74661_Auto_Edited.docx

Usefulness of artificial intelligence in early gastric cancer

Alba Panarese

Abstract

Gastric cancer (GC) is a major cancer worldwide, with high mortality and morbidity. Endoscopy, important for the early detection of GC, requires trained skills, high-quality technologies, surveillance and screening programs. Early diagnosis allows a better prognosis, through surgical or curative endoscopic therapy. Magnified endoscopy with virtual chromoendoscopy remarkably improve the detection of early gastric cancer (EGC) when endoscopy is performed by expert endoscopists. Artificial intelligence (AI) has also been introduced to GC diagnostics to increase diagnostic efficiency. AI improves the early detection of gastric lesions because it supports the non-expert and experienced endoscopist in defining the margins of the tumor and the depth of infiltration. AI increases the detection rate of EGC, reduces the rate of missing tumors, and characterizes EGCs, allowing clinicians to make the best therapeutic decision, that is, one that ensures curability. AI has had a remarkable evolution in medicine in recent years, moving from the research phase to clinical practice. In addition, the diagnosis of GC has markedly progressed. We predict that AI will allow great evolution in the diagnosis and treatment of EGC by overcoming the variability in performance that is currently a limitation of chromoendoscopy.

THE RELEVANCE DIAGNOSIS OF GASTRIC CANCER

Gastric cancer (GC), the fourth leading cause of cancer in men and seventh in women, is still third for cancer-related deaths worldwide[1]. Its 5-year survival rate is less than

40%[2] and its prognosis is related to the stage at the time of detection. The 5-year survival rate of patients with early gastric cancer (EGC) is 91.5%, whereas it is 16.4% for patients in the advanced stage[2-4]. The screening programs are cost effective in high-incidence regions[1,5] and advanced endoscopic technologies allow endoscopists to diagnose EGC[6-8]; however, optical diagnosis requires a period of training[9].

Recently, the practice of medicine has changed with the development of artificial intelligence (AI) based on image recognition with deep learning (DL) using the convolutional neural network (CNN), which, in upper endoscopy, is trained with endoscopic images and detects GC accurately[10-14]. Several AI-assisted CNN computer-aided diagnosis (CAD) systems have been built, with diagnostic precision in the detection of GC based on different types of endoscopic images. AI helps endoscopists to achieve the accuracy needed for GC screening, surveillance of precancerous, as well as for detecting the depth of invasion of gastric lesions, and when applied to radiological imaging techniques, lymph node and peritoneal metastasis[11-14].

OPTICAL ENDOSCOPIC DIAGNOSIS OF EGC

While computed tomography, endoscopic ultrasound, and positron emission tomography are important for the diagnosis and staging of advanced GC, endoscopy plays an essential role in the early detection of EGC, as it allows the gastric mucosa to be examined directly. Endoscopy with targeted biopsies is the gold standard method for diagnosing EGC, and the accurate diagnosis of EGC through endoscopic imaging is a primary goal for improving the poor prognosis of patients[4,15-17]. Although the quality and accuracy of endoscopic detection are variable between centers and endoscopists, endoscopy is crucial because many early-stage tumors (*i.e.* intramucosal cancer) can be resected endoscopically in a curative manner, with an excellent prognosis at 5 years[4,18-19].

Unfortunately, few endoscopists are experts in advanced endoscopic imaging, and diagnostic accuracy depends largely on the clinical experience of the experts and is influenced by multiple factors, such as training and technologies[9,20]. Ultimately, early

diagnosis and curative treatment are important for prognosis but can be difficult to achieve depending on the endoscopist[10,21]. The false negative rate of GC detected by esophagogastroduodenoscopy is 4.6-25.8[22-24], with higher values for inexperienced endoscopists[9,25]. The diagnostic capacity of endoscopists, due to the endoscopic appearance of EGC, which is usually very subtle, varies widely with regard to the differentiation between GC and gastritis, the prediction of the horizontal extension of GC and the depth of invasion[26].

As lesions of the gastric mucosa develop according to the Correa cascade, from atrophy to intestinal metaplasia, intraepithelial neoplasia and invasive neoplasia [27-28]; improving the accuracy of endoscopic diagnosis of precancerous lesions and EGC through screening and surveillance programs, is useful to reduce the incidence and mortality of GC[29-31]. The standard modality for the detection of EGC is endoscopy with white light imaging (WLI), but its overall sensitivity is not satisfactory (40%-60%)[32]. Magnified endoscopy (ME) with image-enhanced endoscopy techniques such as narrow-band imaging (NBI; Olympus Co., Tokyo, Japan), flexible spectral imaging color enhancement (FICE; Fujifilm Co., Tokyo, Japan), and blue laser imaging (BLI; Fujifilm), improve the accuracy of the detection of gastric lesions[26,33-34]. In particular, ME-NBI, the most frequent technology used in AI studies, achieves significantly better sensitivity, specificity, and accuracy than WLI, facilitating examination of the glandular epithelium in the stomach by observing the microvascular architecture and structure of the microsurface[32,35-39].

However, the virtual chromoendoscopic diagnosis of EGC requires considerable skill and experience[9,38,40-41]. The diagnostic effectiveness of endoscopists non yet trained in differentiating EGC from non-cancerous lesions with ME-NBI is disappointing[9,36,41]. Optical diagnosis can improve with AI-assisted CNN, which has been mainly applied to ME-NBI[14].

AI FOR THE DIAGNOSIS OF EGC

AI, which mimics human cognitive function[42] with its efficient computational power and learning capabilities, can be applied to GC because it processes and analyzes large

amounts of data with systems that classify and recognize lesion images without the need to write complicated image processing algorithms[43]. Therefore, AI could help gastroenterologists in clinical diagnosis and decision-making. Technically, the DL method approximates complex information using a multilayer system (*e.g.,* CNN), in which neural layers connect only to the next layer (Figure 1), overcoming the limitation of the "black box" of previous systems because it shows the reasons for the decisions made[44]. Over the years, new CNN-based systems have been introduced to analyze lesions of the gastric mucosa, using higher quality images and image selection strategies based on evidence from previous experiences. CNN systems in the initial training phase take a few hours to generate the identification system, which can then be used repeatedly; and has a good adaptability as it can be used on multiple platforms for the real-time analysis of JPEG images or video captured by chromoendoscopy. Magnifying chromoendoscopic images can improve the speed and accuracy of CNN diagnostics compared to conventional endoscopy alone[45-46]. Typically, training images are judged by experienced endoscopists and pathologically confirmed, and only endoscopic and chromoendoscopic images with appropriate magnification and typical manifestation for learning the CNN model are selected.

In recent studies, other important outcomes have been added to the main outcome to establish endoscopic resectability, namely the identification of the margins and depth of the lesion[47,49]. Gastric tumors of differentiated intramucous type (m) or infiltrating only the superficial layer of the submucosal (≤ 500 µm: Sm1) can be resected endoscopically, while those that deeply invade the submucosal (> 500 µm: Sm2) are surgically resected because of the risk of lymph node and distant metastases. The optical differentiation between m/Sm1 and Sm2 is often difficult[19].

Using PubMed, Embase, Web of Science, and Cochrane Library databases to search the literature on CAD systems for the diagnosis of EGC, we identified 26 relevant physician-initiated studies through November 2021. Table 1 summarizes the main characteristics of the studies (two single-center prospective[50-51], two multicenter prospective[49,52], and twenty-two retrospective[14,45-48,53-69]): study design; endoscopic modality; main study

aim; and subjects/lesions/images for validation. Table 2 describes the endpoints of the studies.

Selected studies included a diagnostic test on the application of AI in endoscopy for the diagnosis of EGC; the absolute numbers of true-positive, false-negative, true-negative and false-positive; clear information about data and number of images; the description of the algorithms and the process applied to the EGC diagnosis.

To form a training dataset, 11 studies used only WLI images[47,50-53,55-58,60-61], 9 only virtual chromoendoscopy images[48-49,59,63-68], 1 only WLI and chromoendoscopy images[54], and 5 WLI, chromoendoscopy and NBI images[14,45-46,62,69]. The identified studies were largely published in the last 3 years.

Overall, current CNN systems work quite well in detecting the endoscopic/chromoendoscopic characteristics of EGC and other gastric lesions and could provide diagnostic support to experienced and non-expert endoscopists in future practice. AI-assisted CNN CAD systems can avoid subjectivity during the processing and diagnosis of endoscopic/chromoendoscopic images; moreover, in the screening of GC, they work as a "confirmer" or "corrector," providing a second opinion to reduce the diagnostic errors committed by endoscopists and suggesting optimal treatment. Current studies by Asian authors[54,59] confirm that CAD systems detect EGCs and estimate the depth of infiltration and extension, overcoming the problem of operator training and the subjectivity of diagnosis. Moreover, if the first studies report comparable results between experts and CAD systems, the most recent ones show that AI has reached a sensitivity even higher than that of experts, with similar specificity[46]. Over time, images used for CAD system training have improved and, at present, advanced training strategies and videos are being used.

Namikawa *et al*[58] first reported the usefulness of AI systems in GC detection, developing the "original convolutional neural network (O-CNN)," with a relatively low positive predictive value (PPV). The same authors developed an advanced AI-based diagnostic system, "advanced CNN (A-CNN)", by adding a new training dataset to the O-CNN and evaluated its applicability for the classification of GC and gastric ulcer. The

diagnostic performance of A-CNN was evaluated retrospectively using an independent validation dataset and compared to that of the O-CNN by estimating the overall accuracy of the classification. The sensitivity, specificity, and PPV rates of A-CNN for the classification of GC at the lesion level were 99.0%, 93.3%, and 92.5%, respectively, and 93.3%, 99.0%, and 99.1% for the classification of gastric ulcers. The overall accuracy of O-CNN and A-CNN in the classification of GC and gastric ulcer was 45.9% (GC 100%, gastric ulcer 0.8%) and 95.9% (GC 99.0%, gastric ulcer 93.3%), respectively, at the lesion level. The A-CNN system can effectively classify GC and gastric ulcer. Horiuki et al[36] explored the diagnostic capacity of the CNN system with ME-NBI to distinguish EGC from gastritis. CNN accuracy with ME-NBI images was 85.3% (220 of 258 images correctly diagnosed). Rates of sensitivity, specificity, PPV, and negative predictive value (NPV) were 95.4%, 71.0%, 82.3%, and 91.7%, respectively. In total, 7 of 151 EGC images were identified as gastritis, while 31 of the 107 gastritis images were recognized as EGC. The overall test speed was 51.83 images/s (0.02 s/image). CNN with ME-NBI can differentiate between EGC and gastritis with high sensitivity and NPV in a short period of time. Thus, the A-CNN system can complement current clinical practice of diagnosis with ME-NBI.

Nam et al[47] have developed and validated CNN-based AI models for lesion detection, differential diagnosis (AI-DDx), and depth of invasion (AI-ID; pT1a vs pT1b among EGC). AI-DDx is comparable to experts and outperforms novice and intermediate endoscopists in the differential diagnosis of gastric mucosal lesions. AI-ID performs better than endoscopic ultrasound to assess depth of invasion. Ling et al[48] have developed a system to identify in real time with precision with ME-NBI the state of differentiation and delineate the margins of the EGC, fundamental to determine a surgical strategy and achieve the curative resection. In the unprocessed videos of EGC, the system obtained a real-time diagnosis of EGC differentiation and its margins ME-NBI endoscopy. This system has achieved higher performance than experts and has been successfully tested in real EGC videos.

Zhu et al[54] represented a further step forward because they developed an algorithm capable of differentiating lesions with Sm2 invasion depth from m/Sm1. AI has presented 76% sensitivity and 96% specificity in identifying "Sm2 or deeper" cancers, resulting in significantly higher sensitivity and specificity than those achieved through visual inspection of endoscopists. The specificity of 96% could minimize the overdiagnosis of invasion, which would contribute to a reduction of unnecessary surgeries for m/Sm1 cancers.

Wu et al[52], in a prospective multicenter randomized controlled trial, developed a CNN system to monitor blind spots during esophagogastroduodenoscopy, updating the previous system (ENDOANGEL), verifying efficacy in improving endoscopy quality, and pretesting performance in detecting EGC.

Ultimately, AI is even superior to endoscopists experienced in identifying and classifying ECC, eliminates interobserver variability, and can train inexperienced endoscopists. Yet, it must optimize the ability to recognize all lesions (PPV) and not interpret the inflammatory or benign aspects of the mucosa as neoplastic (NPV). Over time, CAD systems have improved image selection strategies with strict criteria, using high-quality data and videos, and eliminating overlearning and misdiagnosis. Videos improve the performance of AI[55] because they represent real-life scenarios, and compared to static images improve PPV and NPV. Regarding the selection of images, gastritis, that is, the presence of inflammation, reduces the performance of AI[14] and endoscopists[71]. The small (diameter ≤ 5 mm) and depressed EGCs, difficult to distinguish from gastritis even for experienced endoscopists, influence the rate of false negatives;, and gastritis with redness, atrophy and intestinal metaplasia affects the rate of false positives . In dedicated studies, CAD systems detect *Helicobacter pylori (H.pylori)* infection (sensitivity 89%, specificity 87% and diagnostic time 194 s)[71-72], but, regarding the diagnosis of EGC with AI sistems, we propose to evaluate the gastric mucosa after the eradication of *H. pylori* to reduce the intensity of redness of gastritis.

Integrating in appropriate algorithms, through the intersection of engineering and medical expertise, high-quality image sets, poor images, and images from regular sites,

will increase clinical effectiveness. Moreover, the products obtained through collaboration among centers specialized in the diagnosis and treatment of gastric lesions are reproducible and the limitation in applying AI to the diagnosis of EGC is the acquisition of new technologies, which requires investment. Finally, prospective multicenter trials are needed.

CONCLUSION

The application of AI to the clinical practice of the upper digestive tract increases the rate of EGC compared to all GCs, exceeding the subjectivity of the diagnosis and reducing the chance of missing EGCs. AI recognizes those lesions that not even the most experienced endoscopists can detect, as if "illuminating" the images with its third artificial eye. Of course, AI increases the accuracy of endoscopic diagnosis of EGC, especially when combined with the experience of endoscopists. However, since its introduction in this field is very recent, the results in clinical practice must be further validated, considering all possible aspects, both technical and technological concerning endoscopy, and organizational ones.

74661_Auto_Edited.docx

Comparison between the original and newly developed systems", Endoscopy, 2020
Crossref

| 9 | pubmed.ncbi.nlm.nih.gov
Internet | 13 words — 1% |
| 10 | link.springer.com
Internet | 12 words — < 1% |