Gastroenterology Report, 2022, 1–16

https://doi.org/10.1093/gastro/goac064
Review Article

REVIEW ARTICLE

Artificial intelligence in gastric cancer: applications and challenges

Runnan Cao[1,2†], Lei Tang[3†], Mengjie Fang[1,2,5], Lianzhen Zhong[1,2], Siwen Wang[1,2], Lixin Gong[4], Jiazheng Li[3], Di Dong [1,2*] and Jie Tian[5,6*]

[1]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, P. R. China, [2]CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, [3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/ Beijing), Radiology Department, Peking University Cancer Hospital & Institute, Beijing, P. R. China, [4]College of Medicine and Biological Information Engineering School, Northeastern University, Shenyang, Liaoning, P. R. China, [5]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, P. R. China and [6]Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi, P. R. China

*Corresponding authors. Di Dong, CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing 100190, P. R. China. Tel: +86-13811833760; Email: di.dong@ia.ac.cn; Jie Tian, Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, P. R. China. Tel: +86-10-82618465; Email: jie.tian@ia.ac.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Abstract

Gastric cancer (GC) is one of the most common malignant tumors with high mortality. Accurate diagnosis and treatment decisions for GC rely heavily on human experts' careful judgments on medical images. However, the improvement of the accuracy is hindered by imaging conditions, limited experience, objective criteria, and inter-observer discrepancies. Recently, the developments of machine learning, especially deep-learning algorithms, have been facilitating computers to extract more information from data automatically. Researchers are exploring the far-reaching applications of artificial intelligence (AI) in various clinical practices, including GC. Herein, we aim to provide a broad framework to summarize current research on AI in GC. In the screening of GC, AI can identify precancerous diseases and assist in early cancer detection with endoscopic examination and pathological confirmation. In the diagnosis of GC, AI can support tumor-node-metastasis (TNM) staging and subtype classification. For treatment decisions, AI can help with surgical margin determination and prognosis prediction. Meanwhile, current approaches are challenged by data scarcity and poor interpretability. To tackle these problems, more regulated data, unified processing procedures, and advanced algorithms are urgently needed to build more accurate and robust AI models for GC.

Key words: gastric cancer; artificial intelligence; radiomics; endoscopy; computed tomography; pathology

Submitted: 4 July 2022; **Revised:** 27 September 2022; **Accepted:** 18 October 2022

Introduction

Gastric cancer (GC) is the fifth most common type of malignant tumor and ranks fourth in causes of cancer-related death, with >1,000,000 new cases and 769,000 deaths (~1 in 13 deaths) in 2020 [1]. The incidence and mortality rates vary in sex and region, with 2-fold higher rates in men than women and the highest incidence rate in Eastern Asia. The prognosis of advanced gastric cancer (AGC) is poor, with a 5-year-survival rate of ≤30%, while the rate of early gastric cancer (EGC) can reach 90% [2, 3]. However, it is difficult to detect EGC due to its latent and non-specific symptoms. Endoscopic examination is the most common method for early detection with multiple enhanced techniques [4], but definite diagnosis relies on pathological confirmation via biopsy. The following pathology and computed tomography (CT) examination are suggested to determine the subtypes and stages of the tumor, assisting with the treatment decision and prognosis prediction. Patients with EGC are recommended to perform radical resection, while patients in the advanced stages might require a combination of surgery, chemotherapy, and radiotherapy [5, 6]. In addition, immunotherapy and molecular-targeted drugs demonstrate excellent prognosis for some specific types of GC [7].

In the diagnosis and treatment of GC, multiple procedures rely on the observations and judgments of radiologists and pathologists, which might suffer from limited experience, objective criteria, and inter-observer discrepancies [8–10]. Artificial intelligence (AI), which can learn from data, is widely applied in image processing, including medical images. Traditional machine-learning methods rely on handcraft features predefined by domain experts. Radiomics provides a framework to extract imaging features for cancer diagnosis and treatment, including GC [11–14]. With the availability of large-scale data sets and the improvement of algorithms, deep learning has achieved great success in the computer vision [15–17]. Meanwhile, in medical image processing, deep-learning models have been successfully applied with pre-trained models [18] or self-designed networks [19]. In Table 1, we list the clinical trials of AI applications in GC. Most trials were proposed within the last 3 years and are ongoing.

Figure 1 illustrates the roadmap of AI-supported applications in GC. This review will analyse recent representative studies of AI applications in GC and discuss the current abilities, limitations, and future directions.

AI-supported screening

GC screening is the most promising and popular research of AI in GC. The development of GC is accompanied by a series of stages with a growing cancer risk: *Helicobacter pylori* (HP)-induced atrophic gastritis (AG), followed by gastric intestinal metaplasia (GIM), dysplasia, and finally cancer [20, 21]. Identifying these precancerous gastric diseases is an efficient method to screen individuals with high risk, thereby reducing the incidence of GC. In addition, because of the subtle morphological changes, GC is commonly diagnosed in the advanced stage [22]. Patients in the advanced stage of GC have a 5-year survival rate of ≤30%, while individuals in the early stage can achieve a significantly higher rate of 91.5% [2, 3]. Thus, it is crucial to detect GC as early as possible. Endoscopic examination is routinely used for GC screening. However, previous studies have reported that miss rates of GC detection ranged from 4.6% to 25.8% by endoscopists [10, 23–26]. Misses were related to the types and positions of the tumor, also occurring more for endoscopists with limited experience. Advanced image-enhanced endoscopy techniques can improve GC detection [27], but its universal application is hindered by additional training and expertise. In addition, the definite diagnosis of GC relies on visual examination of the whole slide imaging (WSI) pathological images obtained by biopsy or resection [28]. However, due to the large scale of WSI and various sizes of cancerous regions, it requires long-term concentration and tedious efforts for pathologists to detect GC in WSI. AI could assist automatic, precise, and speedy endoscopic detection, and histopathological examination considering all of the issues.

Table 2 lists some applications of AI in the screening of GC. Studies vary in the number of patients, imaging modality, applied method, and also predicted performance. Researchers tried to detect the status of HP infection from endoscopic images [29–33]. Yan *et al.* [30] highlighted the three-category classification, including the eradicated state. All research reached a performance rate of >0.8, which is comparable to expert endoscopists. For diagnosis of AG and GIM, researchers achieved an accuracy of ~0.9 [34–36], which is more accurate than predicting HP infection. Meanwhile, researchers achieved moderate performance in classifying multiple precancerous statuses and gastric neoplasm [37–39]. They explored the traditional machine-learning method [40, 41] and applied deep-learning models [42, 44, 47] to detect GC in endoscopic images. To improve the detection accuracy, researchers utilized advanced image-enhanced endoscopy [43, 49] and a self-designed network structure [45, 46, 48], respectively. Moreover, some researchers focused on the detection of GC in pathological images [50–53].

Precancerous diseases identification

In 2015, more than half of the world's population was infected with HP [54]. According to the International Agency of Research on Cancer, HP infection is categorized as a first-class carcinogen, which increases the risk of GC [55]. Moreover, even individuals after eradicating HP show a higher risk of cancer than healthy people without HP infection [56, 57]. Thus, detecting the status of HP infection can play a significant role in preventing GC. Endoscopy can predict HP infection, while its accuracy is hindered by the experience of endoscopists and inter-observer variability. Endoscopic biopsy is favored as the definitive diagnosis, while it requires an invasive operation. Therefore, researchers are looking for an accurate and automatic method for diagnosis from endoscopic images.

Huang *et al.* [29] performed pioneering research of applying AI to detect HP infection in 2004. They trained the refined feature selection with a neural network on endoscopic images from 30 patients and predicted HP infection of 74 patients, achieving a sensitivity of 85.4% and a specificity of 90.9%. Although the result is promising, it is validated on a relatively small data set. Shichijo *et al.* [30] used a 22-layer convolutional neural network (CNN) to diagnose the infection of HP on 32,208 training images and 11,481 test images. The CNN algorithm demonstrated powerful capability by achieving higher accuracy and shorter time than the endoscopist's visual inspection. They extended this work by including patients after HP eradication and achieved 80% accuracy of negative diagnoses, 84% eradicated, and 48% positive [32]. Besides the studies on white light imaging (WLI) images, research was conducted on color-enhanced endoscopic images. Yasuda *et al.* [33] explored the CNN model on blue laser imaging (BLI) images, while Nakashima *et al.* [31] used light color imaging (LCI), BLI, and WLI

Table 1. Clinical trials of artificial intelligence in gastric cancer

Study	Sponsor	Date	Recruitment status	ClinicalTrials.gov identifier
Radiomics for Prediction of Lymph Node Metastasis in Gastric Cancer	Nanfang Hospital of Southern Medical University, China	5 April 2018	Recruiting	NCT03488446
Radiomics and Molecular Expression Predictive Model for Esophago-gastric Junction and Gastric Cancer TRG	University of Roma La Sapienza, Italy	7 May 2021	Recruiting	NCT04878783
Study on the Effectiveness of Gastroscope Operation Quality Control Based on Artificial Intelligence Technology	Peking University, China	12 May 2020	Recruiting	NCT04384575
Quality Control of Esophagogastroduodenoscopy (EGD) Using Real-time EGD Auxiliary System	Shandong University, China	20 March 2019	Completed	NCT03883035
AQCS for Detection of Early Cancer and Precancerous Lesions on Upper Gastrointestinal Tract	Shandong University, China	22 January 2021	Recruiting	NCT04720924
Prediction of Gastric Cancer in Intestinal Metaplasia and Atrophic Gastritis	Chinese University of Hong Kong, China	9 April 2021	Not yet recruiting	NCT04840056
Artificial Intelligence Versus Expert Endoscopists for Diagnosis of Gastric Cancer	Tokyo University, Japan	31 July 2019	Completed	NCT04040374
Validation of an Artificial Intelligence System Based on Raman Spectroscopy for Diagnosis of Gastric Premalignant Lesions and Early Gastric Cancer	Changi General Hospital, Singapore	3 May 2021	Recruiting	NCT04869618
Development of a Clinical Decision Support System with Artificial Intelligence for Cancer Care	National University Hospital, Singapore	19 December 2020	Recruiting	NCT04675138
The Research of Constructing a Risk Assessment Model for Gastric Cancer Based on Machine Learning	Second Affiliated Hospital, Zhejiang University, China	12 July 2021	Recruiting	NCT04957407
Multicenter and Multi-modal Deep Learning Study of Gastric Cancer	First Hospital of China Medical University, China	11 August 2021	Not yet recruiting	NCT05001321

We searched "gastric cancer"/"stomach cancer" with "artificial intelligence"/"machine learning"/"deep learning" in ClinicalTrials.gov, and eliminated the items not related to gastric cancer or artificial intelligence methods, finally obtaining the 11 trials.

images to train GoogLeNet. The results showed that the area under curve (AUC) obtained from BLI and LCI was 0.96 and 0.95, respectively, and was significantly higher than 0.66 for WLI, revealing the benefits of utilizing color-enhanced images in the automatic diagnosis of HP infection.

AG is a crucial step in the progression of GC. However, a multicenter national study in China shows that the sensitivity of diagnosis by endoscopists is only 42% [58]. Meanwhile, histological confirmation through multi-point biopsy is painful, costly, and time-consuming. Guimarães *et al.* [34] first presented a CNN approach to classifying endoscopic images in the presence of AG. They achieved a diagnostic accuracy of 93% on 70 test images, which outperformed experienced endoscopists. Zhang *et al.* [35] confirmed the potential for the deep-learning method in the task on a larger data set containing 5,470 images from 1,699 patients. Furthermore, the diagnostic accuracy, sensitivity, and specificity were 0.9424, 0.9458, and 0.9491, respectively, which exceeded human experts. The accurate diagnosis of GIM also suffered from similar limitations [59]. Yan *et al.* [36] built an intelligent diagnostic system based on multiple CNN architecture to achieve a better performance than human experts including sensitivities (91.9% vs 86.5%), specificities (86.0% vs 81.4%), and accuracies (88.8% vs 83.8%).

Since there exist multiple differentiated gastric lesions, Cho *et al.* [38] applied CNN models to classify gastric neoplasms into five categories: non-neoplasm, low-grade dysplasia, high-grade dysplasia, EGC, and AGC. The Inception-ResNet-v2 model showed lower performance than experienced endoscopists with five-category accuracy (76.4% vs 87.6%). To explore the potential for deep learning in differentiating gastric ulcers and cancer, Lee *et al.* [39] applied CNN models to classify endoscopic images into three categories: normal, ulcer, and cancer. Classification of ulcer and cancer resulted in relatively poor performance compared with cases involving normal subjects, resulting from the modest difference between cancer tissue and ulcer regions. This study reveals the potential for deep learning in spotting gastric neoplasms. Moreover, to overcome the deficiency of training big models with too many parameters in a relatively

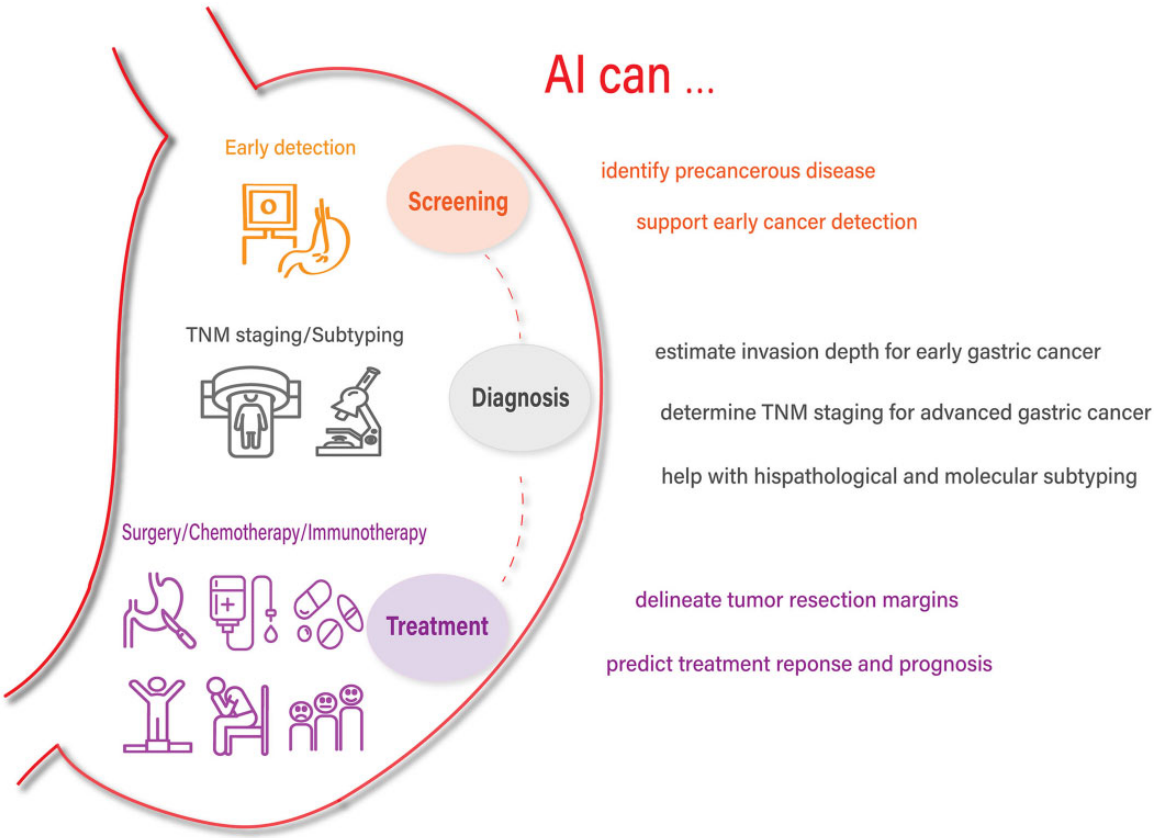

Figure 1. Roadmap of artificial intelligence (AI)-supported applications in gastric cancer

small gastric imaging data set, Zhang *et al.* [37] built a concise full convolutional network called the Gastric Precancerous Disease Network by introducing modules from SqueezeNet. GPDNet was fine-tuned by a proposed novel algorithm called the iterative reinforced learning method and improved the accuracy to 88.9% in the classification of three-class gastric precancerous disease (polyp, erosion, and ulcer).

Early cancer detection

Assisting endoscopic diagnosis

Traditional machine-learning methods require extracting the most robust and representative handcraft features from images. Miyaki *et al.* [40] explored the quantitative identification of mucosal GC on magnifying endoscopic images with flexible spectral imaging color enhancement (FICE). They used densely sampled scale-invariant feature transform (SIFT) descriptors as local features. A hierarchical k-means clustering algorithm obtained the bag-of-features representation, followed by a support vector machine (SVM)-based classifier, yielding a cancer detection sensitivity of 84.8%, specificity of 87.0%, and the accuracy of 85.9% using the cut-off value of 0.59 [65]. They further extended this method on magnifying endoscopy with BLI. The SVM output showed significant discrepancy for cancerous lesions (0.846 ± 0.220), reddened lesions (0.381 ± 0.349), and surrounding tissue (0.219 ± 0.277). These works reveal the potential for AI in automatic GC diagnosis, but its reliability is hampered by the human-selected cut-off value and limited size of the data set.

The developments of deep learning have enabled an end-to-end training manner that combines feature extraction and classification automatically. Hirasawa *et al.* [42] first explored deep-learning-based detection of GC using a single-shot multibox detector (SSD) [60] to predict the presence or absence of GC, its stage (EGC or AGC), and its position. The CNN model achieved automatic GC lesion diagnosis with a sensitivity of 92.2%, where the missed cases were of differentiated type and similar to gastritis. However, only 30.6% precision was reached due to many misdiagnoses of AG and GIM as GC. To further validate the power of deep learning, they compared the diagnosis ability of the CNN model with 67 endoscopists [47]. For diagnostic performance on a single image compared with endoscopists, CNN showed better performance with less diagnostic time (45.5 seconds vs 173 minutes) and significantly higher sensitivity (58.4% vs 31.9%), while its predicted precision (26% vs 46.2%) and specificity (87.3% vs 97.4%) were slightly lower. In addition, they analysed the causes of false-positive and false-negative cases predicted by CNN and endoscopists in detail, which is worthwhile for future research. Considering the limitation of small data sets, Luo *et al.* [44] developed and validated a Gastrointestinal Artificial Intelligence Diagnostic System (GRAIDS) on 1,036,496 endoscopic images from 84,424 patients. The DeepLab-based model was applied to identify the existence of tumors and segment the tumor region, which achieved performance similar to expert endoscopists and even more superior to novice endoscopists. Furthermore, in a randomly selected prospective cohort, results showed that trainee endoscopists and low-volume hospitals could benefit more from the assistance of GRAIDS. This study is the largest multicenter

Table 2. The application of AI in the screening of gastric cancer

Authors	Aim	Data	Method	Result
Huang *et al.* (2004) [29]	Detect HP	104 patients; WLI	RFSNN	Sen: 85.4%, Spe: 90.9%
Shichijo *et al.* (2017) [30]	Detect HP	2,165 patients; WLI	GoogLeNet	Acc: 83.1%, Sen: 81.9%, Spe: 83.4%
Nakashima *et al.* (2018) [31]	Detect HP	222 patients; WLI, LCI, and BLI	GoogLeNet	AUC: 0.66 (WLI), 0.95 (LCI), 0.96 (BLI)
Shichijo *et al.* (2019) [32]	Classify HP state: positive, negative, or eradicated	6,083 patients; WLI	GoogLeNet	Acc: 80% (negative), 84% (eradicated), 48% (positive)
Yasuda *et al.* (2020) [33]	Detect HP	105 patients; LCI	SVM	Acc: 87.6%, Sen: 90.4%, Spe: 85.7%
Guimarães *et al.* (2020) [34]	Diagnose AG	270 patients; WLI	VGG16	Acc: 92.9%, Sen: 100%, Spe: 87.5%
Zhang *et al.* (2020) [35]	Diagnose AG	1,699 patients; WLI	DenseNet	Acc: 94.2%, Sen: 94.5%, Spe: 94.0%
Yan *et al.* (2020) [36]	Diagnose GIM	416 patients; NBI and ME-NBI	EfficientNetB4	Acc: 88.8%, Sen 91.9%, Spe: 86.0%
Zhang *et al.* (2017) [37]	Classify gastric precancerous diseases	1,331 images; WLI	GPDNet	Acc: 88.9%
Cho *et al.* (2019) [38]	Classify gastric neoplasm	1,469 patients; WLI	Inception-ResNet-v2	Acc: 76.4%
Lee *et al.* (2019) [39]	Classify gastric neoplasm	787 images; WLI	ResNet-50	Acc: 96.5% (normal vs cancer), 92.6% (normal vs ulcer), 77.1% (cancer vs ulcer)
Miyaki *et al.* (2013) [40]	Identify GC	46 patients; ME-FICE	SIFT features; SVM	Acc: 85.9%, Sen: 84.8%, Spe: 87.0%
Miyaki *et al.* (2015) [41]	Identify GC	95 patients; ME-BLI	SIFT features; SVM	SVM outputs: 0.846 (cancer), 0.381 (redness), 0.219 (surrounding)
Hirasawa *et al.* (2018) [42]	Identify and segment GC	2,716 lesions; WLI	SSD	Sen: 92.2%, PPV: 30.6%
Liu *et al.* (2018) [43]	Identify GC	1,120 images; ME-NBI	InceptionV3	Acc: 85.9%, Sen: 84.8%, Spe: 87.0%
Luo *et al.* (2019) [44]	Identify and segment GC	84,424 patients; 6 centers; WLI	DeepLabV3+	Acc: 92.8%, Sen: 94.2%, Spe: 92.3%, PPV: 81.4%, NPV: 97.8%
Hsu *et al.* (2019) [45]	Identify GC	473 images; ME-NBI	SSSNet	Acc: 91.7%, Sen: 90.0%, Spe: 93.3%, PPV: 93.1%
Yoon *et al.* (2019) [46]	Identify and segment GC	11,539 images; WLI	Lesion-based VGG	AUC: 0.981
Ikenoyama *et al.* (2021) [47]	Identify and segment GC	16,524 images; WLI	SSD	Sen: 58.4%, Spe: 87.3%, PPV: 26.0%, NPV: 96.5%
Nguyen *et al.* (2020) [48]	Identify pathological site	7,894 images; WLI	Ensemble of deep-learning models	Acc: 70.7%
Hu *et al.* (2021) [49]	Identify GC	295 patients; 3 centers; ME-BLI	VGG19	Acc: 77.0%, Sen: 79.2%, Spe: 74.5%
Li *et al.* (2018) [50]	Identify GC	700 slices; pathological image	GastricNet	Acc: 97.9% (patch), 100% (slice)
Li *et al.* (2018) [51]	Identify and segment GC	700 slices; pathological image	GT-Net	F1 score: 90.9%
Sun *et al.* (2019) [52]	Identify and segment GC	500 images; pathological image	Multi-scale embedding networks	Acc: 81.6 (pixel), mIoU: 82.65%
Wang *et al.* (2019) [53]	Identify GC	608 slices; pathological image	RMDL	Acc: 86.5%

Acc, accuracy; AG, atrophic gastritis; AI, artificial intelligence; AUC, area under the receiver-operating characteristic curve; BLI, blue laser imaging; GC, gastric cancer; GIM, gastric intestinal metaplasia; GPDNet, Gastric Precancerous Disease Network; HP, *Helicobacter pylori*; LCI, light color imaging; mIoU, mean Intersection over Union; ME-BLI, magnifying endoscopy with blue laser imaging; ME-FICE, magnifying endoscopy with flexible spectral imaging color enhancement; ME-NBI, magnifying endoscopy with narrow-band imaging; NBI, narrow-band imaging; NPV, negative predictive value; PPV, positive predictive value; RFSNN, refined feature selection with neural network; RMDL, recalibrated multi-instance deep learning; Sen, sensitivity; SIFT, scale-invariant feature transform; Spe, specificity; SSD, single-shot multi-box detector; SSSNet, small-scale-aware Siamese network; SVM, support vector machine; WLI, white light imaging.

The data statistics contain the training and test data set. The results are the prediction performances on the test cohorts. If the data are given on a patient or image level, the results show patient-level or image-level prediction, respectively. If not explicitly stated, the data are single-center. For the studies that contain multiple contrast experiments, the results show the best performances of the proposed methods.

experiment to confirm the effectiveness of deep learning in GC detection during actual clinical practice.

Furthermore, researchers have tried various methods to improve the accuracy of detection, including advanced endoscopy techniques, and innovative model designs.

Advanced endoscopy technique

Magnifying endoscopy with narrow-band imaging (ME-NBI) has a higher detection rate for EGC than the conventional WLI endoscopic images. Hu *et al.* [49] developed a computer-aided diagnostic model to assist GC detection on ME-NBI images. VGG-19 was trained on 1,777 images and achieved comparable performance with experienced endoscopists. Given that applying the deep-learning model on medical images is always combined with the transfer-learning method by fine-tuning a pre-trained model, Liu *et al.* [43] explored the effect of transfer learning on the application of CNNs in the classification of GC in ME-NBI images. They found that CNNs performed better on coarse data sets than on fine data sets without noise, while Inception-v3 outperformed VGG-16 and Inception-ResNet-v2. The unfrozen fine-tuning method obtained the best performance, indicating that low-level and high-level features of natural images might not be suitable for gastric endoscopic images. Besides, all CNN models showed significantly better performance than the traditional ML method with handcraft features. This study can provide a guide for applying deep learning via transfer learning in GC detection.

Model design

The above studies utilized the state-of-the-art deep-learning model pre-trained on a natural imaging data set directly. Several researchers tried to seek the specific redesigns of the deep-learning algorithm for better detection of GC. As fewer medical images can be obtained compared with natural images, Hsu *et al.* [45] came up with a small-scale-aware Siamese network (SSSNet) for GC detection on the few-shot data set. A Siamese network using a modified DenseNet as the backbone network was trained to learn the discriminative features between cancer and normal images, combined with a small CNN classifier to facilitate end-to-end training. In the experiments on images of typical cases and complicated cases containing misleading inflamed cases, the proposed SSSNet outperformed the state-of-the-art CNN models and Siamese network, demonstrating its effectiveness and efficiency of GC detection on a relatively small data set. Nguyen *et al.* [48] classified endoscopic images into pathological sites and non-pathological sites by the ensemble of multiple deep-learning models. Different models such as VGG-based, Inception-based, and DenseNet-based networks with various ensemble rules like MAX, AVERAGE, and VOTING were applied on a publicly available endoscopic data set from the Hamlyn center for robotic surgery [61]. The results showed the power to improve classification by an ensemble of deep-learning models. In order to guide the deep-learning model to learn the specific feature of GC rather than general gastric tissue, Yoon *et al.* [46] came up with a novel lesion-based CNN for detecting EGC by simultaneously minimizing classification error and localization error (between real lesion masks and activated Grad-CAM [62]). Experiments on 11,539 endoscopic images resulted in an AUC of 0.981 for EGC detection, which validated the efficacy of the proposed lesion-based model.

Automatic pathological confirmation

The definite diagnosis of GC relies on visual examination on WSI pathological images. Considering that the WSI image is usually too large for computers to process directly, the common practice is to crop the slice into multiple patches first, make patch classification, and finally provide the slice-level prediction based on the patch-level prediction. The challenge of WSI detection lies in extracting the most representative patch-level feature and aggregating multiple patch results to the overall prediction efficiently. Li *et al.* [50] proposed a model called GastricNet for GC detection on WSI. The model combines shallow layers that extract multi-scale features via dilated convolutional with deep layers consisting of small convolution kernels for feature fusion. Considering that even a gastric slice contains multiple normal patches, the slice-based classification is calculated by averaging the prediction score of 10 patches with the highest possibility of cancer. Experiments on the public data set achieved better performance than all the other state-of-the-art CNN networks, with an accuracy of 100% for slice-level prediction. Although the result is promising, it used a relatively brutal way to combine the patch-level prediction and lacked wide validations. Wang *et al.* [53] came up with a recalibrated multi-instance deep-learning method (RMDL) to detect GC. They also applied a two-stage framework: first, they used a ResNet-based fully conventional network to generate the discriminative feature and abnormal probability for each patch instance; second, they utilized a local–global feature fusion manner and an attention-based method to aggregate the instance feature. They validated the efficacy of the proposed RMDL method on a collected data set to classify gastric slices into cancer, dysplasia, and normal state, which resulted in the best performance among all the multi-instance-learning algorithms.

The above studies focus on the classification of patches, while even a tiny patch can contain both cancer and normal regions, resulting in confusing predictions. Li *et al.* [51] designed a deep-learning model, namely GT-Net, for the segmentation of gastric slices, where segmentation equals classification for each pixel. They continued with the multi-scale module [50] to extract features, combined with a feature pyramid by concatenating features from different sizes of average pooling, and finally used a dense up-sampling convolutional module to generate the final segmentation prediction. The performance of GT-Net outperformed other state-of-the-art segmentation models like FCN-8s and U-net. Meanwhile, they validated the improvements of different modules by ablation analysis. Furthermore, Sun *et al.* [52] proposed an encoder–decoder deep-learning model for accurate segmentation of gastric tumors in digital pathology images. Considering the non-grid morphology of a cancerous region and multi-scale context feature, dilated convolution and deformable convolution were applied to partially replace the original convolution operation in ResNet as the encoder, followed by a dense up-sampling module, which contained identity mapping and a depth-wise separable operation for feature fusion as the decoder. They achieved an accuracy of 91.6% and a mean Intersection over Union (mIoU) of 82.65% for tumor segmentation.

AI-assisted diagnosis

After detecting GC in the endoscopic examination, the following crucial step is classifying it into different stages and subtypes to assist treatment selection. Table 3 shows some applications of AI in classifying GC. Some researchers focused on estimating

Table 3. The application of AI in the diagnosis of gastric cancer

Authors	Aim	Data	Method	Result
Kubota *et al.* (2012) [63]	Estimate tumor invasion depth	344 patients; WLI endoscopy	Back propagation	Acc: 77.2% (T1), 49.1% (T2), 51% (T3), 55.3% (T4)
Zhu *et al.* (2019) [64]	Estimate tumor invasion depth	993 patients; WLI endoscopy	ResNet	AUC: 0.94, Acc: 89.16%, Sen: 76.47%, Spe: 95.56%, PPV: 89.66%, NPV: 88.97% (T1a/T1b vs deeper than T1b)
Yoon *et al.* (2019) [46]	Estimate tumor invasion depth	11,539 images; WLI endoscopy	Lesion-based VGG	AUC: 0.851 (T1a vs T1b)
Nagao *et al.* (2020) [65]	Estimate tumor invasion depth	1,084 patients; WLI, NBI, Indigo endoscopy	SSD	Acc: 94.5% (WLI), 94.3% (NBI), 95.5% (Indigo)
Wang *et al.* (2019) [66]	Estimate tumor invasion depth	244 patients; CT	Radiomics	AUC: 0.899 (train), 0.825 (test) (T2 vs T3/4)
Sun *et al.* (2020) [67]	Estimate tumor invasion depth	572 patients; CT	Deep-learning radiomics	AUC: 0.87 (test1), 0.90 (test2) (T4)
Dong *et al.* (2020) [68]	Predict lymph-node metastasis	730 patients; CT, multicenter	Deep-learning radiomics	C-index: 0.797 (external), 0.822 (international)
Li *et al.* (2020) [69]	Predict lymph-node metastasis	204 patients; Dual-energy CT	Deep-learning radiomics	AUC: 0.82
Jin *et al.* (2020) [70]	Predict lymph-node metastasis	1,699 patients; CT	ResNet-18	Median AUC: 0.876
Dong *et al.* (2019) [71]	Identify occult peritoneal metastasis	554 patients; four centers; CT	Radiomics	AUC: 0.928–0.920
Huang *et al.* (2020) [72]	Identify occult peritoneal metastasis	544 patients; CT	CNN	AUC: 0.900, Sen: 81.0%, Spe: 87.5%
Jiang *et al.* (2021) [73]	Identify occult peritoneal metastasis	1,978 patients; three centers; CT	PMetNet	AUC: 0.920–0.946; Sen: 75.4%–87.5%, Spe: 92.9%–98.2%
Sharma *et al.* (2016) [74]	Subtype of Her2	11 slices; H&E WSI	Graph-based model	Acc: 58.47%
Sharma *et al.* (2017) [75]	Subtype of Her2	11 slices; H&E WSI	9-layer CNN	Acc: 69.90%
Chen *et al.* (2021) [76]	Immune subtype	808 patients; H&E WSI	ResNet-18	Acc: 80.39% (validation), 76.47% (test)
Kather *et al.* (2019) [77]	Subtype of MSI	1,616 patients; H&E WSI; multicenter	ResNet-18	Acc: 84% (TCGA-CRC-DX), 77% (TCGA-CRC-KR), 84% (DACHS), 69% (KCCH)
Valieris *et al.* (2020) [78]	Subtype of MSI	1,616 patients; H&E WSI	CNN+RNN	Acc: 81%
Lai *et al.* (2019) [79]	Radiogenomics; subtype of CIN	58 patients; CT	Radiomics	AUC: 0.89, Acc: 88.9%, Spe: 88.9%, Sen: 88.9%

Acc, accuracy; AUC, area under the receiver-operating characteristic curve; CIN, chromosomal instability; CNN, convolutional neural network; CT, computed tomography; Her2, Human epidermal growth factor receptor 2; H&E WSI, hematoxylin and eosin-stained whole slide imaging; Indigo, Indigo–carmine dye contrast imaging; MSI, microsatellite instability; NBI, narrow-band imaging; NPV, negative predictive value; PPV, positive predictive value; RNN, recurrent convolutional neural network; Sen, sensitivity; Spe, specificity; SSD, single-shot multi-box detector; WLI, white light imaging.
For estimating tumor invasion depth, the result is given by different classification criteria. Besides, if the test data set contains multiple sources, the result is given on each separate data set. If not explicitly mentioned, the data are single-center, and the result is given an overall performance in all the validation and test data sets.

tumor invasion depth on endoscopic images [46, 63–65]. Results suggested that AI could achieve a relatively high T-staging accuracy for EGC, which might provide suggestions for endoscopic resection. However, its performance is poor for AGC. CT is used more for the diagnosis of patients with AGC. Researchers used AI to estimate the invasion depth [66, 67], predict the lymph-node metastasis (LNM) [68–70], while detecting occult peritoneal metastasis for gastric tumors in advanced stages [71–73]. The research mentioned above focused on TNM staging for gastric tumors. Meanwhile, other biological factors are also found to be related to the prognosis. Researchers classified the histopathological images based on Human epidermal growth factor receptor 2 (Her2) [74, 75], microsatellite instability (MSI) [76], and immune biomarkers [77, 78]. In addition, researchers explored the radiogenomics method to identify the chromosomal instability (CIN) state based on CT images [79]. Overall, these studies demonstrated the power of AI in classifying GC, which can assist in further treatment selection and prognosis prediction.

Invasion depth estimation via endoscopic images

Tumor invasion depth is a significant diagnostic factor for GC, where shallow invasion (within the mucosa or the submucosa) allows minimally invasive treatment such as endoscopic resection to avoid open surgery [80]. However, accurate estimation of invasion depth is difficult through visual examination due to subtle morphological disparities and subjective judgments, while a definite diagnosis cannot be made until invasive histopathological examinations are conducted. There exists no commonly accepted consensus on the best procedure for invasion depth estimation. Endoscopic ultrasonography was once

considered a preferable technique with more objective morphological evidence; however, it was reported to be not superior to conventional endoscopy with ~70% overall accuracy [81]. Therefore, there exists an unmet need for accurate methods to estimate tumor invasion depth in endoscopy.

Kubota *et al.* [63] applied the back propagation algorithm to train multi-layer perception to predict invasion depth of GC. The diagnosis accuracy for tumors in T1, T2, T3, and T4 were 77.2%, 49.1%, 51%, and 55.3%, specifically. The accuracy achieved 68.9% in T1a staging and 63.6% in T1b staging, which demonstrated similar performance with expert endoscopists, although its prediction for advanced lesions should be improved with further research. Considering that only patients in T1a or T1b are recognized as suitable for endoscopic resection, Zhu *et al.* [64] developed a system to classify the lesion as P0 and P1, defined as tumor invasion depth of T1a or T1b and deeper than T1b, respectively. ResNet was trained on 790 images and tested on 203 images, which achieved an AUC score of 0.94 that had significantly higher sensitivity and specificity than endoscopists. They suggested a relatively higher threshold for classification to avoid overdiagnosis, considering that overdiagnosis is more unacceptable than underdiagnosis for the irreversible consequences caused by over-treatment. To predict the invasion depth for EGC, Yoon *et al.* [46] applied the lesion-based VGG to classify T1a-EGC and T1b-EGC, achieving an AUC of 0.851. Furthermore, they found that incorrect prediction was significantly associated with undifferentiated-type histology and T1b staging, consistently with previous findings in endoscopic ultrasonography [81]. Since previous studies focused on the depth diagnosis on conventional WLI endoscopic images, Nagao *et al.* [65] extended exploration to multiple modalities containing NBI and Indigo–carmine dye contrast imaging (Indigo). Three independent ResNet-50-based models were trained individually, reaching an AUC of 94.5%, 94.3%, and 95.5% for WLI, NBI, and Indigo images, respectively. This work demonstrated that automatic invasion depth prediction for GC can be applied to endoscopic images of various modalities.

TNM staging on CT images

Accurate staging is a crucial step in determining the extent of tumor invasion for GC. According to the American Joint Committee on Cancer (AJCC) [82], TNM staging is accessed based on the primary tumor wall invasion depth (T), lymph-node spread (N), and the presence of metastasis (M). CT is a routine preoperative diagnostic modality for patients with GC. The CT accuracy of T staging varied from 77.8% to 93.5% [83, 84]. To discriminate T1/2 from T3/4 stage tumors, Wang *et al.* [66] developed an arterial phase-based radiomics model for the prediction of tumor invasion depth in GC. The model exhibited an AUC of 0.899 in the training set and 0.825 in the test set. For T4a GC staging, the reported diagnosis accuracy by radiologists was only 76.6% [85]. To improve the diagnosis of T4a GC, Sun *et al.* [67] applied the CT-based deep-learning radiomics analysis on 572 GC patients. The deep-learning radiomics nomogram showed accurately discriminating serosa invasion with AUCs of 0.90, 0.87, and 0.90 on the training set and two test sets, respectively. The study demonstrated that AI could assist T staging for patients with AGC.

LNM is categorized into N0 (no LNM), N1 (1–2 LNMs), N2 (3–6 LNMs), N3a (7–15 LNMs), and N3b (>15 LNMs) [82]. CT is applied routinely for preoperative N staging. However, its prediction by visual examination of radiologists is not very convincing, with an accuracy of 50%–70% [86]. Dong *et al.* [68] built a deep-learning radiomic nomogram based on preoperative CT images to predict the number of LNMs in local AGC, which showed considerable discrimination power with AUCs of 0.821, 0.797, and 0.822 in the primary, external validation, and international validation data sets, respectively. Furthermore, Li *et al.* [69] applied dual-energy CT-based deep-learning radiomics to predict LNM for GC. Deep-learning and radiomics features were combined to build the radiomics signature, which achieved an AUC of 0.82 in the test set. The previous work focuses on the overall evaluation of LNM, while multiple lymph-node stations make up the complex lymphatic network in the stomach. Therefore, Jin *et al.* [70] used 11 isomorphic sub-networks based on ResNet-18 to predict station-level LNMs for 11 lymph-node stations and applied transfer learning to address the problem of limited samples for some stations. Experiments on the training cohort of 1,172 patients and the external validation cohort of 527 patients demonstrated powerful prediction ability on all lymph-node stations, with a median AUC of 0.876. Furthermore, they applied the Grad-CAM to visualize the activated region, finding that the network responds to intra-tumor heterogeneity and the invasive margin. Combining the clinical factors did not significantly improve the prediction result, suggesting that LNM risk prediction largely depends on the primary tumor.

Peritoneal metastasis (PM) is one of the most frequent stage IV manifestations of GC. According to the guidelines of the European Society for Medical Oncology [87] and the National Comprehensive Cancer Network [88], PM-positive patients are not suggested to perform surgery since the presence of PM precludes the possibility of R0 resection. Therefore, it is necessary to develop a reliable preoperative method to identify patients with PM. Laparoscopy is recommended as a reliable preoperative method to identify PM but requires an invasive operation and extra costs. CT is the most common non-invasive method to detect PM. However, it suffers from low sensitivity in the case of occult PM, which is usually difficult to be recognized by human eyes on CT. PM-positive, which is missed by CT radiologists and revised by following laparoscopy or surgery, is defined as occult PM [89]. Researchers are exploring a non-invasive and accurate way to detect occult PM preoperatively. Dong *et al.* [71] identified occult PM in 554 AGC patients from four centers. They built radiomic signatures of the primary tumor (RS1) and peritoneum (RS2) based on 266 imaging features. RS1, RS2, and the Lauren types were combined to predict the occult PM with AUCs of 0.941, 0.928, and 0.920 for the training, validation, and test data sets. While previous studies have focused on the peritoneal region to identify PM, this study found that occult PM was associated with imaging features of both the primary tumor and the peritoneum, which also provided imaging evidence for the “soil-and-seed” theory. Huang *et al.* [72] applied a CNN model to predict the occult PM in 544 patients. The CNN classifier achieved an AUC of 0.900, a sensitivity of 81.0%, and a specificity of 87.5% using the Youden index as the cut-off value outperforming the clinical model with an AUC of 0.670. In addition, Jiang *et al.* [73] proposed in the Peritoneal Metastasis Network (PMetNet) to predict occult PM based on preoperative CT images. The model is composed of a densely connected CNN with long–short connections. They trained and validated the proposed PMetNet in three centers, achieving an AUC of 0.920–0.946, sensitivity of 75.4%–87.5%, and specificity of 92.9%–98.2%. They further utilized the Grad-CAM to visualize the activated region and found that intratumoral heterogeneity might significantly detect PM.

Histopathological and molecular subtyping

Classifying the subtype of tumor is crucial for determining the treatment decision for GC. According to the Cancer Genome Atlas (TGGA), GC is classified into four subtypes based on the molecular characteristics: Epstein-Barr virus-positive, MSI, CIN, and genomically stable tumors, where CIN subtype accounts for nearly half of all GC cases [90]. Traditional detection of the CIN state requires invasive biopsy specimens and complex genomic analyses. Recent developments in radiogenomics provide a non-invasive alternative way to detect the molecular subtype of the tumor by building associations between imaging features, clinical findings, and molecular phenotypes [91]. Lai *et al.* [79] applied radiogenomics analysis to predict the CIN status based on CT images. An acute tumor transition angle was found to be a significant predictive imaging feature with a classification AUC of 0.89. This study revealed the potential to non-invasively identify tumor molecular subtypes with the assistance of AI.

Her2, which has recently been seen as a significant predictive biomarker for GC [92], is usually accessed by time-consuming and tedious visual examinations on the histopathological WSI images. The Her2 immunohistochemical stain is limited by less common practice and high costs, while routinely used hematoxylin and eosin (H&E) stain is difficult for pathologists to identify Her2 status [93]. To overcome the current dilemma, Sharma *et al.* [74] proposed a graph-based method to classify the Her2 status of gastric tumors in H&E-stained histology slides. The cell nucleus was segmented to build a nuclei-attributed relational graph, which was used to extract global features and classify them into three types (non-tumor, Her2-positive, and Her2-negative) by ensemble learning. Experimental results on 11 WSI images from a representative case showed that the proposed method outperformed other machine-learning methods using multiple texture features with an overall accuracy of 0.5847. They extended the study by utilizing a deep-learning model for GC classification [75]. A self-designed nine-layer CNN reported an accuracy of 0.699 for cancer classification outperforming all the traditional machine-learning methods, which demonstrated the effectiveness of deep learning in identifying the Her2 state of a gastric tumor.

Chen *et al.* [76] explored the tumor immune microenvironment in gastric tumors. They applied unsupervised consensus clustering to identity three immune subtypes with different immune scores and prognoses, which differ in the components of tumor-infiltrating immune cells and molecular features. The immune subtyping was validated on two GC data sets and six pan-tumor data sets. Finally, they utilized ResNet-18 to predict the immune subtypes in pathological images with an accuracy of 85.71%, 80.39%, and 76.47% in the training, validation, and test cohorts, separately. This work revealed the potential for machine learning and deep learning in exploring tumor immune microenvironments and immune subtypes to assist with immunotherapeutic strategies for GC.

In addition, MSI is crucial for determining the patient response for immunotherapy in gastrointestinal cancer [94]. However, due to technical restrictions, MSI identification by immunohistochemistry or genetic analyses is not available for all patients. Accordingly, Kather *et al.* [77] applied the deep residual network to predict MSI status from H&E images. The deep-learning model was trained and validated elaborately on multiple data sets from various regions. The patient-level AUCs for four gastrointestinal test data sets, TCGA-CRC-DX, TCGA-CRC-KR (from the USA) [95], DACHS (from Germany) [96], and KCCH (from Asia, Japan) [97], were 0.84, 0.77, 0.84, and 0.69, respectively, where the poor performance in the last data set might be caused by different histology of GC between Asian and non-Asian patients. Since MSI is a pan-tumor factor, they also validated the effectiveness of the proposed method on the endometrial cancer data set, yielding an AUC of 0.75. Besides, MSI prediction correlated with a lymphocyte gene expression signature, programmed death-ligand 1 expression, and an interferon-γ signature, which overlapped with a poorly differentiated and lymphocyte-rich region. These findings are consistent with histopathological knowledge and can reversely inspire research in understanding the molecular mechanisms of GC. Furthermore, Valieris *et al.* [78] proposed a deep-learning method to detect the DNA repair deficiency [98] in pathology images, which contains mismatch repair deficiency (dMMR) leading to MSI in GC. The network is composed of a CNN classifier to predict the probability of dMMR for cropped tiles and a recurrent neural network aggregator to establish the slice-level prediction based on the top-k tiles. Since dMMR is common in different cancers, the proposed method was first validated on two independent breast cancer data sets with AUCs of 0.80 and 0.70, then performed a 0.81 AUC for detecting dMMR in GC. This research demonstrated the power of deep models in identifying the DNA repair deficiency from pathology images, revealing the generality of the genomic molecule and histological features in multiple tumors.

AI-aided treatment decision

For EGC, resection is recommended as curative therapy, while adjuvant chemotherapy and targeted molecular therapy have been recommended for patients with AGC. Adjuvant immunotherapy has also been introduced into perioperative treatment protocols. Table 4 summarizes some of the applications of AI in the treatment of GC. In several studies [99–102], researchers explored the application of AI methods in resection surgery, chemotherapy, and molecular drug selection, while in other studies [103–107], they predicted the prognosis of treatment with clinicopathologic features, CT, immunohistochemical stain, and lymph-node WSIs. These applications demonstrated the potential for AI in different practices of GC treatment.

Resection margin delineation

Endoscopic resection (ER) is recommended as the standard treatment for patients with EGC for its minimal invasiveness, curative resection, and better prognosis [5]. Precise delineation of the tumor margin is the first step in determining the treatment strategy for ER. An *et al.* [99] proposed a deep-learning model to delineate the resection margin under chromoendoscopy (CE) and WLI endoscopy. The proposed model achieved similar performances using CE and WLI with an accuracy of 85.7% and 88.9%, respectively, on 1,244 images from 536 enrolled patients. They also validated this in an endoscopic submucosal dissection (ESD) video, outperforming the resection margin marked on ME-NBI images. In addition, Ling *et al.* [100] developed a deep-learning-based system to delineate the margins of EGC in ME-NBI images. They reported an accuracy of 82.7% in differentiated EGC and 88.1% in undifferentiated EGG. Besides, the system achieved superb performance in a real ESD video compared with endoscopic experts.

Table 4. The application of AI in the treatment decision of gastric cancer

Authors	Aim	Data	Method	Result
An *et al.* (2020) [99]	Delineate resection margin for EGC	1,244 images and ESD videos	UNet++	IoU: 67.6% (image), 70.4% (video); Sen: 81.7% (image), 89.5% (video)
Ling *et al.* (2020) [100]	Delineate resection margin for EGC	1,670 images and ESD videos	UNet++	Acc: 82.7% (differentiated), 88.1% (undifferentiated)
Tan *et al.* (2020) [101]	Predict chemotherapy response	116 patients	Delta radiomics	Acc: 0.728–0.828
Joo *et al.* (2019) [102]	Predict molecular drug response	GDSC, CCLE, TGGA dataset	DeepIC50	–
Hyung *et al.* (2017) [103]	Prognosis prediction	1,549 patients; clinicopathologic factors	Five-layer neural network	AUC: 0.844–0.852 (five-year survival)
Zhang *et al.* (2020) [104]	Prognosis prediction	640 patients; CT	ResNet	C-index: 0.78 (OS)
Jiang *et al.* (2020) [105]	Prognosis prediction	1,615 patients; CT	S-net	C-index: 0.719 (DFS), 0.724 (OS)
Meier *et al.* (2020) [106]	Prognosis prediction	248 patients; IHC-stained TMAs	GoogLeNet	Hazard ratio: 1.273 (Cox), 1.234 (Uno), 1.149 (Logrank)
Wang *et al.* (2021) [107]	Prognosis prediction	1,164 patients; lymph-node pathological images	U-net, ResNet	Hazard ratio: 2.04 (univariable), C-index: 0.694

Acc, accuracy; AUC, area under the receiver-operating characteristic curve; C-index, concordance index; CT, computed tomography; DFS, disease-free survival; EGC, early gastric cancer; ESD, endoscopic submucosal dissection; IHC, immunohistochemistry; IoU, Intersection over Union; OS, overall survival; Sen, sensitivity; Spe, specificity; TMA, tissue microarray.

Treatment response prediction

Most GC patients are diagnosed at an advanced stage when radical resection is not available. Systemic chemotherapy is suggested to prolong survival, while individuals have various tumor responses for single drugs and combinations [108]. Tan *et al.* [101] built a delta radiomics model to predict the chemotherapeutic response for far-AGC patients, where delta radiomics is defined as the changes of radiomic features in the dynamic process of cancer therapy. Furthermore, they applied a V-net-based model to perform semi-automatic segmentation to replace tedious manual segmentation. The proposed model reported AUCs of 0.728 and 0.828 in the testing and external validation data sets, with no significant accuracy variance in the two chemotherapy regimens. Recently, with a deeper understanding of the molecular basis of the tumor, molecular target therapies have been developed for specific subtypes of tumor. As mentioned before, histopathological examination or genomic analysis is conducted preoperatively to classify GC into different subtypes, assisting with molecular target therapy selection. Joo *et al.* [102] proposed a deep-learning model to predict the drug response using its half-maximal inhibitory concentration (IC50) as the indicator. Genomics profiles of cancer cells and molecular features of drugs are concatenated as a 1D input vector. The model is trained and validated on the Genomics of Drug Sensitivity in Cancer (GDSC) data set [109] and the Cancer Cell Line Encyclopedia (CCLE) [110]. Furthermore, they validated the model efficacy to predict the patients' responses to drugs in the TGGA data set. The proposed model showed better performance than all the baseline models.

Survival analysis

The prognosis evaluates tumor malignancy and predicts patient survival. TNM staging is an important prognosis factor for GC. However, it is limited since patients in the same stage could have various survival times. Cox regression is the conventional model for survival analysis. Hyung *et al.* [103] proposed the 5-year survival prediction model with a five-layer neural network using eight characteristics (age, sex, histology, depth of tumor, number of metastatic and examined lymph nodes, presence of distant metastasis, and resection extent). The proposed model outperformed the Cox regression model with an accuracy of 83.5%. Zhang *et al.* [104] proposed a ResNet-based model for predicting the overall survival of AGC patients. The proposed model achieved the best performance with a concordance index (C-index) of 0.78 compared with 0.71 (radiomics model) and 0.72 (clinical model) in the external validation dataset. In addition, Jiang *et al.* [105] built a deep-learning-based imaging signature (DeLIS) to predict the disease-free survival and overall survival for GC patients based on preoperative CT images. The proposed DeLIS showed an independent prognostic value from traditional clinicopathologic factors in a multicenter study. By combining imaging signatures and clinical factors, the integrated model improved the performance with a C-index of 0.792–0.802 and net reclassification improvement of 10.1%–28.3%. Besides, they found that patients with higher DeLIS values could benefit more from adjuvant chemotherapy.

Researchers also explored AI methods for prognosis in digital pathology. Traditionally, pathologists count positive cells in selected views and classify them into different grades for prognosis. However, its accuracy suffered from subjectivity and inter-observer variability [111]. Meier *et al.* [106] developed a hypothesis-free deep-learning model to predict risk based on immunohistochemistry-stained tissue microarrays. To adapt to the time-to-event feature of survival data, loss functions (Cox loss, Uno loss, Logrank loss) were applied on the modified GoogLeNet. They also explored the tumor microenvironment by incorporating a panel of immune cell markers (CD8, CD20, CD68) and a proliferation marker (Ki67). They found that the

derived immune-related CNN score can provide additional evidence for prognosis prediction through qualitative analysis. In addition, N staging is a significant prognostic factor, while its precise assessment requires tedious examination and counting of metastatic lymph nodes (MLNs) by pathologists. Wang *et al.* [107] proposed a deep-learning framework to analyse the lymph-node WSI and developed an independent prognostic factor, the tumor-area-to-MLN-area ratio (T/MLN). The framework is composed of U-net for segmentation, ResNet for classification, and T/MLN for calculation. Experiments in multicenter data sets demonstrated that T/MLN is an independent prognostic value with a hazard ratio of 2.05 and an improved C-index of 0.694 over 0.646 using only N stages.

Challenges and future direction

Data scarcity

Most studies are conducted on a single-center, small data set. Although the performance of research efforts looks excellent, they were only validated on samples of small size, which might result in limited and less representative data sets and overfitting AI algorithms. Therefore, more attention has to be paid to the research design on how to validate the performance of AI algorithms to make more convincing research.

Considering the demographic differences and varied annotation criteria, experiments between small, single-center data sets are often challenging to evaluate and compare. Also, the state-of-the-art deep-learning algorithms were trained on massive natural images, while medical images are always limited to a single site, compared with natural images. Therefore, there is a need to collect more cross-regional, multicenter, and larger data sets to validate the ability of AI. Data from different sites often have inevitable distribution differences caused by varied scanning protocols, imaging devices, population geography, etc. It brings obstacles for training algorithms on multicenter data since the basic assumption of machine learning is that all data are from the same distribution. Therefore, it is necessary to formulate and promote unified scanning regulations for the stomach. Meanwhile, domain adaptation can be applied to mitigate data shift from multiple sources and improve model generalization. It has been explored for prostate segmentation [112] and colonoscopy malignant tissue detection [113], which could be explored in more applications for computer-assisted GC diagnosis.

Also, current approaches are mainly trained by supervised learning, which relies on finely human-labeled data. However, it is difficult to obtain a large amount of high-quality labeled data in medical tasks considering limited domain experts, while more unlabeled data remain unexplored. Recently, various self-supervised learning and semi-supervised learning methods have been proposed to exploit the neglected unlabeled data, which could be applied in AI in GC to address the problem of limited labeled data.

Interpretability

One concern of AI in GC is how to explain the predicted result to the doctors. Medical decisions usually require an interpretable solution. However, a common criticism of deep learning is poor interpretability due to its over-parameterized black-box nature. From this perspective, traditional handcrafted features designed by experts are favored as more understandable. Multiple interpretation tools have been proposed to explain how deep-learning models make decisions [114]. Also, current approaches focus on spotting the correlation between clinical data and diagnostic results. Causal inference [115] provides a new framework to explore causality rather than a simple correlation, which could guide AI prediction for medical purposes in a more convincing way [116].

Meanwhile, the development of automatic medical diagnosis algorithms relied on in-depth medical theories to connect the visual features of medical images with the disease. Therefore, the breakthrough of AI in GC depends not only on larger data sets and better algorithms, but also on the ability to build AI-aided GC screening, diagnosis, and treatment decisions with high correlation with solid medical theory.

Multi-modality and multi-task algorithm

Medical images of different modalities can play individual and complementary roles in GC early detection, diagnosis, and treatment decisions. Endoscopic examination is suitable for the early detection of GC, while its confirmation needs to be validated by pathologic examination. Also, it is not suitable for diagnosing AGC, which can be identified through CT scanning. Judgments based on a single modality might lead to misdiagnoses. Therefore, it is valuable to build multi-modality algorithms to make more comprehensive diagnostic decisions. Besides, large-scale imaging data can be relatively easily collected compared with costly molecular data. Radiogenomics has emerged as a new research direction, which aims to uncover the relationship between imaging phenotypes and gene expression patterns of a disease [91]. It has been widely explored in glioblastoma [117], breast cancer [118], and lung cancer [119], which could provide guidance in better treatment decisions, prognosis prediction, and understanding of cancer in general. Considering pan-cancer molecules exist in various types of tumors, cross-cancer researchers could facilitate deeper knowledge of tumor-related molecules. TGGA has come up with the Pan-cancer Atlas, which reclassifies human cancer types by molecular similarity [120]. Knowledge from similar types of cancer might be transferred to the research of GC.

Moreover, most of the existing work has focused on single modeling problems such as identifying precancerous diseases or tumors, classifying the stages or subtypes of the tumor, and predicting the treatment response or the survival states. However, these tasks are inherently correlated. Multi-task-learning algorithms could leverage helpful information by learning multiple related tasks simultaneously, which could help build a more robust GC diagnosis [121]. Also, the development of the tumor is a dynamic process over time. The changes of images over time might be essential in predicting the treatment response and prognosis. Therefore, it might be helpful to build time-related models based on imaging data from different stages of patients, such as first detection, pretreatment, post-treatment, etc. It could help us better understand the dynamic process of cancer development. Also, considering different modalities are applied in different stages of GC, the combinations of multi-modality and multi-task algorithms could result in better prediction power.

Model evaluation and clinical application

In the standard procedure for training AI models in medicine, the whole data set is split into the training and test data sets.

The model is trained on the training data set and its generalization validated in the test data set. To build a robust and available AI system for clinical application, the model should be tested on data sets at different institutions from those in the training data set. In order to facilitate training and testing data from multiple sites, traditional machine-learning and deep-learning methods require gathering data on a primary data center, which burdens data transmission and privacy protection. Federated learning is proposed to enable training on multicenter decentralized data to guarantee data privacy and security [122]. Also, rather than simply focusing on the comparisons of evaluation metrics, it might be more helpful to analyse the predicted results by AI and human experts in detail. What are the similarities and differences in the patterns detected by AI and human experts? How much could experts improve performance or save time with the assistance of AI? Elaborate case studies might build more clinically useful AI applications.

Conclusions

Studies have explored the applications of AI on multiple tasks in the diagnosis and treatment of GC. Figure 2 summarizes general uses for medical images from three different modalities. Here, we focus on endoscopy, CT, and pathology. Other modalities like positron emission tomography and magnetic resonance imaging can also be applied in GC diagnosis. In endoscopic examinations, AI can help detect precancerous lesions and early tumors, estimate tumor invasion depth, and delineate resection margins. In pathological diagnosis, AI can significantly reduce the time for diagnosis and play an auxiliary role in subtyping with WSI images. Based on CT images, AI can assist with TNM staging, treatment selection, and prognosis prediction for patients with AGC. These studies demonstrated the potential for AI in GC. Meanwhile, current approaches face similar challenges, such as data scarcity and poor interpretability, which could be improved by data regularization and advanced algorithms. Also, by developing a multi-modality, cross-modality

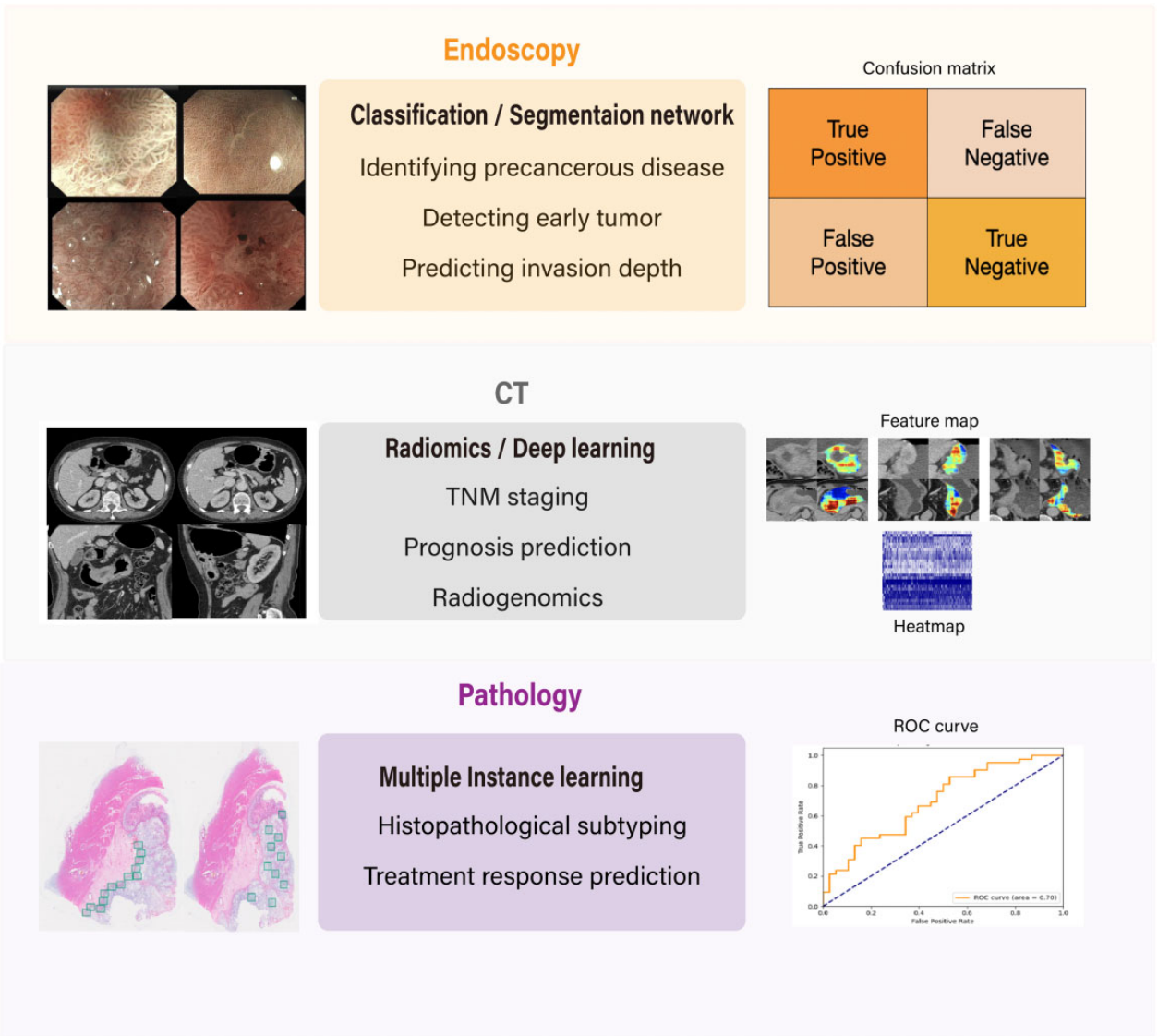

Figure 2. Applications of artificial intelligence in gastric cancer for images from endoscopy, computed tomography (CT) and pathology

algorithm and improving the procedures of model evaluation and clinical application, we hope to build more clinically useful AI applications for GC.

Authors' Contributions

R.C., L.T., D.D., and J.T. conceived and designed this review. R.C., M.F., L.Z., S.W., L.G., and J.L. did the investigation and research. R.C. and L.T. drafted the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 82022036, 91959130, 81971776, 62027901, 81930053], National Key R&D Program of China [grant number 2017YFA0205200], the Beijing Natural Science Foundation [grant number Z20J00105], Strategic Priority Research Program of Chinese Academy of Sciences [grant number XDB38040200], and the Youth Innovation Promotion Association CAS [grant number Y2021049].

Acknowledgements

The authors would like to acknowledge the instrumental and technical support of the Multi-modal biomedical imaging experimental platform, Institute of Automation, Chinese Academy of Sciences.

Conflict of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**: 209–49.
2. Ajani JA, Bentrem DJ, Besh S *et al.*; National Comprehensive Cancer Network. Gastric cancer, version 2.2013. *J Natl Compr Canc Netw* 2013;**11**:531–46.
3. 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**:301–16.
4. Veitch AM, Uedo N, Yao K *et al.* Optimizing early upper gastrointestinal cancer detection at endoscopy. *Nat Rev Gastroenterol Hepatol* 2015;**12**:660–7.
5. Pimentel-Nunes P, Dinis-Ribeiro M, Ponchon T *et al.* Endoscopic submucosal dissection: European society of gastrointestinal endoscopy (ESGE) guideline. *Endoscopy* 2015;**47**: 829–54.
6. Wagner AD, Syn NL, Moehler M *et al.* Chemotherapy for advanced gastric cancer. *Cochrane Database Syst Rev* 2017;**8**: CD004064.
7. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Ann Oncol* 2008;**19**: 1523–9.
8. Sauerbruch T, Schreiber M, Schüssler P *et al.* Endoscopy in the diagnosis of gastritis. *Endoscopy* 1984;**16**:101–4.
9. Watanabe K, Nagata N, Shimbo T *et al.* Accuracy of endoscopic diagnosis of Helicobacter pylori infection according to level of endoscopic experience and the effect of training. *BMC Gastroenterol* 2013;**13**:128.
10. Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. *Endosc Int Open* 2014;**2**:E46–50.
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;**278**:563–77.
12. Lambin P, Leijenaar RT, Deist TM *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;**14**:749–62.
13. Meng L, Dong D, Chen X *et al.* 2D and 3D CT radiomic features performance comparison in characterization of gastric cancer: a multi-center study. *IEEE J Biomed Health Inform* 2021;**25**:755–63.
14. Tian J, Dong D, Liu Z *et al. Radiomics and Its Clinical Application: Artificial Intelligence and Medical Big Data*. Pittsburgh, PA, USA: Academic Press, **2021**, 1–328.
15. Lazăr DC, Avram MF, Faur AC *et al.* The role of computer-assisted systems for upper-endoscopy quality monitoring and assessment of gastric lesions. *Gastroenterol Rep (Oxf)* 2021;**9**:185–204.
16. Deng J, Dong W, Socher R, *et al.*; Imagenet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, 248–55.
17. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 2012;**25**:1097–105.
18. Esteva A, Kuprel B, Novoa RA *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;**542**:115–8.
19. Ronneberger O, Fischer P, Brox T, U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, 234–41.
20. Correa P. A human model of gastric carcinogenesis. *Cancer Res* 1988;**48**:3554–60.
21. 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**:945–52.
22. Van Cutsem E, Sagaert X, Topal B *et al.* Gastric cancer. *Lancet Digital Health* 2016;**388**:2654–64.
23. Amin A, Gilmour H, Graham L *et al.* Gastric adenocarcinoma missed at endoscopy. *J R Coll Surg Edinb* 2002;**47**:681–4.
24. Yalamarthi S, Witherspoon P, McCole D *et al.* Missed diagnoses in patients with upper gastrointestinal cancers. *Endoscopy* 2004;**36**:874–9.
25. Voutilainen ME, Juhola MT. Evaluation of the diagnostic accuracy of gastroscopy to detect gastric tumours: clinicopathological features and prognosis of patients with gastric cancer missed on endoscopy. *Eur J Gastroenterol Hepatol* 2005; **17**:1345–9.
26. Hosokawa O, Hattori M, Douden K *et al.* Difference in accuracy between gastroscopy and colonoscopy for detection of cancer. *Hepatogastroenterology* 2007;**54**:442–4.
27. Diao W, Huang X, Shen L *et al.* Diagnostic ability of blue laser imaging combined with magnifying endoscopy for early esophageal cancer. *Dig Liver Dis* 2018;**50**:1035–40.
28. Robbins SL, Cotran RS. Pathologic Basis of Disease 2nd ed. Philadelphia, PA: Saunders, 1979.
29. Huang C-R, Sheu B-S, Chung P-C *et al.* Computerized diagnosis of Helicobacter pylori infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network. *Endoscopy* 2004;**36**:601–8.

30. 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**:106–11.
31. Nakashima H, Kawahira H, Kawachi H *et al*. Artificial intelligence diagnosis of Helicobacter pylori infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018;**31**:462–8.
32. Shichijo S, Endo Y, Aoyama K *et al*. Application of convolutional neural networks for evaluating Helicobacter pylori infection status on the basis of endoscopic images. *Scand J Gastroenterol* 2019;**54**:158–63.
33. Yasuda T, Hiroyasu T, Hiwa S *et al*. Potential of automatic diagnosis system with linked color imaging for diagnosis of Helicobacter pylori infection. *Dig Endosc* 2020;**32**:373–81.
34. Guimarães P, Keller A, Fehlmann T *et al*. Deep-learning based detection of gastric precancerous conditions. *Gut* 2020;**69**:4–6.
35. Zhang Y, Li F, Yuan F *et al*. Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence. *Dig Liver Dis* 2020;**52**:566–72.
36. Yan T, Wong PK, Choi IC *et al*. Intelligent diagnosis of gastric intestinal metaplasia based on convolutional neural network and limited number of endoscopic images. *Comput Biol Med* 2020;**126**:104026.
37. Zhang X, Hu W, Chen F *et al*. Gastric precancerous diseases classification using CNN with a concise model. *PloS One* 2017;**12**:e0185508.
38. Cho B-J, Bang CS, Park SW *et al*. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy* 2019;**51**:1121–9.
39. Lee JH, Kim YJ, Kim YW *et al*. Spotting malignancies from gastric endoscopic images using deep learning. *Surg Endosc* 2019;**33**:3790–7.
40. Miyaki R, Yoshida S, Tanaka S *et al*. Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement. *J Gastroenterol Hepatol* 2013;**28**:841–7.
41. Miyaki R, Yoshida S, Tanaka S *et al*. A computer system to be used with laser-based endoscopy for quantitative diagnosis of early gastric cancer. *J Clin Gastroenterol* 2015;**49**:108–15.
42. Hirasawa T, Aoyama K, Tanimoto T *et al*. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018;**21**:653–60.
43. Liu X, Wang C, Hu Y, *et al*.; Transfer learning with convolutional neural network for early gastric cancer classification on magnifiying narrow-band imaging images. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, 1388–92.
44. 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**:1645–54.
45. Hsu C-C, Ma H-T, Lee J-Y, SSSNet: Small-Scale-Aware Siamese Network for Gastric Cancer Detection. *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2019, 1–5.
46. Yoon HJ, Kim S, Kim J-H *et al*. A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. *JCM* 2019;**8**:1310.
47. 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**:141–50.
48. Nguyen DT, Lee MB, Pham TD *et al*. Enhanced image-based endoscopic pathological site classification using an ensemble of deep learning models. *Sensors* 2020;**20**:5982.
49. 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**:1333–41. e3.
50. Li Y, Li X, Xie X, *et al*.; Deep learning based gastric cancer identification. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, 182–5.
51. Li Y, Xie X, Liu S, *et al*.; Gt-net: a deep learning network for gastric tumor diagnosis. In: *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2018, 20–4.
52. Sun M, Zhang G, Dang H *et al*. Accurate gastric cancer segmentation in digital pathology images using deformable convolution and multi-scale embedding networks. *IEEE Access* 2019;**7**:75530–41.
53. 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.
54. Hooi JK, Lai WY, Ng WK *et al*. Global prevalence of Helicobacter pylori infection: systematic review and meta-analysis. *Gastroenterology* 2017;**153**:420–9.
55. Herrero R, Park JY, Forman D. The fight against gastric cancer—the IARC Working Group report. *Best Pract Res Clin Gastroenterol* 2014;**28**:1107–14.
56. Lee Y-C, Chiang T-H, Chou C-K *et al*. Association between Helicobacter pylori eradication and gastric cancer incidence: a systematic review and meta-analysis. *Gastroenterology* 2016;**150**:1113–24. e5.
57. Tsukamoto T, Nakagawa M, Kiriyama Y *et al*. Prevention of gastric cancer: eradication of Helicobacter pylori and beyond. *Ijms* 2017;**18**:1699.
58. Du Y, Bai Y, Xie P *et al*.; Chinese Chronic Gastritis Research Group. Chronic gastritis in China: a national multi-center survey. *BMC Gastroenterol* 2014;**14**:21–9.
59. Capelle LG, Haringsma J, de Vries AC *et al*. Narrow band imaging for the detection of gastric intestinal metaplasia and dysplasia during surveillance endoscopy. *Dig Dis Sci* 2010;**55**:3442–8.
60. Liu W, Anguelov D, Erhan D *et al*. Ssd: Single shot multibox detector. In: *European Conference on Computer Vision*, Springer, 2016, 21–37.
61. Ye M, Giannarou S, Meining A *et al*. Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations. *Med Image Anal* 2016;**30**:144–57.
62. 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*, 2017, 618–26.
63. Kubota K, Kuroda J, Yoshida M *et al*. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. *Surg Endosc* 2012;**26**:1485–9.
64. Zhu Y, Wang Q-C, Xu M-D *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**:806–15. e1.
65. Nagao S, Tsuji Y, Sakaguchi Y *et al*. Highly accurate artificial intelligence systems to predict the invasion depth of gastric cancer: efficacy of conventional white-light imaging,

nonmagnifying narrow-band imaging, and indigo-carmine dye contrast imaging. *Gastrointest Endosc* 2020;**92**:866–73. e1.
66. Wang Y, Liu W, Yu Y *et al*. Prediction of the depth of tumor invasion in gastric cancer: potential ROLE of CT radiomics. *Acad Radiol* 2020;**27**:1077–84.
67. Sun R-J, Fang M-J, Tang L *et al*. CT-based deep learning radiomics analysis for evaluation of serosa invasion in advanced gastric cancer. *Eur J Radiol* 2020;**132**:109277.
68. Dong D, Fang M-J, Tang L *et al*. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol* 2020;**31**:912–20.
69. Li J, Dong D, Fang M *et al*. Dual-energy CT–based deep learning radiomics can improve lymph node metastasis risk prediction for gastric cancer. *Eur Radiol* 2020;**30**:2324–33.
70. Jin C, Jiang Y, Yu H *et al*. Deep learning analysis of the primary tumour and the prediction of lymph node metastases in gastric cancer. *Br J Surg* 2021;**108**:542–9.
71. Dong D, Tang L, Li Z-Y *et al*. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol* 2019;**30**:431–8.
72. Huang Z, Liu D, Chen X *et al*. Deep convolutional neural network based on computed tomography images for the preoperative diagnosis of occult peritoneal metastasis in advanced gastric cancer. *Front Oncol* 2020;**10**:601869.
73. Jiang Y, Liang X, Wang W *et al*. Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning. *JAMA Netw Open* 2021;**4**:e2032269.
74. Sharma H, Zerbe N, Heim D, *et al*. Cell nuclei attributed relational graphs for efficient representation and classification of gastric cancer in digital histopathology. In: *Medical Imaging 2016: Digital Pathology*, International Society for Optics and Photonics, 2016, 97910X.
75. 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.
76. 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.
77. Kather JN, Pearson AT, Halama N *et al*. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* 2019;**25**:1054–6.
78. Valieris R, Amaro L, Osório C *et al*. Deep learning predicts underlying features on pathology images with therapeutic relevance for breast and gastric cancer. *Cancers* 2020;**12**:3687.
79. Lai Y-C, Yeh T-S, Wu R-C *et al*. Acute tumor transition angle on computed tomography predicts chromosomal instability status of primary gastric cancer: radiogenomics analysis from TCGA and independent validation. *Cancers* 2019;**11**:641.
80. Wang J, Yu J-C, Kang W-M *et al*. Treatment strategy for early gastric cancer. *Surg Oncol* 2012;**21**:119–23.
81. Choi J, Kim S, Im J *et al*. Comparison of endoscopic ultrasonography and conventional endoscopy for prediction of depth of tumor invasion in early gastric cancer. *Endoscopy* 2010;**42**:705–13.
82. Amin MB, Greene FL, Edge SB et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: A Cancer Journal for Clinicians* 2017;**67**:93–99.
83. Ba-Ssalamah A, Prokop M, Uffmann M *et al*. Dedicated multidetector CT of the stomach: spectrum of diseases. *Radiographics* 2003;**23**:625–44.
84. Kim JW, Shin SS, Heo SH *et al*. Diagnostic performance of 64-section CT using CT gastrography in preoperative T staging of gastric cancer according to 7th edition of AJCC cancer staging manual. *Eur Radiol* 2012;**22**:654–62.
85. Hwang SW, Lee DH, Lee SH *et al*. Preoperative staging of gastric cancer by endoscopic ultrasonography and multidetector-row computed tomography. *J Gastroenterol Hepatol* 2010;**25**:512–8.
86. Kim HJ, Kim AY, Oh ST *et al*. Gastric cancer staging at multidetector row CT gastrography: comparison of transverse and volumetric CT scanning. *Radiology* 2005;**236**:879–85.
87. Smyth E, Verheij M, Allum W *et al*.; ESMO Guidelines Committee. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2016;**27**:v38–v49.
88. Coit DG, Andtbacka R, Anker CJ *et al*.; National Comprehensive Cancer Network (NCCN). Melanoma, version 2.2013: featured updates to the NCCN guidelines. *J Natl Compr Canc Netw* 2013;**11**:395–407.
89. Burbidge S, Mahady K, Naik K. The role of CT and staging laparoscopy in the staging of gastric cancer. *Clin Radiol* 2013;**68**:251–5.
90. Bass AJ, Thorsson V, Shmulevich I *et al*. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;**513**:202.
91. Mazurowski MA. Radiogenomics: what it is and why it is important. *J Am Coll Radiol* 2015;**12**:862–6.
92. Behrens HM, Warneke VS, Böger C *et al*. Reproducibility of HER2/neu scoring in gastric cancer and assessment of the 10% cut-off rule. *Cancer Med* 2015;**4**:235–44.
93. Warneke V, Behrens H-M, Böger C *et al*. Her2/neu testing in gastric cancer: evaluating the risk of sampling errors. *Ann Oncol* 2013;**24**:725–33.
94. Le DT, Uram JN, Wang H *et al*. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 2015;**372**:2509–20.
95. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;**487**:330.
96. Hoffmeister M, Jansen L, Rudolph A *et al*. Statin use and survival after colorectal cancer: the importance of comprehensive confounder adjustment. *J Natl Cancer Inst* 2015;**107**:djv045.
97. Aoyama T, Hutchins G, Arai T *et al*. Identification of a high-risk subtype of intestinal-type Japanese gastric cancer by quantitative measurement of the luminal tumor proportion. *Cancer Med* 2018;**7**:4914–23.
98. Hakem R; DNA-damage repair; the good, the bad, and the ugly. *EMBO J* 2008;**27**:589–605.
99. An P, Yang D, Wang J *et al*. A deep learning method for delineating early gastric cancer resection margin under chromoendoscopy and white light endoscopy. *Gastric Cancer* 2020;**23**:884–92.
100. Ling T, Wu L, Fu Y *et al*. A deep learning-based system for identifying differentiation status and delineating margins of early gastric cancer in magnifying narrow-band imaging endoscopy. *Endoscopy* 2021;**53**:469–77. (AAM).
101. Tan J-W, Wang L, Chen Y *et al*. Predicting chemotherapeutic response for far-advanced gastric cancer by radiomics with deep learning semi-automatic segmentation. *J Cancer* 2020;**11**:7224–36.

102. Joo M, Park A, Kim K *et al*. A deep learning model for cell growth inhibition ic50 prediction and its application for gastric cancer patients. *Ijms* 2019;**20**:6276.
103. Hyung WJ, Son T, Park M *et al*. Superior prognosis prediction performance of deep learning for gastric cancer compared to Yonsei prognosis prediction model using Cox regression. *Journal of Clinical Oncology* 2017;35:164–164.
104. Zhang L, Dong D, Zhang W *et al*. A deep learning risk prediction model for overall survival in patients with gastric cancer: a multicenter study. *Radiother Oncol* 2020;**150**:73–80.
105. Jiang Y, Jin C, Yu H *et al*. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: a multicenter, retrospective study. *Ann Surg* 2021;**274**:e1153–61.
106. Meier A, Nekolla K, Hewitt LC *et al*. Hypothesis-free deep survival learning applied to the tumour microenvironment in gastric cancer. *J Pathol Clin Res* 2020;**6**:273–82.
107. 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**:1–13.
108. Wagner AD, Grothe W, Haerting J *et al*. Chemotherapy in advanced gastric cancer: a systematic review and meta-analysis based on aggregate data. *J Clin Oncol* 2006;**24**:2903–9.
109. Yang W, Soares J, Greninger P *et al*. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013;**41**: D955–D961.
110. Barretina J, Caponigro G, Stransky N *et al*. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7.
111. Rizzardi AE, Johnson AT, Vogel RI *et al*. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol* 2012;**7**:42– 10.
112. Liu Q, Dou Q, Yu L *et al*. MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. *IEEE Trans Med Imaging* 2020;**39**:2713–24.
113. Zhu C, Mei K, Peng T *et al*. Multi-level colonoscopy malignant tissue detection with adversarial CAC-UNet. *Neurocomputing* 2021;**438**:165–83.
114. Montavon G, Samek W, Müller K-R. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* 2018;**73**:1–15.
115. Pearl J. Causal inference in statistics: an overview. *Statist Surv* 2009;**3**:96–146.
116. Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* 2020;**11**:1–9.
117. Zinn PO, Majadan B, Sathyan P *et al*. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PloS One* 2011;**6**:e25451.
118. Yamamoto S, Maki DD, Korn RL *et al*. Radiogenomic analysis of breast cancer using MRI: a preliminary study to define the landscape. *Ajr Am J Roentgenol* 2012; **199**:654–63.
119. Gevaert O, Xu J, Hoang CD *et al*. Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology* 2012;**264**: 387–96.
120. Hoadley KA, Yau C, Hinoue T *et al*.; Cancer Genome Atlas Network. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 2018;**173**:291–304.e6. e6.
121. Zhang Y, Yang Q, A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* 2017.
122. Yang Q, Liu Y, Chen T *et al*. Federated machine learning: Concept and applications. *ACM Trans Intell Syst Technol* 2019; **10**:1–19.