Digestive Endoscopy 2025; **37**: 933–947 doi: 10.1111/den.15047

Review

Computer-aided diagnosis for colorectal polyp in comparison with endoscopists: Systematic review and meta-analysis

Satoshi Shinozaki,[1,2] Jun Watanabe,[3,4] Takeshi Kanno,[5,6] Yuhong Yuan,[7,8] Tomonori Yano[2] and Hironori Yamamoto[2]

[1]Shinozaki Medical Clinic, [2]Division of Gastroenterology, Department of Medicine, [3]Division of Gastroenterological, General and Transplant Surgery, Department of Surgery, [4]Division of Community and Family Medicine, [5]R & D Division of Career Education for Medical Professionals, Medical Education Center, Jichi Medical University, Tochigi, [6]Division of Gastroenterology, Tohoku University Graduate School of Medicine, Miyagi, Japan, [7]Department of Medicine, London Health Science Centre, London and [8]Department of Medicine, McMaster University, Hamilton, ON, Canada

Objectives: Computer-aided diagnosis (CADx) is anticipated to enhance the prediction of colorectal polyp histology. This study aims to compare the diagnostic accuracy of CADx in the optical diagnosis of colorectal polyps, evaluating its performance against that of both experienced and inexperienced endoscopists.

Methods: The protocol of this study was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (ID: CRD42024585097). Three electronic databases including MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL) were searched in September 2024. A bivariate random effects model was employed. The primary outcome was the comparison of sensitivity and specificity between CADx and experienced endoscopists; the secondary outcome was the comparison between CADx and inexperienced endoscopists.

Results: Twenty-one studies involving 5477 polyps were included. The pooled sensitivities of CADx and experienced endoscopists were 0.87 (95% confidence interval [CI] 0.82–0.91) and 0.88 (95% CI 0.83–0.91), respectively ($P = 0.93$). The pooled specificities of CADx and experienced endoscopists were 0.85 (95% CI 0.78–0.90) and 0.87 (95% CI 0.82–0.92), respectively ($P = 0.53$). In nine studies comparing CADx with inexperienced endoscopists, the pooled sensitivities were 0.88 (95% CI 0.82–0.92) for CADx and 0.85 (95% CI 0.78–0.90) for inexperienced endoscopists ($P = 0.46$). The pooled specificities were 0.84 (95% CI 0.78–0.88) for CADx and 0.77 (95% CI 0.70–0.83) for inexperienced endoscopists ($P = 0.16$).

Conclusion: Computer-aided diagnosis does not demonstrate superior diagnostic accuracy in optical diagnosis of colorectal polyps compared to endoscopists, regardless of their experience level.

Key words: artificial intelligence, colonoscopy, colorectal polyp, computer-aided diagnosis, endoscopy

INTRODUCTION

COLORECTAL CANCER IS the second leading cause of cancer-related death worldwide, as reported by the global cancer statistics 2020.[1] Many colorectal cancers develop from small polyps, and colonoscopic resection of adenomatous polyps has been shown to reduce colorectal cancer mortality by 53%.[2] Accurate detection and removal of neoplastic polyps are therefore crucial in preventing the progression to cancer. Precise diagnosis based on surface pattern recognition through magnifying observation with image-enhanced endoscopy (IEE) is necessary to distinguish between neoplastic and nonneoplastic lesions.

However, endoscopic differentiation between neoplastic and nonneoplastic polyps is sometimes challenging, even for experienced endoscopists. The risk of false-negatives—failing to identify a neoplastic polyp—can lead to missed opportunities for early intervention, potentially resulting in cancer progression. Conversely, false-positives—incorrectly identifying a nonneoplastic polyp as neoplastic—can lead to unnecessary polypectomies, exposing patients to unwarranted procedural risks and health-care costs. These diagnostic inaccuracies underscore the need for technologies that can enhance diagnostic performance and minimize the

Corresponding: Takeshi Kanno, R & D Division of Career Education for Medical Professionals, Medical Education Center, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 3290498, Japan. Email: kanno.takeshi@jichi.ac.jp
Received 23 November 2024; accepted 21 April 2025.

risks associated with false-negatives and false-positives. Predicting pathological diagnosis based on meticulous observation is sometimes difficult for both experienced and inexperienced endoscopists.

The development and dissemination of computer-aided diagnosis (CADx) systems utilizing artificial intelligence (AI) have the potential to revolutionize endoscopic diagnostics. CADx systems can assist in accurately distinguishing between neoplastic and nonneoplastic polyps, potentially reducing interobserver variability and bridging the skill gap between experienced and inexperienced endoscopists. Recent studies have reported on the distinguishing ability of CADx between neoplastic and nonneoplastic polyps,[3] but the reported sensitivity and specificity vary widely.[4,5] While CADx may reduce the risk of false-negatives by enhancing diagnostic accuracy, there is also a concern that it may increase false-positives, leading to overtreatment.

Despite the growing body of research, few systematic reviews[4,6] have comprehensively investigated the usefulness of CADx in the diagnosis of colorectal polyps. This study aims to compare the diagnostic accuracy of CADx in the optical diagnosis of colorectal polyps, evaluating its performance against that of both experienced and inexperienced endoscopists. We sought to evaluate whether CADx can enhance diagnostic precision, reduce misdiagnosis risks, and ultimately improve patient outcomes.

METHODS

THIS STUDY WAS conducted according to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy[7] and the Preferred Reporting Items for Systematic Review and Meta-analysis of Diagnostic Test Accuracy (PRISMA-DTA) (Appendix S1).[8] This study protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (Registration ID: CRD42024585097) and published this protocol in the Open Science Framework (OSF) (https://osf.io/2bajy/).

Eligibility criteria

In this review, CADx was defined as the use of deep-learning algorithms to analyze images or video during colonoscopy, assisting in the histological prediction between neoplastic and nonneoplastic polyps. The reference standard was pathological diagnosis, and therefore only resected colorectal polyps were included. Patients with colorectal polyps evaluated by CADx and endoscopists were included.

The primary outcome was a comparison of sensitivity and specificity between CADx and experienced endoscopists, and the secondary outcome was comparisons of the sensitivity and specificity between CADx and inexperienced endoscopists. In this study, the definition of experienced endoscopists was limited to the following criteria: (i) experience of more than 200 colonoscopies; (ii) more than 5 years of experience as an endoscopist; or (iii) board-certified endoscopists.[9] Inexperienced endoscopists were defined as those who did not meet the criteria for experienced endoscopists, as specified above. We accepted the original authors' definitions of an experienced endoscopist to account for variations in experience levels and certification requirements across different countries and institutions. We considered sessile serrated lesions (SSLs) as part of the neoplastic or nonneoplastic classification based on the definitions used in each included study.

The inclusion criteria of the systematic review were as follows: (i) studies designed to assess the diagnostic performance of CADx alone and endoscopists for prediction of histopathology based on endoscopic observations including randomized controlled trials, case–control studies, and cohort studies; (ii) studies involving patients with colorectal polyps evaluated by CADx and endoscopists; (iii) neoplastic or nonneoplastic (binary) determination based on colonoscopic images of colorectal polyps, regardless of the use of IEE; (iv) sensitivity and specificity were reported or could be calculated; and (v) comparison between CADx and endoscopists was evaluated. We excluded the following studies: (i) patients under 18-years-old; (ii) patients with inflammatory bowel disease; (iii) use of endocytoscopy or endomicroscopy; (iv) no comparison with pathological diagnosis; (v) combination diagnosis of AI and endoscopists (comparison between CADx-assisted and CADx-unassisted groups); (vi) studies that did not report critical data; and (vii) review articles or case reports. We did not exclude studies based on real-time or still imaging. Any clinical indications were included. The exclusion of studies comparing CADx-assisted and CADx-unassisted groups was based on the fact that a past systematic review has already extensively evaluated this topic.[6]

Search strategy

We conducted a comprehensive bibliographic search of the following databases from 2015 to September 3, 2024: MEDLINE (Ovid), Embase (Ovid), and the Cochrane Central Register of Controlled Trials (CENTRAL) (Ovid). This search strategy was designed by a systematic review methodologist (Y.Y.). We searched for text words as well as controlled vocabulary (MeSH or Emtree terms) related to colonic polyps or adenomas and colonoscopy or endoscopy, and AI or computer-assisted diagnosis or detection, using

Boolean operators (AND, OR, NOT). Synonyms and variations of root words were also searched. The search was limited to studies fully published in English. The starting year of 2015 was chosen to reflect the availability of AI or computer-assisted trials in colonoscopy diagnosis.[10] The detailed search strategy is included in Appendix S2. A recursive manual search of the bibliographies of eligible studies and systematic reviews was conducted to identify additional studies. Additionally, experts in this field were contacted to ensure the identification of all eligible studies.

We also searched the Guidelines of European Society of Gastrointestinal Endoscopy (ESGE),[11] American Society for Gastrointestinal Endoscopy (ASGE),[12] Japan Gastroenterological Endoscopy Society (JGES),[13] the World Health Organization International Clinical Trials Platform Search Portal (ICTRP), and ClinicalTrials.gov for ongoing trials.

Study selection

Two authors (S.S. and J.W.) independently searched and screened the titles and abstracts of identified articles. After initial screening, full-text articles were independently assessed for eligibility based on the inclusion and exclusion criteria. Duplicate studies were excluded. After the first author (S.S.) abstracted the data, another author (J.W.) verified the data. Any discrepancies during the screening and data extraction process were resolved through discussion between the two reviewers. If consensus could not be reached, the third author (T.K.) adjudicated.

Data extraction and quality assessment

We extracted the following data from published articles: year of publication, country, location of colorectal polyps, prevalence of adenoma, definition of experienced endoscopists, type of IEE, type of CADx system, and type of endoscopic imaging. When critical data were not clearly stated, we requested detailed information from the corresponding authors by direct contact.

Risk of bias

The first and second authors independently evaluated the risk of bias and applicability using the Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) according to our review questions.[14] The QUADAS-2 tool assesses the risk of bias in four domains: patient selection, index test, reference standard, and flow and timing. Applicability concerns are also evaluated for each domain. Each study was rated as having "low," "high," or "unclear" risk of bias in each domain. Two reviewers (S.S. and J.W.) piloted the QUADAS-2 tool on a sample of studies to ensure consistency in interpretation. The results of the quality assessment were presented in both a "Risk of Bias" summary table and a graphical representation. In case of disagreements between two authors, the third author (T.K.) was involved in reaching a consensus. Detailed strategy is described in Appendix S3.

Data synthesis and statistical analysis

The two authors (S.S. and J.W.) independently extracted the designated data and cross-checked each other's work. In case of discrepancies, they discussed them with the third author (T.K.) to reach a consensus. We performed a meta-analysis using a bivariate random-effects model to calculate pooled estimates of sensitivity and specificity with corresponding 95% confidence intervals (CIs).[15] Statistical analyses were performed using Meta-DiSc 2.0 software (the Clinical Biostatistics Unit of the Ramon y Cajal Research Institute, Madrid, Spain). Pooled odds ratios with 95% CIs were also calculated where appropriate. A random-effects model was used because significant variability among studies was expected due to differences in study design, populations, and CADx systems. Based on the Cochrane Handbook, we did not perform univariate tests for sensitivity and specificity or calculate estimates of the I^2 statistic, as these methods do not account for heterogeneity attributable to phenomena such as threshold effects.[7] We assessed heterogeneity through visual inspection of forest plots and summary receiver operating characteristic (SROC) plots by analyzing the variation in study results. Statistical significance was defined as a P-value <0.05.

Subgroup analyses

To mitigate the influence of variability and explore potential sources of heterogeneity, we performed the following subgroup analyses between still and real-time imaging groups and between studies conducted in Japan and those conducted in other countries.

Sensitivity analysis

To assess whether the results were robust enough for the conclusions drawn in the review, we performed a sensitivity analysis including only studies with clear definitions of experienced endoscopists. This helped determine if the inclusion of studies with varying definitions of expertise affected the overall findings.

RESULTS

Study selection

THE STUDY SELECTION process is detailed in Figure 1. As of September 3, 2024, 1056 records were identified from databases and registers. After reviewing titles and abstracts, 1018 records were excluded, leaving 38 records for full-text screening. Following the screening, 20 additional records were excluded. Citation and reference searching identified three additional articles. In total, 21 studies involving 5477 polyps were included in the quantitative analysis.[16–36]

Characteristics of studies

Among the 21 studies, 17 studies evaluated polyps located throughout the colorectum (Table 1). Except for five studies, at least one IEE modality was used. Real-time imaging became more prevalent in the latter half of the timeline. The prevalence of adenoma ranged from 13% to 83%. Data comparing CADx with inexperienced endoscopists were available from nine studies (43%). SSLs were classified as neoplastic and nonneoplastic lesions in five (24%) and 10 studies (48%), respectively. The definition of experienced endoscopists was clearly provided in 20 of 21 studies (95%), and only these studies were analyzed in sensitivity analysis.

Risk of bias

The risk of bias assessed using the QUADAS-2 tool is presented in Appendix S4. Six of the 21 studies were rated as having an unclear risk for patient selection due to insufficient reporting of exclusion criteria. All studies were rated as having a low risk of bias regarding applicability concerns.

Diagnostic quality comparison between CADx and endoscopists

The pooled sensitivities of CADx and experienced endoscopist diagnoses from 21 studies were 0.87 (95% CI 0.82–0.91) and 0.88 (95% CI 0.83–0.91), respectively, with no significant difference (P = 0.93) (Fig. 2a). The pooled specificities of CADx and experienced endoscopist diagnoses were 0.85 (95% CI 0.78–0.90) and 0.87 (95% CI 0.82–0.92), respectively (P = 0.53) (Fig. 2b).

In nine studies comparing CADx and inexperienced endoscopist performance, the pooled sensitivities for CADx and inexperienced endoscopists were 0.88 (95% CI

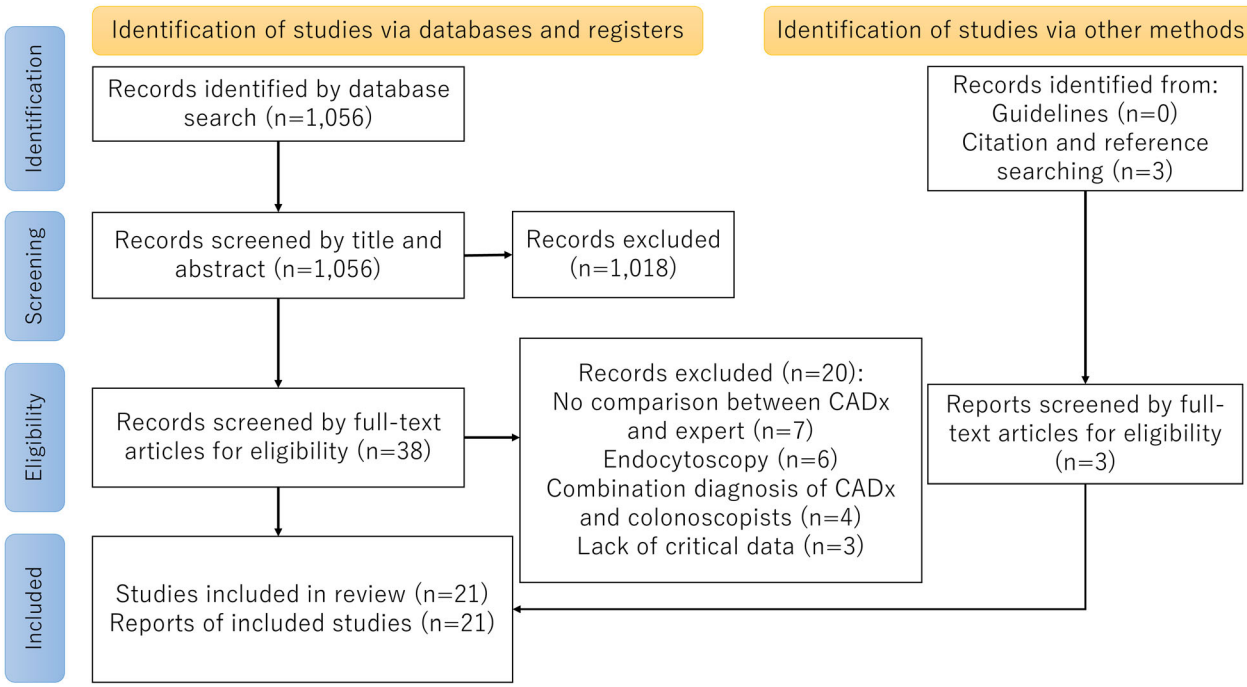

Figure 1 Flowchart of the selection process. CADx, computer-aided diagnosis.

Table 1 Characteristics of 21 studies included in this meta-analysis

First author	Year	Country	Organ	Image-enhanced endoscopy	CADx system	Images
Kominami	2016	Japan	Colorectum	NBI	Support vector machine	Real-time
Tamai	2017	Japan	Colorectum	NBI	Moving average method	Still
Chen	2018	Taiwan	Colorectum	NBI	DNN	Still
Renner	2018	Germany	Colorectum	WLI/NBI	DNN	Still
Sánchez-Montes	2019	Spain	Rectosigmoid	WLI	Support vector machine	Still
Lui	2019	China	Colorectum	WLI/NBI	CNN	Still
Min	2019	China	Colorectum	LCI	Gaussian mixture model	Still
Jin	2020	Korea	Colorectum	NBI	CNN	Still
Yoshida	2021	Japan	Colorectum	BLI	CAD EYE	Real-time
Weigt	2022	Germany	Colorectum	BLI	CAD EYE	Real-time
Hassan	2022	Italy	Rectosigmoid	WLI/BLI/NBI	GI genius	Real-time
Minegishi	2022	Japan	Colorectum	NBI	NBI-CAD	Real-time
Nemoto	2022	Japan	Proximal colon	WLI	CNN (ResNet-50)	Still
Biffi	2022	Italy	Colorectum	WLI	GI genius	Real-time
Hossain	2023	UK	Colorectum	WLI	WISE VISION	Still
Rondonotti	2023	Italy	Rectosigmoid	BLI	CAD EYE	Real-time
Li	2023	Singapore	Colorectum	BLI	CAD EYE	Real-time
Dos Santos	2023	Brazil	Colorectum	BLI	CAD EYE	Real-time
Houwen	2023	Netherland	Colorectum	NBI	CNN (POLAR)	Real-time
Baumer	2023	Germany	Colorectum	WLI	GI genius	Real-time
Lange	2024	Switzerland	Colorectum	BLI	CAD EYE	Real-time

First author	Polyp, *n*	Adenoma, *n*	Size of polyp	Data of inexperienced endoscopist	Classification of sessile serrated lesions	Definition of experienced endoscopists
Kominami	118	73 (62%)	Not reported	Absent	Excluded	>6-year experience
Tamai	121	100 (83%)	Not reported	Absent	Not mentioned	>5000 colonoscopies
Chen	284	188 (66%)	<5 mm	Present	Excluded	>5-year experience
Renner	100	52 (52%)	Mean 4 mm	Absent	Excluded	>200 colonoscopies
Sánchez-Montes	225	142 (63%)	Mean 10.4 mm	Absent	Nonneoplastic	Not described
Lui	76	56 (75%)	Mean 27 mm	Present	Neoplastic	>2000 colonoscopies
Min	181	115 (64%)	Mean 8 mm	Present	Neoplastic	>5000 colonoscopies
Jin	300	180 (60%)	<5 mm	Present	Excluded	Board certified endoscopist
Yoshida	100	55 (55%)	Mean 4 mm	Present	Nonneoplastic	>5000 colonoscopies
Weigt	134	97 (72%)	Not reported	Present	Nonneoplastic	>10,000 colonoscopies
Hassan	295	39 (13%)	<5 mm	Absent	Nonneoplastic	>2000 colonoscopies
Minegishi	395	259 (66%)	<5 mm	Absent	Neoplastic	Board certified endoscopist with >5-year experience
Nemoto	215	87 (40%)	Median 7 mm	Absent	Neoplastic	Board certified endoscopist with >10,000 colonoscopies
Biffi	513	198 (39%)	Not reported	Present	Nonneoplastic	>5-year experience
Hossain	115	80 (70%)	Mean 5 mm	Absent	Nonneoplastic	>500 colonoscopies
Rondonotti	596	259 (43%)	<5 mm	Present	Nonneoplastic	Board certified endoscopist
Li	661	408 (62%)	Median 4 mm	Absent	Excluded	Credentialed endoscopist who underwent >3-year structured training program
Dos Santos	110	80 (73%)	Mean 4 mm	Absent	Nonneoplastic	14-year experience

Table 1 (*Continued*)

First author	Polyp, *n*	Adenoma, *n*	Size of polyp	Data of inexperienced endoscopist	Classification of sessile serrated lesions	Definition of experienced endoscopists
Houwen	423	341 (81%)	<5 mm	Absent	Neoplastic	Endoscopists participating in the Dutch or Barcelona Bowel Cancer Screening program
Baumer	262	158 (60%)	<10 mm	Present	Neoplastic	>5-year experience
Lange	253	152 (60%)	Mean 5 mm	Absent	Nonneoplastic	>10-year experience

First author	Number of endoscopists	Study design	Setting
Kominami	2	Retrospective	Single center
Tamai	2	Retrospective	Single center
Chen	6	Retrospective	Single center
Renner	2	Prospective	Single center
Sánchez-Montes	9	Prospective	Single center
Lui	3	Retrospective	Single center
Min	4	Prospective	Single center
Jin	22	Retrospective	Multicenter
Yoshida	10	Retrospective	Single center
Weigt	6	Prospective	Multicenter
Hassan	4	Prospective	Single center
Minegishi	11	Prospective	Single center
Nemoto	2	Retrospective	Single center
Biffi	21	Prospective	Single center
Hossain	7	Prospective	Single center
Rondonotti	18	Prospective	Multicenter
Li	21	Prospective	Multicenter
Dos Santos	1	Prospective	Single center
Houwen	20	Prospective	Multicenter
Baumer	9	Prospective	Single center
Lange	9	Prospective	Multicenter

BLI, blue laser imaging; CADx, computer-aided diagnosis; CNN, convolutional neural network; DNN, deep neural network; LCI, linked color imaging; NBI, narrow-band imaging; WLI, white light imaging.

0.82–0.92) and 0.85 (95% CI 0.78–0.90), respectively ($P = 0.46$) (Fig. 2c). The pooled specificities for CADx and inexperienced endoscopists were 0.84 (95% CI 0.78–0.88) and 0.77 (95% CI 0.70–0.83), respectively ($P = 0.16$) (Fig. 2d).

To estimate the heterogeneity, we used SROC curves for visual inspection (Fig. 3). The SROC curve of inexperienced endoscopists shows a broader prediction ellipse compared to that of CADx, indicating greater variability in diagnostic accuracy (Fig. 3b). In contrast, the comparison with experienced endoscopists demonstrates a narrower prediction ellipse, suggesting more consistent performance across studies involving experienced endoscopists (Fig. 3a).

The area under the curve (AUC) was calculated and compared. There was no statistically significant difference in AUC between CADx and experienced endoscopists (0.93 [95% CI 0.91–0.95] vs. 0.94 [95% CI 0.92–0.96], respectively; $P = 0.49$). However, the AUC for CADx (0.92 [95% CI 0.90–0.94]) was significantly higher than that for inexperienced endoscopists (0.87 [95% CI 0.84–0.90]) ($P < 0.01$), despite a slight overlap in confidence intervals.

Subgroup analysis

A subgroup analysis was performed to investigate the impact of sequential endoscopic imaging of CADx data. We compared real-time imaging ($n = 12$) with still imaging ($n = 9$), and the sensitivity and specificity between the two groups showed no significant differences (Fig. S1, Table S1). A country-based subgroup analysis comparing studies conducted in Japan ($n = 5$) and those

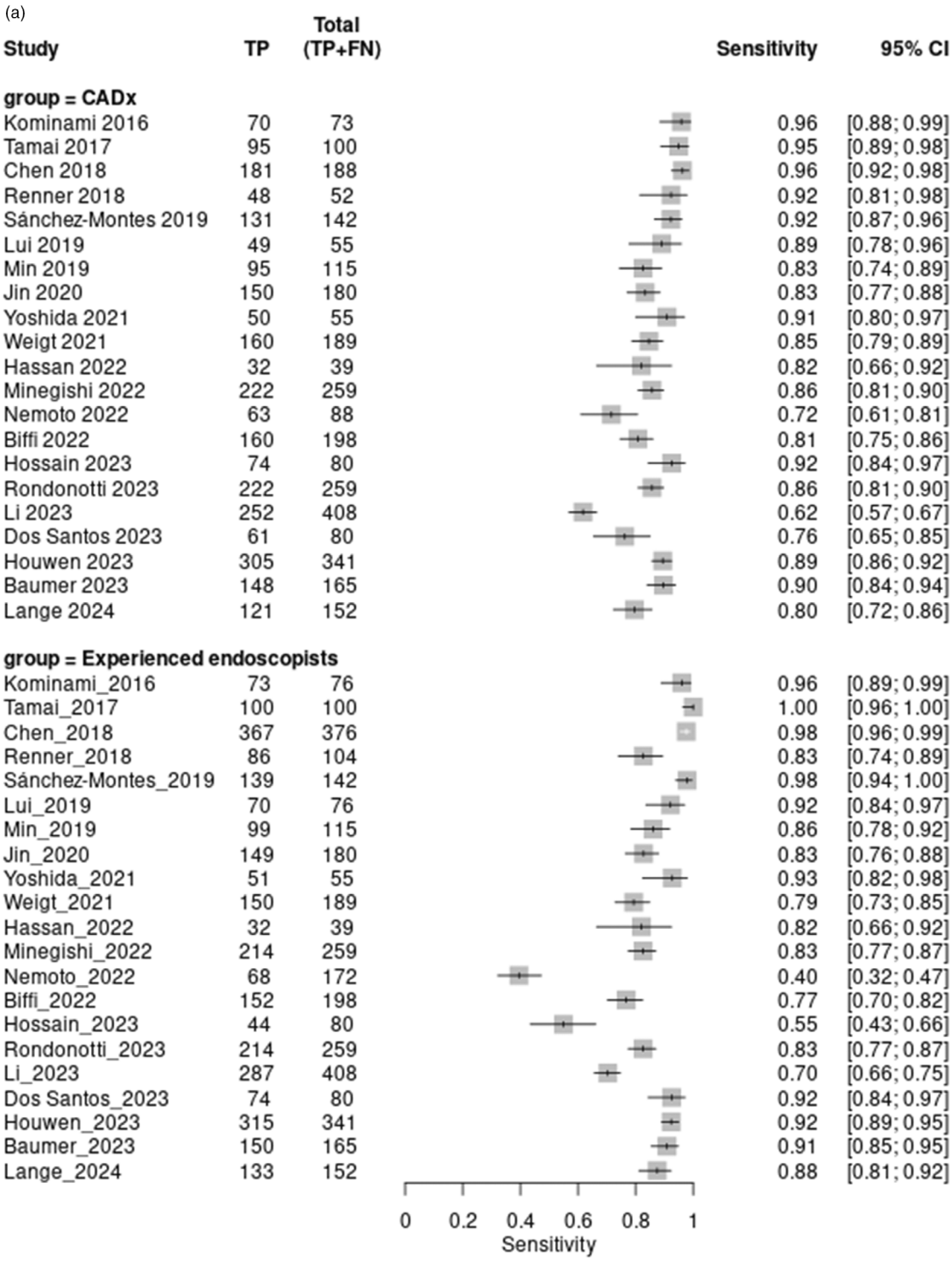

Figure 2 Forest plots: (a) sensitivity comparison between computer-aided diagnosis (CADx) and experienced endoscopists; (b) specificity comparison between CADx and experienced endoscopists; (c) sensitivity comparison between CADx and inexperienced endoscopists; (d) specificity comparison between CADx and inexperienced endoscopists. CI, confidence interval; FN, false-negative; FP, false-positive; TN, true-negative; TP, true-positive.

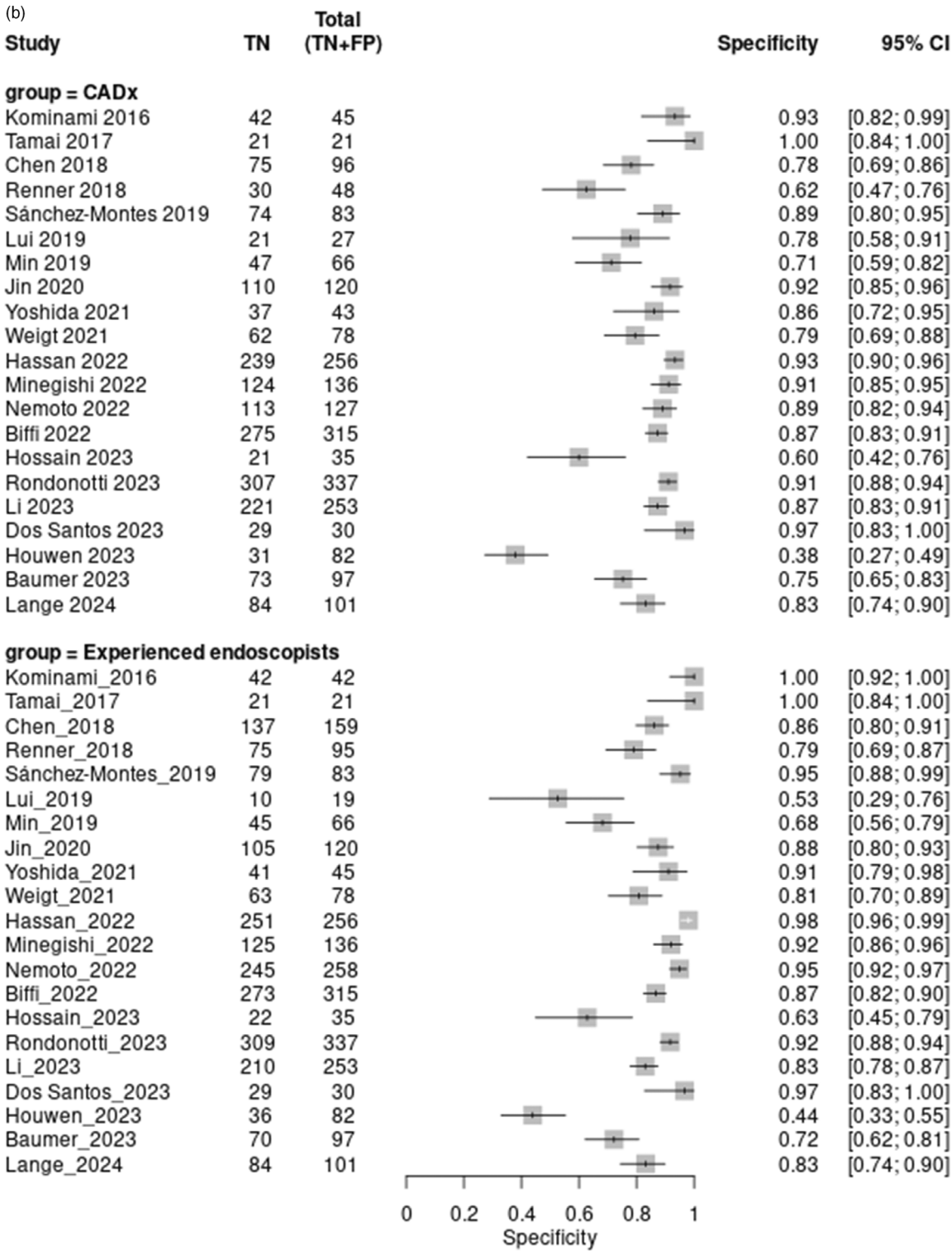

Figure 2 (*Continued*)

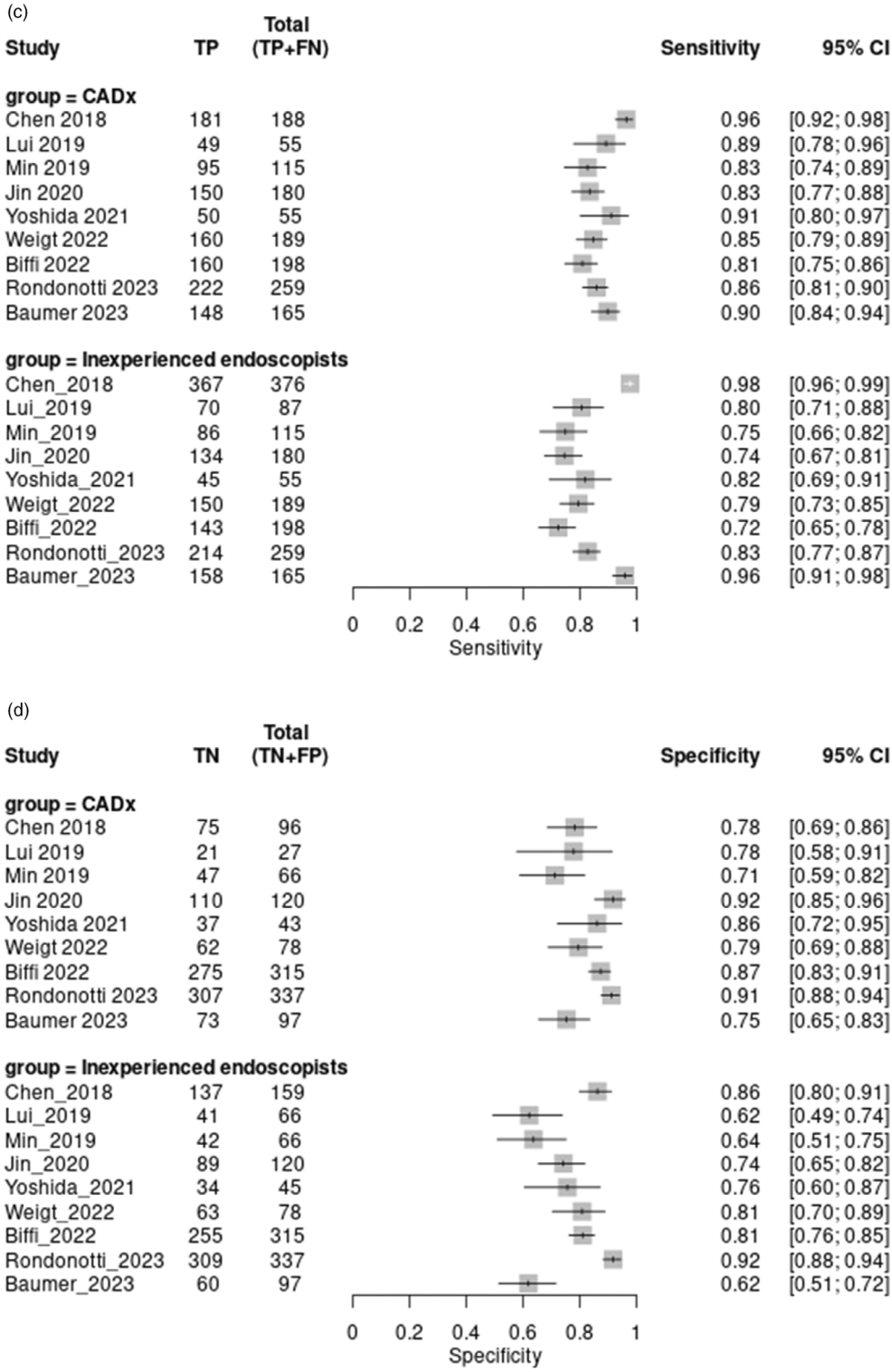

Figure 2 (*Continued*)

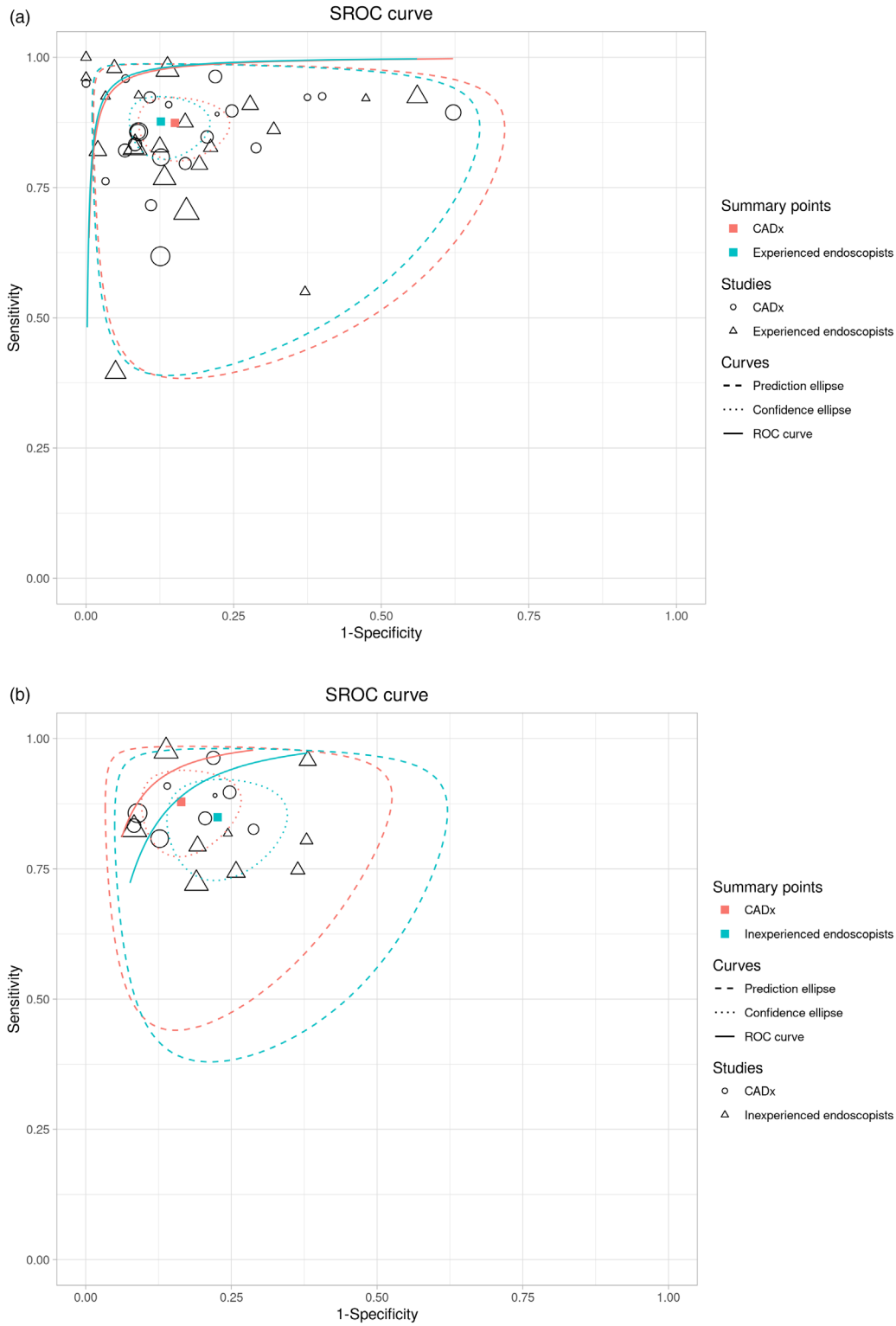

Figure 3 Summary receiver operating characteristic (SROC) plots: (a) comparison between computer-aided diagnosis (CADx) and experienced endoscopists; (b) comparison between CADx and inexperienced endoscopists.

conducted in other countries (n = 16) revealed no significant differences in sensitivity or specificity between the two groups (Fig. S1, Table S1). Therefore, neither the type of endoscopic imaging nor the country of origin appeared to affect the diagnostic accuracy of CADx.

Sensitivity analysis

The definition of an experienced endoscopist was explored as a potential factor influencing the diagnostic accuracy of optical diagnosis of colorectal polyps. We performed a sensitivity analysis, excluding one study without a clear definition of experienced endoscopists. Even with this exclusion, there were no significant differences between CADx and experienced endoscopist groups (Fig. S2, Table S2), consistent with the primary results.

DISCUSSION

THIS META-ANALYSIS REVIEWING the sensitivity and specificity of CADx for colorectal polyp diagnosis comparing those of experienced endoscopists showed no statistically significant differences. Unexpectedly, the comparison of CADx with inexperienced endoscopists also did not show statistically significant differences. Sensitivity analysis supported these results as well.

Despite advancements and excitement in AI diagnosis, the lack of significant differences may be attributed to several factors. First, the performance of human endoscopists, especially experienced endoscopists, is already high, leaving little room for improvement by current CADx systems. Second, the variability in CADx systems used across studies may have diluted potential benefits. The SROC plots indicate the inherent heterogeneity that can be explained by patient population and the experience levels of inexperienced endoscopists. This heterogeneity indicates that the potential utility of CADx in improving diagnostic accuracy may be enhanced in the setting of inexperienced endoscopists. Third, the complexity of polyp morphology and the subtlety of differentiating neoplastic from non-neoplastic lesions may pose challenges for CADx algorithms. A recent multicenter study also reported the suboptimal CADx-assisted algorithm in sensitivity and specificity in comparison between CADx-assisted and CADx-unassisted strategies.[37]

Although CADx did not demonstrate superiority over experienced or inexperienced endoscopists in this study, several strategies could enhance its diagnostic performance. The development of more robust deep-learning algorithms trained on diverse and extensive datasets including a large number of high-definition endoscopic images may address the current limitations in recognizing complex polyp morphologies, such as SSLs.

The AUC analysis revealed a significant difference between CADx and inexperienced endoscopists, despite the lack of significant differences in sensitivity and specificity when compared directly. This discrepancy may be attributed to the way AUC integrates performance across all possible thresholds, providing a more comprehensive measure of diagnostic accuracy than single-point metrics like sensitivity and specificity. Inexperienced endoscopists may exhibit inconsistent performance across various diagnostic thresholds, resulting in a lower AUC. In contrast, CADx systems are designed to maintain consistent performance due to their algorithmic nature. In clinical practice, sensitivity and specificity are the metrics of primary concern. The observed difference in AUC, while statistically significant, is relatively small and unlikely to be clinically meaningful.

The classification of SSLs varied among the included studies, which may have introduced heterogeneity in the pooled results. SSLs exhibit histological features distinct from conventional adenomas, including a serrated crypt pattern, and their classification as neoplastic or nonneoplastic is often inconsistent. Studies that categorized SSLs as neoplastic may have overestimated the sensitivity of CADx, as SSLs can be challenging to diagnose accurately even for experienced endoscopists. Conversely, studies that classified SSLs as nonneoplastic may have underestimated specificity, as SSLs with dysplastic features are clinically relevant. The reported sensitivity for CADx, among experienced and inexperienced endoscopists in this study, falls on the lower end of the competence standards recommended by the ASGE and ESGE for implementing strategies such as “leave in-situ” (>90%). This discrepancy may be explained by the different classification of SSLs among the included studies.

A prior systematic review of diminutive colorectal polyps reported by Bang *et al*. in 2021 showed high sensitivity (96%) and specificity (93%) with CADx.[5] These values were comparatively higher than the results of the present study. Nine of the 13 studies included in their review were different from those in the present study. Furthermore, no comparison with endoscopists was performed. Lui *et al*. reported a systematic review including six studies investigating the comparison between CADx and inexperienced endoscopists,[38] and they reported the superiority of CADx over inexperienced endoscopists. Although their results seem contradictory to those of the present study, three of the six studies used endocytoscopy, and one of the six studies was published in 2011, when deep learning was not developed at that time. The use of endocytoscopy on CADx may have superior performance compared to inexperienced

endoscopists. However, endocytoscopy generally cannot be used in routine clinical practice.

Two sophisticated systematic reviews were published in 2024 from the CADx analysis study group.[4,6] Both reviews included studies using endocytoscopy and focused on specific strategies such as "leave-in-situ" and "resect-and-discard." The former study including 10 studies investigated the effect of CADx on the diagnostic accuracy of rectosigmoid diminutive polyps to evaluate the feasibility of leaving polyps in situ.[4] Their sensitivity (87%) and specificity (89%) were similar to our results, and no incremental benefits or harms were observed with the use of CADx. The present study evaluated whole colorectal polyps other than rectosigmoid and excluded studies with endocytoscopy. The latter study, including 11 studies, compared CADx-assisted strategies with CADx-unassisted strategies to clarify the usefulness of CADx in the resect-and-discard strategy.[6] The CADx-assisted strategy did not confer any benefits or harms.

The implementation of CADx systems in clinical practice has potential economic implications that merit consideration. Although CADx systems may incur high initial costs, including hardware, software, and personnel training, they could lead to cost savings by reducing diagnostic errors and unnecessary procedures. For instance, a reduction in false-positive diagnoses could decrease unnecessary polypectomies, thereby lowering procedural risks and associated health-care costs. Mori *et al*. estimated the cost reduction by using CADx regarding diminutive rectosigmoid polyps.[39] However, current evidence regarding the cost-effectiveness of CADx systems is limited, and further economic analyses are required.

While CADx has not demonstrated significant benefits, AI-assisted detection systems (CADe) have been well received. A recent systematic review reported that CADe-assisted strategies elevate adenoma detection rates compared to CADe-unassisted strategies by 24%, with a 55% risk reduction of polyp miss rate.[40] However, there are both benefits and drawbacks. Nonneoplastic polyps were more frequently resected in the CADe group than in the non-CADe group. When the dissemination of CADe increases adenoma detection rates and polypectomy, it may also lead to more unnecessary polypectomies. This may increase medical costs and the burden on patients and health-care providers, and it may also shorten the surveillance intervals that guidelines recommend for close follow-up with colonoscopy in patients who underwent resection of three or more adenomatous polyps.[13] Without concomitant advancements in CADx to accurately characterize polyps, the benefits of CADe may be offset by these drawbacks. Therefore, integrating improved CADx systems with CADe could help balance the increased detection with accurate characterization, optimizing patient care.

The present study has several strengths. First, we compared CADx with inexperienced endoscopists, a comparison that has been scarcely reported in systematic review. Second, we excluded the studies including data from endocytoscopy, which is only used in limited medical facilities; magnifying endoscopy and/or IEE are generally used worldwide. Endocytoscopy was excluded from this analysis due to its limited availability and use in routine clinical practice. While it demonstrates high diagnostic accuracy, it is predominantly utilized in highly specialized centers, which may not reflect general clinical settings. Our study focused on widely accessible CADx systems to ensure broader applicability of our findings. Third, definition of "experienced endoscopist" was scrutinized, and data of experienced and inexperienced endoscopists were separately analyzed. Sensitivity analysis excluding studies without clear definitions of "experienced endoscopist" supported the main results.

Despite its strengths, our study also has several limitations that warrant discussion. First, the types of CADx systems were not standardized across studies. Different algorithms, training datasets, and image processing techniques can lead to variable performance. Second, the definition of "experienced endoscopist" varied among studies, potentially introducing bias, although our sensitivity analyses suggest the robustness of findings. Third, studies using real-time or still imaging were mixed, but subgroup analysis did not show significant differences. Real-time CADx diagnosis may be affected by variations in magnification and maneuverability determined by the endoscopist's technique and decisions. For example, when an endoscopist estimates a polyp as neoplastic or nonneoplastic during a real-time procedure, the endoscopist may modulate the distance between polyp and tip of endoscope and the magnification to fit their diagnosis. Although this maneuverability is not intentional, the bias is not completely excluded. Fourth, the types of IEEs varied across the studies, which could impact diagnostic accuracy, as different modalities may offer varying levels of detail and contrast. Fifth, reporting or publication bias may exist. Although the PRISMA-DTA statement does not mandate the assessment of reporting or publication bias, we recognize that such biases, if present, can undermine the validity of meta-analytic results.

The effectiveness of CADx systems in clinical practice depends heavily on the interaction between physicians and these AI tools. Reverberi *et al*. highlighted that both overreliance and underreliance on CADx systems can compromise diagnostic accuracy.[41] This underscores the

importance of appropriate trust, particularly for less experienced physicians, who tend to benefit most from CADx assistance. Van der Zander *et al*. demonstrated that confidence scores play a crucial role in guiding physicians' reliance on CADx predictions, helping them balance trust effectively.[42] In a recent randomized controlled trial by Djinbachian *et al*., it was reported that endoscopists disagreed with AI diagnosis in ~25% of cases under AI assistance. This disagreement often led to incorrect changes in diagnosis.[43] To maximize the potential of CADx systems, clear communication of confidence levels is essential.

In conclusion, CADx diagnosis is not superior to that of either experienced or inexperienced endoscopists in the current state. While the evolution and dissemination of CADe technology have increased polyp detection and resection rates, improvements in CADx are highly anticipated to enhance polyp characterization and reduce unnecessary polypectomies. Future research should focus on refining CADx algorithms, standardizing systems, and integrating CADx with CADe to maximize the benefits of AI in colonoscopy.

CONFLICT OF INTEREST

AUTHORS DECLARE NO conflict of interest for this article.

FUNDING INFORMATION

NONE.

DATA AVAILABILITY STATEMENT

STUDY MATERIALS WILL be shared upon reasonable request, after consultation and agreement of the authors.

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 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.
3 Kudo SE, Mori Y, Abdel-Aal UM *et al*. Artificial intelligence and computer-aided diagnosis for colonoscopy: Where do we stand now? *Transl Gastroenterol Hepatol* 2021; **6**: 64.
4 Hassan C, Misawa M, Rizkala T *et al*. Computer-aided diagnosis for leaving colorectal polyps in situ: A systematic review and meta-analysis. *Ann Intern Med* 2024; **177**: 919–28.
5 Bang CS, Lee JJ, Baik GH. Computer-aided diagnosis of diminutive colorectal polyps in endoscopic images: Systematic review and meta-analysis of diagnostic test accuracy. *J Med Internet Res* 2021; **23**: e29682.
6 Hassan C, Rizkala T, Mori Y *et al*. Computer-aided diagnosis for the resect-and-discard strategy for colorectal polyps: A systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2024; **9**: 1010–9.
7 Deeks J, Bossuyt P, Leeflang M, Takwoingi Y. *Cochrane handbook for systematic reviews of diagnostic test accuracy [Internet]*. Version 2.0 (updated July 2023). London: Cochrane; 2023. Available from: https://training.cochrane.org/handbook-diagnostic-test-accuracy/current
8 Salameh JP, Bossuyt PM, McGrath TA *et al*. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): Explanation, elaboration, and checklist. *BMJ* 2020; **370**: m2632.
9 Japan Gastroenterological Endoscopy Society. *Announcement of specialized qualifications in board certified fellow of the Japan Gastroenterological Endoscopy Society 2019 [Internet]*. Tokyo: Japan Gastrointestinal Endoscopy Society; 2019 [cited 2024 Sept 3]. Available from: https://www.jges.net/news/news-specialist/2019/09/27/14680
10 Mitsala A, Tsalikidis C, Pitiakoudis M, Simopoulos C, Tsaroucha AK. Artificial intelligence in colorectal cancer screening, diagnosis and treatment. A new era. *Curr Oncol* 2021; **28**: 1581–607.
11 Ferlitsch M, Moss A, Hassan C *et al*. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) clinical guideline. *Endoscopy* 2017; **49**: 270–97.
12 Kaltenbach T, Anderson JC, Burke CA *et al*. Endoscopic removal of colorectal lesions: Recommendations by the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2020; **115**: 435–64.
13 Tanaka S, Saitoh Y, Matsuda T *et al*. Evidence-based clinical practice guidelines for management of colorectal polyps. *J Gastroenterol* 2021; **56**: 323–35.
14 Whiting PF, Rutjes AW, Westwood ME *et al*. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36.
15 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; **58**: 982–90.
16 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.
17 Tamai N, Saito Y, Sakamoto T *et al*. Effectiveness of computer-aided diagnosis of colorectal lesions using novel software for magnifying narrow-band imaging: A pilot study. *Endosc Int Open* 2017; **5**: E690–4.
18 Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HHS, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018; **154**: 568–75.

19 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.
20 Sánchez-Montes C, Sánchez FJ, Bernal J *et al*. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy* 2019; **51**: 261–5.
21 Lui TKL, Wong KKY, Mak LLY, Ko MKL, Tsao SKK, Leung WK. Endoscopic prediction of deeply submucosal invasive carcinoma with use of artificial intelligence. *Endosc Int Open* 2019; **7**: E514–20.
22 Min M, Su S, He W, Bi Y, Ma Z, Liu Y. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. *Sci Rep* 2019; **9**: 2881.
23 Jin EH, Lee D, Bae JH *et al*. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. *Gastroenterology* 2020; **158**: 2169–79.e8.
24 Yoshida N, Inoue K, Tomita Y *et al*. An analysis about the function of a new artificial intelligence, CAD EYE with the lesion recognition and diagnosis for colorectal polyps in clinical practice. *Int J Color Dis* 2021; **36**: 2237–45.
25 Weigt J, Repici A, Antonelli G *et al*. Performance of a new integrated computer-assisted system (CADe/CADx) for detection and characterization of colorectal neoplasia. *Endoscopy* 2022; **54**: 180–4.
26 Hassan C, Balsamo G, Lorenzetti R, Zullo A, Antonelli G. Artificial intelligence allows leaving-in-situ colorectal polyps. *Clin Gastroenterol Hepatol* 2022; **20**: 2505–13.e4.
27 Minegishi Y, Kudo SE, Miyata Y *et al*. Comprehensive diagnostic performance of real-time characterization of colorectal lesions using an artificial intelligence-assisted system: A prospective study. *Gastroenterology* 2022; **163**: 323–5.e3.
28 Nemoto D, Guo Z, Peng B *et al*. Computer-aided diagnosis of serrated colorectal lesions using non-magnified white-light endoscopic images. *Int J Color Dis* 2022; **37**: 1875–84.
29 Biffi C, Salvagnini P, Dinh NN *et al*. A novel AI device for real-time optical characterization of colorectal polyps. *NPJ Digit Med* 2022; **5**: 84.
30 Hossain E, Abdelrahim M, Tanasescu A *et al*. Performance of a novel computer-aided diagnosis system in the characterization of colorectal polyps, and its role in meeting preservation and incorporation of valuable endoscopic innovations standards set by the American Society of Gastrointestinal Endoscopy. *DEN Open* 2023; **3**: e178.
31 Rondonotti E, Hassan C, Tamanini G *et al*. Artificial intelligence-assisted optical diagnosis for the resect-and-discard strategy in clinical practice: The artificial intelligence BLI characterization (ABC) study. *Endoscopy* 2023; **55**: 14–22.
32 Li JW, Wu CCH, Lee JWJ *et al*. Real-world validation of a computer-aided diagnosis system for prediction of polyp histology in colonoscopy: A prospective multicenter study. *Am J Gastroenterol* 2023; **118**: 1353–64.
33 Dos Santos CEO, Malaman D, Arciniegas Sanmartin ID, Leao ABS, Leao GS, Pereira-Lima JC. Performance of artificial intelligence in the characterization of colorectal lesions. *Saudi J Gastroenterol* 2023; **29**: 219–24.
34 Houwen B, Hazewinkel Y, Giotis I *et al*. Computer-aided diagnosis for optical diagnosis of diminutive colorectal polyps including sessile serrated lesions: A real-time comparison with screening endoscopists. *Endoscopy* 2023; **55**: 756–65.
35 Baumer S, Streicher K, Alqahtani SA *et al*. Accuracy of polyp characterization by artificial intelligence and endoscopists: A prospective, non-randomized study in a tertiary endoscopy center. *Endosc Int Open* 2023; **11**: E818–28.
36 De Lange G, Prouvost V, Rahmi G *et al*. Artificial intelligence for characterization of colorectal polyps: Prospective multicenter study. *Endosc Int Open* 2024; **12**: E413–8.
37 Rex DK, Bhavsar-Burke I, Buckles D *et al*. Artificial intelligence for real-time prediction of the histology of colorectal polyps by general endoscopists. *Ann Intern Med* 2024; **177**: 911–8.
38 Lui TKL, Guo C-G, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: A systematic review and meta-analysis. *Gastrointest Endosc* 2020; **92**: 11–22.e6.
39 Mori Y, Kudo SE, East JE *et al*. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: An add-on analysis of a clinical trial (with video). *Gastrointest Endosc* 2020; **92**: 905–11.e1.
40 Hassan C, Spadaccini M, Mori Y *et al*. Real-time computer-aided detection of colorectal neoplasia during colonoscopy: A systematic review and meta-analysis. *Ann Intern Med* 2023; **176**: 1209–20.
41 Reverberi C, Rigon T, Solari A, Hassan C, Cherubini P, Cherubini A. Experimental evidence of effective human-AI collaboration in medical decision-making. *Sci Rep* 2022; **12**: 14952.
42 van der Zander QEW, Roumans R, Kusters CHJ *et al*. Appropriate trust in artificial intelligence for the optical diagnosis of colorectal polyps: The role of human/artificial intelligence interaction. *Gastrointest Endosc* 2024; **100**: 1070–8.e10.
43 Djinbachian R, Haumesser C, Taghiakbari M *et al*. Autonomous artificial intelligence vs artificial intelligence-assisted human optical diagnosis of colorectal polyps: A randomized controlled trial. *Gastroenterology* 2024; **167**: 392–9.e2.

SUPPORTING INFORMATION

ADDITIONAL SUPPORTING INFORMATION may be found in the online version of this article at the publisher's web site.

Figure S1 Subgroup analysis of the computer-aided diagnosis (CADx) group: (a) sensitivity comparison between real-time and still imaging; (b) specificity comparison between real-time and still imaging; (c) sensitivity comparison between studies from Japan and other countries;

(d) specificity comparison between studies from Japan and other countries.

Figure S2 Sensitivity analysis including only studies with clear definition of "experienced endoscopist": (a) sensitivity comparison between computer-aided diagnosis (CADx) and experienced endoscopists; (b) specificity comparison between CADx and experienced endoscopists.

Table S1 Subgroup analysis of the computer-aided diagnosis (CADx) group.

Table S2 Sensitivity analysis including only studies with clear definitions of "experienced endoscopist."

Appendix S1 Preferred Reporting Items for Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) checklist.

Appendix S2 Search strategy.

Appendix S3 Modified Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.

Appendix S4 Risk of bias and applicability assessment using the modified Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.