Received: 22 January 2021 | Revised: 2 March 2021 | Accepted: 15 April 2021

DOI: 10.1111/1751-2980.12992

META ANALYSIS

Journal of Digestive Diseases WILEY

Accuracy of artificial intelligence-assisted detection of esophageal cancer and neoplasms on endoscopic images: A systematic review and meta-analysis

Si Min Zhang[1,2,3] | Yong Jun Wang[1,2,3] | Shu Tian Zhang[1,2,3]

[1]Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, Beijing, China

[2]National Clinical Research Center for Digestive Diseases, Beijing, China

[3]Beijing Digestive Disease Center, Beijing, China

Correspondence
Shu Tian Zhang, Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, No. 95 Yong'an Road, Xicheng District, Beijing 100050, China.
Email: zhangshutian@ccmu.edu.cn

Objective: To investigate systematically previous studies on the accuracy of artificial intelligence (AI)-assisted diagnostic models in detecting esophageal neoplasms on endoscopic images so as to provide scientific evidence for the effectiveness of these models.

Methods: A literature search was conducted on the PubMed, EMBASE and Cochrane Library databases for studies on the AI-assisted detection of esophageal neoplasms on endoscopic images published up to December 2020. A bivariate mixed-effects regression model was used to calculate the pooled diagnostic efficacy of AI-assisted system. Subgroup analyses and meta-regression analyses were performed to explore the sources of heterogeneity. The effectiveness of AI-assisted models was also compared with that of the endoscopists.

Results: Sixteen studies were included in the systematic review and meta-analysis. The pooled sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratio and area under the summary receiver operating characteristic curve regarding AI-assisted detection of esophageal neoplasms were 94% (95% confidence interval [CI] 92%-96%), 85% (95% CI 73%-92%), 6.40 (95% CI 3.38-12.11), 0.06 (95% CI 0.04-0.10), 98.88 (95% CI 39.45-247.87) and 0.97 (95% CI 0.95-0.98), respectively. AI-based models performed better than endoscopists in terms of the pooled sensitivity (94% [95% CI 84%-98%] vs 82% [95% CI 77%-86%, $P < 0.01$).

Conclusions: The use of AI results in increased accuracy in detecting early esophageal cancer. However, most of the included studies have a retrospective study design, thus further validation with prospective trials is required.

KEYWORDS
artificial intelligence, diagnosis, early esophageal neoplasms, meta-analysis, systemic review

1 | INTRODUCTION

In 2020, esophageal cancer ranks the seventh most common cancer and the sixth most common cause of cancer death worldwide.[1] Histologically, esophageal cancer can mainly be classified into esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC), which accounts for 90% of all cases with esophageal cancer.[2] ESCC is the main histological type of esophageal

cancer in China.[3] The rate of EAC has been increasing, probably due to a high prevalence of gastroesophageal reflux disease, Barrett's esophagus (BE) and obesity in recent years.[4] The 5-year survival rate of patients with early esophageal cancer may reach 85% or higher; however, most patients are diagnosed at an advanced stage, with a 5-year survival rate decreasing to less than 20%.[5] Therefore, early diagnosis of esophageal cancer is vital to improve patient prognosis.

Thanks to the rapid development of artificial intelligence (AI) technology, AI-assisted diagnostic models established through deep learning have been widely applied for the analysis of gastrointestinal endoscopic images. A large number of images are collected for the construction of AI-assisted diagnostic model; the images are divided into the training dataset, which is used to construct the model, and the testing dataset to validate its effectiveness.[6] The AI-based models have been reported to perform well in the detection of colonic polyps, differentiation of gastric and colonic polyps, as well as the identification of early gastrointestinal tumors and *Helicobacter pylori* (*H. pylori*) infection within gastric mucosal layer.[7] Recently, the application of AI-assisted models has been gradually extended to the endoscopic assessment of esophageal diseases.[8] AI models have been found to be accurate in detecting early esophageal cancer via endoscopic images, some of them are even more effective than experienced endoscopists. In this systematic review and meta-analysis, we aimed to investigate systematically the accuracy of AI-assisted diagnostic models in the detection of esophageal neoplasms on endoscopic images so as to provide scientific evidence for their effectiveness.

2 | MATERIALS AND METHODS

2.1 | Search strategy and study selection

This systematic review and meta-analysis were conducted based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement.[9] To identify all studies on the AI-assisted detection of esophageal neoplasms via endoscopic images, a comprehensive literature search was conducted on the PubMed, EMBASE and Cochrane Library databases covering all articles published up to December 2020. The following terms were used for the search: ("artificial intelligence" OR "AI" OR "deep learning" OR "convolutional network" OR "computer aided") AND ("esophageal squamous cell carcinoma" OR "esophageal cancer" OR "Barrett's esophagus" OR "esophageal adenocarcinoma"). Only English-language articles were included.

2.2 | Eligibility criteria

The studies for the systematic review and meta-analysis were English-language prospective or retrospective studies which employed AI-aided diagnostic models in the detection of benign esophageal lesions, early esophageal neoplasms and esophageal cancer with different invasive depths on endoscopy. Only the studies in which the rates of integral true positivity, false positivity, false negativity and true negativity were available were included in the study. Reviews, comments, letters, editorials, meta-analyses and animal studies were excluded. For duplications, only the one with a larger sample size was included.

2.3 | Data extraction

The following information was extracted from each study: first author, publication year, country or region, manuscript type, number of patients and/or endoscopic images, the rates of true positivity, false positivity, false negativity, and true negativity of the AI-assisted diagnostic models and endoscopists in diagnosing the same dataset, histological type of the lesions, type of endoscopy and algorithm used, and whether the study used videos as a part of the dataset, using an external validation dataset or low-quality images, or achieved a real-time diagnosis. Data extraction was conducted by two authors (SMZ and YJW) independently, and any disagreement was resolved through in-depth discussion and consensus.

2.4 | Methodological quality assessment and the evaluation of potential bias

Two authors (SMZ and YJW) assessed the quality and potential bias of the eligible studies in accordance with the revised quality assessment of diagnostic accuracy studies (QUADAS-2).[10] Any disagreement was resolved through discussion. The tool comprises four domains: patient selection, index test, reference standard, and flow and timing. The first three domains were also assessed for concerns regarding applicability. Each section was classified as having a high, low or unclear risk of bias.

2.5 | Study outcomes

The primary outcomes were the diagnostic accuracy, pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and diagnostic odds ratio (DOR) of the AI-assisted models in the detection of esophageal neoplasms in endoscopic images. The secondary outcomes were the comparison of AI-assisted diagnostic models and endoscopists in terms of the pooled sensitivity and specificity in analyzing the same test datasets, as well as the accuracy of these models in different subgroups.

2.6 | Statistical analyses

Statistical analyses were mainly performed by using the Stata 14.0 (STATA, College Station, TX) including the MIDAS packages, and the subgroup analysis of the studies including less than four references

was performed by using the the Meta-DiSc software version 1.4 (Ramón y Cajal Hospital, Madrid, Spain).[11] The RevMan 5.3 (The Nordic Cochrane Centre, Copenhagen, Denmark) was used to plot the figure of the methodological quality assessment. A bivariate mixed-effects regression model following a random effects model was used for the following metrics: pooled sensitivity and specificity, PLR,NLR, DOR, and the area under the summary receiver operating characteristic (SROC) curve (AUROC) of AI-assisted models and endoscopists in detecting esophageal neoplasms. Heterogeneity across the included studies was first assessed by the visual inspection of the pooled SROC curve, with an asymmetric shape suggesting a significant heterogeneity. In addition, the Spearman's correlation coefficient between the logit-transformed sensitivity and specificity was also used to evaluate heterogeneity. The asymmetry parameter β with a significant probability ($P < 0.05$) combined with a positive correlation coefficient indicated a significant heterogeneity. The sources of heterogeneity were explored through subgroup analysis and regression. Among each subgroup, 95% confidence interval (CI) of the AUROC was calculated and compared. A non-overlapping 95% CI of the AUROC between the two subgroups indicated a statistically significant difference. Publication bias was assessed using the Deeks' funnel plot and a P value of <0.1 indicated the asymmetry of the funnel plot. $P < 0.05$ was considered statistically significant.

3 | RESULTS

3.1 | Characteristics of the published studies and assessment of the risk of bias

A total of 1537 articles were identified by the primary literature search. After screening the titles and abstracts, 1485 studies were excluded due to duplications (n = 1027) or irrelevant to the current analysis (n = 458). Fifty-two potentially relevant articles were then retrieved for further review, of which 36 were excluded due to insufficient data (n = 18), not analyzed based on endoscopic image (n = 4), or animal studies (n = 14). Finally, a total of 16 full-text manuscripts[12-27] were enrolled for the meta-analysis. The flowchart of study enrollment is shown in Figure 1.

The results of QUADAS-2 showed that the risk for patient selection was unclear in six studies,[14,20,22,25-27] as shown in Figure 2, and the methodological quality was generally high.

3.2 | Study features

Among the 16 studies endoscopic images or video clips were collected retrospectively as the test dataset in 13 studies[12-19,21,23-25,27] and prospectively in three studies.[20,22,26] Only two studies[20,27] used AI on real patients with non-dysplastic BE and confirmed Barrett's neoplasia. Most studies were conducted in Asian populations,[12-18,23-25] while in the other six studies[19-22,26,27] endoscopic images were obtained from Western populations.

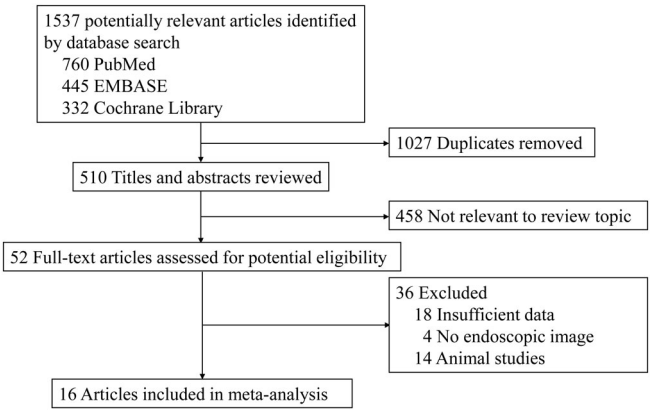

FIGURE 1 Flowchart of study selection for the systematic review and meta-analysis

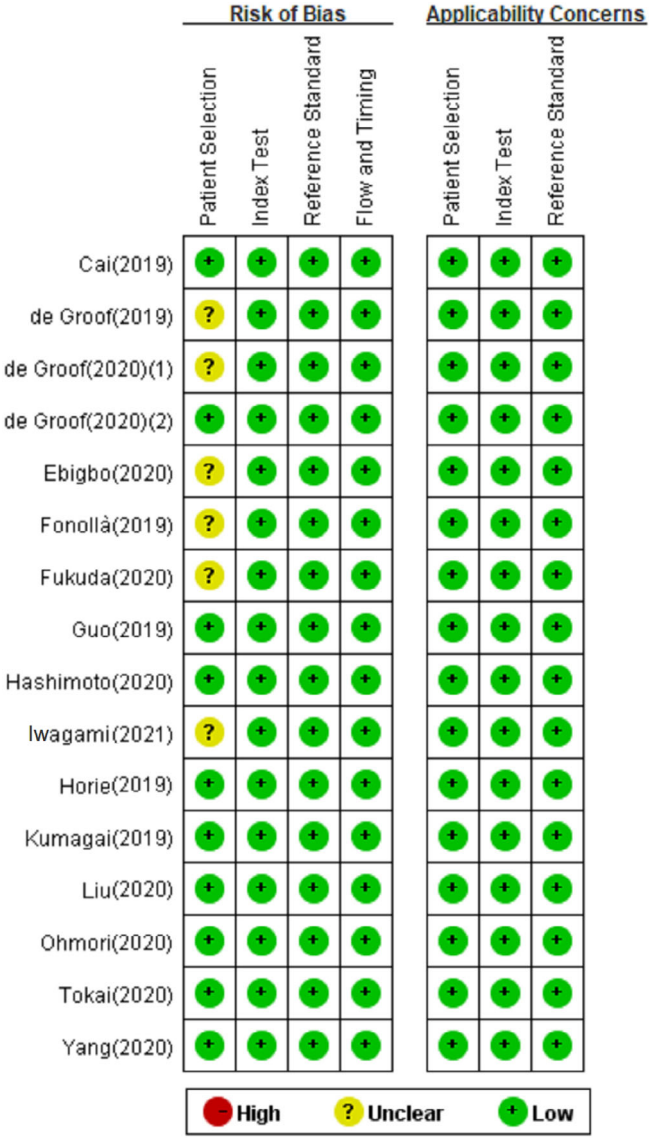

FIGURE 2 Methodological quality assessment

TABLE 1 Characteristics of the selected studies

First author (publication year)	Study type	Country/ region	Histological type	Images (n)	Patients (n)	Endoscopists (n)	Endoscopy type	Algorithm type	Real-time	Videoclips	External validation	Low-quality images	Performance of AI-assisted system			
													TP	FP	FN	TN
Ohmori[12] (2020)	Retrospective	Japan	ESCC	135	102	15	WLI/NBI/ BLI	CNN	No	No	No	No	201	89	8	177
Guo[13] (2020)	Retrospective	China	ESCC	6671	NM	NM	NBI	CNN	Yes	Yes	Yes	No	1451	258	29	4933
Fukuda[14] (2020)	Retrospective	Japan	ESCC	238	NM	13	NBI/BLI	CNN	Yes	Yes	Yes	No	80	53	10	95
Tokai[15] (2020)	Retrospective	Japan	ESCC	279	NM	13	NBI/WLI	CNN	No	No	No	No	159	24	30	66
Kumagai[16,#] (2019)	Retrospective	Japan	ESCC	1520	55	NM	ECS	CNN	No	No	No	NM (possible)	25	3	2	25
Liu[17,*] (2020)	Retrospective	China	ESCC/EAC	127	NM	NM	WLI	CNN	No	No	No	No	71	3	4	49
Horie[18] (2019)	Retrospective	Japan	ESCC/EAC	1118	97	NM	NBI/WLI	CNN	Yes	No	No	No	46	42	1	8
Hashimoto[19] (2020)	Retrospective	USA	BE	458	39	NM	WLI/NBI	CNN	Yes	No	No	No	217	13	8	220
de Groof[20,#] (2020) (1)	Prospective	Europe	BE	144	20	NM	WLI	CNN	Yes	Yes	Yes	NM (possible)	9	3	1	7
de Groof[21] (2020)(2)	Retrospective	Europe	BE	457	255	N	WLI	SVM	Yes	No	Yes	No	186	31	23	217
de Groof[22] (2019)	Prospective	Europe	BE	N	60	10	WLI	SVM	Yes	No	No	NM (possible)	38	3	2	17
Cai[23] (2019)	Retrospective	China	ESCC	187	52	16	WLI	CNN	No	No	No	NM (possible)	89	14	2	82
Yang[24] (2020)	Retrospective	China	ESCC	1203	NM	6	WLI	CNN	Yes	Yes	Yes	Yes	409	13	7	774
Iwagami[25,#] (2021)	Retrospective	Japan	E/J AC	232	79	15	NBI/WLI	CNN	Yes	No	Yes	No	34	25	2	18
Fonollà[26] (2019)	Prospective	Europe	BE	141	NM	NM	VLE	CNN	No	No	No	No	37	12	5	87
Ebigbo[27] (2020)	Retrospective	Europe	BE	62	14	NM	WLI	CNN	Yes	No	No	NM (possible)	30	0	6	26

Note: *Precancerous lesions included in the EC lesions. #Per-case analysis. Abbreviations: AI, artificial intelligence; BE, Barrett's esophagus; BLI, blue-laser imaging; CNN, convolutional neural network; EAC, esophageal adenocarcinoma; ECS, endocytoscopic system images; E/J AC, esophageal and esophagogastric junctional adenocarcinoma; ESCC, esophageal squamous cell cancer; FN, false negative; FP, false positive; FN, false negative; NM, not mentioned; SVM, support vector machine; TN, true negative; TP, true positive; VLE, volumetric laser endomicroscopy; WLI, white light imaging.

Convolutional neural networks (CNN) served as the backbone of the AI-assisted model in 14 studies,[12-20,23-27] while the other two studies[21,22] employed a support vector machine (SVM). The training datasets of seven studies[17,20-24,27] included only white-light imaging (WLI), whereas those of seven studies[12-15,18,19,25] included narrow-band imaging (NBI). One study[16] constructed an AI-assisted model based on endocytoscopic system images of ESCC and the other[26] used volumetric laser endomicroscopy (VLE) images of BE neoplasia to train the AI model. Two studies[15,18] focused mainly on evaluating invasive depth of the esophageal lesions with the assistance of AI. The training datasets could also be histologically classified into ESCC in seven studies,[12-16,23,24] EAC and BE in seven studies,[19-22,25-27] or ESCC and EAC in two studies. [17,18] A comparison between the diagnostic accuracy of AI-assisted models and that of endoscopists was performed in seven studies.[12,14,15,22-25] Low-quality images were excluded from the datasets in 10 studies;[12-15,17-19,21,25,26] one study[24] used both low-quality and high-quality images, and the other five studies[16,20,22,23,27] did not mention whether low-quality images were used or not; these five studies[16,20,22,23,27] and the one with definite low-quality image use[24] were included as the possible or definite use of low-quality images group for analysis. Among these 16 studies, four[13,14,20,24] included video clips to train the AI-assisted models and 10 studies[13,14,18-22,24,25,27] used real-time diagnosis. In all the studies the diagnosis was confirmed by pathology. The identified studies and their characteristics are listed in Table 1.

3.3 | Main analysis

The pooled sensitivity, specificity, PLR, NLR, DOR and AUROC of the AI-assisted model for the diagnosis of esophageal neoplasms were 94% (95% CI 92%-96%), 85% (95% CI 73%-92%), 6.40 (95% CI 3.38-12.11), 0.06 (95% CI 0.04-0.10), 98.88 (95% CI 39.45-247.87) and 0.97 (95% CI 0.95-0.98), respectively (Figures 3-8).

As mentioned above, seven studies[12,14,15,22-25] compared the diagnostic accuracy of AI-assisted model to that of the endoscopists. And the diagnostic efficacy of endoscopists in detecting esophageal neoplasms was reported in four studies.[12,14,15,23] The pooled sensitivity, specificity and AUROC were 82% (95% CI 77%-86%) (Figure 9), 79% (95% CI 66%-88%) (Figure S1) and 0.86 (95% CI 0.82-0.88) (Figure S2), respectively. The sensitivity of the AI-assisted system was higher than that of the endoscopists in detecting esophageal neoplasms using the same datasets (94% [95% CI 84%-98%] vs 82% [95% CI 77%-86%], $P < 0.01$). However, the specificity did not differ between the AI-assisted model and the endoscopists ($P = 0.49$).

The AUROC was calculated and comparisons were performed in the subgroups with at least four studies. Studies reported video clips used for the datasets had a slightly higher AUROC than those using still images only as datasets (0.98 [95% CI 0.97-0.99] vs 0.96 [95% CI 0.94-0.97]). Additionally, the AUROC of AI-assisted system with a possible or definite use of low-quality images was higher than those using high-quality images only (0.98 [95% CI 0.97-0.99] vs 0.96 [95% CI: 0.93-0.97]) (Table 2).

3.4 | Assessment of heterogeneity

The SROC curve was symmetric (Figure 8). The correlation coefficient between the logit-transformed sensitivity and specificity was negative ($r = -0.05$) and the asymmetric β parameter presented insignificancy ($P = 0.85$), which implied that there was no heterogeneity among the included studies.

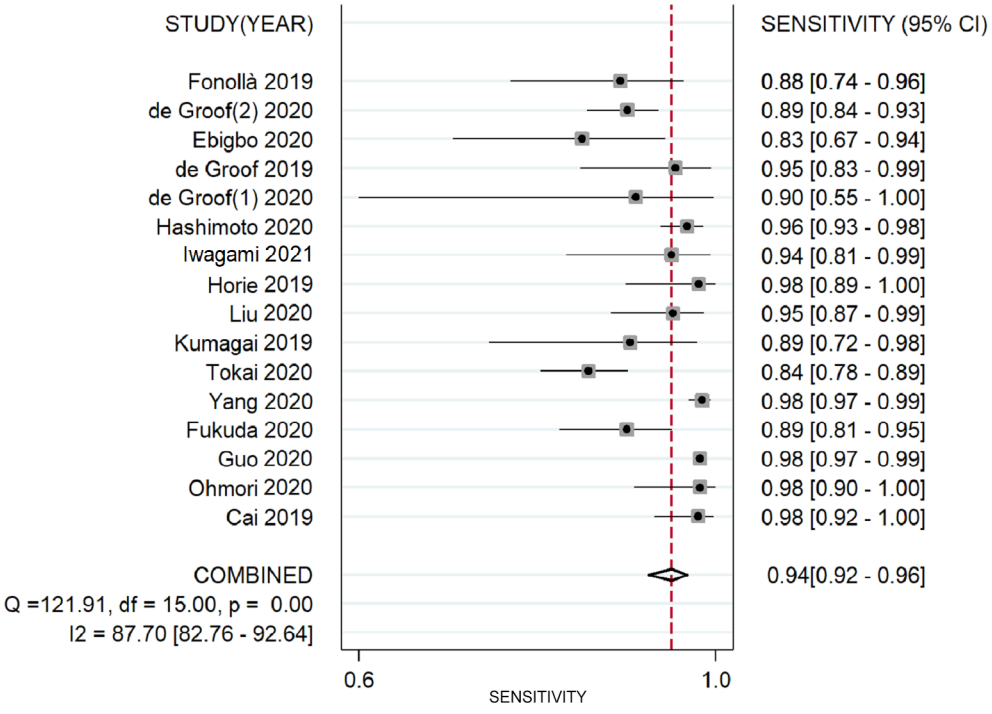

FIGURE 3 Pooled sensitivity of artificial intelligence-assisted model for the diagnosis of esophageal neoplasms [Color figure can be viewed at wileyonlinelibrary.com]

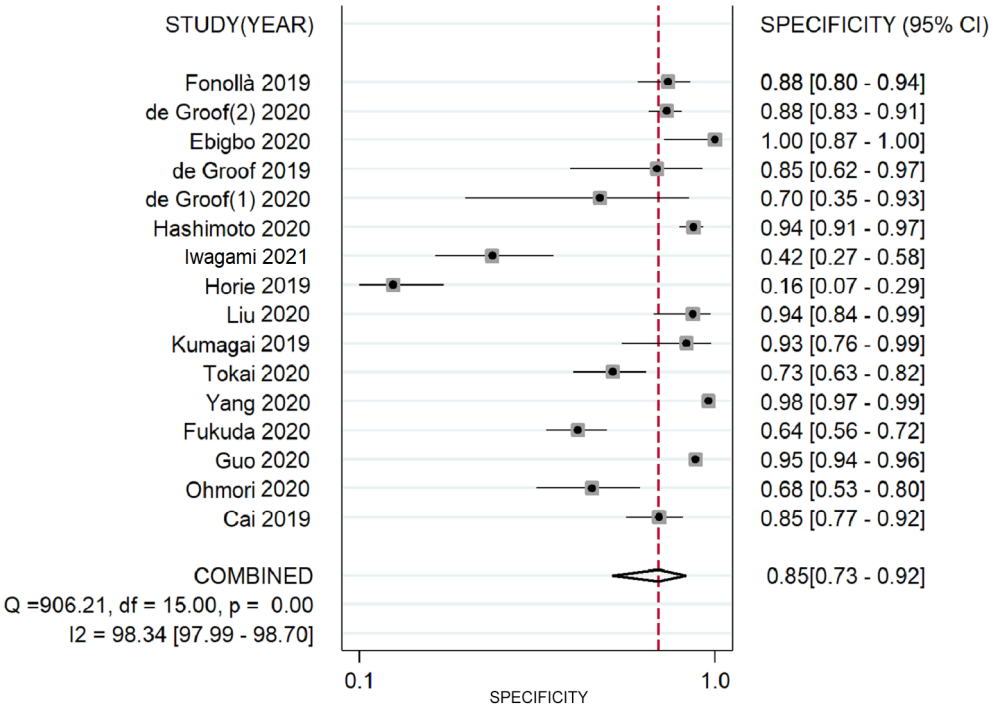

FIGURE 4 Pooled specificity of artificial intelligence-assisted diagnostic model for the diagnosis of esophageal neoplasms [Color figure can be viewed at wileyonlinelibrary.com]

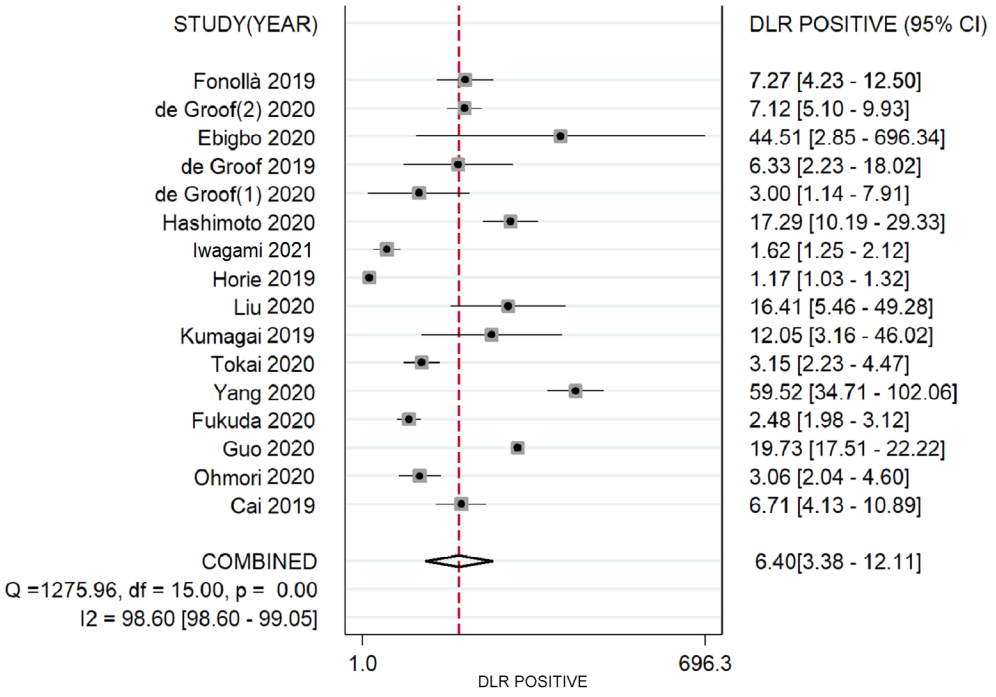

FIGURE 5 Pooled positive likelihood ratio of artificial intelligence-assisted model for the diagnosis of esophageal neoplasms [Color figure can be viewed at wileyonlinelibrary.com]

The results of the meta-regression analysis conducted to explore the possible sources of heterogeneity showed no potential risk factors for heterogeneity (*P* value: 0.84 for study type, 0.14 for region, 0.59 for histological type, 0.77 for algorithm type, 0.57 for the type of endoscopy, 0.39 for video clips, 0.23 for external validation dataset, 0.80 for real-time diagnosis, 0.23 for possible use of low-quality images, and 0.66 for study quality). These results are summarized in Table 2.

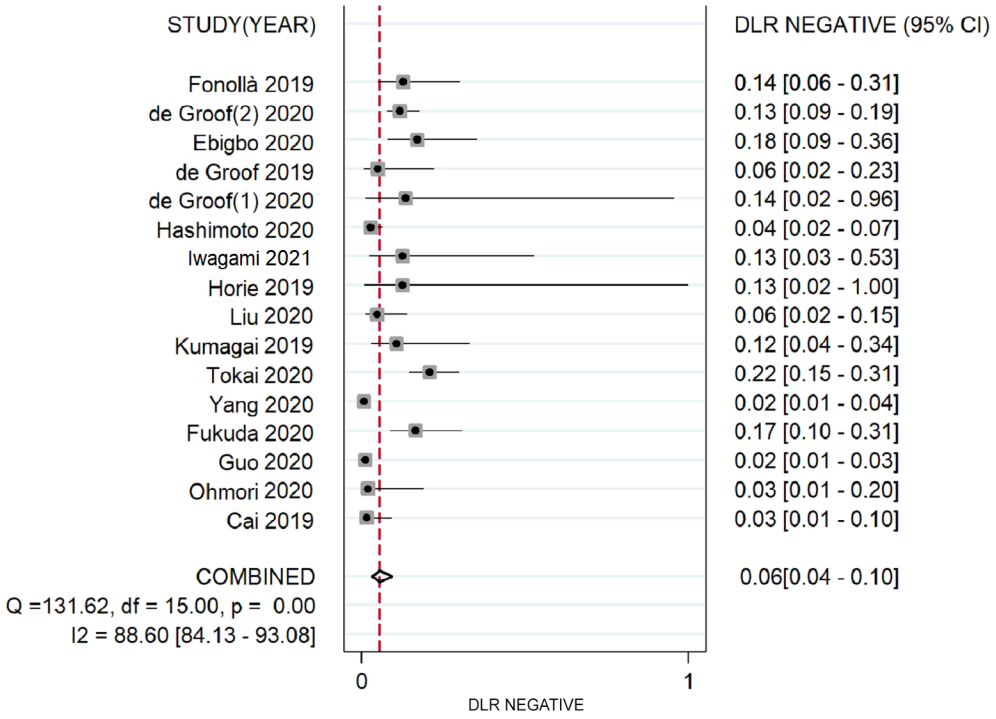

FIGURE 6 Pooled negative likelihood ratio of artificial intelligence-assisted model for the diagnosis of esophageal neoplasms [Color figure can be viewed at wileyonlinelibrary.com]

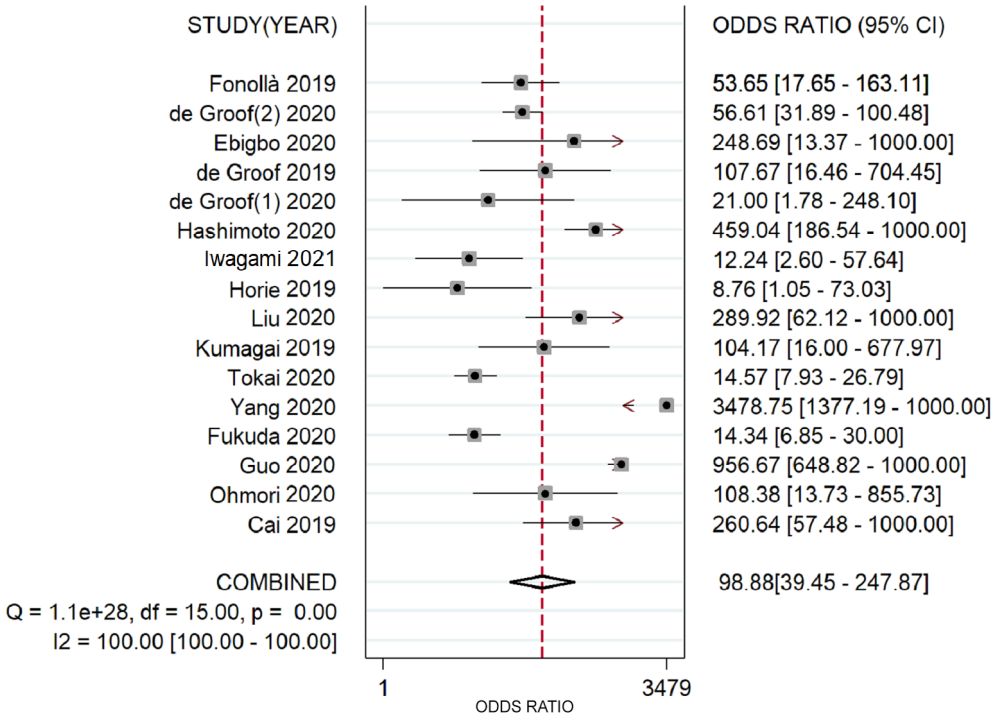

FIGURE 7 Pooled diagnostic odds ratios of artificial intelligence-assisted model for the diagnosis of esophageal neoplasms [Color figure can be viewed at wileyonlinelibrary.com]

3.5 | Publication bias

The Deeks' funnel plot asymmetry test indicated the existence of a publication bias in the included studies ($P < 0.05$; Figure 10).

4 | DISCUSSION

With the rapid development of computer algorithms AI has been increasingly used to improve diagnostic accuracy and identify invasive

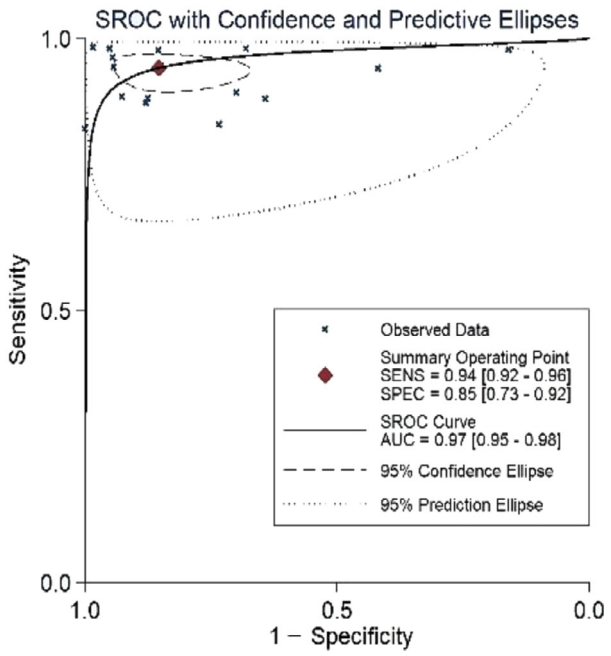

FIGURE 8 Summary receiver operating characteristic of artificial intelligence-assisted system for the diagnosis of esophageal neoplasms

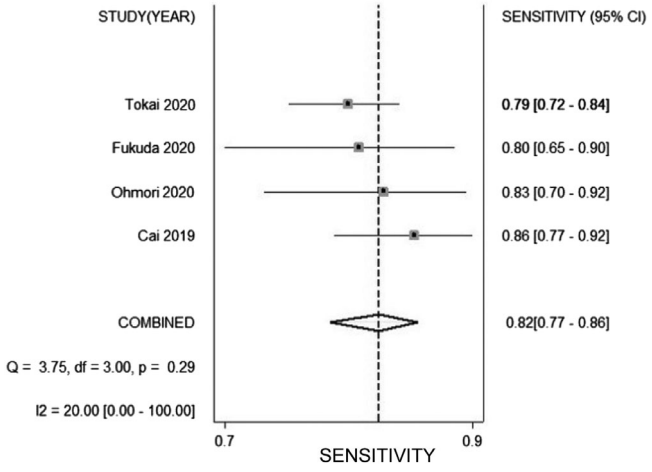

FIGURE 9 Pooled sensitivity of endoscopists for the diagnosis of esophageal neoplasms

depth of the gastrointestinal lesions on endoscopy. A number of studies on AI-assisted diagnostic models have recently been published, which may provide novices with additional assistance in the identification and diagnosis of esophageal cancer. One study[24] demonstrated that the diagnostic performance of novices was considerably improved with the assistance of AI, suggesting that AI may be a practical approach to enhance the diagnostic rate of esophageal cancer and help endoscopists identify precisely invasive depth of the lesions. As for epithelial-submucosal 1 (EP-SM1) lesions, endoscopic resection is recommended due to their relatively low risk of lymph node metastasis (<10%).[28-30] However, SM2-SM3 lesions, which are at a high risk of lymph node metastasis (>25%), should be treated with esophagectomy or chemoradiotherapy.[28,29,31] Therefore, it is important to identify the invasive depth of cancerous lesions accurately. One study[15] reported that the AI-based diagnostic system achieved a better diagnostic accuracy for invasive depth in ESCC than endoscopists, especially for EP-SM1 lesions. Furthermore, AI has also been used in the indirect classification of invasive depth of the lesions by identifying intrapapillary capillary loops, which is a special morphological feature of early esophageal carcinoma,[32] and has shown promising performance.[33] In terms of advanced imaging techniques, AI-assisted models have shown considerably good performance in detecting esophageal neoplasms on endocytoscopic system[16] and volumetric laser endomicroscopy images.[26]

The current meta-analysis demonstrated that AI had excellent accuracy in the detection of esophageal neoplasms on endoscopic images, which was generally comparable to that of the endoscopists. A subgroup analysis further demonstrated that use of video clips as a part of the training and validation datasets might contribute to a higher AUROC of the AI-assisted model compared with those using still images alone (0.98 [95% CI 0.97-0.99] vs 0.96 [95% CI 0.94-0.97]). And the AUROC of AI-assisted diagnostic models with a possible use of low-quality images was higher than those using high-quality images only (0.98 [95% CI 0.97-0.99] vs 0.96 [95% CI 0.93-0.97]). Therefore, incorporating more low-quality images into the datasets may be a way to improve the diagnostic accuracy of AI models.

To the best of our knowledge, our study is the first systematic review and meta-analysis to evaluate the comprehensive effectiveness of AI and to compare its accuracy with that of endoscopists in diagnosing esophageal neoplasms on endoscopy. Some meta-analyses have assessed the effectiveness of AI-assisted system in diagnosing esophageal lesions. Lui et al[34] calculated the pooled sensitivity, specificity and AUROC of AI in diagnosing ESCC, BE or EAC, and compared the performance of AI with that of endoscopists. However, with many studies adopting both ESCC and EAC as datasets, a comprehensive result regardless of the histological type of the esophageal lesions should also be generated. Bang et al[35] performed a meta-analysis to evaluate the accuracy of computer-aided diagnosis of esophageal cancer and neoplasms on endoscopy. However, due to insufficient data they could not compare the effectiveness of computer-aided algorithms with that of endoscopic physicians. In another meta-analysis[36] the metrics of pooled sensitivity, specificity, PLR, NLR, DOR and AUROC for AI could not be obtained due to significant heterogeneity of the included studies and those based on SVM were excluded. One study[34] concluded that the use of NBI resulted in a higher AUROC of the AI models than the use of WLI in the diagnosis of ESCC, while in our study we found no statistical difference in AUROC between the use of WLI and NBI. The reason may lie in the heterogeneity of different study inclusions and use of varying statistical methods, etc. Thus, further overall meta-analyses should be conducted to assess the accuracy of AI models based on WLI and NBI. We also confirmed that including video clips in the training and test datasets may improve the

TABLE 2 Subgroup analysis and meta-regression

Subgroups	Studies (n)	SEN (95% CI)	SPE (95% CI)	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)	AUROC (95% CI)	*P* value for heterogeneity
Region								.14
West	6	0.92 (0.87-0.95)	0.90 (0.86-0.93)	9.09 (6.19-13.36)	0.09 (0.06-0.16)	96.37 (42.71-217.43)	0.96 (0.94-0.97)	
Asia	10	0.96 (0.92-0.98)	0.81 (0.61-0.92)	5.06 (2.18-11.73)	0.06 (0.03-0.11)	92.25 (24.26-350.83)	0.97 (0.95-0.98)	
Study type								.84
Retrospective	13	0.95 (0.92-0.97)	0.86 (0.71-0.94)	6.68 (3.05-14.65)	0.06 (0.04-0.10)	111.64 (37.63-331.24)	0.97 (0.95-0.98)	
Prospective	3	0.91 (0.84-0.96)	0.86 (0.79-0.92)	5.75 (3.46- 9.54)	0.11 (0.06-0.22)	55.55 (22.76-135.56)	0.94	
Algorithm type								.77
CNN	14	0.95 (0.92-0.97)	0.852 (0.71-0.93)	6.35 (3.06-13.20)	0.06 (0.04-0.10)	104.32 (36.85-295.28)	0.97 (0.95-0.98)	
SVM	2	-	-	-	-	-	-	
Endoscopy type								.57
WLI	7	0.95 (0.90-0.97)	0.92 (0.84-0.96)	12.24 (5.75-26.09)	0.06 (0.03-0.12)	203.72 (59.96-692.14)	0.98 (0.96-0.99)	
Including NBI	7	0.95 (0.91-0.98)	0.71 (0.44-0.89)	3.30 (1.47-7.40)	0.07 (0.03-0.15)	48.22 (11.28-206.15)	0.95 (0.93-0.97)	
Histological type								.59
ESCC	7	0.95 (0.91-0.98)	0.87 (0.74-0.94)	7.54 (3.37-16.89)	0.075 (0.03-0.12)	138.33 (30.53-626.78)	0.97 (0.950.98)	
EAC/BE	7	0.92 (0.88-0.95)	0.87 (0.71-0.95)	6.91 (2.97-16.09)	0.09 (0.06-0.14)	77.76 (27.68-218.43)	0.95 (0.93-0.97)	
ESCC/EAC	2	-	-	-	-	-	-	
Real-time								.80
Yes	10	0.95 (0.92- 0.97)	0.85 (0.64-0.95)	6.38 (2.37-17.18)	0.06 (0.03- 0.11)	109.75 (28.35-424.91)	0.97 (0.95-0.98)	
No	6	0.93 (0.87-0.97)	0.85 (0.76- 0.91)	6.09 (3.74-9.91)	0.08 (0.04-0.16)	76.78 (30.13-195.63)	0.95 (0.93-0.97)	
Video clips								.39
Yes	4	0.96 (0.91-0.99)	0.90 (0.68-0.97)	9.25 (2.51-34.13)	0.04 (0.01-0.13)	223.12 (20.51-2427.32)	0.98 (0.97-0.99)	
No	12	0.94 (0.90-0.96)	0.80 (0.68-0.92)	5.64 (2.77-11.48)	0.08 (0.05- 0.12)	73.75 (31.42-173.14)	0.96 (0.94-0.97)	
External validation								.23
Yes	6	0.95 (0.89-0.97)	0.85 (0.63-0.95)	6.28 (2.18-18.04)	0.06 (0.03-0.15)	98.00 (15.41-623.34)	0.97 (0.95-0.98)	
No	10	0.94 (0.90-0.97)	0.86 (0.70-0.94)	6.61 (2.89-15.12)	0.07 (0.04-0.12)	95.76 (36.50-51.27)	0.96 (0.94-0.98)	
Using low-quality images								.23
Possible or definite	6	0.95 (0.89-0.98)	0.93 (0.83-0.98)	13.93 (5.15-37.68)	0.05 (0.02-0.12)	261.81 (62.36-1099.11)	0.98 (0.97-0.99)	
No	10	0.94 (0.90-0.97)	0.79 (0.60-0.90)	4.41 (2.18-8.91)	0.07 (0.04-0.13)	50.39 (19.99-176.48)	0.96 (0.93-0.97)	
Quality								.66
High	10	0.96 (0.93-0.98)	0.87 (0.73-0.95)	7.53 (3.26-17.40)	0.05 (0.03-0.09)	154.10 (45.67- 519.98)	0.98 (0.96-0.99)	
Low	6	0.90 (0.85-0.93)	0.80 (0.57-0.92)	4.49 (1.90-10.64)	0.13 (0.09-0.19)	35.08 (12.76-96.43)	0.92 (0.89-0.94)	

Abbreviations: AUROC, area under the summary receiver operating characteristic curve; BE, Barrett's esophagus; CI, confidence interval; CNN, convolutional neural network; DOR, diagnostic odds ratio; EAC, esophageal adenocarcinoma; ESCC, esophageal squamous cell cancer; NBI, narrow-band imaging; NLR, negative likelihood ratio; PLR, positive likelihood ratio; SEN, sensitivity; SPE, specificity; SVM, support vector machine; WLI, white-light imaging.

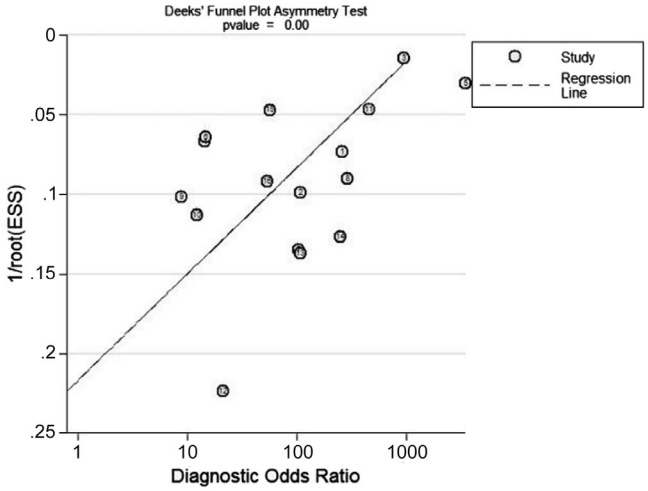

FIGURE 10 Deeks' plot of publication bias

accuracy of AI, as studies containing video clips had a higher AUROC than those only using still images (0.98 [95% CI 0.97-0.99] vs 0.96 [95% CI 0.94-0.97]), which is in line with the previously reported higher accuracy of AI in colonoscopy using video clips.[37] This may be explained by the larger amount of information provided by a series of video frames, showing the appearance of esophageal lesions from various angles, positions and sizes.

There were some limitations to this study. First, a significant publication bias was identified in the included studies. The increasing proportion of positive results in the medical literatures has entailed a decline in the scale of negative results.[38] Therefore, there is a need to accumulate data involving both positive and negative results to confirm the validity of this meta-analysis. Second, due to insufficient data we were not able to perform subgroup analysis on the AUROC between using CNN and SVM. However, it has been shown that CNN, which represents a domain part in the field of deep learning,[39] is more suitable than any other algorithm in terms of diagnosis.[40] As a result, we can reasonably speculate that the AUROC of studies with CNN may be higher than that with SVM. Third, due to the lack of data we could not evaluate the pooled sensitivity, specificity, PLR, NLR, DOR and AUROC of the classification of EP-SM1 and SM2-SM3 lesions. Additional studies aiming at distinguishing the invasive depth of esophageal lesions are therefore needed.[32] Fourth, one study each included in our study used ECS and volumetric laser endomicroscopy images as datasets, which were excluded from the subgroup analysis. Fifth, among the 16 studies, one found apparent group-based diagnostic discrepancies among senior, mid-level and junior endoscopists, and the accuracy of AI models was similar to that of the endoscopists in the senior endoscopist group.[23] Furthermore. another study demonstrated that the diagnostic accuracy of AI models was similar to that of experienced endoscopists and higher than that of novices.[24] However, due to the limited number of included studies we could not compare the pooled diagnostic accuracy of AI models with that of expert, medium experienced and novice endoscopists. Finally, there are so far few prospective studies that validate the diagnostic accuracy of AI in real-time clinical setting. Therefore, our results may not genuinely reflect the performance of AI in actual patients. Further randomized prospective clinical trials for improvement and validation are expected[7] to address the accuracy gap between experimental scenarios and clinical practice.

5 | CONCLUSIONS

In conclusion, AI-assisted systems showed high accuracy in detecting esophageal cancer, which was comparable to that of endoscopists. However, because most studies included in the current systematic review and meta-analysis were retrospectively designed, large prospective studies are needed to further validate our results.

CONFLICT OF INTEREST

None.

ORCID

Si Min Zhang https://orcid.org/0000-0002-6684-1676
Shu Tian Zhang https://orcid.org/0000-0003-2356-4397

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(3):209–249.
2. Huang FL, Yu SJ. Esophageal cancer: risk factors, genetic association, and treatment. *Asian J Surg.* 2018;41(3):210-215.
3. Lin Y, Totsuka Y, Shan B, et al. Esophageal cancer in high-risk areas of China: research progress and challenges. *Ann Epidemiol.* 2017;27(3): 215-221.
4. Klingelhöfer D, Zhu Y, Braun M, Brüggmann D, Schöffel N, Groneberg DA. A world map of esophagus cancer research: a critical accounting. *J Transl Med.* 2019;17(1):150. https://doi.org/10.1186/s12967-019-1902-7.
5. Codipilly DC, Qin Y, Dawsey SM, et al. Screening for esophageal squamous cell carcinoma: recent advances. *Gastrointest Endosc.* 2018; 88(3):413-426.
6. de Souza LA Jr, Palm C, Mendel R, et al. A survey on Barrett's esophagus analysis using machine learning. *Comput Biol Med.* 2018;96: 203-213.
7. Thakkar SJ, Kochhar GS. Artificial intelligence for real-time detection of early esophageal cancer: another set of eyes to better visualize. *Gastrointest Endosc.* 2020;91(1):52-54.
8. Lazăr DC, Avram MF, Faur AC, et al. The impact of artificial intelligence in the endoscopic assessment of premalignant and malignant esophageal lesions: present and future. *Medicina (Kaunas).* 2020;56(7): 364. https://doi.org/10.3390/medicina56070364.
9. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34.
10. Whiting PF, Rutjes AWS, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536.

11. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31. https://doi.org/10.1186/1471-2288-6-31.
12. Ohmori M, Ishihara R, Aoyama K, et al. Endoscopic detection and differentiation of esophageal lesions using a deep neural network. *Gastrointest Endosc*. 2020;91(2):301-309.e1.
13. Guo LJ, Xiao X, Wu CC, et al. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest Endosc*. 2020;91(1):41-51.
14. Fukuda H, Ishihara R, Kato Y, et al. Comparison of performances of artificial intelligence versus expert endoscopists for real-time assisted diagnosis of esophageal squamous cell carcinoma (with video). *Gastrointest Endosc*. 2020;92(4):848-855.
15. Tokai Y, Yoshio T, Aoyama K, et al. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. *Esophagus*. 2020;17(3):250-256.
16. Kumagai Y, Takubo K, Kawada K, et al. Diagnosis using deep-learning artificial intelligence based on the endocytoscopic observation of the esophagus. *Esophagus*. 2019;16(2):180-187.
17. Liu G, Hua J, Wu Z, et al. Automatic classification of esophageal lesions in endoscopic images using a convolutional neural network. *Ann Transl Med*. 2020;8(7):486. https://doi.org/10.21037/atm.2020.03.24.
18. Horie Y, Yoshio T, Aoyama K, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc*. 2019;89(1):25-32.
19. Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest Endosc*. 2020;91(6):1264-1271.e1.
20. de Groof AJ, Struyvenberg MR, Fockens KN, et al. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointest Endosc*. 2020;91(6):1242-1250.
21. de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology*. 2020;158(4):915-929.e4.
22. de Groof J, van der Sommen F, van der Putten J, et al. The Argos project: the development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterol J*. 2019;7(4):538-547.
23. Cai SL, Li B, Tan WM, et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest Endosc*. 2019;90(5):745-753.e2.
24. Yang XX, Zhen L, Shao XJ, et al. Real-time artificial intelligence for endoscopic diagnosis of early esophageal squamous cell cancer (with video) [Epub ahead of print December 4, 2020]. *Dig Endosc*. https://doi.org/10.1111/den.13908. Epub ahead of print.
25. Iwagami H, Ishihara R, Aoyama K, et al. Artificial intelligence for the detection of esophageal and esophagogastric junctional adenocarcinoma. *J Gastroenterol Hepatol*. 2021;36(1):131-136.
26. Fonollà R, Scheeve T, Struyvenberg MR, et al. Ensemble of deep convolutional neural networks for classification of early Barrett's neoplasia using volumetric laser endomicroscopy. *Appl Sci*. 2019;9:2183. https://doi.org/10.3390/app9112183.
27. Ebigbo A, Mendel R, Probst A, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut*. 2020;69(4):615-616.
28. Yamashina T, Ishihara R, Nagai K, et al. Long-term outcome and metastatic risk after endoscopic resection of superficial esophageal squamous cell carcinoma. *Am J Gastroenterol*. 2013;108(4):544-551.
29. Akutsu Y, Uesato M, Shuto K, et al. The overall prevalence of metastasis in T1 esophageal squamous cell carcinoma: a retrospective analysis of 295 patients. *Ann Surg*. 2013;257(6):1032-1038.
30. Hölscher AH, Bollschweiler E, Schröder W, Metzger R, Gutschow C, Drebber U. Prognostic impact of upper, middle, and lower third mucosal or submucosal infiltration in early esophageal cancer. *Ann Surg*. 2011;254(5):802-808.
31. Kitagawa Y, Uno T, Oyama T, et al. Esophageal cancer practice guidelines 2017 edited by the Japan Esophageal Society: part 1. *Esophagus*. 2019;16(1):1-24.
32. Kumagai Y, Kawada K, Yamazaki S, et al. Prospective replacement of magnifying endoscopy by a newly developed endocytoscope, the 'GIF-Y0002'. *Dis Esophagus*. 2010;23(8):627-632.
33. Everson M, Herrera L, Li W, et al. Artificial intelligence for the real-time classification of intrapapillary capillary loop patterns in the endoscopic diagnosis of early oesophageal squamous cell carcinoma: a proof-of-concept study. *United European Gastroenterol J*. 2019;7(2):297-306.
34. Lui TKL, Tsui VWM, Leung WK. Accuracy of artificial intelligence-assisted detection of upper GI lesions: a systematic review and meta-analysis. *Gastrointest Endosc*. 2020;92(4):821-830.e9.
35. Bang CS, Lee JJ, Baik GH. Computer-aided diagnosis of esophageal cancer and neoplasms in endoscopic images: a systematic review and meta-analysis of diagnostic test accuracy. *Gastrointest Endosc*. 2021;93(5):1006-1015.e13.
36. Mohan BP, Khan SR, Kassab LL, Ponnada S, Dulai PS, Kochhar GS. Accuracy of convolutional neural network-based artificial intelligence in diagnosis of gastrointestinal lesions based on endoscopic images: a systematic review and meta-analysis. *Endosc Int Open*. 2020;8(11):E1584-E1594.
37. Misawa M, Kudo SE, Mori Y, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology*. 2018;154(8):2027-2029.e3.
38. Mlinarić A, Horvat M, Šupak Smolčić V. Dealing with the positive publication bias: why you should really publish your negative results. *Biochem Med (Zagreb)*. 2017;27(3):030201. https://doi.org/10.11613/BM.2017.030201.
39. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol*. 2019;29(7):R231-R236.
40. Ebigbo A, Palm C, Probst A, et al. A technical review of artificial intelligence as applied to gastrointestinal endoscopy: clarifying the terminology. *Endosc Int Open*. 2019;7(12):E1616-E1623.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang SM, Wang YJ, Zhang ST. Accuracy of artificial intelligence-assisted detection of esophageal cancer and neoplasms on endoscopic images: A systematic review and meta-analysis. *J Dig Dis*. 2021;22(6):318–328. https://doi.org/10.1111/1751-2980.12992