



Current Evidence and Future Perspective of Accuracy of Artificial Intelligence Application for Early Gastric Cancer Diagnosis With Endoscopy: A Systematic and Meta-Analysis

Kailin Jiang¹, Xiaotao Jiang¹, Jinglin Pan², Yi Wen³, Yuanchen Huang¹, Senhui Weng¹, Shaoyang Lan³, Kechao Nie¹, Zhihua Zheng¹, Shuling Ji¹, Peng Liu¹, Peiwu Li^{3*} and Fengbin Liu^{3*}

OPEN ACCESS

Edited by:

Abhilash Periseti,
University of Arkansas for Medical
Sciences, United States

Reviewed by:

Mahesh Gajendran,
Texas Tech University Health Sciences
Center El Paso, United States
Rahul Shekhar,
University of New Mexico,
United States

*Correspondence:

Peiwu Li
doctorlipw@gzucm.edu.cn
Fengbin Liu
liufb163@vip.163.com

Specialty section:

This article was submitted to
Gastroenterology,
a section of the journal
Frontiers in Medicine

Received: 13 November 2020

Accepted: 20 January 2021

Published: 15 March 2021

Citation:

Jiang K, Jiang X, Pan J, Wen Y,
Huang Y, Weng S, Lan S, Nie K,
Zheng Z, Ji S, Liu P, Li P and Liu F
(2021) Current Evidence and Future
Perspective of Accuracy of Artificial
Intelligence Application for Early
Gastric Cancer Diagnosis With
Endoscopy: A Systematic and
Meta-Analysis. *Front. Med.* 8:629080.
doi: 10.3389/fmed.2021.629080

¹ First College of Clinic Medicine, Guangzhou University of Chinese Medicine, Guangzhou, China, ² Department of Spleen-Stomach and Liver Diseases, Traditional Chinese Medicine Hospital of Hainan Province Affiliated to Guangzhou University of Chinese Medicine, Haikou, China, ³ Department of Gastroenterology, First Affiliation Hospital, Guangzhou University of Chinese Medicine, Guangzhou, China

Background & Aims: Gastric cancer is the common malignancies from cancer worldwide. Endoscopy is currently the most effective method to detect early gastric cancer (EGC). However, endoscopy is not infallible and EGC can be missed during endoscopy. Artificial intelligence (AI)-assisted endoscopic diagnosis is a recent hot spot of research. We aimed to quantify the diagnostic value of AI-assisted endoscopy in diagnosing EGC.

Method: The PubMed, MEDLINE, Embase and the Cochrane Library Databases were searched for articles on AI-assisted endoscopy application in EGC diagnosis. The pooled sensitivity, specificity, and area under the curve (AUC) were calculated, and the endoscopists' diagnostic value was evaluated for comparison. The subgroup was set according to endoscopy modality, and number of training images. A funnel plot was delineated to estimate the publication bias.

Result: 16 studies were included in this study. We indicated that the application of AI in endoscopic detection of EGC achieved an AUC of 0.96 (95% CI, 0.94–0.97), a sensitivity of 86% (95% CI, 77–92%), and a specificity of 93% (95% CI, 89–96%). In AI-assisted EGC depth diagnosis, the AUC was 0.82(95% CI, 0.78–0.85), and the pooled sensitivity and specificity was 0.72(95% CI, 0.58–0.82) and 0.79(95% CI, 0.56–0.92). The funnel plot showed no publication bias.

Conclusion: The AI applications for EGC diagnosis seemed to be more accurate than the endoscopists. AI assisted EGC diagnosis was more accurate than experts. More prospective studies are needed to make AI-aided EGC diagnosis universal in clinical practice.

Keywords: artificial intelligence, machine learning, deep learning, early gastric cancer, endoscopy

INTRODUCTION

Gastric cancer is ranked as the third leading cause of death from cancer worldwide (1). Most gastric cancers are diagnosed at advanced stages because their symptoms and signs tend to be inconspicuous and non-specific, leading to an overall poor prognosis, whereas in the case of early detection, the 5-years survival rate can exceed 90% (2–4). Endoscopic examination is still considered the most effective method for EGC detection (5). However, early gastric cancer (EGC) is particularly difficult to identify since it usually exhibits a subtle elevation or depression with faint redness, which is likely recognized as normal mucosa or gastritis. In addition, the invasion depth within the gastric wall is also hard to predict. Ten studies involving 3,787 patients who received an upper gastrointestinal endoscopy examination revealed an 11.3% miss rate of upper gastrointestinal cancers up to 3 years before diagnosis (6). A meta-analysis involving 2,153 lesion images showed that the area under the receiver operating characteristic curve (AUC) for the diagnosis of EGC using white light imaging (WLI) endoscopy was only 0.48 (7).

In the past decade, the application of artificial intelligence (AI) in medicine has attracted extensive attention. AI-assisted endoscopic diagnosis is a hot spot of research. AI refers to the capacity of a computer to execute a task associated with intelligent beings, such as the “learn” function that mimics the cognitive ability of human beings (8). AI subfields contain machine learning and deep learning (Figure 1). Machine learning, a term originally created by Arthur Samuel in 1959, is a field of computer science, whereby a system is able to develop the ability to “learn” from the input data without a certain program (9). Common machine-learning methods in classification model training comprise ensemble trees, decision trees, support vector machines, k-nearest neighbors, etc. (10).

Deep learning, which was initially applied in the image processing field in 1998, refers to the application of layers in non-linear processing based on machine learning algorithms used for feature extraction and transformation (11). Neural networks, similar to the human brain, particularly mimic closely interconnected neurons to recognize patterns, extract features or “learn” things about the input data to predict a result (12). Different model training paradigms, such as scaled-conjugate gradient, Levenberg-Marquardt and Bayesian regularization, have been termed “neural networks” (13). Several computer aided detection (CAD) algorithms for automatic early gastric cancer detection have been recommended for images from standard endoscopes. The performance improvements of original image classification models mainly depend on visual features and large-scale datasets, which are difficult to implement in EGC

detection models. Although the invasion depth in EGC is defined differently, visual characteristics such as textures, colors, shapes, and regions are similar.

To date, the existing data on the diagnostic value of AI for EGC diagnosis are scattered. Jin et al. (14) reviewed the current studies on AI application for gastric cancer, while the definite diagnostic ability of AI application for EGC was still unclear. The aim of this study was to systematically summarize the recent available studies on the diagnostic accuracy of AI on EGC diagnosis to address the current status of this area and discuss future perspectives.

METHODS

Search Strategy and Study Selection

Electronic databases (PubMed, Web of Science, EMBASE, and the Cochrane Library) were searched from initiation to November 2020 using presupposed search terms. The following medical subject terms and keywords were used: “endoscopy,” “Endoscopic Diagnosis,” “early gastric cancer,” “artificial intelligence,” “computer-assisted diagnosis,” “Deep learning,” and “Convolutional neural network.” The full texts of potentially appropriate studies were then reviewed after the screenings of citations and abstracts exported from the electronic databases. The search strategy was shown as follows: (1) (artificial intelligence [Title/Abstract]) OR (computer-assisted diagnosis [Title/Abstract]) OR (Deep learning [Title/Abstract]) OR (Convolutional neural network [Title/Abstract]) (2) (endoscopy [Title/Abstract]) OR (Endoscopic Diagnosis [Title/Abstract]) OR (early gastric cancer [Title/Abstract]) (3) (1) AND (2).

Study Eligibility Criteria

The eligible studies fulfilled the following criteria: (1) the study was a diagnosis test about AI application in endoscopy for EGC diagnosis. Diagnosis test included AI detection of EGC from other gastric disease or distinguishment of invasion depth; (2) the absolute numbers of true-positive, false-negative, true-negative, and false-positive observations for EGC diagnosis were reported directly or were able to be calculated; (3) the study provided clear information about the database and number of images; (4) the study clearly described the CAD or CNN algorithms and the process applied in the EGC diagnosis.

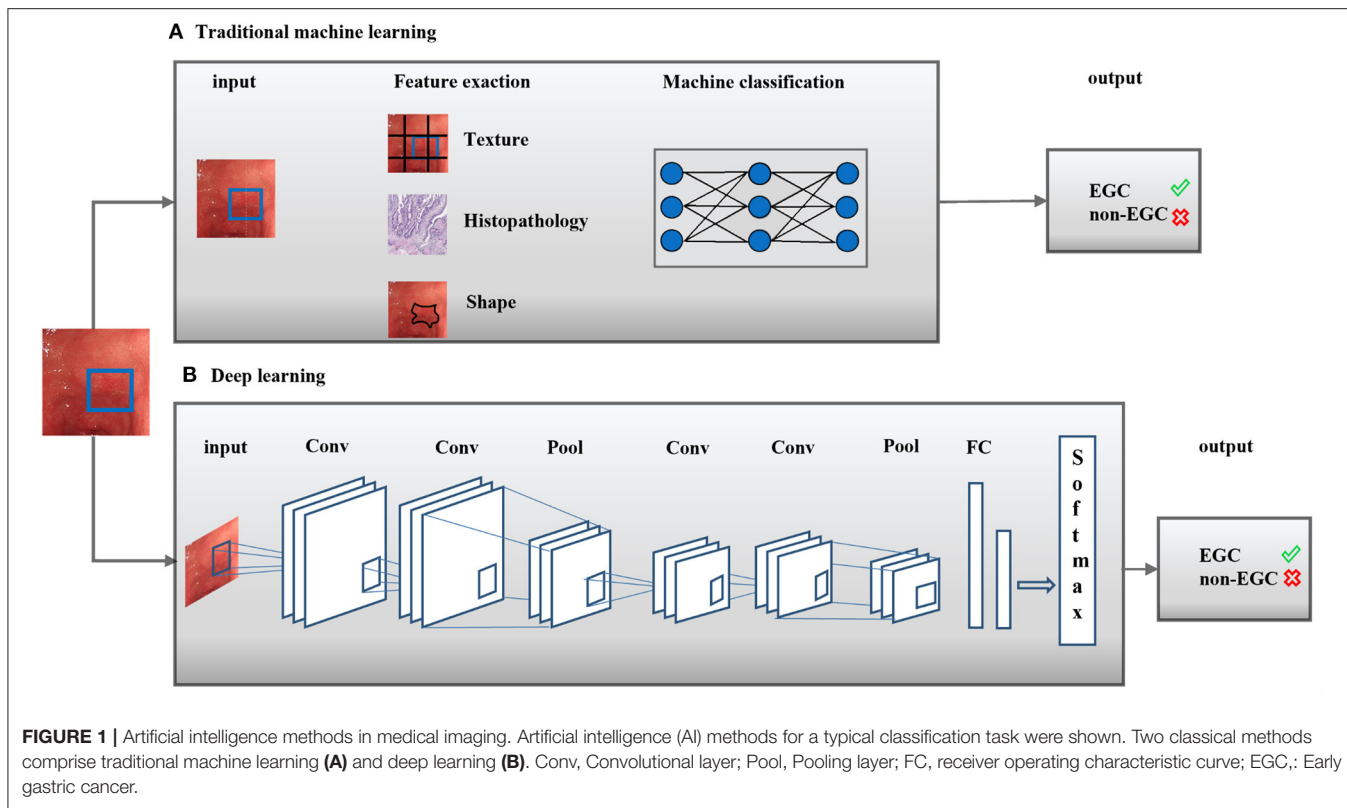
Data Extraction

Two reviewers (Jiang X. T., Wen Y.) independently extracted information, including the author, publication year, region, study type, endoscopy modality, algorithm gold standard and dataset, and used the quality assessment of diagnostic accuracy studies-2 instrument to assess the quality of the study (15). Divergence was resolved through discussion and the involvement of the third reviewer (Li P. W.).

Statistical Analysis

Stata, version 14.2 (StataCorp, College Station, TX) was used for all statistical analyses. Graphpad Prism 8.2.1 was used to delineate the histogram. The TP, FP, FN, and TN observations of each study were input, and the pooled sensitivity and specificity

Abbreviations: EGC, Early gastric cancer; AUC, Area under the receiver operating characteristic curve; ROC, Receiver operating characteristic; WLI, White light imaging; WLE, White light endoscopy; NBI, Narrow band imaging; BLI, blue-laser imaging; EMR, Endoscopic mucosal resection; ESD, Endoscopic submucosal dissection; WHO, World Health Organization; AI, Artificial intelligence; CNN, Convolutional Neural Network; CAD, Computer aided detection; CIs, Confidence intervals; VGG-16, Visual Geometry Group-16; SSD, Single Shot MultiBox Detector; SVM, Support vector machines; DRL, Deep reinforcement learning; Grad-CAM, gradient-weighted class activation mapping.



with the 95% confidence intervals (CIs) for EGC diagnosis with AI were thus calculated. The forest plot was delineated. The inconsistency index (I²) test was used to evaluate the heterogeneity between studies using sensitivity (16). A fixed-effects model would be used with a I² value <50%. More than 50% of the I² values indicated significant heterogeneity. Under this situation, a random-effects model would be applied, and subgroup analysis and influence analysis were performed. A summary receiver operating characteristic (ROC) curve was plotted (17). The area under the curve (AUC) was calculated to estimate the diagnostic accuracy. When the AUC reaches 1.0, it suggests an excellent performance diagnostic test, while if the AUC approaches 0.5, it suggests a poor performance test. Publication bias was evaluated by the Deeks test.

RESULT

Literature Search and Characteristic of Studies

A total of 3,714 studies were retrieved after the search. After removing duplicated studies and excluding improper studies, 17 studies were reserved in this systematic analysis. While Ling et al. (18) distinguished differentiated and undifferentiated type EGC with a sensitivity and specificity of 88.6 and 78.6%, thus was finally excluded in our meta-analysis. A total of 16 studies were finally included in the meta-analysis according to the PRISMA flowchart (Supplementary Figure 1). Three studies were from Korea, eight studies were from Japan, four studies were from

China, and one was from Pakistan. Nine studies used white light endoscopy (WLE) images to establish a training dataset, five studies used narrow band imaging (NBI) images, and two used both WLE and NBI images. Four studies distinguished the invasion depth of EGC. Seven studies compared the diagnostic ability of AI with endoscopists. Two studies applied video to train the dataset. No prospective studies were carried out currently. The general algorithm methods were Visual Geometry Group-16 (VGG-16), ResNet-50, GoogLeNet, Single Shot MultiBox Detector (SSD), Inception neural network and Support vector machines (SVM) classifier. Yoon et al. applied two kinds of algorithm models in his study. The basic characteristics of the included studies and the risk of bias using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool are presented in Table 1 and Supplementary Figure 2.

Diagnostic Performance of AI on EGC Diagnosis

A total of 170,8519 images were utilized for machine training. A total of 22,621 EGC images from the 16 studies were included in the meta-analysis of EGC diagnosis. The diagnostic ability of AI-assisted endoscopy in each study is shown in Supplementary Table 1. The AUC of the AI-assisted endoscopy diagnosis in EGC detection was 0.96 (95% CI, 0.94–0.97) with heterogeneity I² value of 0.98, thus the random effect model was applied. The pooled sensitivity was 86% (95% CI, 77–92%), and the specificity was 93% (95% CI, 89–96%). While the AUC, sensitivity and specificity of AI-assisted depth distinction was

TABLE 1 | Basic characteristic of the included studies.

	Year	Nation	Study type	Endoscopy for training	Image type	Format	Processing image size	DL	Algorithm	Affiliated tools	Gold standard	Training database	Endoscopist involvement	Real-time
Yoon et al. (19)	2019	Korea	Retrospective	WLE	Image	Not mentioned	Not mentioned	CNN	VGG-16(20)	Grad-CAM	WHO classification of Tumors (21), Japanese classification (22)	Gangnam Severance Hospital, Yonsei University College of Medicine, Korea	No	No
Cho et al. (23)	2019	Korea	Retrospective	WLE	Image	JPEG	1,280 × 640 pixels	CNN	Inception-Resnet-v2	SGD	Histopathology	Endoscopically biopsied or EMR/ESD lesions from Chuncheon and Dongtan Sacred Heart Hospitals, Korea	Yes	No
Sakai et al. (24)	2018	Japan	Retrospective	WLE	image	Not mentioned	224 × 224 pixels	CNN	GoogLeNet(25)	No	Histopathology	Not mentioned	No	No
Horiuchi et al. (26)	2019	Japan	Retrospective	ME-NBI	Image	Not mentioned	224 × 224 pixels	CNN	GoogLeNet	No	Histopathology	Cancer Institute Hospital, Ariake, Koto-ku, Japan	No	No
Lan et al. (27)	2019	China	Retrospective	ME-NBI	Image	Not mentioned	299 × 299 pixels to 512 × 512 pixels	CNN	Inception-v3	Keras deep learning framework	Revisited Vienna classification of gastrointestinal epithelial neoplasia(28)	Four hospitals in four areas of Zhejiang province	Yes	No
Toshiaki et al. (29)	2018	Japan	Retrospective	WLE, Chromoendoscopy and NBI	image	Not mentioned	300 × 300 pixels	CNN	SSD(30)	No	Japanese classification	Cancer Institute Hospital Ariake, Japan, Tokatsu Tsujinaka Hospital, Japan and Tomohiro Institute of Gastroenterology and Proctology, Japan, Lalaport Yokohama Clinic, Japan	No	No
Yan et al. (31)	2019	China	Retrospective	WLE	image	Not mentioned	299 × 299 pixels	CNN	ResNet50(32)	No	Japanese classification	Endoscopy Center of Zhongshan Hospital, China	Yes	No
Kanesaka et al. (33)	2017	Japan	Retrospective	ME-NBI	image	Not mentioned	40 × 40 pixels	CAD	SVM classifier	No	pathology-proven EGCS resected by ESD	Ethics Committee of the Osaka International Cancer Institute	No	No
Wu et al. (34)	2018	China	Retrospective	WLE, NBI, BLE	video	Not mentioned	224 × 224 pixels	CNN	VGG-16, ResNet-50	No	Histopathology	Renmin Hospital of Wuhan University, China	Yes	Yes
Miyaki et al. (35)	2013	Japan	Retrospective	magnifying endoscope	image	Not mentioned	1,280 × 1024 pixels	CAD	SVM classifier	No	Histopathology	Hiroshima University Hospital	No	No
Ikenoyama et al. (36)	2020	Japan	Retrospective	WLE	image	Not mentioned	300 × 300 pixels	CNN	SSD	SGD	Histopathology	Cancer Institute Hospital Ariake, Tokatsu-Tsujinaka Hospital, Tada Tomohiro Institute of Gastroenterology and Proctology, Lalaport Yokohama Clinic, Japan	Yes	No
Ali et al. (37)	2018	Pakistan	Retrospective	Chromoendoscopy	Image	Not mentioned	Not mentioned	CAD	SVM classifier	G2LCM descriptors	Not mentioned	Public data-set at the Portuguese Institute of Oncology	No	No
Bun-Joo et al. (38)	2020	Korea	Retrospective	WLE	Image	JPEG	480 × 480 pixels	CNN	Inception-ResNet-v2 and DenseNet-161	Class activation map (CAM)	Histopathology	Chuncheon Sacred Heart Hospital	No	No
Horiuchi et al. (39)	2020	Japan	Retrospective	ME-NBI	Video	Not mentioned	224 × 224 pixels	CNN	GoogLeNet	SGD	Histopathology	Lesions initially treated with ESD at the Cancer Institute Hospital	Yes	No
Ueyama et al. (40)	2020	Japan	Retrospective	ME-NBI	Image	Not mentioned	224 × 224 pixels	CNN	ResNet50	SGD	Japanese Classification	Department of Gastroenterology, Juntendo University School of Medicine	No	No
Zhang et al. (41)	2020	China	Retrospective	WLE	Image	Not mentioned	Not mentioned	CNN	ResNet34	DeepLabv3 structure	Histopathology	Gastric cases admitted to Peking University People's Hospital	Yes	Yes

WLE, White Light Endoscopy; NBI, Narrow Band Imaging; BLI, blue-laser imaging; WHO, World Health Organization; SVM, support vector machine; SSD, Single Shot MultiBox Detector; CNN, Convolutional Neural Network, CAD, Computer-aided diagnosis; Grad-CAM, gradient-weighted class activation mapping; VGG-16, Visual Geometry Group-16, SVM, Support vector machines, SGD, Stochastic gradient descent.

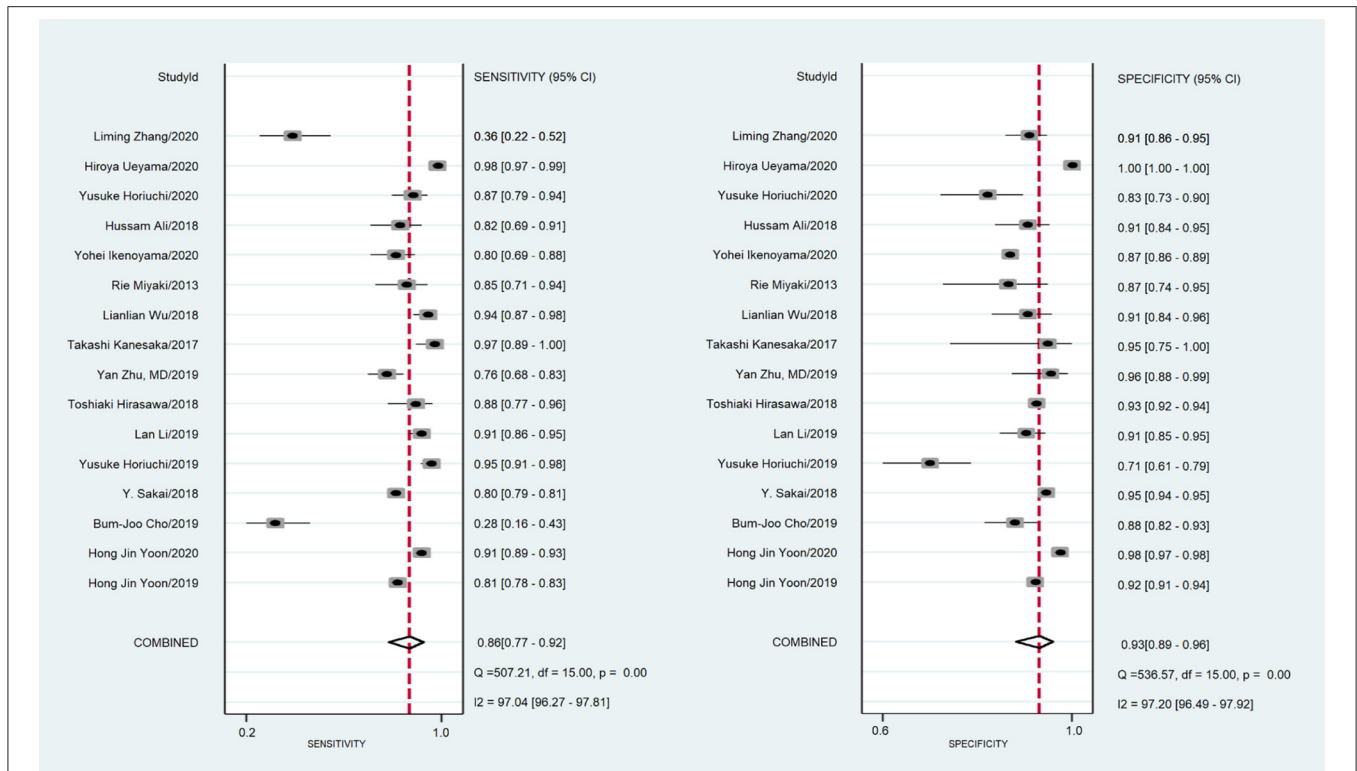


FIGURE 2 | The forest plot of pooled sensitivity and specificity of AI detection on EGC. The pooled sensitivity was 86% (95% CI, 77–92%) and specificity was 93% (95% CI, 89–96%).

0.82 (95% CI, 0.78–0.85), 72% (95% CI, 58–82%), and 79% (95% CI, 56–92%). The forest plots of sensitivity, specificity of AI detection and depth distinction are shown in **Figures 2, 3**. ROC of detection and depth distinction are shown in **Figure 4**. Influence analysis showed that Bum-Joo Cho, Hiroya Ueyama, and Yusuke Horiuchi’s study had the greatest impact on the results (**Supplementary Figure 3**). After rejecting them, the pooled AUC, sensitivity and specificity were 0.95 (95% CI, 0.93–0.97), 85% (95% CI, 78–90%), and 92% (95% CI, 90–94%), respectively, which still indicated an accurate diagnostic ability of AI-aided diagnosis of EGC. The funnel plot asymmetry with a *p*-value of 0.81 showed the absence of publication bias for the included studies (**Supplementary Figure 4**).

Other Factors That Have an Impact on the Accuracy of AI

The effects of the original images from WLE or NBI on the AI diagnostic ability were compared. The sensitivity of the NBI image application was 95% (95% CI, 91–97%), while that of WLE was 73% (95% CI, 57–85%), and the specificity was 96% (95% CI, 70–100%) and 93% (95% CI, 90–95%).

When the number of training images was more than 10,000, the sensitivity and specificity were 88% (95% CI, 83–92%) and 94% (95% CI, 91–96%), respectively, more than that of the sensitivity 85% (95% CI, 69–93%) and specificity 93% (95% CI, 82–97%) of the group that had >10,000 training images.

For the control group, sensitivity and specificity of the expert endoscopist vs. non-expert endoscopist diagnosis were 79% (95% CI, 61–90%) vs. 73% (95% CI, 61–82%), 85% (95% CI, 77–90%) vs. 83% (95% CI, 67–92%), respectively. Here, the general expert endoscopists were those who had clinical experience with endoscopy examination for more than 10 years. **Figure 5** shows the subgroup results.

DISCUSSION

Japanese researchers published a minimum required standard for the “systematic screening protocol for the stomach,” which comprised 22 images of the stomach to precisely discover suspicious cancerous lesions (42). In 2016, the European Society of Gastrointestinal Endoscopy (ESGE) published a protocol comprising 10 images of the stomach (43). However, these protocols could not be carried out absolutely, and endoscopists may miss some regions during the examination due to individual operative levels and subjective factors, which can lead to the misdiagnosis of EGC (44–46).

Deep learning (47, 48), which is typically based on artificial neural networks, aims at learning multilevel manifestations of data to make predictions. The development of deep convolutional neural networks has particularly altered the computer vision field (49, 50).

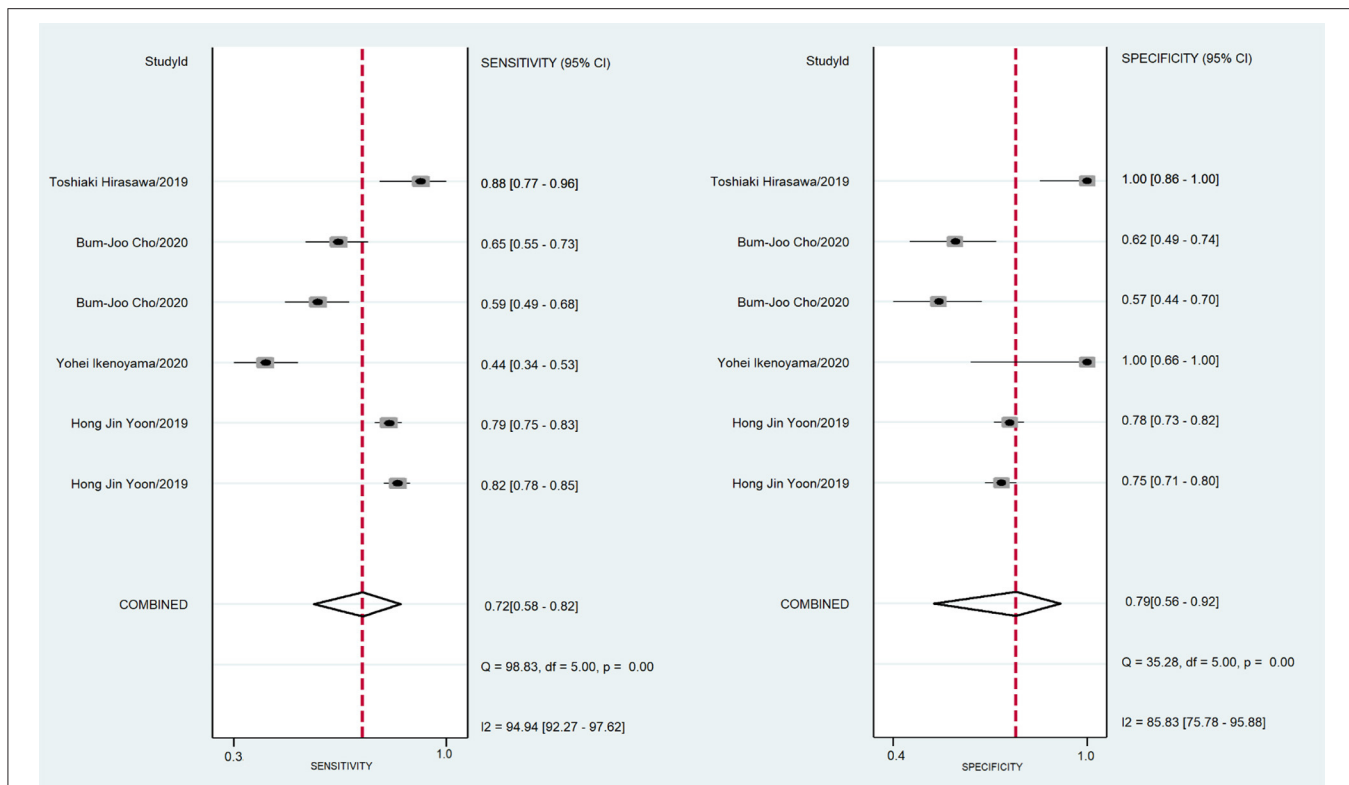


FIGURE 3 | The forest plot of pooled sensitivity and specificity of AI distinction depth on EGC. The pooled sensitivity was 72% (95% CI, 58–82%) and specificity was 79% (95% CI, 56–92%).

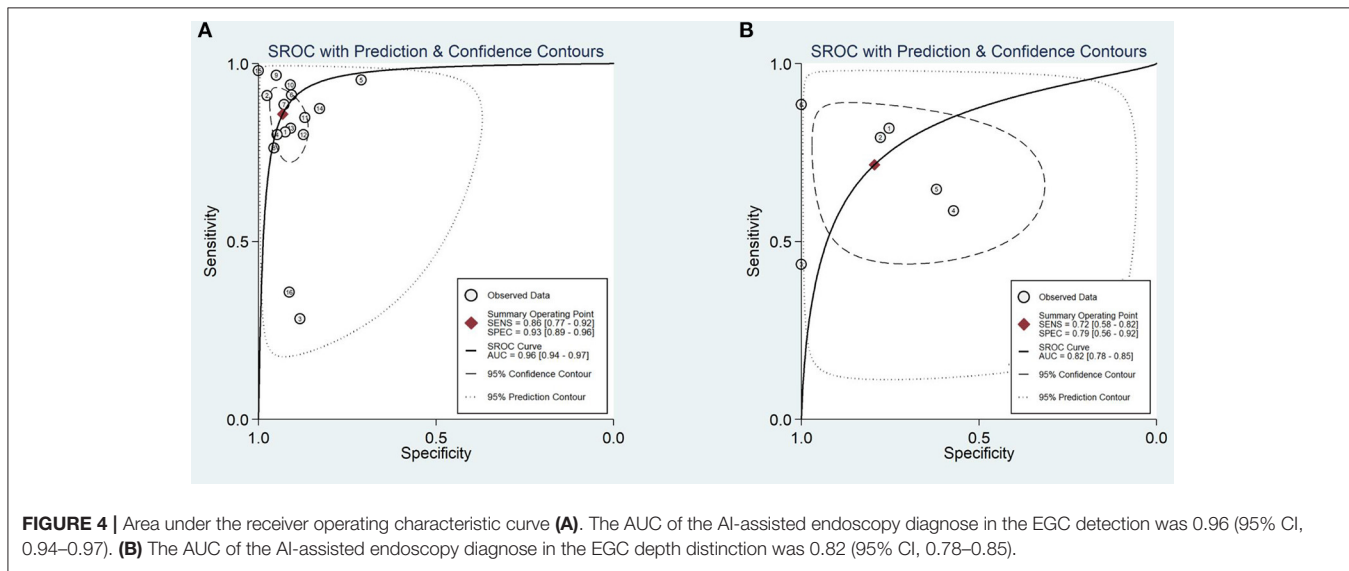
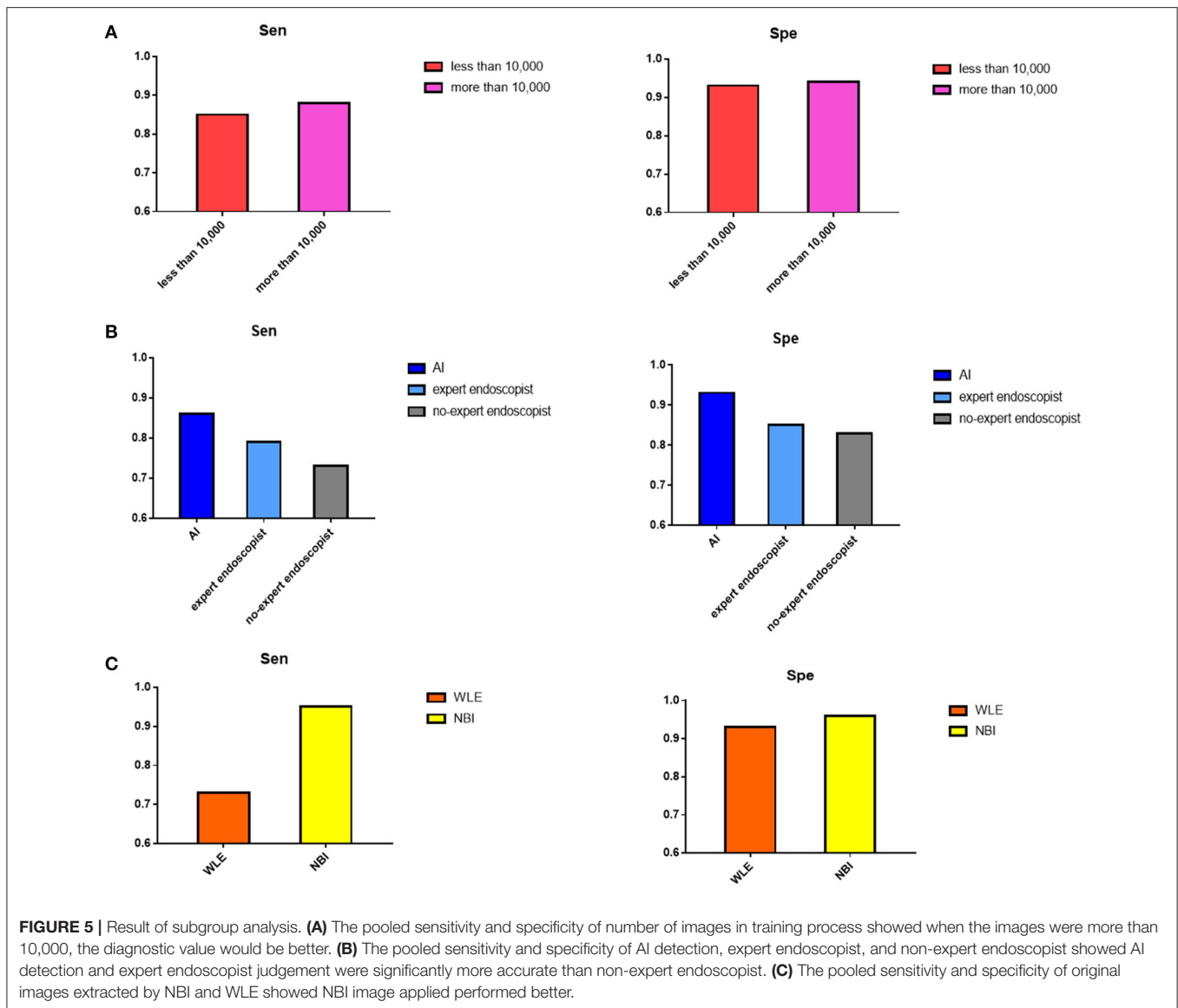


FIGURE 4 | Area under the receiver operating characteristic curve (A). The AUC of the AI-assisted endoscopy diagnose in the EGC detection was 0.96 (95% CI, 0.94–0.97). (B) The AUC of the AI-assisted endoscopy diagnose in the EGC depth distinction was 0.82 (95% CI, 0.78–0.85).

Application of AI recognition with endoscopic images to detect the depth of wall invasion of gastric cancer was initially reported by Keisuke Kubota with an accuracy of 64.7% (51). Soon afterwards, several studies have shown excellent results for advanced technology. Hence, it is necessary to summarize the existing studies to realize the probable ability of AI on EGC detection and discuss what factors may influence the results.

This is the first meta-analysis on the performance of AI on EGC diagnosis with endoscopy. In this article, we indicated that the application of AI in endoscopic detection of EGC achieved an AUC of 0.96 (95% CI, 0.94–0.97), a sensitivity of 86% (95% CI, 77–92%), and a specificity of 93% (95% CI, 89–96%), which manifested a more accurate diagnostic ability than independent detection by endoscopists, while the depth distinction was dissatisfied with a sensitivity, specificity and AUC of 0.82 (95%



CI, 78–85%), 72% (95% CI, 58–82%), and 79% (95% CI, 56–92%). The common reasons for misdiagnosis were lesions of gastritis or flat or depressed texture and anatomical structure which was hard to identify. The cancer invasion depth was classically distinguished by morphologically evaluating several findings such as the concentration of stomach wall folds, the marginal ridge, the elasticity and thickness of the lesion, and the presence of variant of the stomach wall due to the volume of insufflation air in the stomach with WLE (52–54). Furthermore, the accuracy of discriminating EGC depth by conventional endoscopy was reported to be 62–80% (55). Thus, the AI applied endoscopy performed well on EGC depth determination. Bum-Joo Cho, Hiroya Ueyama and Yusuke Horiuchi’s study (23, 26, 40) showed significant heterogeneity. Cho et al. used the Inception-Resnet-v2 model with an AUC of 74.5 (95% CI, 67.9–80.4) and a sensitivity of 28.3 (95% CI, 16.0–43.5). The included poor-quality images, composition of the database, and pathological

classification criteria may cause poor diagnostic performance. In addition, we performed several subgroup analyses to delineate the probable influencing factors of AI performance.

For the algorithm model, Simonyan et al. (56) investigated the value of the convolutional network depth on its accuracy in large-scale image recognition setting. The result showed that when the depth was pushed to 16–19 weight layers, it would have a significant improvement on the prior-art configurations. VGG-16 had 16 convolutional and three fully connected layers, which were carried out by five max-pooling layers and used filters with a small receptive field to achieve a low error rate in practice. On the other hand, SVM also performed excellently in the included studies. SVM is utilized in distinguishing two classes and creating the boundary line to maximize the distance between the hyperplane and the nearest sample. Compared to other mathematical models (57–59), SVMs are utilized to model physical systems by adapting their parameters

(60–63). SVMs are widely known for their application in classification (64).

The endoscopic image modality of validation set should be same to the training set. For training images from different endoscopy modalities, the sensitivity of studies using images from NBI seemed to be better than those using images from WLE (96 vs. 93%). A model which was trained with NBI images could only recognize NBI images in practice. However, a multicenter randomized controlled trial that compared a non-magnifying NBI with WLI indicated no significant difference in gastric cancer detection (65). Although NBI is currently regarded as the most broadly applied image-enhanced modality in AI research, the impact of other imaging modalities, such as the lately available linked-color imaging or blue-laser imaging modalities, need more studies for verification.

For the number of training images, it seemed that the more images the machine trained, the more accurate the AI detection would be. The concept that a large number of images are a prerequisite to structure a learning model was also certified in the research conducted by Seguí et al. (66) for motility movement classification in wireless capsule endoscopy. A recent meta-analysis similarly indicated that a ten-fold increase in training data size could improve the accuracy of AI detection by 3% (67).

Neural networks have the potential capacity for clinical practice and can be significantly popularized in the gastrointestinal field. However, CNN detection is temporarily in the stage of research. This study also had some limitations. A limited number of available studies fit the inclusion criteria since the novel technology has just been developed in recent years. Thus, the subgroup results were not completely reliable due to the limited number of studies. All the included studies were retrospective, which may lead to selection bias of included images, particularly in the validation dataset. In addition, few studies provided a solution to multiple gastrointestinal abnormalities as comparison, while most studies only researched the detection of a single abnormality, including Barrett's esophagus, *Helicobacter pylori* infection, early gastric cancer, atrophic gastritis, etc. (68–70), which is insufficient for clinical application. Moreover, an AI EGC detection model based on full-length videos was scarce, which postpones its general application in clinical practice.

To overcome these limitations, several projects can be carried out in the future. More prospective studies can be designed for strict images, including criteria, high-definition image extraction and expert endoscopist involvement to prove higher level evidence. Luo et al. (71) has carried out a multicenter, case-control, prospective real-time diagnostic study on artificial intelligence for detection of esophagus and gastric cancer with accuracy of 0.955 (95% CI 0.952–0.957). GRAIDS algorithm, which was based on the concept of DeepLab's V3+ (72, 73), was utilized in this prospective study. Expanding the training image number is necessary to improve the machine recognition ability. On the other hand, the validation images are supposed to be larger. Training images extracted from different endoscopy modalities still need to be investigated to establish a popularized dataset. Currently, limited data have shown that the VGG-16, SSD, and SVM classifier models are credible computer-aided diagnosis algorithms. Another branch of deep learning, deep

reinforcement learning (DRL), recently performed at the top level in the GO game in 2016 (74). DRL is likely to be applied in the EGC detection field. DRL combines deep learning with reinforcement learning, incorporating not only the excellent perception and distinguishing abilities of deep learning in visual tasks but also the decision-making capabilities of reinforcement learning (75). DRL has performed well in dealing with dynamic decision problems (74–76). However, DRL has not yet been used in clinical trials. Wu et al. (77) reported that the application of WISENSE, a mechanism that utilizes aspects of both CNN and DRL, could decrease the number of blind spots during an upper endoscopy, initially achieving an accuracy of 90.02%. The exploration of accurate algorithms is worthy of being explored.

CONCLUSION

This is the first meta-analysis to summarize current evidence of AI applications in EGC diagnosis. The AI applications seemed to be more accurate in parts of EGC detection than the endoscopists. The VGG-16, SSD, and SVM classifier models probably performed better according to the limited studies. When the number of training images is expanded, the accuracy will be improved. More strictly designed perspective studies with different reliable CNN algorithms are needed to make AI universal in clinical practice.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

KJ: study concept and design and analysis and interpretation of data. XJ and YW: acquisition of data. SW, KN, ZZ, SJ, and PLiu: literature search. XJ, JP, and YH: figure processing. PLi and SL: critical revision of the manuscript for important intellectual content. FL: study supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Major Program of the National Natural Science Foundation of China (81973819), National Natural Science Foundation for Young Scholars of China (81904139), National Natural Science Foundation for Young Scholars of China (81904145), Guangdong Province Natural Science Foundation of China (2019A1515011145), Guangdong Province medical Program of Scientific technology Foundation (A2020186), High level Hospital development Program of First Affiliated Hospital of Guangzhou University (2019QN01), Specific Clinical Study of the Second Innovation Hospital Program (2019IIT19), and Young Scholars of QiHuang of Traditional Chinese Medicine of China (20207).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.629080/full#supplementary-material>

Supplementary Figure 1 | Literature screening flow according to PRISMA.

Supplementary Figure 2 | Quality of included studies according to QUADAS-2 scale.

Supplementary Figure 3 | Influence analysis showed the significantly heterogeneity in Bum-Joo Cho, Hiroya Ueyama, and Yusuke Horiuchi's study.

Supplementary Figure 4 | Funnel plot showed no publication bias among the included studies.

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA*. (2018) 68:394–424. doi: 10.3322/caac.21492
- Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, 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 Cancer J Clin*. (2017) 67:93–9. doi: 10.3322/caac.21388
- Sano T, Coit DG, Kim HH, Roviello F, Kassab P, Wittekind C, et al. Proposal of a new stage grouping of gastric cancer for TNM classification: international gastric cancer association staging project. *Gastric Cancer*. (2017) 20:217–25. doi: 10.1007/s10120-016-0601-9
- Rice TW, Ishwaran H, Hofstetter WL, Kelsen DP, Apperson-Hansen C, Blackstone EH, et al. Recommendations for pathologic staging (pTNM) of cancer of the esophagus and esophagogastric junction for the 8th edition AJCC/UICC staging manuals. *Dis Esophagus*. (2016) 29:897–905. doi: 10.1111/dote.12533
- Hamashima C, Okamoto M, Shabana M, Osaki Y, Kishimo T. Sensitivity of endoscopic screening for gastric cancer by the incidence method. *Int J Cancer*. (2013) 133:653–9. doi: 10.1002/ijc.28065
- Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. *Endosc Int Open*. (2014) 2:46–50. doi: 10.1055/s-0034-1365524
- Qiang Z, Fei W, Zhen YC, Wang Z, Zhi FC, Liu SD, et al. Comparison of the diagnostic efficacy of white light endoscopy and magnifying endoscopy with narrow band imaging for early gastric cancer: a meta-analysis. *Gastric Cancer*. (2016) 19:543–52. doi: 10.1007/s10120-015-0500-5
- Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. (2018) 18:500–10. doi: 10.1038/s41568-018-0016-5
- Samuel A L. Some studies in machine learning using the game of checkers, II-recent progress. *Ibm J. Res. Dev*. (1988) 11:335–6. doi: 10.1007/978-1-4613-8716-9_14
- Dey A. Machine learning algorithms: a review. *IJCSIT*. (2016) 7:1174–9. doi: 10.21275/ART20203995
- Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. (1998) 86:2278–324. doi: 10.1109/5.726791
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003
- Castellino RA. Computer aided detection (CAD): an overview. *Cancer Imag*. (2005) 5:17–19. doi: 10.1102/1470-7330.2005.0018
- Jin P, Ji XY, Kang WZ, Liu H, Ma F, Ma S, et al. Artificial intelligence in gastric cancer: a systematic review. *J Cancer Res Clin Oncol*. (2020) 146:2339–50. doi: 10.1007/s00432-020-03304-9
- Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. (2011) 155:529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
- Higgins JPT, Thompson SG, Deeks JJ, DG Altman. Measuring inconsistency in meta-analyses. *BMJ*. (2003) 327:557–60. doi: 10.1136/bmj.327.7414.557
- Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. (2005) 58:982–90. doi: 10.1016/j.jclinepi.2005.02.022
- Ling TS, Wu LL, Fu YW, 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*. (2020). doi: 10.1055/a-1229-0920. [Epub ahead of print].
- Yoon HJ, Kim S, Kim JH, Keum JS, Oh SI, Jo J, Chun J, et al. A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. *J Clin Med*. (2019) 8:1310–20. doi: 10.3390/jcm8091310
- Liu SY, Deng WH. Very deep convolutional neural network based image classification using small training sample size. In: *3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, (2015). p. 730–4. doi: 10.1109/ACPR.2015.7486599
- Jung H, Lee HH, Song KY, Jeon HM, Park CH. Validation of the seventh edition of the American Joint Committee on Cancer TNM staging system for gastric cancer. *Cancer*. (2011). 117:2371–8. doi: 10.1002/cncr.25778
- Japanese Gastric Cancer Association, Japanese Classification of Gastric Carcinoma - 2nd English Edition. *Gastric Cancer*. (1998) 1:10–24. doi: 10.1007/PL00011681
- Bum-Joo Cho, Chang SB, Se WP, Yang YJ, Seo SI, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy*. (2019) 51:1121–9. doi: 10.1055/a-0981-6133
- Sakai Y, Takemoto S, Hori K, Nishimura M, Ikematsu H, Yano T, et al. Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc*. (2018) 2018:4138–41. doi: 10.1109/EMBC.2018.8513274
- Aswathy, Siddhartha, Mishra. Deep GoogLeNet Features for Visual Object Tracking. In: *IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*. Trivandrum (2018). p. 60–6. doi: 10.1109/ICIINFS.2018.8721317
- Horiuchi Y, Aoyama K, Tokai Y, Hirasawa T, Yoshimizu S, Ishiyama A, et al. Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. *Digestive Dis Sci*. (2020) 65:1355–63. doi: 10.1007/s10620-019-05862-6
- Lan L, Yishu C, Zhe S, Zhang X, Sang J, Ding Y, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer*. (2020) 23:126–32. doi: 10.1007/s10120-019-00992-2
- Dixon MF. Gastrointestinal epithelial neoplasia: Vienna revisited. *Gut*. (2002) 51:130–1. doi: 10.1136/gut.51.1.130
- Toshiaki H, Kazuharu A, Tetsuya T, Ishihara S, Shichijo S, Ozawa 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. doi: 10.1007/s10120-018-0793-2
- Chen Z, Zhang L, Peng. Fast single shot multibox detector and its application on vehicle counting system. *IET Intelligent Transport Syst*. (2018). 12:1406–13. doi: 10.1049/iet-its.2018.5005
- Yan Z, Wang QC, Xu MD, Zhang Z, Cheng J, Zhong YS, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastroint Endosc*. (2019) 89:806–15. doi: 10.1016/j.gie.2018.11.011
- He K, Zhang X, Ren S, Jian S. Deep residual learning for image recognition. *Proc IEEE Conf Comput Vision Pattern Recogn*. (2016) 770–78. doi: 10.1109/CVPR.2016.90
- Kanesaka T, Lee TC, Uedo N, Lin KP, Chen HZ, Lee JY, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band images. *Gastroint Endosc*. (2017) 87:1339–44. doi: 10.1016/j.gie.2017.11.029

34. Wu LL, Zhou W, Wan XY, Zhang J, Shen L, Hu S, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. *Endoscopy*. (2019) 51:522–31. doi: 10.1055/a-0855-3532
35. Miyaki R, Yoshida S, Tanaka S, Kominami Y, Sanomura Y, Matsuo T, et al. Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement. *Gastroenterol Hepatol*. (2013) 28:841–7. doi: 10.1111/jgh.12149
36. Ikenoyama Y, Hirasawa T, Ishioka M, Namikawa K, Yoshimizu S, Horiuchi Y, et al. Detecting early gastric cancer: comparison between the diagnostic ability of convolutional neural networks and endoscopists. *Dig Endosc*. (2020). doi: 10.1111/den.13688. [Epub ahead of print].
37. Ali H, Yasmin M, Sharif M, Rehmani MH. Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images. *Comp Methods Programs Biomed*. (2018) 157:39–47. doi: 10.1016/j.cmpb.2018.01.013
38. Bum-Joo Cho, Bang CS, Lee JJ, Seo CW, Kim JH. Prediction of submucosal invasion for gastric neoplasms in endoscopic images using deep-learning. *J Clin Med*. (2020) 9:1858–72. doi: 10.3390/jcm9061858
39. Horiuchi Y, Hirasawa T, Ishizuka N, Tokai Y, Namikawa K, Yoshimizu S, et al. Performance of a computer-aided diagnosis system in diagnosing early gastric cancer using magnifying endoscopy videos with narrow-band imaging (with videos). *Gastrointest Endosc*. (2020) 92:856–65. doi: 10.1016/j.gie.2020.04.079
40. Ueyama H, Kato Y, Akazawa Y, Yatagai N, Komori H, Takeda T, 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*. (2020). doi: 10.1111/jgh.15190. [Epub ahead of print].
41. Zhang LM, Zhang Y, Wang L, Wang J, Liu Y. Diagnosis of gastric lesions through a deep convolutional neural network. *Dig Endosc*. (2020). doi: 10.1111/den.13844. [Epub ahead of print].
42. Yao K. The endoscopic diagnosis of early gastric cancer. *Ann Gastroenterol*. (2013) 26:11–22. doi: 10.1016/S0016-5107(79)73384-0
43. Bisschops R, Areia M, Coron E, Adler S, Cash BD, Fernández-Urién I, et al. Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. *Endoscopy*. (2016) 48:843–64. doi: 10.1055/s-0042-113128
44. Yao K, Uedo N, Muto M, Ishikawa H. Development of an e-learning system for teaching endoscopists how to diagnose early gastric cancer: basic principles for improving early detection. *Gastric Cancer*. (2017) 20:28–38. doi: 10.1007/s10120-016-0680-7
45. Scaffidi MA, Grover SC, Carnahan H, Khan R, Amadio JM, Yu JJ, et al. Impact of experience on self-assessment accuracy of clinical colonoscopy competence. *Gastrointest Endosc*. (2018) 87:827–36. doi: 10.1016/j.gie.2017.10.040
46. O'Mahony S, Naylor G, Axon A. Quality assurance in gastrointestinal Endoscopy. *Endoscopy*. (2000) 32:483–8. doi: 10.1055/s-2000-649
47. Iecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015). 521:436–44. doi: 10.1038/nature14539
48. Christian Robert. Machine Learning, a Probabilistic Perspective, CHANCE, (2014). 27:62–3. doi: 10.1080/09332480.2014.914768
49. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Adv Neural Inform Proc Syst*. (2012) 25:1097–105.
50. Szegedy c Vanhoucke V, ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *IEEE Conf Comp Vision Pattern Recogn*. (2016) 28:18–26. doi: 10.1109/CVPR.2016.308
51. Kubota K, Kuroda J, Yoshida M, Ohta K, Kitajima M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. *Surg Endosc*. (2012) 26:1485–9. doi: 10.1007/s00464-011-2036-z
52. Maruyama Y, Shimamura T, Koda K. Diagnosis of the depth of early gastric cancer by conventional and dying endoscopy—from the viewpoint of the size and macroscopic type. *Stomach Intestine*. (2014) 49:35–46.
53. Nagahama T, Yao K, Imamura K, Kojima T, Ohtsu K, Chuman K, et al. Diagnostic performance of conventional endoscopy in the identification of submucosal invasion by early gastric cancer: the “non-extension sign” as a simple diagnostic marker. *Gastric Cancer*. (2017) 20:304–13. doi: 10.1007/s10120-016-0612-6
54. Takeda T, So S, Sakurai T, Nakamura S, Yoshikawa I, Yada S, et al. Learning effect of diagnosing depth of invasion using non-extension sign in early gastric cancer. *Digestion*. (2020) 101:191–7. doi: 10.1159/000498845
55. Watari J, Ueyama S, Tomita T, Ikehara H, Hori K, Hara K, et al. What types of early gastric cancer are indicated for endoscopic ultrasonography staging of invasion depth? *World J Gastrointest Endosc*. (2016). 8:558–67. doi: 10.4253/wjge.v8.i16.558
56. Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition. (2014) *arXiv [Preprint]*. arXiv:1409.1556.
57. Antón JCA, Nieto PJG, Juez FJ, Lasheras S, Roquei Gutierrez, et al. Battery state-of-charge estimator using the MARS technique. *IEEE Trans Power Electron*. (2013) 28:3798–805. doi: 10.1109/TPEL.2012.2230026
58. Juez FJ, Sánchez Lasheras F, García Nieto PJ, MAS Suárez. A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. *Int J Comput Math*. (2009) 86:1878–87. doi: 10.1080/00207160902783557
59. Sánchez-Lasheras F, Andrés J, Lorca P, Juez FJDC. A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Syst Appl*. (2012) 39:7512–23. doi: 10.1016/j.eswa.2012.01.135
60. Osborn J, De Cos Juez JF, Guzmán D, Butterley T, Myers R, et al. Using artificial neural networks for open-loop tomography. *Opt Express*. (2012) 20:2420–34. doi: 10.1364/OE.20.002420
61. Guzmán D, Juez FJ, Myers R, Guesalaga A, Lasheras FS. Modeling a MEMS deformable mirror using non-parametric estimation techniques. *Opt Express*. (2010) 18:21356–69. doi: 10.1364/OE.18.021356
62. Guzmán D, Juez FJ, Sánchez L, FS Lasheras, R Myers, L Young. Deformable mirror model for open-loop adaptive optics using multivariate adaptive regression splines. *Opt Express*. (2010) 18:6492–505. doi: 10.1364/OE.18.006492
63. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. (1995) 20:273–97. doi: 10.1007/BF00994018
64. Juez FJ, García Nieto PJ, Martínez Torres J. Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model. *Math Comput Model*. (2010) 52:1177–84. doi: 10.1016/j.mcm.2010.03.017
65. Ang TL, Pittayanon R, Lau JY, Rerknimitr R, Ho SH, Singh R, et al. A multicenter randomized comparison between high-definition white light endoscopy and narrow band imaging for detection of gastric lesions. *Eur J Gastroenterol Hepatol*. (2015) 27:1473–8. doi: 10.1097/MEG.0000000000000478
66. Segui S, Drozdal M, Pascual G, Radeva P, Malagelada C, Azpiroz F, et al. Generic feature learning for wireless capsule endoscopy analysis. *Comp Biol Med*. (2016) 79:163–72. doi: 10.1016/j.compbiomed.2016.10.011
67. Soffer S, Klang E, Shimon O, Nachmias N, Eliakim R, Ben-Horin S, et al. Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. *Gastrointest Endosc* (2020) 92:831–9. doi: 10.1016/j.gie.2020.04.039
68. Jisu H, Bo-Yong P, Hyunjin P. Convolutional neural network classifier for distinguishing barrett's esophagus and neoplasia endomicroscopy images. *Conf Proc IEEE Eng Med Biol Soc*. (2017) 2017:2892–5. doi: 10.1109/EMBC.2017.8037461
69. Nakashima H, Kawahira H, Kawachi H, Sakaki N. 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. doi: 10.20524/aog.2018.0269
70. Guimar aes P, Keller A, Fehlmann T, Lammert F, Casper M. Deep-learning based detection of gastric precancerous conditions. *Gut*. (2020) 69:4–6. doi: 10.1136/gutjnl-2019-319347
71. Luo H, Xu GL, Li CF, He L, Luo L, Wang Z, 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. doi: 10.1016/S1470-2045(19)30637-0
72. Pohlen T, Hermans A, Mathias M, Leibe B. Full-resolution residual networks for semantic segmentation in street scenes. *arXiv*. (2017) 1:3309–3318. doi: 10.1109/CVPR.2017.353

73. Bei-Bei L, Bei H. Encoder-decoder for semi-supervised image semantic segmentation. *Comp Syst Appl.* (2019) 28:182–7. doi: 10.15888/j.cnki.csa.007159
74. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Schrittwieser J, et al. Mastering the game of go with deep neural networks and tree search. *Nature.* (2016) 529:484–9. doi: 10.1038/nature16961
75. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature.* (2015) 518:529–33. doi: 10.1038/nature14236
76. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. (2013) *arXiv [Preprint]. arXiv:1312.5602.*
77. Wu Lianlian, Zhang Jun, Zhou Wei, An P, Shen L, Liu J, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut.* (2019). 68:2161–9. doi: 10.1136/gutjnl-2018-317366

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Jiang, Jiang, Pan, Wen, Huang, Weng, Lan, Nie, Zheng, Ji, Liu, Li and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.