23(6). doi:10.1007/s10120-020-01077-1
17. Zheng W, Zhang X, Kim JJ, et al. High Accuracy of Convolutional Neural Network for Evaluation of Helicobacter pylori Infection Based on Endoscopic Images: Preliminary Experience. *Clin Transl Gastroenterol*. 2019;10(12):e00109. doi:10.14309/ctg.0000000000000109
18. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. *Endoscopy International Open*. 2018;06(02):E139-E144. doi:10.1055/s-0043-120830
19. Yang X, Wang H, Dong Q, et al. An Artificial Intelligence System for Distinguishing Between Gastrointestinal Stromal Tumors and Leiomyomas Using Endoscopic Ultrasonography (with video). *Endoscopy*. 2021;(16). doi:10.1055/a-1476-8931

20. Dimas G, Spyrou E, Iakovidis DK, Koulaouzidis A. Intelligent visual localization of wireless capsule endoscopes enhanced by color information. *Computers in Biology and Medicine*. 2017;89(August):429-440. doi:10.1016/j.compbiomed.2017.08.029
21. Sana MK, Hussain ZM, Maqsood MH, Shah PA. Artificial intelligence in celiac disease. *Computers in Biology and Medicine*. 2020;125(August):103996. doi:10.1016/j.compbiomed.2020.103996
22. Mosli MH, Feagan BG, Sandborn WJ, et al. Histologic evaluation of ulcerative colitis: A systematic review of disease activity indices. *Inflammatory Bowel Diseases*. 2014;20(3):564-575. doi:10.1097/01.MIB.0000437986.00190.71
23. Gubatan J, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions. *World Journal of Gastroenterology*. 2021;27(17):1920-1935. doi:10.3748/wjg.v27.i17.1920
24. Stidham RW, Liu W, Bishu S, et al. Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Network Open*. 2019;2(5):1-10. doi:10.1001/jamanetworkopen.2019.3963
25. Bhambhvani HP, Zamora A. Deep learning enabled classification of Mayo endoscopic subscore in patients with ulcerative colitis. *European Journal of Gastroenterology and Hepatology*. Published online 2021:645-649. doi:10.1097/MEG.0000000000001952
26. Rutter MD, Beintaris I, Valori R, et al. World Endoscopy Organization Consensus Statements on Post-Colonoscopy and Post-Imaging Colorectal Cancer. *Gastroenterology*. 2018;155(3):909-925.e3. doi:10.1053/j.gastro.2018.05.038
27. el Dallal M, Chen Y, Lin Q, et al. Meta-Analysis of virtual-based chromoendoscopy compared with dye-spraying chromoendoscopy standard and high-definition white light endoscopy in patients with inflammatory bowel disease at increased risk of colon cancer. *Inflammatory Bowel Diseases*. 2020;26(9):1319-1329. doi:10.1093/ibd/izaa011
28. Hassan C, Spadaccini M, Iannone A, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointestinal Endoscopy*. 2021;93(1):77-85.e6. doi:10.1016/j.gie.2020.06.059
29. Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: A systematic review and meta-analysis. *Endoscopy*. 2021;53(3):277-284.

doi:10.1055/a-1201-7165

30. Bisschops R, East JE, Hassan C, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2019. *Endoscopy*. 2019;51(12):1155-1179. doi:10.1055/a-1031-7657
31. Fernández-Esparrach G, Bernal J, López-Cerón M, et al. Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. *Endoscopy*. 2016;48(9):837-842. doi:10.1055/s-0042-108434
32. Sánchez-Montes C, Sánchez FJ, Bernal J, et al. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy*. 2019;51(3):261-265. doi:10.1055/a-0732-5250
33. Houwen BBSL, Hassan C, Coupé VMH, et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. 2022;54(1):88-99. doi:10.1055/a-1689-5130
34. Zachariah R, Samarasena J, Luba D, et al. Prediction of Polyp Pathology Using Convolutional Neural Networks Achieves "resect and Discard" Thresholds. *American Journal of Gastroenterology*. 2020;115(1):138-144. doi:10.14309/ajg.0000000000000429
35. Ichimasa K, Kudo SE, Mori Y, et al. Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. *Endoscopy*. 2018;50(3):230-240. doi:10.1055/s-0043-122385
36. London AJ. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Center Report*. 2019;49(1):15-21. doi:10.1002/hast.973
37. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI Extension. *The BMJ*. 2020;370:1-14. doi:10.1136/bmj.m3210
38. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health*. 2020;2(10):e537-e548. doi:10.1016/S2589-7500(20)30218-1

39. Zhang Y, Li F, Yuan F, et al. Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence. *Digestive and Liver Disease*. 2020;52(5). doi:10.1016/j.dld.2019.12.146

40. Namikawa K, Hirasawa T, Nakano K, et al. Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: Comparison between the original and newly developed systems. *Endoscopy*. 2020;52(12):1077-1083. doi:10.1055/a-1194-8771

41. Soffer S, Klang E, Shimon O, et al. Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. *Gastrointestinal Endoscopy*. 2020;92(4):831-839.e8. doi:10.1016/j.gie.2020.04.039

42. Mascarenhas Saraiva M, Afonso J, Ribeiro T, et al. Artificial intelligence and capsule endoscopy: automatic detection of enteric protruding lesions using a convolutional neural network. *Revista Española de Enfermedades Digestivas*. Published online 2021. doi:10.17235/reed.2021.7979/2021

43. 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. *Gastrointestinal Endoscopy*. 2020;92(1):144-151.e1. doi:10.1016/j.gie.2020.01.054

44. Mohan BP, Khan SR, Kassab LL, et al. High pooled performance of convolutional neural networks in computer-aided diagnosis of GI ulcers and/or hemorrhage on wireless capsule endoscopy images: a systematic review and meta-analysis. *Gastrointestinal Endoscopy*. 2021;93(2):356-364.e4. doi:10.1016/j.gie.2020.07.038

45. Repici A, Badalamenti M, Maselli R, et al. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology*. 2020;159(2). doi:10.1053/j.gastro.2020.04.062

46. Su JR, Li Z, Shao XJ, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). *Gastrointestinal Endoscopy*. 2020;91(2):415-424.e4. doi:10.1016/j.gie.2019.08.026

47. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *The Lancet*

Gastroenterology and Hepatology. 2020;5(4):352-361. doi:10.1016/S2468-1253(19)30413-3

48. Aziz M, Fatima R, Dong C, Lee-Smith W, Nawras A. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: A systematic review with meta-analysis. *Journal of Gastroenterology and Hepatology (Australia)*. 2020;35(10):1676-1683. doi:10.1111/jgh.15070
49. Lui TKL, Guo CG, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointestinal Endoscopy*. 2020;92(1):11-22.e6. doi:10.1016/j.gie.2020.02.033
50. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*. 2019;68(1):94-100. doi:10.1136/gutjnl-2017-314547

Conflict of interest.

Oswaldo Ortiz is funded through Rio Hortega research grant form Instituto Carlos III. Other authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report.

Table 1. Artificial intelligence technologies applied to endoscopy for esophageal lesions.

Author(s)	Lesion	Study design	Type of artificial intelligence	Data sets	Results AUC Se/Sp/Accuracy
Esophagus					
De Groof et al. Gastroenterology 2020[5]	Early neoplasia in BE	Validation study. Retrospective	Deep learning CAD system GastroNet	5 data sets. **Pre-training** 494,364 endoscopic images from all intestinal segments. **Training** 1704 esophageal WLE HD images of early-stage neoplasia in BE and non-dysplastic BE.	Classify between neoplasms and non-dysplastic BE from 80 patients images: 89% Accuracy, Se 90%, Sp 88%88%
Swager et al GIE 2017[6]	Dysplasia in BE with the use of Volumetric laser endomicroscopy (VLE)	Retrospective	Computer aided algorithm to detect dysplasia on ex vivo images	60 VLE images from ex vivo VLE-histology correlations, obtained from BE patients (30 non-dysplastic BE and 30 high-grade dysplasia/early adenocarcinoma).	Layer: 83%-93% Signal intensity distribution: 83%-87%
Van de Sommen F.et al Endoscopy 2016[7]	Early neoplastic lesions in Barrets esophagus	Validation study	Machine learning, algorithm, which employed specific texture, color filters.	100 images from 44 patients with Barrett's esophagus, a computer.	Per-image analysis with a Se and Sp of 0.83. At patient level, Se and Sp of 0.86 and 0.87, respectively.

Table 2. Artificial Intelligence applied to endoscopy for gastric pathology.

Stomach					
Author(s)	**Evaluated indication**	**Study design**	**Type of artificial intelligence**	**Training process/dataset**	**Results AUC Se/Sp/Accuracy**
Wu L. t al Endoscopy 2021[10]	AI (ENDOANGEL) for **endoscopy quality and detection of early gastric cancer.**	RCT	DCNN and deep reinforcement learning.	1050 patients were randomized. 498 ENDOANGEL and 504 control group.	ENDOANGEl group had fewer blinded spots and longer inspection time. Accuracy 84.7%, Se 100%, and Sp of 84.3% for detecting gastric cancer.
Zhang Y., et al Dig Liv Dis 2020.[39]	Improve the **diagnostic of Chronic atrophic gastritis.**	Retrospective	CNN system	5470 images of antrum from 1699 patients of which 3042 had atrophic gastritis.	Accuracy 94.2% Se: 94.5%, Sp: 94%
Ikenoyama Y. Digestive Endoscopy 2021[11]	To compare diagnostic ability between CNN and endoscopist for **detection of early gastric cancer**.	Retrospective and test dataset	CNN system	13,584 WLE without magnification endoscopic images from 2639 of gastric cancer. test dataset (2940 images from 140 cases).	Se, Sp, PPV and NPV for the CNN were 58.4%, 87.3%, 26.0%, and 96.5%, respectively. For the endoscopists were 31.9%, 97.2%, 46.2%, and 94.9%, respectively.
Li L., et al. Gastric Cancer 2020.[12]	**Detection of EGC** with magnifying NBI.	Retrospective	CNN system	386 images with NBI and magnification of non-cancerous lesions and 1702 images of early gastric cancer for training. Then 341 images (171 non-cancerous and 170 EGC) to validate and compare with endoscopists	Se, Sp, and accuracy of CNN system were 91.18%, 90.64%, and 90.91%, respectively
Yoon HJ et al J Clin Med 2019[19]	**Invasion depth in early gastric cancer.**	Retrospective	CNN system based on a visual geometry group-16 model	11539 WLE endoscopic images (896 T1a-EGC, 809 T1b-EGC, and 9834 non-EGC).	AUC for EGC detection and depth prediction 0.981 and 0.851, respectively
Zhu Y. et al. GIE 2019[15]	**Invasion depth in gastric cancer**	Retrospective	CNN system developed through transfer learning leveraging a state-of-the-art pretrained CNN architecture, ResNet50	790 images as development and 203 test dataset.	**AUC:** 0.94 **Se:** 46.5% **Sp:** 95.5% **Accuracy:** 89%
Hu H. et al GIE 2021[13]	**Detection of EGC** with NBI with magnifying endoscopy.	Cohort	VGG-19 architecture (Visual Geometry Group [VGG], Oxford University	1777 Magnifying-NBI images from 295 cases. training cohort (n = 170), an internal test cohort (n = 73)	**Similar predictive performance as the senior endoscopists** (Acu: 0.77 vs 0.755, P = .355; Se: 0.792 vs 0.767, P = .183; Sp: 0.745 vs .0742, P = .931) **but better than**

AI: Artificial intelligence**, AUC:** area under the receiver operating characteristic curve**, CAD:** Computer aided diagnosis**, CNN:** Deep convolutional neural network, **EGC**: Early gastric cancer. **NBI**: Narrow-band imaging, **NPV:** Negative predictive value, **PPV:** Positive predictive value, **RCT:** randomized controlled trial, **RR:** relative risk, **Se:** sensitivity, **Sp:** specificity, **VCE:** Video-capsule endoscopy

Table 3. Artificial intelligence technologies applied to small bowel pathology.

Small bowel pathology					
Author(s)	**Lesion**	**Study design**	**Type of artificial intelligence**	**Training process**	**Results AUC Se/Sp/Accuracy**
Soffer S. et al Gastrointestinal Endoscopy 2020[41]	Deep learning implementation in wireless capsule endoscopy. included detection of ulcers, polyps, celiac disease, bleeding, and hookworm.	Systematic review and meta-analysis	-	19 retrospective studies using deep learning technologies.	**Accuracy** was above 90% for most studies. Pooled Se and Sp for ulcer detection 0.95 and 0.94 respectively. **Pooled Se and Sp** for bleeding or bleeding source were 0.98 and 0.99 respectively.
Mascarenhas M. et al *Rev Esp Enferm Dig. 2021*[42]	Protruding lesions in small bowel during video capsule endoscopy	Tested retrospectively	CNN based system	18625 VCE images (2830 showing protruding lesions and 15795 normal mucosa) Training and validation datasets were built in a 80%/20% distribution	**Accuracy** 92.5% with a sensitivity of 93.8 and specificity of 96.5%. The system analyzed the image with a rate of 70 frames per second.
Saito H. et al. GIE 2020.[42]	Protruding lesions (polyps and tumors	Tested retrospectively	CNN based system	30584 images of 292 patients	AUC: 0.911 Se 90.7%, 7Sp 99.8%
Mohan B.P. et al GIE 2021[43]	Diagnosis of gastrointestinal ulcers and/or hemorrhage	Systematic review and meta-analysis	-	9 retrospectives studies evaluating performance of CAD systems using wireless capsule endoscopy	Accuracy 95% Se 95.5%, Sp s 95.8%, PPV 95.8%, and NPV 96.8%. I2% heterogeneity was negligible except for the pooled positive predictive value.

REVISTA ESPAÑOLA DE
ENFERMEDADES DIGESTIVAS
The Spanish Journal of Gastroenterology

Table 4. Performance of Artificial Intelligence technologies based on deep learning evaluating colonic lesions.

Colon					
Author(s)	**Lesion**	**Study design**	**Type of artificial intelligence**	**Training process**	**Results AUC Se/Sp/Accuracy**
Repici A. et al Gastroenterology 2020.[44]	ADR	Multicenter RCT comparing HD-WLE vs. CADe (Gi Genius, Medtronic) aided colonoscopy	CCN trained and validated using 2684 videos of histologically confirmed polyps from 840 patients	685 subjects	Increase in ADR in CAD group 54.8% vs. 40.4% RR 1.30 APC higher in CADe group (mean, 1.07 ±1.54) than in the control group (mean 0.71 ± 1.20) (incidence rate ratio, 1.46; 95% CI, 1.15–1.86)
Su J. et al Gastrointestinal endoscopy 2020[45]	ADR	RCT	5 DCCN automatically time the withdrawal phase, supervise stability, evaluate bowel preparation, and detect colorectal polyps in real time.	659 patients	DCCN significantly increased the ADR (0.289 vs 0.165, P < .001) and the mean number of adenomas per procedure (0.367 vs 0.178, P < .001)
Gong D.Wu L. Lancet Gastroenterol Hepatol 2020[46]	ENDOANGEL to monitor withdrawal and blind spots.	RCT comparing ADR between unassisted colonoscopy and assisted with ENDOANGEL system.	DCCN using three datasets came from stored data for more than 5000 patients	704 patients	In the intention-to-treat, ADR was greater in the ENDOANGEL group 16% vs 8% [OR] 2·30, 95%; p=0·001). per-protocol analysis, findings were similar, 17% vs 8% assigned control having one or more adenomas detected (OR 2·18, 95% CI 1·31–3·62; p=0·0026).
Hassan C. GIE 2021[28]	Performance of CADe systems pooled ADR of 5 RCT.	Systematic review and meta-analysis	-	4354 patients	Pooled ADR was significantly higher in the CADe group than in the control group 36.6% vs 25.2%; RR, 1.44; 95% [CI], 1.27-1.62; P < .01; I2 Z 42%).
Barua I et al Endoscopy 2021[29]	To compare AI with standard colonoscopy for detection of polyps, adenomas and colorectal cancer.	Systematic review and Meta-analysis	-	3175 patients from 5 RCT	Increase in ADR (29.6% vs.19.3 %) in CAD group with high certainty. PDR (45.4 % vs. 30.6%); RR 1.48 with high certainty. There was no difference in detection of advanced

					adenomas.

REVISTA ESPAÑOLA DE
ENFERMEDADES DIGESTIVAS
The Spanish Journal of Gastroenterology

					There was no difference in detection of advanced adenomas.
Aziz M. et al J Gastroenterol Hepatol 2020[47]	To assess impact of DCNN-based AI-assisted colonoscopy in improving the ADR and PDR	Systematic review and Meta-analysis	-	3 studies with 2815 patients	AI colonoscopy resulted in significantly improved ADR (32.9% vs 20.8%, RR: 1.58, 95% CI 1.39–1.80, P = < 0.001) and PDR (43.0% vs 27.8%, RR: 1.55, 95% CI 1.39–1.72, P = < 0.001) compared with standard colonoscopy
Lui T.K.L. et al. Gastrointestinal Endoscopy 2020[48]	Accuracy of AI on histology prediction and detection of polyps.	Systematic review and meta-analysis.	-	7680 images of colorectal polyps with NBI and without magnification from 18 studies	For characterization of diminutive polyps with non-magnifying NBI, the pooled NPV was 95.1%. For polyp detection, the pooled AUC was 0.90 with a Se of 95.0 and a Sp of 88.0%.
Zachariah R. et al. Am J Gastroenterol 2020[34]	**Prediction of polyp histology and achieve PIVI thresholds.**	Retrospective validation study.	CNN-based optical pathology model using TensorFlow and pre-trained on ImageNet	6,223 images of colorectal polyps underwent 5-fold cross-training (80%) and validation (20%)	NPV for adenomas was 97% among diminutive recto-sigmoid polyps. Results were independent of use of NBI or white light. Surveillance interval concordance was 93%.
Byrne M.F. et al Gut 2019[49]	**Optical diagnosis** assessment of video images of colorectal polyps	Retrospective training study.	CNN system	125 videos of diminutive polyps	No sufficient confidence to predict histology of 19 (15%) polyps in the test set. For the remaining 106 polyps, the accuracy was 94%, Se 98%, Sp was 83% NPV 97% and PPV of 90%.
Stidham R.W. et al. JAMA Netw Open 2019[24]	**Endoscopic disease severity in ulcerative colitis.**	Retrospective	CNN system	6514 images from 3082 patients	AUC of 0.966; a PPV of 0.87 with a Se of 83.0% and Sp of 96.0% and NPV of 0.94 for distinguishing endoscopic remission from moderate to severe disease.

Table 4. Performance of Artificial Intelligence technologies based on deep learning evaluating colonic lesions.

ADR: adenoma detection rate, **AI:** Artificial intelligence**, AUC:** area under the receiver operating characteristic curve**, CAD:** Computer aided diagnosis**, CNN:** Deep convolutional neural network, **NBI**: Narrow-band imaging, **NPV:** Negative predictive value, **PIVI:** Preservation and Incorporation of Valuable Endoscopic Innovations, **PPV:** Positive predictive value, **RCT:** randomized controlled trial, **RR:** relative risk, **Se:** sensitivity, **Sp:** specificity, **VCE:** Video-capsule endoscopy.

REVISTA ESPAÑOLA DE
ENFERMEDADES DIGESTIVAS
The Spanish Journal of Gastroenterology

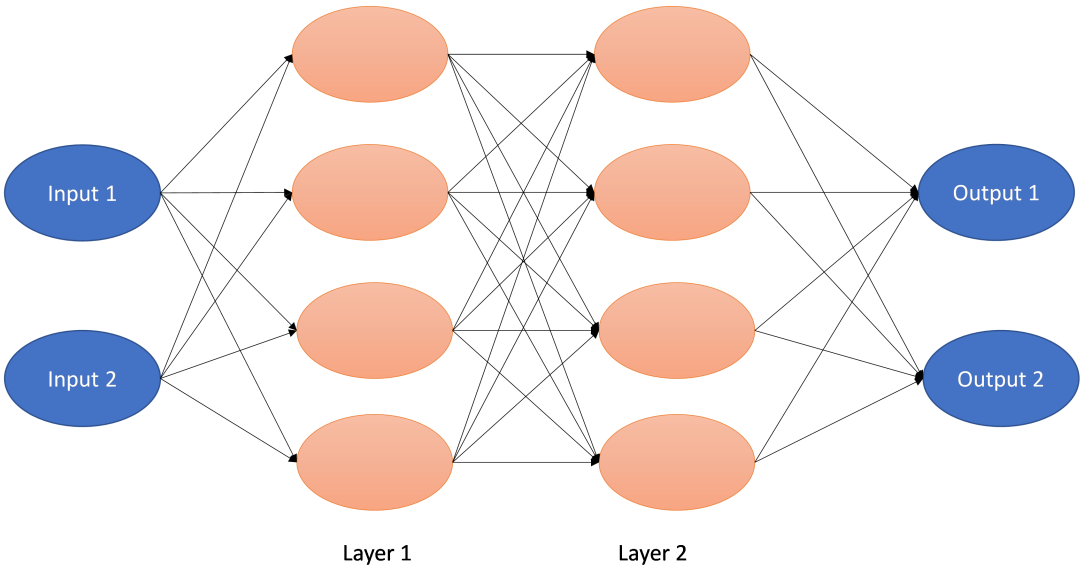

Figure 1. *Deep convolutional network is composed of a number of interconnected hierarchic layers that results in an output emulating the neuronal connections.*

Figure 1. Deep convolutional network is composed of a number of interconnected hierarchic layers that results in an output emulating the neuronal connections.

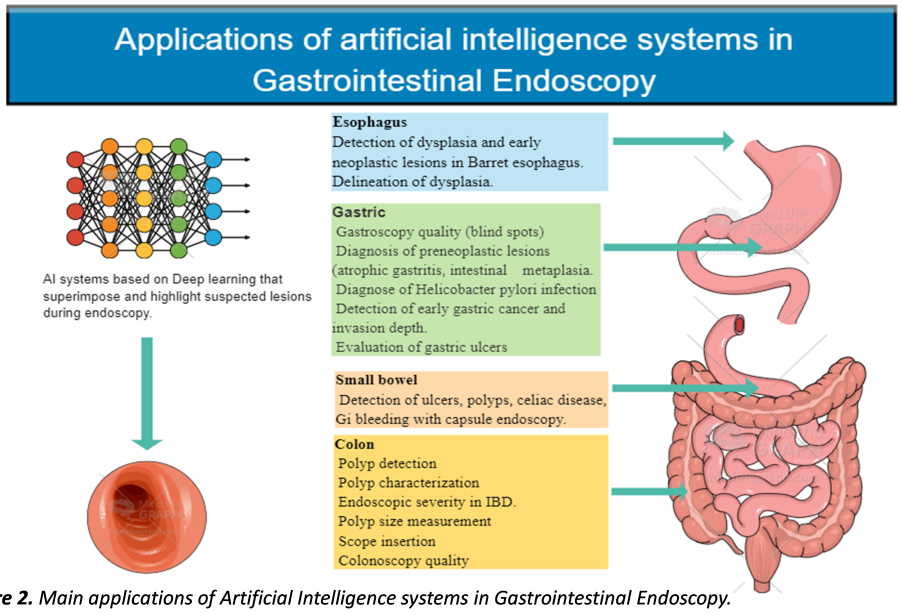

Figure 2. *Main applications of Artificial Intelligence systems in Gastrointestinal Endoscopy.*

Figure 2. Main applications of Artificial Intelligence systems in Gastrointestinal Endoscopy.