Original Article

Artificial intelligence-assisted detection and classification of colorectal polyps under colonoscopy: a systematic review and meta-analysis

Aling Wang[1#], Jiahao Mo[1#], Cailing Zhong[2#], Shaohua Wu[1], Sufen Wei[2], Binqi Tu[1], Chang Liu[1], Daman Chen[3], Qing Xu[3], Mengyi Cai[3], Zhuoyao Li[3], Wenting Xie[3], Miao Xie[3], Motohiko Kato[4], Xujie Xi[2], Beiping Zhang[2]

[1]The Second Clinical Medical School, Guangzhou University of Chinese Medicine, Guangzhou, China; [2]Department of Gastroenterology, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China; [3]Guangzhou University of Chinese Medicine, Guangzhou, China; [4]Division of Gastroenterology and Hepatology, Department of Internal Medicine Keio University School of Medicine, Tokyo, Japan

Contributions: (I) Conception and design: A Wang, J Mo, C Zhong; (II) Administrative support: M Kato, X Xi, B Zhang; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: C Liu, D Chen, Q Xu, M Cai, Z Li, W Xie, M Xie; (V) Data analysis and interpretation: A Wang, J Mo, S Wu, S Wei, B Tu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Xujie Xi; Beiping Zhang. Department of Gastroenterology, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, 111 Dade Road, Yuexiu District, Guangzhou 510120, China. Email: xixujiemuzi@qq.com; doctorzbp@163.com.

Background: Artificial intelligence (AI) is used to solve the problem of missed diagnosis of polyps in colonoscopy, which has been proved to improve the detection rate of adenomas. The aim of this review was to evaluate the diagnostic performance of AI-assisted detection and classification of polyps in colonoscopy.

Methods: The literature search was undertaken on 4 electronic databases (PubMed, Web of Science, Embase, and Cochrane Library). The inclusion criteria were as follows: studies reporting AI-assisted detection and classification of polyps; studies containing patients, images, or videos receiving AI-assisted diagnosis; studies which included AI-assisted diagnosis and reported classification based on histopathology; and studies providing accurate diagnostic data. Non-English language studies, case-reports, reviews, meeting abstracts and so on were excluded. The Quality Assessment of Diagnostic Accuracy Studies-2 scale was used to evaluate the quality of literature and the Stata 13.0 software was used to perform meta-analysis.

Results: Twenty-six articles were included with all of medium quality. Meta-analysis showed none of literature had any obvious publication bias. The application of AI in detection of colorectal polyps achieved a sensitivity of 0.95 [95% confidence interval (CI): 0.89–0.98] and an area under the curve (AUC) of 0.79 (95% CI: 0.79–0.82). In the AI-assisted classification, the sensitivity was 0.92 (95% CI: 0.88–0.95) with a specificity of 0.82 (95% CI: 0.71–0.89) and an AUC of 0.94 (95% CI: 0.92–0.96). For the classification of diminutive polyps, the AI-assisted technique yielded a sensitivity of 0.95 (95% CI: 0.94–0.97), a specificity of 0.88 (95% CI: 0.74–0.95), and an AUC of 0.97 (95% CI: 0.95–0.98). For AI-assisted classification under magnifying endoscopy, the sensitivity was 0.954 (95% CI: 0.92–0.96) with a specificity of 0.95 (95% CI: 0.80–0.99) and an AUC of 0.97 (95% CI: 0.95–0.98).

Discussion: The AI-assisted technique demonstrates impressive accuracy for the detection and characterization of colorectal polyps and can be expected to be a novel auxiliary diagnosis method. Our study has inevitable limitations including heterogeneity due to different AI systems and the inability to further analyze the specificity and sensitivity of AI for different types of endoscopes.

Keywords: Artificial intelligence (AI); colorectal polyps; colonoscopy; meta-analysis

Submitted Sep 05, 2021. Accepted for publication Nov 12, 2021.
doi: 10.21037/atm-21-5081
View this article at: https://dx.doi.org/10.21037/atm-21-5081

Introduction

Colorectal cancer (CRC), is the third most common cancer worldwide and poses a considerable threat to public health due to its high mortality (1). Colorectal adenoma (CRA) and serrated polyps have been proven to be precancerous lesions of CRC. Colonoscopy is performed for the detection and resection of these lesions and been demonstrated to reduce the incidence and mortality of CRC (2,3). A large US cohort study (4) showed that the mortality rate of CRC was reduced by approximately 70% by colonoscopy screening and on-demand therapeutics. There is evidence suggesting that the adenoma detection rate (ADR) can indicate the colonoscopy quality and that ADR is inversely proportional to postcolonoscopy CRC risk (5,6). However, due to operator-dependent limitations, polyps smaller than 5 mm may be missed at colonoscopy with an overall missed diagnosis rate for adenomas as high as 27% (7-9). Colorectal polyps can be divided into neoplastic and nonneoplastic polyps and require different treatment strategies. Therefore, there is an urgent need to reduce the miss rate of polyps and improve the accuracy of polyp pathology evaluation under endoscopy.

Artificial intelligence (AI) emerged as a scientific discipline in 1956, but what is now shown to people is more of a technology, which refers to systems with the ability to reason, discover meaning, generalize, or learn from past experience, thus able to perform tasks normally requiring human interaction (10). At present, artificial intelligence has been applied to many aspects of human life, such as transportation, entertainment, trade, medical care and so on. In contemporary society, Artificial intelligence has been gradually applied to the field of digestive endoscopy. Notably, it has been employed in the detection and classification of colorectal polyps. However, sensitivity and specificity differences have been reported in the results of AI-assisted colorectal polyp diagnosis (11-13). Although there have been meta-analysis articles on the diagnostic performance of AI-assisted colonoscopy for colorectal polyps, most articles only focus on the detection of adenomas or only study one type of AI system (14-19), and the original research articles are constantly updated. Thus, we aimed to systematically review and meta-analyze the diagnostic quality of AI-based technologies in both the detection and characterization of colorectal polyps combining with updated articles. This review has been registered on PROSPERO: Diagnostic performance of artificial intelligence in the detection and classification of colorectal polyp: a systematic review and meta-analysis; ID: CRD42021256884. We present the following article in accordance with the PRISMA reporting checklist (available at https://dx.doi.org/10.21037/atm-21-5081).

Methods

Search strategy

We searched all published articles evaluating the diagnostic performance of AI-assisted detection and classification of colorectal polyps in PubMed, Web of science, Embase, and Cochrane Library until April 2021. The search strategy was based on the following keywords: {[“artificial intelligence”] OR [“convolutional neural networks”] OR [“deep learning”] OR [“computer-aided”]} AND {[“colonoscopy”] OR [“endoscopy”]} AND {[“colon”] OR [“colonic”] OR [“colorectal”]} AND {[“polyp”] OR [“polyps”] OR [“adenoma”] OR [“adenomas”]}.

Inclusion and exclusion criteria

The inclusion criteria were as follows: (I) studies reporting AI-assisted detection and classification of colorectal polyps in international publications; (II) studies containing patients, endoscopic images, or videos receiving AI-assisted diagnosis of colorectal polyps with definite diagnostic results; (III) studies whose diagnostic methods included AI-assisted diagnosis (including detection and classification of colorectal polyps) without restrictions of algorithms, with those studies reporting the classification of colorectal polyps being based on histopathological diagnosis; and (IV) studies providing accurate diagnostic data.

The exclusion criteria were as follows: (I) non-English language studies; (II) case-reports, reviews, meeting abstracts, comments, letters, systematic reviews, or study protocols; (III) studies with an irrelevant subject; (IV) studies with incomplete data; and (V) studies with a small sample size.

Study selection and data extraction

Study selection and data extraction were completed independently by 2 investigators (Wang and Mo). Based on the inclusion and exclusion criteria, the candidate articles were screened by reviewing their titles and abstracts at first. Relevant studies were then further evaluated through a reading of the full text. Finally, search results were cross-checked by 2 investigators, and the discrepancies were resolved by a third investigator (Zhong).

Data extracted from studies were placed onto a standard spreadsheet template using Microsoft Excel. For each study, the following data were extracted: the first author's name, publication year, country where the study was conducted, data source, type of study (detection or classification of colorectal polyps), type of observation (image and video verification or real-time monitoring), AI algorithms, test objects, sample group, and original data reflecting the diagnostic performance [i.e., true positive (TP), false positive (FP), true negative (TN), and false negative (FN)]. For studies involving multiple AI structure verification, the method for merging all structures was applied to raw data processing. For studies verifying the same AI system in different databases, the method for merging all databases was used for raw data processing until the original data were complete. For the studies splitting the same database into different subdatabases, only the original data of the original database were included. For studies that listed the original data of colorectal polyps diagnosed by experts and nonexperts, the data of experts and nonexperts were entered separately before being included in the meta-analysis. For the studies listing the diagnosis from experts and nonexperts one by one, the data of experts and nonexperts were added separately before being included in the meta-analysis.

Quality assessment

RevMan 5.4 (Cochran Training) was used to assess the quality of all included literature, and the risk of bias was evaluated by 2 investigators (Wang and Mo) independently adopting Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) criteria. For each item, an evaluation of "yes" or "unclear" or "no" was given, and each item was classified as "high risk" or "unclear" or "low risk". In terms of the applicability evaluation, each item was classified as "high concern" or "unclear concern" or "low concern".

Objective and outcome indicators

The objective of this study was to explore the diagnostic performance of AI-assisted detection and classification of colorectal polyps. The outcome indicators consisted of pooled sensitivity, pooled specificity, pooled positive likelihood ratio (PLR), pooled negative likelihood ratio (NLR), pooled diagnosis odds ratio (DOR), summary receiver operating characteristic curves (SROC), and the area under the curve (AUC), all which were calculated based on TP, FP, TN and FN.

Statistical analysis

Statistical analysis was performed by Stata 13.0 (StataCorp). The heterogeneity caused by a threshold effect was tested by Spearman correlation analysis, and the heterogeneity caused by a non-threshold effect was tested by Cochran-Q and I^2 value, where <50% was low and >50% was high; the fixed effects model and the random effects model were used to merge respectively. Four grid tables for AI-assisted detection and classification of colorectal polyps were listed, and the sensitivity (SEN), specificity (SPE), PLR, NLR, DOR, and their 95% confidence interval (95% CI) were calculated. The probabilities before and after the test were observed through Bayesian analysis, and the changes of positive and negative results were evaluated. The sensitivity analysis of our study was to eliminate studies with low quality or different efficacy evaluation criteria, and then conducted merger analysis to compare with the merger effect before elimination, so as to explore the impact of the elimination study on the merger effect. If there was no significant change in the amount of merger effect before and after elimination, the result was relatively stable. If there was a large difference or even an opposite conclusion, it indicated that the stability of the results was poor. Furthermore, we drew the SROC curve, calculated the AUC, and evaluated the diagnostic value. The AUC values were interpreted as follows: no diagnostic value if AUC <0.5, low diagnostic value if 0.5≤ AUC <0.7, high diagnostic value if 0.7≤ AUC <0.9, and extremely high diagnostic value if AUC >0.9. Finally, the publication bias of the included studies was quantitatively assessed by bias analysis.

Results

Literature screening

According to the above retrieval strategy, a total of 709 articles were identified from the databases (PubMed 210, Web of Science 100, Embase 326, Cochrane Library 73). In addition, 5 articles were obtained after screening the published articles of related systematic reviews and meta-analysis, which totaled 714 records. After 284 articles were excluded as duplicates and 372 articles were excluded on the basis of titles and abstracts, 15 studies on polyps detection (9,12,13,20-31), 10 studies on polyps classification (32-41), and 1 (42) which was a combination of both were identified as being appropriate for full-text review. The process of literature screening and inclusion is shown in *Figure 1*.

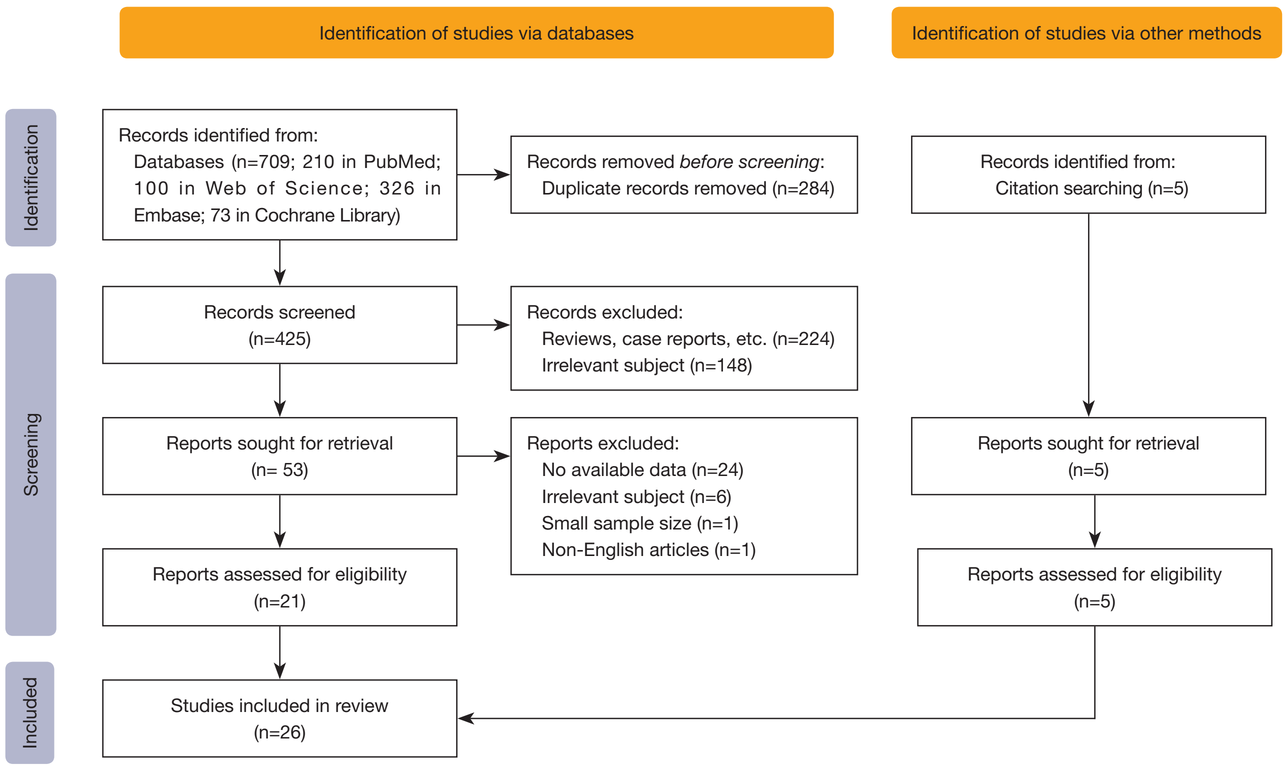

Figure 1 The process of literature screening and inclusion.

The basic characteristics of the included literature

Studies in this systematic review included 15 preclinical studies on polyps detection (9,12,13,20-31), 10 preclinical studies on polyps classification (32-41), and 1 (42) which was a combination of both. In terms of the detection of polyps, 5 studies exploring the (9,23,26-28) the performance of real-time AI-assisted detection reported no TN. There were 4 studies that (13,20,22,30) used pictures or videos with polyps to verify AI-assisted detection performance, in which the reported number of TN was 0. However, the other trials all reported TN. In terms of the classification of polyps, except for a (37) real-time AI-assisted classification study, all AI-assisted classification studies used pictures or videos. A further 5 studies (33,35-38) compared the diagnostic performance of AI, experts, and non-experts, while 1 study (40) only compared the diagnostic performance of AI with that of experts. Among all the literature, only the studies of Jia (22) and Patel (39) assessed the diagnostic performance of convolutional neural network (CNN) systems with different structures, which was a kind of feedforward neural networks with depth structure including convolution calculation, and also one of the representative algorithms of deep learning. The basic characteristics and diagnostic characteristics of the included literature are shown in *Table 1* and *Table 2*, respectively.

Results of literature quality evaluation

Among the 26 included articles, the overall quality of the research was medium. Nine studies (20,22,29,32,34-36,38,40) were classified as "high risk" in terms of patient selection due to the lack of indication of whether the included cases or polyp images were continuous and randomized and due to the exclusion criteria of the inappropriate cases. One study (40) was rated as "high risk" in terms of flow and timing because not all endoscopic images were included in the outcome analysis. Four studies (35-38) were listed as "high concern" in terms of patient selection, mainly because enlarged endoscopic images were included in the studies. One study (12) was "high concern" in terms of the reference standard because the existence of polyps was confirmed by different endoscopists. The quality evaluation results of the included literature are shown in *Figure 2*.

Table 1 The basic characteristics of the included literature

Author	Year	Region	Field focused	Method of study	Types of AI systems	Type of lesions	Type of images	Testing objects
Liu WN (9)	2020	China	Detection	Real-time use	3D-CNN	Polyps of any size	NA	AI system
Misawa (12)	2021	Japan	Detection	Videos verification	YoloV3	Polyps of any size	WLI	AI system
Urban (13)	2018	USA	Detection	Images and videos verification	DCNN	Polyps of any size	NA	AI system
Qadir (20)	2021	Norway	Detection	Image verification	F-CNN	Polyps of any size	NA	AI system
Guo (21)	2021	Japan	Detection	Videos verification	YoloV3	Polyps of any size	NA	AI system/expert/nonexpert
Jia (22)	2020	Hong Kong, China	Detection	Image verification	CNN	Polyps of any size	NA	AI system
Liu P (23)	2020	China	Detection	Real-time use	Deep learning	Polyps of any size	NA	AI system
Poon (24)	2020	Hong Kong, China	Detection	Images and videos verification	CNN	Polyps of any size	NA	AI system
Shin (25)	2018	Norway	Detection	Images verification	Dictionary learning scheme	Polyps of any size	NA	AI system
Su (26)	2020	China	Detection	Real-time use	DCNN	Polyps of any size	NA	AI system
Wang (27)	2019	China	Detection	Real-time use	DCNN	Polyps of any size	NA	AI system
Wang (28)	2020	China	Detection	Real-time use	Deep learning	Polyps of any size	NA	AI system
Wang (29)	2018	China	Detection	Image and video verification	Deep learning	Polyps of any size	NA	AI system
Yu (30)	2017	Hong Kong, China	Detection	Image verification	3D-FCN	Polyps of any size	NA	AI system
Zhang (31)	2018	Hong Kong, China	Detection	Images verification	DCNN	Polyps of any size	NA	AI system
Byrne (32)	2019	Canada	Classification	Video verification	DCNN	Polyps that ≤5 mm	NA	AI system
Chen (33)	2018	Taiwan, China	Classification	Video verification	DCNN	Polyps that ≤5 mm	NA	AI system/expert/nonexpert
Kominami (34)	2016	Japan	Classification	Image verification	SVM	Polyps of any size	NA	AI system
Kudo (35)	2020	Japan	Classification	Image verification	NA	Polyps that ≤10 mm	WLI/EC NBI/EC methylene blue staining	AI system/expert/nonexpert
Mori (36)	2016	Japan	Classification	Image verification	SVM	Polyps of any size	EC images	AI system/expert/nonexpert

Table 1 *(continued)*

Table 1 *(continued)*

Author	Year	Region	Field focused	Method of study	Types of AI systems	Type of lesions	Type of images	Testing objects
Mori (37)	2018	Japan	Classification	Real-time use	NA	Polyps that ≤5 mm	EC NBI/EC methylene blue staining	AI system/expert/nonexpert
Mori (38)	2015	Japan	Classification	Image verification	NA	Polyps that ≤10 mm	WLI/EC images	AI system/expert/nonexpert
Patel (39)	2020	America	Classification	video verification	CNN	Polyps of any size	NA	AI system
Renner (40)	2018	Germany	Classification	Image verification	DCNN	Polyps of any size	NA	AI system/expert
Yamada (41)	2019	Japan	Classification	Image verification	NA	Polyps of any size	NA	AI system
Ozawa (42)	2020	Japan	Detection and Classification	Image verification	CNN	Polyps of any size	NA	AI system

DCNN, deep convolutional neural network; CNN, convolutional neural network; YoloV3, a deep learning–based common object detection algorithm; NBI, narrow band imaging; 3D-FCN, three-dimensional fully convolutional network; F-CNN, fully convolutional neural network; 3D-CNN, three-dimensional convolutional neural network; SVM, support vector machine; WLI, white light imaging; EC, endocytoscopy; NA, not available.

Meta-analysis

Meta-analysis of AI-assisted detection of colorectal polyps

A total of 16 studies reported the performance of AI-assisted detection of colorectal polyps. The TN was set to 0 in studies reporting no TN. For the pooled analysis of 16 studies, the heterogeneity (I^2) of the Sen was 99.85 (P<0.01), and the Sen was 0.95 (95% CI: 0.89–0.98), as shown in *Figure 3*. In terms of literature analysis, the 19% probability after the test was calculated from the probability before test and PLR [1] in the positive test results, while the 97% probabilities before and after the test were calculated from the pretest probability and NLR (114.31) in the negative test results (*Figure 4*). The AUC under the SROC curve was estimated to be 0.79 (95% CI: 0.79–0.82), as shown in *Figure 5*. Moreover, the publication bias of included literature was quantitatively analyzed, and the results are shown in *Figure 6* (P=0.07>0.05) and suggested no significant publication bias.

AI-assisted detection of colorectal polyps: a subgroup meta-analysis of studies with TN

A total of 7 studies with TN reported the performance of AI-assisted detection of colorectal polyps. In the pooled analysis of the 7 studies, the heterogeneity (I^2) of the sensitivity was 99.95 (P<0.01), and the sensitivity was 0.88 (95% CI: 0.81–0.92). The heterogeneity (I^2) of the specificity was 99.99 (P<0.01), and the specificity was 0.95 (95% CI: 0.94–0.96), as shown in *Figure 7*. In the SROC curve, the AUC was 0.97 (95% CI: 0.95–0.98), as shown in *Figure 8*.

Meta-analysis of AI-assisted classification of colorectal polyps

A total of 11 studies reported the performance of AI-assisted classification of colorectal polyps for distinguishing neoplastic and nonneoplastic polyps. The heterogeneity (I^2) of the sensitivity was 99.37 (P<0.01), and the heterogeneity (I^2) of the specificity was 99.17 (P<0.01). The sensitivity was 0.92 (95% CI 0.88–0.95), and the specificity was 0.82 (95% CI: 0.71–0.89). The PLR was 5.0 (95% CI: 3.1–8.2), and the NLR was 0.10 (95% CI: 0.06–0.15). The DOR was 51 (95% CI: 22–117), as shown in *Figure 9*. In terms of literature analysis, the 57% of the posttest probability was calculated from the pretest probability and PLR [5] in the positive test results, while the 2% of the posttest probability was calculated from the pretest probability and NLR (0.09) in the negative test results (*Figure 10*). In the SROC curve, the AUC was 0.94 (95% CI: 0.92–0.96), as shown in *Figure 11*. The publication bias of included literature

Table 2 Diagnostic characteristics of the included literature

Studies	Different grouping methods	AI systems				Expert				Nonexpert			
		TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN
Polyp detection													
Liu WN(9)		486	36	0	NA								
Misawa (12)		44,472	5,964	4,668	88,075								
Urban (13)		113	127	5	NA								
Qadir (20)	Dataset 1	180	28	28	NA								
	Dataset 2	273	36	27	NA								
Guo (21)	Long videos	37,938	5,590	5,672	78,658								
	Short videos	44	NA	6	NA	88	0	12	100	80	17	20	83
Jia (22)	Architecture 1	524	116	122	NA								
	Architecture 2	535	96	111	NA								
	Architecture 3	549	239	97	NA								
	Architecture 4	557	3,608	89	NA								
	Architecture 5	595	107	51	NA								
Liu P (23)		421	29	0	NA								
Poon (24)	Dataset 1	3,206	480	1,207	12,880								
	Dataset 2	47,877	277,407	18,082	3,363,076								
Shin (25)		188	8	7	163								
Su (26)		177	62	0	NA								
Wang (27)		498	39	0	NA								
Wang (28)		501	50	0	NA								
Wang (29)	Dataset 1	6,233	1,297	413	20,691								
	Dataset 2	55,822	49,334	5,092	1,023,149								
Yu (30)		3,062	414	1,251	NA								
Zhang (31)		3,087	398	1226	13,057								
Ozawa (42)	All images	1,073	173	99	5,732								
	WLI	787	161	87	5,713								
	NBI	289	9	9	22								
Polyp classification													
Byrne (32)		65	7	1	33								
Chen (33)		181	21	7	75	367	55	9	137	671	95	81	289
Kominami (34)	All polyps	70	3	3	42								
	Polyps ≤5 mm	40	3	3	42								
Kudo (35)	Polyps ≤10 mm in stained mode	1,260	0	40	700	603	20	20	330	920	240	380	460
	Polyps ≤10 mm in NBI mode	1,260	40	40	660	608	12	42	338	807	100	493	600
	Polyps ≤5 mm in stained mode	960	0	40	680	453	20	47	320	690	236	310	444

Table 2 *(continued)*

Table 2 *(continued)*

Studies	Different grouping methods	AI systems				Expert				Nonexpert			
		TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN
Mori (36)		131	7	16	51	408	20	33	154	1,128	153	342	427
	Polyps ≤5 mm in NBI mode	960	40	40	640	459	12	41	328	578	97	422	583
Mori (37)	All polyps in NBI mode	268	18	17	159								
	All polyps in stained mode	263	19	23	157								
	Proximal-to-rectosigmoid polyps ≤5 mm in NBI mode	170	13	10	21								
	Proximal-to-rectosigmoid polyps ≤5 mm in stained mode	167	14	9	24								
	Rectosigmoid polyps ≤5 mm in NBI mode	98	5	7	138								
	Rectosigmoid polyps ≤5 mm in stained mode	96	5	14	133								
	Proximal-to-rectosigmoid polyps ≤5 mm in NBI mode	167	9	12	21	300	12	58	48	278	20	80	40
	Rectosigmoid polyps ≤5 mm in NBI mode	95	6	5	135	176	14	24	268	161	30	39	252
Mori (38)	EC images	126	8	11	31	254	7	20	71	224	19	50	59
	WLI	126	8	11	31	242	26	32	52	228	34	46	44
Patel (39)	Architecture 1	2,424	680	466	1,149								
	Architecture 2	2,071	389	819	1,440								
	Architecture 3	2,350	607	540	1,222								
	Architecture 4	2,246	547	644	1,282								
	Architecture 5	2,230	509	660	1,320								
	Architecture 6	2,239	616	651	1,213								
Renner (40)	All polyps	48	18	4	30	86	21	18	75				
	Polyps ≤5 mm	8	6	0	21	12	8	4	46				
Yamada (41)		732	64	20	638								
Ozawa (42)	WLI	562	64	14	59								
	NBI	197	31	5	37								

WLI, white light imaging; EC, endocytoscopy; NBI, narrow-band imaging.

was quantitatively analyzed, and the results are shown in *Figure 12* (P=0.13>0.05) and suggested no significant publication bias.

AI-assisted classification of colorectal polyps: a subgroup meta-analysis of diminutive polyps (≤5 mm)

A total of 8 studies reported the performance of AI-assisted classification of diminutive polyps (≤5 mm). The heterogeneity (I^2) of the sensitivity was 69.22 (P<0.01), and the heterogeneity (I^2) of the specificity was 96.86 (P<0.01). The sensitivity was 0.95 (95% CI: 0.94–0.97), and the specificity was 0.88 (95% CI: 0.74–0.95). The PLR was 8.2 (95% CI: 3.5–19.3), the NLR was 0.05 (95% CI: 0.04–0.07), and DOR was 155 (95% CI: 60–400), as shown in *Figure 13*. The AUC under SROC curve was estimated to be 0.97 (95% CI: 0.95–0.98), as shown in *Figure 14*.

Figure 2 Literature quality evaluation map.

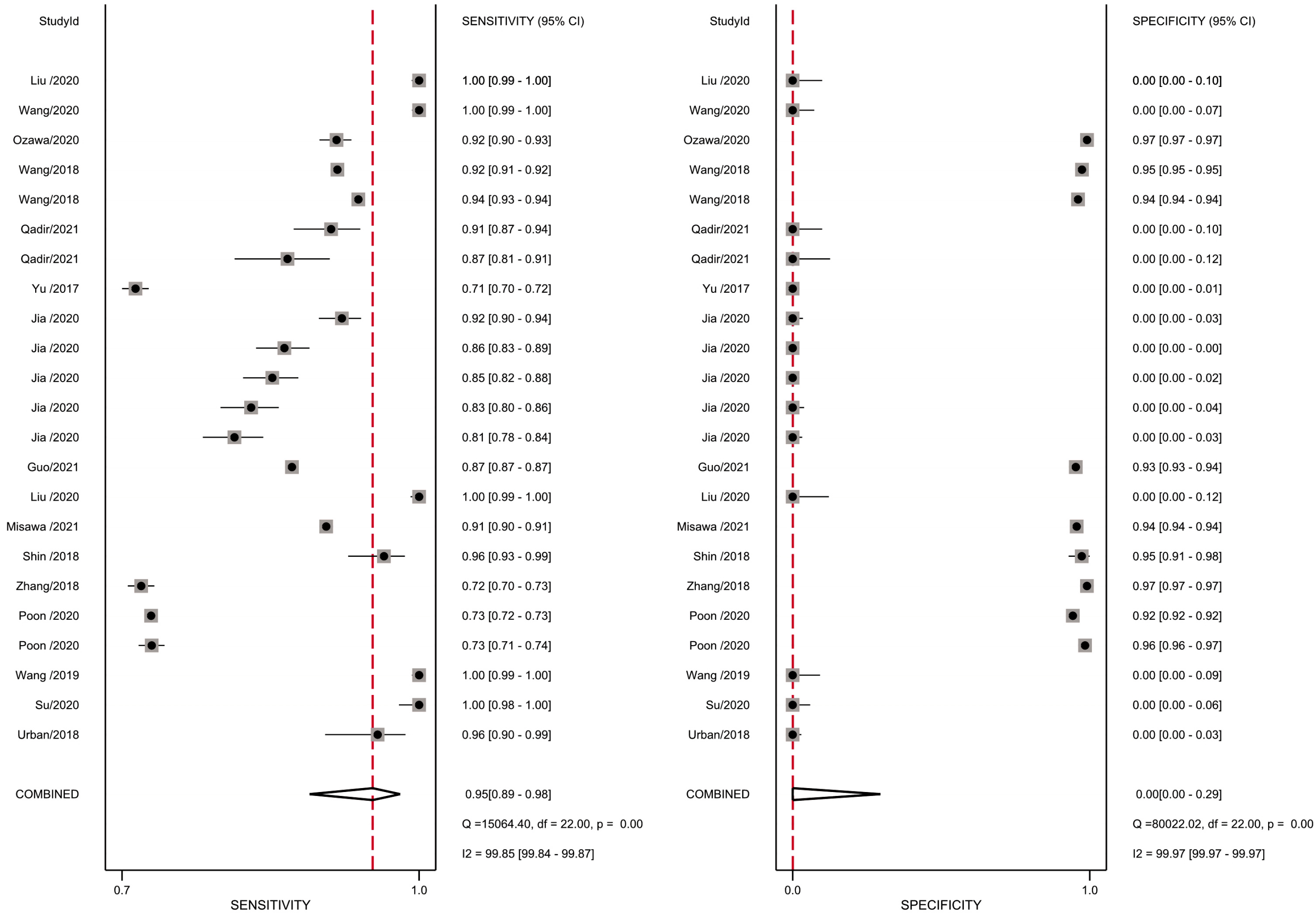

Figure 3 Meta-analysis of the sensitivity and specificity of AI-assisted polyp detection. AI, artificial intelligence.

AI-assisted classification of colorectal polyps: a subgroup meta-analysis of magnification endoscopy

A total of 4 studies reported the performance of AI-assisted classification of colorectal polyps under magnification endoscopy. The heterogeneity (I^2) of the sensitivity was 89.49 (P<0.01), and the heterogeneity (I^2) of the specificity was 93.28 (P<0.01). The sensitivity was 0.94 (95% CI: 0.92–0.96), and the specificity was 0.95 (95% CI: 0.80–0.99). The PLR was 17.4 (95% CI: 4.4–69.3), the NLR was 0.06 (95% CI: 0.04–0.09), and the DOR was 293 (95% CI: 51–1,673), as shown in *Figure 15*. The AUC under the SROC curve was estimated to be 0.97 (95% CI: 0.95–0.98), as shown in *Figure 16*.

Discussion

AI technology has been applied in many areas of clinical diagnosis and treatment, including intelligent inspection, diagnosis, treatment, monitoring, and prevention, with the common purpose of improving the quality of medical health (43). In the diagnosis and treatment of colorectal polyps under colonoscopy, the current applications of AI mainly include polyp detection and classification (44-46). The former mainly aims at improving the detection rate for polyps and adenomas, while the latter mainly focuses on the classification of neoplastic polyps and nonneoplastic polyps, with the goal of improving the quality of colonoscopy and the accuracy of endoscopists (especially young endoscopists). For the classification of colorectal polyps, AI is usually used to capture the local features of polyps involving texture, shape and color from the endoscopic target area, and summarize the hidden features in the image. The local features and hidden features are fused into AI data analysis to classify the images of neoplastic polyps and non-neoplastic polyps.

We conducted a meta-analysis to examine the current

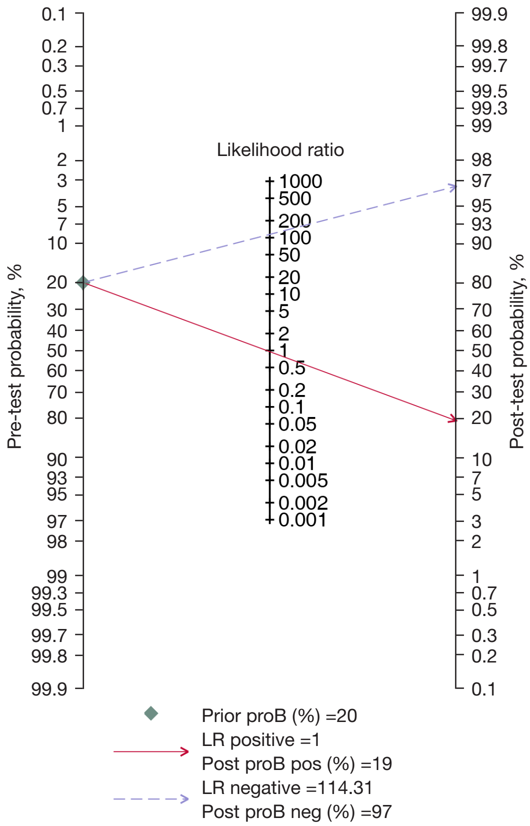

Figure 4 Bayesian analysis of posttest probability and pretest probability (polyp detection).

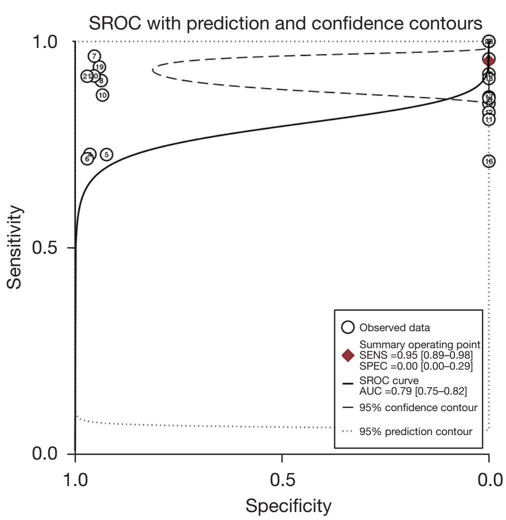

Figure 5 SROC curve of AI-assisted polyp detection. SROC, summary receiver operating characteristic curve; AI, artificial intelligence.

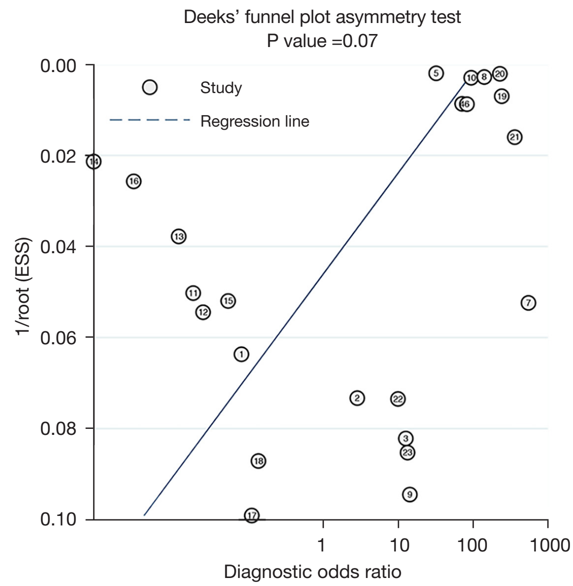

Figure 6 Funnel plot of included literature (polyp detection).

status of diagnostic performance for AI-assisted technologies in the detection and classification of colorectal polyps. We found several machine learning methods being applied for polyp detection and characterization in numerous studies. In terms of the detection of colorectal polyps, although the meta-analysis showed no prominent publication bias in the included literature, the heterogeneity was statistically significant, which may be relevant to the absence of TN in some studies. Our results highlight a high diagnostic accuracy of AI-assisted polyp detection, with a sensitivity of 95% and an AUC of 0.79. Results concerning the reliability of specificity were suspect, as there was no reported TN in some studies. Thus, we performed a subgroup analysis in studies reporting TN, and results demonstrated a sensitivity of 88% and a specificity of 95% with an AUC of 0.97, indicating a missed diagnosis rate and a misdiagnosis rate of 12% and 5%, respectively. These outcomes demonstrated good results for AI techniques in detecting polyps. Our results suggested an increase of 10% in ADR in patients with the use of AI for polyp detection compared with patients who achieved standard colonoscopy.

Various of factors may contribute to the lack of applicability of the AI techniques in clinical practice. A considerable proportion of research into AI-assisted polyp detection and has been carried out in China and Japan, but differences in polyp biology and tumorigenesis may limit the application of findings in endoscopic practice. Furthermore, only AI technologies that enable real-time detection have

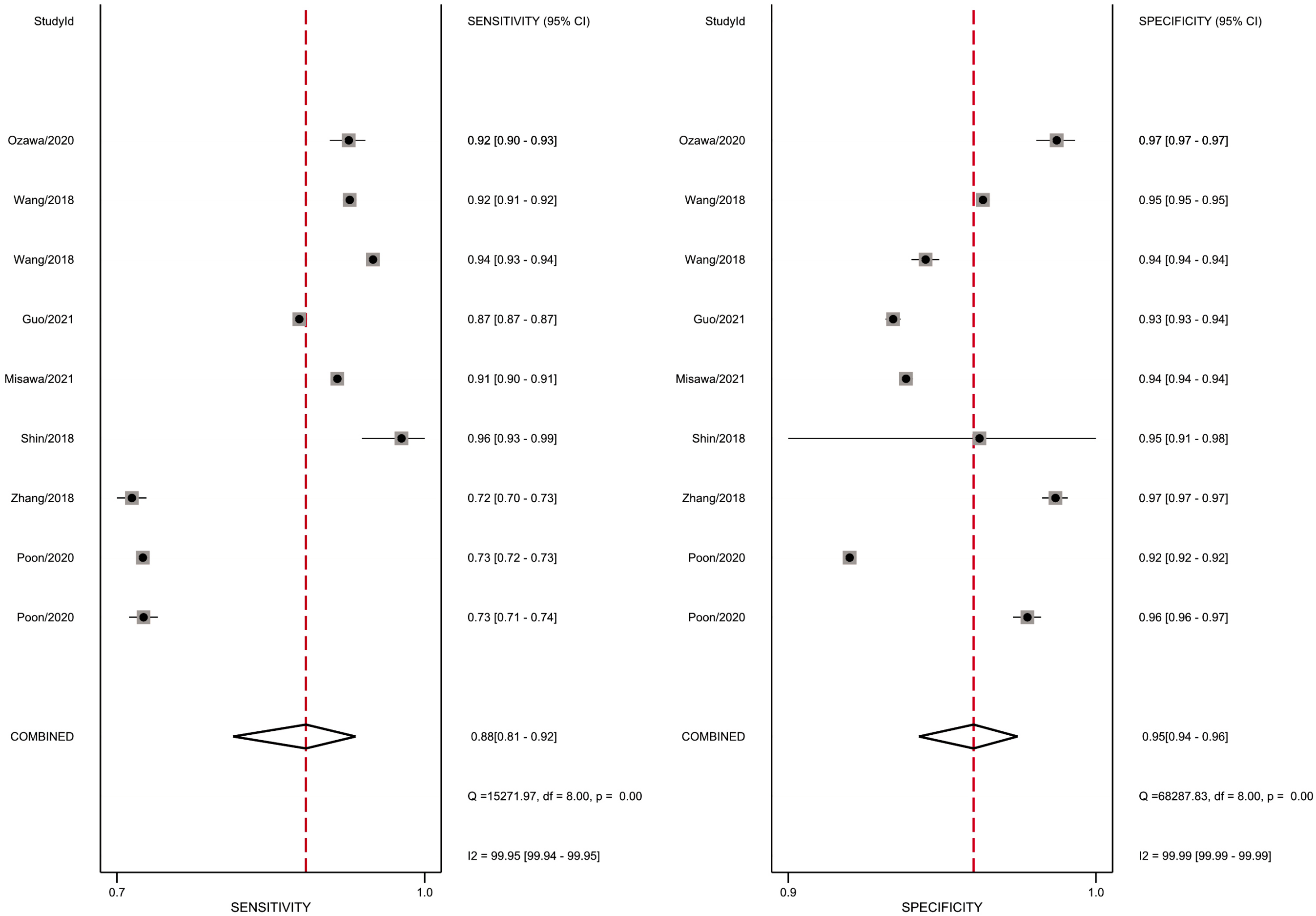

Figure 7 Meta-analysis of the sensitivity and specificity of AI-assisted polyp detection (including TN subgroup). AI, artificial intelligence; TN, true negative.

clinical application value in endoscopy. However, most recent studies used endoscopic high-quality images or videos to train and verify the performance of AI-assisted detection, which might have led to an overestimation of the AI's detection performance. Meanwhile, several published clinical studies (23,24,35) have shown that for real-time detection, the AI may be affected by the quality of intestinal preparation, intestinal mucosal folds or other intestinal diseases, and foreign bodies, resulting in false positives. Therefore, a further development of AI diagnostic models is needed to reduce interference factors in real-time detection.

Results evaluating the classification performance of AI in colorectal polyps showed no significant publication bias in the included literature. More importantly, our meta-analysis demonstrated a high diagnostic accuracy of AI-assisted polyp classification with a sensitivity of 92% and a specificity of 82%, indicating a missed diagnosis rate of 8% and a misdiagnosis rate of 18%. The pooled PLR was 5.0, suggesting that the probability of correctly classifying colorectal polyps was 5 times more than that of misclassifying. Moreover, the pooled NLR was 0.10, revealing that the probability of incorrect classification is 0.1 times higher than that of correct classification. DOR, the diagnostic odds ratio, indicated the strength of the association between the diagnostic results of tests and diseases. Our study yielded a pooled DOR of 51, indicating the high diagnostic value of AI-assisted detection and classification in polyps. Additionally, Bayesian test analysis showed that the overall correct diagnostic rate of endoscopy increased by 37% and the overall false diagnostic rate decrease d by 18% with the use of AI. The AUC of the SROC curve was 0.94, which confirmed the high value of AI in the classification of colorectal polyps. Considering the obvious heterogeneity of included studies, which may

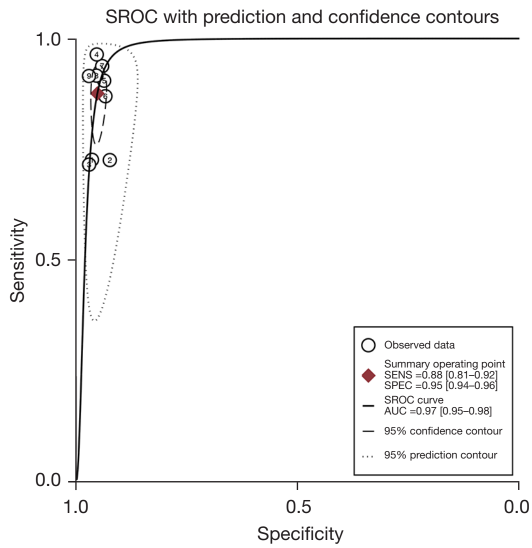

Figure 8 SROC curve of AI-assisted polyp detection (including TN subgroup). SROC, summary receiver operating characteristic curve; AI, artificial intelligence; TN, true negative.

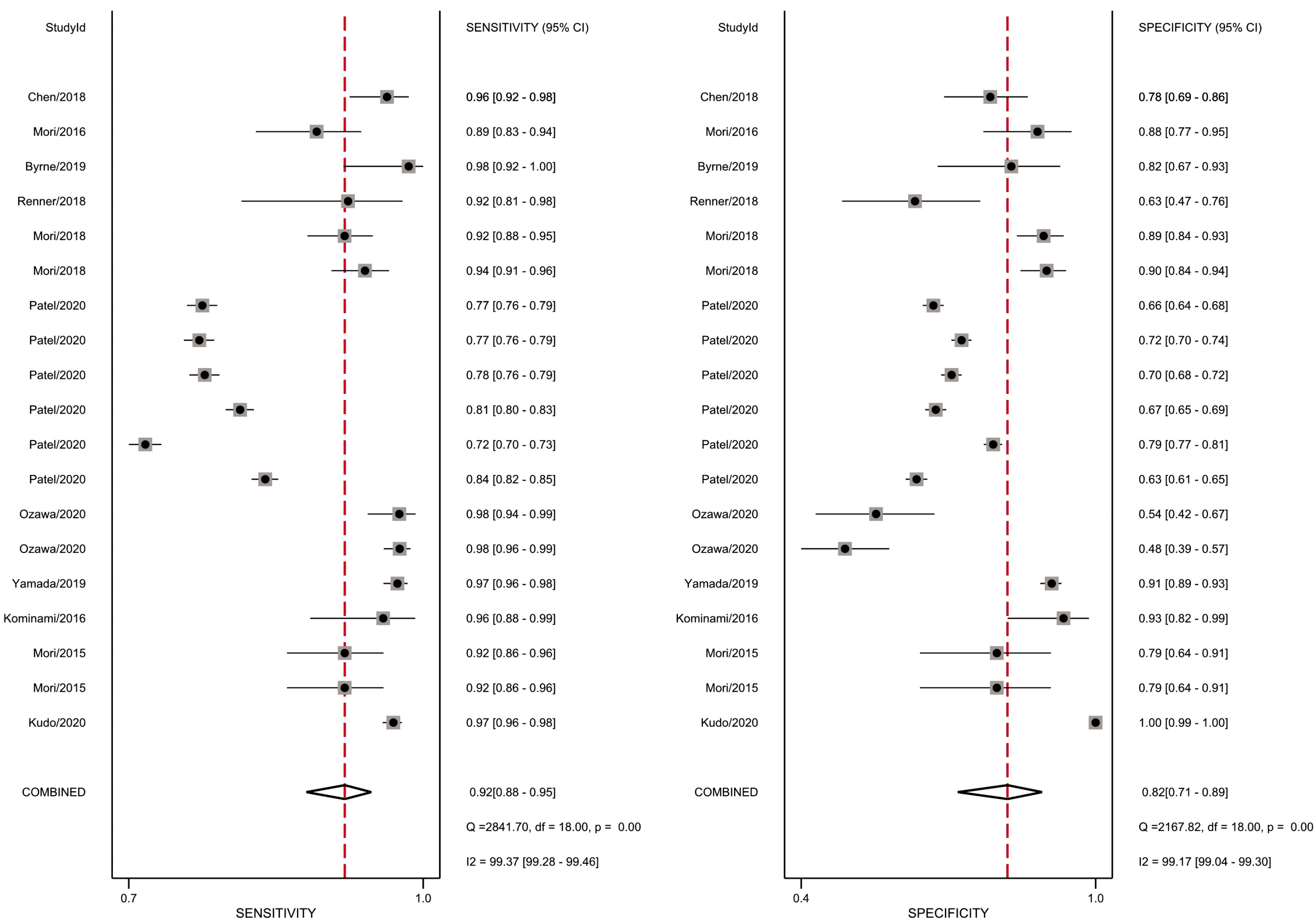

Figure 9 Meta-analysis on sensitivity and specificity of AI-assisted polyp classification. AI, artificial intelligence.

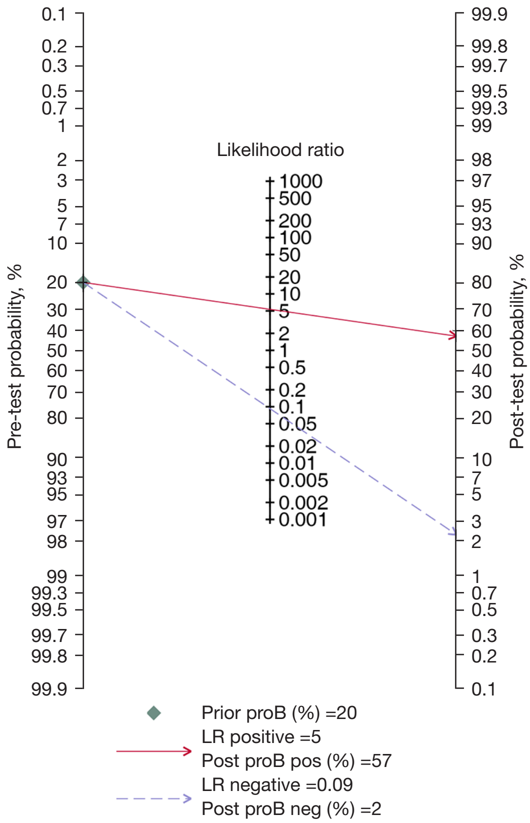

Figure 10 Bayesian analysis of posttest probability and pretest probability (polyp classification).

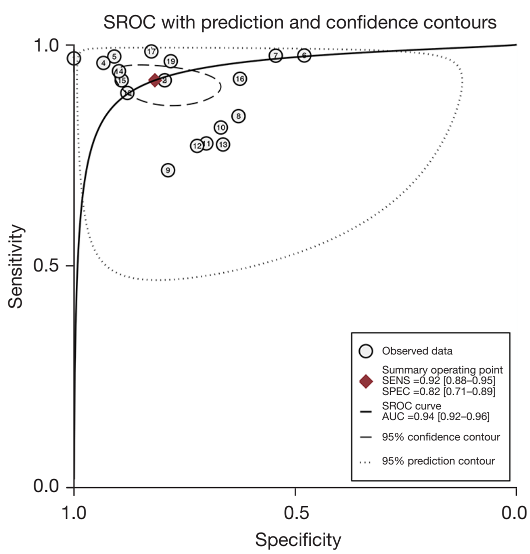

Figure 11 SROC curve of AI-assisted endoscopic polyp classification. SROC, summary receiver operating characteristic curve; AI, artificial intelligence.

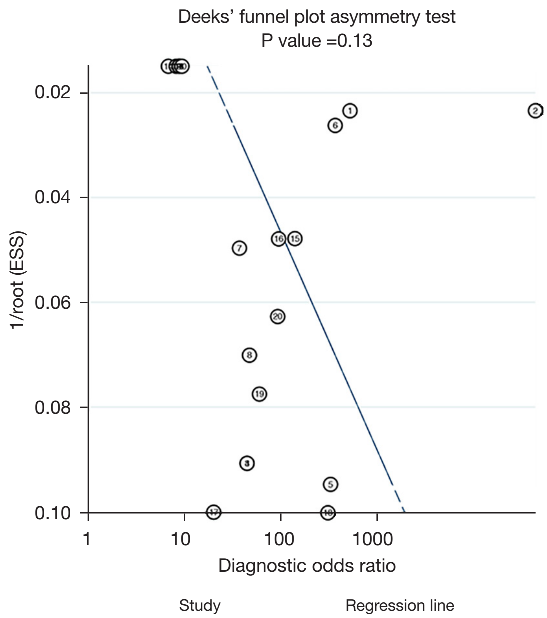

Figure 12 Funnel plot of included literature (polyp classification).

be related to differences in the size of polyps, we performed a subgroup analysis of diminutive polyps (≤5 mm). The results showed a lower heterogeneity than before, and no significant publication bias in the included literature. The sensitivity of 95% and a specificity of 88% indicated a missed diagnosis rate and misdiagnosis rate of 5% and 12%, respectively. Meanwhile, an AUC of 0.97 suggested that AI-assisted classification of diminutive polyps also has high auxiliary diagnostic value.

In the comparison of the diagnostic performance of AI, endoscopic experts, and nonexperts in the classification of colorectal polyps, a previously published meta-analysis (14) had shown the diagnostic performance of AI to be equivalent to that of endoscopic experts and significantly better than that of nonexperts. Moreover, the AUC obtained from our meta-analysis showed that AI had an extremely high diagnostic performance in the classification of polyps, while current studies comparing the classification performance of AI with experts and nonexperts seem to require further investigation.

The subgroup analysis of different types of endoscopies produced a sensitivity of 94% and a specificity of 95%, indicating a missed diagnosis rate and misdiagnosis rate of 6% and 5%, respectively. The AUC was estimated to be 0.97, suggesting a high auxiliary diagnostic value of AI-assisted classification under magnification endoscopy. Cell endoscopy is currently used in clinic, and research into AI

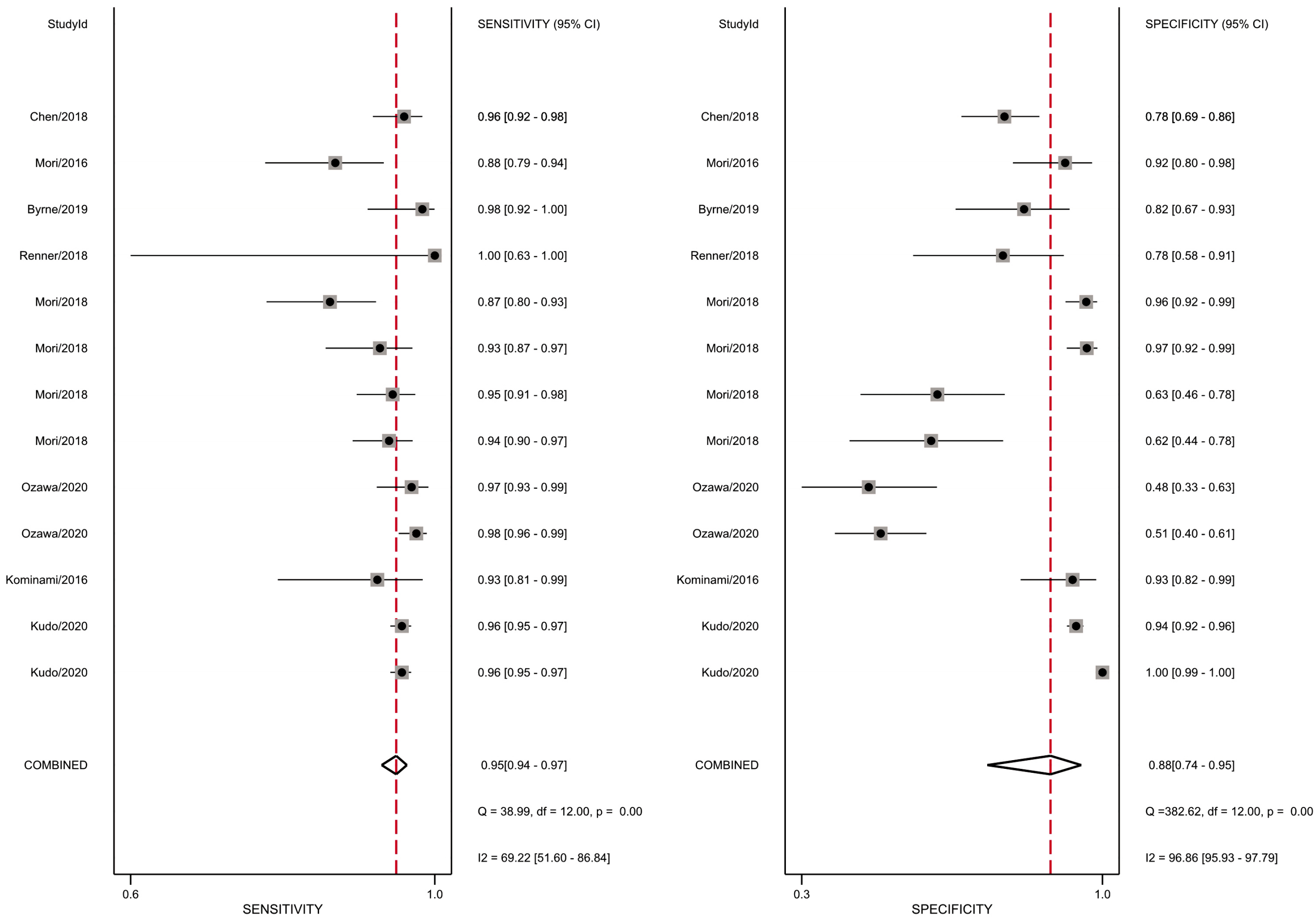

Figure 13 Meta-analysis of the sensitivity and specificity of AI-assisted diminutive polyp classification. AI, artificial intelligence.

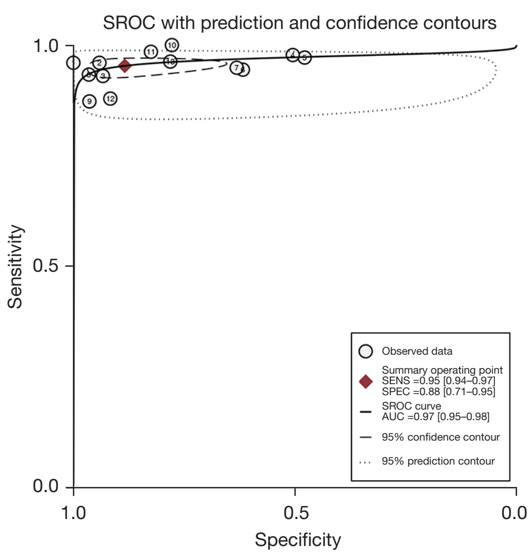

Figure 14 SROC curve of AI-assisted classification of diminutive polyps. SROC, summary receiver operating characteristic curve; AI, artificial intelligence.

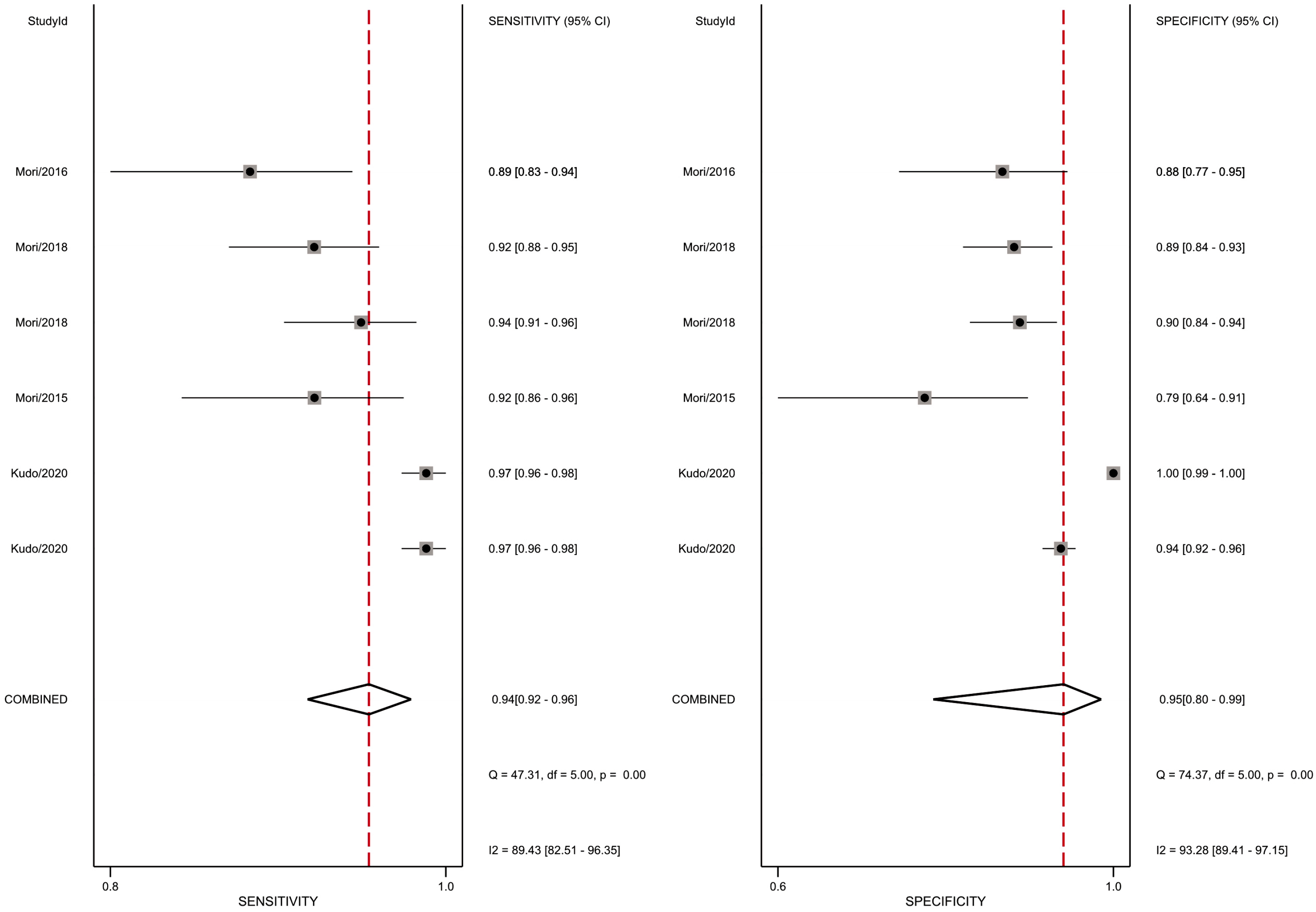

Figure 15 Meta-analysis of the sensitivity and specificity of the AI-assisted magnification endoscopy subgroup. AI, artificial intelligence.

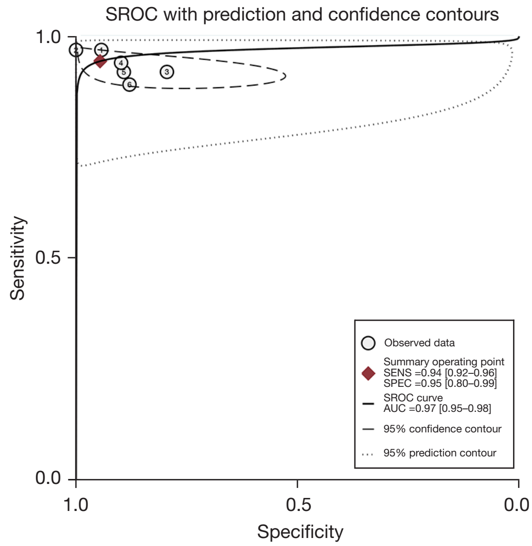

Figure 16 SROC curve of the AI-assisted magnification endoscopy subgroup. SROC, summary receiver operating characteristic curve; AI, artificial intelligence.

for polyp classification and evaluation of infiltration depth under cell endoscopy may intensify substantially in the near future.

Two inevitable limitations to our study should be acknowledged. First, due to the differences in AI systems, a large degree of heterogeneity was found among the included study groups, and thus the results should be further scrutinized. Second, a few of the including studies did not clarify the specific types of endoscopies, and the specificity and sensitivity of AI for different types of endoscopes could not be further analyzed.

In conclusion, our study demonstrated the high clinical value of AI in the detection and classification of colorectal polyps, suggest that AI may be used as a novel auxiliary diagnostic method in the upcoming years. Looking to the future, AI-assisted diagnosis should be developed to be more accurate and rapid, which will be more conducive to the real-time detection and classification of colorectal polyps and the evaluation infiltration depth. Only in this way can the application of AI in endoscopy improve the detection rate and classification accuracy of colorectal polyps and lighten the workload of endoscopists, and promote the diversified and balanced development of medical resources.

Acknowledgments

Funding: Special funding for this study was received from the Guangdong Provincial Hospital of Traditional Chinese Medicine (No. YN10101914) and the Guangzhou University of Chinese Medicine "Double First -Class" and High-level University Discipline Collaborative Innovation Team (No. 2021xk58).

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at https://dx.doi.org/10.21037/atm-21-5081

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/atm-21-5081). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Shaukat A, Kahi CJ, Burke CA, et al. ACG Clinical Guidelines: Colorectal Cancer Screening 2021. Am J Gastroenterol 2021;116:458-79.
3. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012;366:687-96.
4. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 2013;369:1095-105.
5. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370:1298-306.
6. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010;362:1795-803.
7. Ahn SB, Han DS, Bae JH, et al. The Miss Rate for Colorectal Adenoma Determined by Quality-Adjusted, Back-to-Back Colonoscopies. Gut Liver 2012;6:64-70.
8. Mahmud N, Cohen J, Tsourides K, et al. Computer vision and augmented reality in gastrointestinal endoscopy. Gastroenterol Rep (Oxf) 2015;3:179-84.
9. Liu WN, Zhang YY, Bian XQ, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26:13-9.
10. Cantisani V, Grani G, Tovoli F, et al. Artificial Intelligence: What Is It and How Can It Expand the Ultrasound Potential in the Future? Ultraschall Med 2020;41:356-60.
11. Min M, Su S, He W, et al. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. Sci Rep 2019;9:2881.
12. Misawa M, Kudo SE, Mori Y, et al. Development of a computer-aided detection system for colonoscopy and a

publicly accessible large colonoscopy video database (with video). Gastrointest Endosc 2021;93:960-967.e3.
13. Urban G, Tripathi P, Alkayali T, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology 2018;155:1069-1078.e8.
14. Xu Y, Ding W, Wang Y, et al. Comparison of diagnostic performance between convolutional neural networks and human endoscopists for diagnosis of colorectal polyp: A systematic review and meta-analysis. PLoS One 2021;16:e0246892.
15. Hassan C, Spadaccini M, Iannone A, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021;93:77-85.e6.
16. Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: A systematic review and meta-analysis. Endoscopy 2021;53:277-84.
17. Ashat M, Klair JS, Singh D, et al. Impact of real-time use of artificial intelligence in improving adenoma detection during colonoscopy: A systematic review and meta-analysis. Endosc Int Open 2021;9:E513-E521.
18. Li J, Lu J, Yan J, et al. Artificial intelligence can increase the detection rate of colorectal polyps and adenomas: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol 2021;33:1041-48.
19. Aziz M, Fatima R, Dong C, et al. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: A systematic review with meta-analysis. J Gastroenterol Hepatol 2020;35:1676-83.
20. Qadir HA, Shin Y, Solhusvik J, et al. Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction. Med Image Anal 2021;68:101897.
21. Guo Z, Nemoto D, Zhu X, et al. Polyp detection algorithm can detect small polyps: Ex vivo reading test compared with endoscopists. Dig Endosc 2021;33:162-9.
22. Jia X, Mai X, Cui Y, et al. Automatic Polyp Recognition in Colonoscopy Images Using Deep Learning and Two-Stage Pyramidal Feature Prediction. IEEE Transactions on Automation Science and Engineering 2020;17:1570-84.
23. Liu P, Wang P, Glissen Brown JR, et al. The single-monitor trial: an embedded CADe system increased adenoma detection during colonoscopy: a prospective randomized study. Therap Adv Gastroenterol 2020;13:1756284820979165.
24. Poon CCY, Jiang Y, Zhang R, et al. AI-doscopist: a real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices. NPJ Digit Med 2020;3:73.
25. Shin Y, Balasingham I. Automatic polyp frame screening using patch based combined feature and dictionary learning. Comput Med Imaging Graph 2018;69:33-42.
26. Su JR, Li Z, Shao XJ, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020;91:415-424.e4.
27. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813-9.
28. Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020;5:343-51.
29. Wang P, Xiao X, Glissen Brown JR, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng 2018;2:741-8.
30. Yu L, Chen H, Dou Q, et al. Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos. IEEE J Biomed Health Inform 2017;21:65-75.
31. Zhang R, Zheng Y, Poon CCY, et al. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. Pattern Recognit 2018;83:209-19.
32. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019;68:94-100.
33. Chen PJ, Lin MC, Lai MJ, et al. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. Gastroenterology 2018;154:568-75.
34. Kominami Y, Yoshida S, Tanaka S, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointest Endosc 2016;83:643-9.
35. Kudo SE, Misawa M, Mori Y, et al. Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. Clin Gastroenterol Hepatol 2020;18:1874-81.e2.
36. Mori Y, Kudo SE, Chiu PW, et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal

lesions: an international web-based study. Endoscopy 2016;48:1110-8.
37. Mori Y, Kudo SE, Misawa M, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med 2018;169:357-66.
38. Mori Y, Kudo SE, Wakamura K, et al. Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). Gastrointest Endosc 2015;81:621-9.
39. Patel K, Li K, Tao K, et al. A comparative study on polyp classification using convolutional neural networks. PLoS One 2020;15:e0236452.
40. Renner J, Phlipsen H, Haller B, et al. Optical classification of neoplastic colorectal polyps - a computer-assisted approach (the COACH study). Scand J Gastroenterol 2018;53:1100-6.
41. Yamada M, Saito Y, Imaoka H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. Sci Rep 2019;9:14465.
42. Ozawa T, Ishihara S, Fujishiro M, et al. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therap Adv Gastroenterol 2020;13:1756284820910659.
43. Ahmed Z, Mohamed K, Zeeshan S, et al. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford) 2020;2020:baaa010.
44. Rees CJ, Koo S. Artificial intelligence - upping the game in gastrointestinal endoscopy? Nat Rev Gastroenterol Hepatol 2019;16:584-5.
45. Kim KO, Kim EY. Application of Artificial Intelligence in the Detection and Characterization of Colorectal Neoplasm. Gut Liver 2021;15:346-53.
46. Misawa M, Kudo SE, Mori Y, et al. Current status and future perspective on artificial intelligence for lower endoscopy. Dig Endosc 2021;33:273-84.

Cite this article as: Wang A, Mo J, Zhong C, Wu S, Wei S, Tu B, Liu C, Chen D, Xu Q, Cai M, Li Z, Xie W, Xie M, Kato M, Xi X, Zhang B. Artificial intelligence-assisted detection and classification of colorectal polyps under colonoscopy: a systematic review and meta-analysis. Ann Transl Med 2021;9(22):1662. doi: 10.21037/atm-21-5081