






# Optical diagnosis of colorectal polyp images using a newly developed computer-aided diagnosis system (CADx) compared with intuitive optical diagnosis

## Authors

Quirine E. W. van der Zander<sup>1,2,\*</sup> , Ramon M. Schreuder<sup>3,\*</sup> , Roger Fonollà<sup>4</sup>, Thom Scheeve<sup>4</sup> , Fons van der Sommen<sup>4</sup>, Bjorn Winkens<sup>5</sup>, Patrick Aepfl<sup>6</sup>, Bu'Hussain Hayee<sup>7</sup> , Andreas B. Pischel<sup>8</sup>, Milan Stefanovic<sup>9</sup>, Sharmila Subramaniam<sup>10</sup> , Pradeep Bhandari<sup>10</sup>, Peter H. N. de With<sup>4</sup>, Ad A. M. Masclee<sup>1</sup>, Erik J. Schoon<sup>2,3</sup>

## Institutions

- 1 Division of Gastroenterology and Hepatology, Maastricht University Medical Center + Maastricht, the Netherlands
- 2 GROW, School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands
- 3 Division of Gastroenterology and Hepatology, Catharina Hospital Eindhoven, Eindhoven, the Netherlands
- 4 Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands
- 5 Department of Methodology and Statistics, CAPHRI, Care and Public Health Research Institute, Maastricht University, Maastricht, the Netherlands
- 6 Division of Gastroenterology and Hepatology, Luzerner Kantonsspital, Lucerne, Switzerland
- 7 Division of Gastroenterology and Hepatology, King's College Hospital, London, United Kingdom
- 8 Division of Gastroenterology and Hepatology, University Hospital Gothenburg, Gothenburg, Sweden
- 9 Division of Gastroenterology and Hepatology, Diagnostični Center Bled, Ljubljana, Slovenia
- 10 Division of Gastroenterology and Hepatology, Queen Alexandra Hospital, Portsmouth, United Kingdom

submitted 10.6.2020

accepted after revision 23.12.2020

published online 23.12.2020

## Bibliography

Endoscopy 2021; 53: 1219–1226

DOI 10.1055/a-1343-1597

ISSN 0013-726X

© 2020, Thieme. All rights reserved.

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



Fig. 1s, Tables 1s, 2s

Supplementary material is available under

<https://doi.org/10.1055/a-1343-1597>



Scan this QR-Code for the author commentary.



## Corresponding author

Quirine E. W. van der Zander, MD, Division of Gastroenterology and Hepatology, Maastricht University, P. Debyelaan 25, 6229 HX Maastricht, The Netherlands  
q.vanderzander@maastrichtuniversity.nl

## ABSTRACT

**Background** Optical diagnosis of colorectal polyps remains challenging. Image-enhancement techniques such as narrow-band imaging and blue-light imaging (BLI) can improve optical diagnosis. We developed and prospectively validated a computer-aided diagnosis system (CADx) using high-definition white-light (HDWL) and BLI images, and compared the system with the optical diagnosis of expert and novice endoscopists.

**Methods** CADx characterized colorectal polyps by exploiting artificial neural networks. Six experts and 13 novices optically diagnosed 60 colorectal polyps based on intuition. After 4 weeks, the same set of images was permuted and optically diagnosed using the BLI Adenoma Serrated International Classification (BASIC).

**Results** CADx had a diagnostic accuracy of 88.3% using HDWL images and 86.7% using BLI images. The overall diagnostic accuracy combining HDWL and BLI (multimodal imaging) was 95.0%, which was significantly higher than that of experts (81.7%,  $P=0.03$ ) and novices (66.7%,  $P<0.001$ ). Sensitivity was also higher for CADx (95.6% vs. 61.1% and 55.4%), whereas specificity was higher for experts compared with CADx and novices (95.6% vs. 93.3% and 93.2%). For endoscopists, diagnostic accuracy did not increase when using BASIC, either for experts (intuition

\* These authors contributed equally to this manuscript.

79.5% vs. BASIC 81.7%,  $P=0.14$ ) or for novices (intuition 66.7% vs. BASIC 66.5%,  $P=0.95$ ).

**Conclusion** CADx had a significantly higher diagnostic accuracy than experts and novices for the optical diagnosis of colorectal polyps. Multimodal imaging, incorporating both

HDWL and BLI, improved the diagnostic accuracy of CADx. BASIC did not increase the diagnostic accuracy of endoscopists compared with intuitive optical diagnosis.

## Introduction

Optical diagnosis of colorectal polyps, the *in vivo* characterization of the histology by endoscopists [1], is of increasing interest for clinical endoscopy practice. Correct characterization of colorectal polyps with high levels of confidence is of utmost importance for adequate and cost-effective treatment strategies. Diminutive ( $\leq 5$  mm) hyperplastic polyps in the rectosigmoid can be left in place according to the “diagnose-and-leave” strategy. Diminutive adenomatous polyps can be endoscopically resected but not sent for histological evaluation, in what is known as the “resect-and-discard” strategy. To adopt these strategies in clinical practice, the Preservation and Incorporation of Valuable Endoscopic Innovations (PIVI) criteria are proposed [2]. These PIVI criteria demand a negative predictive value (NPV) of  $\geq 90\%$  for the *in vivo* diagnosis of diminutive adenomatous polyps and an agreement of  $\geq 90\%$  between optical and histological diagnosis in determining surveillance intervals. Although colorectal polyps contain valuable visual clues about their histology, current optical diagnostic procedures are limited in their predictive power and are subject to considerable interobserver variability. Recent studies have also shown that while the PIVI criteria are met in highly selected groups of expert endoscopists [3], the same is not true in community endoscopy practices [4,5].

To improve the optical diagnosis of colorectal polyps, image-enhancement technologies, which optimize the visualization of superficial vascular and mucosal patterns, have been developed, such as narrow-band imaging (NBI; Olympus, Tokyo, Japan), I-SCAN (Pentax, Tokyo, Japan), and blue-light imaging (BLI; Fujifilm, Tokyo, Japan). The BLI technology optimizes the visualization of superficial vascular and mucosal patterns by emitting light with a short wavelength (410 nm), which is then selectively absorbed by hemoglobin. White light is used for contrast enhancement (450 nm) [6]. Based on these technologies, different classifications have been developed to increase the viability of optical diagnosis: the NBI International Colorectal Endoscopic (NICE) classification based on NBI [7] and the BLI Adenoma Serrated International Classification (BASIC) [8] based on BLI. BASIC differentiates hyperplastic polyps from adenomas, sessile serrated lesions (SSLs), and adenocarcinomas (CRCs) using surface, pit pattern, and vessel characteristics. Compared with NBI and NICE, optical diagnosis using BLI and BASIC shows similar rates of diagnostic accuracy [6,9–11].

The most recent developments proposed to improve optical diagnostics are computer techniques based on artificial intelligence (AI). Studies investigating AI have mainly focused on (NBI) magnification images or endocytoscopy [12–15] systems

that are not readily available in current clinical practice. The diagnostic accuracies of the developed systems range from 90.0% to 94.9%. Although these results are promising, data on the performance of computer-aided diagnosis systems using BLI are lacking. This also applies to multimodal imaging, which combines the different advantages of multiple imaging methods in order to improve the accuracy of imaging diagnostics [16].

The aims of the current study were i) to develop a computer-aided diagnosis system (CADx) for colorectal polyp characterization and to investigate its diagnostic performance using high-definition white-light (HDWL) imaging, BL imaging, and multimodal imaging (both HDWL and BLI), and ii) to compare CADx with the diagnostic performance of optical diagnosis by expert and novice endoscopists using their intuition and BASIC.

## Methods

This prospective, noninterventional study was conducted at the Maastricht University Medical Center+ and Catharina Hospital Eindhoven, the Netherlands, from November 2019 to March 2020. The department of Electrical Engineering of the Eindhoven University of Technology was responsible for CADx development. The study was conducted in accordance with the declaration of Helsinki [17] and the General Data Protection Regulation [18]. The Medical Ethical Review Committee of Maastricht UMC+ approved the study (METC2019–1231).

### Image database

Two image databases were created for this study: a training database and a testing database. The training database consisted of 2449 polyp images from 398 unique colorectal polyps. The 2449 images included 344 benign images (including 76 hyperplastic polyps) and 2105 (pre)malignant images (including 291 adenomas, 24 SSLs, and 7 T1 CRCs). Of the 398 colorectal polyps, 45.8% were diminutive ( $\leq 5$  mm), 18.0% were small (6–9 mm), and 36.2% were large ( $\geq 10$  mm). The SSLs were categorized as premalignant because of their malignant potential [19]. In the training database, each colorectal polyp was represented by one HDWL image and one or more optically enhanced images (either I-SCAN or BLI) without magnification. All training images were retrospectively obtained.

The testing database consisted of 60 prospectively obtained, unique colorectal polyps represented by one HDWL and one BLI image (without magnification). These colorectal polyps were different from those included in the training database. Colorectal polyps were selected by one endoscopist (R.M.S.) based on good image quality and availability of the corresponding histology results (gold standard).

The images in both databases were obtained from national Dutch screening and surveillance colonoscopies in regular care and were fully anonymized. A one-on-one link between polyp images and histology results was confirmed by means of corresponding polyp number and histology number, corresponding size, and corresponding morphology. If any ambiguity existed, the polyp was excluded. Colorectal polyps were histologically characterized as hyperplastic polyp, adenoma, SSL, or T1 adenocarcinoma by expert pathologists according to the revised Vienna classification [20]. The pathologists were unaware of the study protocol. The histology distribution of colorectal polyps in the testing database corresponded to the natural occurrence in the screening population [21]. The endoscopists were blinded regarding the histological distribution and diagnosis.

### Computer-aided diagnosis system

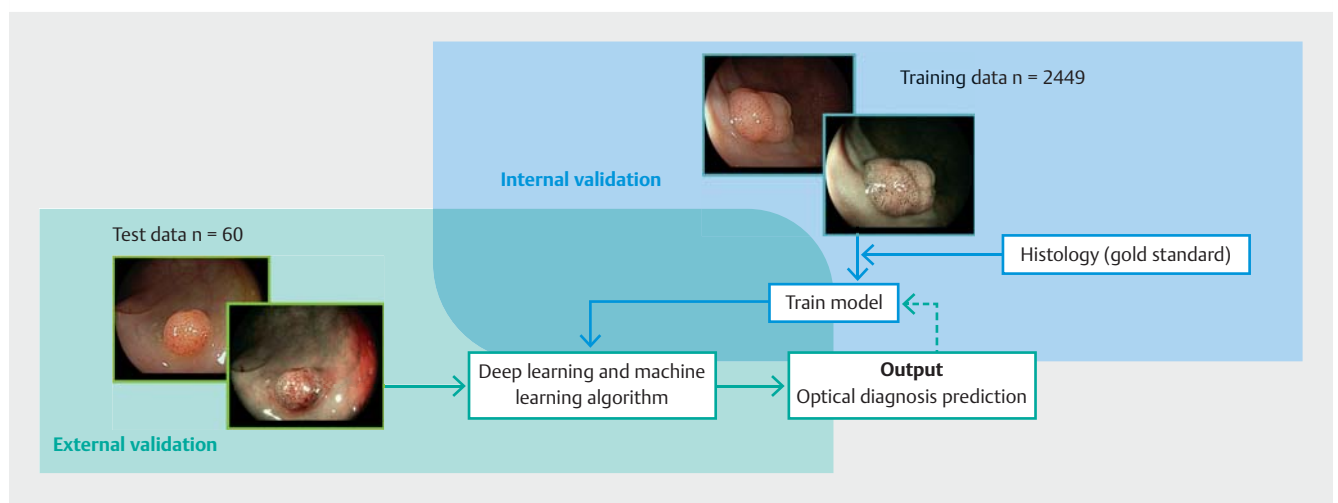
CADx was developed to differentiate between benign and (pre) malignant colorectal polyps, using artificial neural networks and the training database, with histology as the gold standard. The artificial neural networks automatically determined the most discriminative features of the colorectal polyps in the training set. A subset of 20% of the training data was used as validation to monitor the progress of the algorithm. After training, the deep learning algorithm was evaluated using the testing database for external validation and verification of model performance (► Fig. 1). As outputs, CADx generated a region of interest represented by a heatmap (generated using Grad-CAM [22]) and a corresponding probability. These measures were generated for each modality. A weighted mean probability was calculated, using both HDWL and BLI modalities, as a final prediction for each colorectal polyp. A cutoff value of 0.35 was used as a boundary decision in order to differentiate between the presence or absence of neoplasia.

### Algorithm details

The architecture used in our CADx was EfficientNet [23]. This family of models achieved state-of-the-art accuracy on the open dataset ImageNet. EfficientNet models employ a simple but powerful concept that models should not only be scaled in depth, but also in width and resolution. For our CADx we employed the variant B4, which is a neural network architecture with 19 million parameters. The network was pretrained with ImageNet and subsequently trained on the training dataset using Stochastic Gradient Descent with momentum 0.9 and a batch size of 8. We chose to use an exponential learning rate, with hard restarts at every two epochs, ranging from 0.01 to 0.004. We trained a unique model containing both imaging modalities, ensuring that the information learned was domain invariant. The images were resized to 299×299×3 pixels, and the model was trained until convergence on the validation set (the subset of 20% of the training data).

### Optical diagnosis by endoscopists

Colorectal polyps from the testing database were optically diagnosed by six expert endoscopists from the international BLI expert group. All experts were experienced in using BLI and BASIC, and had performed >2000 colonoscopies. In addition, 13 Dutch novices with limited colonoscopy experience (four with 100–200, five with 200–300, and 4 with 300–400 colonoscopies) and no prior experience of using BLI or BASIC optically diagnosed the colorectal polyps. First, endoscopists optically diagnosed colorectal polyps based on intuition, with a time limit of 30 seconds. Intuition was defined as “the optical diagnosis that comes first to your mind” based on knowledge and experience without systematically going through classification schemes [24]. After a “washout period” of 4 weeks in order to minimize recall bias, both the experts and novices were additionally trained in BLI and BASIC through a previously validated



► Fig. 1 Overview of the development of the computer-aided diagnosis system. The training image database, consisting of 2449 colorectal polyps (344 benign and 2105 [pre]malignant), was used for internal validation. The testing image database, consisting of 60 colorectal polyps (15 benign and 45 [pre]malignant) was used for external validation.

training module [25] by two expert endoscopists (E.J.S. and R. M.S.). The training module consisted of an introduction into optical diagnosis, the PIVI criteria, and the BLI technology, an elaboration on the individual BASIC descriptors, and examples with direct feedback. Afterwards, the same set of colorectal polyps was permuted and optically diagnosed using BASIC. Endoscopists had to complete each individual BASIC descriptor and a final diagnosis for each colorectal polyp: hyperplastic polyp, adenoma, SSL, or CRC. In addition, endoscopists reported their level of confidence (high [ $\geq 90\%$ ] vs. low) per colorectal polyp. All endoscopists participated in both diagnostic phases (intuition and BASIC). This cross-sectional design was chosen because of the high reported interobserver variability in performance of optical diagnosis among endoscopists [3, 7]. The optical diagnoses were made using an online portal (see **Fig. 1s** in the online-only Supplementary material).

## Outcomes

The diagnostic performance of CADx was compared with the diagnostic performance of expert and novice endoscopists. Optical diagnoses of the endoscopists were dichotomized into benign vs. (pre)malignant to make an accurate comparison with CADx. The diagnostic performances were investigated in terms of diagnostic accuracy (defined as the percentage of correctly characterized colorectal polyps), sensitivity, specificity, negative and positive predictive values, and area under the receiver operating characteristic curve (AUC). AUC values  $>0.91$  were interpreted as excellent,  $0.81$ – $0.90$  as good, and  $0.71$ – $0.80$  as fair. The diagnostic performance of experts and novices for optical diagnosis based on intuition and BASIC were compared. In addition, a subgroup analysis for high-confidence diagnoses was performed. End points reflecting computational time of CADx and interobserver agreement (Fleiss' kappa) were also evaluated. Fleiss' kappa values of  $0.81$ – $1.00$  were interpreted as very good,  $0.61$ – $0.80$  as good,  $0.41$ – $0.60$  as moderate,  $0.21$ – $0.40$  as fair, and  $<0.20$  as poor [26].

## Statistical analyses and sample size

A sample size calculation based on a desired CADx accuracy was not possible owing to a lack of data. Because the study design involved experts and novices characterizing colorectal polyps in two phases, the McNemar's test for paired proportions was used for the power calculation. Discordant pairs were set at 15%. Assuming a difference in diagnostic accuracy of 10% and using a power of 80% with a 5% significance level, 15 discordant pairs and 101 observations per phase were needed. Correcting for the interobserver difference, with an intraclass correlation of 0.05 [27] and using 60 colorectal polyps, 399 observations were needed per phase [28, 29]. We recruited 19 endoscopists, resulting in 900 observations without missing data.

Descriptive statistics are presented as mean and standard deviation (SD) or as number of optical diagnosis and percent, with or without 95% confidence interval (CI). Differences between intuition and BASIC were compared using paired samples *t* test or Wilcoxon signed-rank test where appropriate. Interobserver agreement between endoscopists was calculated using Fleiss' kappa statistics. Sensitivity analyses were per-

**► Table 1** Characteristics of the colorectal polyps used in the testing database.

Polyp characteristics	n = 60
Size, n (%)	
▪ $\leq 5$ mm	33 (55.0)
▪ 6–9 mm	14 (23.3)
▪ $\geq 10$ mm	13 (21.7)
Location, n (%)	
▪ Cecum	2 (3.3)
▪ Ascending colon	9 (15.0)
▪ Transverse colon	11 (18.3)
▪ Descending colon	7 (11.7)
▪ Sigmoid	15 (25.0)
▪ Rectum	16 (26.7)
Histology, n (%)	
▪ Hyperplastic polyp	15 (25.0)
▪ Adenoma	39 (65.0)
▪ Sessile serrated lesion	4 (6.7)
▪ Adenocarcinoma	2 (3.3)
Morphology, n (%)	
▪ Sessile	41 (68.3)
▪ Flat elevated	19 (31.7)

formed to account for potential correlation for patients with more than one colorectal polyp ( $n=6$ ). In all cases, the first colorectal polyp was chosen and the duplicate was removed. Repeating the analyses with 54 colorectal polyps yielded no differences for the conclusion. Two-sided *P* values of  $<0.05$  were considered statistically significant.

All statistical analyses were performed with IBM SPSS Statistics for Windows version 25 (IBM Corp., Armonk, New York, USA).

## Results

### Colorectal polyps

Overall, 60 colorectal polyps from 54 patients were included in the analyses. These consisted of 15 hyperplastic polyps (25.0%), 39 adenomas (65.0%), 4 SSLs (6.7%), and 2 CRCs (3.3%) (**► Table 1**). Of the 60 colorectal polyps, 33 were diminutive ( $\leq 5$  mm), 14 were small (6–9 mm), and 13 were large ( $\geq 10$  mm).

### Computer-aided diagnosis system

CADx showed a diagnostic accuracy of 88.3% for HDWL images, 86.7% for BLI images, and peaked at 95.0% when both HDWL and BLI images (multimodal imaging) were used (**► Table 2**). The overall sensitivity, specificity, PPV, and NPV of CADx were



► **Table 2** Diagnostic results of the computer-aided diagnosis system based on high-definition white-light (HDWL) images, blue-light imaging (BLI) images, and multimodal imaging (combining HDWL and BLI) images for benign (hyperplastic) vs. (pre)malignant (adenomas, sessile serrated lesions, and adenocarcinomas) colorectal polyps.

	HDWL, % [n/N] (95%CI)	BLI, % [n/N] (95%CI)	Multimodal imaging, % [n/N] (95%CI)
Diagnostic accuracy	88.3 [53/60] (78.3–95.0)	86.7 [52/60] (76.7–93.3)	95.0 [57/60] (86.7–98.3)
Sensitivity	84.4 [38/45] (71.1–93.3)	84.4 [38/45] (71.1–93.3)	95.6 [43/45] (86.7–100)
Specificity	100.0 [15/15] (100–100)	93.3 [14/15] (66.7–100)	93.3 [14/15] (66.7–100)
PPV	100.0 [38/38] (100–100)	97.4 [38/39] (84.6–100)	97.7 [43/44] (86.4–100)
NPV	68.2 [15/22] (50.0–86.4)	66.7 [14/21] (47.6–85.7)	87.5 [14/16] (68.8–100)
AUC	0.92	0.89	0.94

HDWL, high-definition white light; BLI, blue-light imaging; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve.

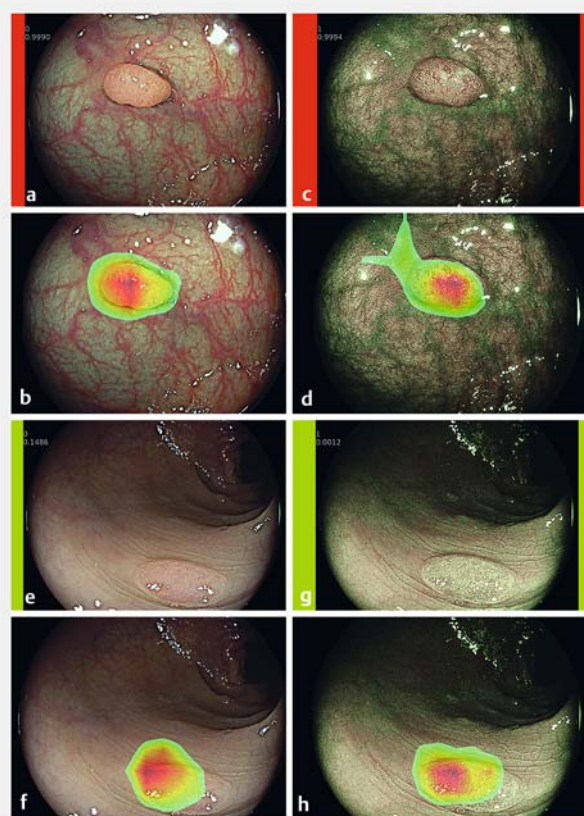
95.6%, 93.3%, 97.7%, and 87.5%, respectively. CADx demonstrated an overall AUC of 0.94. CADx misclassified three colorectal polyps when the imaging modalities were combined: one hyperplastic polyp was misclassified as a premalignant polyp and two adenomas were misclassified as benign polyps. Endoscopists also incorrectly characterized these three colorectal polyps when using intuition (in 28 out of a total of 57 optical diagnoses; 49.1%) and when using BASIC (27/57 optical diagnoses; 47.4%). The CRC cases were correctly recognized by CADx. CADx misclassified seven polyps when using HDWL alone and eight polyps when using BLI alone (**Table 1s**). Re-evaluation by an expert pathologist, who was blinded to the initial histology, did not change the histological decision.

The mean computation time per image was 0.0258 seconds (SD 0.0148 seconds). For each prediction made by CADx, an image with a corresponding region of interest depicted with a heatmap was generated (► **Fig. 2**) to ensure that the network was not exploiting potential bias in the images. This heatmap image revealed that the network was indeed focusing on the colorectal polyp area in all cases.

### Intuition vs. BASIC

► **Table 3** shows the diagnostic performance of experts and novices for optical diagnosis based on intuition and BASIC. The diagnostic accuracy of experts increased nonsignificantly for optical diagnosis made with BASIC (intuition 79.5% [95%CI 72.6–86.4] vs. BASIC 81.7% [95%CI 77.3–86.1],  $P = 0.14$ ). Sensitivity increased from 50.0% [95%CI 33.0–67.0] when using intuition to 61.1% [95%CI 55.9–66.3] when using BASIC ( $P = 0.22$ ). These increases in diagnostic accuracy and sensitivity were accompanied by a nonsignificant decrease in specificity (intuition 95.6% [95%CI 90.5–100] vs. BASIC 94.1% [95%CI 92.2–96.0],  $P = 0.54$ ).

For novices, the diagnostic accuracy did not improve when using BASIC (intuition 66.7% [95%CI 61.6–71.8] vs. BASIC 66.5% [95%CI 60.9–72.1],  $P = 0.95$ ). Sensitivity increased nonsignificantly from 46.2% (95%CI 36.1–56.3) when using intuition to 55.4% (95%CI 46.5–64.3) when using BASIC ( $P = 0.09$ ). Again, this was accompanied by a nonsignificant decrease in specificity (intuition 93.2% [95%CI 90.8–95.6] vs. BASIC 92.1% [95%CI 88.4–95.8];  $P = 0.55$ ). The overall sensitivity (all endos-



► **Fig. 2** Endoscopic images of an adenoma (a–d) and a hyperplastic polyp (e–h) in different imaging modalities. Adenoma: a High-definition white light (HDWL) image. b HDWL with corresponding heatmap. c Blue-light imaging (BLI) image. d BLI with corresponding heatmap. The prediction of the computer-aided diagnosis system (CADx) can be seen from the cutoff value being  $\geq 0.35$  and the red color surrounding the colorectal polyp image. Hyperplastic polyp: e HDWL. f HDWL with corresponding heatmap. g BLI. h BLI with corresponding heatmap. The prediction of CADx can be seen from the cutoff value being  $< 0.35$  and the green color surrounding the colorectal polyp image.

**► Table 3** Diagnostic results based on intuition and on the BLI Adenoma Serrated International Classification for benign (hyperplastic) vs. (pre)malignant (adenoma, sessile serrated lesion, and adenocarcinoma) colorectal polyps for both experts and novices.

Group	Intuition, mean % <sup>1</sup> (95%CI)	BASIC, mean % <sup>1</sup> (95%CI)	P value <sup>2</sup>
Experts (n = 6)			
▪ Accuracy	79.5 (72.6–86.4)	81.7 (77.3–86.1)	0.14
▪ Sensitivity	50.0 (33.0–67.0)	61.1 (55.9–66.3)	0.22
▪ Specificity	95.6 (90.5–100)	94.1 (92.2–96.0)	0.54
▪ PPV	83.7 (67.9–99.5)	77.8 (73.0–82.6)	0.44
▪ NPV	85.4 (81.5–89.3)	87.9 (86.5–89.3)	0.21
Novices (n = 13)			
▪ Accuracy	66.7 (61.6–71.8)	66.5 (60.9–72.1)	0.95
▪ Sensitivity	46.2 (36.1–56.3)	55.4 (46.5–64.3)	0.09
▪ Specificity	93.2 (90.8–95.6)	92.1 (88.4–95.8)	0.55
▪ PPV	69.3 (60.9–77.7)	71.9 (62.8–81.0)	0.60
▪ NPV	84.0 (81.5–86.5)	86.2 (83.8–88.6)	0.16
Overall (n = 19)			
▪ Accuracy	70.7 (66.0–75.4)	71.3 (66.1–76.5)	0.69
▪ Sensitivity	47.4 (39.6–55.2)	57.2 (51.1–63.3)	0.03
▪ Specificity	93.9 (91.8–96.0)	92.7 (90.2–95.2)	0.38
▪ PPV	73.9 (66.4–81.4)	73.8 (67.6–80.0)	0.99
▪ NPV	84.4 (82.5–86.3)	86.7 (85.0–88.4)	0.05

BASIC, BLI Adenoma Serrated International Classification; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; BLI, blue-light imaging.

<sup>1</sup> Mean calculated for the number of experts, novices, and overall observers, respectively.

<sup>2</sup> P value based on paired sample t test for total group of experts, novices, and overall observers.

copists) increased significantly from 47.4% (95%CI 39.6–55.2) when using intuition to 57.2% (95%CI 51.1–63.3) when using BASIC ( $P=0.03$ ).

The proportion of optical diagnoses made with high confidence increased when using BASIC, from 82.5% to 85.0% for experts ( $P=0.30$ ), and from 52.1% to 53.7% for novices ( $P=0.49$ ). When only high-confidence diagnoses were considered, the diagnostic accuracy did not show a significant change when using BASIC (85.6% [SD 12.0]) compared with using intuition (84.1% [SD 17.1],  $P=0.55$ ) (Table 2s). The mean NPV reached  $\geq 90\%$  for experts using intuition (91.0%) and BASIC (91.1%), and for novices using BASIC (92.5%).

### CADx vs. endoscopists

CADx showed a significantly superior overall diagnostic accuracy compared with that of experts (95.0% vs. 81.7%,  $P=0.03$ ) and novices (95.0% vs. 66.7%,  $P<0.001$ ). CADx also achieved a higher sensitivity compared with the endoscopists (experts

**► Table 4** Interobserver agreement for optical diagnosis based on intuition and on the BLI Adenoma Serrated International Classification.

Observer group	Intuition, Fleiss' kappa* (95%CI)	BASIC, Fleiss' kappa* (95%CI)
Experts	0.61 (0.57–0.66)	0.64 (0.59–0.68)
Novices	0.43 (0.41–0.45)	0.43 (0.41–0.45)
Overall	0.45 (0.44–0.46)	0.46 (0.45–0.47)

CI, confidence interval; BASIC, BLI Adenoma Serrated International Classification; BLI, blue-light imaging.

\* Fleiss kappa values: 0.81–1.00 = very good; 0.61–0.80 = good; 0.41–0.60 = moderate; 0.21–0.40 = fair; and  $<0.20$  = poor.

95.6% vs. 61.1%,  $P=0.03$ ; novices 95.6% vs. 55.4%,  $P<0.001$ ). The NPV of experts using BASIC was slightly higher compared with CADx (87.9% vs. 87.5%,  $P=0.31$ ). The two CRC cases were correctly recognized as CRC by all experts, but not by two novices using intuition and not by four novices using BASIC.

### Interobserver agreement

The interobserver agreement (Fleiss' kappa) for experts showed an increase from 0.61 (95%CI 0.57–0.66) using intuition to 0.64 (95%CI 0.59–0.68) using BASIC (► Table 4). For novices, the interobserver agreement was the same for both phases (Fleiss' kappa 0.43).

### Discussion

The key finding of this study was the significantly higher diagnostic accuracy of CADx compared with the optical diagnosis of experts and novices for colorectal polyps. The diagnostic performance values of CADx were highest when using multimodal imaging (combining HDWL and BLI). We found no significant improvement in diagnostic accuracy when using BASIC compared with intuitive optical diagnoses by expert and novice endoscopists.

The current study demonstrated high diagnostic accuracies for CADx: 88.3% based on HDWL images and 86.7% based on BLI images. Combining these imaging modalities (multimodal imaging) further improved the diagnostic accuracy to 95.0%. This combined diagnostic accuracy was comparable to or even higher than that reported in previous studies [12, 13, 30, 31]. A CADx developed by Min et al. (2019), which used linked color imaging, showed a diagnostic accuracy of 78.4% [30]. A real-time image recognition system by Kominami et al. (2016) showed a diagnostic accuracy of 94.9% [12]. Our CADx did not reach an NPV of  $\geq 90\%$ , unlike the system developed by Chen et al. (2018) [15]. In order to meet this PIVI criterion for the “diagnose-and-leave” strategy, a larger image dataset, consisting of diminutive colorectal polyps, is required. The study was not powered to evaluate the performance of CADx. However, we found a significant difference in diagnostic accuracy of CADx compared with endoscopists of  $>10\%$  (the assumed difference used for the sample size calculation), making it adequately powered.

The use of a full-image deep learning approach for CADx development alleviated the need for manual selection of the region of interest, as was done in most previous studies [15, 30]. The heatmap visualization of what CADx is predicting and whether it is truly the colorectal polyp area is highly attractive. This control mechanism supports the endoscopists to rely on the prediction of CADx. The European Society of Gastrointestinal Endoscopy recommends the use of AI-based techniques if acceptable and reproducible accuracies can be reached [32]. Along with our finding that multimodal imaging improves the diagnostic performance of CADx, standardized imaging protocols need to be developed and used in clinical practice in order to train endoscopists to perform high-quality multimodal imaging for future real-time use of CADx.

The finding that the diagnostic accuracy of experts did not increase significantly for optical diagnosis made using BASIC (intuition 79.5% vs. BASIC 81.7%) is not in line with the study by Subramaniam et al. (2019), which found a significant increase in the diagnostic accuracy of experts using BASIC (87% vs. 97%,  $P < 0.001$ ) [25]. Although BASIC incorporates SSLs and CRCs in its classification, previous studies excluded these histological subtypes [25, 33]. In contrast, the present study included SSLs and CRCs. Optical diagnosis of SSLs is known to be difficult, which might explain why the current study found no significant increase in diagnostic accuracy using BASIC and why the absolute diagnostic accuracies were lower than those reported by Subramaniam et al. As intuition depends on the quality of the endoscopist's knowledge and experience [24] and the experts in the current study were considered highly experienced, another reason might be that the experts had already incorporated BASIC into their intuitive diagnosis.

The reasons for the lack of improvement in diagnostic accuracy among novices are not clear. This finding might be explained by the fact that the novices were unfamiliar with BLI and BASIC, and the validated training module was therefore not extensive enough. In addition, BASIC might incorporate too many descriptors, making it a complex classification model. Recently, Hassan et al. (2019) proposed new BASIC classes involving fewer descriptors, after the current ones were shown to have different strengths in predicting histology [33].

Subgroup analyses on high-confidence diagnoses showed a nonsignificant increase in diagnostic accuracy when using BASIC compared with intuition (85.6% vs. 84.1%). However, this study was not powered to detect differences in high-confidence diagnoses and the number was too low to draw conclusions; the same applies to subgroup analyses for diminutive colorectal polyps diagnosed with high confidence. Consequently, it was not possible to analyze whether the PIVI criteria for the "diagnose-and-leave" strategy were met.

Previous studies on this topic have reported moderate to very good interobserver agreements for experts and novices [15, 25, 26]. In the current study, interobserver agreements were good for experts (intuition 0.61 vs. BASIC 0.64) and moderate for novices (intuition 0.43 vs. BASIC 0.43). An explanation for the comparatively lower agreement scores might be that we used four rather than two histology categories (hyperplastic polyp, adenoma, SSL, and CRC).

This study has several strengths. First, both experts and novices performed the optical diagnosis of colorectal polyps, while previous studies mainly focused on highly selected groups of experts [3–5]. The inclusion of novices increased the generalizability of the study. Second, for the evaluation of BASIC, we included SSLs and CRCs. The addition of these subtypes strengthens the accuracy of the polyp classification and the generalizability of findings compared with other studies [25, 33]. Third, the CADx system was developed using a sufficiently sized image database of 2449 images [30, 31]. The decision to include 60 colorectal polyps for the testing database was deliberate and took account of the time investment and endoscopist concentration required to optically diagnose colorectal polyps. Fourth, state-of-the-art deep learning architectures were used to develop CADx.

This study also has certain limitations. First, endoscopists and CADx were provided with prospectively selected still colorectal polyp images. Although video-based systems have been developed [13], this approach cannot easily be translated to multimodal imaging, as recording videos in two modalities simultaneously is not yet possible and recording videos one after another results in unequal conditions, adding bias. Second, selection bias may have occurred as only high-quality images were selected. Third, to increase generalizability, validation of our CADx system should be performed in a prospective real-time clinical trial. Fourth, it should be noted that the image databases were not balanced, with only 344 (14.0%) benign images in the training database and 15 (25.0%) benign images in the testing database. To counter the effect of unbalanced classes, awaited data augmentation was performed to enforce a uniform distribution over the classes. The number of SSLs and CRCs was limited, making it only possible to analyze benign vs. (pre)malignant with CADx rather than differentiation between the different (pre)malignant histological subgroups.

In conclusion, CADx diagnosed colorectal polyps with a significantly higher diagnostic accuracy than experts and novices. The use of multimodal imaging, incorporating both HDWL and BLI, improved the diagnostic accuracy of CADx. BASIC did not increase the diagnostic accuracy of endoscopists compared with their intuitive optical diagnosis in this setting. The above-mentioned overview stresses the importance of continuing our efforts to improve optical diagnosis, high-quality multimodal imaging, and research into AI-based techniques for future implementation into daily endoscopy practice.

## Clinical trial

Trial Registration: ClinicalTrials.gov (<http://www.clinicaltrials.gov/>) | Registration number (trial ID): NCT04349787 | Type of study: Prospective, non-interventional study

## Competing interests

M. Stefanovic has received speaker fees from Fujifilm Inc. S. Subramaniam has received speaker fees from Fujifilm Inc. P. Bhandari has received research grants and speaker fees from Olympus, Fujifilm, Pentax, 3-D matrix, and Boston Scientific. A.A.M. Masclee has received a

health care efficiency grant from ZON MW (Organization for Health Research and Development, The Netherlands), an unrestricted research grant from Will Pharma S.A., Research funding from Allegan and Grünenthal. He has also provided scientific advice to Bayer, Kyowa Kirin, and Takeda, and received a research grant from Pentax Europe GmbH and the Dutch Cancer Society. E.J. Schoon has received speaker fees and financial support to conduct research from Fujifilm Inc. All other authors declare that they have no conflicts of interest.

## References

- [1] 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.e01–502.e16
- [2] 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
- [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
- [4] 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
- [5] van de Wetering AJP, Meulen LWT, Bogie RMM et al. Optical diagnosis of diminutive polyps in the Dutch Bowel Cancer Screening Program: are we ready to start? *Endosc Int Open* 2020; 8: E257–e265
- [6] Yoshida N, Yagi N, Inada Y et al. Ability of a novel blue laser imaging system for the diagnosis of colorectal polyps. *Dig Endosc* 2014; 26: 250–258
- [7] Djinbachian R, Dube AJ, von Renteln D. Optical diagnosis of colorectal polyps: recent developments. *Curr Treat Options Gastroenterol* 2019; 17: 99–114
- [8] Bisschops R, Hassan C, Bhandari P et al. BASIC (BLI Adenoma Serrated International Classification) classification for colorectal polyp characterization with blue light imaging. *Endoscopy* 2018; 50: 211–220
- [9] Nakano A, Hirooka Y, Yamamura T et al. Comparison of the diagnostic ability of blue laser imaging magnification versus pit pattern analysis for colorectal polyps. *Endosc Int Open* 2017; 5: E224–e231
- [10] Rondonotti E, Paggi S, Amato A et al. Blue-light imaging compared with high-definition white light for real-time histology prediction of colorectal polyps less than 1 centimeter: a prospective randomized study. *Gastrointest Endosc* 2019; 89: 554–564
- [11] Subramaniam S, Hayee B, Aepli P et al. Optical diagnosis of colorectal polyps with blue light imaging using a new international classification. *United European Gastroenterol J* 2019; 7: 316–325
- [12] Kominami Y, Yoshida S, Tanaka S et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016; 83: 643–649
- [13] Byrne MF, Chapados N, Soudan F et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019; 68: 94–100
- [14] Misawa M, Kudo SE, Mori Y et al. Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy. *Gastroenterology* 2016; 150: 1531–1532
- [15] Chen PJ, Lin MC, Lai MJ et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018; 154: 568–575
- [16] Erickson-Bhatt SJ, Boppart SA. Biophotonics for assessing breast cancer. In: Meglinski I, ed. *Biophotonics for medical applications*. Cambridge: Woodhead Publishing; 2015: 175–214
- [17] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013; 310: 2191–2194
- [18] European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* 2016; L119: 1–88
- [19] Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* 2010; 138: 2088–2100
- [20] Schlemper RJ, Riddell RH, Kato Y et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000; 47: 251–255
- [21] Foss FA, West KP, McGregor AH. Pathology of polyps detected in the bowel cancer screening programme. *Diagn Histopathol* 2011; 17: 495–504
- [22] Selvaraju RR, Cogswell M, Das A et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; 2017 Oct 22–29; Venice, Italy 2017: 618–626
- [23] Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*; 2019. *Proceedings of Machine Learning Research PMLR* 2019; 97: 6105–6114
- [24] Van den Brink N, Holbrechts B, Brand PLP et al. Role of intuitive knowledge in the diagnostic reasoning of hospital specialists: a focus group study. *BMJ Open* 2019; 9: e022724
- [25] Subramaniam S, Hayee B, Aepli P et al. Optical diagnosis of colorectal polyps with blue light imaging using a new international classification. *United European Gastroenterol J* 2019; 7: 316–325
- [26] Altman DG. *Practical statistics for medical research*. New York: Chapman & Hall/CRC Press; 1999
- [27] Ijspeert JEG, Bastiaansen BA, van Leerdam ME et al. Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. *Gut* 2016; 65: 963–970
- [28] Rosner B. *Fundamentals of biostatistics*. Mason, OH, United States: Cengage Learning, Inc; 2015
- [29] McCarthy WF. Assessment of sample size and power for the analysis of clustered matched-pair data. *COBRA Preprint Series* 2007: 28
- [30] Min M, Su S, He W et al. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. *Sci Rep* 2019; 9: 2881
- [31] Song EM, Park B, Ha CA et al. Endoscopic diagnosis and treatment planning for colorectal polyps using a deep-learning model. *Sci Rep* 2020; 10: 30
- [32] Bisschops R, East JE, Hassan C et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2019. *Endoscopy* 2019; 51: 1155–1179
- [33] Hassan C, Bisschops R, Bhandari P et al. Predictive rules for optical diagnosis of <10-mm colorectal polyps based on a dedicated software. *Endoscopy* 2020; 52: 52–60