



Impact of User's Background Knowledge and Polyp Characteristics in Colonoscopy with Computer-Aided Detection

Jooyoung Lee¹, Woo Sang Cho², Byeong Soo Kim², Dan Yoon², Jung Kim¹, Ji Hyun Song¹, Sun Young Yang¹, Seon Hee Lim¹, Goh Eun Chung¹, Ji Min Choi¹, Yoo Min Han¹, Hyoun-Joong Kong^{3,4,5,6}, Jung Chan Lee^{3,7,8}, Sungwan Kim^{3,7,8}, Jung Ho Bae¹

¹Department of Internal Medicine and Healthcare Research Institute, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Korea; ²Interdisciplinary Program in Bioengineering, Graduate School, Seoul National University, Seoul, Korea; ³Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul, Korea; ⁴Medical Big Data Research Center, Seoul National University College of Medicine, Seoul, Korea; ⁵Artificial Intelligence Institute, Seoul National University, Seoul, Korea; ⁶Transdisciplinary Department of Medicine and Advanced Technology, Seoul National University Hospital, Seoul, Korea; ⁷Institute of Medical and Biological Engineering, Medical Research Center, Seoul National University, Seoul, Korea; ⁸Institute of Bioengineering, Seoul National University, Seoul, Korea

Article Info

Received February 13, 2024

Revised May 3, 2024

Accepted May 9, 2024

Published online July 26, 2024

Corresponding Author

Jung Ho Bae

ORCID <https://orcid.org/0000-0001-7669-1213>

E-mail bjh@snuh.org

Sungwan Kim

ORCID <https://orcid.org/0000-0002-9318-849X>

E-mail sungwan@snu.ac.kr

Jooyoung Lee and Woo Sang Cho contributed equally to this work as first authors.

Background/Aims: We investigated how interactions between humans and computer-aided detection (CADe) systems are influenced by the user's experience and polyp characteristics.

Methods: We developed a CADe system using YOLOv4, trained on 16,996 polyp images from 1,914 patients and 1,800 synthesized sessile serrated lesion (SSL) images. The performance of polyp detection with CADe assistance was evaluated using a computerized test module. Eighteen participants were grouped by colonoscopy experience (nurses, fellows, and experts). The value added by CADe based on the histopathology and detection difficulty of polyps were analyzed.

Results: The area under the curve for CADe was 0.87 (95% confidence interval [CI], 0.83 to 0.91). CADe assistance increased overall polyp detection accuracy from 69.7% to 77.7% (odds ratio [OR], 1.88; 95% CI, 1.69 to 2.09). However, accuracy decreased when CADe inaccurately detected a polyp (OR, 0.72; 95% CI, 0.58 to 0.87). The impact of CADe assistance was most and least prominent in the nurses (OR, 1.97; 95% CI, 1.71 to 2.27) and the experts (OR, 1.42; 95% CI, 1.15 to 1.74), respectively. Participants demonstrated better sensitivity with CADe assistance, achieving 81.7% for adenomas and 92.4% for easy-to-detect polyps, surpassing the standalone CADe performance of 79.7% and 89.8%, respectively. For SSLs and difficult-to-detect polyps, participants' sensitivities with CADe assistance (66.5% and 71.5%, respectively) were below those of standalone CADe (81.1% and 74.4%). Compared to the other two groups (56.1% and 61.7%), the expert group showed sensitivity closest to that of standalone CADe in detecting SSLs (79.7% vs 81.1%, respectively).

Conclusions: CADe assistance boosts polyp detection significantly, but its effectiveness depends on the user's experience, particularly for challenging lesions. (*Gut Liver* 2024;18:857-866)

Key Words: Colonoscopy; Polyps; Artificial intelligence

INTRODUCTION

With the progression in artificial intelligence (AI) technology, various computer-aided detection (CADe) systems have been developed for colonoscopy to overcome the limitation of human recognition errors.¹ CADe is designed

to assist endoscopists by automatically identifying polyp patterns, thereby minimizing operator variability and enhancing the efficacy of colonoscopy screening. Clinical studies have demonstrated that CADe effectively improves polyp detection, leading to an increase in the adenoma detection rate and a decrease in the adenoma miss rate.²⁻⁷

© Gut and Liver.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Furthermore, a recent meta-analysis including 12 previous randomized controlled trials (RCTs) has indicated the effectiveness of CADe in adenoma detection, regardless of the size, location, and morphology of the adenomas.⁸

However, recent studies examining the practical application of CADe systems have raised concerns about the discrepancy between the expected benefits of CADe demonstrated in previous RCTs and the actual outcomes observed in real-world settings.⁹⁻¹² Those real-world studies have not shown an improved adenoma detection rate in AI-assisted colonoscopy compared to the standard colonoscopy. Particularly, three studies employing the GI-Genius system, recognized as one of the most authentic CADe systems, indicated a declining trend in detecting adenomas or advanced adenomas during the colonoscopies using CADe.^{9,11,12} This observation raises a question for the expectation that endoscopist-CADe collaboration invariably leads to better outcomes than those achieved using either modality independently.

Despite the advancement in AI development, the current CADe system still faces several algorithmic challenges, including a high rate of false positives (approximately 27 false alarms per colonoscopy) and reduced efficacy in detecting elusive lesions, such as sessile serrated lesions (SSLs).¹³ Ultimately, these limitations can have a detrimental impact on endoscopists' perceptions of CADe, potentially leading to distraction or fatigue during procedures and diminishing the reliability of CADe alarms.¹²⁻¹⁴ These potential drawbacks of CADe can impact the interaction between human operators and the AI system, resulting in various outcomes. Additionally, these results might be further influenced by endoscopist-specific factors, such as their overall experience with colonoscopy and the level of their background knowledge of various colon polyps and AI technologies.

This study investigated the impact of CADe on detection performance among 18 staff members with varying levels of colonoscopy experience and background knowledge in multicenter endoscopy departments. We aimed to evaluate the differences in responses to CADe alerts based on the correctness of CADe system assistance and the participants' career. Additionally, we explored the various interactions between humans and CADe in polyp detection, which vary depending on the polyp characteristics.

MATERIALS AND METHODS

1. Study design and participants

In this experimental study, we used a computerized testing module to assess the accurate localization of co-

lon polyps. This module was developed using MATLAB (MATLAB R2020b; MathWorks Inc., Natick, MA, USA). The test was conducted two times on the same set of test images, depending on the CADe assistance.

Overall, 18 participants were enrolled in this study; they were categorized into three groups based on their experience with colonoscopy. Group 1 comprised six nurses who had no experience in performing colonoscopies but had worked in the endoscopy unit as assistant nurses for >3 years. They also had prior experience assisting CADe-assisted colonoscopy. Before participating in this study, all participants in group 1 completed 30-minute educational courses on the clinical knowledge of various colorectal polyps. Group 2 included six fellows from the Department of Gastroenterology at Seoul National University Hospital (SNUH) with >1 year but <3 years of experience with colonoscopy. Group 3 comprised six specialized board-certified gastroenterologists from SNUH Healthcare System Gangnam Center, each with >6 years of colonoscopy experience and had conducted >5,000 colonoscopies (Supplementary Table 1).

This study was conducted as a part of the SNUH Colonoscopy AI (SCAI) project. The study protocol adhered to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the SNUH Institutional Review Board (IRB number: H-2107-235-1240).

2. Test process for polyp detection

In test 1, participants were required to detect polyps within each test image. When the participants could not detect the polyp, their answers were automatically interpreted as the absence of a polyp. If the participants identified a polyp on the test screen, they were required to accurately delineate the boundaries of the polyp using a bounding box. To mitigate recall bias, approximately 2 months after completing test 1, test 2 was conducted. The images used in the test remained unchanged; however, their sequence was randomly altered. In test 2, the same participants conducted the test with CADe assistance. In cases where the CADe detected polyps, the auxiliary AI screen displayed the boundaries of the polyps using a bounding box. Additionally, the confidence probability indicating the potential presence of polyp was explicitly specified (Fig. 1).

3. The composition and characteristics of the test set

The colonoscopy images used in the test were extracted from our prospective database. The prospective database comprised colonoscopy videos and still images obtained from patients who underwent colonoscopy at SNUH Healthcare System Gangnam Center from January 2020

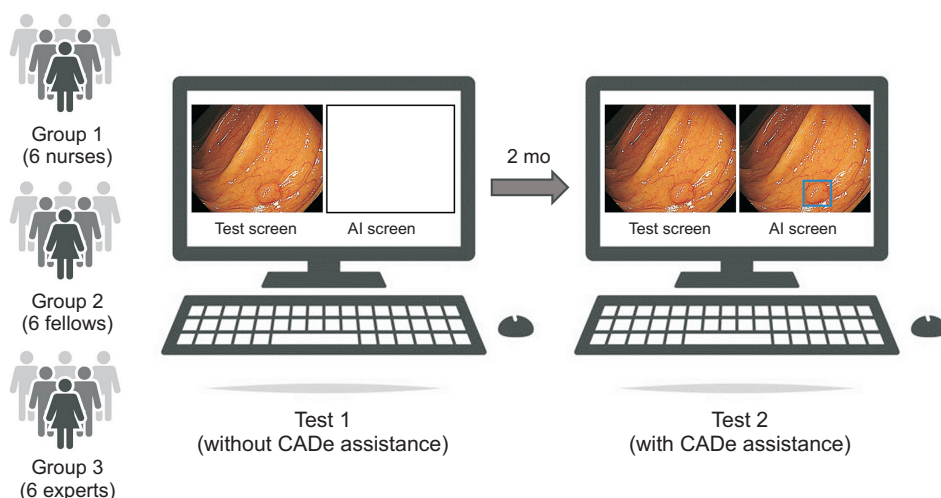


Fig. 1. Study design. CADE, computer-aided detection; AI, artificial intelligence.

to December 2021. All image collections were conducted with written informed consent.

Two experts who did not participate in the test reviewed all images and selected a total of 300 high-quality white-light colonoscopy images, comprising 219 polyps and 81 normal images. The normal images included normal bowel walls (n=63, 77.3%) or bowel contents (n=18, 22.2%), such as bubbles and fecal materials. Among the 219 polyp images, 152, 30, and 37 were adenomatous polyps, hyperplastic polyps, and SSLs, respectively. To investigate the effect of the CADE according to the level of difficulty for polyp detection, all polyp images were classified as “difficult-to-detect polyp” and “easy-to-detect polyp.” A polyp was considered a “difficult-to-detect polyp” if it met at least one of the following criteria: (1) polyps occupying <5% of the total frame; (2) subtle polyps exhibiting only color changes compared to the adjacent normal mucosa; (3) polyps resembling normal structures, such as folds; and (4) polyps situated at the four corner edges of the frame (Supplementary Fig. 1). Among the polyps, 160 (73.1%) and 59 (26.9%) were difficult-to-detect and easy-to-detect polyps, respectively (Table 1).

To establish ground truth for polyps, these two experts separately marked the boundaries of all polyps using the LabelImg version 1.8.6 software (<https://github.com/tzutalin/labelImg>), and assessed the level of detection difficulty. A third expert made minimal adjustments if there were significant disagreements between the two experts’ assessments.

4. Outcome measures and statistical analysis

We assessed the changes in polyp detection performance for each group according to the CADE assistance, using accuracy, sensitivity, and false positive rate (FPR) as evaluation metrics. The difference in the effects of CADE

Table 1. Details of the Test Set Images

Image	Number (n=300)
Polyp image	219
Pathology	
Adenoma (adenocarcinoma)*	152
Hyperplastic polyp	30
Sessile serrated lesion	37
Morphology	
Protruded (Is and Isp)	31
Flat (IIa, IIb, and IIc)	188
Percentage of area occupied by polyps on the screen	
<5%	148
≥5%	71
Levels of difficulty for polyp detection	
Difficult-to-detect	160
Easy-to-detect	59
Normal image	81

*Adenoma (n=151) and adenocarcinoma (n=1).

assistance based on participant groups and polyp characteristics was also evaluated. Additionally, we compared the change in polyp localization performance according to the CADE assistance using the intersection over union (IoU) to measure localization accuracy. An IoU value <0.5 indicated the polyp absence.

The changes in performance metrics (accuracy, sensitivity, FPR, and IoU) were compared using a one-sample proportional test. A generalized linear mixed-effect model was used to measure the effect of CADE on the polyp detection performance. A receiver operating characteristics curve was also plotted to evaluate the detection performance of our CADE system. Statistical significance was set at p<0.05. Categorical variables are presented as frequency counts and percentages. Continuous variables are expressed as means and standard deviations. All statistical analyses were

performed using the R statistical programming software (R Core Team 2022; R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>).

5. CADe model

Colonoscopy images for developing the CADe algorithm were collected from retrospective and prospective databases, as mentioned above. The retrospective dataset was constructed by extracting still images of polyps collected for the Gangnam-Real-Time Optical Diagnosis program, as described in detail by Bae *et al.*¹⁵

We used the YOLOv4 to develop a computer-aided polyp detection algorithm with the following conditions: 100 epochs, 64 batch size, 0.001 learning rate, 0.949 momentum, 0.0005 decay, and nine sizes of anchors optimized with our training dataset. Regarding training, we used 16,996 high-quality images (8,481, 2,956, and 5,559 images of adenomatous polyps, SSLs, and hyperplastic polyps, respectively) from 1,914 patients. The SSL polyp synthesis augmentation technique was used to overcome the low detection rate due to the relative lack of SSL images.¹⁶ We used 1,800 SSL synthetic images for training. Subsequently, 15,532 false-positive images, including fecal or remnant food material from 197 normal videos, were extracted and used for training to reduce the FPR.

To evaluate the performance of CADe, 51 and 39 polyp and normal images, respectively, were used as validation data. The sensitivity and specificity of CADe were 98.0% and 97.0%, respectively; the positive and negative predictive values were 98.0% and 97.0%, respectively.

RESULTS

1. The effect of CADe across different participating groups

For the test set, the CADe system demonstrated an accuracy of 79.0% in detecting polyps. The overall accuracy of all three groups in polyp detection increased significantly from 69.7% to 77.7% ($p < 0.001$) (Table 2). Group 1 (nurses) and group 2 (fellows) showed accuracy increases of 14.7% and 5.7%, respectively, while group 3 (experts) exhibited a 3.6% improvement in accuracy (Table 2, Fig 2A). The overall accuracy in polyp localization assessed using IoU also significantly increased from 0.52 to 0.61 with CADe assistance ($p < 0.001$). The performance of polyp localization according to participating groups is described in Supplementary Table 2.

The FPR of the CADe system for the test set was 19.8%. The overall FPR across the three groups slightly decreased from 21.5% to 20.5% with CADe assistance; however, sta-

tistical significance was not observed ($p = 0.381$). The FPR of group 1 increased from 13.6% to 13.8% ($p = 0.914$), while that of groups 2 and 3 decreased by 1.4% and 1.8% with CADe assistance, respectively (Fig. 2B). These differences were not statistically significant.

The CADe system exhibited a sensitivity of 78.5% in detecting all polyps within the test set. The overall sensitivity in polyp detection increased with CADe assistance by 20.3%, 7.3%, and 4.2% in groups 1, 2, and 3, respectively ($p < 0.001$) (Fig. 2C).

2. Human-AI interaction according to the correctness of CADe

The area under the curve for the CADe system was 0.87 (95% confidence interval [CI], 0.83 to 0.91), indicating good detection performance (Fig. 3). When CADe detected polyps accurately, a significant increase was found in the accuracy of polyp detection among the total participants (odds ratio [OR], 2.87; 95% CI, 2.52 to 3.28; $p < 0.001$) (Table 3). However, the accuracy of total participants in polyp detection decreased when CADe did not accurately detect polyps (OR, 0.72; 95% CI, 0.59 to 0.87; $p < 0.001$). Of the cases that the participants detected incorrectly in test 1, 48.9% were correctly detected with CADe assistance in test 2. Among the cases correctly detected by the participants in test 1, 9.7% were misguided with CADe assistance in test 2 (Table 4).

Table 2. The Effect of CADe on the Accuracy of Polyp Detection

Group	Participants, %	Participants+CADe, %	p-value
Total (n=18)	69.7	77.7	<0.001
Group 1 (nurses)	60.1	74.8	<0.001
N1	48.7	74.7	
N2	59.7	77.7	
N3	62.3	65.7	
N4	66.7	77.0	
N5	56.0	73.0	
N6	67.0	80.7	
Group 2 (fellows)	70.2	75.9	<0.001
F1	58.7	70.7	
F2	74.3	77.7	
F3	69.3	68.7	
F4	75.0	80.0	
F5	73.3	82.7	
F6	70.3	76.0	
Group 3 (experts)	78.9	82.5	0.010
E1	81.3	83.7	
E2	81.7	83.3	
E3	79.3	82.7	
E4	75.3	80.7	
E5	75.7	83.7	
E6	80.3	81.0	

CADe, computer-aided detection.

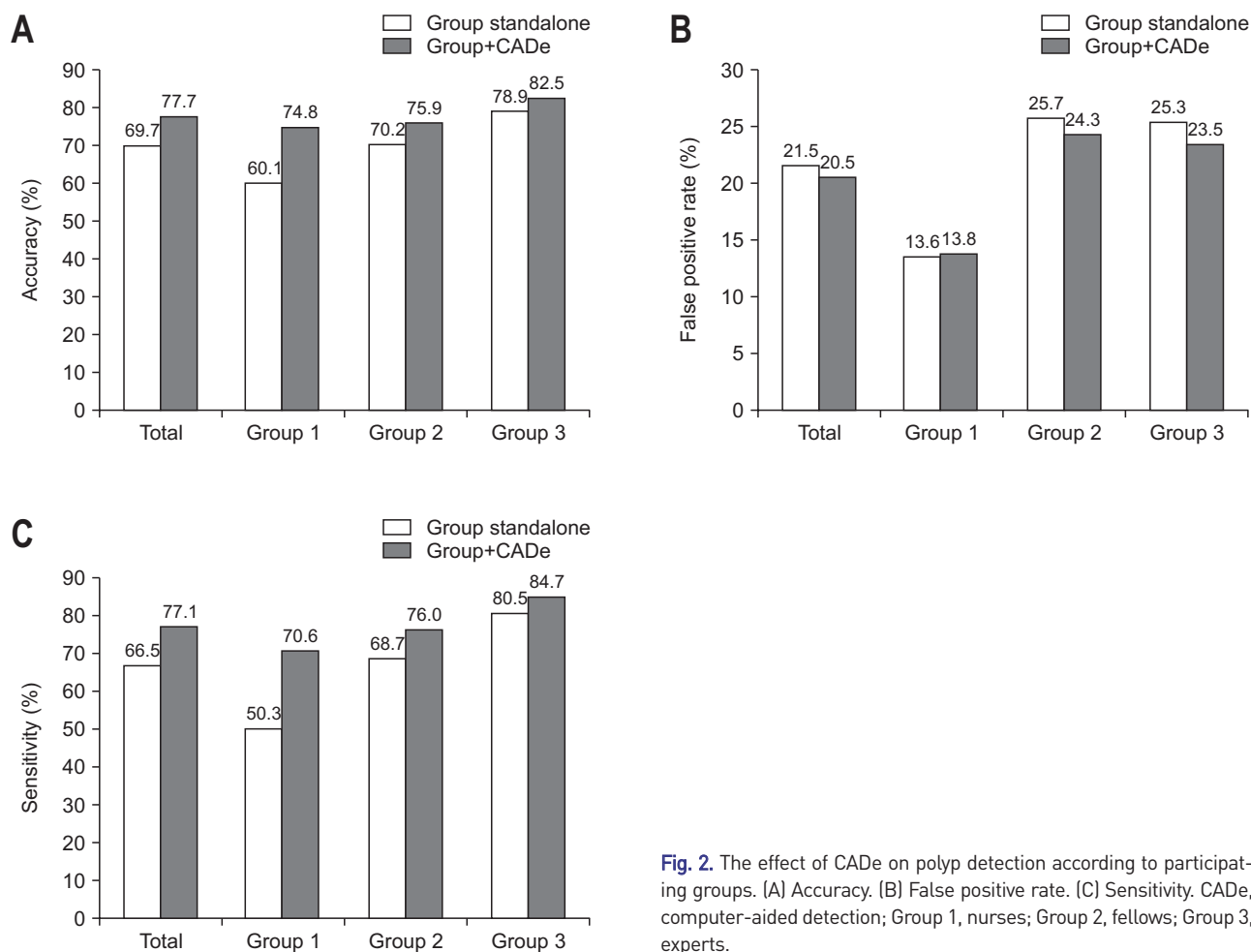


Fig. 2. The effect of CADe on polyp detection according to participating groups. (A) Accuracy. (B) False positive rate. (C) Sensitivity. CADe, computer-aided detection; Group 1, nurses; Group 2, fellows; Group 3, experts.

Table 3. The Interaction between CADe and Participating Groups According to True or False Guidance by the CADe

Group	OR (95% CI)	p-value
Overall		
Total	1.88 (1.69–2.09)	<0.001
Group 1	1.97 (1.71–2.27)	<0.001
Group 2	1.60 (1.33–1.94)	<0.001
Group 3	1.42 (1.15–1.74)	<0.001
Correct CADe assistance		
Total	2.87 (2.52–3.28)	<0.001
Group 1	6.78 (5.25–8.84)	<0.001
Group 2	2.15 (1.71–2.70)	<0.001
Group 3	2.18 (1.68–2.83)	<0.001
Incorrect CADe assistance		
Total	0.72 (0.59–0.87)	<0.001
Group 1	0.66 (0.45–0.96)	0.026
Group 2	0.80 (0.57–1.14)	0.206
Group 3	0.60 (0.42–0.87)	0.005

CADe, computer-aided detection; OR, odds ratio; CI, confidence interval; Group 1, nurses; Group 2, fellows; Group 3, experts.

3. The impact of CADe assistance based on polyp characteristics

The area under the curve of CADe were 0.87 (95% CI,

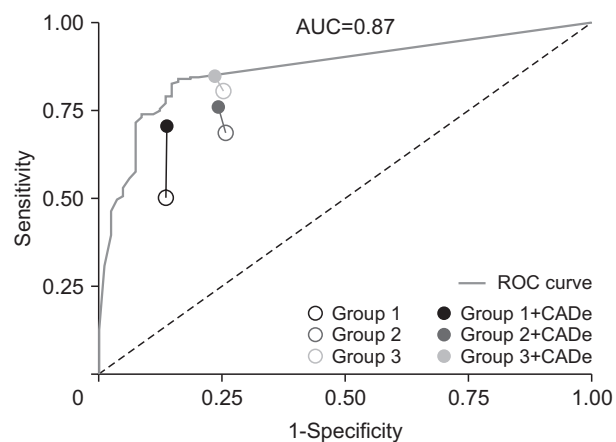


Fig. 3. The ROC curve evaluating detection performance of the CADe system. ROC, receiver operating characteristic; CADe, computer-aided detection; AUC, area under the curve; Group 1, nurses; Group 2, fellows; Group 3, experts.

0.82 to 0.91) for adenomas and 0.92 (95% CI, 0.86 to 0.98) for SSLs, respectively, indicating good-to-excellent performance. For easy-to-detect and difficult-to-detect polyps, the area under the curves of the CADe system were 0.94

Table 4. The Effect of the CADe on Changes of Confusion Matrix

Distribution	Total	Group 1 (nurses)	Group 2 (fellows)	Group 3 (experts)
Corrected cases when assisted/incorrect cases when unassisted	48.9 (799/1,635)	52.2 (375/719)	43.2 (232/537)	50.7 (192/379)
(FP to TN)/FP	49.0 (154/314)	63.6 (42/66)	39.2 (49/125)	51.2 (63/123)
(FN to TP)/FN	48.8 (645/1,321)	51.0 (333/653)	44.4 (183/412)	50.4 (129/256)
Misguided cases when assisted /originally correct cases when unassisted	9.7 (366/3,765)	10.2 (110/1,081)	10.1 (128/1,263)	9.0 (128/1,421)
(TP to FN)/TP	8.7 (227/2,621)	10.1 (67/661)	9.5 (86/902)	7.0 (74/1,058)
(TN to FP)/TN	12.2 (139/1,144)	10.2 (43/420)	11.6 (42/361)	14.9 (54/363)

Data are presented as percent [number/number].

CADe, computer-aided detection; FP, false positive; TN, true negative; FN, false negative; TP, true positive.

Table 5. The Influence of CADe on the Sensitivity According to Pathology and Levels of Difficulty for Polyp Detection

Group	CADe, %	Participants, %	Participants+CADe, %	p-value
Total				
Adenoma	79.5	73.5	81.7	<0.001
Sessile serrated lesion	81.1	49.6	66.5	<0.001
Easy-to-detect	89.8	86.4	92.4	<0.001
Difficult-to-detect	74.4	59.2	71.5	<0.001
Group 1 (nurses)				
Adenoma	79.5	60.6	76.7	<0.001
Sessile serrated lesion	81.1	26.6	58.1	<0.001
Easy-to-detect	89.8	75.4	91.0	<0.001
Difficult-to-detect	74.4	41.0	63.0	<0.001
Group 2 (fellows)				
Adenoma	79.5	75.4	82.2	<0.001
Sessile serrated lesion	81.1	50.9	61.7	0.001
Easy-to-detect	89.8	89.6	91.5	0.237
Difficult-to-detect	74.4	60.9	70.3	<0.001
Group 3 (experts)				
Adenoma	79.5	84.4	86.2	0.151
Sessile serrated lesion	81.1	71.2	79.7	0.003
Easy-to-detect	89.8	94.1	94.6	0.655
Difficult-to-detect	74.4	75.5	81.0	<0.001

CADe, computer-aided detection.

(95% CI, 0.90 to 0.98) and 0.85 (95% CI, 0.80 to 0.89), respectively (Supplementary Fig. 2).

For adenoma detection, the integration of CADe with participants also demonstrated an increase in sensitivity. In group 3, no significant difference was observed in sensitivity for adenoma detection with CADe assistance (84.4% vs 86.2%, $p=0.151$). However, groups 1 and 2 exhibited statistically significant increases in sensitivity by 16.1% and 6.8%, respectively ($p<0.001$) (Table 5, Fig. 4A). When interaction occurred between participants and CADe, the overall sensitivity for adenoma detection exceeded that observed when participants or CADe operated alone (Table 5, Fig. 4A). This tendency was also observed in the detection of easy-to-detect polyps (Table 5, Fig. 4C).

In detecting SSLs and difficult-to-detect polyps, all three groups showed statistically significant increase in sensitivity with CADe assistance (Table 5, Fig 4B and D). However,

participants failed to achieve the performance of CADe as a standalone tool. Nonetheless, group 3 demonstrated results closest to the sensitivity of CADe alone in detecting SSLs (79.7% vs 81.1%, respectively), compared to the other two groups.

DISCUSSION

In this study, we evaluated the interaction between humans and CADe-assisted colonoscopy using a sophisticated test platform involving participants with various levels of expertise. Notably, no report exists on the relationship between the endoscopist and CADe according to the characteristics of the user and the polyps to be detected. Our findings provide a pivotal insight into integrating AI tools in real practice. These data indicate that participants were

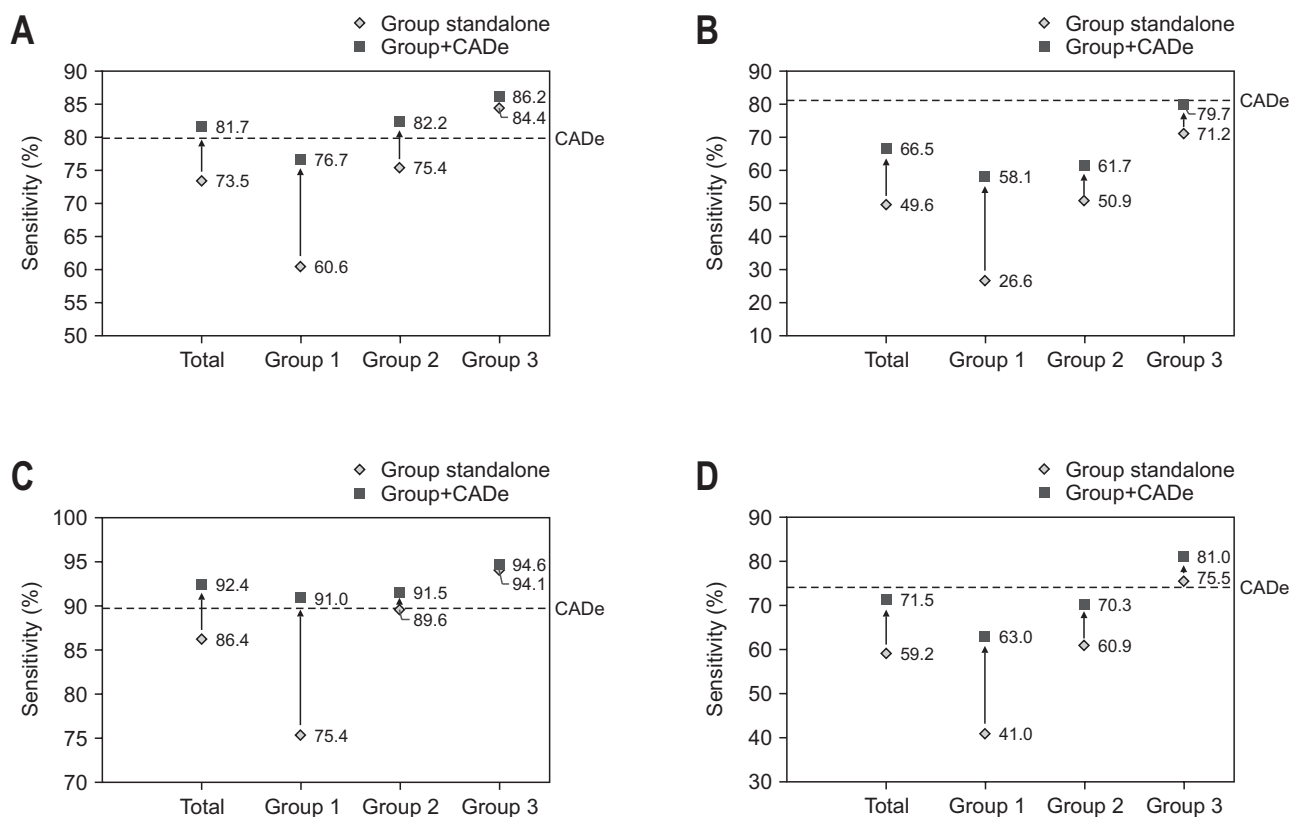


Fig. 4. The influence of CADe assistance in polyp detection: an analysis of sensitivity based on polyp characteristics. (A) Adenoma. (B) Sessile serrated lesion. (C) Easy-to-detect polyp. (D) Difficult-to-detect polyp. CADe, computer-aided detection; Group 1, nurses; Group 2, fellows; Group 3, experts.

more receptive to CADe guidance when it was correct (OR, 2.87) than instances where CADe provided incorrect advice (OR, 0.72). This suggests that endoscopists can discern and adopt appropriate recommendations from CADe. Particularly, the advantage of correct guidance from CADe outweighs the risks associated with inaccurate guidance. Additionally, our results reveal that the interpretation of CADe outputs varies depending on the user's background knowledge (or skill level) and the specific characteristics of the polyp. Consequently, the optimal effect achieved through the collaboration between humans and CADe differs across the groups. This highlights the importance of comprehensive training in foundational colonoscopy knowledge for endoscopists, even in an era increasingly dominated by AI technologies.

A recent meta-analysis including 12 RCTs with 11,340 patients found evidence of CADe effect on adenoma detection, which resulted in 26% and 8.4% increases in the relative and absolute adenoma detection rate, respectively.⁸ Another meta-analysis for miss rates of colon polyps also demonstrated 65% and 78% reduction in adenoma and SSL miss rates, respectively.¹⁷ Nonetheless, recent studies examining CADe performance in real-world settings have

revealed that implementing AI has not improved quality metrics in clinical practice. This striking discrepancy between real-world data and previous RCTs suggested new insight and opportunity for study, enabling us to understand what transformed a useful AI tool into a bothersome assistant in terms of the human-AI interaction.

In this study, the CADe system exhibited a substantial impact on the accuracy of detecting colon polyps across all groups. The human-AI interaction was also most and least significant in the nurse group (OR, 1.97) and the expert group (OR, 1.42), respectively. This finding deviates from our initial expectations, which anticipated that nurses, given their relatively lower level of colonoscopy expertise compared to fellows and experts, would derive limited benefits from CADe assistance. Previous research applying the Unified Theory of Acceptance and Use of Technology model suggests a strong correlation between physicians' overall perceptions of AI-assisted technology and their acceptance and adoption of AI systems.¹⁸ In the preliminary survey of this study, the nurse group demonstrated a relatively higher familiarity with AI technology due to their experience assisting CADe-assisted colonoscopy, which seemed to lead to significant interaction with CADe

despite their relatively lower baseline performance. Conversely, the expert group tended to adhere to their original decisions, potentially reflecting a reliance on their established expertise. This finding underscores the necessity for further research to investigate the impact of user attitudes, particularly their reliance on AI, on the efficacy of the system.

Our study revealed that the nurse and fellow groups exhibited limited performance with the CADe system, notably in detecting SSLs and other challenging lesions, failing to achieve the performance of CADe when used alone. This observation suggests that accumulating in-depth knowledge about complex polyp types, such as SSLs, would significantly enhance the efficacy of AI-assisted colonoscopy. When CADe assisted in recognizing subtle and challenging lesions, the performance (assessed using the sensitivity) of the fellow group was lower than that of the expert group. This observation highlights the need for advanced expertise in the field, even with the assistance of sophisticated technology, while also emphasizing the necessity for more targeted and comprehensive education on challenging lesions for trainees. An increasing body of evidence suggests a correlation between improved SSL detection and the extent of lesion-specific knowledge.¹⁹⁻²¹ In a preliminary questionnaire, all members of the nurse group reported no prior knowledge of optical diagnosis for colon polyps, and only half of the individuals in the fellow group indicated a high level of familiarity with the Workgroup Serrated Polyps and Polyposis classification. In contrast, members of the expert group were well acquainted with the Workgroup Serrated Polyps and Polyposis classification and had undergone training in optical diagnosis for SSLs.^{15,20}

Interestingly, this study is the first to identify the effects of the CADe system on nurses' detection performance. Among the three groups, the nurse group had the most significant improvement in polyp detection. In the early 2000s, the interest in and demand for nurses who perform colonoscopies in many countries increased because of the growing demand for colonoscopies as a screening tool for colorectal cancer.^{22,23} In Korea, nurses do not perform colonoscopies because of medicolegal issues. However, their assistance with colonoscopy procedures and therapeutic interventions is essential.²⁴ Improvement in the ability of nurses to recognize polyps is crucial to the effectiveness and safety of colonoscopy procedures. CADe implementation in clinical practice might positively influence the nurses' performance because they are less likely to have sufficient training opportunities than physicians.

This study had some limitations. First, the evaluation of the interaction between the CADe system and participants

was conducted using an *ex vivo* approach, employing a computerized test module rather than a real-time study. Nonetheless, the sophisticated test module assessed the impact of the CADe system on recognition accuracy among participants with differing levels of knowledge and experience. This decision was made due to the impracticality of conducting *in vivo* studies for the design of this study, which aimed to be performed under controlled conditions with consistent polyp characteristics, excluding proficiency in the exposure technique. Second, more than half of the polyp images utilized in the test set were classified as difficult-to-detect, following consensus among three experts. This prevalence may not precisely represent the typical clinical setting and could exaggerate or diminish the effectiveness of the CADe system. Consequently, the CADe system exhibited lower detection performance on the test set compared to the validation phase. However, it is important to note that missed polyps frequently presented as flat or small lesions. We prioritized the occurrence of missed polyps as criterion for identifying difficult-to-detect polyps in categorizing the polyp images within the test set. This approach in our test design underscores the necessity for developing a more sophisticated CADe system capable of more effectively detecting challenging lesions. Additionally, in contrast to previous studies, "the proportion of the lesion in the entire frame," rather than the actual size of the polyp, was used to classify polyps, which reflects the distance of the polyp observed from the operator. This can provide a more realistic assessment of the interaction between humans and CADe systems. Lastly, the performance of CADe systems can play a crucial role in determining their effectiveness. In fact, Nehme *et al.*¹² reported that endoscopists have cited the FPR, a key performance indicator of CADe systems, as a major concern in their incorporation into colonoscopy practice. Furthermore, a recent study that directly compare the performance of different commercial CADe systems indicated that clinicians might take these performance discrepancies into account when selecting a CADe system to meet specific requirements.²⁵ These findings suggest that CADe systems with varying performance levels could have diverse impacts on user's behaviors and attitudes towards CADe. The CADe system used in this study was an early-stage model that demonstrated substantial performance levels. Further research is necessary to determine how the interaction between CADe and users may differ when a higher-performance CADe system is employed.

In conclusion, CADe assistance enhanced performance for polyp detection across various groups, encompassing nurses, fellows, and experts. However, the extent of its impact on polyp recognition and the synergistic interaction

between human operators and CADe systems appears to be influenced by the career stage of the participants and the specific characteristics of the polyps. This variability may stem from differences in the endoscopists' foundational knowledge of colonoscopy or their attitudes toward adopting new technologies. Establishing a harmonious collaboration between humans and AI necessitates a focus on continuous education in both basic colonoscopy skills and AI technology. Moreover, fostering a balanced attitude (neither over-reliance nor under-reliance) towards AI technology is crucial for the successful integration of CADe systems into colonoscopy practices.

CONFLICTS OF INTEREST

J.H.B. holds equity in AINEX corporation. All the other authors declare that they do not have any competing interest.

ACKNOWLEDGEMENTS

This work was supported by Seoul National University Medical Big Data Research Center (MBRC) and AINEX Corporation.

The authors are grateful to Gu-Cheol Jung for statistical analysis of this work.

AUTHOR CONTRIBUTIONS

Study concept and design: J.H.B. Data acquisition: J.K., J.H.S., S.H.L., G.E.C., J.M.C., Y.M.H. Data analysis and interpretation: J.L., W.S.C., S.Y.Y. Drafting of the manuscript: J.L., W.S.C. Critical revision of the manuscript for important intellectual content: J.H.B. Statistical analysis: J.L. Obtained funding: J.H.B., S.K. Administrative, technical, or material support: W.S.C., B.S.K., D.Y., H.J.K., J.C.L. Study supervision: J.H.B., S.K. Approval of final manuscript: all authors.

ORCID

Jooyoung Lee <https://orcid.org/0000-0003-1652-088X>
 Woo Sang Cho <https://orcid.org/0000-0003-0032-5836>
 Byeong Soo Kim <https://orcid.org/0000-0001-8767-9842>
 Dan Yoon <https://orcid.org/0000-0002-5657-5984>
 Jung Kim <https://orcid.org/0000-0001-9509-0851>
 Ji Hyun Song <https://orcid.org/0000-0001-9459-9250>

Sun Young Yang <https://orcid.org/0000-0003-4766-3752>
 Seon Hee Lim <https://orcid.org/0000-0001-6174-7165>
 Goh Eun Chung <https://orcid.org/0000-0001-8344-7737>
 Ji Min Choi <https://orcid.org/0000-0001-8611-4647>
 Yoo Min Han <https://orcid.org/0000-0001-5580-8427>
 Hyoun-Joong Kong <https://orcid.org/0000-0001-5456-4862>
 Jung Chan Lee <https://orcid.org/0000-0002-4546-0962>
 Sungwan Kim <https://orcid.org/0000-0002-9318-849X>
 Jung Ho Bae <https://orcid.org/0000-0001-7669-1213>

SUPPLEMENTARY MATERIALS

Supplementary materials can be accessed at <https://doi.org/10.5009/gnl240068>.

REFERENCES

1. Kamitani Y, Nonaka K, Isomoto H. Current status and future perspectives of artificial intelligence in colonoscopy. *J Clin Med* 2022;11:2923.
2. 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.
3. 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.
4. 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.
5. 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;68:1813-1819.
6. Wang P, Liu P, Glissen Brown JR, 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.
7. 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.
8. Shiha MG, Oka P, Raju SA, et al. Artificial intelligence-assisted colonoscopy for adenoma and polyp detection: an updated systematic review and meta-analysis. *iGIE* 2023;2:333-343.
9. Levy I, Bruckmayer L, Klang E, Ben-Horin S, Kopylov U.

- Artificial intelligence-aided colonoscopy does not increase adenoma detection rate in routine clinical practice. *Am J Gastroenterol* 2022;117:1871-1873.
10. Quan SY, Wei MT, Lee J, et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. *Sci Rep* 2022;12:6598.
 11. Ladabaum U, Shepard J, Weng Y, Desai M, Singer SJ, Manalithara A. Computer-aided detection of polyps does not improve colonoscopist performance in a pragmatic implementation trial. *Gastroenterology* 2023;164:481-483.
 12. Nehme F, Coronel E, Barringer DA, et al. Performance and attitudes toward real-time computer-aided polyp detection during colonoscopy in a large tertiary referral center in the United States. *Gastrointest Endosc* 2023;98:100-109.
 13. Spadaccini M, Hassan C, Alfarone L, et al. Comparing the number and relevance of false activations between 2 artificial intelligence computer-aided detection systems: the NOISE study. *Gastrointest Endosc* 2022;95:975-981.
 14. Hassan C, Wallace MB, Sharma P, et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. *Gut* 2020;69:799-800.
 15. Bae JH, Lee C, Kang HY, et al. Improved real-time optical diagnosis of colorectal polyps following a comprehensive training program. *Clin Gastroenterol Hepatol* 2019;17:2479-2488.
 16. Yoon D, Kong HJ, Kim BS, et al. Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network. *Sci Rep* 2022;12:261.
 17. Shah S, Park N, Chehade NEH, et al. Effect of computer-aided colonoscopy on adenoma miss rates and polyp detection: a systematic review and meta-analysis. *J Gastroenterol Hepatol* 2023;38:162-176.
 18. Zhai H, Yang X, Xue J, et al. Radiation oncologists' perceptions of adopting an artificial intelligence-assisted contouring technology: model development and questionnaire study. *J Med Internet Res* 2021;23:e27122.
 19. Bleijenberg AGC, van Leerdam ME, Bargeman M, et al. Substantial and sustained improvement of serrated polyp detection after a simple educational intervention: results from a prospective controlled trial. *Gut* 2020;69:2150-2158.
 20. Lee J, Bae JH, Chung SJ, et al. Impact of comprehensive optical diagnosis training using Workgroup serrated polyps and Polyposis classification on detection of adenoma and sessile serrated lesion. *Dig Endosc* 2022;34:180-190.
 21. Li D, Woolfrey J, Jiang SF, et al. Diagnosis and predictors of sessile serrated adenoma after educational training in a large, community-based, integrated healthcare setting. *Gastrointest Endosc* 2018;87:755-765.
 22. Shum NF, Lui YL, Choi HK, Lau SC, Ho JW. A comprehensive training programme for nurse endoscopist performing flexible sigmoidoscopy in Hong Kong. *J Clin Nurs* 2010;19:1891-1896.
 23. van Putten PG, Ter Borg F, Adang RP, et al. Nurse endoscopists perform colonoscopies according to the international standard and with high patient satisfaction. *Endoscopy* 2012;44:1127-1132.
 24. Verschuur EM, Kuipers EJ, Siersema PD. Nurses working in GI and endoscopic practice: a review. *Gastrointest Endosc* 2007;65:469-479.
 25. Troya J, Sudarevic B, Krenzer A, et al. Direct comparison of multiple computer-aided polyp detection systems. *Endoscopy* 2024;56:63-69.