Artificial Intelligence in Upper Gastrointestinal Diagnosis

https://www.helicojournal.org

Corresponding author
Khek Yu Ho, MD
Division of Gastroenterology and
Hepatology, Department of Medicine,
National University Hospital,
1E Kent Ridge Road 119228, Singapore
E-mail: mdchoky@nus.edu.sg

Sabrina Xin Zi Quek[1] and Khek Yu Ho[1,2]

[1]Division of Gastroenterology and Hepatology, Department of Medicine, National University of Singapore, Singapore
[2]Division of Gastroenterology and Hepatology, Department of Medicine, National University Hospital, Singapore

Artificial intelligence (AI) has revolutionized upper gastrointestinal (GI) endoscopy by enhancing the detection, characterization, and management of GI diseases. In this review, we explore the transformative role of AI technologies, including machine learning and deep learning, in improving diagnostic accuracy and streamlining clinical workflows. AI systems such as convolutional neural networks have shown remarkable potential for identifying subtle lesions, assessing tumor margins, and reducing interobserver variability. By providing real-time decision-making support, AI minimizes unnecessary biopsies and improves patient outcomes. We also explore the applications of AI in detecting precancerous conditions such as Barrett's esophagus, atrophic gastritis, and gastric intestinal metaplasia, as well as its role in guiding therapy for early gastric cancer. Non-image-based AI tools such as Raman spectroscopy complement traditional imaging by offering molecular-level insights for real-time tissue characterization. Despite its promise, the adoption of AI in endoscopy faces challenges, including the need for robust validation, user-centric design, and targeted training for endoscopists. Concerns regarding overreliance and deskilling underscore the importance of balancing AI integration with the preservation of clinical expertise. Lastly, we examine the future of AI in upper GI diagnosis and how image-based and non-image-based AI technologies can be integrated to enable comprehensive diagnosis and personalized therapeutic planning. By addressing current limitations and fostering collaboration between clinicians and technologists, AI has the potential to redefine the standards of care for upper GI diagnosis and treatment.

Keywords Artificial intelligence; Gastroscopy; Endoscopy.

INTRODUCTION

Endoscopy plays an essential role in the diagnosis and management of gastrointestinal (GI) diseases. While the first successful endoscope was developed in 1805, it was not until the invention of the fiberoptic endoscope in 1957[1] that endoscopy evolved beyond rudimentary rigid instruments, paving the way for broader advancements and widespread applications. The integration of artificial intelligence (AI) has marked the latest transformative shift in endoscopy, enhancing pathology, diagnosis, and clinical decision-making. These advancements are particularly crucial in the upper GI tract, where early detection of esophageal[2] and gastric[3,4] cancers, as well as their premalignant lesions, can significantly improve patient outcomes.

The applications of AI in upper GI endoscopy (esophagogastroduodenoscopy [EGD]) can be broadly classified into image-based and non-image-based AI, each addressing unique clinical challenges. Image-based AI encompasses tools such as computer-aided detection (CADe), computer-aided diagnosis (CADx), and computer-aided quality improvement (CADq), which improve real-time lesion identification, characterization, and procedural standardization. In contrast, non-image-based AI leverages clinical, molecular, and spectroscopic data to provide insights beyond the visual field of endoscopy, with Raman spectroscopy emerging as a promising technology for

molecular-level diagnostics.

In this review, we explore the current state and future directions of AI in upper GI endoscopy, with a focus on the detection and diagnosis of malignant and pre- malignant lesions. However, alongside these advancements, we address critical challenges, including the potential pitfalls of overreliance on AI, ethical and legal considerations, and technical limitations. Furthermore, we discuss the concerns raised by endoscopists, such as the impact of AI on clinical judgment, trust in automated systems, and barriers to widespread adoption, including costs, training requirements, and the need for robust validation in diverse clinical settings.

TERMINOLOGY

Before discussing the use of AI in upper GI diagnosis, one must first understand the common terminology used in this field. AI is a broad term that refers to machines or computers that can perform tasks requiring human intelligence. Machine learning is one of its key subfields[5,6] that involves training algorithms to analyze data and make predictions. Traditional machine learning requires feature extraction, which means that humans must first identify and define important characteristics in the data before an algorithm can process them.

Deep learning, a more advanced subset of machine learning, eliminates the need for manual feature extraction using artificial neural network-multilayered structures designed to automatically learn patterns from raw data. In this architecture, earlier layers recognize simple features (e.g., edges in an image), whereas deeper layers identify more complex patterns (e.g., lesion morphology or vascular patterns in endoscopic images). This capability makes deep learning particularly powerful for tasks such as image analysis, which is central to AI applications in upper GI endoscopy. In upper GI endoscopy, the most commonly used deep learning model is a convolutional neural network (CNN). CNNs are specifically designed for image analysis; they use convolutional layers to extract spatial features from images, followed by pooling layers to reduce dimensionality, and fully connected layers to classify the data. This design enables CNNs to detect, classify, and assess abnormalities in real time with high accuracy.

IMAGE-BASED AI IN UPPER GI ENDOSCOPY

Although white-light endoscopy (WLE) with biopsy is the gold standard modality for the diagnosis of upper GI pathologies,[7,8] there are many pitfalls in relying on this modality alone. Even with extensive training, endoscopists encounter challenges such as interobserver variability and the risk of missed lesions, especially in pre-neoplastic and early malignant cases, where subtle changes can be easily overlooked, yet have a profound impact on a patient's prognosis. Image-based AI systems address these limitations by providing consistent, fatigue-free,[9] and highly accurate analyses. Kamran et al.[10] conducted a root cause analysis to establish possible explanations for post-endoscopic upper GI cancer and identified inadequate endoscopy quality, inadequate assessment of premalignant or focal lesions, and poor decision-making around surveillance or follow-up plans as common explanations. AI could potentially help address all of these, as discussed below and summarized in Table 1.

CADq

A key role of AI in upper GI endoscopy is to reduce interobserver variability and ensure that all procedures achieve high-quality complete mucosal inspection, a concept known as CADq.[11] Wu et al.[12] demonstrated the effectiveness of the ENDOANGEL system (previously known as WISENSE) in reducing blind-spot rates during endoscopy. Their study showed that the use of ENDOANGEL significantly decreased blind spot rates from 22.46% to 5.86% ($p<0.001$) through real-time prompting, thereby improving mucosal visualization. This finding was further supported by Chen et al.[13] who conducted a single-blind, three-parallel-group, randomized, single-center trial involving 437 patients. This study compared the performances of unsedated ultrathin transoral endoscopy, unsedated conventional EGD, and sedated conventional EGD with or without AI assistance. The results revealed that the AI-assisted subgroups consistently had lower blind spot rates than the control subgroups across all three groups ($p<0.001$).

An AI-based system for real-time photo documentation during EGD, termed the Automated Photodocumentation Task (APT), was developed using a training and testing dataset of 102798 endoscopic images from 3309 EGD examinations conducted at Seoul National University Hospital. The APT utilizes a Swin Transformer (Shifted Window Transformer), which is a hierarchical vision transformer architecture designed for computer vision tasks. Unlike traditional CNNs, the Swin Transformer uses a self-attention mechanism to understand the relationships between different parts of an image, even when they are far apart, making it well suited for complex tasks such as the classification, detection, and segmentation of endoscopic images, where subtle details and spatial relationships are critical. In this study,[14] virtual endoscopy was performed by seven endoscopists and an APT with the goal of capturing 11 anatomical landmarks from endoscopic videos.

Table 1. Summary of AI applications in upper GI diagnosis, with a focus on malignant and premalignant lesions

AI type	Subcategory/tool	Primary uses/explanation	Study examples
Image-based AI	CADq	Quality assessment of endoscopy: enhances procedural quality by evaluating blind spots, gastric area coverage, and photo documentation using deep learning.	Endoangel: (12, 13) Automated Photodocumentation Task: (14)
	CADe/x–esophagus	Identify dysplastic areas, assist in lesion classification, and reduce miss rates.	Esophageal SCC: (24-26) Esophageal adenocarcinoma and Barrett's esophagus: (27-35)
	CADe/x–stomach		GC: (37-42) AG: (44) GIM: (45)
Non-image-based AI	Raman spectroscopy+AI	Real-time molecular characterization of gastrointestinal tissue during endoscopy by analyzing biochemical signatures to distinguish between normal, dysplastic, and cancerous tissue. Its integration with AI enhances diagnostic accuracy and supports the feasibility of in vivo use for real-time decision-making.	Ex vivo: (52-55) In vivo: (56-59)

AI, artificial intelligence; GI, gastrointestinal; CADq, computer-aided quality improvement; CADe, computer-aided detection; CADx, computer-aided diagnosis; SCC, squamous cell carcinoma; GC, gastric cancer; AG, atrophic gastritis; GIM, gastric intestinal metaplasia.

The primary endpoints were the completeness of landmark capture and image quality. APT achieved an average accuracy of 98.16% for capturing landmarks, demonstrating completeness similar to that of endoscopists (87.72% vs. 85.75%, p=0.258). However, the combined use of endoscopists and APT resulted in significantly higher completeness (91.89% vs. 85.75%, p<0.001). Additionally, APT-captured images had higher mean opinion scores than those captured by endoscopists (3.88 vs. 3.41, p<0.001), indicating a superior image quality. However, further prospective, real-time studies are required to validate these findings.

With its ability to alert users of blind spots, identify anatomical landmarks, and obtain standardized protocol views[15,16] in accordance with endoscopic guidelines, AI can help trainees develop the skills necessary for thorough and accurate gastroscopy.[17] Furthermore, built-in documentation tools such as automated photo capturing reduce the need for repeated freezing and capturing, thereby reducing the procedural burden.[18] This not only improves overall efficiency and consistency, but also streamlines clinical workflows in daily practice.

CADe/x

Even with optimal mucosal exposure, premalignant and early malignant lesions in the esophagus and stomach can be challenging to detect owing to their subtle and often inconspicuous appearances. Thus, CADe systems play a crucial role in augmenting an endoscopist's ability to identify these subtle changes. Once a lesion is detected, CADx systems can further assist in characterizing the degree of neoplasia, providing valuable insights into the nature of the lesion and guiding appropriate clinical management. Together, CADe and CADx enhance diagnostic accuracy and improve patient outcomes by ensuring the early detection and precise characterization of lesions.

Early detection is a key factor in the prognosis of both esophageal cancer and gastric cancer (GC), in which missing an early lesion that may be surgically or even endoscopically resectable with curative intent would result in a much poorer prognosis, with 5-year survival rates of <5%–28%[19,20] in the case of advanced (stage III-IV) esophageal cancer and 7%–35% in the case of advanced GC.

Esophagus

The rate of missed esophageal cancer ranges from 6.4% to 8.0% in previous reports,[21-23] in which missed esophageal cancer is defined as esophageal cancer diagnosed at 6–36 months after non-diagnostic upper endoscopy. In a multicenter, double-blind, randomized control trial, Yuan et al.[24] evaluated an AI system designed to assist in detecting superficial esophageal squamous cell carcinoma (ESCC) and precancerous lesions using WLE and non-magnified narrow-band imaging (NBI). Their results indicated lower miss rates with AI assistance (1.7% per lesion, 1.9% per patient) than with routine endoscopy (6.7% per lesion, 5.1% per patient), suggesting potential benefits. Nevertheless, further assessment of effectiveness and cost-benefit in real-world settings is needed. Meng et al.[25] developed a deep learning-based CAD system using the YOLO v5 algorithm to detect superficial ESCC with high diagnostic performance (area under the curve [AUC], 0.982; accuracy, 92.9%). The system significantly improved the accuracy of non-

expert endoscopists (from 78.3% to 88.2%), demonstrating its potential to enhance detection, particularly for less experienced practitioners. However, challenges remain in identifying certain lesion types, such as the Paris classification 0-IIb.

Interestingly, a prospective study by Nakao et al.[26] did not demonstrate a significant benefit in ESCC detection using an AI diagnostic support system. Possible reasons for this negative result included the user interface, as the AI alert was on a separate monitor, and the endoscopist had to shift focus between the two monitors with the potential for missing lesions. Another reason for this was lesion characteristics. The AI system may have been less effective in detecting certain lesion types, such as flat or small lesions, which are inherently more challenging to identify.

Barrett's esophagus[27] is a well-established precursor of esophageal adenocarcinoma. Patients with Barrett's esophagus often undergo EGD surveillance because of their increased risk of esophageal adenocarcinoma. Early detection[28] of dysplasia and intramucosal carcinoma in Barrett's esophagus compared to neoplasia with submucosal involvement would indicate a difference between minimally invasive endoscopic curative options and esophagectomy, which is associated with significant morbidity and mortality risks.

A systematic review of 14 studies by Patel et al.[29] examined the use of AI for the diagnosis of Barrett's esophagus and related neoplasias. Five out of these studies[30-34] with sample sizes ranging from 20 to 1229 patients evaluated CADe systems for Barrett's esophagus, demonstrating high sensitivity (84%–100%) and variable specificity (64%–90.7%), outperforming non-expert endoscopists in diagnosing Barrett's esophagus and related neoplasias. The BONS-AI consortium[35] involving 15 international centers developed and validated a CADx system for Barrett's esophagus and related neoplasias. The system, trained on 3596 NBI images from 525 patients, achieved a standalone sensitivity and specificity of 100%/98% for images and 93%/96% for videos. With CADx assistance, the diagnostic performance of general endoscopists improved significantly, matching that of experts in Barrett's esophagus while increasing their confidence in lesion characterization.

Beyond the detection and characterization of premalignant and early malignant lesions, Römmele et al.[36] demonstrated that an AI algorithm for eosinophilic esophagitis could effectively detect this condition with excellent performance. The algorithm achieved a sensitivity, specificity, and accuracy of 0.93, which further improved to 0.96, 0.94, and 0.95, respectively, when the Eosinophilic Esophagitis Endoscopic Reference Score (EREFS) criteria were incorporated. The model outperformed less-experienced endoscopists and showed results comparable to those of experts, highlighting its potential for enhancing diagnostic precision and reducing variability in clinical practice.

Stomach

With respect to GC, several studies[37] have demonstrated the benefits of AI in detecting and diagnosing early GC (EGC) using both WLE and image-enhanced endoscopy.[37] In addition to identifying and classifying lesions as neoplastic, AI has been shown to assess the depth of GC invasion and delineate the margins of neoplastic lesions. These capabilities are critical for guiding therapeutic decisions and ensuring adequate resection margins.

The development of AI-based diagnostic support tools for EGC detection is driven by the challenge of identifying subtle mucosal changes that are often missed because of their inconspicuous nature. One such tool, Tango,[38] has shown promising results. In comparative studies, Tango achieved superior sensitivity over specialists (84.7% vs. 65.8%; difference, 18.9%; 95% confidence interval [CI], 12.3%–25.3%) and demonstrated non-inferior accuracy (70.8% vs. 67.4%). Additionally, Tango outperformed non-specialists in both sensitivity (84.7% vs. 51.0%) and accuracy (70.8% vs. 58.4%), highlighting its potential to enhance diagnostic performance across varying levels of expertise.

In a single-center randomized controlled trial using the EN-DOANGEL-lesion detection system, same-day tandem upper GI endoscopy was performed in which participants first underwent either AI-assisted or routine WLE. Wu et al.[39] demonstrated a significantly reduced rate of missed gastric neoplasm in the group that underwent AI-assisted endoscopy first (6.1%, 95% CI: 1.6–17.9 [3/49] vs. 27.3%, 95% CI: 15.5–43.0 [12/44]; relative risk, 0.224, 95% CI: 0.068–0.744; p=0.015). The same group[40] also evaluated ENDOANGEL with magnified NBI on endoscopists in a multicenter prospective trial. A total of 46 endoscopists were compared with ENDOANGEL. The sensitivity rates of the system for detecting neoplasms and diagnosing EGC were 87.81% and 100%, respectively, which were significantly higher than those of endoscopists (83.51%, 95% CI: 81.23–85.79 vs. 87.13%, 95% CI: 83.75–90.51). The accuracy rates of the system for predicting EGC invasion depth and differentiation status were 78.57% and 71.43%, respectively, which were slightly higher than those of endoscopists (63.75%, 95% CI: 61.12–66.39 vs. 64.41%, 95% CI: 60.65–68.16).

Nam et al.[41] developed a CNN-based AI using three models: lesion detection, differential diagnosis, and invasion depth (pT1a vs. pT1b in EGC). Their AI-lesion detection model performed similarly to that of expert endoscopists with >5 years of experience. The diagnostic performance of their AI-differential diagnosis model (area under the receiver operating

characteristic curve [AUROC]: 0.86, 95% CI: 0.84–0.89) was significantly better than that of novice endoscopists with <1 year of experience (AUROC: 0.78, 95% CI: 0.76–0.80, p<0.001) and intermediately experienced endoscopists with 2–3 years of experience (AUROC: 0.84, 95% CI: 0.83–0.86, p=0.035), but was comparable to that of expert endoscopists (AUROC: 0.86, 95% CI: 0.84–0.88, p=0.942) in the internal validation set, with similar trends in the external validation set. Among patients with EGC, the AI-ID model showed fair performances in both the internal (AUROC: 0.78) and external validation sets (AUROC: 0.73), which were significantly better than the results of endoscopic ultrasound performed by experts (AUROC: 0.62 in the internal validation set, AUROC: 0.56 in the external validation set; both p<0.001).

Complete endoscopic resection with adequate margins is critical for the treatment of EGC, as it prevents unnecessary repeat endoscopies, which are often more challenging because of scarring or even require surgical intervention. Although methods such as magnified NBI and indigo carmine endoscopy have been used to delineate tumor margins, more precise tools are required to ensure accurate lesion sizing and enhance the planning and success of endoscopic submucosal dissection. In a review by Lei et al.,[42] six studies evaluating AI for boundary identification demonstrated accuracy rates ranging from 82.7% to 96.3%. However, further prospective studies are required to validate these promising results and to establish the role of AI in optimizing the outcomes of endoscopic submucosal dissection.

Beyond the early detection of GC, identifying patients at risk of developing GC is crucial for determining appropriate surveillance strategies. While risk factors such as family history play a significant role, precancerous changes such as atrophic gastritis (AG) and gastric intestinal metaplasia (GIM) also influence the timing and frequency of surveillance endoscopies. The progression of GC through Correa's cascade[43] is well established, and AI tools have been developed to detect these precancerous conditions, enabling earlier intervention and personalized patient management. A systematic review and meta-analysis[44] of 8 studies evaluating AI for AG detection demonstrated a sensitivity of 94% (95% CI: 0.88–0.97) and a specificity of 96% (95% CI: 0.88–0.98), with an area under the summary receiver operating characteristic (SROC) curve of 0.98 (95% CI: 0.96–0.99). These results indicate that AI significantly outperformed endoscopists in the diagnosis of AG. Similarly, another meta-analysis[45] of 12 studies focusing on AI for GIM detection reported a pooled sensitivity of 94% (95% CI: 0.92–0.96) and a specificity of 93% (95% CI: 0.89–0.95), with an SROC curve of 0.97. AI demonstrated superior diagnostic performance compared with endoscopists, with a sen-

sitivity of 95% versus 79% for human experts. These findings underscore the potential of AI in enhancing the early detection of precancerous gastric lesions.

However, both meta-analyses exhibited substantial heterogeneity, particularly in the definition of diagnostic criteria and grading of AG and GIM severity. Additionally, variations in the endoscopic equipment, imaging techniques, datasets, and AI algorithms further limit the generalizability of their findings. Notably, the majority of the studies included were retrospective in design—5 out of 8 for AG and 9 out of 12 for GIM—and all but one study was conducted in non-Asian populations, which may affect the applicability across different geographic regions. In real-world clinical settings, AI performance has not consistently replicated the high sensitivity and specificity reported in controlled environments. This highlights the need for well-designed prospective multicenter studies that apply standardized diagnostic criteria and clinically relevant endpoints to accurately assess the utility of AI in routine practice. To facilitate broader applicability and external validation, algorithm code sharing should also be encouraged to ensure more robust and generalizable data across diverse populations.

The role of *Helicobacter pylori* in GC has been well studied; since 1994, it has been labeled as a human carcinogen by the World Health Organization's International Agency for Research on Cancer.[46,47] While the gold standard[48] for the diagnosis of *H. pylori* infection is histopathological examination, this requires biopsies with inherent risks of complications, such as bleeding. AI-based methods for detecting *H. pylori* infections using endoscopic images have shown excellent diagnostic performance. A meta-analysis[49] reported pooled sensitivity and specificity of 0.90 (95% CI: 0.80–0.95) and 0.92 (95% CI: 0.88–0.95), with an AUC of 0.97 (95% CI: 0.96–0.99). Individual studies, such as Lin et al.,[50] achieved a sensitivity of 1.00 and specificity of 0.82, whereas others, such as Yacob et al.,[51] reported both metrics at 0.98. These findings suggest that AI can reduce the need for invasive biopsies and improve diagnostic confidence. However, further validation in diverse populations is required.

NON-IMAGE-BASED AI IN UPPER GI ENDOSCOPY AND DIAGNOSIS

Non-image-based AI technologies for upper endoscopy are emerging as powerful tools for enhancing diagnostic accuracy and providing real-time tissue characterization. One such technique is Raman spectroscopy, which uses laser light to analyze the molecular composition of tissues (Fig. 1). When integrated with AI, Raman spectroscopy can differentiate normal, precancerous, and cancerous tissues by detecting subtle
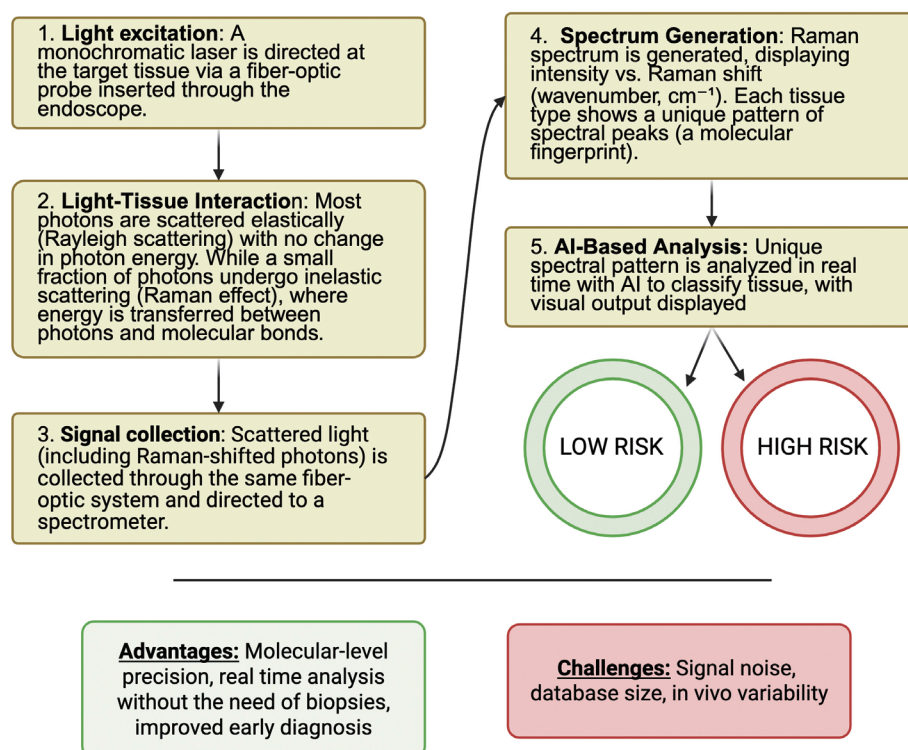
Fig. 1. Flowchart explaining Raman spectroscopy and AI for real-time tissue characterization in upper GI endoscopy. AI, artificial intelligence; GI, gastrointestinal.

biochemical changes that are not visible to the naked eye. Although numerous studies have demonstrated the ability of Raman spectroscopy to differentiate normal tissue, dysplasia, and cancer with high sensitivity and specificity, the majority of these studies have been conducted ex vivo.[52-55]

Several studies have evaluated the feasibility of Raman spectroscopy for clinical application in vivo. Significant advancements have been made since Shim et al.[56] first demonstrated the use of Raman spectroscopy during endoscopy in 2000. For example, Bergholt et al.[57] showed that real-time image-guided Raman endoscopy combined with AI diagnostic algorithms could achieve a diagnostic sensitivity and specificity of 94.6% for the in vivo diagnosis of gastric neoplasia. In a feasibility proof-of-concept study comparing Raman spectroscopy-based AI (SPECTRA IMDx™)[58] with high-definition WLE for classifying gastric lesions as low or high risk for neoplasia, the Raman spectroscopy system achieved a sensitivity, specificity, and accuracy of 100%, 80%, and 89% by patient, and 100%, 80%, and 92% by lesion, respectively—performance comparable to that of expert endoscopists. Similarly, Noh et al.[59] identified the biomolecular differences between benign gastric tissues and gastric adenocarcinoma and evaluated the diagnostic potential of Raman spectroscopy combined with machine learning. Their model achieved diagnostic accuracy, sensitivity, specificity, and AUC values of 0.905, 0.942, 0.787,

and 0.957, respectively.

Despite the promising clinical trial results, practical limitations remain. Real-time Raman implementation is challenged by the need for miniaturized fiber-optic probes, rapid spectral acquisition, and robust AI algorithms that can handle the variability in in vivo tissue spectra. Moreover, current studies often involve small patient cohorts and are predominantly conducted in single-center settings, limiting their generalizability.

To enable real-world adoption, further modifications to both software and hardware are required to improve spectral data collection. Additionally, future studies should prioritize multicenter trials with diverse patient populations, seamless integration into existing endoscopic platforms with training provided for end users, and standardized reporting of diagnostic thresholds. Despite these challenges, Raman spectroscopy is a promising approach that complements traditional endoscopic imaging by providing molecular-level insights. This enables real-time detection of malignancies and has the potential to reduce the need for unnecessary biopsies.

POTENTIAL PITFALLS

One of the major concerns regarding the use of AI in endoscopy is the risk of overreliance and subsequent deskilling by endoscopists. This concern appears to be supported by ret-

rospective data from Budzyń et al.,[60] which examined the impact of AI on adenoma detection rates (ADR) during colonoscopy. In this nested study of the ACCEPT trial, participants undergoing screening colonoscopy were randomized into AI-assisted or standard colonoscopy groups. The ADR for standard, non-AI-assisted colonoscopy significantly decreased from 28.4% (226/795) before AI exposure to 22.4% (145/648) after, representing a 6% absolute reduction (95% Cl: -10.5% to -1.6%). This decline raises concerns that endoscopists may become complacent when AI is not using AI. However, further robust prospective studies are required to confirm these findings. Additionally, as evidence supporting the benefits of AI continues to grow, and its use becomes more widespread, the following question arises: "Should deskilling remain a primary concern?" For instance, with the advent of advanced GPS technology, the ability to read paper maps is likely to diminish among younger generations. Similarly, rather than focusing on the potential for deskilling, it may be more productive to invest resources in training young endoscopists to effectively utilize AI, while emphasizing the importance of maintaining vigilance and ensuring that they maintain the capability to process and interpret visual information.

Another significant concern is the issue of liability[8,61,62] when AI-assisted decisions lead to adverse outcomes. Clear guidance from government agencies and regulatory bodies is essential to define the appropriate applications of AI in clinical practice. It is crucial to emphasize that the role of AI should not be to replace clinicians but to augment their diagnostic capabilities, ensuring that the final decision-making authority remains with trained healthcare professionals.

Third, is a potential impact on health equity, both between wealthy and developing nations and within countries where disparities exist between urban and rural areas.[63,64] Adopting AI requires significant resources to acquire the necessary equipment and train endoscopists, which could further widen the gap in endoscopic practice between the affluent and underserved regions. As AI is becoming more widely adopted, efforts must be made to ensure its affordability and formally evaluate its cost-effectiveness in day-to-day practice.

Finally, beyond ensuring the security of sensitive patient data, safeguarding the integrity of machine learning algorithms is important. Malicious tampering by ill-intentioned individuals can compromise the reliability of AI systems, potentially leading to severe clinical consequences. Robust measures must be implemented to protect both the data and algorithms from unauthorized access or manipulation.

FUTURE ADVANCEMENTS

While most studies have largely shown benefits in the use of AI in clinical practice, Nakao et al.,[26] which found no significant improvement in detection rates when endoscopists used the AI system compared with standard endoscopic procedures, provided valuable insights into areas for improvement.

First, developing AI systems with more extensive and diverse datasets could improve diagnostic accuracy across various clinical scenarios, particularly in detecting premalignant or early malignant lesions, where mucosal changes can be very subtle. To ensure robustness and generalizability, AI systems must undergo rigorous testing across diverse populations and ethnicities to validate their external validity before being approved for commercial use.

Secondly, AI systems must adopt a user-centric design. They should be integrated seamlessly into clinical workflows, providing intuitive and non-disruptive alerts to enhance the endoscopists' ability to effectively utilize AI. An important consideration is the risk of excessive false positive detections, which can contribute to alarm fatigue and potentially diminish the clinical utility of CADe systems. Evidence from colorectal polyp detection studies[65,66] has shown that frequent false alarms may desensitize endoscopists, slow procedures, and reduce overall diagnostic confidence. Translating these concerns to upper GI endoscopy, careful tuning of sensitivity-specificity thresholds, and smarter alert prioritization will be critical for maintaining trust and optimizing performance in real-world applications.

Third, targeted training of both non-expert and expert endoscopists on how to interpret and act upon AI-generated alerts could further improve detection rates. With the growing evidence supporting the benefits of AI, its use is likely to become increasingly prevalent in daily practice. Therefore, establishing proper guidelines for incorporating AI into endoscopic training and teaching novice endoscopists how to use AI effectively is essential.

Lastly, looking further into the future, we may expect the development of a comprehensive "all-in-one" system. Such a system could seamlessly combine CADq, CADe, and CADx for both benign and malignant conditions, assess the depth of malignant invasion, and assist with therapeutic planning. Furthermore, the integration of image-based and non-image-based AI promises to revolutionize endoscopy by combining macroscopic visualization with microscopic precision. Imagine a scenario during screening or surveillance endoscopy in which an endoscopist identifies a subtle lesion with the assistance of AI and characterizes it as potentially malignant. The endoscopist then confirmed this diagnosis in real time using

AI-enhanced Raman spectroscopy while simultaneously employing AI to assess the depth of invasion and determine whether the lesion can be resected endoscopically. Thus, the system could assist in planning en bloc curative resections, eliminating the need for unnecessary biopsies. Such advancements would not only reduce risks to patients, but also save time, streamline workflows, and empower clinicians to make more informed treatment decisions, ultimately improving patient management and outcomes.

CONCLUSION

Endoscopic procedures, traditionally reliant solely on operator experience, are now being transformed using AI technologies that enhance diagnostic accuracy, efficiency, and decision-making. The growing role of AI in endoscopy is inevitable, and those resistant to change risks have fallen behind in this rapidly evolving field of research. As endoscopists, we must embrace an open mindset, and leverage AI to improve patient outcomes while remaining vigilant to ensure that our diagnostic skills do not deteriorate through overreliance. Future research should focus on the synergistic potential of combining image-based and non-image-based AI, unlocking new possibilities for comprehensive and precise upper GI diagnosis.

Authors' Contribution
Conceptualization: Khek Yu Ho, Sabrina Xin Zi Quek. Writing—original draft: Sabrina Xin Zi Quek. Writing—review & editing: Khek Yu Ho, Sabrina Xin Zi Quek. Approval of final manuscript: Khek Yu Ho, Sabrina Xin Zi Quek.

ORCID iDs
Sabrina Xin Zi Quek   https://orcid.org/0000-0001-5377-1528
Khek Yu Ho   https://orcid.org/0000-0002-2932-1962

REFERENCES

1. Hunt RH. A brief history of endoscopy. Gastroenterology 2001;121:738-739.
2. Tustumi F, Kimura CM, Takeda FR, et al. Prognostic factors and survival analysis in esophageal carcinoma. Arq Bras Cir Dig 2016;29:138-141.
3. National Cancer Institute. Stomach cancer survival rates and prognosis [accessed on March 2, 2025]. Available from: https://www.cancer.gov/types/stomach/survival.
4. Hu HM, Tsai HJ, Ku HY, et al. Survival outcomes of management in metastatic gastric adenocarcinoma patients. Sci Rep 2021;11:23142.
5. Chahal D, Byrne MF. A primer on artificial intelligence and its application to endoscopy. Gastrointest Endosc 2020;92:813-820.e4.
6. van der Sommen F, de Groof J, Struyvenberg M, et al. Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. Gut 2020;69:2035-2045.
7. Tang Y, Anandasabapathy S, Richards-Kortum R. Advances in optical gastrointestinal endoscopy: a technical review. Mol Oncol 2021;15:2580-2599.
8. Yu H, Singh R, Shin SH, Ho KY. Artificial intelligence in upper GI endoscopy - current status, challenges and future promise. J Gastroenterol Hepatol 2021;36:20-24.
9. Quek SXZ, Lee JWJ, Feng Z, et al. Comparing artificial intelligence to humans for endoscopic diagnosis of gastric neoplasia: an external validation study. J Gastroenterol Hepatol 2023;38:1587-1591.
10. Kamran U, King D, Abbasi A, et al. A root cause analysis system to establish the most plausible explanation for post-endoscopy upper gastrointestinal cancer. Endoscopy 2023;55:109-118.
11. Renna F, Martins M, Neto A, et al. Artificial intelligence for upper gastrointestinal endoscopy: a roadmap from technology development to clinical practice. Diagnostics (Basel) 2022;12:1278.
12. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019;68:2161-2169.
13. Chen D, Wu L, Li Y, et al. Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: a prospective, single-blind, 3-parallel-group, randomized, single-center trial. Gastrointest Endosc 2020;91:332-339.e3.
14. Ahn BY, Lee J, Seol J, Kim JY, Chung H. Evaluation of an artificial intelligence-based system for real-time high-quality photodocumentation during esophagogastroduodenoscopy. Sci Rep 2025;15:4693.
15. Yao K. The endoscopic diagnosis of early gastric cancer. Ann Gastroenterol 2013;26:11-22.
16. Rey JF, Lambert R; ESGE Quality Assurance Committee. ESGE recommendations for quality control in gastrointestinal endoscopy: guidelines for image documentation in upper and lower GI endoscopy. Endoscopy 2001;33:901-903.
17. An P, Wang Z. Application value of an artificial intelligence-based diagnosis and recognition system in gastroscopy training for graduate students in gastroenterology: a preliminary study. Wien Med Wochenschr 2024;174:173-180.
18. Arif AA, Jiang SX, Byrne MF. Artificial intelligence in endoscopy: overview, applications, and future directions. Saudi J Gastroenterol 2023;29:269-277.
19. GBD 2017 Oesophageal Cancer Collaborators. The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet Gastroenterol Hepatol 2020;5:582-597.
20. The American Cancer Society medical and editorial content team. Survival rates for esophageal cancer [accessed on August 13, 2025]. Available from: https://www.cancer.org/cancer/types/esophagus-cancer/detection-diagnosis-staging/survival-rates.html.
21. Rodríguez de Santiago E, Hernanz N, Marcos-Prieto HM, et al. Rate of missed oesophageal cancer at routine endoscopy and survival outcomes: a multicentric cohort study. United European Gastroenterol J 2019;7:189-198.
22. Chadwick G, Groene O, Hoare J, et al. A population-based, retrospective, cohort study of esophageal cancer missed at endoscopy. Endoscopy 2014;46:553-560.
23. Straum S, Wollan K, Rekstad LC, Fossmark R. Esophageal cancers missed at upper endoscopy in Central Norway 2004 to 2021 – apopulation-based study. BMC Gastroenterol 2024;24:279.
24. Yuan XL, Liu W, Lin YX, et al. Effect of an artificial intelligence-assisted system on endoscopic diagnosis of superficial oesophageal squamous cell carcinoma and precancerous lesions: a multicentre, tandem, double-blind, randomised controlled trial. Lancet Gastroenterol Hepatol 2024;9:34-44.
25. Meng QQ, Gao Y, Lin H, et al. Application of an artificial intelligence system for endoscopic diagnosis of superficial esophageal squamous cell carcinoma. World J Gastroenterol 2022;28:5483-5493.
26. Nakao E, Yoshio T, Kato Y, et al. Randomized controlled trial of an artificial intelligence diagnostic system for the detection of esophageal squamous cell carcinoma in clinical practice. Endoscopy 2025;

57:210-217.

27. Gonzalez-Haba M, Waxman I. Red flag imaging in Barrett's esophagus: does it help to find the needle in the haystack? Best Pract Res Clin Gastroenterol 2015;29:545-560.

28. Wang VS, Hornick JL, Sepulveda JA, Mauer R, Poneros JM. Low prevalence of submucosal invasive carcinoma at esophagectomy for high-grade dysplasia or intramucosal adenocarcinoma in Barrett's esophagus: a 20-year experience. Gastrointest Endosc 2009;69:777-783.

29. Patel A, Arora GS, Roknsharifi M, Kaur P, Javed H. Artificial intelligence in the detection of Barrett's esophagus: a systematic review. Cureus 2023;15:e47755.

30. Fockens KN, Jukema JB, Boers T, et al. Towards a robust and compact deep learning system for primary detection of early Barrett's neoplasia: initial image-based results of training on a multi-center retrospectively collected data set. United European Gastroenterol J 2023;11: 324-336.

31. de Groof AJ, Struyvenberg MR, Fockens KN, et al. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). Gastrointest Endosc 2020;91:1242-1250.

32. 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;158:915-929.e4.

33. de Groof J, van der Sommen F, van der Putten J, et al. The Argos project: the development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. United European Gastroenterol J 2019;7:538-547.

34. Abdelrahim M, Saiko M, Maeda N, et al. Development and validation of artificial neural networks model for detection of Barrett's neoplasia: a multicenter pragmatic nonrandomized trial (with video). Gastrointest Endosc 2023;97:422-434.

35. Jukema JB, Kusters CHJ, Jong MR, et al. Computer-aided diagnosis improves characterization of Barrett's neoplasia by general endoscopists (with video). Gastrointest Endosc 2024;100:616-625.e8.

36. Römmele C, Mendel R, Barrett C, et al. An artificial intelligence algorithm is highly accurate for detecting endoscopic features of eosinophilic esophagitis. Sci Rep 2022;12:11115.

37. Klang E, Sourosh A, Nadkarni GN, Sharif K, Lahat A. Deep learning and gastric cancer: systematic review of AI-assisted endoscopy. Diagnostics (Basel) 2023;13:3613.

38. Ishioka M, Osawa H, Hirasawa T, et al. Performance of an artificial intelligence-based diagnostic support tool for early gastric cancers: retrospective study. Dig Endosc 2023;35:483-491.

39. Wu L, Shang R, Sharma P, et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomised controlled trial. Lancet Gastroenterol Hepatol 2021;6:700-708.

40. Wu L, Wang J, He X, et al. Deep learning system compared with expert endoscopists in predicting early gastric cancer and its invasion depth and differentiation status (with videos). Gastrointest Endosc 2022;95:92-104.e3.

41. Nam JY, Chung HJ, Choi KS, et al. Deep learning model for diagnosing gastric mucosal lesions using endoscopic images: development, validation, and method comparison. Gastrointest Endosc 2022;95: 258-268.e10.

42. Lei C, Sun W, Wang K, Weng R, Kan X, Li R. Artificial intelligence-assisted diagnosis of early gastric cancer: present practice and future prospects. Ann Med 2025;57:2461679.

43. Correa P, Piazuelo MB. The gastric precancerous cascade. J Dig Dis 2012;13:2-9.

44. Shi Y, Wei N, Wang K, Tao T, Yu F, Lv B. Diagnostic value of artificial intelligence-assisted endoscopy for chronic atrophic gastritis: a systematic review and meta-analysis. Front Med (Lausanne) 2023;10: 1134980.

45. Li N, Yang J, Li X, Shi Y, Wang K. Accuracy of artificial intelligence-assisted endoscopy in the diagnosis of gastric intestinal metaplasia: a systematic review and meta-analysis. PLoS One 2024;19:e0303421.

46. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Schistosomes, liver flukes and *Helicobacter pylori*. Lyon: International Agency for Research on Cancer, 1994.

47. Parsonnet J, Friedman GD, Vandersteen DP, et al. *Helicobacter pylori* infection and the risk of gastric carcinoma. N Engl J Med 1991;325: 1127-1131.

48. Miller JM, Binnicker MJ, Campbell S, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2018 update by the Infectious Diseases Society of America and the American Society for Microbiology. Clin Infect Dis 2018;67:e1-e94.

49. Parkash O, Lal A, Subash T, et al. Use of artificial intelligence for the detection of *Helicobacter pylori* infection from upper gastrointestinal endoscopy images: an updated systematic review and meta-analysis. Ann Gastroenterol 2024;37:665-673.

50. Lin CH, Hsu PI, Tseng CD, et al. Application of artificial intelligence in endoscopic image analysis for the diagnosis of a gastric cancer pathogen-*Helicobacter pylori* infection. Sci Rep 2023;13:13380.

51. Yacob YM, Alquran H, Mustafa WA, Alsalatie M, Sakim HAM, Lola MS. *H. pylori* related atrophic gastritis detection using enhanced convolution neural network (CNN) learner. Diagnostics (Basel) 2023;13: 336.

52. Yin F, Zhang X, Fan A, et al. A novel detection technology for early gastric cancer based on Raman spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2023;292:122422.

53. Li C, Liu S, Zhang Q, et al. Combining Raman spectroscopy and machine learning to assist early diagnosis of gastric cancer. Spectrochim Acta A Mol Biomol Spectrosc 2023;287(Pt 1):122049.

54. Mahadevan-Jansen A, Richards-Kortum RR. Raman spectroscopy for the detection of cancers and precancers. J Biomed Opt 1996;1:31-70.

55. Fang S, Xu P, Wu S, et al. Raman fiber-optic probe for rapid diagnosis of gastric and esophageal tumors with machine learning analysis or similarity assessments: a comparative study. Anal Bioanal Chem 2024;416:6759-6772.

56. Shim MG, Song LM, Marcon NE, Wilson BC. In vivo near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy. Photochem Photobiol 2000;72:146-150.

57. Bergholt MS, Zheng W, Lin K, et al. In vivo diagnosis of gastric cancer using Raman endoscopy and ant colony optimization techniques. Int J Cancer 2011;128:2673-2680.

58. Soong TK, Kim GW, Chia DKA, et al. Comparing Raman spectroscopy-based artificial intelligence to high-definition white light endoscopy for endoscopic diagnosis of gastric neoplasia: a feasibility proof-of-concept study. Diagnostics (Basel) 2024;14:2839.

59. Noh A, Quek SXZ, Zailani N, et al. Machine learning classification and biochemical characteristics in the real-time diagnosis of gastric adenocarcinoma using Raman spectroscopy. Sci Rep 2025;15:2469.

60. Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol 2025 Aug 12 [Epub]. https://doi.org/10.1016/S2468-1253(25)00133-5.

61. Christou CD, Tsoulfas G. Challenges involved in the application of artificial intelligence in gastroenterology: the race is on! World J Gastroenterol 2023;29:6168-6178.

62. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology- promises and pitfalls. Am J Gastroenterol 2019;114:422-428.

63. El-Sayed A, Salman S, Alrubaiy L. The adoption of artificial intelli-

gence assisted endoscopy in the Middle East: challenges and future potential. Transl Gastroenterol Hepatol 2023;8:42.

64. Anirvan P, Meher D, Singh SP. Artificial intelligence in gastrointestinal endoscopy in a resource-constrained setting: a reality check. Euroasian J Hepatogastroenterol 2020;10:92-97.

65. Zhang C, Yao L, Jiang R, et al. Assessment of the role of false-positive alerts in computer-aided polyp detection for assistance capabilities. J Gastroenterol Hepatol 2024;39:1623-1635.

66. Chung GE, Lee J, Lim SH, et al. A prospective comparison of two computer aided detection systems with different false positive rates in colonoscopy. NPJ Digit Med 2024;7:366.