91336_Auto_Edited.docx

WORD COUNT	3477	TIME SUBMITTED	16-MAY-2024 09:01AM
		PAPER ID	108998167

4

Name of Journal: *Artificial Intelligence in Gastroenterology*
Manuscript NO: 91336
Manuscript Type: EDITORIAL

Will Artificial intelligence reach any limit in gastroenterology?

Artificial intelligence in gastroenterology.

Joseph Bou Jaoude, Rose Al Bacha, Bassam Abboud

Abstract

Artificial intelligence is gaining ground in all areas. The medicosurgical specialties, more particularly gastroenterology, have not escaped this evolution. The aim of this work was to see if this artificial intelligence will reach its limit in the field of the digestive tract and endoscopy.

Key Words: Artificial intelligence; digestive tract; gastroenterology; gastroscopy; coloscopy.

Bou Jaoude J, Al Bacha R, Abboud B. Will Artificial intelligence reach any limit in gastroenterology? . *Artif Intell Gastroenterol* 2024; 0(0): 91336
URL: https://www.wjgnet.com/2644-3236/full/v0/i0/91336.htm
DOI: https://dx.doi.org/10.35712/aig.v0.i0.91336

Core Tip: The field of gastrointestinal endoscopy is an essential tool in the management of digestive diseases. Despite the ongoing development of endoscopic technology, the incidence of missed pre-neoplastic and neoplastic lesions remains high. This is attributed to the operator-dependent nature of endoscopy, resulting in variability in detection rates and the characterization of lesions among endoscopists. In order to

enhance the performance of the endoscopic procedure, it is imperative to minimize the "cognitive errors" made by the endoscopist. Artificial Intelligence, being operator-independent, could potentially serve as an unlimited solution.

INTRODUCTION

The field of gastrointestinal endoscopy is an essential tool in the management of digestive diseases. Technology is essential for the advancement of endoscopy. Presently, white-light endoscopy (WLE) with high resolution stands as the standard technology that enables endoscopists to detect and characterize lesions more accurately. However, despite this, even expert endoscopists can overlook several lesions, including small and flat ones.

In an effort to morphologically predict the malignant potential of digestive lesions in real-time, several classification systems have been endorsed by scientific societies. These systems categorize lesions based on morphology (sessile, slightly raised, or excavated) or through a detailed examination of vascular and mucosal patterns using optical image-enhancing technology known as virtual chromo-endoscopy. Consequently, the assessment of invasion depth or lymph node involvement plays a crucial role in clinical decision-making, determining whether the lesion is surgically or endoscopically resectable.

Despite the ongoing development of endoscopic technology, the incidence of missed pre-neoplastic and neoplastic lesions remains high. This is attributed to the operator-dependent nature of endoscopy, resulting in variability in detection rates and the characterization of lesions among endoscopists. The existence of this skills gap can be explained by the extended learning curve associated with adopting new technologies.

In order to enhance the performance of the endoscopic procedure, it is imperative to minimize the "cognitive errors" made by the endoscopist. Artificial Intelligence (AI), being operator-independent, could potentially serve as an unlimited solution.

As endoscopy fundamentally depends on high-quality images, it presents an appealing domain for AI, which comprises computer processes performing complex tasks to simulate the human brain. Alan Turing, one of its founders, defined AI as "the ability of a computer to achieve human performance in cognitive tasks." Thus, this concept combined the fields of medical knowledge and machine tools.In order to allow the machine to learn and make decisions on its own, deep learning technique (DL) was innovated as a major and prominent transformation of machine learning (ML). DL automatically extracts input features from targeted images, demonstrating the ability to explore all pixels without experiencing transitory lapses in attention or fatigue. As a result, DL emerges as a promising technology, serving as a reliable "second observer" independent of the endoscopist's performance. DL encompasses two primary tasks: real-time detection or computer-aided detection (CADe) and real-time characterization or computer-aided diagnosis (CADx). Given that navigation software enhances mucosal exposure, CADe assists endoscopists in reducing the miss rate of lesion detection. Simultaneously, CADx aims to predict the histologic and optical diagnosis of pre-neoplastic lesions without the need for biopsy, as well as estimating the depth of invasion in malignant lesions to facilitate optimal therapeutic decision-making.

Moreover, DL can reduce the cost and the procedure time by abandoning of random biopsies in favor to targeted one and by avoiding unnecessary resection of non-neoplastic lesions. Also, DL can evaluate the quality of our endoscopic procedure by identifying many parameters: landmarks, blind spots, measurement of withdrawal speed and assessment of the level of mucosal cleansing making our surveillance protocol more effective.

Thus, AI allows human-machine interaction leading to a transfer of human knowledge from experts to the entire gastroenterological community.

AI applications in diagnostic gastro-intestinal endoscopy.

AI applications in clinical gastrointestinal diseases are continuously expanding and evolving into new areas. AI has been embraced for its robust self-learning capability and unbiased nature. Real-time AI assists endoscopists throughout the entire digestive tract, including the upper, middle, and lower parts, as well as the hepato-biliary tree and pancreatic gland.

A- Lower Part

1. Colorectal polyps:

In the gastrointestinal field, the primary application of AI involves the development of DL CNN models for the detection and diagnosis of polyps during colonoscopy.

i. Detection of Colorectal Polyps Using CADe

It has been established that the removal of pre-neoplastic polyps reduces the risk of colorectal cancer (CRC). However, endoscopy is operator-dependent, and the adenoma detection rate (ADR) varies widely from 7% to 53% among colonoscopists[1] while post-colonoscopy interval CRC constitutes nearly 8% of all diagnosed CRC (2). The initial application of AI technology in gastrointestinal endoscopy (GE) was the detection of colorectal polyps, with most research focusing on the management of colorectal polyps. In 2018, Urban *et al* (3) and Misawa *et al* (4) reported the two earliest applications of CADe on video clips. Their algorithms demonstrated an accuracy of ≥90%. In 2019, Wang *et al* (5) conducted the first randomized controlled trial. Since then, numerous prospective randomized controlled trials (6-10), as well as meta-analyses(11), have been published, involving different AI systems and training ([]). Consequently, CADe for polyp detection has been shown to increase the adenoma detection rate, at least comparable to that assessed by experienced endoscopists, as recommended by the European Society of Gastrointestinal Endoscopy (ESGE) guidelines (12).

ii. Characterization of Colorectal Polyps Using CADx

1. Polyp ≤ 5 mm

According to the current European Society of Gastrointestinal Endoscopy (ESGE) guidelines, polyps ≤ 5 mm with adenomatous structures need to be removed and sent for histopathological analysis. Diminutive polyps located in the recto-sigmoid, characterized as hyperplastic by virtual chromo-endoscopy, can either be "left in situ" or undergo the "resect and discard" approach. CADx tools, when combined with CADe, can assist endoscopists in real-time colonoscopy by distinguishing between neoplastic (adenoma or serrated) and non-neoplastic (hyperplastic) polyps. Consequently, in the case of non-neoplastic polyps, the "diagnose and leave" strategy reduces the need for polypectomy. Similarly, for neoplastic diminutive polyps, the "resect and discard" strategy minimizes the necessity for histopathological processing. In clinical practice, these two strategies, supported by CAD systems, contribute to reduced costs and procedure time. Indeed, many centers have developed CADx tools with white-light endoscopy (WLE), narrow-band imaging (NBI), and endocystoscopy (13-15). Their published results align with the parameters outlined by the American Society of Gastrointestinal Endoscopy Preservation and Incorporation of Valuable Endoscopic Innovation (ASGE PIVI).

2. Advanced Subtle Neoplastic (Flat and Serrated)

The increased detection of non-advanced adenomas alone cannot reduce the interval colorectal cancer. Consequently, developing AI systems to enhance the detection of advanced polyps is now considered a priority, as they pose the highest risk of developing colorectal cancer (CRC). Most CADx studies lack data about sessile serrated lesions (SSL) and flat polyps. When SSL are described, the majority is located in the recto-sigmoid and is diminutive. Only one recent prospective study, utilizing video datasets enriched with flat, SSL, and advanced colorectal polyps, evaluated AI performance against endoscopists. The AI-based algorithm achieves high per-polyp sensitivities for the diagnosis of advanced polyps (16)

3. Malignancy in Colorectal Polyps

Endoscopists must assess the level of submucosal invasion in T1 colorectal cancer without resorting to biopsy to decide whether to perform endoscopic or surgical resection. AI emerges as an ideal tool to offer valuable guidance to endoscopists 2 japanese AI studies were conducted using CNN algorithms to differentiate between T1a and T1b. the initial study was a randomized one and achieved 94% of accuracy; however the second one ranged only 81.4% of accuracy (17-18).

iii. Computer-aided quality assessment of colonoscopy technique.

AI, functioning as a virtual endoscopist, can complement the expertise of endoscopists in reducing the rate of missed polyps visible on the screen. However, the quality of the colonoscopy procedure relies on additional parameters such as incomplete mucosal exposure, blind spots, withdrawal speed, and the degree of bowel cleansing.Currently, AI is developing new systems to measure these parameters during the procedure, alongside CADe and CADx, to address exposure errors. Consequently, computer-aided quality assessment objectively evaluates the time spent exploring different segments of the colon, the quality of fold examination, and mucosal cleansing (19). Therefore, in the future, we can objectively determine the quality of colonoscopy for the optimal surveillance protocol of colorectal cancer (CRC).

B-Upper body

1. Esophagus

In a recent multicenter study of upper gastrointestinal endoscopies, a 6.4% esophageal cancer miss rate was reported (20). Due to the capability of DL to explore images beyond the reach of the human eye, it has been employed in the analysis of endoscopic images related to esophageal and stomach diseases. Wu *et al* (21) utilized a DL model and demonstrated promising outcomes in the classification and segmentation of individual esophageal lesions. Consequently, several CAD systems have been recently tested in clinical settings.

i. Precursor Lesion of Esophageal Squamous Cell Neoplasia (ESCC)

Intrapapillary capillary loops (IPCL) observed through virtual chromoendoscopy (NBI) have been classified as a precancerous lesion of ESCC, correlating with depth invasion. Everson *et al* demonstrated that a DL model was an efficient, accurate, and reliable tool for classifying IPCL patterns as normal or abnormal (22). In two separate studies, Zhao *et al* (23) and Yuan *et al* (24) compared the accuracy of AI systems to that of endoscopists. AI models significantly enhance the ability of junior endoscopists to diagnose IPCL abnormalities and depth invasion of ESCC

ii. ESCC

A recent literature review demonstrated high diagnostic accuracy for AI in esophageal squamous cell carcinoma (ESCC) (25). Over the years, extensive datasets have supported the overall diagnostic performance of AI for both superficial and advanced esophageal squamous cancer. Numerous studies have indicated that, in comparison to the experience of endoscopists, AI accuracy in detection was comparable or even higher than that of experts (26-28). In therapeutic decisions for ESCC, which depend on the depth of invasion, Zhang *et al* conducted a multicenter study using an AI-based computer-aided diagnosis (CAD) model that simulated radiologists' diagnoses of lymph node metastasis (29). The results from AI systems significantly outperformed those of human diagnostics. Additionally, Tokai *et al* published a comparative study between a DL CNN model and endoscopists to determine ESCC depth invasion. The results demonstrated that AI algorithms surpassed the performance of all endoscopists (30). Given these promising results, AI-assisted diagnostic techniques should be considered for adoption in future clinical practice.

iii. Barett’s esophagus-related neoplasia (BE)

It is established that Barrett's esophagus (BE) is a precursor of esophageal adenocarcinoma (EAC). BE represents an exemplary application of AI systems, showcasing their capability in lesion identification and determining the degree of malignancy.

Pan *et al* demonstrated the ability of an AI model in identifying and classifying BE according to the Prague classification (31). To enable endoscopists to successfully detect dysplasia or EAC in BE, several AI studies have achieved high sensitivity, specificity, and accuracy, meeting the parameters outlined by the Preservation and Incorporation of Valuable Endoscopic Innovation (PIVI) (32-35). Two meta-analyses have reached similar conclusions (36-37).

In order to choose the optimal treatment, the identification of sub-mucosal invasion of BE-related neoplasia is mandatory. A retrospective multicenter study evaluated the performance of DL algorithms in discriminating between T1a and T1b cancer (38). The AI model demonstrated comparable performance to experienced endoscopists.

2. Stomach

Gastric Precancerous lesions

Helicobacter pylori infection can produce chronic atrophic gastritis (CAG) and gastric intestinal metaplasia (GIM). CAG and GIM are precancerous lesions associated with an increased risk of gastric cancer (CG) development (39). Thus endoscopic surveillance of the precancerous lesions is mandatory to detect CG in an early stage, early gastric cancer (EGC).The diagnosis of EGC is difficult because the sensitivity of endoscopic diagnosis of CAG is only 42% in a large study and the overall rate of missed neoplasia at endoscopy varies between 8.3 and 10% (40).

AI models may improve the diagnosis accuracy and aid the endoscopist in the detection and staging of precancerous lesions.

AI in the detection of gastric precancerous lesions and HP infection

Concerning CAG, in 2 studies, AI models were compared to endoscopists. Zhang *et al* (41) used the CNN model to detect CAG in 1699 patients. It outperformed 3 expert endoscopists with a sensitivity, specificity, and accuracy of 95%, 94%, and 94% respectively. Guimaraes *et al* (42) reported a 93% accuracy with WLE images.

Concerning GIM, Yan *et al* (43) developed a CNN-CAD model with ME-NBI. It reached an accuracy of 89% compared to 84% accuracy for expert endoscopists.

Concerning HP infection, Zheng *et al* (44) developed a CAD system to detect HP infection status based on endoscopic images without the need for biopsies. The CNN systems reached an accuracy of 92%. Nakashima *et al* (45) used a DL model with WLE and blue light imaging (BLI). The DL model had a AUC of 0.96 with BLI, and 0.66 with WLE.

AI in the detection of EGC

Li *et al* (46) developed a CNN model on images of benign lesion and EGC. The AI model has a diagnostic accuracy of 91% compared to an accuracy of 87% when used by expert and 70-74% for non-expert endoscopists. Horiuchi *et al* (47) used a CAD system to detect EGC using NBI videos and compared to 11 expert endoscopists in NBI. Only 2 endoscopists were outperformed by the CAD systems.

AI in the prediction of invasion depth of EGC

Nagao *et al* (48) developed a CNN-CAD system by using images of GC that underwent endoscopic resection or radical surgery to evaluate the accuracy of AI to determine invasion depth. They found that the CAD system can predict the invasion depth with a sensitivity of 84%-75%, specificity of 99%-80% and accuracy of 94%-94% during WLE and NBI images respectively. Yoon *et al* (49) analyzed images of GC (T1a and T1b) to predict invasion depth with AUC of 0.85. this accuracy was significantly lower in undifferentiated lesion.

C- Small Bowel

1. Inflammatory bowel disease

Recently, the therapeutic goals for patients with inflammatory bowel disease (IBD) have shifted toward mucosal healing, defined by endoscopic evaluation. However, histologic evaluation is essential to predict the risk of relapse and colon cancer.

This gastrointestinal field has emerged as a new area for AI, utilizing data from endoscopic images, video capsule endoscopy images, histology, magnetic resonance imaging images, laboratory studies, and genetics. Numerous studies with meta-analyses using machine learning (ML) and deep learning (DL) systems have aimed to detect Crohn's disease and ulcerative colitis with high sensitivity and accuracy (49). Additionally, AI studies utilizing ML and DL CNN systems have achieved a high level of accuracy in predicting disease severity for IBD (50).

2. Villous atrophy

Celiac disease is the primary cause of villous atrophy and still be undiagnosed in 50% of cases. A study conducted by Gadermayr *et al* (51)achieved a high accuracy of 94% but that requires water immersion. Also, studies with video capsule endoscopy showed an accuracy > 90% (52-53). These studies were done under special conditions with high probability of suspicion.It is mandatory to make the diagnosis in routine endoscopy. A recent retrospective study done by Scheppach *et al* (54) compared AI algorithms to performance of fellows and experts on routine endoscpopy. The results showed that AI improve significantly the performance of all endoscopists with a stable performance.

D- Pancreas

Endoscopic ultrasound (EUS) is a reliable tool for the detection and staging of pancreatic lesions, particularly pancreatic cancer. EUS-FNB is a well-established diagnostic tool for pancreatic cancer, demonstrating a specificity and sensitivity greater than 90%. However, the EUS technique is operator-dependent, exhibiting inter-observer variability, making it an ideal platform for AI applications.

1. AI in EUS for detection of pancreatic cancer

Three retrospective studies were conducted using DL algorithms, demonstrating high sensitivity, specificity, and accuracy in diagnosing pancreatic cancer (55-57). Additionally, Goyal *et al* conducted a systematic review of 11 studies on the role of AI-assisted EUS models in diagnosing pancreatic cancer, revealing that AI algorithms had a high potential for detecting pancreatic cancer (58).

2. AI in EUS differentiation pancreatic cancer from benign lesions

i. Chronic Pancreatitis

Chronic pancreatitis still mimics pancreatic cancer in radiologic features and is also considered a risk factor for the development of pancreatic cancer. Five studies were conducted with DL algorithms, reporting high accuracy, sensitivity, and specificity (59-

63). However, these studies were heterogeneous with a small patient population. Hence, two recent prospective multicenter studies using DL models were published, validating the aforementioned findings (64-65). Therefore, AI-assisted EUS can be a validated tool in clinical practice to differentiate pancreatic cancer from chronic pancreatitis with accepted results.

ii. Auto-immune pancreatitis

Marya *et al* (66) using DL model to differentiate between PC and AIP conducted the unique study published till now. The high sensitivity and specificity encourage the use of AI to assist EUS endoscopists in this field.

II- Limitations

A- Input Data

DL tasks rely on a database that has been used to train AI algorithms, which must be manually annotated and propagated through frames using dedicated software. The development of DL may be affected by selection biases, wherein the chosen disease, its prevalence, the endoscopic center characterize the patient population, and the number of patients enrolled. Spectrum biases in DL performance can arise from variations in patient population, the number and skills of endoscopists, and the technical characteristics used, such as white-light endoscopy (WLE), advanced imaging, and optical magnification. Consequently, a database from a single institution, lacking diversity and failing to capture all possible variations, can impact the quality of input data, as well as the reproducibility and generalizability of the results. To mitigate these biases, it is essential to establish a quality-monitored central data collection server that aggregates data from all institutions.

B- Algorithm

Studies utilize different AI-assisted models that require images prepared in specific ways. These algorithms may not consistently achieve a high degree of accuracy.

Therefore, it is essential to establish a universal protocol for input data to enhance the efficacy and accuracy of AI-assisted models.

C- Validation

AI findings must undergo clinical validation before being introduced into clinical practice. AI has a valuable advantage when the reference standard is based on histologic verification. However, if not, the reference standard relies on expert endoscopist raters, introducing potential bias. Therefore, AI systems should be validated through randomized trials comparing the standard and new endoscopic modalities. Additionally, these algorithms must be tested on large and cross-institutional datasets.Long-term data on the accuracy of AI-assisted models is lacking. Consequently, there are no results regarding the impact of AI on reducing the incidence and mortality of gastrointestinal cancer.Clinical efficacy evaluation must adhere to established guidelines. The two recommended guidelines are the Preservation and Incorporation of Valuable Endoscopic Innovation (PIVI) statement as a guide for new imaging technology and the European Society of Gastrointestinal Endoscopy (ESGE) guidelines. For example, in cases requiring targeted biopsies, PIVI recommends a per-patient sensitivity of 90% or greater and a specificity of 80% or greater to allow a reduction in biopsies. Therefore, studies must meet these parameters to be approved for clinical practice. Additionally, according to ESGE, the results of AI studies must be comparable to those of experts.

D- Output

There are "black boxes" in the logic of DL algorithm decision-making processes that are not understood or controlled by humans (67). Consequently, AI can make mistakes, and humans cannot explain or justify the computer's decisions. For instance, physicians have concerns regarding the number of false-positive signals generated by AI. This may cause distraction, prolong procedure time, and be frustrating to the endoscopist, making some users hesitant to use it. Therefore, humans must make the final decision and should not become entirely dependent on AI technology for both diagnostic and therapeutic endoscopies; otherwise, they risk losing their cognitive abilities.

CONCLUSION

In conclusion, because the gastrointestinal (GI) field relies on imaging, AI-assisted algorithms continue to explore new GI organs and diseases. The growth and applications of AI increase exponentially with the development of computer science and may reach no limit. However, we must be careful about how we use it and preserve our independence in the final decision. Additionally, to achieve better results in AI studies in the future, collaboration between academic and private gastroenterologists and the industry must be closer, aiming to improve the quality, utility, ease of use, and accuracy of AI models. We hope that AI-assisted diagnostic techniques will be widely used in GI diseases because AI is an unavoidable tool in GI endoscopy.

91336_Auto_Edited.docx

ORIGINALITY REPORT

7%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.karger.com Internet	157 words — 4%
2	Chin Hock Nicholas Tee, Rajesh Ravi, Tiing Leong Ang, James Weiquan Li. "Role of artificial intelligence in Barrett’s esophagus", Artificial Intelligence in Gastroenterology, 2023 Crossref	33 words — 1%
3	www.ncbi.nlm.nih.gov Internet	31 words — 1%
4	f6publishing.blob.core.windows.net Internet	14 words — < 1%
5	Prateek Sharma, Cesare Hassan. "ARTIFICIAL INTELLIGENCE AND DEEP LEARNING FOR UPPER GASTRO-INTESTINAL NEOPLASIA", Gastroenterology, 2021 Crossref	13 words — < 1%
6	Prateek Sharma, Cesare Hassan. "Artificial Intelligence and Deep Learning for Upper Gastrointestinal Neoplasia", Gastroenterology, 2021 Crossref	13 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES < 12 WORDS