Thieme

Accuracy of polyp characterization by artificial intelligence and endoscopists: a prospective, non-randomized study in a tertiary endoscopy center

OPEN ACCESS

Authors

Sebastian Baumer[1‡], Kilian Streicher[1‡], Saleh A. Alqahtani[2,3], Dominic Brookman-Amissah[1], Monika Brunner[1], Christoph Federle[1], Klaus Muehlenberg[1], Lukas Pfeifer[1], Andrea Salzberger[1], Wolfgang Schorr[1], Jozef Zustin[4,5], Oliver Pech[1]

Institutions

1 Department of Gastroenterology and Interventional Endoscopy, Krankenhaus Barmherzige Brüder Regensburg, Regensburg, Germany
2 Department of Gastroenterology and Hepatology, Johns Hopkins Hospital, Baltimore, United States
3 Liver Transplant Center, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia
4 Private Practice, Histopathology Service Private Practice, Regensburg, Germany
5 Gerhard-Domagk-Institute of Pathology, Universitätsklinikum Münster, Munster, Germany

received 17.2.2023
accepted after revision 8.5.2023

Bibliography
Endosc Int Open 2023; 11: E818–E828
DOI 10.1055/a-2096-2960
ISSN 2364-3722

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

Corresponding author
Prof. Dr. Oliver Pech, Krankenhaus Barmherzige Brüder Regensburg, Department of Gastroenterology and Interventional Endoscopy, Regensburg, Germany
oliver.pech@barmherzige-regensburg.de

‡ These authors contributed equally.

ABSTRACT

Background and study aims Artificial intelligence (AI) in gastrointestinal endoscopy is developing very fast. Computer-aided detection of polyps and computer-aided diagnosis (CADx) for polyp characterization are available now. This study was performed to evaluate the diagnostic performance of a new commercially available CADx system in clinical practice.

Patients and methods This prospective, non-randomized study was performed at a tertiary academic endoscopy center from March to August 2022. We included patients receiving a colonoscopy. Polypectomy had to be performed in all polyps. Every patient was examined concurrently by an endoscopist and AI using two opposing screens. The AI system, overseen by a second observer, was not visible to the endoscopist. The primary outcome was accuracy of the AI classifying the polyps into "neoplastic" and "non-neoplastic." The secondary outcome was accuracy of the classification by the endoscopists. Sessile serrated lesions were classified as neoplastic.

Results We included 156 patients (mean age 65; 57 women) with 262 polyps ≤10 mm. Eighty-four were hyperplastic polyps (32.1%), 158 adenomas (60.3%), seven sessile serrated lesions (2.7%) and 13 other entities (normal/inflammatory colonmucosa, lymphoidic polyp) (4.9%) on histological diagnosis. Sensitivity, specificity and accuracy of AI were 89.70% (95% confidence interval [CI]: 84.02%-93.88%), 75.26% (95% CI: 65.46%-83.46%) and 84.35% (95% CI: 79.38%-88.53%), respectively. Sensitivity, specificity and accuracy for less experienced endoscopists (2–5 years of endoscopy) were 95.56% (95% CI: 84.85%-99.46%), 61.54% (95% CI: 40.57%-79.77%) and 83.10% (95% CI: 72.34%-90.95%) and for experienced endoscopists 90.83% (95% CI: 84.19%-95.33%), 71.83% (95% CI: 59.90%-81.87%) and 83.77% (95% CI: 77.76%-88.70%), respectively.

Conclusion Accuracy for polyp characterization by a new commercially available AI system is high, but does not fulfill the criteria for a "resect-and-discard" strategy.

Introduction

Colonoscopy is the most important screening examination for the prevention of colon cancer. Colonoscopy had led to a significant risk reduction for carcinoma incidence [1]. For years, great efforts have been made to improve the quality of colonoscopy. For example, a European training and validation program for endoscopists was introduced to perform optical diagnosis of colorectal lesions [2]. Data show good results overall for validated participants in the program, but the performance levels differed by individuals [3, 4]. Other working groups do not achieve the set goals [5, 6].

Another point to improve the quality of colonoscopy is the development of artificial intelligence (AI) systems for supporting endoscopy. First, AI systems for polyp detection were introduced. Multiple studies in recent years have shown very good adenoma detection rates (ADR) for these systems [7, 8]. However, an increase in the detection of small lesions of questionable relevance was also reported [9]. Removing the small, often non-neoplastic lesions and examining them histologically leads to increasing costs with no relevance for the patient.

In a further step, updates to the existing AI systems and new developments for polyp characterization were introduced. The first image and video studies of these new systems are very promising. Two systematic reviews [10, 11] give a good overview of the reported performance of the systems. The accuracy of the different systems varies between 75% to 95%.

However, there are hardly any data from clinical practice. Real-world data published to date focus on diminutive (≤5 mm) rectosigmoid polyps.[12, 13, 14, 15]. In compliance with the Preservation and Incorporation of Valuable endoscopic Innovations (PIVI) criteria [16] from the American Society for Gastrointestinal Endoscopy (ASGE) or the Simple Optical Diagnosis Accuracy (SODA) competence criteria [17] from the European Society of Gastrointestinal Endoscopy (ESGE), a resect-and-discard strategy or a diagnose-and-leave strategy can be carried out. This procedure would guarantee the highest level of patient protection while being cost-effective and environmentally friendly [18, 19, 20].

An ESGE position statement [21] was recently published to classify AI in this context. Here it is stated that the above criteria for the resect-and-discard or diagnose-and-leave strategy must also be observed for AI systems. The studies published so far on this topic have shown different results. Three of the five studies met the PIVI criteria [12, 13, 14, 15, 22].

In addition, it could be shown that young endoscopists in particular with little experience can benefit from the AI and improve their examination results to an almost expert level [23]. This is probably the greatest strength of AI in the current state of development.

The present study on polyp characterization by a commercially available AI system and by experienced endoscopists is a further study of our group following a polyp detection study, that showed an ADR of AI at expert level [8].

Patients and methods

Study design

This was a prospective, non-randomized single-center study (St. John of God Hospital, Regensburg, Germany). The protocol was approved by a local ethics committee (ethics committee of the University Hospital Regensburg).

Patients

The prospective, non-randomized study was conducted from March 2022 to August 2022 at the St. John of God Hospital, Germany, a maximum care facility with a tertiary endoscopy department. The study included inpatients and outpatients who presented for a diagnostic colonoscopy or planned polypectomy and were able to consent and were at least 18 years old. Exclusion criteria were polyps with a diameter greater than ten millimeters, patients with chronic inflammatory bowel disease, coagulation disorders or drugs that excluded polypectomy, poor general condition (from ASA IV) and pregnancy.

AI system

The AI system for this study was the CE-certified, commercially available GI Genius (Medtronic, Minneapolis, Minnesota, United States). The AI detects and analyzes polyps. It differentiates between "adenoma" and "non-adenoma." If the AI could not make a classification into the previously mentioned categories, "no prediction" was displayed. GI Genius is compatible with while light mode and narrow-band imaging mode.

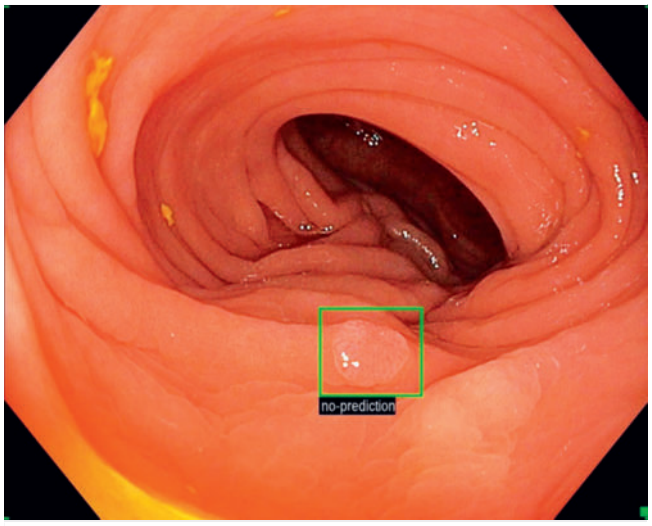
Colonoscopy

A standard bowel preparation was performed. The assessment of the quality of bowel preparation was carried out using the Boston Bowel Preparation Scale (BBPS). Olympus endoscopes of the CF-H 190 and CF-HQ 1100 series with an Olympus Evis X1 processor were used.

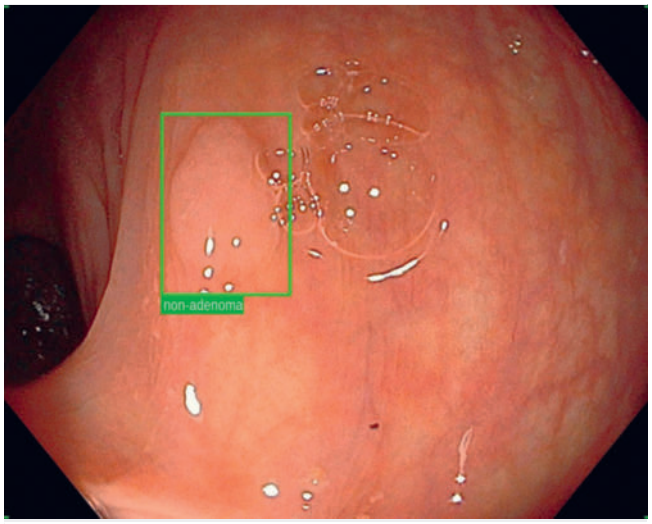
The endoscopies were performed by nine doctors of the department. Of these, three doctors had less than 5 years of experience in endoscopy, another 5 to 10 years of experience, and five more than 10 years of experience.

Study setting

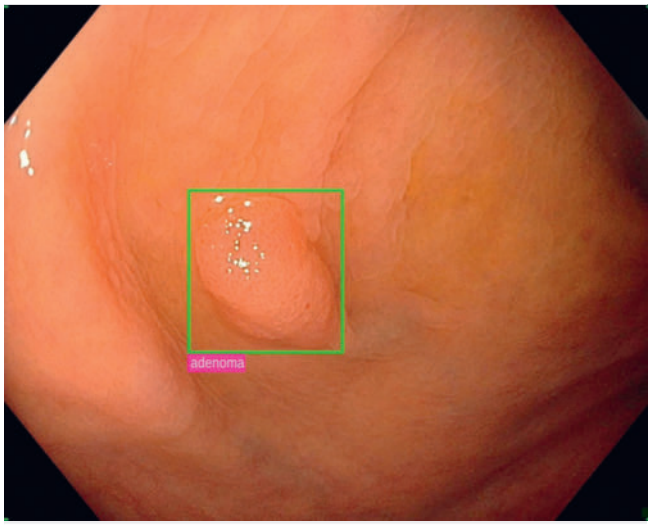
The endoscopy image was displayed on two monitors, the endoscopist could not observe the AI monitor. The endoscopist (investigator 1) only had a view of Monitor 1 (image of processor without AI). A specially trained doctoral student (investigator 2) assessed the image with AI on Monitor 2. The endoscopist performed the colonoscopy as usual. As soon as the endoscopist detected a polyp, location, size, Paris classification, NBI International Colorectal Endoscopic (NICE) classification and workgroup serrated polyps and polyposis (WASP) classification were queried and a final assessment ("adenoma," "carcinoma," "hyperplastic polyp" and "sessile serrated lesion") were documented. After 160 included polyps, the confidence of endoscopic diagnosis was additionally documented (high confidence vs low confidence). The AI marks the polyps with green frames and analyzes them simultaneously during the examination, differentiating between "adenoma", "non-adenoma" and



▶ **Fig. 1** Small polyp marked by AI - no prediction.



▶ **Fig. 2** Small polyp marked by AI - non-adenoma.



▶ **Fig. 3** Small polyp marked by AI – adenoma.

"no prediction" (▶ **Fig. 1, Fig. 2, Fig. 3**). If the AI detected a polyp that the endoscopist had overlooked, Investigator 1 was asked to carefully inspect the marked area again.

The decision for a potential resection of the polyps was based solely on the assessment of the endoscopist.

Histology

Following retrieval of the tissue, the samples were individually preserved and histologically examined by two experienced pathologists. The standard method of histological processing of the polyps is to examine six to eight sections per polyp. These results are the reference to which the AI and endoscopist diagnoses were compared. In the presence of a sessile serrated lesion (SSL), the AI result "adenoma" was rated as truly positive. Among endoscopists, the assessment of SSLs "SSL" and "adenoma" was rated as truly positive, as both have the potential for neoplasia and this results in the same recommendations regarding follow-up intervals.

All initially false-positive polyps, where AI or endoscopists rated "adenoma" while histology was negative, were further examined by the pathologists in a second step. The remaining paraffin blocks were completely stepped up and all cuts were examined. In the section with the most adenoma content, the number of adenoma-positive crypts and their depth was indicated.

Endpoints

The primary endpoint of the study was the rate of correctly classified polyps by AI (accuracy).

Secondary endpoints were the rate of correctly classified polyps by endoscopists and the influence of endoscopists' clinical experience on the classification outcome.

Statistical analysis

Based on our CADe previous study, a patient number of 150 was chosen. A sample size calculation was not possible because no basic clinical data was available for the AI system used.

The McNemar test was used to compare sensitivity, specificity and accuracy. $P < 0.05$ was considered statistically significant. The 95% confidence intervals were calculated as exact Clopper-Pearson confidence intervals.

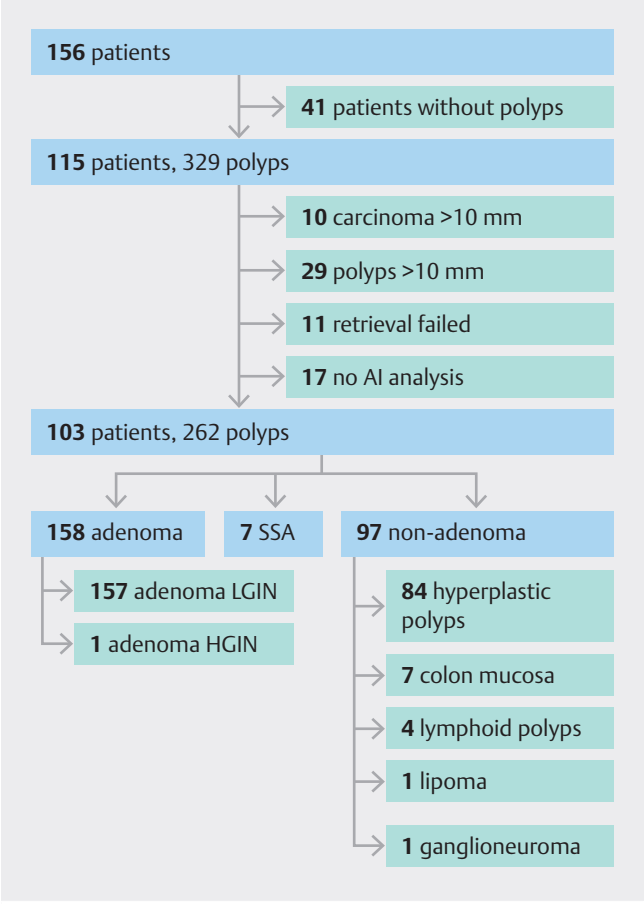
In the primary analysis, SSLs were considered adenomatous. In the secondary analysis, they were considered non-adenomatous and completely excluded in the tertiary analysis. The values were calculated using SPSS Statistics Version 29.0 (IBM Corp., Armonk, New York, United States).

Results for negative predictive values have been given. Due to a dependency on prevalence, which is not consistent across our subgroups, the results are difficult to interpret.

Results

Patients and colonoscopy

One hundred fifty-six patients were included. The quality of bowel preparation was measured using the BBPS. With a mean BBPS of 7.92 (standard deviation [SD] 1.267, median 9) the quality of bowel preparation was very good. Forty-one of 156



▶ **Fig. 4** Overview of patients and polyps.

patients had no polyp. A total of 329 polyps were detected in 115 patients. Of these, 10 (3.0%) were carcinomas, 29 (8.8%) polyps >10 mm. Eleven (3.3%) polyps could not be retrieved after endoscopic resection. Taking into account the exclusion criteria (▶ **Fig. 4**), 103 patients with 262 polyps ≤10 mm were included in the analyses. The median age of included patients was 67.0 ± 12.3 years. 68.9% were female. The most common indications were follow-up colonoscopy (23.3%), screening colonoscopy (14.6%), tumor search (13.6%), planned polypectomy (10.7%) and abdominal pain (10.7%). A detailed overview is shown in ▶ **Table 1**.

AI analysis

In 290 polyps (no carcinomas or polyps >10 mm) the AI analysis was documented. In 17 cases no characterization by the AI system was achieved (no prediction), 14 polyps were assessed as undifferentiated, in three cases there was no stable conclusion of the analysis process. Thus, the optical diagnosis of the AI was made in 94.2% of the polyps.

Polyps

After histological assessment of the 262 polyps, 165 (63.0%) "adenomas" and 97 (37.0%) "non-adenomas" was carried out analogously to the AI characterization (▶ **Fig. 4**). There were 157 adenomas with low-grade intraepithelial neoplasia (LGIN), one adenoma with high-grade intraepithelial neoplasia (HGIN) and seven SSL. Non-adenomatous polyps included 84 hyperplastic polyps, seven pieces of colonic mucosa, four lymphoid polyps, one lipoma and one ganglioneuroma. The average polyp size was 5.5 mm (SD 2.64).

Of the polyps, 59.9% were ≤5 mm. The most common polyp morphology according to the Paris classification was Paris Is (50%) and Paris IIa (43.4%), only 5.8% were classified as Paris Ip and 0.8% as Paris IIb. 30.2% of polyps were removed in the sigmoid colon and rectum. Looking at the colonic section, most polyps in the C. transversum were removed (30.9%). A detailed list can be found in ▶ **Table 1**.

Performance of AI and endoscopist in the entire colon

In the evaluation for the entire colon, the AI showed a sensitivity of 89.7%, a specificity of 75.3%, an accuracy of 84.4% and a negative predictive value (NPV) of 81.1%.

The sensitivity of all endoscopists was 92.1%, specificity 69.1%, accuracy 83.6% and NPV 83.8%. Looking at the group of endoscopists with many years of experience, a sensitivity of 90.8%, a specificity of 71.8%, an accuracy of 83.8% and a NPV of 82.3% was calculated.

For endoscopists with less than 5 years of experience, sensitivity was 95.6%, specificity 61.5%, accuracy 83.1% and NPV 88.9%. The results showed only minor differences, no significant correlation between the different groups could be detected. ▶ **Table 2** gives a detailed overview of the performance of the AI and endoscopists for all included polyps.

Performance of AI and endoscopists by polyp localization and size

Of the 262 enclosed polyps ≤10 mm, 183 (69.8%) were removed proximal to the sigmoid colon. Of these, 137 adenomas (74.8%) were LGIN, one adenoma (0.6%) was HGIN, seven SSL (3.8%) and 38 were non-adenomatous polyps (18.0%). In the sigmoid colon and rectum, 79 polyps ≤10 mm were removed (30.2%). Of these, 20 (25.3%) were adenomas LGIN, 54 (68.4%) were hyperplastic polyps, 4 (5.0%) were colonic mucosa and one (1.2%) was a ganglioneuroma.

Comparing the performance of the AI for rectosigmoidal and proximal polyps ≤10 mm, better results were seen for the more proximal polyps. An accuracy of 85.8% vs 81% for rectosigmoidal polyps was shown. Surprisingly, endoscopists show exactly the opposite. The accuracy for proximal polyps is 82.5% and for rectosigmoidal 86.1%.

Looking at the size of all polyps in two groups (≤5 mm and 6 to 10 mm), a clearly better result can be found with larger polyps. The accuracy of the AI or endoscopist for larger polyp group is 89.5% and 86.7%, respectively. In contrast, an accuracy for small polyps (≤5 mm) of 80.9% and 81.5%, respectively, can be seen. A detailed overview of the performance in these subgroups is shown in ▶ **Table 3**.

▶ **Table 1** Patient and polyp features.

	Endoscopist experience		All
	Experienced	Non- experienced	
Patient features			
Number of patients, n (%)	74 (71.8)	29 (28.2)	103 (100.0)
Age, mean (SD), years	64.9 (12.4)	72.2 (10.3)	67.0 (12.3)
Sex, n (%)			
Female	21 (28.4)	11 (37.9)	31.1
Male	53 (71.6)	18 (62.1)	71 (68.9)
Indication for colonoscopy, n (%)			
Polypectomy	6 (8.1)	5 (17.2)	10.7
Screening	14 (18.9)	1 (3.4)	14.6
Surveillance	22 (29.7)	2 (6.9)	23.3
Tumor screening	9 (12.2)	5 (17.2)	13.6
Follow-up after EMR	3 (4.1)	3 (10.3)	(5.8)
Follow-up after surgery	1 (1.4)	0 (0.0)	(1.0)
IFOBT	2 (2.7)	2 (6.9)	(3.9)
Visible bleeding	2 (2.7)	5 (17.2)	(6.8)
Anemia	0 (0.0)	2 (6.9)	(1.9)
Abdominal pain	9 (12.2)	2 (6.9)	11 (10.7)
Diarrhea, constipation, change of bowel habits	2 (2.7)	1 (3.4)	(2.9)
Before op	2 (2.7)	1 (3.4)	3 (2.9)
Surveillance after radiochemotherapy	1 (1.4)	0 (0.0)	1 (1.0)
Surveillance after recurrent diverticulitis	1 (1.4)	0 (0.0)	1 (1.0)
Polyp features			
Size, n (%)			
1–5 mm	108 (56.5)	49 (69.0)	9.9
6–10 mm	83 (43.5)	22 (31.0)	105 (40.1)
Macroscopic type, n (%)			
Ip	11 (5.9)	4 (5.6)	5.8
Is	84 (44.9)	45 (63.4)	0.0
IIa	90 (48.1)	22 (31.0)	3.4
IIb	2 (1.1)	0 (0.0)	(0.8)
IIc	0 (0.0)	0 (0.0)	(0.0)
III	0 (0.0)	0 (0.0)	0 (0.0)
Location, n (%)			
Cecum	20 (10.5)	10 (14.1)	11.5
Ascending colon	31 (16.2)	7 (9.9)	14.5
Transverse colon	60 (31.4)	21 (29.6)	30.9
Descending colon	24 (12.6)	10 (14.1)	34 (13.0)
Sigmoid colon	49 (25.7)	12 (16.9)	23.3
Rectum	7 (3.7)	11 (15.5)	18 (6.9)
Histology, n (%)			

▶ **Table 1** (Continuation)

	Endoscopist experience		All
	Experienced	Non- experienced	
Adenoma LGIN	112 (58.6)	45 (63.4)	9.9
Adenoma HGIN	1 (0.5)	0 (0.0)	(0.4)
Carcinoma	0 (0.0)	0 (0.0)	(0.0)
SSA	7 (3.7)	0 (0.0)	(2.7)
Hyperplastic polyp	61 (31.9)	23 (32.4)	32.1)
Colon mucosa	5 (2.6)	2 (2.8)	(2.7)
Lymphoid polyp	3 (1.6)	1 (1.4)	(1.5)
Lipoma	1 (0.5)	0 (0.0)	1 (0.4)
Ganglioneuroma	1 (0.5)	0 (0.0)	1 (0.4)

SD, standard deviation; EMR, endoscopic mucosal resection; IFOBT, immunochemical fecal occult blood test; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; SSA, sessile serrated lesion.

▶ **Table 2** Accuracy parameters (95% CIs) for colorectal polyps ≤10 mm.

	AI	Endoscopists (all)	Experienced endoscopists	Non-experienced endoscopists
Accuracy	84.4% (79.4%–88.5%)	83.6% (78.5%–87.9%)	83.8% (77.8%–88.7%)	83.1% (72.3%–91.0%)
Sensitivity	89.7% (84.0%–93.9%)	92.1% (86.9%–95.7%)	90.8% (84.2%–95.3%)	95.6% (84.9%–99.5%)
Specificity	75.3% (65.5%–83.5%)	69.1% (58.9%–78.1%)	71.8% (59.9%–81.9%)	61.5% (40.6%–79.8%)
Positive predictive value*	86.0% (81.3%–89.8%)	83.5% (79.0%–87.3%)	84.5% (78.9–88.8%)	81.1% (72.5%–87.5%)
Negative predictive value*	81.1% (73.0%–87.2%)	83.8% (75.1%–89.8%)	82.3% (72.2–89.2%)	88.9% (66.6%–97.0%)

No statistically significant differences of the different parameters were found, P >0.005.
*Difficult interpretation due to influence of prevalence.

Performance of AI and endoscopist for diminutive rectosigmoid polyps

Fifty-four polyps (20.6%) met the criteria for diminutive recetosigmoid polyps (DRSPs) and are eligible for the leave-in-situ strategy. Histologically, 11 adenomas LGIN (20.4%), 39 hyperplastic polyps (72.2%), and four colonic mucosal pieces (7.4%) were shown.

The AI showed a sensitivity for DRSP of 72.7%, a specificity of 81.4%, an accuracy of 79.6% and an NPV of 92.1%.

The pooled results for all endoscopists were slightly but not significantly better than for AI. A sensitivity of 72.7%, specificity of 88.4%, accuracy of 85.2% and NPV of 92.7% were detected. ▶ **Table 4** gives a detailed overview of the accuracy of DRSP.

Performance AI and endoscopists in diminutive non-rectosigmoid polyps

A total of 103 polyps (39.3%) ≤5 mm were removed proximal to the sigma. These would fall under the resect-and-discard strategy. Seventy-five (72.8%) were adenomas LGIN, three (2.9%) were SSLs, 20 (19.4%) were hyperplastic polyps, two (1.9%) were pieces of colonic mucosa, and three (2.9%) were lymphoid polyps. The sensitivity, specificity, and accuracy of the AI were 87.2%, 64.0%, and 81.6%, respectively.

In contrast, the endoscopists showed sensitivity, specificity, and accuracy of 89.7%, 48.0% and 79.6%, respectively.

Performance of AI and endoscopists for high-confidence polyps

Eighty polyps (30.5% of all polyps, 78.4% of polyps from the introduction of the parameter) were assessed by the endoscopists as a "high-confidence" decision. Looking at the performance of the categorization for "high-confidence" polyps, this is significantly higher than the previously reported values. The sensitivities for the AI and all endoscopists were 93.5% and 97.8%, the specificities were 79.4% and 73.5%, the accuracy was 87.5% and 87.5%, and the NPVs were 90.0% and 96.2%, respectively (▶ **Table 3**).

Histology

In 33 (12.6%) of the 262 polyps, AI or endoscopists predicted adenoma while initial histology stated non-neoplastic tissue. Therefore, the paraffin blocks of these polyps were completely sliced in a second step. The 33 polyps were removed from 26

▶ **Table 3** Accuracy parameters (95% CIs) for colorectal polyps by location and size.

	AI	Endoscopists (all)
Localization of polyps ≤10 mm		
Proximal (cecum – descending c.)		
Accuracy	85.8% (79.9%–91.0%)	82.5% (76.2%–87.7%)
Sensitivity	90.3% (84.3%–94.6%)	93.1% (87.7%–96.6%)
Specificity	68.4% (51.4%–82.5%)	42.1% (26.3%–59.2%)
PPV	91.6% (87.2%–94.6%)	86.0% (82.3%–89.0%)
NPV	65.0% (51.9%–76.2%)	61.5% (44.2%–76.4%)
Distal (sigmoid c. – rectum)		
Accuracy	81.0% (70.6%–89.0%)	86.1% (76.5%–92.8%)
Sensitivity	85.0% (62.1%–96.8%)	85.0% (62.1%–96.8%)
Specificity	79.7% (67.2%–89.0%)	86.4% (75.0%–94.0%)
PPV*	58.6% (45.3%–70.8%)	68.0% (52.1%–80.6%)
NPV*	94.0% (84.6%–97.8%)	94.4% (85.6%–98.0%)
Size		
≤5 mm		
Accuracy	80.9% (73.9%–86.7%)	81.5% (74.6%–87.3%)
Sensitivity	85.4% (76.3%–92.0%)	87.6% (79.0%–93.7%)
Specificity	75.0% (63.0%–84.7%)	73.5% (61.4%–83.5%)
PPV*	81.7% (74.6%–87.2%)	81.3% (74.3%–86.7%)
NPV*	79.7% (70.0%–86.9%)	82.0% (72.0%–89.0%)
>5 mm		
Accuracy	89.5% (82.0%–94.7%)	86.7% (78.6%–92.5%)
Sensitivity	94.7% (87.1%–98.6%)	97.4% (90.8%–99.7%)
Specificity	75.9% (56.5%–89.7%)	58.6% (38.9%–76.5%)
PPV*	91.1% (84.3%–95.2%)	86.0% (80.0%–90.5%)
NPV*	84.6% (67.5%–93.6%)	89.5% (67.7%–97.2%)
Confidence		
High		
Accuracy	87.5% (78.2%–93.8%)	87.5% (78.2%–93.8%)
Sensitivity	93.5% (82.1%–98.6%)	97.8% (88.5%–99.9%)
Specificity	79.4% (62.1%–91.3%)	73.5% (55.6%–87.1%)
PPV*	86.0% (76.0%–92.3%)	83.3% (74.0%–89.8%)
NPV*	90.0% (74.8%–96.5%)	96.2% (78.1%–99.4%)
Low		
Accuracy	81.8% (59.7%–94.8%)	77.3% (54.6%–92.2%)
Sensitivity	92.3% (64.0%–99.8%)	92.3% (64.0%–99.8%)
Specificity	66.7% (29.9%–92.5%)	55.6% (21.2%–86.3%)
PPV*	80.0% (61.0%–91.1%)	75.0% (58.7%–86.4%)
NPV*	85.7% (46.3%–97.7%)	83.3% (41.0%–97.3%)

No statistically significant differences in the different parameters were found, P >0.005.
*Difficult interpretation due to influence of prevalence.
PPV, positive predictive value; NPV, negative predictive value.

▶ **Table 4** Accuracy of optical diagnosis in DRSP and DnRSP.

	AI	Endoscopists (all)
DRSP		
Accuracy	79.6% (66.5%–89.4%)	85.2% (72.9%–93.4%)
Sensitivity	72.7% (39.0%–94,0%)	72.7% (39.0%–94.0%)
Specificity	81.4% (66.6%–91.6%)	88.4% (74.9%–96.1%)
PPV*	50.0% (32.7%–67.3%)	61.5% (39.4%–79.7%)
NPV*	92.1% (81.5%–96.9%)	92.7% (82.8%–97.1%)
DnRSP		
Accuracy	81.6% (72.7%–88.5%)	79.6% (70.5%–86.9%)
Sensitivity	87.2% (77.7%–93.7%)	89.7% (80.8%–95.5%)
Specificity	64.0% (42.5%–82.0%)	48.0% (27.8%–68.7%)
PPV*	88.3% (81.7%–92.8%)	84.3% (78.6%–88.8%)
NPV*	61.5% (45.5%–75.4%)	60.0% (40.9%–76.5%)

No statistically significant differences of the different parameters were found, P >0.005.
*Difficult interpretation due to influence of prevalence.
DRSP, diminutive rectosigmoid polyp; DnRSP, diminutive non-rectosigmoid polyp; PPV, positive predictive value; NPV, negative predictive value.

patients. In 11 of 33 polyps (33.3%), adenomatous components could be found by further histopathological processing, so that the initial diagnosis changed from non-adenomatous to tubular adenoma (LGIN) and in one polyp (3.0%) to SSL.

The number of positive crypts varied from five to 70 (mean 21.42 SD 18.63), the depth of positive crypts varied from 20% to 100% (mean 65, SD 29.69).

The remaining non-adenomatous polyps were 18 (54.5%) hyperplastic polyps, two (6.1%) lymph follicles, one (3.0%) colonic mucosa, and one (3.0%) ganglioneuroma. Overall, the initial diagnosis changed to "adenomatous" in 36.6% of the additionally graded polyps. This change affected 10 (9.7%) patients.

Sessile serrated lesions

Seven (2.7%) SSLs were detected, all proximal to rectum/sigmoid colon. For the primary evaluation, these were considered adenomas. The AI as well as the endoscopists characterized three SSL as adenomas. The four false-negative SSLs correspond to approximately 23.5% of all false negatives in the performance calculation for AI.

A calculation was made for the scenario SSL evaluated as "non-adenomatous". This showed a sensitivity of 91.7%, a specificity of 74.1%, an accuracy of 84.7% and an NPV of 85.6%. These results show hardly any difference from the per protocol calculation.

In a third scenario, the SSLs were completely excluded from the calculation. With a sensitivity of 91.8%, a specificity of 75.3%, an accuracy of 85.5% and a NPV of 84.9%, there are hardly any divergent values for the remaining polyps.

Discussion

With a sensitivity of 89.7%, a specificity of 75.3% and an accuracy of 84.4%, the performance of AI in a routine clinical setting for the complete colon is in the range of the so far determined values. Five real-time studies examining AI performance have been published. The accuracy for colonic polyps was 83.3% to 86.8% [13, 15, 22].

Two of these studies used AI systems with endocytoscopy mode. In one single- center study, NPVs for DRSP of 93.7% to 96.4% were reported [14]. In another multicenter setting, the achieved sensitivity for DRSP was 90.4%, the specificity 85.9% and the NPV 92.8% [12], and thus, significantly higher than in our study.

However, the endocytoscopy technique uses a high magnification with a dedicated endoscope and is only available in highly specialized centers and, therefore, not widely available [24]. The results are also not directly comparable with ours due to the different technology. Furthermore, these studies focused mainly on DRSPs.

Mori et al. [14] already reported significantly decreased accuracy for proximal polyps (DnRSP). This observation was confirmed by Hassan et al [15] with an accuracy of 91.8% for DRSP vs. 77.9% for DnRSP. The endoscopists also showed a significantly lower accuracy for DnRSP in this study (96.1% for DRSP vs 80.3% DnRSP). Another study shows peculiarities in the proximal colon. Rondonotti et al. [13] show approximately the same accuracy for DRSP (87%) and DnRSP (88.4%) for AI-assisted colonoscopy, but a significantly lower specificity (66.7%) and lower NPV (72.4%) for proximal polyps. The reported values for DnRSP are very similar to our results.

An explanation for why there is limited specificity of the AI systems in the proximal colon is still a matter of debate. Hassan

et al [15] suggest that their detected lack of accuracy in the proximal colon could be due to a worse relationship between visible features for optical diagnostic and histological outcomes. In our study, 69.8% of the polyps were located proximal to the sigmoid colon. In reference to the Italian study group [15], this is probably one of the reasons for the lower accuracy with leading lack of specificity in our study. Low specificity, due to the false-positive non-adenomatous polyps would be of minor clinical relevance.

Interestingly, our analyses show better AI performance for proximal vs. rectosigmoidal polyps ≤10 mm. These results conflict with the performance of endoscopists who had better scores for rectosigmoidal polyps. AI performance for DnRSP vs DRSP, on the other hand, shows little difference.

Compared to the results of other study groups that focused on DRSP, the accuracy for DnRSP is the same, but our accuracy for DRSP is lower.

When we compared the accuracy of two polyp size-based groups in the proximal colon (DnRSP 1–5 mm; all polyps 1–10 mm), the accuracy of DnRSP (81.6%) was lower than the accuracy for all removed polyps in the proximal colon (85.8%). This fact of increasing accuracy corresponding with increasing polyp size was probably due to better optical diagnosis by AI with larger-diameter lesions. With regard to polyp size, there were better performance values for polyps >5 mm. This was probably due to the fact that a larger surface area with more mucosal and vascular structure offers better characterization properties. This is probably one of the reasons for better AI performance with polyps ≤10 mm in the proximal colon in our study.

Compared to most study groups who have published real-time studies thus far, endoscopists at our tertiary endoscopy center are not specially trained for optical diagnosis, but map the clinical care situation through a tertiary endoscopy center. Our study was preceded by a training session with a dedicated training module mandatory for all participating endoscopists on the classification systems used. Certification of examiners according to ESGE protocol for optical diagnosis [2] was not available. Comparative values for endoscopist performance have been determined by Pecere et al. [25]. They determined the pooled sensitivity (84.5%) and specificity (83%) of endoscopists in six AI validation studies. Our results for sensitivity (92.1%) significantly exceeded the above values, but our specificity (69.1%) was lower.

In a recently published real-time study from Spain [22] the endoscopists showed very good specificity (95.2%) but also worse sensitivity (75.4%). Hossain et. al [26] recently published an image/video study comparing characterization by AI and endoscopists. Seven general endoscopists evaluated the images. Sensitivity, specificity and accuracy achieved were 55.2%, 61.9%, 57.1% for white light and 76.4%, 59.3% and 71.0% for image-enhanced endoscopy. This shows that optical diagnosis can be a major challenge for non-specially trained endoscopists outside the specialized study centers.

Looking at the performance for “high-confidence” polyps, the diagnostic values were significantly improved for AI and endoscopists with an accuracy of 87.5% for polyps up to 10 mm.

The number of DRSPs in our study was significantly lower than in the majority of the studies previously mentioned. Compared with Hassan et al. [15] and Rondonotti et al. [13], significantly fewer patients with primary screening were included in our study and most of our patients had a surveillance colonoscopy or were admitted to polypectomy of large polyps, so often small lesions were removed in advance.

Resect-and-discard/leave-in-situ strategy

Due to dependency on prevalence, which is not consistent in our analysis, the interpretation of positive predictive value (PPV) and NPV is difficult. We, therefore, do not focus on these figures, especially because there is reduced power due to the small number of DRSPs in our study. Thus, we postulate that the established PIVI-1 criteria [27] could be achieved for a leave-in-situ strategy. An analysis of the agreement about the follow-up interval was not planned in this study because of the focus on the entire colon and polyp size up to 10 mm. The PIVI criteria for a resect-and-discard strategy were not met in our study. Also, the SODA criteria for the resect-and-discard strategy for diminutive colorectal polyps (≥ 80% sensitivity and ≥ 80% specificity) [17] were not met.

Only three of the five published studies [12, 13, 14, 15, 22] were able to reach the thresholds for the resect-and-discard or leave-in-situ strategy. One working group [15] stated that the criteria could be met in a single-center setting. In a subsequent multicenter study [13], the AI algorithm alone barely reached the threshold values. However, this study aimed at the interaction between endoscopists and AI. For AI-assisted colonoscopy, the threshold values could then be met.

Experience of endoscopists

The three endoscopists with less than 5 years of experience in our study achieved very good sensitivity for all polyps ≤10 mm (95.6% vs. 90.8% in long-term experienced patients), but their specificity was lower (61.5% vs. 71.8%). Accuracy was 83.1% vs. 83.8%. A significant difference in the accuracy of AI was not shown here. There was a higher rate of patients with planned polypectomy in the group seen by non-experienced endoscopists. These patients had at least one larger (>15 mm) polyp that was not removed by the referring colleagues in private practice a few weeks before. Normally, referring gastroenterologists resect smaller polyps immediately, so we do not think that there was major bias due to the higher rate of patients planned for polypectomy.

An evaluation of the interaction between AI and endoscopists was not carried out due to our blinded study design.

Rondonotti et al. [13] published data for non-experts at DRSPs. Here, a sensitivity, specificity and accuracy of 81.8%, 83.3% and 82.7% were achieved. As mentioned previously, these results are not directly comparable.

SSL

The SSLs occupy a special situation. Our AI's algorithm has not been trained on SSLs. These are classified as non-neoplastic in the majority of studies [13, 23, 28, 29]. However, because they have potential for malignant degeneration [30] and are particu-

larly relevant in the right-sided colon, they were evaluated as neoplastic in our study. In our work, seven (2.7%) SSLs were found. They were all found proximal to the sigmoid colon. This puts us in the range of the expected occurrence of SSLs [31]. However, some authors also report a higher prevalence (5% to 7%) [32].

In our study, three SSLs were categorized by both AI and endoscopists as "adenoma" and four as "non-adenoma." In a second (SSL as "non-neoplastic") and third (SSL excluded) evaluation scenario, only slightly different values were found.

Rondonotti et al. [13] also performed a secondary analysis after regrouping the SSLs as neoplastic, which resulted in hardly any difference in values. Given the small number of SSLs, only limited conclusions can be drawn. Especially with regard to the relevance in the right colon, further technical developments and studies in this field are necessary.

Histology

The histological assessment was carried out by two experienced pathologists. The standard histopathological work up method is cutting six to eight sections of the paraffin block per polyp. In the course of the study, we discovered some discrepant findings, so together with our colleagues from the Pathology Department, we developed a method of further investigating the polyps.

The remaining paraffin blocks were completely cut in multiple serial sections. It turned out that in the deeper incisions, 36.6% of the polyps had an adenoma that could not be detected in the initial sections. This statement must make you sit up and take notice!

Some working groups have already published similar results. These studies [33, 34, 35, 36] showed 21.5% to 36.9% adenomas in deeper sections with initially negative biopsies. Currently, there are no standardized requirements for histological sample processing. Especially with polyps that are characterized by AI and endoscopists as adenoma, in the case of negative histology, cutting deeper sections is essential to making a correct diagnosis and to recommending the correct follow-up interval for a patient.

In our study, the change in histology affected 10 patients. Further studies and the definition of universal standards in this area are urgently needed.

The strengths of the study include the real-time character of the data collection and its use in the daily clinical routine at a large tertiary hospital. In contrast to many other image- or video-based studies, AI was performed live in everyday clinical practice. This did not result in a selection bias due to high quality images and the results were much more meaningful. Another strength was the blinded setting, because it prevented AI from influencing the decision-making of endoscopists.

Our study also has some limitations. First, it was single-center and non-randomized. Second, two-thirds of endoscopists were highly experienced but are not certified experts in the field of optical diagnosis. Third, the assessment of polyp size was subject to interindividual differences. Thus, no absolutely exact values can be specified here. Currently, there is no widely used technical tool to objectively measure polyp size. Thus, all previous comparable studies are subject to this bias. Fourth, an early software version of the AI algorithm was used, which will be improved over time. Fifth, only the two latest series of Olympus endoscopes were used. Thus, no statement can be made about how well the AI algorithm performs with the older generation of Olympus endoscopes, which are still very common in our region, or with endoscopes from other manufacturers.

Conclusions

In summary, the commercially available AI system used in our study showed good accuracy in classifying colorectal polyps, which was comparable with that for experienced endoscopists. However, the SODA and PIVI thresholds for a resect-and-discard or leave-in-situ strategy were not met.

Conflict of Interest

Oliver Pech has received speaker honorarium from Medtronic Norgine, Olympus and Fujifilm. The remaining authors declare that they have no conflict of interest.

Clinical trial

Trial registry: German Clinical Trials Register (https://drks-neu.uniklinik-freiburg.de/)
Registration number (trial ID): DRKS00028894
Type of Study: prospective, non-randomized study

References

[1] Bretthauer M, Løberg M, Wieszczy P et al. Effect of colonoscopy screening on risks of colorectal cancer and related death. N Engl J Med 2022; 387: 1547–1556 doi:10.1056/NEJMoa2208375

[2] Dekker E, Houwen BBSL, Puig I et al. Curriculum for optical diagnosis training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2020; 52: 899–923

[3] Vleugels JLA, Dijkgraaf MGW, Hazewinkel Y et al. Effects of training and feedback on accuracy of predicting rectosigmoid neoplastic lesions and selection of surveillance intervals by endoscopists performing optical diagnosis of diminutive polyps. Gastroenterology 2018; 154: 1682–1693.e1

[4] Vleugels JLA, Hazewinkel Y, Dijkgraaf MGW et al. Optical diagnosis expanded to small polyps: post-hoc analysis of diagnostic performance in a prospective multicenter study. Endoscopy 2019; 51: 244–252 doi:10.1055/a-0759-1605

[5] Ladabaum U, Fioritto A, Mitani A et al. Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions. Gastroenterology 2013; 144: 81–91

[6] Rees CJ, Rajasekhar PT, Wilson A et al. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. Gut 2017; 66: 887–895 doi:10.1136/gutjnl-2015-310584

[7] Hassan C, Spadaccini M, Iannone A et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021; 93: 77–85.e6

[8] Zippelius C, Alqahtani SA, Schedel J et al. Diagnostic accuracy of a novel artificial intelligence system for adenoma detection in daily practice: a prospective nonrandomized comparative study. Endoscopy 2022; 54: 465–472 doi:10.1055/a-1556-5984

[9] Barua I, Vinsard DG, Jodal HC et al. Artificial intelligence for polyp detection during colonoscopy: a systemati review and meta-analysis. Endoscopy 2021; 53: 277–284

[10] Kader R, Hadjinicolaou AV, Georgiades F et al. Optical diagnosis of colorectal polyps using convolutional neural networks. World J Gastroenterol 2021; 27: 5908 doi:10.3748/wjg.v27.i35.5908

[11] Parsa N, Byrne MF. Artificial intelligence for identification and characterization of colonic polyps. Ther Adv Gastrointest Endosc 2021; 14: 1–12

[12] Barua I, Wieszczy P, Kudo S et al. Real-Time Artificial Intelligence-Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy. NEJM Evid 2022; 1: doi:10.1056/EVIDoa2200003

[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 doi:10.1055/a-1852-0330

[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: 357–366

[15] Hassan C, Balsamo G, Lorenzetti R et al. Artificial intelligence allows leaving-in-situ colorectal polyps. Clin Gastroenterol Hepatol 2022; 20: 2505–2513

[16] Rex DK, Kahi C, O'Brien M et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2011; 73: 419–422 doi:10.1016/j.gie.2011.01.023

[17] Houwen BBSL, Hassan C, Coupé VMH et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2022; 54: 88–99

[18] Mori Y, Kudo SE, East JE et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video). Gastrointest Endosc 2020; 92: 905–911.e1

[19] Hassan C, Pickhardt PJ, Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. Clin Gastroenterol Hepatol 2010; 8: 865–869 doi:10.1016/j.cgh.2010.05.018

[20] Vleugels J, Greuter M, Hazewinkel Y et al. Implementation of an optical diagnosis strategy saves costs and does not impair clinical outcomes of a fecal immunochemical test-based colorectal cancer screening program. Endosc Int open 2017; 5: E1197–E1207

[21] Messmann H, Bisschops R, Antonelli G et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2022; 54: 1211–1231 doi:10.1055/a-1950-5694

[22] García-Rodríguez A, Tudela Y, Córdova H et al. In vivo computer-aided diagnosis of colorectal polyps using white light endoscopy. Endosc Int Open 2022; 10: E1201 doi:10.1055/a-1881-3178

[23] Weigt J, Repici A, Antonelli G et al. Performance of a new integrated computer-assisted system (CADe/CADx) for detection and characterization of colorectal neoplasia. Endoscopy 2022; 54: 180–184 doi:10.1055/a-1372-0419

[24] Milluzzo SM, Cesaro P, Grazioli LM et al. Artificial intelligence in lower gastrointestinal endoscopy: the current status and future perspective. Clin Endosc 2021; 54: 329 doi:10.5946/ce.2020.082

[25] Pecere S, Antonelli G, Dinis-Ribeiro M et al. Endoscopists performance in optical diagnosis of colorectal polyps in artificial intelligence studies. United Eur Gastroenterol J 2022; 10: 817 doi:10.1002/ueg2.12285

[26] Hossain E, Abdelrahim M, Tanasescu A et al. Performance of a novel computer-aided diagnosis system in the characterization of colorectal polyps, and its role in meeting Preservation and Incorporation of Valuable Endoscopic Innovations standards set by the American Society of Gastrointestinal Endoscopy. DEN open 2022; 3: e178 doi:10.1002/deo2.178

[27] Abu Dayyeh BK, Thosani N, Konda V et al. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2015; 81: 502.e1–502.e16

[28] 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: 3976–3983

[29] 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: 2237–2245

[30] East JE, Vieth M, Rex DK. Serrated lesions in colorectal cancer screening: detection, resection, pathology and surveillance. Gut 2015; 64: 991–1000 doi:10.1136/gutjnl-2014-309041

[31] Huang J, Chan PSF, Pang TWY et al. Rate of detection of serrated lesions at colonoscopy in an average-risk population: a meta-analysis of 129,001 individuals. Endosc Int open 2021; 9: E472–E481 doi:10.1055/a-1333-1776

[32] Fan C, Younis A, Bookhout CE et al. Management of serrated polyps of the colon. Curr Treat Options Gastroenterol 2018; 16: 182–202 doi:10.1007/s11938-018-0176-0

[33] Schick BA, McLean CA, Driman DK. Negative colorectal polyp biopsies: the utility of cutting deeper levels. Virchows Arch 2015; 467: 635–640 doi:10.1007/s00428-015-1866-9

[34] Nash JW, Niemann T, Marsh WL et al. To step or not to step: an approach to clinically diagnosed polyps with no initial pathologic finding. Am J Clin Pathol 2002; 117: 419–423 doi:10.1309/27NQ-A1R2-DQ41-20Y7

[35] Nielsen JA, Lager DJ, Lewin M et al. Incidence of diagnostic change in colorectal polyp specimens after deeper sectioning at 2 different laboratories staffed by the same pathologists. Am J Clin Pathol 2013; 140: 231–237 doi:10.1309/AJCPPXU66QZAZNYN

[36] Warnecke M, Engel UH, Bernstein I et al. Biopsies of colorectal clinical polyps–emergence of diagnostic information on deeper levels. Pathol Res Pract 2009; 205: 231–240 doi:10.1016/j.prp.2008.08.003