Systematic Review/Meta-Analysis

Convolutional neural network-based artificial intelligence for the diagnosis of early esophageal cancer based on endoscopic images: A meta-analysis

Hongbiao Ma, Longlun Wang[1], Yilin Chen, Lu Tian[1]

Department of Thoracic Surgery, Chongqing General Hospital, No.118, Xingguang Avenue, Liangjiang New Area, Chongqing, China, [1]Department of Radiology, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Pediatrics, Chongqing, China

Abstract

Background: Early screening and treatment of esophageal cancer (EC) is particularly important for the survival and prognosis of patients. However, early EC is difficult to diagnose by a routine endoscopic examination. Therefore, convolutional neural network (CNN)-based artificial intelligence (AI) has become a very promising method in the diagnosis of early EC using endoscopic images. The aim of this study was to evaluate the diagnostic performance of CNN-based AI for detecting early EC based on endoscopic images.

Methods: A comprehensive search was performed to identify relevant English articles concerning CNN-based AI in the diagnosis of early EC based on endoscopic images (from the date of database establishment to April 2022). The pooled sensitivity (SEN), pooled specificity (SPE), positive likelihood ratio (LR+), negative likelihood ratio (LR−), diagnostic odds ratio (DOR) with 95% confidence interval (CI), summary receiver operating characteristic (SROC) curve, and area under the curve (AUC) for the accuracy of CNN-based AI in the diagnosis of early EC based on endoscopic images were calculated. We used the I^2 test to assess heterogeneity and investigated the source of heterogeneity by performing meta-regression analysis. Publication bias was assessed using Deeks' funnel plot asymmetry test.

Results: Seven studies met the eligibility criteria. The SEN and SPE were 0.90 (95% confidence interval [CI]: 0.82–0.94) and 0.91 (95% CI: 0.79–0.96), respectively. The LR+ of the malignant ultrasonic features was 9.8 (95% CI: 3.8–24.8) and the LR− was 0.11 (95% CI: 0.06–0.21), revealing that CNN-based AI exhibited an excellent ability to confirm or exclude early EC on endoscopic images. Additionally, SROC curves showed that the AUC of the CNN-based AI in the diagnosis of early EC based on endoscopic images was 0.95 (95% CI: 0.93–0.97), demonstrating that CNN-based AI has good diagnostic value for early EC based on endoscopic images.

Conclusions: Based on our meta-analysis, CNN-based AI is an excellent diagnostic tool with high sensitivity, specificity, and AUC in the diagnosis of early EC based on endoscopic images.

Keywords: Artificial Intelligence, convolutional neural network, early esophageal cancer, endoscopic, meta-analysis

Address for correspondence: Dr. Lu Tian, Department of Radiology, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Pediatrics, Chongqing, 400014, China.
E-mail: 195792179@qq.com

Submitted: 21-Apr-2022 **Revised:** 31-May-2022 **Accepted:** 03-Jun-2022 **Published:** 08-Jul-2022

Access this article online	
Quick Response Code:	Website: www.saudijgastro.com
	DOI: 10.4103/sjg.sjg_178_22

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Ma H, Wang L, Chen Y, Tian L. Convolutional neural network-based artificial intelligence for the diagnosis of early esophageal cancer based on endoscopic images: A meta-analysis. Saudi J Gastroenterol 2022;28:332-40.

INTRODUCTION

Esophageal cancer (EC) is the eighth most common malignant tumor in the world and the sixth in terms of mortality.[1-3] It has characteristics of high invasiveness and lymph node metastasis.[4,5] The early signs and symptoms of EC are often latent and nonspecific. Ninety percent of EC patients would have reached the middle and advanced stage when they are diagnosed with it, which leads to poor prognosis and high costs. The 5-year survival rate does not exceed 20%.[6-8] Therefore, early screening and treatment of EC is particularly important for the survival and prognosis of patients, which has always been a hot spot for clinical medical research. Early detection of EC by endoscopy is a widely adopted strategy to prevent cancer-related morbidity and mortality.[9-11] White-light endoscopy (WLE) and narrow-band imaging (NBI) are the most common techniques for detecting EC.[12,13] However, endoscopic features of these early lesions are subtle and easily missed with conventional endoscopy. With the development of endoscopic technology, the appearance of high-definition WLE with or without chromoendoscopy, NBI with or without magnification, confocal laser endomicroscopy, and endocytoscopic imaging system improved the detection rate of lesions.[14-16] But the diagnosis rate is still low and detection of these subtle changes in endoscopic images relies on the expertise of endoscopists, and is inevitably affected by differences in their experience. The emergence of convolutional neural network (CNN)-based artificial intelligence (AI) has brought hope to solve this problem. CNN is a type of deep learning technology that can be used to build a computer-aided diagnosis model. It usually consists of an input layer, multiple convolutional layers, a pooling layer, a fully connected layer, a normalization layer, and an output layer. It directly extracts the most representative features from a large amount of given image data and performs automatic learning and classification recognition, and makes intelligent decisions accordingly, with high recognition accuracy.[17-20] Therefore, it is applied to the classification, detection, and segmentation of medical images. In recent years, the application of CNN in vision system and medical image analysis has attracted more and more attention. In the field of digestive endoscopy image analysis, CNN is mainly used to assist in detecting colon polyps, judging the degree of differentiation of gastric and colon polyps, and identifying early gastrointestinal tumors and gastric mucosal *Helicobacter pylori* infection.[21-24] In recent years, the development and research on the clinical application of CNN has gradually expanded into the field of esophageal diseases. Previously published studies on CNN-based AI for the diagnosis of early EC on endoscopic images show promising yet divergent results, and the methodology varies considerably among studies, particularly in endoscopic approach and CNN structure.[25-31] The results are, therefore, difficult to compare between studies. In this context, we suggest that the diagnostic accuracy of CNN-based AI for early EC on endoscopic images should be fully explored through a meta-analysis of existing studies.

MATERIALS AND METHODS

This meta-analysis was performed in accordance with the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[32] and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines.[33]

Search strategy

A comprehensive search of PubMed, EMBASE, the Cochrane Library, Web of Science, and Wiley Online Library was performed to identify relevant English articles (from the date of database establishment to April 2022). The search principle followed the population, intervention, comparison, outcome, and study design (PICOS) principle (P: "early esophageal cancer," I: "convolutional neural network," S: "diagnostic test").[34]

A combination of subject words and free words was used for the search. The included references were searched twice to reduce the chance of missing relevant articles. Please refer to Appendix 1 for the complete search strategy used for PubMed.

Selection criteria

All studies concerning the CNN-based AI for diagnosis of early EC on endoscopic images, with pathology report results as the gold standard, were considered eligible for inclusion. Furthermore, studies from which a 2 × 2 table could be constructed for true-positive, false-positive, true-negative, and false-negative values were included. Studies were excluded if they lacked an explicitly stated reference standard or had insufficient data to calculate the study outcomes. Animal experiments, case reports, meta-analyses, and reviews were excluded from this study.

Study selection and data extraction

Two investigators independently reviewed the titles and abstracts of the studies for which the inclusion criteria were satisfied for a full-text assessment. The data were independently extracted by two investigators with an agreement kappa value of 96.9%. Differences were resolved by mutual agreement or, if an agreement could not be reached, by discussion with a third reviewer. The extracted data included the following: first author, year,

country, research center, imaging modality, architecture of CNN, training set (number of patients, number of images), text set (number of patients, number of images), and results (sensitivity, specificity, and accuracy). The corresponding authors of the study were contacted when additional information was needed. If no response was received after sending a reminder, the study was excluded.

Quality of the studies

The risk of bias in individual studies was assessed in accordance with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist.[35] Each article was independently evaluated by the reviewers using these criteria, and disagreements were resolved by discussion.

Statistical analysis

Stata software (version 14; Stata Corporation, College Station, TX, USA) was used to draw the plots and perform some calculations. The pooled sensitivity (SEN), pooled specificity (SPE), positive likelihood ratio (LR+), negative likelihood ratio (LR−), diagnostic odds ratio (DOR) with 95% confidence interval (CI), summary receiver operating characteristic (SROC) curve, and area under the curve (AUC) for the accuracy of CNN-based AI for early EC were calculated. Pooling was conducted using a bivariate generalized linear mixed model.[36] Heterogeneity was assessed using the Chi-square and Cochran's Q tests. If I^2 >50%, substantial heterogeneity was considered.[37] If heterogeneity among the studies was recorded, the potential source of heterogeneity was investigated by performing subgroup analyses and a meta-regression analysis. Deeks' funnel plot asymmetry test was used to investigate the publication bias in all included studies, and publication bias was considered significant if $P < 0.01$.

RESULTS

Literature search results

We identified 390 articles by searching the databases, of which 123 (duplicate articles), 209 (after reviewing the titles and abstracts), and 51 (after a full-text review) studies were excluded, leaving seven studies for inclusion in our analysis [Figure 1].

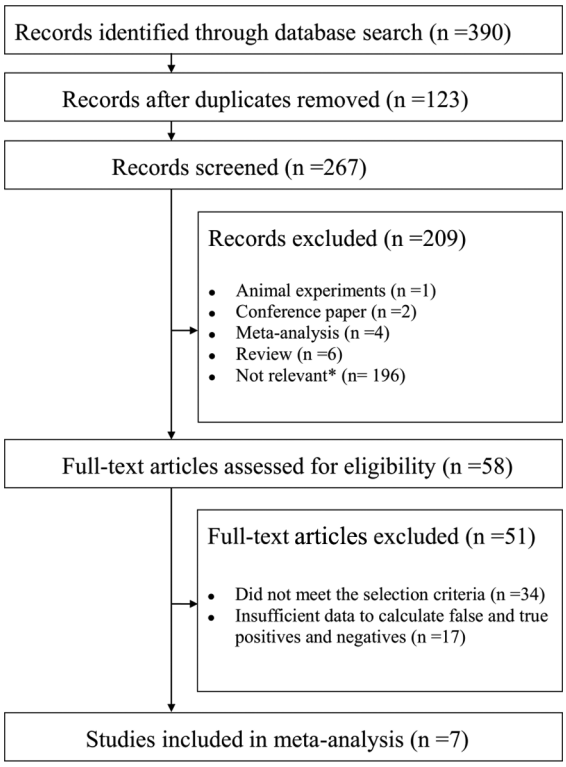


Figure 1: Flowchart summarizing the study selection process

Characteristics of the included studies

The study characteristics are presented in Table 1. Articles included in the analysis were published recently, with the first study published in 2017. This result reflects the fact that CNN-based AI is a relatively novel concept in the field of early EC. All studies[25-31] have training sets and validation sets. The number of patients and images in the training set ranged from 17 to 804 and from 129 to 8660, respectively. The number of patients and images in the validation set ranged from 17 to 155 and from 26 to 1437, respectively. Methodology varied considerably among studies, particularly in endoscopic approaches and CNN structure. All studies[25-31] did not indicate whether the data was obtained from consecutive or random patients, and if all studies[25-31] were retrospective. All studies adopted blinding methods, and all studies used pathological analyses as their gold standard. As noted in Table 1, the sensitivity (range, 0.80–0.97) and specificity (range, 0.73–1.00) of these studies were different.

Quality of the studies

The risk of bias and applicability concerns for the studies included are shown in Table 2. The information is composed of 14 items divided into four parts: patient selection, index test, reference standard, flow and timing. None of the studies fulfilled all items, but all studies fulfilled at least 10 items. The high-risk items were mainly reflected in the patient selection part. The remaining parts were associated with a low risk of bias.

Main results

In our study, the pooled sensitivity and specificity of CNN-based AI based on endoscopic images for predicting early EC were 0.90 (95% CI: 0.82–0.94) and 0.91 (95% CI: 0.79–0.96), respectively [Figure 2]. SROC curves showed that the accuracy of the AUC was 0.95 [Figure 3]. As noted in Supplementary Figure 1, CNN-based AI

Table 1: Characteristics of the included studies

First author	Year	Country	Research center	Imaging modality	Architecture of CNN	Training set		Test set		Results		
						Patients (N1)	Images (N2)	Patients (N1)	Images (N2)	SEN	SPE	ACC
Everson[25]	2019	UK	Multicenter	ME-NBI	Self-designed CNN	17	3962	17	1437	0.89	0.98	0.93
Tokai[26]	2020	Japan	Unicenter	WLE/NBI	GoogLeNet	55	1751	55	291	0.84	0.73	0.80
Cai[27]	2019	China	Multicenter	WLE	Self-designed CNN	746	2428	52	187	0.97	0.85	0.91
Nakagawa[28]	2019	Japan	Unicenter	WLE/NBI	Single-shot multibox detector CNN	804	8660/5678	155	405/509	0.90	0.95	0.91
Hashimoto[29]	2020	Japan	Unicenter	WLE/NBI	GoogLeNet	65	916	39	458	0.96	0.94	0.95
Hong[30]	2017	Korea	Unicenter	CLE	Self-designed CNN	236	262	26	26	0.80	0.77	0.80
Ebigbo[31]	2019	Germany	Unicenter	HD-WLE	ResNet	-	129	14	62	0.83	1.00	0.90

Acc = accuracy, CLE = confocal laser endomicroscopy, CNN = convolutional neural network, ME-NBI = magnification endoscopy with narrow-band imaging, N1 = number of patients, N2 = number of images, NBI = narrow-band imaging, SEN = sensitivity, SPE = specificity, WLE = white-light endoscopy, – = not mentioned in article, HD-WLE = High definition white light endoscopy

based on endoscopic images had a high LR+ (9.8) and a low LR− (0.11), revealing that CNN-based AI exhibited an excellent ability to confirm or exclude early EC on endoscopic images.

Publication bias

We identified publication bias by performing Deeks' regression test of asymmetry ($t = -1.04$; $P = 0.35$) [Supplementary Figure 2]. Deeks' funnel plots for CNN-based AI indicated no publication bias ($P > 0.01$).

Heterogeneity and meta-regression analyses

Substantial heterogeneity was detected among the studies ($I^2 = 96.00\%$, 95% CI: 93.00–99.00). We performed subgroup analyses and meta-regression analysis to identify the source of heterogeneity [Table 3]. Research center (unicenter vs. multicenter), imaging modality (WLE/NBI vs. advanced endoscopic), training set (number of images: ≥1000 vs.<1000), and test set (number of images: ≥1000 vs.<1000) were used as covariates of meta-regression analysis of the effect of heterogeneity. The meta-regression analysis showed that no source of heterogeneity was identified among the covariates we selected.

DISCUSSION

Early EC refers to cancer tissue confined to mucosa and submucosa, regardless of lymph node metastasis.[38] Endoscopic resection is the main treatment method, the prognosis is good, and the survival rate can reach about 90%.[39] However, endoscopic features of these early lesions are subtle and easily missed with conventional endoscopy. In addition, the discovery of early EC relies on the expertise of endoscopists and is inevitably affected by differences in their experience. Therefore, the requirement for more efficient methods of detection and characterization of early EC has led to intensive research in the field of AI. CNN is a branch of machine learning in AI that appears to be better than others AI techniques in image recognition and classification.[40] It has powerful data processing capabilities and is suitable for medical image recognition and complex clinical data analysis. After it is combined with digestive endoscopic imaging technology, it can learn and train a large number of endoscopic images and analyze the relationship between endoscopic images and disease diagnosis. So as to achieve the level of imitating human cognition, it can help doctors to complete fast and accurate diagnosis. In this meta-analysis, we investigated the diagnostic value of CNN-based AI based on endoscopic images for early EC. Our meta-analysis showed that CNN-based AI has good diagnostic value for early EC based on endoscopic

Table 2: Risk of bias and applicability concerns of the included studies

Study	Risk of bias[1]				Applicability[2]		
	Patient selection[3]	Index text[4]	Reference standard[5]	Flow and timing[6]	Patient selection	Index text	Reference standard
Everson[25]	H	L	L	L	L	L	L
Tokai[26]	H	L	L	L	L	L	L
Cai[27]	H	L	L	L	L	L	L
Nakagawa[28]	H	L	L	L	L	L	L
Hashimoto[29]	H	L	L	L	L	L	L
Hong[30]	H	L	L	L	L	L	L
Ebigbo[31]	H	L	L	L	L	L	L

H=high risk, L=low risk, U=unclear. [1]Risk of bias is judged as "low," "high," or "unclear." If the answers to all item questions for a part are "yes," then the risk of bias can be judged low. If any item question is answered "no," potential for bias exists. The "unclear" category should be used only when insufficient data are reported to permit a judgment. [2]Applicability sections are structured in a way similar to that of the bias sections, but do not include signaling questions. Review authors record the information on which the judgment of applicability is made and then rate their concern that the study does not match the review question. Concerns about applicability are rated as "low," "high," or "unclear." The "unclear" category should be used only when insufficient data are reported. [3]Part 1: Patient selection. Risk of bias: Item 1: Was a consecutive or random sample of patients enrolled? Item 2: Was a case-control design avoided? Item 3: Did the study avoid inappropriate exclusions? Applicability: Item 4: Are there concerns that the included patients and setting do not match the review question? [4]Part 2: Index test. Risk of bias: Item 5: Were the index test results interpreted without knowledge of the results of the reference standard? Item 6: If a threshold was used, was it prespecified? Applicability: Item 7: Are there concerns that the index test, its conduct, or its interpretation differ from the review question? [5]Part 3: Reference standard. Risk of bias: Item 8: Is the reference standard likely to correctly classify the target condition? Item 9: Were the reference standard results interpreted without knowledge of the results of the index test? Applicability: Item 10: Are there concerns that the target condition as defined by the reference standard does not match the question? [6]Part 4: Flow and timing. Risk of bias: Item 11: Was there an appropriate interval between the index test and reference standard? Item 12: Did all patients receive a reference standard. Item 13: Did all patients receive the same reference standard? Item 14: Were all patients included in the analysis?

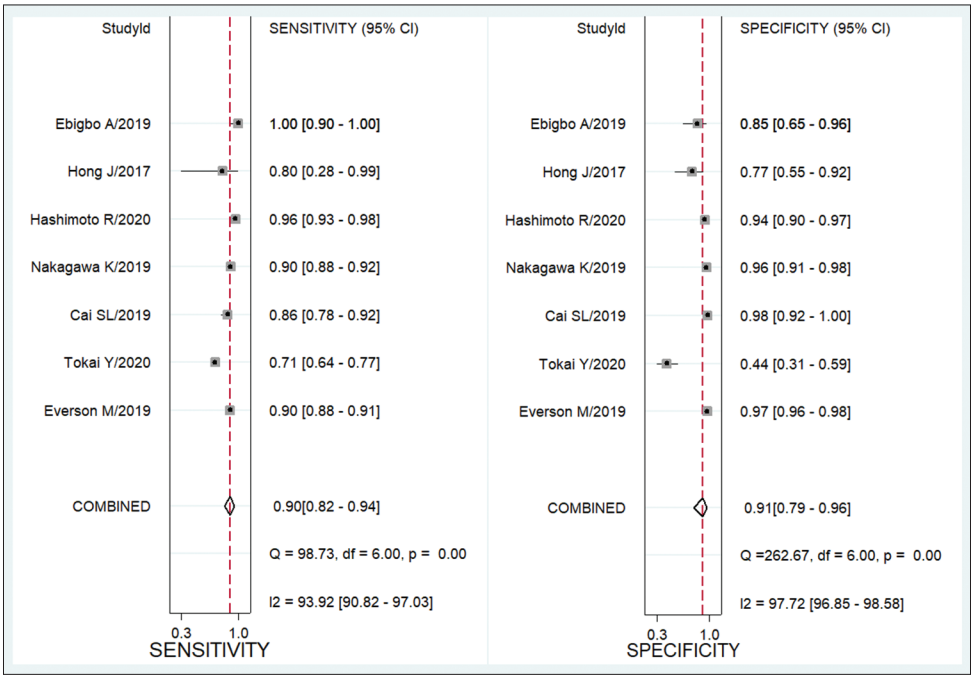


Figure 2: Forest plots of the sensitivity and specificity of CNN-based AI based on endoscopic images for the diagnosis of early EC. The dots correspond to the individual studies included in this analysis, and both sides of the line represent the 95% CI. The narrower the line is, the greater the accuracy of the study and the greater the weight. The diamond corresponds to the pooled result. The intermediate vertical line represents an invalid line. Q statistic test card square value (Chi-square), degree of freedom (df), *P* values, and I^2 statistic test results (inconsistency [I^2]) correspond to heterogeneity test results. The Q test was used to assess heterogeneity, while the I^2 test was used to measure the size of heterogeneity. Heterogeneity was considered when *P* was less than 0.01. If I^2 was <25%, no heterogeneity was noted. If the value of I^2 was between 25% and 50%, the degree of heterogeneity was considered to be small. If the value of I^2 was between 50% and 75%, heterogeneity was noted. If I^2 was >75%, large heterogeneity was noted. AI = artificial intelligence, CI = confidence interval, CNN = Convolutional neural network, EC = esophageal cancer, FN = false negatives, FP = false positives, TN = true negatives, TP = true positives

images with high sensitivity and specificity. In addition, the high AUC value also increases our confidence in the prediction of early EC, which will greatly improve the prognosis of patients, reduce the death rate of EC, and reduce the working intensity of doctors. Another finding of this analysis was that CNN-based AI can not only qualitatively diagnose early EC on endoscopic images, but also quantitatively diagnose it. CNN-based

Table 3: Univariate and multivariate meta-regression analyses for identifying covariates to explain heterogeneity among studies on CNN-based AI for the diagnosis of early EC

Covariates	Multivariate meta-regression			SEN and 95% CI	SPE and 95% CI
	LR (Chi-square test)	*P*	I^2 index (%)		
Research center	7.29	0.03	73	-	-
Unicenter (*n*=5)					
Multicenter (*n*=2)					
Imaging modality	0.03	0.96	0	-	-
WLE/NBI (*n*=4)					
Advanced endoscopic (*n*=3)					
Training set (number of images)	8.21	0.05	87	-	-
≥1000 (*n*=4)					
<1000 (*n*=3)					
Test set (number of images)	1.93	0.38	0	-	-
≥1000 (*n*=1)					
<1000 (*n*=6)					

AI=artificial intelligence, CI=confidence interval, CNN=Convolutional neural network, EC=esophageal cancer, LR=likelihood ratio, SEN=sensitivity, SPE=specificity. *P* value of<0.01 was considered statistically significant

AI can assist in judging the depth of early EC lesions on endoscopic images. Nakagawa *et al*.[28] developed a CNN-based AI system that can be used to distinguish epithelium–submucosal microinvasive (EP-SM1) and submucosal deep invasive (SM2/3) lesions on endoscopic images. The diagnostic performance was similar to that of 16 experienced endoscopists. Tokai *et al*.[26] used 1751 esophageal squamous cell carcinomas with different invasion depths (EP-SM1, SM2) to create a diagnostic model through CNN. Subsequently, CNN-based AI and 13 endoscopists simultaneously examined 291 test images to evaluate diagnostic efficiency. The results showed that the diagnostic performance of CNN exceeded that of 12 experienced endoscopists. A well-known disadvantage of CNN is its black-box nature. To interpret the CNN results visually, seven studies included in this meta-analysis used a single-shot multibox detector, deepLab V.3+, explicit class activation maps, U-Net, and other methods. These techniques could mark the lesion sites with rectangular boxes or outline lesion boundaries with curves to help endoscopists clarify CNN-labeled lesion sites, thus potentially improving the detection rate of early EC with positive biopsy sampling rates. Our meta-analysis does not facilitate the development of clear recommendations regarding the CNN method. In high-dimensional spaces, no method generally outperforms others. Currently, the potential utility of CNN in clinical practice appears promising, despite the use of different approaches in the included studies. Our meta-analysis showed that CNN-based AI has good diagnostic value for early EC based on high-quality endoscopic images.

Substantial heterogeneity was detected in the included studies. According to meta-regression analysis, no source of heterogeneity was identified among the variables we selected. The source of heterogeneity may be related to differences in endoscopic approach and CNN structure in the included studies. However, these variables were not included as covariates in the regression analysis because the grouping conditions were not satisfactory.

We assessed the quality of the included studies using the following four components: patient selection, index test, reference standard, and flow and timing, among which the high-risk items were mainly reflected in the patient selection component. The potential explanation for this finding is that the inclusion criteria of diagnostic trials are often based on case–control trials rather than randomized

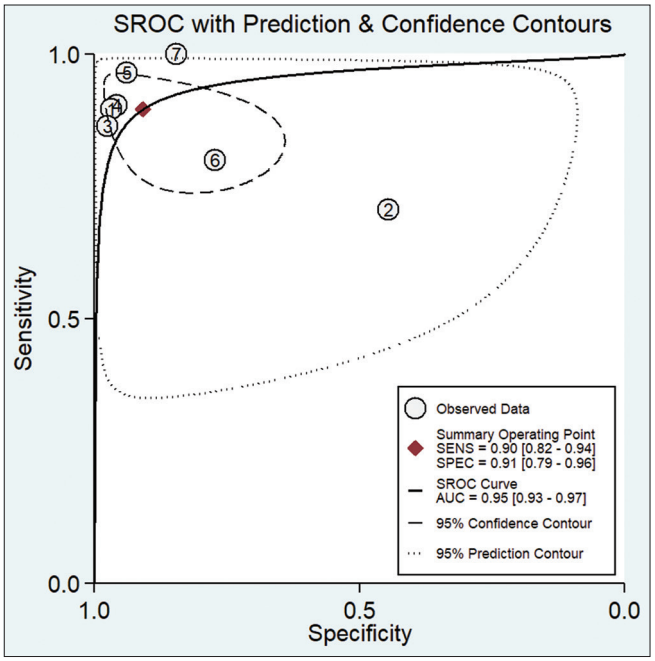


Figure 3: Hierarchical summary of SROC plots of CNN-based AI for the diagnosis of early EC. The ellipse represents 95% CI for this estimate. Numbers correspond to the sensitivity and specificity of individual studies included in this analysis. AI = artificial intelligence, CI = confidence interval, CNN = Convolutional neural network, EC = esophageal cancer

controlled trials, and patients included in the study reported only the time period and did not indicate whether they were consecutive cases. The CNN method and the gold standard method were performed without knowing the results of the other test, and pathological analyses were used as the gold standard in the included studies. Therefore, the selection bias was small and the results were reliable, indicating that these factors were associated with a low risk of bias. Additionally, Deeks' funnel plots for CNN-based endoscopy revealed no publication bias among the studies.

Our meta-analysis has several limitations. Firstly, most studies included in the meta-analysis employed a retrospective design and, therefore, were subject to selection bias and prone to data loss. Secondly, substantial heterogeneity was detected in the included studies. The source of heterogeneity may be related to differences in endoscopic approach and CNN structure in the included studies. Finally, most studies have relatively small sample sizes and are single-center studies. Therefore, further prospective studies using a larger and more balanced population from multiple centers are required. In summary, CNN is a relatively new concept and the variety of presented approaches is worth investigating. All weak points should be addressed in future trials comprising larger patient cohorts from multicenter, multivendor studies.

In conclusion, our meta-analysis showed that CNN-based AI has good diagnostic value for early EC based on endoscopic images with high sensitivity, specificity, and AUC. More importantly, our study validated the feasibility of using CNN-based AI to conclusively identify a disease prone to missed diagnosis and misdiagnosis. This finding is very important for determining the diagnosis, treatment, and prognosis of early EC on endoscopic images.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *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. Yamamoto S, Kato K. Pembrolizumab for the treatment of esophageal cancer. Expert Opin Biol Ther 2020;20:1143-50.
3. Zheng BZ, Liu TD, Chen G, Zhang JX, Kang X. The effect of curcumin on cell adhesion of human esophageal cancer cell. Eur Rev Med Pharmacol Sci 2018;22:551-60.
4. Liu X, Wu W, Zhang S, Tan W, Qiu Y, Liao K, *et al*. Effect of miR-630 expression on esophageal cancer cell invasion and migration. J Clin Lab Anal 2021;35:e23815. doi: 10.1002/jcla.23815.
5. Cai W, Li Y, Huang B, Hu C. Esophageal cancer lymph node metastasis-associated gene signature optimizes overall survival prediction of esophageal cancer. J Cell Biochem 2019;120:592-600.
6. Shin A, Won YJ, Jung HK, Kong HJ, Jung KW, Oh CM, *et al*. Trends in incidence and survival of esophageal cancer in Korea: Analysis of the Korea central cancer registry database. J Gastroenterol Hepatol 2018;33:1961-8.
7. Xie SH, Santoni G, Malberg K, Lagergren P, Lagergren J. Prediction model of long-term survival after esophageal cancer surgery. Ann Surg 2021;273:933-9.
8. Hirst J, Smithers BM, Gotley DC, Thomas J, Barbour A. Defining cure for esophageal cancer: Analysis of actual 5-year survivors following esophagectomy. Ann Surg Oncol 2011;18:1766-74.
9. Monig S, Chevallay M, Niclauss N, Zilli T, Fang W, Bansal A, *et al*. Early esophageal cancer: The significance of surgery, endoscopy, and chemoradiation. Ann N Y Acad Sci 2018;1434:115-23.
10. Domper Arnal MJ, Ferrandez Arenas A, Lanas Arbeloa A. Esophageal cancer: Risk factors, screening and endoscopic treatment in Western and Eastern countries. World J Gastroenterol 2015;21:7933-43.
11. Yang H, Hu B. Recent advances in early esophageal cancer: Diagnosis and treatment based on endoscopy. Postgrad Med 2021;133:665-73.
12. Arantes V, Albuquerque W, Salles JM, Freitas Dias CA, Alberti LR, *et al*. Effectiveness of unsedated transnasal endoscopy with white-light, flexible spectral imaging color enhancement, and lugol staining for esophageal cancer screening in high-risk patients. J Clin Gastroenterol 2013;47:314-21.
13. Ebi M, Shimura T, Yamada T, Mizushima T, Itoh K, Tsukamoto H, *et al*. Multicenter, prospective trial of white-light imaging alone versus white-light imaging followed by magnifying endoscopy with narrow-band imaging for the real-time imaging and diagnosis of invasion depth in superficial esophageal squamous cell carcinoma. Gastrointest Endosc 2015;81:1355-61.e2.
14. Kawada K, Arima M, Miyahara R, Tsunomiya M, Kikuchi M, Yamamoto F, *et al*. Effect of adding magnifying BLI, magnifying NBI, and iodine staining to white light imaging in diagnosis of early esophageal cancer. Endosc Int Open 2021;9:E1877-85.
15. Yu T, Geng J, Song W, Jiang Z. Diagnostic accuracy of magnifying endoscopy with narrow band imaging and its diagnostic value for invasion depth staging in esophageal squamous cell carcinoma: A systematic review and meta-analysis. BioMed Res Int 2018;2018:8591387. doi: 10.1155/2018/8591387.
16. Mashimo H, Gordon SR, Singh SK. Advanced endoscopic imaging for detecting and guiding therapy of early neoplasias of the esophagus. Ann N Y Acad Sci 2020;1482:61-76.
17. Wu Y, Lin L, Wang J, Wu S. [Application of semantic segmentation based on convolutional neural network in medical images]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2020;37:533-40.
18. Liu J, Zhao H. Application of convolution neural network in medical image processing. Technol Health Care 2021;29:407-17.
19. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. Proc Natl Acad Sci U S A 2018;115:254-9.
20. Hassanzadeh T, Essam D, Sarker R. Evolutionary deep attention convolutional neural networks for 2D and 3D medical image segmentation. J Digit Imaging 2021;34:1387-404.
21. Rasouli P, Dooghaie Moghadam A, Aghajanpoor Pasha M, Asadzadeh Aghdaei H, Mehrvar A, *et al*. The role of artificial intelligence in colon polyps detection. Gastroenterol Hepatol Bed Bench 2020;13:191-9.
22. Guo XM, Zhao HY, Shi ZY, Wang Y, Jin ML. [Application and progress of convolutional neural network-based pathological diagnosis of gastric cancer]. Sichuan da Xue Xue Bao Yi Xue Ban 2021;52:166-9.
23. Xiao Z, Ji D, Li F, Li Z, Bao Z. Application of artificial intelligence in

early gastric cancer diagnosis. Digestion 2022;103:69-75.
24. 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 2021;36:482-9.
25. Everson M, Herrera L, Li W, Luengo IM, Ahmad O, Banks M, *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;7:297-306.
26. Tokai Y, Yoshio T, Aoyama K, Horie Y, Yoshimizu S, Horiuchi Y, *et al*. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. Esophagus 2020;17:250-6.
27. Cai SL, Li B, Tan WM, Niu XJ, Yu HH, Yao LQ, *et al*. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). Gastrointest Endosc 2019;90:745-53.e2.
28. Nakagawa K, Ishihara R, Aoyama K, Ohmori M, Nakahira H, Matsuura N, *et al*. Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists. Gastrointest Endosc 2019;90:407-14.
29. Hashimoto R, Requa J, Dao T, Ninh A, Tran E, Mai D, *et al*. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). Gastrointest Endosc 2020;91:1264-71.e1.
30. Jisu H, Bo-Yong P, Hyunjin P. Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images. Annu Int Conf IEEE Eng Med Biol Soc 2017;2017:2892-5.
31. Ebigbo A, Mendel R, Probst A, Manzeneder J, Prinz F, de Souza LA Jr, *et al*. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. Gut 2020;69:615-6.
32. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, *et al*. Updated guidance for trusted systematic reviews: A new edition of the cochrane handbook for systematic reviews of interventions. Cochrane Database Syst Rev 2019;10:ED000142. doi: 10.1002/14651858.ED000142.
33. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; and the PRISMA-DTA Group, *et al*. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. JAMA 2018;319:388-96.
34. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014;14:579. doi: 10.1186/s12913-014-0579-0.
35. Qu YJ, Yang ZR, Sun F, Zhan SY. [Risk on bias assessment: (6) A revised tool for the quality assessment on diagnostic accuracy studies (QUADAS-2)]. Zhonghua Liu Xing Bing Xue Za Zhi 2018;39:524-31.
36. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T. Bivariate random effects meta-analysis of ROC curves. Med Decis Making 2008;28:621-38.
37. Zhang C, Chen M, Wang X. Statistical methods for quantifying between-study heterogeneity in meta-analysis with focus on rare binary events. Stat Interface 2020;13:449-64.
38. Wei MT, Friedland S. Early esophageal cancer what the gastroenterologist needs to know. Gastroenterol Clin N 2021;50:791-808.
39. Lim H, Kim DH, Jung HY, Gong EJ, Na HK, Ahn JY, *et al*. Clinical significance of early detection of esophageal cancer in patients with head and neck cancer. Gut Liver 2015;9:159-65.
40. Huang LM, Yang WJ, Huang ZY, Tang CW, Li J. Artificial intelligence technique in detection of early esophageal cancer. World J Gastroenterol 2020;26:5959-69.

APPENDIX 1

PubMed search strategy

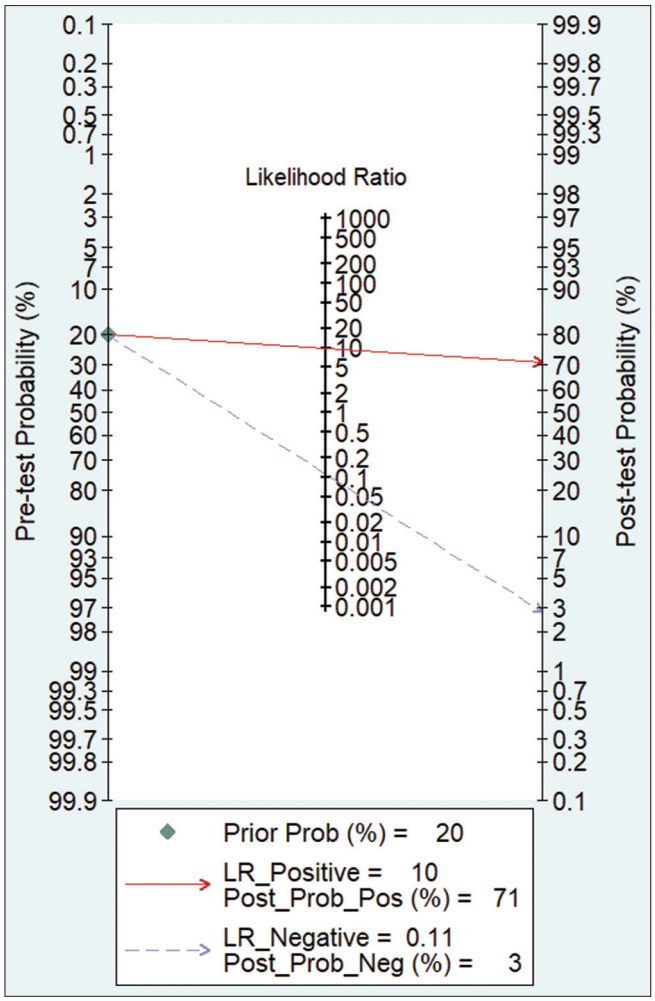
#1 ("Esophageal Neoplasms"[Mesh]) OR ((((((((((((((((Esophageal Neoplasm[Title/Abstract]) OR (Neoplasm, Esophageal[Title/Abstract])) OR (Esophagus Neoplasm[Title/Abstract])) OR (Esophagus Neoplasms[Title/Abstract])) OR (Neoplasm, Esophagus[Title/Abstract])) OR (Neoplasms, Esophagus[Title/Abstract])) OR (Neoplasms, Esophageal[Title/Abstract])) OR (Cancer of Esophagus[Title/Abstract])) OR (Cancer of the Esophagus[Title/Abstract])) OR (Esophagus Cancer[Title/Abstract])) OR (Cancer, Esophagus[Title/Abstract])) OR (Cancers, Esophagus[Title/Abstract])) OR (Esophagus Cancers[Title/Abstract])) OR (Esophagus Cancers[Title/Abstract])) OR (Cancer, Esophageal[Title/Abstract])) OR (Cancers, Esophageal[Title/Abstract])) OR (Esophageal Cancers[Title/Abstract])).

#2 ((((((((((((convolutional neural network) OR (convolutional network)) OR (neural network)) OR ("Deep Learning"[Mesh])) OR (artificial intelligence)) OR (machine learning)) OR (computer-aided)) OR (computer aided)) OR (hierarchical learning)) OR (computational intelligence)) OR (machine intelligence)) OR (computer reasoning)) OR (classification algorithm)) OR (feed-forward neural network).

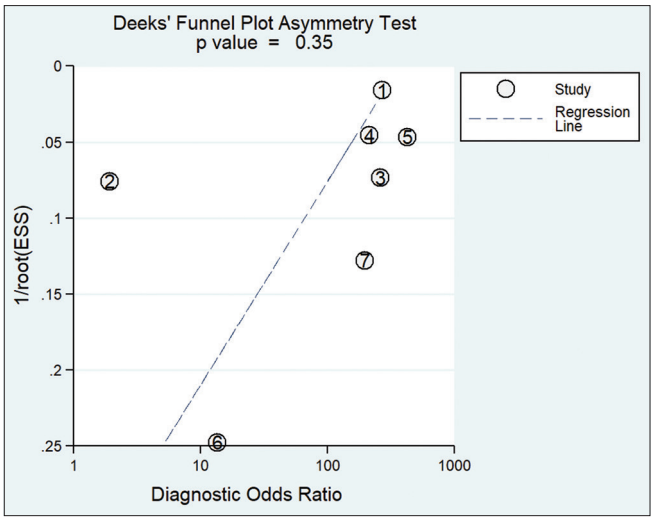
#3 ("sensitivity and specificity"[MeSH] OR predict*[text] OR diagnos*[text] OR accura*[text]).

#1 AND #2 AND #3.

Number of articles: 264.



Supplementary Figure 1: Fagan nomogram of CNN-based AI for the diagnosis of early EC. AI = artificial intelligence, CNN = Convolutional neural network, EC = esophageal cancer



Supplementary Figure 2: Deeks' funnel plot asymmetry test for publication bias. Numbers correspond to the individual studies included in this analysis, ESS=Effective sample size