BMJ Open Gastroenterology

Multicentre study to assess the performance of an artificial intelligence instrument to support qualitative diagnosis of colorectal polyps

Keigo Sato,[1] Mizuki Kuramochi,[2] Akihiko Tsuchiya,[3] Akihiro Yamaguchi,[2] Yasuo Hosoda,[2] Norio Yamaguchi,[3] Naohiro Nakamura,[3] Yuki Itoi,[1] Yu Hashimoto,[1] Kengo Kasuga,[1] Hirohito Tanaka,[1] Shiko Kuribayashi,[1] Yoji Takeuchi,[1] Toshio Uraoka [1]

To cite: Sato K, Kuramochi M, Tsuchiya A, *et al*. Multicentre study to assess the performance of an artificial intelligence instrument to support qualitative diagnosis of colorectal polyps. *BMJ Open Gastroenterol* 2024;**11**:e001553. doi:10.1136/bmjgast-2024-001553

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjgast-2024-001553).

Received 6 August 2024
Accepted 30 September 2024

[1]Department of Gastroenterology and Hepatology, Gunma University Graduate School of Medicine School of Medicine Faculty of Medicine, Maebashi, Gunma, Japan
[2]Department of Gastroenterology, National Hospital Organization Saitama Hospital, Wako, Saitama, Japan
[3]Ageo Central General Hospital, Ageo, Saitama, Japan

Correspondence to
Dr Toshio Uraoka;
toshi_urao@yahoo.co.jp

ABSTRACT

Objective Computer-aided diagnosis (CAD) using artificial intelligence (AI) is expected to support the characterisation of colorectal lesions, which is clinically relevant for efficient colorectal cancer prevention. We conducted this study to assess the diagnostic performance of commercially available CAD systems.

Methods This was a multicentre, prospective performance evaluation study. The endoscopist diagnosed polyps using white light imaging, followed by non-magnified blue light imaging (non-mBLI) and mBLI. AI subsequently assessed the lesions using non-mBLI (non-mAI), followed by mBLI (mAI). Eventually, endoscopists made the final diagnosis by integrating the AI diagnosis (AI+endoscopist). The primary endpoint was the accuracy of the AI diagnosis of neoplastic lesions. The diagnostic performance of each modality (sensitivity, specificity and accuracy) and confidence levels were also assessed.

Results Overall, 380 lesions from 139 patients were included in the analysis. The accuracy of non-mAI was 83%, 95% CI (79% to 87%), which was inferior to that of mBLI (89%, 95% CI (85% to 92%)) and mAI (89%, 95% CI (85% to 92%)). The accuracy (95% CI) of diagnosis by expert endoscopists using mAI (91%, 95% CI (87% to 94%)) was comparable to that of expert endoscopists using mBLI (91%, 95% CI (87% to 94%)) but better than that of non-expert endoscopists using mAI (83%, 95% CI (75% to 90%)). The level of confidence in making a correct diagnosis was increased when using magnification and AI.

Conclusions The diagnostic performance of mAI for differentiating colonic lesions is comparable to that of endoscopists, regardless of their experience. However, it can be affected by the use of magnification as well as the endoscopists' level of experience.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Computer-aided diagnosis (CADx) using artificial intelligence (AI) can support the characterisation of colorectal lesions. To date, most published CADx-focused studies have evaluated polyps of ≤5 mm in size in the rectum to sigmoid colon based on the preservation and incorporation of valuable endoscopic innovations statement at the advanced institute where AI was developed.

WHAT THIS STUDY ADDS

⇒ The results illustrate that the diagnostic performance of AI with magnified blue light imaging for differentiating colonic lesions is comparable to the diagnostic performance of endoscopists. Furthermore, magnifying endoscopy can enhance the diagnostic performance of both endoscopists and AI, and AI can improve the diagnostic accuracy of non-expert endoscopists.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ CADx may be more effective when combined with magnified endoscopy and when used by non-experts, although good image quality is required.

INTRODUCTION

The adenoma-carcinoma sequence[1] has been recognised as a central pathway for the development of colorectal cancer, and endoscopic resection of neoplastic polyps in the colorectum has been performed. The National Polyp Study showed that endoscopic resection of neoplastic polyps reduces the incidence of colorectal cancer as well as reducing mortality.[2] Because typical hyperplastic polyps are considered to be unrelated to future cancer development[3] and endoscopic treatment for such indolent polyps only increases the burden, cost and risk of adverse events, neoplastic polyps should be differentiated from hyperplastic polyps to avoid unnecessary endoscopic interventions. In addition, the American Society for Gastrointestinal Endoscopy's Preservation and Incorporation of Valuable Endoscopic Innovations statement mentioned that the 'resect and discard' strategy,[4–7] which can decide the

next surveillance interval after polypectomy without a pathological diagnosis, can be realised when the negative predictive value for adenoma is >90%. Therefore, better differentiation accuracy between neoplastic and non-neoplastic polyps is crucial for efficacious management of colorectal polyps. Currently, endoscopic diagnosis using magnifying endoscopy with image enhancement endoscopy (IEE) is commonly used.[8] Still, conventional endoscopic observation sometimes makes it difficult to distinguish between neoplastic and non-neoplastic polyps, especially for non-experts.

Computer-aided detection (CADe) and computer-aided diagnosis (CADx) using artificial intelligence (AI) technology are expected to support the detection and characterisation of colorectal lesions during routine colonoscopy,[9] and a few AI systems have been launched in the past few years. The clinical question is whether AI may be a useful alternative to conventional diagnostic modalities, such as magnified endoscopic diagnosis by IEE. However, these AI systems have not been widely used because there is a lack of information on their effectiveness in different patient groups and in different clinical situations. To date, it has been suggested that the diagnostic performance of AI is comparable to that of endoscopists. However, there are few prospective studies.[9–11] Additionally, most of the published CADx-focused studies have evaluated polyps of ≤5 mm in size in the rectum to sigmoid colon based on PIVI at the advanced institute where AI was developed.[12–15] We conducted this study to investigate the diagnostic capability of a commercially available CAD system for polyps and to evaluate its usefulness in daily clinical practice.

METHODS

Study design and patients

This prospective, performance evaluation study was conducted at the endoscopy units of three centres (Gunma University Hospital, Ageo Central General Hospital and National Hospital Organisation Saitama Hospital) between March 2022 and December 2022. Patients who were scheduled to undergo colonoscopy and who were between the ages of 20 and 85 years at the time of enrolment were eligible. Patients with positive faecal immunochemistry tests who were planned for surveillance colonoscopy after colorectal polypectomy; with digestive symptoms, such as abdominal pain or constipation; who underwent screening colonoscopy; or in whom the endoscopist in charge deemed colonoscopy as necessary, had the contents of the consent document explained to them, and they provided voluntary written informed consent. Patients with inflammatory bowel disease, familial adenomatous polyposis, a history of colorectal resection other than an appendectomy, colorectal stenosis, pregnancy, abnormal blood coagulation function, an inability to manage anticoagulation and antiplatelet medication according to Japanese guidelines,[16] or severe organ failure, as well as patients who were deemed inappropriate for enrolment by the endoscopist, were excluded. The history of previous colonoscopies was not considered.

The study protocol was approved by the Clinical Research Network Fukuoka Certified Review Board and was performed in accordance with the ethical standards outlined in the 1964 Declaration of Helsinki and its later amendments. The study is registered in the Japan Registry of Clinical Trials (jRCTs072220003). This manuscript follows the reporting of diagnostic accuracy studies (STARD) guidelines.[17]

Computer-aided diagnosis system

Fujifilm Corporation developed and launched CAD EYE in November 2020. CAD EYE was developed using deep learning from vast clinical data. CAD EYE consists of a lesion detection support function for CADe and a lesion characterisation support function for CADx of colorectal polyps and other lesions. When CADe detects a suspected lesion, such as a polyp, it supports the endoscopists' detection by displaying a frame around the target area on the endoscope screen. CADx assists endoscopists in their diagnosis by displaying the results of differentiation of suspected lesions, such as polyps, as either neoplastic or non-neoplastic lesions, including hyperplastic polyps. If the differential result is neoplastic, the diagnosis is displayed in yellow; if non-neoplastic, it is displayed in green at the outer edge of the endoscopic image. CAD EYE can be used with white light imaging (WLI) and blue light imaging (BLI) with or without magnification (online supplemental figure 1).

We used high-definition colonoscopes (EC-L600ZP, EC-L600ZP7, EC-L600ZP7_L, EC-L600ZW7, EC-760ZP-V_L, EC-760ZP-V_M, EC-760Z-V_L or EC-760Z-V_M) with light sources (LL-7000 and BL-7000), a video processor (VP-7000) and CAD EYE (EX-1, EW10-EC02 and EW10-SC01).

Procedures

Preparation began on the day before colonoscopy and included a low-fibre diet and administration of preparatory medicines, such as antispasmodic agents, to achieve adequate bowel preparation as per the institution's protocol. Sedatives were also administered according to each institution's protocol. Fifteen endoscopists (4 non-experts and 11 experts) who were available in clinical practice were allocated. The endoscopists were divided into expert and non-expert endoscopists according to whether they had performed greater or fewer than 1500 colonoscopy procedures. Both experts and non-experts routinely performed endoscopies using magnification. The colonoscope was first inserted into the caecum under WLI. The endoscopist observed the colorectal mucosa to detect lesions from the cecum to the anus under WLI while withdrawing the scope with or without the CADe system. Once the lesion was detected, after checking that the CADx had not been activated, the endoscopist made the diagnosis using WLI, non-magnified BLI (non-mBLI)

and magnified BLI (mBLI), respectively, without using CADx (online supplemental figure 2). The diagnosis was classified as neoplastic, hyperplastic or others (eg, juvenile polyps and inflammatory polyps), and the confidence level was evaluated as high or low. The diagnosis using WLI was made based on the endoscopists' experience; the diagnosis using BLI was made based on the NBI International Colorectal Endoscopic classification,[18 19] and the diagnosis using mBLI was made using the Japan NBI Expert Team classification.[8] Hyperplastic polyps and other lesions, such as juvenile polyps and inflammatory polyps, diagnosed by the endoscopists were categorised together as non-neoplastic polyps to compare with the AI diagnosis. When the endoscopists judged the lesions as whitish diminutive polyps in the rectosigmoid colon, invasive cancers or submucosal tumours, they were excluded from the analysis. Because CAD EYE cannot distinguish between sessile serrated lesions (SSLs) and hyperplastic polyps, lesions diagnosed as SSLs were classified as hyperplastic. After diagnosis by the endoscopists and activation of AI, AI diagnosis using non-mBLI (non-mAI) was performed, followed by AI diagnosis using mBLI (mAI). When CADx did not respond to the lesion within 5 s, it was classified as 'no response'. After the AI diagnosis was made, the endoscopists re-evaluated their diagnosis and judged their confidence level (AI+endoscopist). Each diagnosis was noted on the case report form by an independent nurse immediately after the endoscopist and AI made their diagnosis and then was entered into the electronic data capture (EDC) system on the same day. In cases where bowel preparation was judged as poor or inadequate by the endoscopist in charge, insertion into the cecum could not be performed, no colorectal lesion was deemed eligible for the study, or the eligible lesion could not be diagnosed by AI, the patients were excluded from the protocol and treated according to the endoscopist's decision outside of the protocol.

All evaluated lesions were endoscopically removed immediately or within 3 months, including by cold polypectomy, snare polypectomy, endoscopic mucosal resection (EMR) or endoscopic submucosal dissection (ESD). The resected lesions were retrieved, fixed in 10% formalin buffer, and subjected to histological processing and evaluation. Histopathological evaluation was performed by at least two pathologists (concealed to the endoscopic images at each institution) according to the ninth edition of the Japanese Classification of Colorectal Carcinoma.[20] If the pathological diagnosis differed between the two pathologists, they consulted with each other to reach a consensus on the diagnosis. The histopathological diagnosis was entered into the EDC system approximately 1–2 weeks after the endoscopic procedure.

According to the Paris classification, the location, size and macroscopic type of all detected lesions were documented.[21] The size of the detected lesion was measured using 2.2-mm closed biopsy forceps or an opened electrosurgical snare. The reference standard was histopathology, and the diagnostic performance (sensitivity, specificity and accuracy) was evaluated in comparison with the reference standard.

Endpoints

The primary endpoint was the accuracy of AI diagnoses in the full analysis set (FAS). The accuracy of AI diagnoses was defined as the percentage of lesions for which the CADx matched the pathologists' diagnosis. The secondary endpoints included the sensitivity and specificity of the AI diagnoses for neoplastic lesions; the diagnostic performance (accuracy, sensitivity and specificity) of the endoscopists using each modality for neoplastic lesions; the diagnostic performance of the endoscopists based on their colonoscopy experience; changes in confidence before and after the AI diagnosis; and adverse events in the safety analysis set (SAF). Adverse events were evaluated according to the common terminology criteria for adverse events, V.5.0.

Sample size calculation

As the main objective of this study was to prospectively investigate the diagnostic performance of CADx, the target number of cases was set after considering the possibility of case accumulation within the planned enrolment period and the estimation accuracy. The accuracy of AI diagnosis, which was the primary endpoint of this study, was expected to be around 0.85–0.90 based on a previous study.[3] To consider the data as sufficiently accurate, we considered that the 95% CI of the accuracy should be within 0.1. We used the Clopper-Pearson method to calculate the CI, and if the point estimate of the accuracy was ≥0.85, more than 211 cases were required. On the basis of in-house data from Gunma University Hospital, the number of target lesions per patient was calculated to be 1.46, and the number of required cases was 144.5. Considering dropout cases, the sample size was set at 170 cases.

Statistical analysis

Eligible patients were those enrolled in the study, excluding those who violated enrollment. After excluding the patients who could not complete colonoscopy, all eligible patients were included in the SAF. The FAS included all patients who had lesions that were eligible for the analysis.

Point estimates of accuracy, sensitivity and specificity for neoplastic lesions with their 95% CIs were calculated using the Clopper-Pearson method for each of the diagnoses. Statistical analyses were performed using Statistical Analysis System software, V.9.4 (SAS Institute, Cary, North Carolina, USA).

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting, or dissemination plans at any stage of this research.

RESULTS

Study participant flow

A total of 176 available patients were enrolled. 167 were eligible, while nine were excluded because they did not

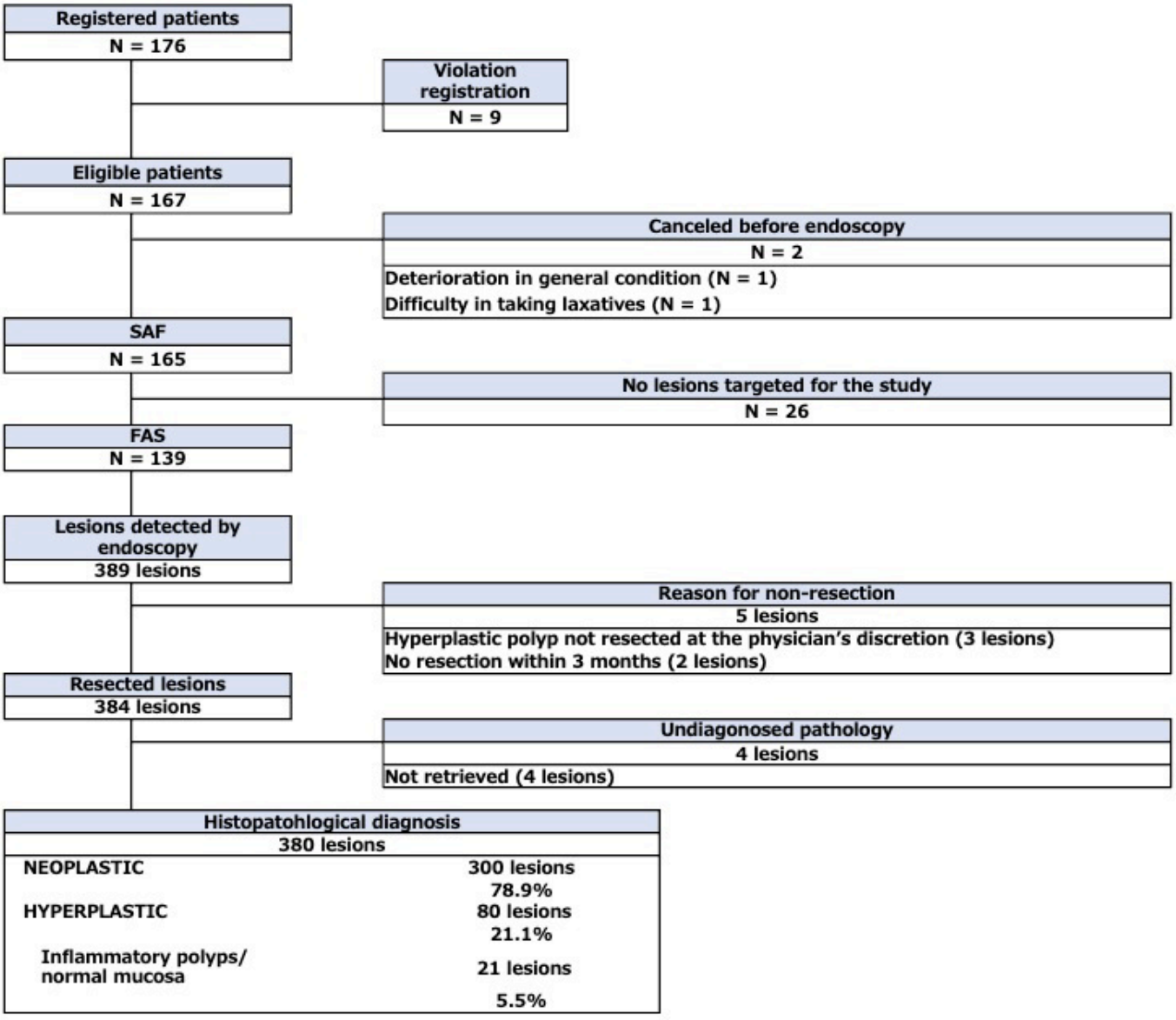

Figure 1 Patient flowchart. FAS, full analysis set; SAF, safety analysis set.

meet the inclusion criteria. Of the enrolled patients, two did not undergo colonoscopy; therefore, 165 patients were included in the SAF. Of these, 26 had no eligible lesions for the analysis, and thus 139 patients were included in the FAS. Among the 389 lesions detected in the FAS, five were not resected and three could not be retrieved, resulting in no pathological diagnosis. The diagnostic performance was evaluated using these 380 lesions (figure 1).

Patient and lesion characteristics

The characteristics of the patients and lesions are shown in table 1. Expert endoscopists performed 70% of the colonoscopies. The lesions were most frequently located in the right-sided colon, and protruded lesions were more frequent than flat lesions. Approximately two-thirds of the lesions were diminutive (1–5 mm). Pathological diagnosis indicated that 291/380 lesions (77%) were adenomas, 9/380 (2.4%) were adenocarcinomas, 48/380 (13.3%) were hyperplastic polyps, 11/380 (2.9%) were SSLs and 21/380 (5.5%) were inflammatory polyps or normal mucosa.

Endoscopic and pathological diagnosis

Online supplemental table 1 shows the endoscopic and pathological diagnoses. The endoscopists diagnosed 294/380 lesions (77%) as neoplastic by WLI, 302/380 (80%) by non-mBLI and 307/380 (81%) by mBLI. CADx diagnosed 278 lesions (73%) as neoplastic by non-mBLI (non-mAI) and 299 lesions (79%) by mBLI (mAI). The endoscopists judged 304 lesions (80%) as neoplastic after considering the diagnosis of AI (AI+endoscopist).

The diagnostic performance of each modality is shown in table 2. The diagnostic accuracy of non-mAI was 82.9%. When AI was used with magnification (mAI), the diagnostic accuracy increased to 88.7%. Both the sensitivity and specificity of mAI were better than those of non-mAI. The accuracy and sensitivity of the endoscopists'

Table 1 Patient and lesion characteristics

		Patients, lesions (N=139, n=380)
Sex	Male, n (%)	95 (68%)
Age	Years, median (range)	70 (32–85)
Endoscopists' experience	<1500, n (%)	114 (30%)
	≥1500, n (%)	266 (70%)
Location	Right-sided colon, n (%)	214 (56%)
	Left-sided colon, n (%)	142 (38%)
	Rectum, n (%)	24 (6%)
Macroscopic type	0–I (protruded), n (%)	274 (72%)
	0–II (superficial), n (%)	106 (28%)
Polyp size	mm, median (range)	4 (1–25)
	1–5 mm, n (%)	250 (66%)
	6–9 mm, n (%)	71 (19%)
	≥10 mm, n (%)	59 (16%)
Histology	Adenoma, n (%)	291 (76.6%)
	Adenocarcinoma, n (%)	9 (2.4%)
	Hyperplastic, n (%)	48 (13.3%)
	Sessile serrated lesion, n (%)	11 (2.9%)
	Inflammation polyps/ normal mucosa, n (%)	21 (5.5%)

diagnoses improved when non-mBLI and mBLI were used. However, the specificity of the endoscopists' diagnoses did not change regardless of whether non-mBLI and mBLI were used. The final endoscopists' diagnoses with consideration of the AI diagnosis (AI+endoscopist) showed the best diagnostic performance, and the sensitivity and accuracy were almost the same as the endoscopists' diagnoses with mBLI.

Comparison between non-expert and expert endoscopists

In the subgroup analysis according to the endoscopists' experience (table 3), the diagnostic performance of AI when evaluated by non-experts was worse than when evaluated by experts. Furthermore, the specificity and accuracy of the diagnoses by non-experts without AI were worse than the diagnoses of experts without AI. The diagnostic performance of non-experts+AI was better than that of non-experts without AI or AI alone. The diagnostic performance of experts after considering the AI diagnosis was similar to or worse than the diagnostic performance of the experts when using mBLI without AI.

Changes in the confidence level of endoscopists in reaching an accurate diagnosis

The changes in the confidence level of endoscopists in making an accurate diagnosis and diagnostic accuracy are shown in figure 2. The percentage of lesions

Table 2 Sensitivity, specificity and accuracy of the diagnosis for all lesions

	Sensitivity (95% CI) (%) N=300			Specificity (95% CI) (%) N=80			Accuracy (95% CI) (%) N=380		
	WLI	BLI without magnification	BLI with magnification	WLI	BLI without magnification	BLI with magnification	WLI	BLI without magnification	BLI with magnification
Endoscopist diagnosis	90.0 (86.0 to 93.2) 270	93.0 (89.5 to 95.6) 279	94.3 (91.1 to 96.7) 283	68.8 (57.4 to 78.7) 55	70.0 (58.7 to 79.7) 56	68.8 (57.4 to 78.7) 55	85.5 (81.6 to 88.9) 325	88.2 (84.5 to 91.2) 335	88.9 (85.4 to 91.9) 338
AI diagnosis	–	86.3 (81.9 to 90.0) 259	92.7 (89.1 to 95.3) 278	–	70.0 (58.7 to 79.7) 56	73.8 (62.7 to 83.0) 59	–	82.9 (78.7 to 86.5) 315	88.7 (85.1 to 91.7) 337
AI+endoscopist diagnosis	94.3 (91.1 to 96.7) 283			71.3 (60.0 to 80.8) 57			89.5 (85.9 to 92.4) 340		

AI, artificial intelligence; BLI, blue light imaging; WLI, white light imaging.

Table 3 Subgroup analysis according to the level of colonoscopy experience (<1500 cases vs ≥1500 cases)

		Sensitivity(95% CI)(%)			Specificity (95% CI) (%)			Accuracy (95% CI) (%)		
		WLI	BLI without magnification	BLI with magnification	WLI	BLI without magnification	BLI with magnification	WLI	BLI without magnification	BLI with magnification
Non-expert<1500	Endoscopist diagnosis	90.1 (82.1 to 95.4) 82	91.2 (83.4 to 96.1) 83	92.3 (84.8 to 96.9) 84	43.5 (23.2 to 65.5) 10	43.5 (23.2 to 65.5) 10	47.8 (26.8 to 69.4) 11	80.7 (72.3 to 87.5) 92	81.6 (73.2 to 88.2) 93	83.3 (75.2 to 89.7) 95
	AI diagnosis	–	87.9 (79.4 to 93.8) 80	90.1 (82.1 to 95.4) 82	–	52.2 (30.6 to 73.2) 12	56.5 (34.5 to 76.8) 13	–	80.7 (72.3 to 87.5) 92	83.3 (75.2 to 89.7) 95
	AI+endoscopist diagnosis	92.3 (84.8 to 96.9) 84			60.9 (38.5 to 80.3) 24			86.0 (78.2 to 91.8) 98		
Expert≥1500	Endoscopist diagnosis	90.0 (85.1 to 93.7) 188	93.8 (89.6 to 96.6) 196	95.2 (91.4 to 97.7) 199	78.9 (66.1 to 88.6) 45	80.7 (68.1 to 90.0) 46	77.2 (64.2 to 87.3) 44	87.6 (83.0 to 91.3) 233	91.0 (86.9 to 94.1) 242	91.4 (87.3 to 94.4) 243
	AI diagnosis	–	85.6 (80.1 to 90.1) 179	93.8 (89.6 to 96.6) 196	–	77.2 (64.2 to 87.3) 44	80.7 (68.1 to 90.0) 46	–	83.8 (78.8 to 88.0) 223	91.0 (86.9 to 94.1) 242
	AI+endoscopist diagnosis	95.2 (91.4 to 97.7) 199			75.4 (62.2 to 85.9) 43			91.0 (86.9 to 94.1) 242		

AI, artificial intelligence; BLI, blue light imaging; WLI, white light imaging.

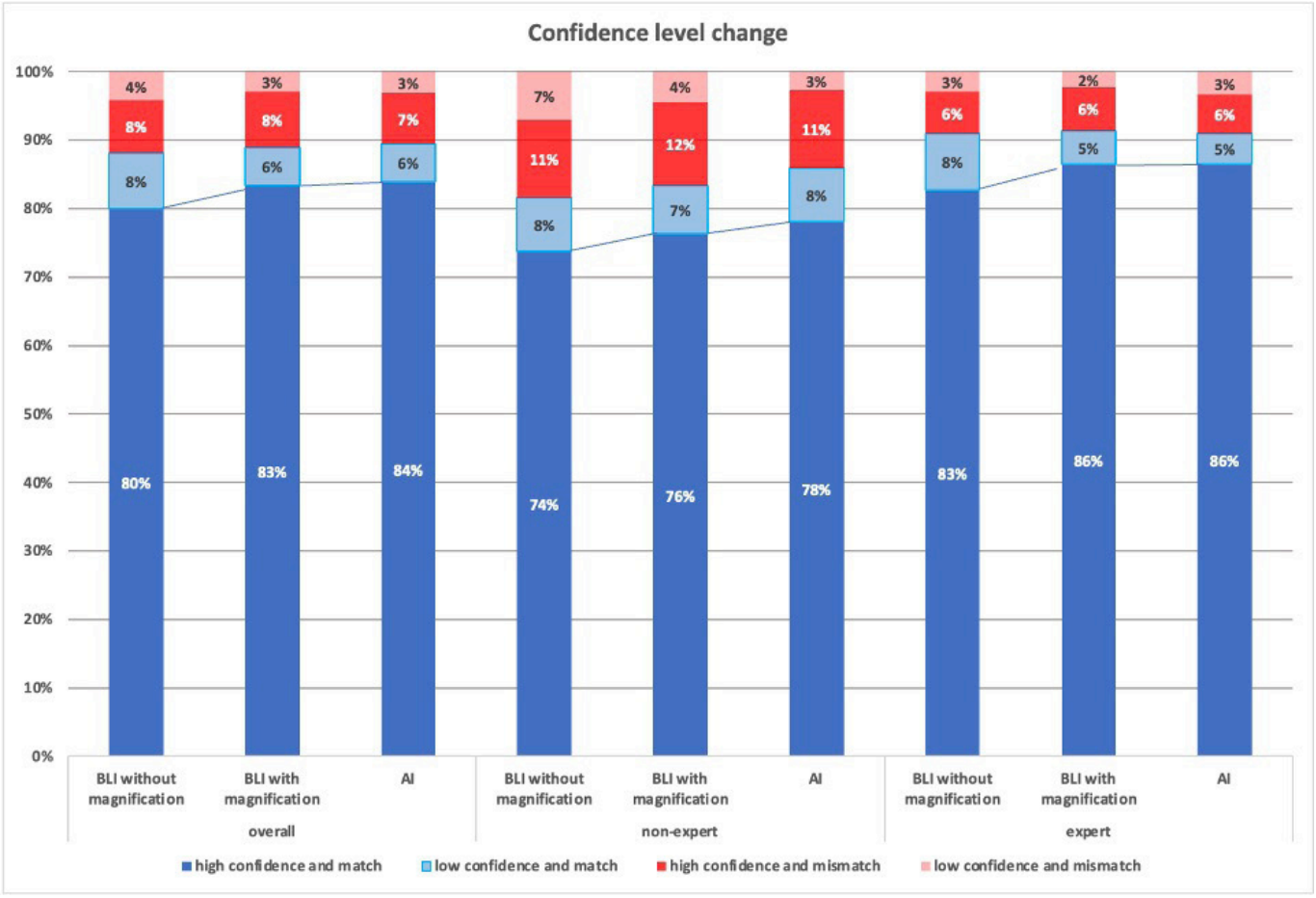

Figure 2 Change in the level of confidence of endoscopists in making an accurate diagnosis and diagnostic accuracy by modality. AI, artificial intelligence; BLI, blue light imaging.

diagnosed with high confidence and with a correct diagnosis was increased when using magnification and AI. The difference seemed more apparent when magnification was used by experts, but AI also improved accuracy. The number of lesions with an accurate diagnosis with a high level of confidence was increased slightly, while the number of lesions with an inaccurate diagnosis with a high level of confidence was decreased in the non-expert group. The proportions were almost unchanged in the expert group.

Adverse events

Among the SAF, one patient had a Grade 2 adverse event of lower gastrointestinal bleeding after endoscopic resection. No other adverse events were observed.

DISCUSSION

In this prospective multicentre pilot trial, we investigated the diagnostic performance of AI and magnification for differentiating neoplastic and non-neoplastic colorectal lesions. The novelty of the present study lies in its prospective design, which differentiates it from previous studies on CADx for similar lesions that were conducted retrospectively. The results indicate that the diagnostic performance of AI is comparable to that of expert endoscopists. Moreover, although the lower limit of the 95% CI for the accuracy of mAI was >85%, which was the threshold for sample size calculation, the lower limit of the 95% CI for the accuracy of non-mAI was lower than this threshold. Magnifying endoscopy can enhance the diagnostic performance of both endoscopists and AI. The experience of the endoscopist can affect the diagnostic performance, especially the specificity. Furthermore, AI can improve the diagnostic accuracy of non-expert endoscopists, and AI can increase the level of confidence of endoscopists in making a correct diagnosis. In addition, while the diagnostic performance of AI should theoretically be the same whether used by non-experts or experts, it was shown that the diagnostic performance of AI was inferior when used by non-experts.

The primary endpoint of this study was the diagnostic accuracy of AI when used with non-mBLI (non-mAI; 82.9% (95% CI 78.7% to 86.5%)) and mBLI (mAI; 88.7% (95% CI 85.1% to 91.7%)). Although CAD EYE can be used with or without magnification, combining it with magnification might help to achieve a more accurate diagnosis. The sensitivity, specificity and accuracy of non-mAI for differentiating neoplastic lesions from non-neoplastic lesions in previous studies were 85%–92%, 79%–87% and 84%–89%, respectively.[9 11] The

sensitivity, specificity and accuracy of mAI in a previous study were 91%, 85% and 88%, respectively.[9] In addition, mBLI showed better accuracy than WLI and non-mBLI by endoscopists. Similar trends have been reported in previous studies.[22 23] In the present study, AI diagnosis was improved by magnification. The diagnostic performance in the current study was comparable with earlier data obtained at advanced centres for AI. Our data may contribute to the more widespread use of mAI on the basis that we assessed its performance in a practical setting with endoscopists of varying levels of experience at community hospitals.

In the subgroup analysis, the expert endoscopists showed better diagnostic performance than the non-expert endoscopists. Although the experts did not demonstrate an improvement after considering the AI diagnosis, the non-experts did, especially in terms of specificity. Therefore, AI (CADx) may be more beneficial when used by non-experts, and there is no apparent advantage when used by experts. In previous studies, the diagnostic accuracy of non-experts, when supported by AI diagnosis, was comparable to that of experts without AI diagnosis or AI diagnosis alone, and diagnosis by non-experts supported by AI showed better sensitivity and lower specificity than diagnosis by experts or AI alone.[11] Although the definition of 'experts' differs among studies and it is not easy to compare them directly, the tendency is consistent with the current study. Our study showed that the diagnostic performance of AI was inferior when used by non-experts, although theoretically, it should be the same, irrespective of whether it is used by non-experts or experts. Our findings may suggest that AI diagnosis is affected by the quality of the recorded image, and the ability of experts to obtain better-quality images may improve AI diagnosis. Therefore, endoscopists should continuously improve their level of skill, even in the AI era. Overall, the accuracy of non-mAI was 82.9%, mAI was 88.7% and that for endoscopists with mBLI was 88.9%. For experts, the accuracy of non-mAI was 83.8%, mAI was 91.0% and that for experts with mBLI was 91.4%. Therefore, the results of mAI and diagnosis by an endoscopist using mBLI were similar in all cases. Therefore, mAI may be a useful alternative to endoscopic diagnosis with IEE and magnification when combined with magnification. The increase in diagnostic accuracy with mAI compared with non-mAI was less pronounced for non-experts than for experts. In addition, the increase in the accuracy of AI with mBLI compared with non-mAI was less pronounced for non-experts than for experts. This means that the barrier to non-experts is the magnification technique. As mentioned above, it is essential to be skilled in endoscopic manipulation, but endoscopic devices have been developed, and it has become possible to easily perform magnified observation. The development of endoscopic instruments has had a synergistic effect on accurate AI diagnosis.

In terms of the level of confidence of endoscopists in making an accurate diagnosis, mBLI showed an increase in the level of confidence in diagnostic accuracy and AI, although the difference was small. This tendency was more commonly observed in the non-expert group than in the expert group. Although only minor changes were observed, AI tended to make endoscopists more confident in making a correct diagnosis and increased the level of confidence in avoiding an incorrect diagnosis. The current insufficiency of evidence on the diagnostic capabilities of AI may have caused some participants to have less confidence in AI. A better understanding of the capabilities of AI by endoscopists may improve their confidence in endoscopic diagnosis using AI.

CAD EYE has several limitations that should be considered. First, current CAD EYE cannot distinguish between SSLs and hyperplastic polyps. However, CAD EYE will be able to distinguish between SSLs and hyperplastic polyps in the future. Second, the AI diagnosis was usually made immediately after observation, but in some cases, AI could not determine the diagnosis immediately. Finally, CAD EYE can distinguish between neoplastic and hyperplastic lesions, but it would be better to assess the grade of dysplasia or to further evaluate the depth of invasion to decide the treatment strategy for detected lesions, including cold polypectomy, EMR, ESD and surgery. If possible, we can make treatment decisions with AI without a pathological diagnosis.

One of the strengths of this study was the participation of endoscopists with different levels of experience from community hospitals, which made it possible to investigate the efficacy of AI in the clinical practice setting. Furthermore, we showed that the diagnostic performance of AI could be affected by the endoscopists' experience.

Despite these strengths, this study had several limitations. First, the sample size was calculated to show that the accuracy of AI diagnosis would be 85%–90%, not to compare this approach with other modalities. A future confirmatory study should be conducted to validate our results. Second, each pathologist at the three different institutions made each pathological diagnosis, and thus the diagnoses were not unified by central review. However, this study aimed to investigate the diagnostic performance of AI in a practical setting, which was feasible for the concept of this study. In addition, the pathological diagnosis in this study was simple (differentiating between neoplastic and non-neoplastic lesions), and thus the intraobserver difference would not be significant. Third, pathological diagnosis for diminutive polyps (<5 mm) is not always correct because we remove diminutive polyps by cold polypectomy, including normal mucosa, and tiny neoplastic tissue cannot be correctly sectioned in the evaluated specimens.[24] Therefore, the specificity when evaluating many diminutive polyps can be low compared with the sensitivity. However, the additional value of AI in terms of improving the specificity in the current study was reliable because we evaluated the same lesions. Fourth, there were more expert endoscopists than non-expert endoscopists. The overall results were more influenced by expert endoscopists. Finally, colonoscopy time was not

investigated in this study because we needed to wait for 5 s before the AI diagnosis settled down and to record the diagnosis of each modality on the CRF.

CONCLUSIONS

The diagnostic performance of AI for differentiating colonic neoplastic lesions from non-neoplastic lesions is comparable to diagnosis by endoscopists, regardless of their experience. It may be more effective when combined with magnified endoscopy and when used by non-experts, although good image quality is required.

Acknowledgements The authors are grateful to Fujifilm Japan for supporting this study. We also thank Takeshi Takajo, Yoriyuki Takamori, Yuji Mishina, Masahiro Akashi, and Ryutaro Sumi for performing the endoscopy procedures. We thank Emily Woodhouse, PhD, from Edanz (https://jp.edanz.com/achttps://jp.edanz.com/ac) for editing a draft of this manuscript.

Contributors KS and TU: designed the study. KS, MK, AT, AY, YH, NY, NN, YI, YH, KK, HT, SK, YT and TU: collected data. KS, YT and TU: analysed and interpreted data. KS, YT and TU: drafted the manuscript. All authors approved the final manuscript. TU is the guarantor.

Funding This work was supported by Fujifilm Japan (award/grant number is not applicable).

Competing interests We received research funding from Fujifilm Japan. There are no other conflicts of interest.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. The study protocol was approved by the Clinical Research Network Fukuoka Certified Review Board with approval reference number CRB7180004. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

ORCID iD

Toshio Uraoka http://orcid.org/0000-0002-4425-4331

REFERENCES

1 Vogelstein B, Fearon ER, Hamilton SR, *et al*. Genetic alterations during colorectal–tumor development. *N Engl J Med* 1988;319:525–32.
2 Winawer SJ, Zauber AG, Ho MN, *et al*. Prevention of Colorectal Cancer by Colonoscopic Polypectomy. *N Engl J Med* 1993;329:1977–81.
3 Laiyemo AO, Murphy G, Sansbury LB, *et al*. Hyperplastic Polyps and the Risk of Adenoma Recurrence in the Polyp Prevention Trial. *Clin Gastroenterol Hepatol* 2009;7:192–7.
4 Shaukat A, Kaltenbach T, Dominitz JA, *et al*. Endoscopic Recognition and Management Strategies for Malignant Colorectal Polyps: Recommendations of the US Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2020;115:1751–67.
5 Hassan C, Antonelli G, Dumonceau J-M, *et al*. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2020. *Endoscopy* 2020;52:687–700.
6 Takeuchi Y, Hanafusa M, Kanzaki H, *et al*. Proposal of a new "resect and discard" strategy using magnifying narrow band imaging: pilot study of diagnostic accuracy. *Dig Endosc* 2014;26 Suppl 2:90–7.
7 Takeuchi Y, Hanafusa M, Kanzaki H, *et al*. An alternative option for "resect and discard" strategy, using magnifying narrow-band imaging: a prospective "proof-of-principle" study. *J Gastroenterol* 2015;50:1017–26.
8 Sano Y, Tanaka S, Kudo S-E, *et al*. Narrow-band imaging (NBI) magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI Expert Team. *Dig Endosc* 2016;28:526–33.
9 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 Colorectal Dis* 2021;36:2237–45.
10 Sakamoto T, Nakashima H, Nakamura K, *et al*. Performance of Computer-Aided Detection and Diagnosis of Colorectal Polyps Compares to That of Experienced Endoscopists. *Dig Dis Sci* 2022;67:3976–83.
11 Weigt J, Repici A, Antonelli G, *et al*. Performance of a new integrated computerassisted system (CADe/CADx) for detection and characterization of colorectal neoplasia. *Endoscopy* 2022;54:180–4.
12 Barua I, Wieszczy P, Kudo S, *et al*. Real-time artificial intelligence-based optical diagnosis of neoplastic polyps during colonoscopy. *NEJM Evid* 2022.
13 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.
14 Mori Y, Kudo SE, Misawa M, *et al*. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Ann Intern Med* 2018;169:357–66.
15 Hassan C, Balsamo G, Lorenzetti R, *et al*. Artificial Intelligence Allows Leaving-In-Situ Colorectal Polyps. *Clin Gastroenterol Hepatol* 2022;20:2505–13.
16 Fujimoto K, Fujishiro M, Kato M, *et al*. Guidelines for gastroenterological endoscopy in patients undergoing antithrombotic treatment. *Dig Endosc* 2014;26:1–14.
17 The updated STARD 2015 reporting guideline for diagnostic accuracy studies. 2023. Available: https://www.equator-network.org/reporting-guidelines/stard/ [Accessed 26 Feb 2024].
18 Hewett DG, Kaltenbach T, Sano Y, *et al*. Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. *Gastroenterology* 2012;143:599–607.
19 Hayashi N, Tanaka S, Hewett DG, *et al*. Endoscopic prediction of deep submucosal invasive carcinoma: validation of the narrow-band imaging international colorectal endoscopic (NICE) classification. *Gastrointest Endosc* 2013;78:625–32.
20 Japanese Society for Cancer of the Colon and Rectum. *Japanese classification of colorectal, appendiceal, and anal carcinoma*. 9th edn. Tokyo: Kanehara & Co, 2018.
21 Endoscopic Classification Review Group. Update on the Paris Classification of Superficial Neoplastic Lesions in the Digestive Tract. *Endoscopy* 2005;37:570–8.
22 Sano Y, Ikematsu H, Fu KI, *et al*. Meshed capillary vessels by use of narrow-band imaging for differential diagnosis of small colorectal polyps. *Gastrointest Endosc* 2009;69:278–83.
23 Wada Y, Kudo SE, Kashida H, *et al*. Diagnosis of colorectal lesions with the magnifying narrow-band imaging system. *Gastrointest Endosc* 2009;80:522–31.
24 Ponugoti P, Rastogi A, Kaltenbach T, *et al*. Disagreement between high confidence endoscopic adenoma prediction and histopathological diagnosis in colonic lesions ≤3 mm in size. *Endoscopy* 2019;51:221–6.