

Computer-aided detection of colorectal polyps using a newly generated deep convolutional neural network: from development to first clinical experience

Lukas Pfeifer^a, Clemens Neufert^a, Moritz Leppkes^a, Maximilian J. Waldner^a, Michael Häfner^b, Albert Beyer^c, Arthur Hoffman^d, Peter D. Siersema^e, Markus F. Neurath^a and Timo Rath^a

Aim The use of artificial intelligence represents an objective approach to increase endoscopist's adenoma detection rate (ADR) and limit interoperator variability. In this study, we evaluated a newly developed deep convolutional neural network (DCNN) for automated detection of colorectal polyps ex vivo as well as in a first in-human trial.

Methods For training of the DCNN, 116 529 colonoscopy images from 278 patients with 788 different polyps were collected. A subset of 10 467 images containing 504 different polyps were manually annotated and treated as the gold standard. An independent set of 45 videos consisting of 15 534 single frames was used for ex vivo performance testing. In vivo real-time detection of colorectal polyps during routine colonoscopy by the DCNN was tested in 42 patients in a back-to-back approach.

Results When analyzing the test set of 15 534 single frames, the DCNN's sensitivity and specificity for polyp detection and localization within the frame was 90% and 80%, respectively, with an area under the curve of 0.92. In vivo, baseline polyp detection rate and ADR were 38% and 26% and significantly increased to 50% ($P = 0.023$) and 36% ($P = 0.044$), respectively, with the use of the DCNN. Of the 13 additionally with the DCNN detected lesions, the majority were diminutive and flat, among them three sessile serrated adenomas.

Conclusion This newly developed DCNN enables highly sensitive automated detection of colorectal polyps both ex vivo and during first in-human clinical testing and could potentially increase the detection of colorectal polyps during colonoscopy. *Eur J Gastroenterol Hepatol* 33: e662–e669

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc.

Introduction

It is well established that the adenoma detection rate (ADR), defined as the percentage of individuals undergoing screening colonoscopy in which at least one adenoma is found, is inversely associated with the incidence of interval colorectal cancers (CRCs). Analyzing data from more than 300 000

colonoscopies, it has been shown that a 1% increase of the ADR results in a decrease of interval CRC incidence by 3% [1]. At the same time, as shown in large-scale prospective long-term studies, colonoscopy with subsequent removal of adenomatous polyps reduces not only the incidence but also the mortality of CRC up to 70% [2–4]. Based on these considerations, ADR has been implemented as a key benchmark criterion to assess quality during screening colonoscopy in clinical practice guideline across the globe [5,6].

At the same time, colonic neoplasia can frequently be missed during screening colonoscopy with adenoma miss rates reaching up to 26%, as shown in a recent meta-analysis [7] and it is thought that at least 50% of all interval carcinomas arise from missed lesions [8]. Several factors are considered to contribute to these high miss rates, and among them are blind spots and human error. Attempts to address human error have been made, for example by including a second observer, and while initial experiences were promising, the effect on ADR remains controversial [9–12]. Furthermore, dual observation remains a labor-intensive and observer-dependent method.

For unbiased and operator-independent real time detection of colorectal polyps, several computer-aided detection (CAD) systems have recently been developed.

Apart from initial experiences of AI systems on ex vivo offline images and videos [13–18], performance of various AI system has been assessed in vivo. Since the first

European Journal of Gastroenterology & Hepatology 2021, 33:e662–e669

Keywords: adenoma detection rate, artificial intelligence, colorectal cancer, computer-aided detection, screening colonoscopy

^aDepartment of Internal Medicine 1, Division of Gastroenterology, Ludwig Demling Endoscopy Center of Excellence, Friedrich-Alexander-University, Erlangen-Nuernberg, Germany, ^bDepartment of Gastroenterology, Physiopathology and Endoscopy of the Gastrointestinal Tract, Central Hospital Bolzano, Bolzano, Italy, ^cGastroenterology Outpatient Clinic, Altötting,

^dDepartment of Internal Medicine 3, Division of Gastroenterology, Klinikum Aschaffenburg, Aschaffenburg, Germany and ^eDepartment of Gastroenterology and Hepatology, Radboud University Medical Center, Nijmegen, The Netherlands

Correspondence to Timo Rath, MD, Professor of Endoscopy, Ludwig Demling Endoscopy Center of Excellence, Division of Gastroenterology, Friedrich-Alexander-University, Erlangen Nuernberg, Ulmenweg 18, 91054 Erlangen, Germany

Tel: + 49/91 31/8 53 50 00; fax: + 49/91 31/8 53 52 52;

e-mail: timo.rath@uk-erlangen.de

Received 13 October 2020 **Accepted** 22 February 2021

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

randomized controlled trial on the efficacy of an artificial intelligence system for colorectal polyp detection by Wang and coworkers [19], three other RCT have been performed to assess the efficacy of artificial intelligence systems [20–22]. However, these trials were performed with different AI systems and given the different development, architecture, and training of the different systems, results cannot be transferred or extrapolated from one system to the other.

In the current study, we evaluated a novel deep convolutional neural network (DCNN) for automated detection of colorectal polyps that has been developed by a manufacturer of the healthcare industry (Hoya Corporation, Pentax Medical Division, Digital Endoscopy, Friedberg, Germany) in close collaboration with clinical and scientific partners and assessed the performance of the DCNN ex vivo as well as in a first in-human pilot trial.

Material and methods

Development of the AI algorithm

For this study, a novel object detector for colorectal polyp detection based on DCNNs was developed by a project team of Pentax Medical (Hoya Corporation, Pentax Medical Division, Digital Endoscopy, Friedberg, Germany) in close collaboration with clinical and scientific partners. For training of the DCNN, a total of 116 529 colonoscopy images from 278 patients containing a total of 788 different polyps were collected. Images were acquired with different commercially available high definition (HD) or HD+ video processors (Imagina EPK-i 5500c, EPK-i, Optivista EPK-i7010, all Pentax Medical, Tokyo, Japan). From these images, a subset of 10 467 images containing 504 different polyps were manually annotated by one of the five expert endoscopists. These annotations consisted of bounding boxes delineating the polyp within the image and were treated as the gold standard. Time from beginning of development to CE certification of the DCNN was 11 months. After CE Certification, a brand name of *Discovery* was chosen for the DCNN.

In contrast to image classification where a given image is categorized into classes (e.g. polyp vs. no polyp) [15], the task of object detection includes the localization of object instances within the image. The backbone of the DCNN evaluated in this study is provided by a VGG16-style network [23]. The DCNN processes video data as a sequence of single video frames and generates predictions based on the visual evidence of a single video frame. The predictions of individual consecutive frames are then fused to provide a more stable detection. Single frame detections are provided through the ‘Single Shot MultiBox Detector’ (SSD) framework [24]. The principle of the generation of this DCNN and its diagnostic output are shown in Fig. 1.

Ex vivo test phase

To evaluate the performance of the system, we performed an ex vivo test using a set of 45 independently acquired videos from routine colonoscopies consisting of 15 534 single frames in total. Using this test set, two evaluations were performed: a per-frame-based evaluation and a per-polyp-based evaluation.

For the per-frame-based evaluation, a true positive was defined as an image that contains a polyp, which has been

localized correctly and has been detected with sufficiently high confidence. Localization quality is commonly measured through the intersection over union (*IoU*) criterion.

$$IoU(D, A) = \frac{|D \cap A|}{|D \cup A|}$$

where $|D \cap A|$ is the image area of the detection D that overlaps with the annotation A and $|D \cup A|$ is the union of the image areas of detection and annotation. We followed the commonly used conventions on object detection and considered a detection to be localized correctly if the *IoU* with the annotation is ≥ 0.5 and evaluated the performance of the system over all confidence values to yield a sensitivity vs. specificity curve.

Based on the per-frame evaluation, an operating point was selected and used to perform the per-polyp evaluation. For the per-polyp evaluation, a true positive was defined to be a polyp that has been detected and correctly localized in at least a single video frame.

Patients

The study was approved by the local ethics committee (Friedrich-Alexander-University, Erlangen-Nuernberg, Germany, IRB No.: 166_19B) and was performed in accordance with the declaration of Helsinki. The study was registered at ClinicalTrials.gov under NCT04359355. Patients presenting for colonoscopy at the Ludwig Demling Endoscopy Center were prospectively included from January to June 2020 if they fulfilled the following inclusion criteria: screening or surveillance colonoscopy, work-up of abdominal pain and/or change of bowel habits. Exclusion criteria were as follows: known or suspected inflammatory bowel diseases, known polyps or referral for polypectomy, presence of coagulopathy. Patients with a diagnosis of CRC during colonoscopy were also excluded. Prior to study inclusion, written informed consent was obtained from all participating subjects, and minors were excluded.

Colonoscopy and computer-aided polyp detection

All patients received bowel preparation with low-volume polyethylene glycol-based bowel lavage in a split dose regimen. Colonoscopy was performed using commercially available HD endoscopes and video processors (EC38-i10 and Optivista EPK-i7010, both Pentax Medical, Tokyo, Japan) by four experienced (lifetime experience >1000 colonoscopies and >300 colonoscopies per year) endoscopists in a tandem setting: after reaching the cecum, withdrawal was performed with high definition-white light endoscopy (HD-WLE), followed by a second inspection of the whole colon again with the DCNN for automated polyp detection. To systematically assess the detection of colorectal polyps with HD-WLE and the DCNN within the different segments of the colon, the colon was divided into the following three segments: cecum and ascending colon, transverse colon, descending and sigmoid colon. Each segment was first inspected with HD-WLE followed by inspection of the same segment with the DCNN. Insertion times as well as withdrawal times in each segment under either HD-WLE or the DCNN were recorded using a stop watch. During further cleaning of the colon, polyp

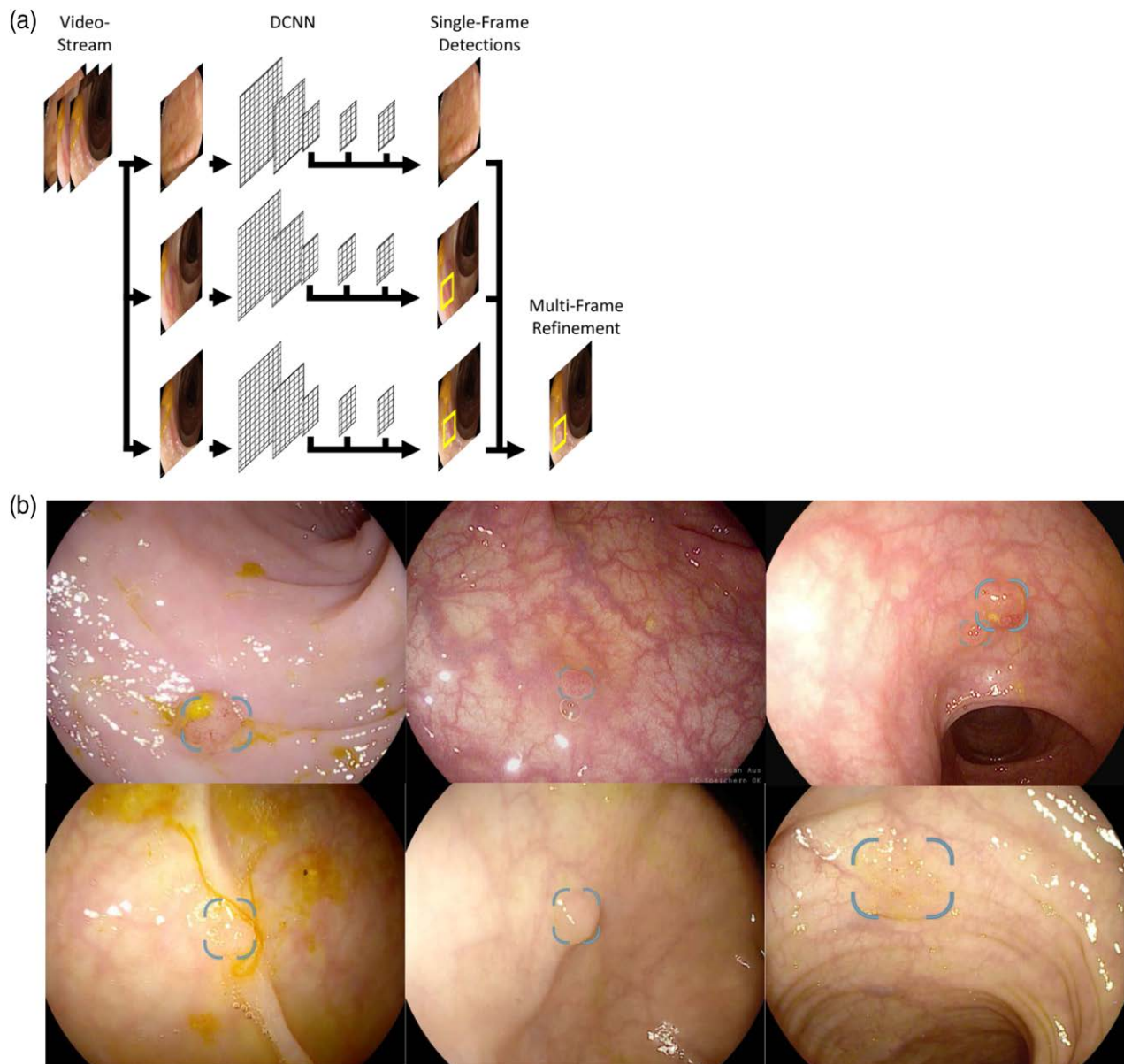


Fig. 1. Development and diagnostic output of the system. (a) The deep convolutional neural network (DCNN) processes video data as a sequence of single video frames and generates predictions based on the visual evidence of a single video frame. The predictions from individual frames are then fused to provide a more stable detection. (b) Different examples of polyp detection with the DCNN during routine colonoscopy. The computer-aided detection (CAD) system generates the diagnostic output on a second screen on which polyps are highlighted by a bounding box. Note that the DCNN is able to detect multiple polyps in a single frame simultaneously (upper right picture).

assessment (morphology and size), and polyp removal, the stop watch was paused. Morphology of polyps and adenomas in each segment were assessed using the Paris criteria [25], polyp size was evaluated against an open biopsy forceps with a diameter of 7 mm. All polyps and adenomas found during colonoscopy were removed using either cold- or hot-snare polypectomy at the discretion of the endoscopist, formalin fixed in separate containers after removal and analysed by experienced gastrointestinal (GI) pathologists.

Endpoints

For ex vivo testing, the primary endpoint was to assess the diagnostic performance of the DCNN for the detection of colorectal polyps. Primary endpoint in the clinical trial

was the ADR under HD-WLE and the DCNN. Secondary endpoint included polyp detection rate (PDR) under HD-WLE and the DCNN, the size and histology of adenomas detected under HD-WLE and the DCNN, number of true and false-positive lesions detected by the DCNN, as well as withdrawal times in each colonic segment.

Statistical analysis

All data are presented as mean, median, SD and range, as indicated in the respective figures and tables. Grouped continuous data were compared using the unpaired *t*-test. For the comparison of ADR and PDR, the Wilcoxon matched pairs test was used. A two-sided $P < 0.05$ was considered to be significant. The exact value was reported with P value between 0.05 and 0.001, whereas $P < 0.001$

was reported for values below it. The statistics were processed using Graph Pad Prism (version 5, GraphPad Software Inc., La Jolla, USA).

Results

Technical characteristics and ex vivo performance of the deep convolutional neural network for automated polyp detection

To gain first insights into the performance of the system, an ex vivo test using a set of 45 independently acquired videos from routine colonoscopies consisting of 15 534 single frames in total was used. The test dataset composition used for this ex vivo testing is shown in Table 1.

When being analyzed at an input frame rate of 30 Hz, the DCNN's sensitivity for polyp detection and localization within the frame was 90% with a specificity of 80%. As shown in Fig. 2, an area under the curve for polyp detection and localization of 0.92 was calculated in ROC analysis.

For per-polyp evaluation, a per-frame sensitivity of 90% was chosen as an operating point. Using the previously described per-polyp evaluation protocol, we were able to achieve 100% sensitivity on our test dataset. The average number of false positives (either false detections or incorrectly localized detections) per frame was approximately 8%.

Table 1. Composition of the test dataset used for this ex vivo of the deep convolutional neural network

| | Test dataset |
|-----------------------------|---------------|
| Patients (n) | 45 |
| Total video frames (n) | 15 534 |
| Polyp containing frames (%) | 60/40 |
| Polyp size (mm) | |
| Mean \pm SD | 6.1 \pm 2.6 |
| Range | 2–12 |
| Paris classification (n) | |
| Is | 20 |
| IIa | 15 |
| IIb | 5 |

Clinical characteristics of the patient cohort and in vivo detection with the deep convolutional neural network for automated polyp detection

A total of 43 patients presenting for colonoscopy were prospectively included in this pilot study. One patient was excluded because two removed polyps were accidentally put in the same histopathologic container; therefore, clear histologic assignment of the polyp detected with the DCNN was not possible. There were no complications during colonoscopy. A total of 42 patients with a mean age of 62 years were included in the final analysis of this pilot trial. Clinical characteristics and withdrawal time under first inspection with HD-WLE and second inspection with the DCNN are shown in Table 2. The mean Boston Bowel Preparation Score (BBPS) was 5.7 with a range from 3 to 8, indicating that bowel preparation was suboptimal in this cohort.

During first withdrawal without the DCNN, the number of polyps detected was 32 and this significantly increased by 26 additionally detected polyps (+81%) to a total of 58 after the second inspection with the DCNN ($P < 0.001$). Corresponding to this, baseline PDR significantly increased from 38% after withdrawal with HD-WLE to a PDR of 50% after the second inspection with the DCNN ($P = 0.023$), as shown in Table 3.

Similarly, the number of adenomas detected was significantly increased during withdrawal with the DCNN: during first withdrawal with HD-WLE, 21 adenomas were found and this increased by 13 additionally detected adenomas (+62%) to a total of 34 during withdrawal with the DCNN ($P = 0.006$). Correspondingly, baseline ADR after the first inspection the DCNN was 26% and significantly increased to an ADR of 36% after second inspection with the DCNN ($P = 0.044$, Table 3).

Characteristics of the polyps detected during first inspection without the DCNN and those additionally detected during second inspection with the DCNN are shown in Table 4. Importantly, among the 13 additionally with the DCNN detected adenomas, three were found to be sessile-serrated adenomas and these were not

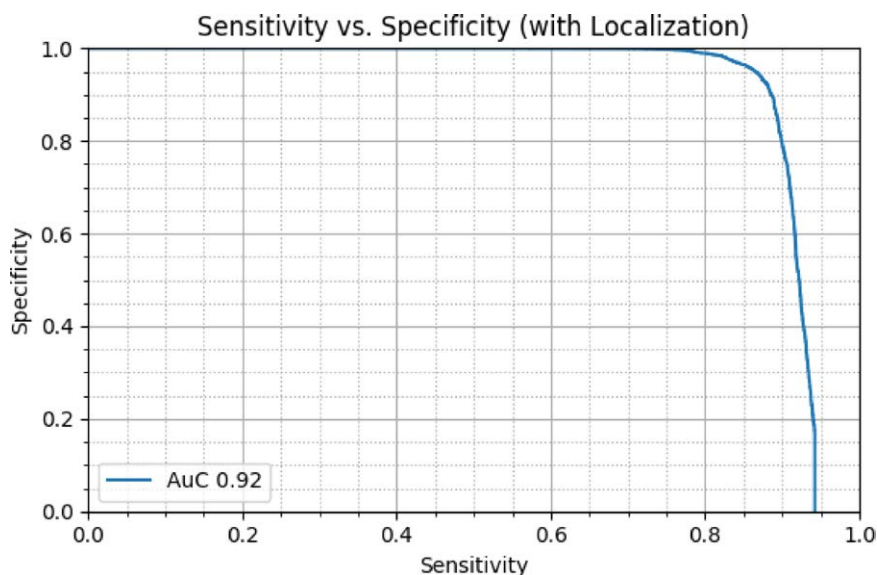


Fig. 2. Diagnostic performance of the system. When being analyzed at an input frame rate of 30 Hz, the deep convolutional neural network's (DCNN's) sensitivity and specificity for polyp detection and localization within the frame were 90% and 80%, respectively with an area under the curve (AUC) of 0.92.

Table 2. Patient characteristics and withdrawal times

| | | | |
|-----------------------------------------|-------------------------------|----------------------------|-----------------|
| Patients, n (m/f) | 42 (26/16) | | |
| Age (years) | | | |
| Mean ± SD | 62 ± 13 | | |
| Range | 34–83 | | |
| BBPS | | | |
| Mean ± SD | 5.7 ± 1.0 | | |
| Withdrawal time in minutes mean (range) | 1st inspection (without DCNN) | 2nd inspection (with DCNN) | <i>P</i> -value |
| Cecum and ascending colon | 1:55 (0:50–3:05) | 1:39 (1:00–3:19) | 0.10 |
| Transverse colon | 1:37 (0:45–3:05) | 1:28 (0:40–2:11) | 0.27 |
| Descending and sigmoid colon | 2:39 (1:30–4:53) | 2:19 (1:04–4:20) | 0.08 |

BBPS, Boston Bowel Preparation Scale; DCNN, deep convolutional neural network; FP, False positives.

Table 3. Total number of polyps and adenomas and polyp detection rate and adenoma detection rate after first (without deep convolutional neural network) and second inspection (with deep convolutional neural network)

| | After 1st inspection | After 1st + 2nd inspection | <i>P</i> -value |
|--------------------------|----------------------|----------------------------|------------------|
| Total number of polyps | 32 | 58 (+81%) | <i>P</i> < 0.001 |
| PDR | 16/42 = 38% | 21/42 = 50% | 0.023 |
| Total number of adenomas | 21 | 34 (+62%) | <i>P</i> = 0.006 |
| ADR | 11/42 = 26% | 15/42 = 36% | 0.044 |

PDR: Number of patients in which at least one polyp was found divided by the total number of patients included.

ADR: Number of patients in which at least one adenoma was found divided by the total number of patients included.

ADR, adenoma detection rate; PDR, polyp detection rate.

recognized without the DCNN even during second inspection. Regarding adenoma size and shape 7 of 13 additionally with the DCNN detected adenomas were diminutive while six were small adenomas between 5 and 10 mm and 46% were of flat morphology (Paris IIA/IIb). Altogether, these pilot data indicate that CAD-assisted colonoscopy indeed facilitates detection of clinically relevant lesions and those adenomas that represent a challenge for detection such as diminutive and flat lesions.

Of note, all adenomas and polyps found during the first inspection were also detected by the DCNN, resulting in a true positive rate of 100% and false-negative rate of 0%. In 18 of the 42 patients included, at least one false-positive detection occurred during second withdrawal with the DCNN (Table 5). However, 10 of 18 patients in whom at least one false-positive detection was observed had a BBPS < 6 with a mean number of false-positive hits of 2.5 in these patients while no false-positive detection occurred in patients with excellent bowel preparation, indicating that adequate bowel cleaning is mandatory to reduce false-positive detections.

Discussion

The global CRC burden is expected to increase to more than 2.2 million new cases and 1.1 million deaths by 2030 [26]. At the same time, it has been shown that colonoscopic polypectomy can reduce CRC mortality by approximately 50% [4]. Nonetheless, despite its proven efficacy with regard to CRC prevention, colonoscopy is not able to detect all colonic lesions and sometimes even misses CRC. A recent meta-analysis analyzing 43 studies and more

Table 4. Characteristics of the polyps detected during first inspection without deep convolutional neural network and those additionally detected during second inspection with deep convolutional neural network

| | 1st inspection (without DCNN) | 2nd inspection (with DCNN) |
|------------------------------|-------------------------------|----------------------------|
| Adenoma size | | |
| <5 mm | 14 | 7 |
| 5–10 mm | 5 | 6 |
| >10 mm | 2 | |
| Adenoma localization | | |
| Cecum and ascending colon | 10 | 6 |
| Transverse colon | 1 | 1 |
| Descending and sigmoid colon | 10 | 6 |
| Histology | | |
| LG IEN | 14 | 10 |
| HG IEN | 1 | |
| SSA without dysplasia | 6 | 3 |
| SSA with dysplasia | | |
| Paris classification | | |
| Is | 10 | 7 |
| Ip | 1 | 4 |
| Ila | 7 | 2 |
| Ilb | 3 | |

DCNN, deep convolutional neural network; HG IEN, high-grade intraepithelial neoplasia; LG IEN, low-grade intraepithelial neoplasia; SSA, sessile serrated adenoma.

than 15 000 tandem colonoscopies calculated a pooled miss rate of 26% for adenomas, 9% for advanced adenomas and 27% for serrated polyps [7].

Blind spots as well as human error are considered to be major factors for missing lesions during colonoscopy. The latter factor can, at least partly, be explained by psychology factors such as ‘inattentive blindness’, wherein an observer fails to observe an image on the screen due to distraction or ‘change blindness’, in which lesions are missed during interruptions in visual scanning or during eye movements [27,28]. Several studies have addressed this issue by including a second observer during colonoscopy. Although participation of a second observer in colonoscopy has been shown to be beneficial in increasing PDR, the effect on ADR is still controversial [9,10,12,29].

In recent years, an interesting solution to the human error in detecting polyps has been developed, i.e. CAD systems. Ideally, a CAD system should exhibit a high sensitivity while at the same time having high specificity [30,31]. Furthermore, it is mandatory that a CAD system has a short diagnostic delay and provides an onscreen alerting system in order to make it useful in clinical practice. To date, several CAD system have been developed, the majority of which use a convolutional or deep neural network for automated and operator-independent detection of colorectal polyps. Although performance of these systems on offline videos and images seems promising, evidence on the ability to increase detection during real-time clinical practice is only scarce to date [15–18,32].

In this work, we present our results from development with ex vivo testing to the first clinical experience with a new CAD system. As shown in our study, our system detected all polyps that were recognized with standard HD-WLE and no lesion was missed with CAD. Although not tested in a randomized controlled setup, this high sensitivity translated into a marked increase of the ADR during withdrawal with the CAD system. Apart from ADR, also the PDR increased from 38% without CAD to 50% with

Table 5. Characteristics of false-positive findings according to bowel preparation status

| BBPS | Total number of patients | Number of patients with ≥ 1 FP | Mean number of FP per patient |
|----------------------|--------------------------|-------------------------------------|-------------------------------|
| Excellent (BBPS 8–9) | 5 | 0 | 0 |
| Good (BBPS 6–7) | 24 | 8 | 1.0 |
| Poor (BBPS 3–5) | 13 | 10 | 2.5 |

BBPS, Boston Bowel Preparation Score; FP, false positive.

the CAD system and out the 26 additionally with the CAD system detected polyps, 50% of them (13 out of 26) were hyperplastic lesions, while 13 were adenomas. Given the high sensitivity of our CAD system, this major increase in the detection of hyperplastic lesions is not surprising but might be regarded as a drawback of CAD systems since this increase in detection of clinically irrelevant lesions can potentially lead to more polypectomies, a higher workload and a prolongation of procedural time. First reports have already used computer-aided diagnosis algorithms for polyp differentiation into neoplastic and non-neoplastic (hyperplastic) lesions [33–35] and in the future, combination of these techniques represents a promising approach for efficient handling of hyperplastic polyps.

As a further consequence of the high sensitivity of our DCNN, the rate of false-positive signal was relatively high in the *in vivo* phase of our study: in 18 of the 42 patients included, at least one false-positive signal occurred. However, when looking at these numbers, it should be kept in mind that we did not exclude patients with insufficient bowel preparation. In fact, of 18 patients in which at least one false-positive detection with the DCNN occurred, 10 (56%) had insufficient bowel cleaning with a BBPS of 5 or below and mean number of false-positive hits was 2.5 in these patients while no false-positive detection occurred in patients with excellent bowel preparation. In light of this, larger multicentric studies in which patients will be stratified and analyzed according to their bowel cleaning score are necessary before final conclusions on false-positive detection can be drawn.

That CAD systems are indeed able to significantly increase ADR has been shown by Wang and coworkers in the first randomized controlled trial published in this field to date [19]. Randomizing >1000 patients to either HD-WLE or HD-WLE with a CAD system resulted in a significant increase in ADR, PDR and mean number of polyps and adenomas per colonoscopy in the CAD group [19]. However, the increase in overall ADR was mainly due to an increased detection of diminutive adenomas with the CAD [19]. These results are consistent with data from a more recent randomized controlled trial by Repici and coworkers using the GI-Genius CAD system [21]. In this trial, ADR was increased by 14.4% in the CAD group with significantly increased detection of diminutive and small adenomas.

In our study, we observed a significant increase of baseline ADR by 10% and of the 13 additionally with the DCNN detected adenomas in our study, almost half were diminutive and of flat morphology (Paris IIa/IIb). These findings might have two implications: at first sight, increased detection of flat and diminutive lesions with CAD indicates a powerful detection of lesions through CAD that are subtle and otherwise hard to detect. At the

same time, diminutive polyps, especially when located in the rectosigmoid, have a very low risk of developing to advanced histologic features and the incidence of cancer in diminutive polyps is extremely low [36–39]. Therefore, the clinical relevance of these lesions may well be limited and at the same time raise the question on the cost-benefit ratio of the implementation of artificial intelligence into daily routine. This question becomes especially relevant when AI-system are not only able to detect lesions, but also are capable of characterizing lesions to predict polyp histology. Just recently, an analysis was by Mori and colleagues provided first evidence that implementation of AI-assisted polyp characterization can indeed result in considerable cost-savings [40]. As shown, a diagnose-and-leave strategy supported by the AI prediction (i.e., diminutive rectosigmoid polyps were not removed when predicted as non-neoplastic) was estimated to reduce the average colonoscopy cost and the gross annual reimbursement for colonoscopies by 18.9% and 149.2 million dollars in Japan, 6.9% and 12.3 million dollars in England, 7.6% and 1.1 million dollars in Norway, and 10.9% and 85.2 million dollars in the United States, respectively, compared to the resect-all-polyps strategy [40].

Together with further studies on the cost-effectiveness, future trials also need to address whether the increased detection of diminutive lesions with CAD system also reduces the incidence of CRC, which is the main object of screening colonoscopy [41,42]. Nevertheless, in our study the use of CAD also led to the detection of three sessile serrated adenomas that were not recognized without the DCNN even during second inspection and these observations provide a glimpse that CAD can very well improve detection of clinically relevant lesions.

Limitations of our study also need to be addressed. As such, first, the number of patients included in the *in vivo* phase of our study is relatively small and therefore our study should be regarded as a pilot study. Second, we used a back-to-back design to analyze the efficacy of the DCNN and while this approach offers the advantage that missed lesions can be characterized very precisely, the possibility that, during second withdrawal, the operator pays special attention to the lesions found during the first inspection represents a potential bias bearing the risk of overestimating the sensitivity of the system. In light of these considerations, it seems clear that large randomized controlled trials are necessary to further evaluate the efficacy of this system.

As shown, this newly developed CAD enables highly sensitive automated detection of colorectal polyps both *ex vivo* and during first *in-human* clinical testing and holds the potential to increase the detection of colorectal polyps during colonoscopy. Future studies should systematically assess whether CAD-based colonoscopy over standard HD-WLE for detection of colorectal polyps and adenomas and therefore can render surveillance colonoscopy more effectively.

Acknowledgements

The authors gratefully acknowledge the work of Hannes Seibt, Christian Eggert and Harald Huber who developed the AI algorithm. Further the authors gratefully

acknowledge the help of the endoscopy staff at the Ludwig Demling Endoscopy Center of Excellence in conducting this study.

We acknowledge support by Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) within the funding programme Open Access Publishing.

The study was registered at ClinicalTrials.gov under NCT04359355.

Conflicts of interest

Timo Rath and Peter D. Siersema served as consultants and received speaking honoraries from Pentax Medical. For the remaining authors, there are no conflicts of interest.

References

- Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014; 370:1298–1306.
- Nishihara R, Wu K, Lochhead P, Morikawa T, Liao X, Qian ZR, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* 2013; 369:1095–1105.
- Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993; 329:1977–1981.
- Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 2012; 366:687–696.
- Kaminski MF, Thomas-Gibson S, Bugajski M, Bretthauer M, Rees CJ, Dekker E, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. *Endoscopy* 2017; 49:378–397.
- Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal cancer screening: recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol* 2017; 112:1016–1030.
- Zhao S, Wang S, Pan P, Xia T, Chang X, Yang X, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. *Gastroenterology* 2019; 156:1661–1674.e11.
- Pohl H, Robertson DJ. Colorectal cancers detected after colonoscopy frequently result from missed lesions. *Clin Gastroenterol Hepatol* 2010; 8:858–864.
- Aslanian HR, Shieh FK, Chan FW, Ciarleglio MM, Deng Y, Rogart JN, et al. Nurse observation during colonoscopy increases polyp detection: a randomized prospective study. *Am J Gastroenterol* 2013; 108:166–172.
- Buchner AM, Shahid MW, Heckman MG, Diehl NN, McNeil RB, Cleveland P, et al. Trainee participation is associated with increased small adenoma detection. *Gastrointest Endosc* 2011; 73:1223–1231.
- Kim TS, Park DI, Lee DY, Yoon JH, Park JH, Kim HJ, et al. Endoscopy nurse participation may increase the polyp detection rate by second-year fellows during screening colonoscopies. *Gut Liver* 2012; 6:344–348.
- Lee CK, Park DI, Lee SH, Hwangbo Y, Eun CS, Han DS, et al. Participation by experienced endoscopy nurses increases the detection rate of colon polyps during a screening colonoscopy: a multicenter, prospective, randomized study. *Gastrointest Endosc* 2011; 74:1094–1102.
- Ahmad OF, Soares AS, Mazomenos E, Brandao P, Vega R, Seward E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *Lancet Gastroenterol Hepatol* 2019; 4:71–80.
- Kudo SE, Mori Y, Misawa M, Takeda K, Kudo T, Itoh H, et al. Artificial intelligence and colonoscopy: Current status and future perspectives. *Dig Endosc* 2019; 31:363–371.
- Misawa M, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018; 154:2027–2029.e3.
- Tajbakhsh N, Gurudu SR, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans Med Imaging* 2016; 35:630–644.
- Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* 2018; 2:741–748.
- Zhang R, Zheng Y, Poon CCY, Shen D, Lau JYW. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognit* 2018; 83:209–219.
- Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; 68:1813–1819.
- Liu WN, Zhang YY, Bian XQ, Wang LJ, Yang Q, Zhang XD, Huang J. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. *Saudi J Gastroenterol* 2020; 26:13–19.
- Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020; 159:512–520.e7.
- Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADE-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol* 2020; 5:343–351.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations 2015, ICLR 2015; 7–9 May, San Diego, USA, arXiv:1409.1556.
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision 2016, ECCV 2016; 8–16 October, Amsterdam, Netherlands, arXiv:1512.02325.
- The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003;58 (6 Suppl):S3–43.
- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017; 66:683–691.
- Simons DJ, Chabris CF. Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception* 1999; 28:1059–1074.
- Simons DJ, Levin DT. Change blindness. *Trends Cogn Sci* 1997; 1:261–267.
- Nishizawa T, Suzuki H, Takahashi M, Kaneko H, Fujiyama Y, Komatsu H, et al. Trainee participation during colonoscopy adversely affects polyp and adenoma detection rates. *Digestion* 2011; 84:245–246.
- Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: the future is almost here. *World J Gastrointest Endosc* 2018; 10:239–249.
- Mori Y, Kudo SE, Berzin TM, Misawa M, Takeda K. Computer-aided diagnosis for colonoscopy. *Endoscopy* 2017; 49:813–819.
- Yamada M, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep* 2019; 9:14465.
- Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HH, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018; 154:568–575.
- Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, 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.
- Sánchez-Montes C, Sánchez FJ, Bernal J, Córdova H, López-Cerón M, Cuatrecasas M, et al. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy* 2019; 51:261–265.
- Butterly LF, Chase MP, Pohl H, Fiarman GS. Prevalence of clinically important histology in small adenomas. *Clin Gastroenterol Hepatol* 2006; 4:343–348.
- Gupta N, Bansal A, Rao D, Early DS, Jonnalagadda S, Wani SB, et al. Prevalence of advanced histological features in diminutive and small colon polyps. *Gastrointest Endosc* 2012; 75:1022–1030.
- Lieberman D, Moravec M, Holub J, Michaels L, Eisen G. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. *Gastroenterology* 2008; 135:1100–1105.

- 39 O'Brien MJ, Winawer SJ, Zauber AG, Gottlieb LS, Sternberg SS, Diaz B, *et al.* The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology* 1990; 98:371–379.
- 40 Mori Y, Kudo SE, East JE, Rastogi A, Bretthauer M, Misawa M, *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.
- 41 Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, *et al.* Colorectal cancer screening: recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2017; 153:307–323.
- 42 Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, *et al.* Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin* 2018; 68:250–281.