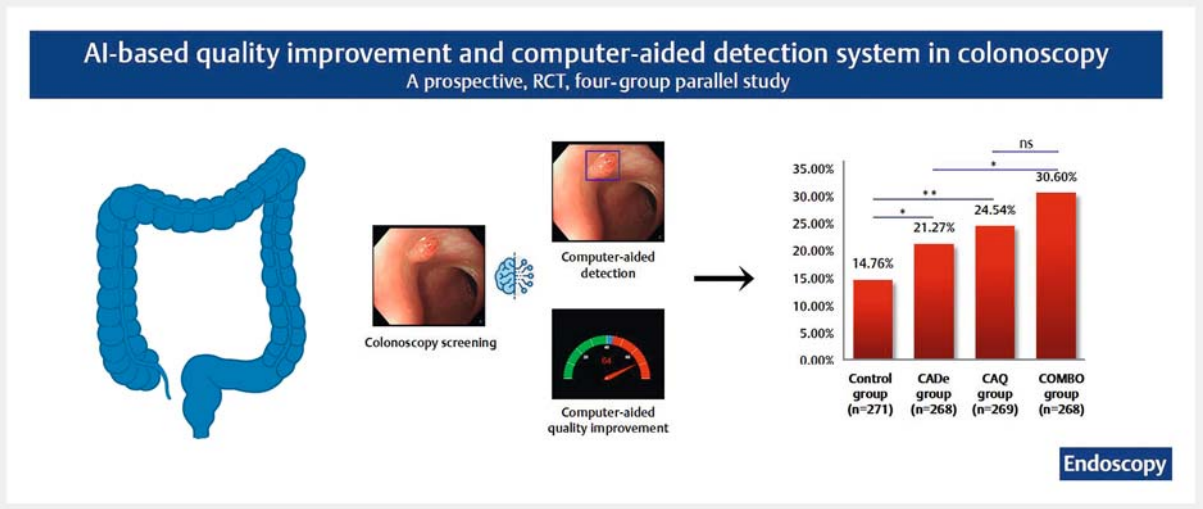
Effect of an artificial intelligence-based quality improvement system on efficacy of a computer-aided detection system in colonoscopy: a four-group parallel study

GRAPHICAL ABSTRACT



Authors

Liwen Yao[1,2,3,*], Lihui Zhang[1,2,3,*], Jun Liu[1,4], Wei Zhou[1,2,3], Chunping He[1,2,3], Jun Zhang[1,2,3], Lianlian Wu[1,2,3], Hongguang Wang[5], Youming Xu[1,2,3], Dexin Gong[1,2,3], Ming Xu[1,2,3], Xun Li[1,2,3], Yutong Bai[1,2,3], Rongrong Gong[1,2,3], Prateek Sharma[6], Honggang Yu[1,2,3]

Institutions

1 Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China
2 Key Laboratory of Hubei Province for Digestive System Disease, Renmin Hospital of Wuhan University, Wuhan, China
3 Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Renmin Hospital of Wuhan University, Wuhan, China
4 Department of Nursing, Renmin Hospital of Wuhan University, Wuhan, China
5 Department of Gastroenterology, Jilin Renmin Hospital, Jilin, China
6 University of Kansas School of Medicine and Veterans Affairs Medical Center, Kansas City, Missouri, United States

submitted 28.6.2021
accepted after revision 25.11.2021
published online 25.11.2021

Bibliography
Endoscopy 2022; 54: 757–768
DOI 10.1055/a-1706-6174
ISSN 0013-726X

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Supplementary material
Supplementary material is available under https://doi.org/10.1055/a-1706-6174

Corresponding author
Honggang Yu, MD, Department of Gastroenterology, Renmin Hospital of Wuhan University, 99 Zhangzhidong Road, Wuhan 430060, Hubei Province, China
yuhonggang@whu.edu.cn

* These authors contributed equally to this work.

ABSTRACT

Background Tandem colonoscopy studies have found that about one in five adenomas are missed at colonoscopy. It remains debatable whether the combination of a computer-aided polyp detection (CADe) system with a computer-aided quality improvement (CAQ) system for real-time monitoring of withdrawal speed results in additional benefits in adenoma detection or if the synergetic effect may be harmed due to excessive visual burden resulting from information overload. This study aimed to evaluate the interaction effect on improving the adenoma detection rate (ADR).

Methods This single-center, randomized, four-group, parallel, controlled study was performed at Renmin Hospital of Wuhan University. Between 1 July and 15 October 2020, 1076 patients were randomly allocated into four treatment groups: control 271, CADe 268, CAQ 269, and CADe plus CAQ (COMBO) 268. The primary outcome was ADR.

Results The ADR in the control, CADe, CAQ, and COMBO groups was 14.76 % (95 % confidence interval [CI] 10.54 to 18.98), 21.27 % (95 %CI 16.37 to 26.17), 24.54 % (95 %CI 19.39 to 29.68), and 30.60 % (95 %CI 25.08 to 36.11), respectively. The ADR was higher in the COMBO group compared with the CADe group (21.27 % vs. 30.6 %, $P = 0.024$, odds ratio [OR] 1.284, 95 %CI 1.033 to 1.596) but not compared with the CAQ group (24.54 % vs. 30.6 %, $P = 0.213$, OR 1.309, 95 %CI 0.857 to 2.000, respectively).

Conclusions CAQ significantly improved the efficacy of CADe in a four-group, parallel, controlled study. No significant difference in the ADR or polyp detection rate was found between CAQ and COMBO.

Introduction

Colorectal cancer (CRC) is the third leading cause of cancer-related death in the world [1]. The mortality rate of CRC could be effectively reduced through colonoscopy by the detection and resection of precancerous lesions [2]. However, a recent meta-analysis revealed that 22 % of colorectal adenomas were missed at screening colonoscopy and that these missed lesions were responsible for the majority of interval CRCs [3]. There are two main independent issues that cause the missed diagnosis of adenoma: 1) failure to recognize polyps (cognitive limitation); 2) existence of a blind spot (technical defect) [4]. The development of computer-aided techniques – computer-aided detection (CADe) and computer-aided quality improvement (CAQ) systems – has allowed the adenoma detection rate (ADR) to be improved; CADe was designed to augment cognitive performance and CAQ was developed to avoid technical defects [5–9]. Although these techniques have shown encouraging results in improving the ADR, existing evidence also revealed flaws in both techniques [7, 10].

Even when the lesions are within the visual field, they may be missed due to human cognitive limitations [11]. For example, polyps in the visual field may be missed due to being inconspicuous, only briefly visible, or appearing at the edge of the screen [12]. CADe systems, based on deep learning, could improve the ADR by displaying visual alerts that identify precancerous polyps on the endoscopy monitor in real time [5–7, 13]. However, despite the effectiveness of CADe, a previous randomized study reported the adenoma missing rate in CADe-assisted colonoscopy to be as high as 18 % [7]. Similarly, nonvisualization is a major cause of missed diagnosis, as lesions may remain hidden behind folds or debris during colonoscopy. Such nonvisualized lesions could be better exposed by meticulous mucosal inspection techniques, which require a steady and low withdrawal speed.

Rapid withdrawal is an important technical failure that leads to colonoscopy blind spots [14]. Calculating the similarity between consecutive frames, and thus monitoring the real-time withdrawal speed during colonoscopy, could assist in maintaining a low and uniform speed throughout the withdrawal process. A previous study reported that the ADR could be doubled by using a CAQ system to ensure steady withdrawal [9]. Nevertheless, visual fatigue and different human visual gaze patterns have indicated that missed diagnosis might happen even during a high quality withdrawal phase [10].

In order to increase the possible benefits from the available techniques, complementary techniques and combinations of modalities have been established in clinical practice [15, 16]. However, the interaction effect between CADe and CAQ remains unknown. While an overload of information may be caused by the combined intervention, it is worth exploring whether the interaction effect between the two techniques is synergetic or mutually exclusive. In this study, we conducted a four-group, parallel, controlled trial of endoscopists equipped with CADe, CAQ, or both, for the detection of adenomas. Based on the evidence for single function effects, we hypothesized that the combination intervention would further increase the ADR compared with the use of CAQ alone and CADe alone, respectively.

Methods

CADe and CAQ systems

We conducted a prospective, single-center, open-label, four-group, randomized parallel study at the endoscopy center in Renmin Hospital of Wuhan University, China.

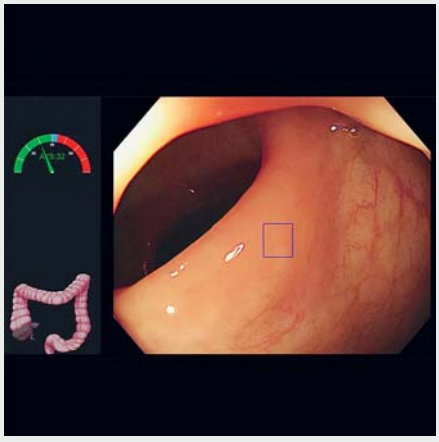
The development and validation of the real-time CADe model (Wuhan EndoAngel Medical Technology Company Co., Ltd., Wuhan, China) were performed using YOLO V3, the details of which are available in the online-only **Supplementary material**. The model achieved a comparable performance to that reported in a previous study, with a sensitivity of 97.31 % and a specificity of 90.75 % in video validation. The performance of the CADe system on the public dataset of the Computer Vision Center (CVC)-

Clinic database achieved a sensitivity of 96.57 %, which was higher than what was previously reported (88.24 %) [4].

The similarity between continuous colonoscopy frames was calculated in the CAQ system based on the perceptual hashing algorithm to obtain the withdrawal speed value, which was presented in the form of a speed dashboard. The system has been described previously [9] and in the **Supplementary material**. Following the previous study, a value of <40 was considered a safe withdrawal speed, 40–44 was defined as the alarm withdrawal speed (i. e. to take note that the speed was slightly too fast), and >44 was considered a dangerous withdrawal speed (i. e. too fast). When the speed was >44, the pointer of the speed dashboard enters the red zone. We have followed this setting in the current study, as shown in ▶ **Video 1**. Negative withdrawal time was defined as the withdrawal time excluding the time taken for biopsy and polypectomy.

Patients

Consecutive patients older than 18 years who attended the endoscopy center in Renmin Hospital of Wuhan University, China, between 1 July and 15 October 2020 were included in the study. Patients with known contraindications to biopsy, bowel obstruction or perforation, or those who were pregnant or lactating, suffering from polyposis syndromes or who had a history of inflammatory bowel disease, CRC, or colorectal surgery were excluded. In addition, any patient whose cecum was not reached, or who had suspicion for polyposis syndromes, inflammatory bowel disease, intestinal tuberculosis, or CRC was also excluded. The study was approved by the Ethics Committee of Renmin Hospital of Wuhan University. Informed consent was obtained from all patients.



Video 1 Interventions in each group. Control group: the system shows the original colonoscopy video. CADe group: the system presents the detected polyp location with a hollow blue alert box directly on a high definition monitor. CAQ: the system presents the real-time withdrawal speed with a dashboard. COMBO: the system presents both alert box and withdrawal speed. CADe, computer-aided detection; CAQ, computer-aided quality improvement; COMBO, CADe plus CAQ.
Online content viewable at:
https://doi.org/10.1055/a-1706-6174

Randomization

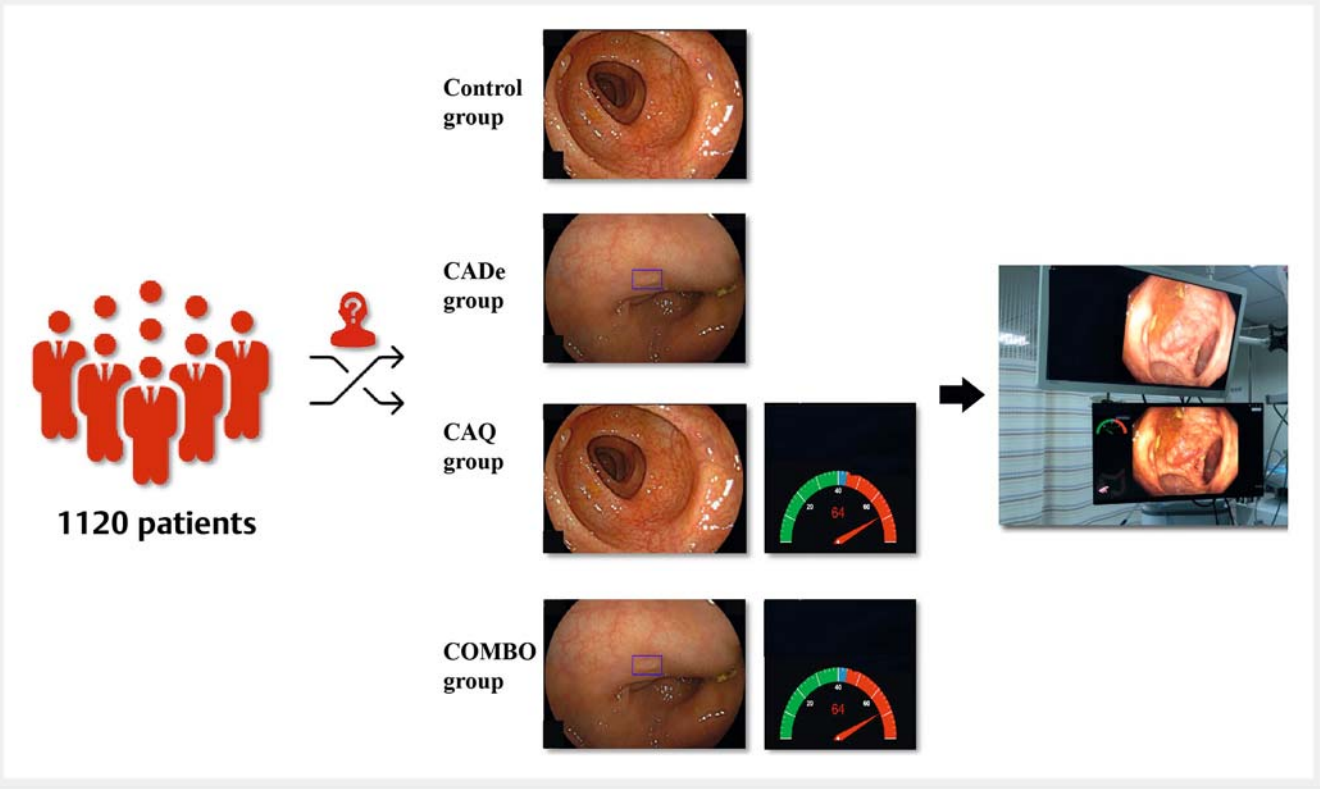
In the parallel controlled study, we evaluated four groups: CAQ, CADe, nonintervention (control), both interventions (COMBO). All patients were randomly assigned to receive one of the four regimens (▶ **Fig. 1**). An electronic digital capture system was used to generate a random number for each patient before the procedure, and randomization was done in blocks of 16. Before inserting the colonoscope, the operators were informed about the allocation of the patient by the researchers. Patients and outcome assessors were blinded to intervention allocation. Details of the colonoscopy procedure and intervention are provided in the **Supplementary material**.

Outcome

The primary outcome of the study was the ADR. Secondary outcomes included the polyp detection rate (PDR), ADR/PDR for adenomas with different sizes (diminutive ≤5 mm, small >5 to <10 mm, large ≥10 mm) and locations (cecum, ascending colon, transverse colon, descending colon, sigmoid colon, and rectum), advanced ADR (≥10 mm diameter, high grade dysplasia, villous histology), benign lesion detection rate (including hyperplastic and inflammatory polyps), sessile serrated lesion (SSL) detection rate, proportion of over-speed frames, as well as false-positive and false-negative rates of CADe. ADR was defined as the proportion of individuals who had at least one histologically confirmed adenoma. SSLs were not included in the ADR calculation [17]. PDR was defined as the proportion of individuals who had at least one polyp detected during a complete screening colonoscopy, and was calculated based on the pathology report (including adenoma, SSL, inflammatory and hyperplastic polyps).

Statistical analysis

Details of the sample size calculation are available in the **Supplementary material**. The baseline characteristics among the study groups were compared using the chi-squared test for categorical variables and one-way analysis of variance test for continuous variables. A logistic regression was applied to assess the effect of the intervention on the lesion detection rate. The analysis of the detection rate was at a patient-based level, such that one positive outcome was registered regardless of the number of lesions. To address possible confounding effects, we built covariate-adjusted regression models by adding group differences into the models as covariates, including the indication for colonoscopy, body mass index, age, bowel cleanliness, sex, status as inpatient or outpatient, family history, hypertension, diabetes, endoscope manufacturers, and use of tobacco and alcohol. A Poisson regression was performed to assess the effect of the intervention on the mean number of adenomas and polyps per patient. The withdrawal time (with or without operation) and proportion of over-speed frames were analyzed using the Kruskal–Wallis test among the four groups and the Mann–Whitney *U* test between each set of two groups. A two-sided *P* value of <0.05 was considered statistically significant.



▶ **Fig. 1** Study design and group intervention. CADe, computer-aided detection; CAQ, computer-aided quality improvement; COMBO, CADe plus CAQ.

The analyses were done using SPSS software version 20 (IBM Corp., Armonk, New York, USA). We performed subgroup analyses according to age (≤49 years and ≥50 years), endoscope manufacturer, and bowel preparation (adequate, defined as a Boston Bowel Preparation Scale [BBPS] score with all segments ≥2; inadequate, defined as a BBPS score with any segments <2) to investigate the universality of the intervention.

Role of the funding source

The funding source had no role in the study design, data collection, data analysis, data interpretation, or report preparation. The corresponding author had full access to all the data in the study and had the final responsibility for the decision to submit the work for publication.

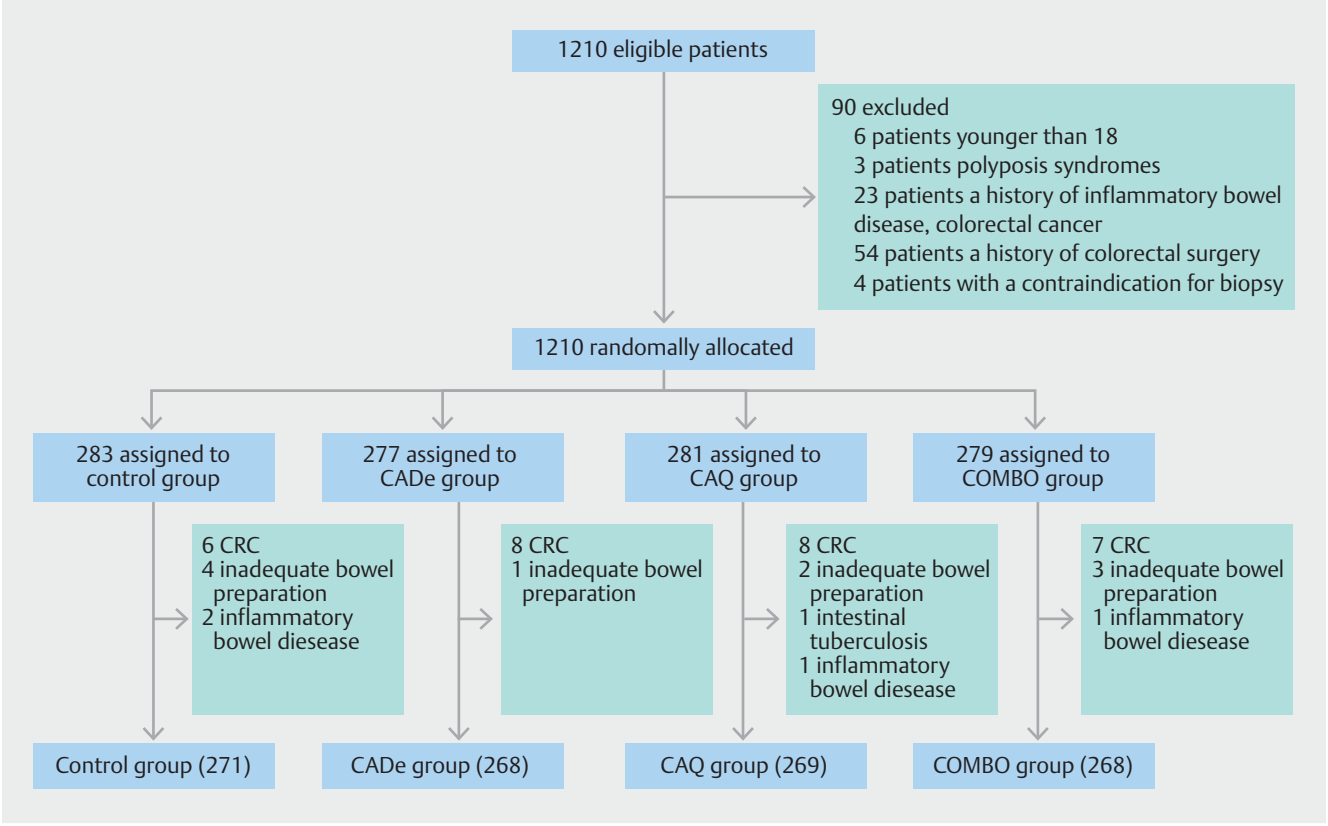
Result

Patient enrollment and baseline data

A total of 1210 consecutive patients were evaluated and 1120 eligible individuals were enrolled in the study. We excluded 44 patients during colonoscopy according to the exclusion criteria: 29 patients due to CRC, 10 due to poor bowel preparation, and 5 due to high suspicion for inflammatory bowel disease or intestinal tuberculosis. Finally, 1076 eligible patients were analyzed, with 271 patients in the control group, 268 in the CADe group, 269 in the CAQ group, and 268 in the COMBO group (▶ **Fig. 2**). Baseline characteristics of patients are presented in ▶ **Table 1**. There was no difference between the four arms in the clinical indication, which was overall screening in 89.13% (959/1076), post-polypectomy surveillance in 9.48% (102/1076), and gastrointestinal symptoms indicating the diagnosis in 1.39% (15/1076) of the cases. There was no statistically significant difference between the four groups with regard to the demographic information or other adenoma risk factors. The primary and secondary outcome analyses are presented in ▶ **Table 2**.

ADR comparison

For adenoma detection, the ADR in the control, CADe, CAQ, and COMBO groups was 14.76% (95% confidence interval [CI] 10.54 to 18.98), 21.27% (95%CI 16.37 to 26.17), 24.54% (95% CI 19.39 to 29.68), and 30.60% (95%CI 25.08 to 36.11), respectively. Compared with the control group, the ADR was significantly higher in the CADe group ($P=0.045$, odds ratio [OR] 1.621, 95%CI 1.011 to 2.601) and CAQ group ($P=0.004$, OR 1.406, 95%CI 1.116 to 1.772). The ADR was significantly higher in the COMBO group than in the CADe group ($P=0.024$, OR 1.284, 95%CI 1.033 to 1.596) (▶ **Fig. 3**). No statistically significant difference was found in the ADR between the CAQ and COMBO groups ($P=0.213$; OR 1.309, 95%CI 0.857 to 2.00), or between the CAQ and CADe groups ($P=0.519$, OR 1.178, 95% CI 0.715 to 1.941). Moreover, the ADR for large adenomas (≥10 mm) was significantly higher in the COMBO group com-



▶ **Fig. 2** Flow chart of patient inclusion. CRC, colorectal cancer.

pared with the CADe group ($P=0.012$, OR 2.669, 95%CI 1.241 to 5.740) (▶ **Table 2**).

PDR comparison

For polyp detection, the PDR in the control, CADe, CAQ, and COMBO groups was 41.70% (95%CI 35.83 to 47.57), 55.60% (95%CI 49.65 to 61.55), 53.53% (95%CI 47.57 to 59.49), and 64.18% (95%CI 58.44 to 69.92), respectively. Compared with the control group, a significantly higher PDR was observed in the CADe group ($P=0.001$, OR 1.847, 95%CI 1.27 to 2.687) and CAQ group ($P=0.003$, OR 1.336, 95%CI 1.103 to 1.618). Compared with the CAQ group, a higher PDR was observed in the COMBO group ($P=0.018$, OR 1.608, 95%CI 1.084 to 2.384). No statistically significant difference was found in the PDR between the CADe and COMBO groups ($P=0.051$, OR 1.215, 95%CI 0.999 to 1.477), or between the CAQ and CADe groups ($P=0.797$, OR 0.952, 95%CI 0.653 to 1.388) (▶ **Fig. 3,** ▶ **Table 2**).

The benign lesion detection rate in the control, CADe, CAQ, and COMBO groups was 34.69% (95%CI 28.98 to 40.39), 47.01% (95%CI 41.00 to 53.03), 43.12% (95%CI 37.17 to 49.08), and 52.24% (95%CI 46.22 to 58.26), respectively. Compared with the control group, a significantly higher benign lesion detection rate was observed in the CADe group ($P=0.005$, OR 1.726, 95%CI 1.184 to 2.515) and CAQ group ($P=0.035$, OR 1.23, 95%CI 1.015 to 1.49). Compared with the CAQ group, a higher benign lesion detection rate was observed in the COMBO group ($P=0.036$, OR 1.494, 95%CI 1.027 to 2.174). No statistically significant difference was found in the benign lesion detection rate between the CADe and COMBO groups ($P=0.199$, OR 1.128, 95%CI 0.939 to 1.356), or between the CAQ and CADe groups ($P=0.347$, OR 0.836, 95%CI 0.576 to 1.214) (▶ **Table 2**).

Proportion of over-speed frames and withdrawal time comparison

The proportion of over-speed frames was 27.19%, 25.35%, 21.67%, and 20.78% in the control, CADe, CAQ, and COMBO groups, respectively. A significantly lower proportion of over-speed frames was observed in the CAQ group and COMBO group compared with the control group and CADe group, respectively. The groups with CAQ had a significantly lower proportion of over-speed frames than those without CAQ (21.24% vs. 26.26%, respectively; $P<0.001$).

The total withdrawal time was 9.71 minutes, 10.52 minutes, 10.14 minutes, and 10.17 minutes in the control, CADe, CAQ, and COMBO groups, respectively. The withdrawal time without operation was 9.36 minutes, 9.94 minutes, 9.46 minutes, and 9.46 minutes, respectively. No statistically significant difference in the withdrawal time without operation was found among the four groups (▶ **Table 3**).

▶ **Table 1** Baseline characteristics of patients.

	Control group (n = 271)	CADe group (n = 268)	CAQ group (n = 269)	COMBO group (n = 268)	*P* value
Age, mean (SD), years	50.85 (13.56)	50.69 (13.15)	49.91 (13.64)	51.17 (12.79)	0.730
Sex, n (%)					
▪ Male	114 (42.07)	121 (45.15)	127 (47.21)	130 (48.51)	0.462
▪ Female	157 (57.93)	147 (54.85)	142 (52.79)	138 (51.49)	
Body mass index, n (%)					0.579
▪ < 25 kg/m²	211 (77.86)	208 (77.61)	207 (76.95)	210 (78.36)	
▪ ≥ 25– < 30 kg/m²	53 (19.56)	58 (21.64)	55 (20.45)	48 (17.91)	
▪ > 30 kg/m²	7 (2.58)	2 (0.75)	7 (2.60)	10 (3.73)	
Indication for colonoscopy, n (%)					0.474
▪ Screening	241 (88.93)	238 (88.81)	239 (88.85)	241 (89.93)	
▪ Diagnostic	3 (1.11)	2 (0.75)	3 (1.12)	7 (2.61)	
▪ Surveillance	27 (9.96)	28 (10.45)	27 (10.04)	20 (7.46)	
Recruitment, n (%)					0.744
▪ Inpatient	48 (17.71)	45 (16.79)	52 (19.33)	54 (20.15)	
▪ Outpatient	223 (82.29)	223 (83.21)	217 (80.67)	214 (79.85)	
Smoke, n (%)	41 (15.13)	51 (19.03)	49 (18.22)	51 (19.03)	0.593
Alcohol, n (%)	36 (13.28)	44 (16.42)	51 (18.96)	46 (17.16)	0.347
High blood pressure, n (%)	41 (15.13)	37 (13.81)	41 (15.24)	40 (14.93)	0.955
Diabetes, n (%)	14 (5.17)	11 (4.10)	18 (6.69)	16 (5.97)	0.579
Family history, n (%)	13 (4.80)	15 (5.60)	14 (5.20)	13 (4.85)	0.978
Bowel preparation, n (%)					0.969
▪ BBPS ≥ 2 in all segments	231 (85.24)	227 (84.70)	227 (84.39)	230 (85.82)	
▪ BBPS < 2 in any segment	40 (14.76)	41 (15.30)	42 (15.61)	38 (14.18)	
Endoscope manufacturer, n (%)					
▪ Fujifilm	96 (35.42)	84 (31.34)	86 (31.97)	101 (37.69)	0.363
▪ Olympus	175 (64.58)	184 (68.66)	183 (68.03)	167 (62.31)	

CADe, computer-aided detection; CAQ, computer-aided quality improvement; COMBO, CADe plus CAQ; BBPS, Boston Bowel Preparation Scale.

False-positive and false-negative rates of the CADe system

A total of 54 false positives were observed in the CADe and COMBO groups. Biopsies were taken and confirmed from the two cases of false-positive inflammations. All polyps in the CADe and COMBO groups were detected by the polyp detection system (**Table 1 s**).

Subgroup analysis

The results were stratified according to the baseline characteristics to determine the presence or absence of an intervention modification. Compared with the control group, the ADR and PDR in both the CADe and CAQ groups significantly improved in the subgroups with Olympus endoscopes, age ≥ 50 years, and adequate bowel preparation. Regarding the comparison between the CADe and COMBO groups, the ADR was significantly higher in the subgroup with Fujifilm endoscopes, age ≤ 49 years, and adequate bowel preparation (**Table 2 s**). No adverse events were reported during the clinical trial.

Discussion

In this randomized, open-label, four-group, parallel, controlled study on the efficacy of the CAQ and CADe systems for adenoma detection, the ADR was significantly improved from 14.76 % (95 %CI 10.54 to 18.98) to 21.27 % (95 %CI 16.37 to 26.17) with the CADe system and to 24.54 % (95 %CI 19.39 to 29.68) with the CAQ system. Based on this enhancement, CADe equipped

▶ **Table 2** Primary and secondary outcome analyses: comparisons between control group and single intervention group, and between single intervention group and combination group.

	n (%) [95 %CI]				*P* value (OR) [95 %CI]				
	Control group (n = 271)	**CADe group (n = 268)**	**CAQ group (n = 269)**	**COMBO group (n = 268)**	**Control vs. CADe**	**Control vs. CAQ**	**CADe vs. CAQ**	**COMBO vs. CADe**	**COMBO vs. CAQ**
ADR[1]	40 (14.76) [10.54 to 18.98]	57 (21.27) [16.37 to 26.17]	66 (24.54) [19.39 to 29.68]	82 (30.60) [25.08 to 36.11]	0.045 (1.621) [1.011 to 2.601]	0.004 (1.406) [1.116 to 1.772]	0.519 (1.178) [0.715 to 1.941]	0.024 (1.284) [1.033 to 1.596]	0.213 (1.309) [0.857 to 2.00]
PDR[1]	113 (41.70) [35.83 to 47.57]	149 (55.60) [49.65 to 61.55]	144 (53.53) [47.57 to 59.49]	172 (64.18) [58.44 to 69.92]	0.001 (1.847) [1.27 to 2.687]	0.003 (1.336) [1.103 to 1.618]	0.797 (0.952) [0.653 to 1.388]	0.051 (1.215) [0.999 to 1.477]	0.018 (1.608) [1.084 to 2.384]
Nonprecancerous PDR[1]	94 (34.69) [28.98 to 40.39]	126 (47.01), [41.00 to 53.03]	116 (43.12) [37.17 to 49.08]	140 (52.24) [46.22 to 58.26]	0.005 (1.726) [1.184 to 2.515]	0.035 (1.23) [1.015 to 1.49]	0.347 (0.836) [0.576 to 1.214]	0.199 (1.128) [0.939 to 1.356]	0.036 (1.494) [1.027 to 2.174]
Adenoma size[1]									
▪ Diminutive (≤ 5 mm)	39 (14.39) [10.21 to 18.57]	54 (20.15) [15.35 to 24.95]	61 (22.68) [17.67 to 27.68]	72 (26.87) [21.56 to 32.17]	0.073 (1.547) [0.959 to 2.495]	0.013 (1.347) [1.066 to 1.702]	0.581 (1.152) [0.697 to 1.906]	0.122 (1.192) [0.954 to 1.489]	0.51 (1.158) [0.749 to 1.789]
▪ Small (> 5 to < 10 mm)	4 (1.48), [0.04 to 2.91]	13 (4.85) [2.28 to 7.42]	11 (4.09) [1.72 to 6.46]	20 (7.46) [4.32 to 10.61]	0.064 (3.01) [0.939 to 9.653]	0.079 (1.706) [0.94 to 3.098]	0.693 (0.832) [0.333 to 2.08]	0.135 (1.337) [0.913 to 1.957]	0.078 (2.039) [0.924 to 4.502]
▪ Large (≥ 10 mm)	2 (0.74) [0 to 1.76]	2 (0.75) [0 to 1.78]	7 (2.60) [0.70 to 4.50]	13 (4.85) [2.28 to 7.42]	0.671 (1.618) [0.176 to 14.882]	0.16 (1.813) [0.791 to 4.158]	0.251 (2.698) [0.495 to 14.69]	0.012 (2.669) [1.241 to 5.740]	0.146 (2.098) [0.772 to 5.701]
Adenoma pathology[1]									
▪ Advanced adenoma	5 (1.85) [0.24 to 3.45]	5 (1.87) [0.25 to 3.49]	11 (4.09) [1.72 to 6.46]	14 (5.22) [2.56 to 7.89]	0.664 (1.346) [0.352 to 5.151]	0.175 (1.478) [0.84 to 2.6]	0.239 (2.15) [0.602 to 7.682]	0.046 (1.737) [1.009 to 2.991]	0.749 (0.861) [0.345 to 2.152]
▪ Other adenoma	37 (13.65) [9.57 to 17.74]	53 (19.78) [15.01 to 24.54]	55 (20.45) [15.63 to 25.27]	68 (25.37) [20.16 to 30.58]	0.058 (1.599) [0.984 to 2.597]	0.036 (1.294) [1.017 to 1.645]	0.966 (1.011) [0.611 to 1.671]	0.207 (1.154) [0.924 to 1.441]	0.321 (1.252) [0.804 to 1.951]
SSL	1 (0.37) [0 to 1.09]	1 (0.37) [0 to 1.1]	1 (0.37) [0 to 1.1]	4 (1.49) [0.04 to 2.94]	0.963 (0.925) [0.035 to 24.686]	0.996 (0.018) [N/A]		0.290 (1.838) [0.595 to 5.674]	0.287 (3.734) [0.33 to 42.247]
Adenoma location[1]									
▪ Cecum	1 (0.37) [0 to 1.09]	0 (0) [0 to 0]	4 (1.49) [0.04 to 2.93]	1 (0.37) [0 to 1.1]	0.986 (0) [0 to .]	0.381 (1.561) [0.576 to 4.228]	0.993, 15093126.815 [N/A]	> 0.99 (6.752) [N/A]	0.226 (0.159) [0.008 to 3.132]
▪ Ascending colon	7 (2.58) [0.69 to 4.47]	16 (5.97) [3.13 to 8.81]	15 (5.58) [2.83 to 8.32]	20 (7.46) [4.32 to 10.61]	0.034 (2.949) [1.082 to 8.032]	0.039 (1.703) [1.028 to 2.821]	0.384 (0.694) [0.306 to 1.578]	0.432 (0.154) [0.808 to 1.647]	0.39 (1.379) [0.663 to 2.865]
▪ Transverse colon	9 (3.32) [1.19 to 5.45]	15 (5.6) [2.84 to 8.35]	25 (9.29) [5.82 to 12.76]	23 (8.58) [5.23 to 11.94]	0.244 (1.665) [0.706 to 3.923]	0.007 (1.737) [1.165 to 2.591]	0.179 (1.643) [0.796 to 3.388]	0.226 (1.241) [0.875 to 1.761]	0.681 (0.875) [0.465 to 1.65]
▪ Descending colon	5 (1.85) [0.24 to 3.45]	7 (2.61) [0.7 to 4.52]	8 (2.97) [0.94 to 5]	12 (4.48) [2 to 6.95]	0.565 (1.433) [0.421 to 4.876]	0.536 (1.206) [0.667 to 2.18]	0.727 (1.232) [0.381 to 3.985]	0.304 (1.300) [0.789 to 2.141]	0.228 (1.839) [0.683 to 4.954]

▶ **Table 2** (Continuation)

	n (%) [95 %CI]				*P* value (OR) [95 %CI]				
	Control group (n = 271)	**CADe group (n = 268)**	**CAQ group (n = 269)**	**COMBO group (n = 268)**	**Control vs. CADe**	**Control vs. CAQ**	**CADe vs. CAQ**	**COMBO vs. CADe**	**COMBO vs. CAQ**
▪ Sigmoid colon	15 (5.54) [2.81 to 8.26]	20 (7.46) [4.32 to 10.61]	18 (6.69) [3.71 to 9.68]	30 (11.19) [7.42 to 14.97]	0.413 (1.345) [0.662 to 2.736]	0.827 (1.042) [0.723 to 1.501]	0.615 (0.832) [0.405 to 1.707]	0.314 (1.173) [0.860 to 1.600]	0.134 (1.643) [0.858 to 3.145]
▪ Rectum	5 (1.85) [0.24 to 3.45]	5 (1.87) [0.25 to 3.49]	9 (3.35) [1.2 to 5.49]	9 (3.36) [1.2 to 5.52]	0.84 (0.865) [0.21 to 3.556]	0.261 (1.396) [0.78 to 2.5]	0.405 (1.822) [0.444 to 7.469]	0.272, 1.412 (0.763 to 2.616)	0.788 (0.868) [0.309 to 2.438]
Polyp size[1]									
▪ Diminutive (≤ 5 mm)	108 (39.85) [34.02 to 45.68]	142 (52.99) [47.01 to 58.96]	138 (51.3) [45.33 to 57.27]	163 (60.82) [54.98 to 66.67]	0.002 (1.789) [1.23 to 2.604]	0.003 (1.336) [1.103 to 1.617	0.966 (1.025) [0.331 to 3.174]	0.070 (1.192) [0.986 to 1.441]	0.051 (1.464) [0.998 to 2.146]
▪ Small (> 5 to < 10 mm)	9 (3.32) [1.19 to 5.45]	17 (6.34) [3.43 to 9.26]	16 (5.95) [3.12 to 8.77]	26 (9.70) [6.16 to 13.25]	0.185 (1.775) [0.76 to 4.15]	0.168 (1.36) [0.879 to 2.104]	0.816 (0.913) [0.424 to 1.965]	0.093 (1.334) [0.953 to 1.867]	0.097 (1.811) [0.899 to 3.648]
▪ Large (≥ 10 mm)	3 (1.11) [0 to 2.35]	6 (2.24) [0.47 to 4.01]	10 (3.72) [1.46 to 5.98]	16 (5.97) [3.13 to 8.81]	0.312 (2.082) [0.503 to 8.619]	0.073 (1.856) [0.944 to 3.649]	0.372 (1.549) [0.593 to 4.046]	0.038 (1.686) [1.029 to 2.762]	0.172 (1.822) [0.77 to 4.311]
Polyp location[1]									
▪ Cecum	6 (2.21) [0.46 to 3.97]	5 (1.87) [0.25 to 3.49]	5 (1.86) [0.24 to 3.47]	3 (1.12) [0 to 2.38]	0.448 (0.598) [0.158 to 2.257]	0.457 (0.78) [0.405 to 1.501]	0.668 (0.741) [0.188 to 2.918]	0.685 (0.833) [0.346 to 2.009]	0.337 (0.425) [0.074 to 2.443]
▪ Ascending colon	22 (8.12) [4.87 to 11.37]	34 (12.69) [8.7 to 16.67]	37 (13.75) [9.64 to 17.87]	37 (13.81) [9.68 to 17.94]	0.074 (1.729) [0.948 to 3.154]	0.024 (1.404) [1.045 to 1.887]	0.749 (1.096) [0.625 to 1.922]	0.613 (1.070) [0.823 to 1.392]	0.996 (0.999) [0.594 to 1.679]
▪ Transverse colon	35 (12.92) [8.92 to 16.91]	45 (16.79) [12.32 to 21.27]	51 (18.96) [14.27 to 23.64]	51 (19.03) [14.33 to 23.73]	0.203 (1.398) [0.835 to 2.339]	0.064 (1.267) [0.987 to 1.627]	0.706 (1.105) [0.658 to 1.856]	0.477 (1.090) [0.860 to 1.381]	0.921 (1.024) [0.644 to 1.628]
▪ Descending colon	21 (7.75) [4.57 to 10.93]	27 (10.07) [6.47 to 13.68]	20 (7.43) [4.3 to 10.57]	30 (11.19) [7.42 to 14.97]	0.372 (1.319) [0.718 to 2.424]	0.778 (0.954) [0.688 to 1.323]	0.135 (0.603) [0.311 to 1.171]	0.767 (1.044) [0.785 to 1.390]	0.146 (1.586) [0.852 to 2.951]
▪ Sigmoid colon	48 (17.71) [13.17 to 22.26]	72 (26.87) [21.56 to 32.17]	59 (21.93) [16.99 to 26.88]	89 (33.21) [27.57 to 38.85]	0.01 (1.763) [1.142 to 2.721]	0.223 (1.15) [0.918 to 1.441]	0.19 (0.722) [0.443 to 1.175]	0.144 (1.156) [0.952 to 1.403]	0.007 (1.748) [1.163 to 2.629]
▪ Rectum	43 (15.87), [11.52 to 20.22]	50 (18.66) [13.99 to 23.32]	54 (20.07) [15.29 to 24.86]	66 (24.63) [19.47 to 29.79]	0.464 (1.191) [0.746 to 1.9]	0.235 (1.153) [0.912 to 1.458]	0.377 (1.255) [0.758 to 2.077]	0.163 (1.169) [0.939 to 1.455]	0.372 (1.221) [0.788 to 1.893]
Adenomas detected per colonoscopy, mean (SD)	0.166 (0.411)	0.24 (0.485)	0.279 (0.526)	0.37 (0.65)	0.061 (−0.073) [−0.149 to 0.003][2]	0.006 (−0.113) [−0.192 to −0.033][2]	0.384 (0.043) [−0.122 to 0.047]	0.007 (−0.134) [−0.232 to −0.037][2]	0.055 (−0.098) [−0.198 to 0.002][2]

▲ **Table 2** (Continuation)

	n (%) [95 %CI]				*P* value (OR) [95 %CI]				
	Control group (n = 271)	CADe group (n = 268)	CAQ group (n = 269)	COMBO group (n = 268)	Control vs. CADe	Control vs. CAQ	CADe vs. CAQ	COMBO vs. CADe	COMBO vs. CAQ
Polyps detected per colonoscopy, mean (SD)	1.23 (2.117)	1.57 (2.328)	1.49 (2.319)	2.06 (3.017)	0.075 (−0.342) [−0.719 to 0.034][2,2]	0.183 (−0.255) [−0.63 to 0.121][2]	0.528 (0.208) [−0.277 to 0.539]	0.039 (−0.481) [−0.939 to −0.024][2]	0.015 (−0.569) [−1.025 to −0.113][2]

CI, confidence interval; OR, odds ratio; CADe, computer-aided detection; CAQ, computer-aided quality improvement; COMBO, CADe plus CAQ; ADR, adenoma detection rate; PDR, polyp detection rate; SSL, sessile serrated lesion; N/A, not applicable due to insufficient number of cases.
[1] Detection rate according to intervention arm, as well as per patient distribution.
[2] These data are shown as *P* value, mean range [95 %CI].

▶ **Fig. 3** Detection rates in each group. **a** Adenoma detection rate. **b** Polyp detection rate.

with CAQ yielded an additional 9.33 % increase in the ADR without prolonging the withdrawal time. As a result, the combination intervention was a reliable way to further increase the ADR.

CRC is an important public health issue, with gradually increasing incidence and mortality rates seen in young adults in recent years [18]. The ADR is known to be inversely associated with the risk of interval CRC [19]. As the ADR can range from 7.4 % to 52.5 % among endoscopists, it is vital to develop a reliable way to mitigate such variation. The recent use of deep learning to improve the cognitive or technical skills of endoscopic lesion detection has increased in colonoscopy studies. For colorectal neoplasia detection, Wang et al. and Repici et al. demonstrated the effectiveness of CADe, while Gong et al. verified the effectiveness of CAQ [5–9]. However, the efficacy of a combination intervention has rarely been investigated. Our trial

▲ **Table 3** Withdrawal time and over-speed analysis.

	Control group	CADe group	CAQ group	COMBO group	P^1 value	P^2 value	P^3 value	P^4 value	Group with computer-aided quality control[5]	Group without computer-aided quality control[6]	*P* value
Withdrawal time, mean (SD), minutes	9.71 (4.13)	10.52 (4.24)	10.14 (4.14)	10.17 (4.11)	0.291	0.056	0.302	0.413	10.12 (4.20)	10.15 (4.12)	0.909
Withdrawal time without operation, mean (SD), minutes	9.36 (4.09)	9.94 (3.93)	9.46 (3.66)	9.46 (3.47)	0.422	0.153	0.794	0.2	9.65 (4.01)	9.46 (3.56)	0.471
Proportion of over-speed, mean (SD), %	27.19 (12.08)	25.35 (11.81)	21.67 (10.34)	20.78 (10.17)	<0.001	0.127	<0.001	<0.001	26.26 (11.97)	21.24 (10.25)	<0.001

CADe, computer-aided detection; CAQ, computer-aided quality improvement; COMBO, CADe plus CAQ.
[1] Analysis of variance between four groups.
[2] *t* test between control group and CADe group.
[3] *t* test between control group and CAQ group.
[4] *t* test between CADe group and COMBO group.
[5] Groups with computer-aided quality control included the CAQ and COMBO groups.
[6] Groups without computer-aided quality control included CADe and control groups.

has shown that CADe equipped with CAQ can yield a higher ADR.

CADe was designed to augment the endoscopists' cognitive ability to discriminate between the candidate lesion and surrounding healthy mucosa, by automatically presenting a box around the putative lesion [5]. However, the efficacy of the CADe system on lesion detection depended on the quality of mucosa exposure. In CADe video validation, Urban et al. found that the polyps that were missed by CADe were actually located in the segments with hurried withdrawal, which suggests that CADe cannot compensate for an unqualified inspection technique [20]. CAQ contributed to a more thorough examination of the mucosa; thus, the combined intervention further enhanced the effect of CADe. In this study, the use of CADe did not significantly improve the efficacy of CAQ. This may be attributed to the low withdrawal speed, which provided sufficient time for the endoscopists to detect the polyps that were at risk of being missed, and thus partially compensated for the insufficient efficacy of CADe. Moreover, perhaps due to similar efficacy, no significant difference was found in ADR between the CAQ and CADe groups.

Regarding the lesion size, the results of previous studies and our study reported that the increase in ADR in the CADe group compared with the control group was mainly due to an increase in diminutive and small adenomas (diminutive: $P=0.073$, OR 1.547, 95 %CI 0.959 to 2.495; small: $P=0.064$, OR 3.01, 95 %CI 0.939 to 9.653; large: $P=0.671$, OR 1.618, 95 %CI 0.176 to 14.882) [6–8]. The observation that more diminutive and small adenomas were detected by the CADe system supported the conventional view that small polyps are more likely to be missed within the visual field compared with bigger and more prominent polyps. In contrast, more advanced and large adenomas were detected using the CAQ system (with or without CADe). A previous study found that patients with advanced adenoma were at a significantly increased risk of developing CRC compared with those with no adenoma [21]. Having more large lesions detected with the CAQ system may be attributed to the low and uniform speed throughout the withdrawal process. The low withdrawal speed allowed the endoscopist to conduct a more comprehensive examination of proximal haustral folds, flexures, and rectal valves, and also to distend the lumen and expose the mucosa more thoroughly [22]. Further studies should address the role of CADe plus CAQ in decreasing interval cancer, which is the main goal of screening colonoscopy.

Regarding the withdrawal time, use of CAQ did not significantly prolong the negative withdrawal time compared with no CAQ (10.12 [SD 4.20] vs. 10.15 [SD 4.12], respectively; $P=0.909$). Previous studies on the relationship between the withdrawal time and lesion detection have had conflicting conclusions. Barclay et al. demonstrated a strong correlation between the withdrawal time and lesion detection, whereas Sawhney et al. and Adler et al. did not [23–25]. In the current study, a negative withdrawal time of more than 6 minutes was required for each procedure. However, it is possible that endoscopists maintained a high speed of withdrawal in the proximal colon while spending more time in the rectum and sigmoid colon to reach a total withdrawal time of 6 minutes. Thus, even if the withdra-

wal time of the whole colon is 6 minutes, the ADR may still significantly differ among endoscopists [26]. In the current study, a significantly lower proportion of over-speed frames was observed in the COMBO group compared with the CADe group, as well as in the CAQ group compared with the control group. This lower proportion may explain the significantly improved ADR when applying the COMBO intervention, and suggests that the proportion of over-speed frames was a reliable indicator of withdrawal quality. Further studies should be designed to confirm the correlation between endoscopist performance, ADR, and proportion of over-speed frames.

In this study, the CADe system showed high fidelity and was competent for consistent automatic detection of colon polyps. The system produced no false-negative results, while the false-positive rate was low. The similar negative withdrawal time among the control, CADe, and COMBO groups suggests that the false-positive cases of the system did not result in additional withdrawal time. The performance of our CADe system was comparable with that reported in previous trials, and thus further confirmed the conclusion of the current study [6].

There are some limitations to this study. First, the study suffers from a lack of external validity. Factors including genetic, dietary, lifestyle, and habitual differences between Chinese and Western populations, as well as differences in the morbidity of colon polyps/adenomas among our patients could have contributed to the differences observed. Hence, an international multicenter study is being designed to investigate the global adaptability and effectiveness of our findings. Second, this was an open-label study and the endoscopists' behavior may have been biased toward improving performance when randomized to the computer-assisted interventions. As there is strong evidence that endoscopists' ADRs vary, tandem studies may be better at characterizing improvements in adenoma detection than the current study design.

To the best of our knowledge, this is the first four-group parallel study to test the efficacy of CADe and CAQ, both alone and combined. The design of the study allowed evaluation of the supplementary effect between two computer vision technologies that have emerged in recent years. Our findings support the conclusion that the combination of CADe with CAQ is an effective method of further increasing the ADR during colonoscopy, without prolonging the withdrawal time.

Acknowledgments

The authors would like to express their gratitude to EditSprings (https://www.editsprings.cn/) for the expert linguistic services provided.

Funding

the Project of Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incisio 2018BCC337
Hubei Province Major Science and Technology Innovation Project 2018–916–000–008

Clinical trial

Trial Registration: ClinicalTrials.gov | Registration number (trial ID): NCT04453956 | Type of study: Prospective, Randomized, Single-Center Study

Competing interests

The authors declare that they have no conflict of interest.

References

[1] Siegel RL, Miller KD, Goding Sauer A et al. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020; 70: 145–164

[2] Doubeni CA, Corley DA, Quinn VP et al. Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. Gut 2018; 67: 291–298

[3] Zhao S, Wang S, Pan P et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. Gastroenterology 2019; 156: 1661–1674

[4] 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: 741–748

[5] Repici A, Badalamenti M, Maselli R et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020; 159: 512–520

[6] Wang P, Berzin TM, Brown JRG et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019; 68: 1813–1819

[7] Wang P, Liu P, Brown JRG et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. Gastroenterology 2020; 159: 1252–1261

[8] 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: 343–351

[9] Gong D, Wu L, Zhang J et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. Lancet Gastroenterol Hepatol 2020; 5: 352–361

[10] Lami M, Singh H, Dilley JH et al. Gaze patterns hold key to unlocking successful search strategies and increasing polyp detection rate in colonoscopy. Endoscopy 2018; 50: 701–707

[11] Kudo T, Saito Y, Ikematsu H et al. New-generation full-spectrum endoscopy versus standard forward-viewing colonoscopy: a multicenter, randomized, tandem colonoscopy trial (J-FUSE Study). Gastrointest Endosc 2018; 88: 854–864

[12] Aslanian HR, Shieh FK, Chan FW et al. Nurse observation during colonoscopy increases polyp detection: a randomized prospective study. Am J Gastroenterol 2013; 108: 166–172

[13] Luo Y, Zhang Y, Liu M et al. Artificial intelligence-assisted colonoscopy for detection of colon polyps: a prospective, randomized cohort study. J Gastrointest Surg 2021; 25: 2011–2018

[14] Shaukat A, Rector TS, Church TR et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. Gastroenterology 2015; 149: 952–957

[15] Suzuki S, Gotoda T, Kusano C et al. Seven-day vonoprazan and low-dose amoxicillin dual therapy as first-line treatment: a multicentre randomised trial in Japan. Gut 2020; 69: 1019–1026

[16] Hung IFN, Lung KC, Tso EYK et al. Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. Lancet 2020; 395: 1695–1704

[17] Crockett SD, Nagtegaal ID. Terminology, molecular features, epidemiology, and management of serrated colorectal neoplasia. Gastroenterology 2019; 157: 949–966

[18] Bailey CE, Hu CY, You YN et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975–2010. JAMA Surg 2015; 150: 17–22

[19] Corley DA, Jensen CD, Marks AR et al. Adenoma detection rate and risk of colorectal cancer and death. New Engl J Med 2014; 370: 1298–1306

[20] 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

[21] Click B, Pinsky PF, Hickey T et al. Association of colonoscopy adenoma findings with long-term colorectal cancer incidence. JAMA 2018; 319: 2021–2031

[22] Rex DK, Schoenfeld PS, Cohen J et al. Quality indicators for colonoscopy. Gastrointest Endosc 2015; 81: 31–53

[23] Barclay RL, Vicari JJ, Doughty AS et al. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. New Engl J Med 2006; 355: 2533–2541

[24] Adler A, Wegscheider K, Lieberman D et al. Factors determining the quality of screening colonoscopy: a prospective study on adenoma detection rates, from 12 134 examinations (Berlin colonoscopy project 3, BECOP-3). Gut 2013; 62: 236–241

[25] Sawhney MS, Cury MS, Neeman N et al. Effect of institution-wide policy of colonoscopy withdrawal time ≥7 minutes on polyp detection. Gastroenterology 2008; 135: 1892–1898

[26] Jung Y, Joo YE, Kim HG et al. Relationship between the endoscopic withdrawal time and adenoma/polyp detection rate in individual colonic segments: a KASID multicenter study. Gastrointest Endosc 2019; 89: 523–530