Gastroenterology 2025;169:396–415

REVIEWS

Artificial Intelligence Applications in Image-Based Diagnosis of Early Esophageal and Gastric Neoplasms



Alanna Ebigbo[1] Helmut Messmann[1] Sung Hak Lee[2,3]

[1]Department of Gastroenterology, University Hospital Augsburg, Augsburg, Germany; [2]Department of Hospital Pathology, College of Medicine, The Catholic University of Korea, Seoul, South Korea; and [3]Seoul St. Mary's Hospital, Seoul, South Korea

Artificial intelligence (AI) holds the potential to transform the management of upper gastrointestinal (GI) conditions, such as Barrett's esophagus, esophageal squamous cell cancer, and early gastric cancer. Advancements in deep learning and convolutional neural networks offer improved diagnostic accuracy and reduced diagnostic variability across different clinical settings, particularly where human error or fatigue may impair diagnostic precision. Deep learning models have shown the potential to improve early cancer detection and lesion characterization, predict invasion depth, and delineate lesion margins with remarkable accuracy, all contributing to effective treatment planning. Several challenges, however, limit the broad application of AI in GI endoscopy, particularly in the upper GI tract. Subtle lesion morphology and restricted diversity in training datasets, which are often sourced from specialized centers, may constrain the generalizability of AI models in various clinical settings. Furthermore, the "black box" nature of some AI systems can impede explainability and clinician trust. To address these issues, efforts are underway to incorporate multimodal data, such as combining endoscopic and histopathologic imaging, to bolster model robustness and transparency. In the future, AI promises substantial advancements in automated real-time endoscopic guidance, personalized risk assessment, and optimized biopsy decision making. As it evolves, it would substantially impact not only early diagnosis and prognosis, but also the cost-effectiveness of managing upper GI diseases, ultimately leading to improved patient outcomes and more efficient health care delivery.

Keywords: Barrett's Esophagus; Squamous Cell Cancer; Gastric Cancer; Early Detection; Artificial Intelligence.

In Barrett's esophagus (BE), a precancerous condition, the natural squamous epithelium of the esophagus is replaced by specialized epithelium with intestinal differentiation. BE development is closely linked to gastroesophageal reflux disease, however, not all patients with gastroesophageal reflux disease develop BE.[1] The latter is a precursor condition for esophageal adenocarcinoma (EAC), whose early diagnosis is associated with better prognosis and cost-effectiveness.

Esophageal squamous cell cancer (ESCC) is equally challenging to diagnose at an early stage, owing to its extremely subtle appearance and lack of specific symptoms in early stages.[2] Conventional endoscopy for ESCC diagnosis is routinely enhanced using Lugo dye chromoendoscopy and advanced imaging techniques, such as narrow-band imaging (NBI), to highlight regions of interest (ROIs) and improve lateral-margin visibility. Population-based screening is usually recommended for high-risk geographical areas and individuals with risk factors, such as prior head and neck tumors. Despite innovations in advanced endoscopic imaging, most ESCC cases are still diagnosed at advanced stages, resulting in poor prognosis.[3]

Gastric cancer (GC) is the fifth most frequently diagnosed cancer and fifth major cause of cancer-related deaths globally.[4] GC incidence varies worldwide, with higher rates observed in Asia, Africa, South America, and Eastern Europe.[4] In Western countries, although the overall incidence has gradually decreased, the number of GC cases in

Abbreviations used in this paper: **AG, atrophic gastritis; AGC, advanced gastric cancer; AI, artificial intelligence; AUC, area under the curve; BE, Barrett's esophagus; BERN, Barrett's esophagus–related neoplasia; CAD, computer-aided detection; CNN, convolutional neural network; DL, deep learning; EAC, esophageal adenocarcinoma; EGC, early gastric cancer; ESCC, esophageal squamous cell cancer; GC, gastric cancer; GI, gastrointestinal; GIM, gastric intestinal metaplasia; HGD, high-grade dysplasia; LGD, low-grade dysplasia; LNM, lymph node metastasis; ME, magnifying endoscopy; M-IEE, magnified image–enhanced endoscopy; ML, machine learning; MSI, microsatellite instability; NBI, narrow-band imaging; NDBE, nondysplastic Barrett's esophagus; NDR, neoplasia detection rate; ROI, region of interest; TIL, tumor-infiltrating lymphocyte; WLI, white-light imaging; WSI, whole-slide image; XAI, explainable artificial intelligence.**

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2025.01.253

the proximal stomach is still on the rise. GC prognosis is generally poor, with a 5-year survival rate <40%; for advanced GC (AGC), it falls below 30%. However, for early stage GC, the 5-year survival rate exceeds 90%, highlighting the critical importance of early detection.[5,6]

Recent developments in artificial intelligence (AI), including deep learning (DL) using convolutional neural networks (CNN), have significantly improved the computing capacity and potential of AI. Integrating DL models into endoscopic procedures can potentially enhance diagnostic accuracy, ensuring that more patients receive appropriate and timely treatment.[7] In addition, AI systems may have the capacity to reduce variability among human endoscopists, providing a more consistent standard of care across clinical settings. DL models, which have been trained on extensive datasets, could provide robust algorithms capable of generalization across patient populations.[8,9] However, most data generated for the application of AI during endoscopy of the upper gastrointestinal (GI) tract are largely preclinical and stem from experimental settings. Although the huge potential of AI has become apparent, there is no evidence that AI has impacted standard clinical practice.

In this review, we outline recent representative studies and possible clinical indications of image-based AI for premalignant and malignant lesions of the esophagus and stomach, and discuss current abilities, challenges, and future directions.

Artificial Intelligence in Barrett's Esophagus Neoplasia Detection

AI support in BE mainly involves identifying subtle morphologic changes associated with early EAC that are often missed by endoscopists.[10] Early EAC diagnosis and better prognosis can be achieved by increasing the neoplasia detection rate (NDR) (Figure 1). For every 1% increase in NDR, the post-endoscopy Barrett's neoplasia rate can be reduced by 3.5%.[11]

Early AI research focused primarily on endoscopic still images for AI training and validation. More recently, AI research has progressed to evaluating model performance using video feeds from endoscopic procedures and implementing real-time systems in the endoscopy suite. This shift from still images to video and real-time applications underscores the rapid advancements in AI computing capabilities. In earlier real-time AI studies, endoscopy often required a brief pause to allow the system to generate an output. However, most modern algorithms now operate without requiring direct endoscopist intervention, enabling seamless integration of AI outputs, which are typically visualized as heatmaps or bounding boxes around ROIs, providing intuitive guidance during procedures (Figure 2). The final step toward commercial deployment involves overcoming the gap from preclinical validation to real-world implementation. Table 1 summarizes the current studies on AI application in BE and early Barrett's neoplasia.

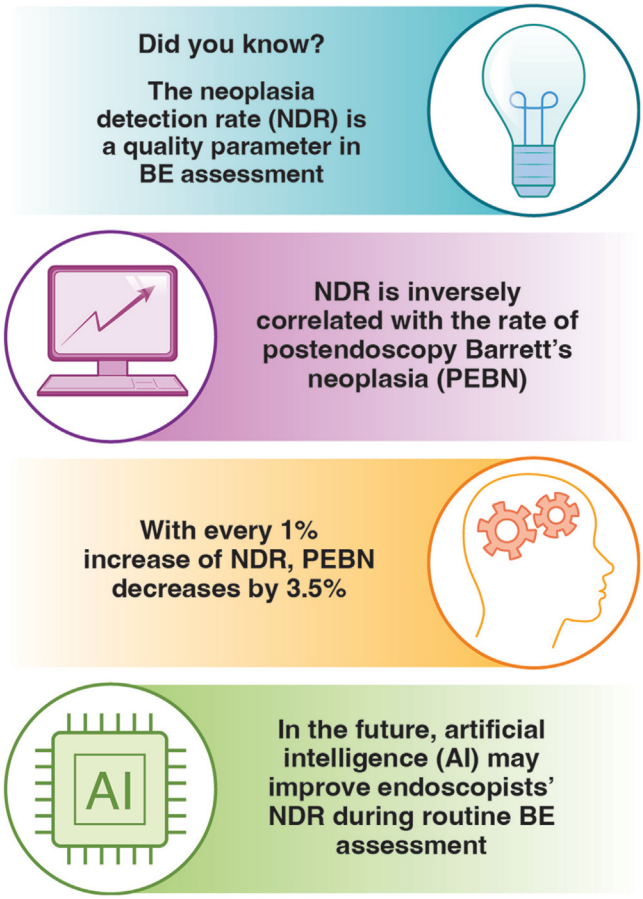

Figure 1. AI may improve NDR during routine endoscopic assessment of BE.

Artificial Intelligence Studies on Still Endoscopy Images

An initial study on AI's application in the assessment of still BE images reported excellent accuracy rates and overlap scores (Dice coefficient) between AI image segmentation and expert annotations. The major limitation was the small number of data used for training and validation, performed as cross-validation on an internal dataset.[13] Nevertheless, this was a pioneer report of DL with neural networks assessing EAC endoscopic images. Another early study developed a pattern recognition–based computer algorithm that could detect EAC in BE images with reasonable accuracy.[12] A landmark study based on extensive collection of images from multiple centers and countries developed an AI system based on a CNN.[14] They performed an image-based trial, comparing the AI algorithm with international experts. Major strengths included the large datasets used for training and external validation and number of endoscopists participating in the study's comparisons. Interestingly, the AI system far outperformed the endoscopists, while showing excellent delineation scores of tumor boundaries in the segmentation task. In a follow-up pilot study performed during live endoscopic procedures, the same group demonstrated the system's integration in real-life

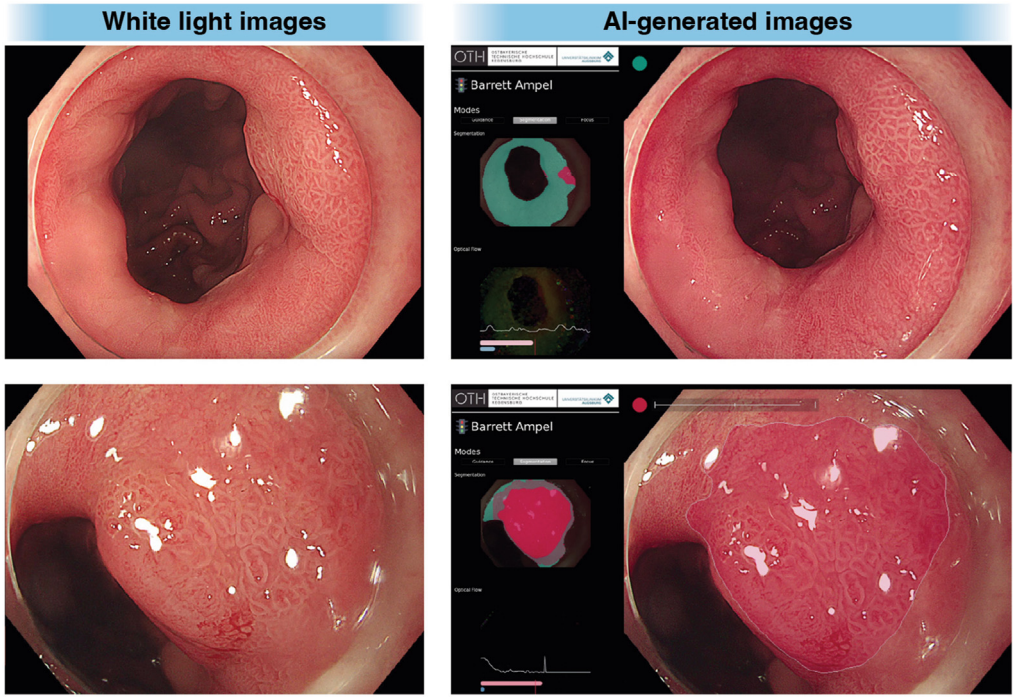

Figure 2. AI-assisted detection and characterization of early Barrett's neoplasia. The *left panel* shows the overview and near-focus WLIs of a subtle lesion located at the level of the cardia; the *right panel* shows the AI-generated heatmaps and overlay within the user interface predicting neoplasia.

endoscopic examinations, importantly showcasing the feasibility of transitioning from preclinical endoscopic image analysis to real-time procedures.[15] However, there may be a gap between AI performance in high-expert academic centers and community hospitals without extensive BE experience.[14,15]

Clinical Decision Support With Artificial Intelligence

Abdelraheem et al[20] published data from a multicenter video-based study in which AI detected subtle BE-related neoplasia (BERN) in short video clips. Expert endoscopists assessed the same videos, and the standalone performance of AI was compared with that of the endoscopists. AI significantly outperformed endoscopists in all counts, including sensitivity, specificity, and accuracy. However, standalone performance studies cannot truly depict endoscopist behavior using AI as a clinical decision support system. The decision made by endoscopists using AI is not driven solely by the AI output, but also by internal and external factors, such as technology aversion, complacency, and automation bias, or simply fatigue and time of the day. Two recent AI studies on the performance of endoscopists in BE videos better mimicked real-world situations with AI used as a clinical decision support system. In the first, non–BE experts, in conjunction with AI, significantly improved their NDR by approximately 12%. In the second, non–BE experts improved their NDR by approximately 9%. Two important insights can be drawn from these studies: first, expert endoscopists did not profit from using AI, and second, the standalone performance of AI systems was still significantly higher than the final performance of nonexperts using AI.[21,28] To better understand the human–AI interaction, research is expected to advance rapidly toward the application of AI in real-time during endoscopic procedures. For example, 2 studies have demonstrated AI's implementation in real-life settings with excellent results.[17,23]

When analyzing these studies, it is important to understand the limitations of data generated from experimental settings, which do not necessarily depict routine situations in everyday endoscopic practice. The excellent results may reflect the inherent limitations associated with experimental situations that may provide more optimized conditions for AI systems. Small patient or data sample sizes and a high proportion of pathologies within the study population could skew study results toward an unreasonably high sensitivity and should be interpreted cautiously. Other issues, such as the Hawthorne effect and overfitting, should also be considered as potential sources of bias.

Characterization, Choice of Lesion, and Infiltration Depth

In the studies described above, most data focused on detecting and improving NDR with AI. However, the further characterization of visible abnormalities detected during endoscopy may be challenging, even with image-enhanced endoscopy. Furthermore, characterization could also involve optical diagnosis of a neoplastic lesion for covert features of high-risk histopathology. Jukema et al[24] developed and validated an AI system for improving the detection and characterization of neoplastic lesions in patients with BE, basically distinguishing between BERN and nondysplastic BE (NDBE) and significantly improving the performance of general endoscopists. The possible future impact is in the

Table 1. Relevant Studies on the Application of Artificial Intelligence in the Evaluation and Assessment of Barrett's Esophagus and Early Barrett's Neoplasia

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
AI studies on endoscopy images						
van der Sommen[12] (2016)	Detection of neoplastic lesions in BE using AI	Observational study	250	50	ML (support vector machine)	Sensitivity and specificity of 0.83 (per-image analysis), Sensitivity and specificity of 0.86 and 0.87 (patient level)
Ebigbo[13] (2019)	AI-based diagnosis in early EAC	Prospective study	200	50	DL (ResNet)	Sensitivity and specificity of 97% and 88% (Augsburg data)
de Groof[14] (2020)	DL system for neoplasia detection in Barrett's with benchmarking	Multistep training and validation	1	250	DL	Accuracy of up to 89% Sensitivity and specificity of 90% and 88%
de Groof[15] (2020)	DL algorithm accuracy in live procedures for Barrett's neoplasia	Pilot study	300	100	DL-based image recognition (Hybrid ResNet/U-Net)	Accuracy of 90% Sensitivity and specificity of 91% and 89%
Hashimoto[16] (2020)	Real-time detection of esophageal neoplasia in BE	Pilot study	350	80	CNN (Inception-ResNet-v2)	Accuracy of 95.4% Sensitivity and specificity of 96.4% and 94.2%
Ebigbo[17] (2020)	Real-time AI use in cancer evaluation in BE	Prospective study	250	60	DL (ResNet with DeepLab V.3+)	Accuracy of 89.9% Sensitivity and specificity of 83.7% and 100%
Ali[18] (2021)	3D quantification of BE for risk stratification	Pilot study	200	70	3D imaging, AI (ResNet-50 and DeepLabv3+)	Accuracy of up to 98.4%
Ebigbo[19] (2021)	AI in predicting submucosal invasion in Barrett's cancer	Pilot study	150	50	DL (101-layer residual CNN)	Accuracy of 0.71 Sensitivity and specificity of 0.77 and 0.64

Table 1. Continued

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
van der Laan[8] (2023)	Optical biopsy of dysplasia in BE using AI	Observational study	300	100	CNN with transfer learning (ResNet)	Accuracy of 77.67% (+13.0% with AI assistance; $P =$ .020) compared with the baseline Sensitivity of 90.0% (+32.67% with AI assistance; $P <$ .001) compared with the baseline
Tsai[9] (2023)	Detection of BE using AI	Pilot study	400	100	CNN (EfficientNetV2B2)	Accuracy of 94.37% Sensitivity and specificity of 94.29% and 94.44%
Abdelrahi[20] (2023)	Neural network model for Barrett's neoplasia detection	Multicenter trial	500	150	Artificial neural network	Accuracy of 94.7% (phase 2) and 92.0% (phase 3) Sensitivity and specificity of 95.3% and 94.5% (phase 2) and 93.8% and 90.7% (phase 3)
Fockens[21] (2023)	DL for early Barrett's neoplasia detection	Model development and validation	600	150	DL (EfficientNet-Lite1 encoder and MobileNetV2 DeepLabV3+ decoder)	Sensitivity of 95% (images) and 97% (videos)
Fockens[22] (2023)	Compact DL system for early Barrett's neoplasia detection	Retrospective study	450	120	DL (EfficientNet-Lite1 encoder and MobileNetV2 DeepLabV3+ decoder)	Sensitivity of 84% (test set 1), 100% (test set 2) and 88% (test set 3)
Fockens[23] (2024)	Real-time CAD system for visible lesions in Barrett's	Pilot study	400	120	CNN (CAD system)	Sensitivity and specificity of 100% and 53% (per-patient), Sensitivity and specificity of 100% and 73% (per-level)

Table 1. Continued

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
Jukema[24] (2024)	AI-aided diagnosis for Barrett's neoplasia	Prospective study	450	120	DL (EfficientNet-Lite1 feature encoder)	Sensitivity and specificity of 100% and 98% for images and 93% and 96% for videos
AI studies on pathology images						
Tomita[25] (2019)	Normal, NDBE, BE with dysplasia, and adenocarcinoma detection	Single center	379 images	123 images	Attention-based model	Mean accuracy of 0.83
Sali[26] (2020)	Dysplastic and nondysplastic BE detection	Multicenter	387 images	535 images	CNN, multiple-instance learning (fully supervised learning, unsupervised learning, weakly supervised learning)	AUCs (up to 0.986)
Faghani[27] (2022)	NDBE, LGD, and HGD classification	Single center	368 slides (8596 bounding boxes)	70 slides (840 bounding boxes)	ResNet101 model	Accuracy of 95.7% in diagnosing BE dysplasia

3D, 3-dimensional.

additive value of such AI systems during BE surveillance and the transition toward AI-guided targeted biopsies.

For obvious BE neoplasia, a possible application of AI is the choice of lesions amenable to endoscopic resection or resection modality. It is often difficult for endoscopists assessing BERN to accurately predict the depth of invasion, grade of differentiation, or presence of lymphatic or vessel infiltration. These criteria, associated with higher risk of lymphonodal metastases, may influence lesion choice for endoscopic resection or the resection modality, such as endoscopic mucosal resection or endoscopic submucosal dissection. A retrospective multivariate analysis showed that the size of the lesion, deep submucosal invasion, and lymphatic or vessel infiltration, especially combined, could increase the metastasis risk by up to 70%. An initial attempt at risk stratification with optical diagnosis showed that expert endoscopists could differentiate between mucosal and submucosal tumors with a sensitivity of only 40%–50% and a low interobserver variability of approximately 41%.[29,30] Hence, AI could improve risk stratification of early Barrett's neoplasia according to the criteria above. A pilot study trained an algorithm to identify lesions with submucosal invasion on endoscopic still images. The overall accuracy of the AI algorithm (71%) was equal to the performance of expert endoscopists, highlighting the potential of AI for lesion risk stratification.[19]

Effect of Lesion Morphology on Artificial Intelligence Performance

To better understand the challenges associated with early esophageal cancer, particularly early Barrett's cancer, it is essential to understand the types of lesions used for training and validation. This is crucial because, in real-world clinical settings, most missed lesions are flat and subtle, typically classified as Paris IIa and IIb. Many studies have incorporated such lesions in their training datasets, and some have reported performance metrics stratified by Paris classifications. In a real-time pilot study, AI achieved a sensitivity of 100% for Paris IIa, IIb, and Is lesions, and per-level specificity was 73%.[23] Another study highlighted the impact of the Paris classification on AI and endoscopist performance, showing that most lesions included were Paris IIa (46.9%) and IIb (28.1%), with AI sensitivity slightly decreasing from 93.3% for Paris IIa lesions to 88.9% for Paris IIb lesions.[20] Sensitivity of 100% was achieved for polypoid lesions (Paris Is), emphasizing the need for enhanced AI training to address the challenges of detecting flat lesions. One of the most comprehensive studies on AI in early Barrett's cancer enriched the dataset with 80% Paris IIa/IIb lesions, thereby focusing on subtle and flat morphologies to reflect real-world challenges; the performance of general endoscopists significantly improved with AI assistance, particularly for flat lesions.[21] Another study exclusively included flat or slightly elevated lesions (Paris IIa/IIb) for a standalone AI sensitivity of 92.2%, further demonstrating AI's effectiveness in challenging scenarios.[28] These findings underscore the importance of incorporating flat and subtle lesions into AI training datasets and refining algorithms to improve the detection of such morphologies in clinical practice.

Applications of Deep Learning in Barrett's Esophagus Histopathologic Diagnosis

Supervised learning is the form of machine learning (ML) most pathologists are likely to encounter in a diagnostic setting, such as the classification of pathology images using CNNs. Although supervised DL methods have achieved remarkable success in digital pathology, recent progress in unsupervised and weakly supervised AI approaches has demonstrated significant potential to reduce dependence on extensive pathologist annotations. These approaches use large volumes of unlabeled or minimally labeled data to develop robust diagnostic models.

In the histologic diagnosis of BE and associated neoplasia, DL is highly relevant, and several studies have investigated its applications. For example, a study introduced an attention-based DL model to dynamically identify ROIs using a CNN-based algorithm to classify histology images into categories of NDBE, dysplastic BE, and EAC, achieving classification accuracies of 0.85, 0.89, and 0.88, respectively.[25] Another study constructed a DL model identifying dysplastic BE and NDBE with high-resolution histopathologic images from esophagus and gastroesophageal junction mucosal biopsy using only tissue-level labeling.[25] This framework achieved a mean classification accuracy of 0.83 (95% CI, 0.80–0.86) across normal, BE without dysplasia, BE with dysplasia, and EAC classes. By eliminating the need for manual ROI annotations, the model streamlined the diagnostic process while maintaining robust performance.[25] Another study[26] used 387 pathologic whole-slide images (WSIs) from biopsy or mucosal resection to construct a DL model for dysplastic from NDBE differentiation; they compared different feature representation approaches, including unsupervised, weakly supervised, and fully supervised learning on model performance and showed that the appropriate setting of unsupervised feature representation could extract more relevant image features from WSIs compared with weakly supervised and fully supervised methods, thereby improving dysplastic BE and NDBE identification. This approach reduced reliance on annotated training data while achieving high diagnostic accuracy.[26]

Furthermore, DL models are useful for predicting dysplasia grades based on pathologic images. A study constructed a DL model to classify NDBE, low-grade dysplasia (LGD), and high-grade dysplasia (HGD) on WSIs, demonstrating high sensitivity and specificity for LGD and excellent performance for NDBE and HGD, potentially enhancing the histologic diagnosis of BE dysplasia and guide the appropriate use of endoscopic therapy.[27]

Image-Based Artificial Intelligence in Esophageal Squamous Cell Cancer

AI systems have demonstrated promising results in detecting and diagnosing ESCC,[31] with some studies showing performance comparable with or even superior to that of experienced endoscopists.[32] In addition, AI models have proven effective in delineating ESCC margins during

white-light imaging (WLI) endoscopy.[33] Most AI systems have been trained on early-stage cancers, characterized by subtle morphology that is challenging to identify using conventional endoscopy. To improve robustness and specificity, some studies have included noncancerous lesions in their training datasets, such as inflammatory lesions or ectopic gastric mucosa.[34] The cancerous lesions used were typically classified based on the Japanese Esophageal Classification, which uses intrapapillary capillary loop patterns to detect and characterize lesions. AI has specifically used intrapapillary capillary loop patterns from the Japanese Esophageal Classification to assess the invasion depth, distinguishing between intramucosal cancers (m) and submucosally invasive cancers (sm1/2/3). AI systems could play a pivotal role in ESCC detection and in aiding endoscopist selection of lesions suitable for endoscopic resection.

Challenges of Artificial Intelligence in Early Esophageal Cancer

Despite the impressive progress in AI-assisted detection of esophageal neoplasia, several challenges remain. A primary example in BE is the inherent complexity of Barrett's lesions compared with other GI neoplasms, for example, colonic adenomas. Barrett's lesions are often multifocal, flat, and subtle, complicating their detection using conventional AI models.[21] This is in stark contrast to the more easily recognizable protruding lesions typically found in the colon.[10,35] In addition, a reliable reference standard is difficult to provide using histopathology. The result of a targeted biopsy revealing neoplasia can probably correlate well with the endoscopic image used for AI training; however, the resection of a wide BE area cannot be fully correlated with the final histopathology, owing to the possibility of different levels of dysplasia lying immediately adjacent to each other.

The reference standard for AI benchmarking in BE is currently under debate. Whether AI systems should be compared with expert endoscopists or histopathologic confirmation remains unclear, and more clinical trials are required to determine the most appropriate benchmark. In addition, the variability in diagnostic thresholds among human endoscopists complicates the widespread adoption of AI in clinical practice. The question remains: What should be the ultimate goal of AI in relation to the current diagnostic reference standard, for example, the Seattle biopsy protocol? Should AI aim to ultimately replace random biopsies, or should its development focus on enhancing diagnostic accuracy by improving the quality of targeted biopsies? This debate is particularly relevant considering that most AI systems have been, and likely will continue to be, trained primarily for detecting HGD and mucosal or submucosal lesions. Conversely, data on its utility for LGD remain limited. For LGD, the substantial interobserver variability among pathologists makes it extremely difficult to establish a reliable reference standard, emphasizing the importance of a second, independent reference pathologist. Histologic diagnoses of LGD are often subject to up- or downgrading on expert review, and its management remains controversial, given the highly variable progression rates. These uncertainties and controversies pose a major obstacle to developing AI systems specifically designed to enhance early BERN detection and characterization.

The lack of large and diverse datasets, even for obvious high-grade or mucosal/submucosal lesions, further complicates the development of robust and generalizable AI systems. Many AI models are trained using data from specialized centers, limiting their ability to be generalized to broader and more diverse patient populations. This poses a challenge for AI systems that seek to accurately identify early-stage neoplasia in BE.

Accurate delineation of lesion boundaries remains a significant hurdle. Although AI systems are adept at detecting lesions, accurately outlining the borders of neoplastic tissues is difficult. This is particularly problematic in cases when lesions are flat or have subtle morphologic changes, often seen in early-stage BE neoplasia.[22] Misidentification of lesion boundaries can lead to inappropriate treatment decisions, such as incomplete resection or overtreatment.

For ESCC, a major challenge lies in integrating AI systems into real-time clinical settings, as highlighted by a prospective study that failed to demonstrate the noninferiority of AI compared with human endoscopists; the AI system achieved an accuracy of 80.6% for ESCC detection, compared with 85.7% for human endoscopists.[36] AI systems must be optimized for rapid data processing and analysis, while maintaining high diagnostic accuracy. Geographical constraints add to the complexity, as the limited incidence of ESCC in certain regions results in scarce data for the development of AI models. Although AI systems may ultimately affect the quality of esophageal cancer diagnosis and management, their integration into routine workflows would require significant changes to existing practices.[37]

Artificial Intelligence in Stomach Neoplasia Detection: Current Data and Clinical Indications

Artificial Intelligence–Assisted Diagnosis Via Endoscopy

Endoscopy is essential for diagnosing gastric lesions. AI has been applied to many medical imaging analyses, including endoscopy. AI-assisted endoscopy for premalignant and malignant gastric lesions is currently one of the most actively researched areas.[38] In addition, even experienced endoscopists may miss or misdiagnose lesions for factors such as heavy workloads, fatigue, or diverse lesion variations during endoscopic examinations.

AI is a noninvasive technology that can support diagnostic workflows by monitoring hidden or hard-to-detect blind spots during endoscopic diagnosis in real time. Use of an AI system during upper GI endoscopy significantly reduces the miss rate in the detection of gastric lesions. Furthermore, the automatic detection and classification of lesions could be beneficial for endoscopists, particularly those with less experience. Table 2 summarizes the current studies on AI in the endoscopy domain.

Table 2. Current Studies on Artificial Intelligence for Gastric Non-Neoplastic and Neoplastic Lesions in Endoscopy Domain

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
Diagnosis of AG and GIM						
Shichijo[39] (2017)	*Helicobacter pylori* infection detection	Multicenter	1750 patients (WLIs)	397 patients (WLIs)	CNN (GoogLeNet)	Accuracy (up to 87.7%) Sensitivity and specificity of up to 88.9% and 87.4%
Guimaraes[40] (2020)	Gastric precancerous lesion detection	Single center	200 images (WLIs)	70 images (WLIs)	CNN	AUC (up to 0.984) Sensitivity and specificity of up to 1.000 and 0.940
Zhang[41] (2020)	Chronic AG detection	Single center	3829 images (WLIs)	1641 images (WLIs)	CNN	Accuracy (0.942) Sensitivity and specificity of up to 0.945 and 0.940
Xu[42] (2021)	Gastric precancerous lesion detection	Multicenter	547 patients (ME-NBI)	213 patients (ME-NBI)	CNN (ENDOANGEL)	Accuracy (up to 0.908) Sensitivity and specificity of up to 0.966 and 0.930
Mu[43] (2021)	Gastric non-neoplastic lesion classification	Multicenter	4587 patients (WLIs)	523 patients (WLIs)	CNN	Accuracy (up to 95.0%)
Diagnosis of gastric neoplasms						
Kanesaka[44] (2018)	GC classification	Single center	126 M-NBI images	81 M-NBI images	Support vector machine	Accuracy (up to 96.3%) Sensitivity and specificity of up to 96.7% and 95%
Yoon[45] (2019)	GC classification, invasion depth prediction	Single center	11,539 images (WLIs)	1050 images (WLIs)	CNN (VGG-16)	AUC (0.981 for GC classification), AUC (0.851 for invasion depth prediction)
Wu[46] (2019)	GC and location classification	Multicenter	9151 images for GC classification, 24,549 images for monitoring blind spots (WLIs)	200 images (WLIs)	CNN	Accuracy (92.5% for GC classification, 90.0% for location classification) Sensitivity and specificity of 94.0% and 91.0% for GC classification

Table 2. Continued

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
Cho[47] (2019)	Gastric neoplasm classification	Multicenter	1057 patients (WLIs)	412 patients (WLIs)	CNN	AUCs (up to 0.927) Sensitivity and specificity of up to 84.0% and 87.3%
Zhu[48] (2019)	Invasion depth prediction	Single center	790 images (WLIs)	203 images (WLIs)	CNN (ResNet50)	AUC (0.94) Sensitivity and specificity of 76.47% and 95.56%
Luo[49] (2019)	Upper GI cancer detection	Multicenter	15,040 individuals (WLIs)	69,384 individuals (WLIs)	CNN (DeepLabV3+)	Accuracy (up to 0.977) Sensitivity and specificity of up to 0.982 and 0.979
Horiuchi[50] (2020)	GC classification	Single center	2570 ME-NBI	258 ME-NBI	CNN	Accuracy (85.3%) Sensitivity and specificity of 95.4% and 71.0%
Li[51] (2020)	GC classification	Multicenter	2088 images (ME-NBI)	341 images (ME-NBI)	CNN (Inception-v3)	Accuracy (90.91%) Sensitivity and specificity of 91.18% and 90.64%
Ling[52] (2020)	GC differentiation and margin prediction	Multicenter	145 patients for differentiation prediction 132 patients for margin prediction (ME-NBI)	139 patients for differentiation prediction 87 patients for margin prediction (ME-NBI)	CNN	Accuracy (83.3% for differentiation prediction), Accuracy (82.7% in differentiated EGC, 88.1% in undifferentiated EGC for margin prediction)
Wu[53] (2021)	Gastric neoplastic lesion detection	Single center	1812 patients (WLIs)	NA	DL (ENDOANGEL-LD)	Significantly lower gastric neoplasm miss rate in the AI-first group (6.1%)
Wu[38] (2021)	GC classification	Multicenter	1050 patients (WLIs)	NA	CNN (ENDOANGEL)	Accuracy (84.7%) Sensitivity and specificity of 100% and 84.3%
Hu[54] (2021)	GC classification	Multicenter	170 cases (ME-NBI)	125 cases (ME-NBI)	CNN (VGG-19)	AUCs (up to 0.813) Sensitivity and specificity of 0.782 and 0.741

Table 2. Continued

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
Ikenoyama[55] (2021)	GC classification	Multicenter	2639 GCs (WLIs)	140 cases (WLIs)	CNN	AUCs (0.757) Sensitivity and specificity of 58.4% and 87.3%
Ueyama[56] (2021)	GC classification	Single center	5574 ME-NBI images	2300 ME-NBI images	CNN (ResNet50)	Accuracy (98.7%) Sensitivity and specificity of 98.7% and 98%
Wu[57] (2022)	Invasion depth and differentiation prediction	Multicenter	100 videos (WLIs)	NA	CNN (ENDOANGEL)	Accuracy (78.57% for invasion depth prediction, 71.43% for differentiation prediction) Sensitivity of 87.81% for detecting neoplasm, 100% for diagnosing EGC
Niikura[58] (2022)	GC classification	Single center	500 patients (WLIs)	NA	CNN	Per-image diagnosis rate (99.87%)
Yuan[59] (2022)	Gastric lesion classification	Single center	8947 patients (WLIs)	496 patients (WLIs)	CNN	Accuracy (85.7%)
He[60] (2022)	GC classification	Multicenter	4667 M-IEE images	6026 M-IEE images	CNN (ENDOANGEL-ME)	Accuracy (up to 93.49%) Sensitivity and specificity of up to 96% and 93.51%
Li[61] (2022)	GC classification	Single center	1630 M-IEE images	267 M-IEE images	CNN (ENDOANGEL-LA)	Accuracy of 88.76% (ENDOANGEL-LA in images) Accuracy of 87.00% (ENDOANGEL-LA in videos)
Dong[62] (2023)	Early gastric neoplasms detection	Multicenter	NA (4482 images and 296 videos in total)	NA	CNN (ENDOANGEL-ED)	Accuracy of up to 88.24%
Zeng[63] (2024)	Gastric whitish lesion classification	Multicenter	NA (4558 images in total)	NA	CNN (ENDOANGEL-WD)	Accuracy of 80.47%

NA, not applicable.

Diagnosis of Atrophic Gastritis and Gastric Intestinal Metaplasia

Among the underlying causes of GC, *Helicobacter pylori* infection plays a pivotal role in GC pathogenesis; it induces atrophic gastritis (AG) and gastric intestinal metaplasia (GIM), eventually leading to GC development.[53] Endoscopic surveillance is generally advised for patients with extensive AG or GIM. However, studies have reported the sensitivity of endoscopic diagnosis of AG to be only 42%–61.5%.[64,65]

Shichijo et al[39] constructed a CNN model to diagnose the infection of *H pylori* based on endoscopic images, with higher accuracy than the endoscopist's visual inspection in a considerably shorter time. Guimaraes et al[40] presented a DL approach using real-world endoscopic images (WLI) of the proximal stomach to assess the endoscopic diagnosis of AG, which was significantly better than the combined results of expert endoscopists. Similarly, Zhang et al[41] constructed a CNN model to improve the diagnostic rate of AG, which was higher than those of the experts, highlighting AI's potential to enhance diagnostic outcomes.

A CNN model could diagnose *H pylori* infection based on endoscopic images with higher accuracy than the endoscopist's visual inspection in a considerably shorter time.[39] A DL approach using real-world endoscopic images (WLI) of the proximal stomach to assess the endoscopic diagnosis of AG was significantly better than the combined results of expert endoscopists.[40] Similarly, a CNN model had a higher diagnostic AG rate compared with experts, highlighting AI's potential to enhance diagnostic outcomes.[41]

Even if the entire stomach is examined, owing to the structural features, some blind spots in the gastric mucosa, for example, the sinus and small curvature of the fundus, may not be detected. Endoscopist ability to accurately identify gastric lesions varies, leading to a significant miss rate of approximately 25% in the detection of gastric neoplasms.[38] A recent multicenter retrospective study conducted in countries with low GC incidence revealed that the sensitivity for endoscopic recognition of AG and GIM recognition was only 48.5% and 16.3%, respectively.[66] Notably, proximal GIM was less likely to be identified endoscopically compared with distal GIM, highlighting the challenge of detecting subtle mucosal changes in the gastric body and fundus.[66]

Diagnosis of Gastric Neoplasms

GC is known to progress through stages of premalignant lesions before eventually developing into cancer.[67] With the advancements in diagnosis and treatment modalities, the overall 5-year survival rate of patients with early GC (EGC) has exceeded 90%, while that of patients with AGC is <50%, given its aggressive nature and swift progression.[68,69] Therefore, early GC detection through endoscopy is critically important. Appropriate histologic sampling facilitates the stratification of patients based on their risk of developing precancerous lesions.[70] DL development has provided valuable support in the diagnosis and treatment of GC. AI can assist in the detection and classification of GC during endoscopy, predict the invasion depth, and determine the resection margins.

The ENDOANGEL is an example of an AI system that helps physicians perform endoscopies by providing real-time assistance. The system outperformed endoscopists in detecting EGCs and was similar to endoscopists in predicting EGC invasion depth and differentiation status.[57] This DL-based system could serve as an effective training tool for unskilled endoscopists aiming to improve their observational skills. Even experienced endoscopists may miss or misdiagnose lesions due to factors such as heavy workloads, fatigue, or diverse variations of the lesions during endoscopic examinations. Wu et al[53] evaluated the effectiveness of an AI system (ENDOANGEL-LD) for the detection of focal lesions and prediction of gastric neoplasms using endoscopy with WLI; they found a significantly lower miss rate in detecting gastric neoplasms than that with standard upper GI endoscopy.

Luo et al[49] proposed a GI AI diagnostic system (GRAIDS) based on WLI for upper GI cancer diagnosis. In a multicenter, case-control, diagnostic study, GRAIDS achieved high diagnostic accuracy in identifying GC, with sensitivity similar to that of expert endoscopists and better than that of nonexpert endoscopists. GRAIDS was the first AI-powered, real-time image recognition system applied in clinical settings for the detection of upper GI cancers during endoscopy. Notably, the model demonstrated consistent diagnostic accuracy, even when validated across hospitals of varying levels, with potential to support nonexpert endoscopists in primary or low-volume hospitals.

The rate of GC diagnosis was higher in an AI group (99.87%) than in an expert endoscopist group (88.17%).[58] Recently, an AI system that could diagnose 6 common gastric lesions, including EGC, AGC, submucosal tumor, polyp, peptic ulcer, and erosion under WLE, was comparable with senior endoscopist, and notably outperformed junior endoscopist, performance.[59] Similarly, CNN models classifying gastric neoplasms into 5 categories—non-neoplasm, LGD, HGD, EGC, and AGC—based on endoscopic images, had classification performance comparable with that of experienced endoscopists.[47] These findings suggest the potential for in situ add-on testing to enhance the accuracy of diverse gastric lesion prediction.

Tumor invasion depth is a significant diagnostic factor for determining the appropriate treatment for GC. Minimally invasive treatments, such as endoscopic resection, can be performed when the cancer is confined to the mucosa or submucosa, reducing the need for surgery.[71] However, predicting the invasion depth during endoscopic examination is often challenging, and pathologic confirmation through invasive procedures remains the only reliable method. For example, recent studies showed that the detection accuracy of conventional endoscopy for predicting the tumor-invasion depth in EGC was only 73.7%,[72] promoting the application of DL using endoscopic images to predict the invasion depth of GC and screen patients for endoscopic resection. A CNN system could classify the GC invasion depth based on endoscopic images.[48] The DL classifier showed significantly higher accuracy and specificity than endoscopists. Likewise, to detect EGC and predict

invasion depth, a lesion-based DL model that classified T1a- and T1b-EGC achieved areas under the curve (AUCs) of 0.981 and 0.851 for EGC detection and depth prediction, respectively.[45]

Diagnosis With Magnified Image–Enhanced Endoscopy and Narrow-Band Imaging Endoscopy

Endoscopic examination using WLI is the standard for detecting GC, however, it relies on the expertise of endoscopists and remains inadequate for identifying EGC.[73,74] To improve diagnostic accuracy, magnified image–enhanced endoscopy (M-IEE), which provides a clear view of the microstructure and microvessels in the gastric mucosa, has been developed and widely applied to detect EGC.[75]

The AI system ENDOANGEL-ME applied to diagnose EGC with M-IEE in a multicenter diagnostic cohort achieved better diagnostic performance in internal and external images and video datasets than senior and junior endoscopists, suggesting great potential for EGC diagnosis under M-IEE in clinical practice.[60]

With rapid advancements in NBI endoscopy, magnifying endoscopy (ME)-NBI has gained significant attention for the diagnosis of GC, demonstrating greater accuracy than conventional WLI. However, interpreting ME-NBI patterns can be challenging, requiring extensive knowledge and specialized training of endoscopists, limiting the broader clinical application of this technology.[76] In addition, ME-NBI requires generating still images of high quality, which general endoscopists may struggle with. Several studies have attempted to apply DL to recognize precancerous conditions and EGC in ME-NBI.

ENDOANGEL, developed using diverse ME-NBI images, achieved high diagnostic accuracy and positive predictive value for detecting precancerous conditions, with performance comparable with that of experts and superior to that of nonexperts.[42] This implies that computer-aided detection (CAD) has significant potential in aiding the diagnosis of gastric precancerous conditions during IEE, while minimizing the need for extensive biopsy sampling in routine clinical practice. In addition, a computer-aided diagnostic model for GC detection in ME-NBI images showed performance comparable with that of experienced endoscopists and superior to that of junior endoscopists, highlighting its potential to enhance endoscopist diagnosis of EGC.[54]

DL systems could serve as useful tools for determining the surgical strategy and achieving curative treatment in EGC. A real-time DL system could predict the differentiation status and delineate the margins of EGC on ME-NBI.[52] The system revealed high accuracy in both predicting GC differentiation and delineating cancer margins, outperforming expert endoscopists.

Integration of Explainable Artificial Intelligence in Endoscopy

Explainable AI (XAI) integration into endoscopy has enhanced the interpretability and clinical utility of AI models, particularly in diagnosing EGC under both WLI and ME-NBI. A logical anthropomorphic diagnostic system incorporating feature extraction with quantitative analysis, ENDOANGEL-LA, was developed using DL and ML approaches to improve the identification of EGC under M-IEE. ENDOANGEL-LA achieved high diagnostic accuracy and provided interpretable outputs that facilitated clinician understanding and decision making, highlighting XAI's potential in bridging the gap between algorithmic outputs and clinical insights.[61] Similarly, an XAI framework using features that were integrated with literature research could diagnose early gastric neoplasms under WLI, facilitating clinician understanding of the rationale behind AI predictions. This approach further validated XAI's role in enhancing trust and adoption of AI systems in endoscopic practice by reducing the "black box" nature of conventional AI.[54]

Moreover, highly specialized AI systems focusing on unique challenges in endoscopic diagnosis, for example, morphologic variability in gastric lesions, have been developed. Gastric whitish lesions, encompassing a wide range of benign or malignant conditions, including AG, scarring, GC, and gastric mucosa–associated lymphoid tissue lymphoma, pose significant diagnostic difficulties, given their heterogeneous appearance. Incorporating domain knowledge into DL algorithms significantly improved the differentiation between diverse benign and malignant whitish gastric lesions in a retrospective study, enhancing diagnostic accuracy and highlighting the importance of integrating tailored AI models for specific lesion characteristics, thereby accommodating the diverse morphologic presentations of gastric neoplasms.[63]

Artificial Intelligence–Assisted Diagnosis in Pathology

The definitive diagnosis of various premalignant and malignant gastric lesions is made through the visual interpretation of histopathologic slide images by pathologists. However, manual pathologic examinations are labor-intensive and time-consuming. AI is expected to assist in performing precise and rapid pathologic assessments automatically, and various relevant studies have been conducted (Supplementary Table 1). AI-assisted pathologic diagnosis involves the automatic identification of gastric adenoma and carcinoma and histologic subclassification of GC.

Recent advancements in unsupervised and weakly supervised DL techniques have significantly improved the histopathologic diagnosis of various tumors. For example, weakly supervised DL achieved clinical-grade accuracy in pathology image analysis, decreasing reliance on expert-labeled datasets, while maintaining high diagnostic precision.[98] Weakly supervised learning combined with transfer learning could classify poorly differentiated adenocarcinomas from endoscopic submucosal dissection WSIs; the model achieved an AUC of up to 0.975, demonstrating robust performance without need for manual tumor-region annotations.[99] Similarly, Liang et al[82] introduced a reiterative learning framework for weakly annotated gastric histopathology images. By integrating predictions with weak annotations, their model achieved a mean Intersection over

Union of 0.883 and a mean accuracy of 91.09%, enabling effective tumor-region segmentation, despite limited labeling. These advances reduce the burden of manual labeling, enhancing generalizability through diverse datasets. Hence, unsupervised and weakly supervised AI systems present a cost-effective and scalable approach for improving GC detection and diagnosis, particularly in resource-limited settings.

A clinically applicable DL system for GC detection using H&E-stained WSIs showed a sensitivity of approximately 100% and a mean specificity of 80.6% on a real-world test dataset, suggesting the feasibility of using histopathologic AI assistance systems in routine workflows.[100] A fully automated DL approach distinguished between differentiated/undifferentiated and nonmucinous/mucinous tumor types in GC using The Cancer Genome Atlas Stomach Adenocarcinoma WSIs.[86] The patch-level AUCs for the differentiated/undifferentiated and nonmucinous/mucinous classifiers were 0.932 and 0.979, respectively, suggesting the DL-based approach as a useful tool for the quantitative analysis of GC histopathology images. CNNs and recurrent neural networks classifying GC, adenoma, and non-neoplastic lesions on 3 independent datasets of biopsy WSIs achieved AUCs up to 0.97 and 0.99 for GC and adenoma, respectively.[85] These advancements highlight AI's potential to transform histopathologic diagnosis by improving efficiency and accuracy, while reducing variability across clinical settings.

AI models have been used to enhance survival rates in GC. In recent years, AI applications in GC prognosis have focused on predicting survival time, lymph node metastasis (LNM), and tumor-infiltrating lymphocytes (TILs).

DL-based models assisting in the diagnosis and overall survival of patients with GC using digitalized pathologic images achieved accuracy comparable with that of expert pathologists and exceeded the performance of junior pathologists in external validation.[92] Moreover, the risk score of the DL system was a strong predictor of overall survival in both univariate and multivariate analyses.

LNM is a significant prognostic factor in GC.[101] The absence of precise methods for predicting LNM in GC has prompted the use of AI-assisted techniques to more effectively assess the risk of metastasis. A recent multi-institutional study proposed an ML algorithm that could predict the LNM status using H&E-stained GC WSIs.[99] The ML classifier achieved an overall AUC of 0.75 for LNM status prediction, hence suggesting its potential to improve the selection of EGC subgroups with high-risk histologic features. An AI-assisted LN assessment workflow using gigapixel images facilitated the routine evaluation of LNM, with a slide-level AUC of 0.9936.[94] Clinical application of this system demonstrated that the workflow could significantly improve the sensitivity of micrometastasis and isolated tumor cell detection in a significantly shorter review time, indicating the feasibility of this algorithm in routine workflows.

TILs play a regulatory role in tumor-associated immune responses and are important prognostic factors for solid organ cancers, including GC.[102,103] An unsupervised consensus clustering algorithm could identify 3 immune subtypes (ie, IS1, IS2, and IS3) with diverse components, including tumor-infiltrating immune cells, molecular features, and clinical outcomes. Three immune subtypes were identified: "immune-inflamed," with active immune responses and better prognosis (IS3); "immune-desert," with low immune activity and intermediate prognosis (IS2); and "immune-excluded," characterized by strong immunosuppressive features and poorer prognosis (IS1). The model showed strong predictive performance across training, validation, and test cohorts using GC WSIs, highlighting DL's potential to investigate tumor immune microenvironments and immune subtypes, offering support for developing immunotherapeutic strategies for GC.[91] For example, tumors in the immune-inflamed subtype (IS3) may respond better to immune checkpoint inhibitors, while other subtypes (IS1 and IS2) may require treatments targeting regulatory immune cells or pathways to enhance immune activation and overcome suppression. A related study presented the DL workflow to generate TIL maps to reveal the abundance and spatial distribution of TILs using GC WSI, with AUCs of up to 0.821.[93]

Microsatellite instability (MSI) is an important prognostic and treatment response indicator for various cancers.[104] Because immune checkpoint inhibitors are much more effective in tumors with MSI-high than microsatellite stable status, the MSI status is routinely assessed in multiple cancer types, including GC.[105] However, because of technical limitations, identifying MSI through immunohistochemistry or genetic analysis is not always feasible. Therefore, several studies have explored the feasibility of DL-based prediction of the MSI status from H&E-stained WSIs. A DL system achieved MSI recognition in GC directly from H&E-stained WSIs, with an accuracy of 83.87%.[95] Similarly, a fully automated DL classifier predicted the MSI status with The Cancer Genome Atlas Stomach Adenocarcinoma tissue WSIs, with AUCs of 0.893 and 0.902 for frozen and formalin-fixed paraffin-embedded GC tissues, respectively,[96] highlighting the potential of DL-based MSI classifiers as effective screening tools for definitive cases.

Driver mutations may cause alterations in cancer cell morphology, such as changes in nuclear and cytoplasmic texture, size, and shape in histologic images. Moreover, cancer cells can alter their adjacent stroma to form a supportive environment for tumor progression, that is, the "reactive" tumor stroma, leading to secondary changes in tumor cell morphology on a micrometer or millimeter scale.[106] Although changes associated with individual oncogenic mutations may be subtle, DL can reliably detect these alterations.[107] Jang et al[89] evaluated the feasibility of DL classifiers for mutations in *CDH1, ERBB2, KRAS, PIK3CA,* and *TP53* in GC tissues, with AUCs of up to 0.862 and 0.858 for frozen and formalin-fixed paraffin-embedded tissues, respectively, demonstrating the power of AI in mutational prediction from H&E WSIs in GC. This finding that DL-based classifiers can directly predict major mutations from H&E-stained WSIs when properly trained.

Challenges of Artificial Intelligence in Stomach Investigations

The main limitation of DL algorithms is the insufficient transparency regarding their underlying mechanisms and how they function.[108] This "black box" nature of algorithms may cause clinicians to not fully trust DL applications.

Various data visualization tools have now enhanced the visual interpretation of algorithm decision making, thereby promoting broader adoption of ML in clinical settings.[109] In addition, the recent advent of XAI can offer explanations for prediction models, providing clinicians with clear insights that help them understand the basis of model decisions.[110] Various data visualization tools have now enhanced the visual interpretation of algorithm decision making, thereby promoting broader ML adoption in clinical settings.[109,110]

Meanwhile, the advancement of automated medical diagnostic algorithms for various gastric diseases requires larger datasets and the ability to build AI-supported screening, diagnosis, and treatment decisions based on in-depth medical theories.

In current research on AI in GC, many studies are based on single-center, small-scale datasets. Although their outcomes often appear highly promising, they are typically validated using limited sample sizes, possibly leading to the risk of overfitting AI algorithms. Consequently, validation of the DL system using external datasets, ideally multicenter datasets, is crucial to ensure more robust and convincing results. Video images can serve as training resources for real-time lesion diagnosis, and the number of images can be further increased.

For image analysis in pathology using DL, the performance of DL systems is known to improve as the number of patients in the training set increases, reaching a plateau after training on 10,000–15,000 histologic WSIs. This underscores the need for large volumes of images and data to develop DL systems with optimal performance.[98]

Furthermore, many current DL algorithms are designed using supervised learning, which depends on data that are precisely labeled by experts. However, in medical image analysis, acquiring a large amount of high-quality labeled data is challenging due to the limited availability of domain experts, leaving much of the unlabeled data unexplored. Recently, several self-supervised and semi-supervised learning approaches have been developed to use unlabeled data that can be applied to AI in GC.[111]

Future Directions and Perspectives

The application of AI in the management of BE and other GI neoplasms is expected to increase significantly. AI's integration into screening and surveillance plans is an exciting area of development. AI may assist in neoplasia detection and in prognosis prediction and treatment planning, offering a more personalized approach to patient care[9] (Figure 3).

Shifting the focus of AI from a mere detection and characterization tool to a more versatile solution may increase its acceptance and applicability in BE. For example, AI could reconstruct the surface area of post-procedure BE, enabling 3-dimensional reassessment of the entire BE.[18] This approach could allow for more precise risk stratification and therapy monitoring.

In addition, applying weakly supervised AI in large-scale histopathologic screening has shown significant potential for improving BE detection. Bouzid et al[112] developed a DL model to analyze Cytosponge samples stained with H&E, eliminating the need for additional immunohistochemical staining. The model achieved high accuracy (AUC of 0.914 on discovery and 0.873 on external datasets) and reduced pathologist workload by 48%, enabling pathologists to prioritize only high-risk cases for manual review. This AI-enhanced approach enables cost-effective large-scale screening, particularly in resource-limited settings, while maintaining diagnostic precision and improving early detection of EAC.

AI applications in endoscopy include far more than pattern recognition or computer vision. Large language models and natural language processing will further

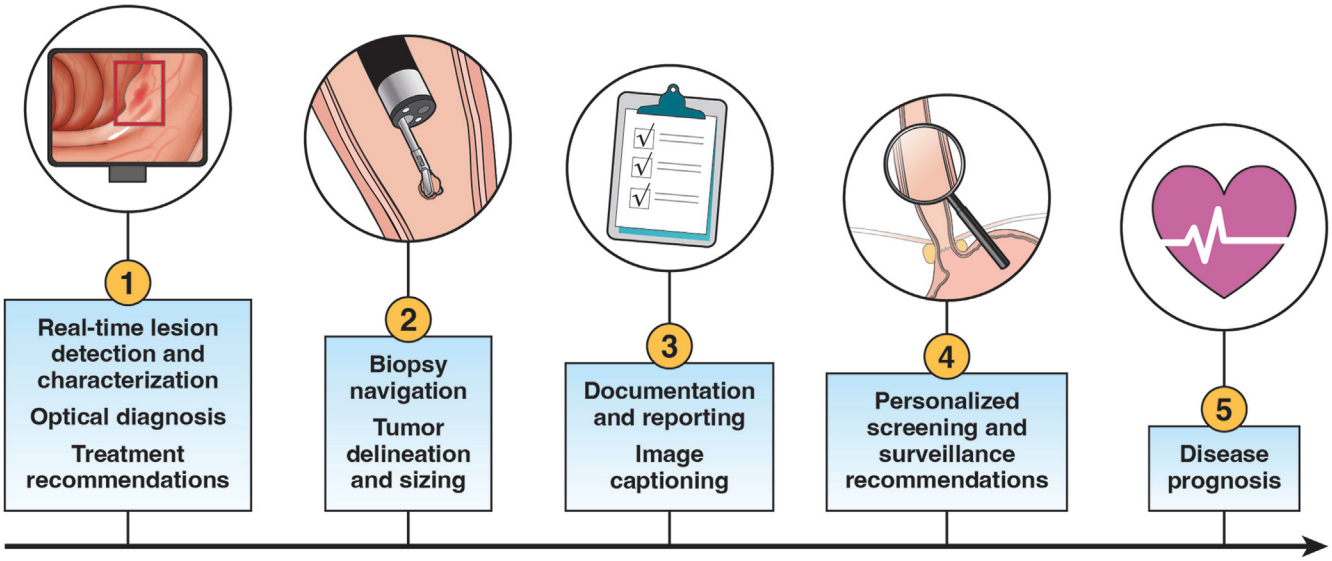

Figure 3. AI in the future: from real-time lesion detection to personalized screening and surveillance strategies.

broaden the possibilities of intelligent and automated clinical decision support by integrating speech- and text-based clinical data. AI might guide endoscopists through the entire procedure of BE inspection, lesion detection, characterization, documentation and reporting, surveillance, and treatment planning. A real-world, 2-center study found that CAD significantly enhanced detection rates when integrated into esophagogastroduodenoscopy workflows. In a high-biopsy-rate setting, CAD increased EGC detection from 0.35% to 0.59% ($P = .028$), while maintaining a stable overall gastric neoplasm detection rate. Conversely, in a low-biopsy-rate setting, CAD nearly doubled the overall gastric neoplasm detection rate from 1.78% to 3.23% ($P < .001$), primarily through improved detection of low-grade intraepithelial neoplasia. These findings emphasize the importance of tailoring CAD applications to the specific clinical context and optimizing their implementation across diverse practice environments.[113]

Current AI research on GI diseases, including GC, typically focuses on analyzing endoscopy images, computed tomography scans, and histopathologic WSIs separately. Therefore, the development of integrated multimodal algorithms that can analyze the areas simultaneously is required. In addition, most studies have addressed single tasks, such as GC detection, staging, or prognosis prediction, individually. A foundation model that can comprehensively connect these tasks would be capable of concurrently learning and analyzing information from various medical knowledge sources.[114] The development of such multimodal algorithms could lead to significant advances in medical AI technology.

The successful adoption of AI devices in clinical practice depends on their technical capabilities and on psychological and social factors that influence their acceptance and effectiveness. Endoscopists' and pathologists' trust in AI systems is crucial, as a cautious and skeptical attitude regarding the reliability of AI outputs can limit widespread implementation. Clear explainability, such as features provided by XAI, can bridge this gap by enhancing confidence in AI-assisted diagnostics. In addition, institutional support, access to clinician training programs, and equitable availability of AI technologies are critical social factors that influence the overall impact of AI systems.

Finally, AI systems must be adapted for use in community health care settings, where resources and expertise may be limited. Ensuring that AI tools are accessible and easy to use in these settings would be critical for their widespread adoption and successful integration into routine clinical practice. Future research should focus on bridging the gap between academic development and real-world applications to ensure that AI tools are effectively deployed across diverse health care environments.

Conclusions

AI can play a transformative role in detecting and managing premalignant and malignant neoplasms of the esophagus and stomach. AI has demonstrated significant potential for improving NDRs, enhancing diagnostic accuracy, providing prognostication, and offering real-time support to endoscopists and pathologists. However, challenges remain, such as the complexity of BE and ESCC lesions, need for larger datasets, and variability among human operators. As AI technologies continue to evolve, they will soon become integral to oncologic treatment plans and offer personalized and effective therapeutic options. By addressing the current limitations and focusing on practical implementation, AI can potentially revolutionize the fields of upper GI endoscopy and pathology, improving the outcomes in patients with early GI neoplasms.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2025.01.253.

References

1. Maret-Ouda J, Markar SR, Lagergren J. Gastroesophageal reflux disease: a review. JAMA 2020;324:2536–2547.
2. Alsop BR, Sharma P. Esophageal cancer. Gastroenterol Clin North Am 2016;45:399–412.
3. Marabotto E, Pellegatta G, Sheijani AD, et al. Prevention strategies for esophageal cancer-an expert review. Cancers (Basel) 2021;13:2183.
4. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74:229–263.
5. Ajani JA, Bentrem DJ, Besh S, et al. Gastric cancer, version 2.2013: featured updates to the NCCN Guidelines. J Natl Compr Canc Netw 2013;11:531–546.
6. Isobe Y, Nashimoto A, Akazawa K, et al. Gastric cancer treatment in Japan: 2008 annual report of the JGCA nationwide registry. Gastric Cancer 2011;14:301–316.
7. Jong MR, de Groof AJ. Advancement of artificial intelligence systems for surveillance endoscopy of Barrett's esophagus. Dig Liver Dis 2024;56:1126–1130.
8. van der Laan JJH, van der Putten JA, Zhao X, et al. Optical biopsy of dysplasia in barrett's oesophagus assisted by artificial intelligence. Cancers (Basel) 2023;15:1950.
9. Tsai MC, Yen HH, Tsai HY, et al. Artificial intelligence system for the detection of Barrett's esophagus. World J Gastroenterol 2023;29:6198–6207.
10. Meinikheim M, Messmann H, Ebigbo A. Role of artificial intelligence in diagnosing Barrett's esophagus-related neoplasia. Clin Endosc 2023;56:14–22.
11. Hamade N, Kamboj AK, Krishnamoorthi R, et al. Systematic review with meta-analysis: neoplasia detection rate and post-endoscopy Barrett's neoplasia in Barrett's oesophagus. Aliment Pharmacol Ther 2021;54:546–559.
12. van der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. Endoscopy 2016;48:617–624.

13. Ebigbo A, Mendel R, Probst A, et al. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. Gut 2019;68:1143–1145.
14. 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.
15. 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.
16. Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). Gastrointest Endosc 2020; 91:1264–1271.e1.
17. Ebigbo A, Mendel R, Probst A, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. Gut 2020;69:615–616.
18. Ali S, Bailey A, Ash S, et al. A pilot study on automatic three-dimensional quantification of Barrett's esophagus for risk stratification and therapy monitoring. Gastroenterology 2021;161:865–878.e8.
19. Ebigbo A, Mendel R, Ruckert T, et al. Endoscopic prediction of submucosal invasion in Barrett's cancer with the use of artificial intelligence: a pilot study. Endoscopy 2021;53:878–883.
20. 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.
21. Fockens KN, Jong MR, Jukema JB, et al. A deep learning system for detection of early Barrett's neoplasia: a model development and validation study. Lancet Digit Health 2023;5:e905–e916.
22. 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.
23. Fockens KN, Jukema JB, Jong MR, et al. The use of a real-time computer-aided detection system for visible lesions in the Barrett's esophagus during live endoscopic procedures: a pilot study (with video). Gastrointest Endosc 2024;100:527–531.e3.
24. 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.
25. Tomita N, Abdollahi B, Wei J, et al. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. JAMA Netw Open 2019;2:e1914645.
26. Sali R, Moradinasab N, Guleria S, et al. Deep learning for whole-slide tissue histopathology classification: a comparative study in the identification of dysplastic and non-dysplastic Barrett's esophagus. J Pers Med 2020; 10:141.
27. Faghani S, Codipilly DC, David V, et al. Development of a deep learning model for the histologic diagnosis of dysplasia in Barrett's esophagus. Gastrointest Endosc 2022;96:918–925.e3.
28. Meinikheim M, Mendel R, Palm C, et al. Influence of artificial intelligence on the diagnostic performance of endoscopists in the assessment of Barrett's esophagus: a tandem randomized and video trial. Endoscopy 2024; 56:641–649.
29. Gupta S, Mandarino FV, Shahidi N, et al. Can optical evaluation distinguish between T1a and T1b esophageal adenocarcinoma: an international expert interobserver agreement study. Endoscopy 2025;57:200–207.
30. Younis F, Rösch T, Beyna T, et al. Expert assessment of infiltration depth and recommendation of endoscopic resection technique in early Barrett cancer. United European Gastroenterol J 2024;12:848–858.
31. Waki K, Ishihara R, Kato Y, et al. Usefulness of an artificial intelligence system for the detection of esophageal squamous cell carcinoma evaluated with videos simulating overlooking situation. Dig Endosc 2021;33:1101–1109.
32. Tajiri A, Ishihara R, Kato Y, et al. Utility of an artificial intelligence system for classification of esophageal lesions when simulating its clinical use. Sci Rep 2022; 12:6677.
33. Liu W, Yuan X, Guo L, et al. Artificial intelligence for detecting and delineating margins of early ESCC under WLI endoscopy. Clin Transl Gastroenterol 2022;13: e00433.
34. Pan Y, He L, Chen W, et al. The current state of artificial intelligence in endoscopic diagnosis of early esophageal squamous cell carcinoma. Front Oncol 2023;13:1198941.
35. Massironi S. Advancements in Barrett's esophagus detection: the role of artificial intelligence and its implications. World J Gastroenterol 2024;30:1494–1496.
36. Tani Y, Ishihara R, Inoue T, et al. A single-center prospective study evaluating the usefulness of artificial intelligence for the diagnosis of esophageal squamous cell carcinoma in a real-time setting. BMC Gastroenterol 2023;23:184.
37. Patel A, Arora GS, Roknsharifi M, et al. Artificial intelligence in the detection of Barrett's esophagus: a systematic review. Cureus 2023;15:e47755.
38. 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 2021;53:1199–1207.
39. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images. EBioMedicine 2017;25:106–111.
40. Guimaraes P, Keller A, Fehlmann T, et al. Deep-learning based detection of gastric precancerous conditions. Gut 2020;69:4–6.
41. Zhang Y, Li F, Yuan F, et al. Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence. Dig Liver Dis 2020;52:566–572.

42. Xu M, Zhou W, Wu L, et al. Artificial intelligence in the diagnosis of gastric precancerous conditions by image-enhanced endoscopy: a multicenter, diagnostic study (with video). Gastrointest Endosc 2021;94:540–548.e4.
43. Mu G, Zhu Y, Niu Z, et al. Expert-level classification of gastritis by endoscopy using deep learning: a multicenter diagnostic trial. Endosc Int Open 2021; 9:E955–E964.
44. Kanesaka T, Lee TC, Uedo N, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. Gastrointest Endosc 2018;87:1339–1344.
45. Yoon HJ, Kim S, Kim JH, et al. A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. J Clin Med 2019;8:1310.
46. Wu L, Zhou W, Wan X, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. Endoscopy 2019;51:522–531.
47. Cho BJ, Bang CS, Park SW, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019; 51:1121–1129.
48. 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. Gastrointest Endosc 2019;89:806–815.e1.
49. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol 2019;20:1645–1654.
50. Horiuchi Y, Aoyama K, Tokai Y, et al. Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. Dig Dis Sci 2020;65:1355–1363.
51. 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; 23:126–132.
52. Ling T, Wu L, Fu Y, et al. A deep learning-based system for identifying differentiation status and delineating the margins of early gastric cancer in magnifying narrow-band imaging endoscopy. Endoscopy 2021;53:469–477.
53. 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.
54. Hu H, Gong L, Dong D, et al. Identifying early gastric cancer under magnifying narrow-band images with deep learning: a multicenter study. Gastrointest Endosc 2021; 93:1333–1341.e3.
55. Ikenoyama Y, Hirasawa T, Ishioka M, et al. Detecting early gastric cancer: comparison between the diagnostic ability of convolutional neural networks and endoscopists. Dig Endosc 2021;33:141–150.
56. Ueyama H, Kato Y, Akazawa Y, et al. Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging. J Gastroenterol Hepatol 2021;36:482–489.
57. 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.
58. Niikura R, Aoki T, Shichijo S, et al. Artificial intelligence versus expert endoscopists for diagnosis of gastric cancer in patients who have undergone upper gastrointestinal endoscopy. Endoscopy 2022;54:780–784.
59. Yuan XL, Zhou Y, Liu W, et al. Artificial intelligence for diagnosing gastric lesions under white-light endoscopy. Surg Endosc 2022;36:9444–9453.
60. He X, Wu L, Dong Z, et al. Real-time use of artificial intelligence for diagnosing early gastric cancer by magnifying image-enhanced endoscopy: a multicenter diagnostic study (with videos). Gastrointest Endosc 2022;95:671–678.e4.
61. Li J, Zhu Y, Dong Z, et al. Development and validation of a feature extraction-based logical anthropomorphic diagnostic system for early gastric cancer: a case-control study. EClinicalMedicine 2022;46:101366.
62. Dong Z, Wang J, Li Y, et al. Explainable artificial intelligence incorporated with domain knowledge diagnosing early gastric neoplasms under white light endoscopy. NPJ Digit Med 2023;6:64.
63. Zeng X, Yang L, Dong Z, et al. The effect of incorporating domain knowledge with deep learning in identifying benign and malignant gastric whitish lesions: a retrospective study. J Gastroenterol Hepatol 2024; 39:1343–1351.
64. Eshmuratov A, Nah JC, Kim N, et al. The correlation of endoscopic and histological diagnosis of gastric atrophy. Dig Dis Sci 2010;55:1364–1375.
65. Du Y, Bai Y, Xie P, et al. Chronic gastritis in China: a national multi-center survey. BMC Gastroenterol 2014; 14:21.
66. Honing J, Keith Tan W, Dieninyte E, et al. Adequacy of endoscopic recognition and surveillance of gastric intestinal metaplasia and atrophic gastritis: a multicentre retrospective study in low incidence countries. PLoS One 2023;18:e0287587.
67. de Vries AC, van Grieken NC, Looman CW, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. Gastroenterology 2008;134:945–952.
68. Necula L, Matei L, Dragu D, et al. Recent advances in gastric cancer early diagnosis. World J Gastroenterol 2019;25:2029–2044.
69. Goetze OT, Al-Batran SE, Chevallay M, et al. Multimodal treatment in locally advanced gastric cancer. Updates Surg 2018;70:173–179.
70. Pimentel-Nunes P, Libanio D, Marcos-Pinto R, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade

Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019. Endoscopy 2019;51:365–388.
71. Wang J, Yu JC, Kang WM, et al. Treatment strategy for early gastric cancer. Surg Oncol 2012;21:119–123.
72. Choi J, Kim SG, Im JP, et al. Comparison of endoscopic ultrasonography and conventional endoscopy for prediction of depth of tumor invasion in early gastric cancer. Endoscopy 2010;42:705–713.
73. Raftopoulos SC, Segarajasingam DS, Burke V, et al. A cohort study of missed and new cancers after esophagogastroduodenoscopy. Am J Gastroenterol 2010;105:1292–1297.
74. Osawa H, Yamamoto H, Miura Y, et al. Diagnosis of depressed-type early gastric cancer using small-caliber endoscopy with flexible spectral imaging color enhancement. Dig Endosc 2012;24:231–236.
75. Miyaoka M, Yao K, Tanabe H, et al. Diagnosis of early gastric cancer using image enhanced endoscopy: a systematic approach. Transl Gastroenterol Hepatol 2020;5:50.
76. Nakayoshi T, Tajiri H, Matsuda K, et al. Magnifying endoscopy combined with narrow band imaging system for early gastric cancer: correlation of vascular pattern with histopathology (including video). Endoscopy 2004; 36:1080–1084.
77. Cosatto E, Laquerre P, Malon C. Automated gastric cancer diagnosis on H&E-stained sections; Itraining a classifier on a large scale with multiple instance machine learning. Proc SPIE 2013;8676:867605.
78. Sharma H, Zerbe N, Klempert I, et al. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Comput Med Imaging Graph 2017;61:2–13.
79. Garcia E, Hermoza R, Castanon CB, et al. Automatic lymphocyte detection on gastric cancer IHC images using deep learning. In: 2017 IEEE 30th international Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2017.
80. Qu J, Hiruta N, Terai K, et al. Gastric pathology image classification using stepwise fine-tuning for deep neural networks. J Healthc Eng 2018;2018:8961781.
81. Yoshida H, Shimazu T, Kiyuna T, et al. Automated histological classification of whole-slide images of gastric biopsy specimens. Gastric Cancer 2018;21:249–257.
82. Liang Q, Nan Y, Coppola G, et al. Weakly supervised biomedical image segmentation by reiterative learning. IEEE J Biomed Health Inform 2019;23:1205–1214.
83. Saritha D, Pavani LLC, Pramidha PS, et al. Accurate gastric cancer segmentation in digital pathology images using deformable convolution. Turk J Comput Mathemat Educ 2023;14:273–281.
84. Wang S, Zhu Y, Yu L, et al. RMDL: recalibrated multi-instance deep learning for whole slide gastric image classification. Med Image Anal 2019;58:101549.
85. Iizuka O, Kanavati F, Kato K, et al. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. Sci Rep 2020;10:1504.
86. Jang HJ, Song IH, Lee SH. Deep learning for automatic subclassification of gastric carcinoma using whole-slide histopathology images. Cancers (Basel) 2021;13:3811.
87. Jiang Y, Xie J, Han Z, et al. Immunomarker support vector machine classifier for prediction of gastric cancer survival and adjuvant chemotherapeutic benefit. Clin Cancer Res 2018;24:5574–5584.
88. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25:1054–1056.
89. Jang HJ, Lee A, Kang J, et al. Prediction of genetic alterations from gastric cancer histopathology images using a fully automated deep learning approach. World J Gastroenterol 2021;27:7687–7704.
90. Wang X, Chen Y, Gao Y, et al. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. Nat Commun 2021;12:1637.
91. Chen Y, Sun Z, Chen W, et al. The immune subtypes and landscape of gastric cancer and to predict based on the whole-slide images using deep learning. Front Immunol 2021;12:685992.
92. Huang B, Tian S, Zhan N, et al. Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: a retrospective multicentre study. EBioMedicine 2021;73:103631.
93. Abousamra S, Gupta R, Hou L, et al. Deep learning-based mapping of tumor infiltrating lymphocytes in whole slide images of 23 types of cancer. Front Oncol 2021;11:806603.
94. Huang SC, Chen CC, Lan J, et al. Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. Nat Commun 2022;13:3347.
95. Su F, Li J, Zhao X, et al. Interpretable tumor differentiation grade and microsatellite instability recognition in gastric cancer using deep learning. Lab Invest 2022; 102:641–649.
96. Lee SH, Lee Y, Jang HJ. Deep learning captures selective features for discrimination of microsatellite instability from pathologic tissue slides of gastric cancer. Int J Cancer 2023;152:298–307.
97. Sung YN, Lee H, Kim E, et al. Interpretable deep learning model to predict lymph node metastasis in early gastric cancer using whole slide images. Am J Cancer Res 2024;14:3513–3522.
98. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019; 25:1301–1309.
99. Tsuneki M, Kanavati F. Weakly supervised learning for poorly differentiated adenocarcinoma classification in gastric endoscopic submucosal dissection whole slide images. Technol Cancer Res Treat 2022;21: 15330338221142674.
100. Song Z, Zou S, Zhou W, et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. Nat Commun 2020;11:4294.
101. Yokota T, Ishiyama S, Saito T, et al. Lymph node metastasis as a significant prognostic factor in gastric cancer: a multiple logistic regression analysis. Scand J Gastroenterol 2004;39:380–384.
102. deLeeuw RJ, Kost SE, Kakal JA, et al. The prognostic value of FoxP3+ tumor-infiltrating lymphocytes in

cancer: a critical review of the literature. Clin Cancer Res 2012;18:3022–3029.
103. Kim HI, Kim H, Cho HW, et al. The ratio of intra-tumoral regulatory T cells (Foxp3+)/helper T cells (CD4+) is a prognostic factor and associated with recurrence pattern in gastric cardia cancer. J Surg Oncol 2011; 104:728–733.
104. Mann SA, Cheng L. Microsatellite instability and mismatch repair deficiency in the era of precision immuno-oncology. Expert Rev Anticancer Ther 2020;20:1–4.
105. Shimozaki K, Nakayama I, Hirota T, et al. Current strategy to treat immunogenic gastrointestinal cancers: perspectives for a new era. Cells 2023;12:1049.
106. Mueller MM, Fusenig NE. Friends or foes - bipolar effects of the tumour stroma in cancer. Nat Rev Cancer 2004; 4:839–849.
107. Kather JN, Heij LR, Grabsch HI, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer 2020;1:789–799.
108. Salahuddin Z, Woodruff HC, Chatterjee A, et al. Transparency of deep neural networks for medical image analysis: a review of interpretability methods. Comput Biol Med 2022;140:105111.
109. Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017.
110. Binzagr F. Explainable AI-driven model for gastrointestinal cancer classification. Front Med (Lausanne) 2024; 11:1349373.
111. Eckardt JN, Bornhauser M, Wendt K, et al. Semi-supervised learning in cancer diagnostics. Front Oncol 2022;12:960984.
112. Bouzid K, Sharma H, Killcoyne S, et al. Enabling large-scale screening of Barrett's esophagus using weakly supervised deep learning in histopathology. Nat Commun 2024;15:2026.
113. Huang L, Xu M, Li Y, et al. Gastric neoplasm detection of computer-aided detection-assisted esophagogastroduodenoscopy changes with implement scenarios: a real-world study. J Gastroenterol Hepatol 2024; 39:2787–2795.
114. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature 2023;616:259–265.

Received November 15, 2024. Accepted January 22, 2025.

Correspondence
Address correspondence to: Alanna Ebigbo, MD, PhD (current affiliation), Department of Internal Medicine and Gastroenterology, St. Josef-Hospital, Ruhr University of Bochum, Gudrunstr. 56, 44791 Bochum, Germany. e-mail: alanna.ebigbo@gmx.de; or Helmut Messmann, MD, PhD, Department of Gastroenterology, University Hospital Augsburg, Stenglinstrasse 2, 86156 Augsburg, Germany. e-mail: helmut.messmann@uk-augsburg.de; or Sung Hak Lee, MD, PhD, Department of Hospital Pathology, College of Medicine, The Catholic University of Korea, Seoul St. Mary's Hospital, Banpo-daero 222, Seoul, South Korea. e-mail: hakjjang@catholic.ac.kr.

CrediT Authorship Contributions
Alanna Ebigbo, MD, PhD (Conceptualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)
Helmut Messmann, MD, PhD (Conceptualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal)
Sung Hak Lee, MD, PhD (Conceptualization: Equal; Writing – original draft: Equal; Writing – review & editing: Equal).

Conflicts of interest
The authors disclose no conflicts.

Funding
Sung Hak Lee reports that this research was funded by a grant from the National Research Foundation of Korea (NRF-2022R1A2C2010644).

Supplementary Table 1. Current Studies on Artificial Intelligence for Gastric Non-Neoplastic and Neoplastic Lesions in Pathology Domain

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
Cosatto[77] (2013)	GC classification	Single center	8000 patients	4000 patients	MIL	AUC (0.96)
Sharma[78] (2017)	GC classification, necrosis detection	Single center	231,000 images	NA	CNN (AlexNet)	Accuracy (0.699 for GC classification), Accuracy (0.814 for necrosis detection)
Garcia[79](2017)	TIL detection	Single center	3257 images	NA	CNN	Accuracy (96.88%)
Qu[80] (2018)	GC classification	Single center	10,800 patches	5400 patches	CNN	AUCs (up to 0.965)
Yoshida[81] (2018)	Gastric neoplasm classification	Single center	3062 biopsy specimens	NA	MIL	Overall concordance rate (55.6%)
Liang[82] (2019)	GC classification	Single center	1900 images	NA	CNN	Accuracy (91.09%)
Saritha[83] (2023)	GC classification	Single center	350 images	150 images	CNN (VGG19 and ResNetV2)	Accuracy (91.6%)
Wang[84] (2019)	GC classification	Single center	408 images	200 images	MIL	Accuracy (86.5%)
Iizuka[85] (2020)	GC classification	Single center	1746 biopsy WSIs	NA	CNN, RNN	AUCs (up to 0.98)
Jang[86] (2021)	GC classification	Single center	371 patients	232 patients	CNN (Inception-v3)	AUCs (up to 0.979)
Jiang[87] (2018)	Prognosis prediction	Multicenter	251 patients	535 patients	SVM	AUCs (up to 0.834)
Kather[88] (2019)	MSI prediction	Multicenter	315 patients	185 patients	CNN (ResNet18)	AUC (0.81)
Jang[89] (2021)	Genetic alteration prediction	Single center	371 patients	96 patients	CNN (Inception-v3)	AUCs (up to 0.862)
Wang[90] (2021)	Prognosis prediction	Multicenter	1164 patients	307 patients	U-Net	Accuracy (96.9%)
Chen[91] (2021)	Immune subtypes prediction	Single center	84 cases	85 cases	CNN (ResNet18)	Accuracy (up to 85.71%)
Huang[92] (2021)	GC classification, Prognosis prediction	Multicenter	1037 patients	91 patients	CNN (GastroMIL and MIL-GC)	Accuracy (up to 0.920)
Abousamra[93] (2022)	TIL detection	Single center	434 WSIs	NA	CNN	AUCs (up to 0.821)

Supplementary Table 1. Continued

Study, first author (year)	Purpose	Design	No. of cases in training set	No. of cases in test set	AI models	Model performance
Huang[94] (2022)	Prognosis prediction	Multicenter	5907 LN images	NA	CNN	AUC (0.9936)
Su[95] (2022)	MSI prediction	Single center	348 patients	119 patients	CNN	Accuracy (83.87%)
Lee[96] (2023)	MSI prediction	Single center	331 patients	383 patients	CNN (Inception-v3)	AUCs (up to 0.902)
Sung[97] (2024)	Prognosis prediction	Multicenter	987 WSIs	250 WSIs	ML (XGBoost)	AUC (0.7487)

MIL, multiple-instance learning; NA, not applicable; RNN, recurrent neural network; SVM, support vector machine.