Journal Pre-proof

Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video)

Albert J. de Groof, MD, Maarten R. Struyvenberg, MD, Kiki N. Fockens, MD, Joost van der Putten, MSc, Fons van der Sommen, PhD, Tim G. Boers, MSc, Sveta Zinger, PhD, Raf Bisschops, MD PhD, Peter H. de With, PhD, Roos E. Pouw, MD PhD, Wouter L. Curvers, MD PhD, Erik J. Schoon, MD PhD, Jacques J.G.H.M. Bergman, MD PhD

PII: S0016-5107(20)30014-6

DOI: <https://doi.org/10.1016/j.gie.2019.12.048>

Reference: YMGE 11917

To appear in: *Gastrointestinal Endoscopy*

Received Date: 24 October 2019

Accepted Date: 23 December 2019

Please cite this article as: de Groof AJ, Struyvenberg MR, Fockens KN, van der Putten J, van der Sommen F, Boers TG, Zinger S, Bisschops R, de With PH, Pouw RE, Curvers WL, Schoon EJ, Bergman JJGHM, Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video), *Gastrointestinal Endoscopy* (2020), doi: <https://doi.org/10.1016/j.gie.2019.12.048>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Copyright © 2020 by the American Society for Gastrointestinal Endoscopy

Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video)

Albert J. de Groof MD¹, Maarten R. Struyvenberg MD¹, Kiki N. Fockens MD¹, Joost van der Putten MSc², Fons van der Sommen PhD², Tim G. Boers MSc², Sveta Zinger PhD², Raf Bisschops MD PhD⁴, Peter H. de With PhD², Roos E. Pouw MD PhD¹, Wouter L. Curvers MD PhD³, Erik J. Schoon MD PhD³, Jacques J.G.H.M. Bergman MD PhD¹

¹ Department of Gastroenterology and Hepatology, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

² Department of Electrical Engineering, VCA group, Eindhoven University of Technology, Eindhoven, the Netherlands

³ Department of Gastroenterology and Hepatology, Catharina Hospital Eindhoven, Eindhoven, the Netherlands

⁴ Department of Gastroenterology and Hepatology, University Hospitals Leuven, Leuven, Belgium

The Netherlands National Trials Registry (NTR) number: NL7544

This research is supported by the Dutch Cancer Society and Technology Foundation STW, as part of their joint strategic research program 'Technology for Oncology'

Corresponding author:

Jacques J.G.H.M. Bergman, MD PhD
Professor of Gastrointestinal Endoscopy
Director of Endoscopy
Amsterdam UMC, location Academic Medical Center
Meibergdreef 9
1105 AZ Amsterdam
The Netherlands
Phone number: +31 (0)20 5663556
Fax number: +31 (0)20 6917033

Author contribution:

Study design: A.J. de Groof, M.R. Struyvenberg, W.L. Curvers, E.J. Schoon, J.J. Bergman

Patient recruitment and data collection: A.J. de Groof, M.R. Struyvenberg, K.N. Fockens, W.L. Curvers, E.J. Schoon, R.E. Pouw, J.J. Bergman

CAD system development: J. van der Putten, F. van der Sommen, T.G. Boers, S. Zinger, P.H. de With

Data analysis and interpretation: A.J. de Groof, M.R. Struyvenberg, K.N. Fockens, W.L. Curvers, J.J. Bergman

Writing of manuscript: A.J. de Groof, M.R. Struyvenberg, K.N. Fockens, F. van der Sommen, J.A. van der Putten, T.G. Boers, R. Bisschops, W.L. Curvers, E.J. Schoon, J.J. Bergman.

All authors reviewed the final manuscript for important intellectual content and agreed to submit.

Disclosures:

A.J. de Groof: *None declared*

M.R. Struyvenberg: *None declared*

K.N. Fockens: *None declared*

J. van der Putten: *None declared*

F. van der Sommen: *None declared*

T.G. Boers: *None declared*

S. Zinger: *None declared*

R. Bisschops: *Fujifilm (research support, consulting fees, speakers fees)*

P.H. de With: *None declared*

R.E. Pouw: *None declared*

W.L. Curvers: *None declared*

E.J. Schoon: *None declared*

J.J. Bergman: *Fujifilm, NinePoint Medical (research support); Fujifilm (speaking fees)*

ABSTRACT

Background & Aims: We assessed the preliminary diagnostic accuracy of a recently developed computer aided detection (CAD) system for detection of Barrett's neoplasia during live endoscopic procedures.

Methods: The CAD system was tested during endoscopic procedures of 10 patients with nondysplastic Barrett's esophagus (NDBE), and 10 patients with confirmed Barrett's neoplasia. White-light endoscopy images were obtained at every 2-centimeter level of the Barrett's segment, and immediately analyzed by the CAD system, providing instant feedback to the endoscopist. At every level, 3 images were evaluated by the CAD system. Outcome measures were as follows: (1) diagnostic performance of the CAD system per-level and per-patient, defined as accuracy, sensitivity, and specificity. Ground truth was established by expert assessment and corresponding histopathology; and (2) concordance of 3 sequential CAD predictions per-level.

Results: Accuracy, sensitivity, and specificity of the CAD system in a per-level analyses were 90%, 91%, and 89%, respectively. In total, 9 out of 10 neoplastic patients were correctly diagnosed. The single lesion not detected by CAD, showed NDBE in the endoscopic resection specimen. In only 1 NDBE patient, the CAD system produced false positive predictions. In 75% of all levels, the CAD system produced 3 concordant predictions.

Conclusion: This is one of the first studies to evaluate a CAD system for Barrett's neoplasia during live endoscopic procedures. The system detected neoplasia with high accuracy, with only a small number of false-positive predictions, and with a high concordance rate between separate predictions. The CAD system is thereby ready for testing in larger, multicenter trials.

Key words: Computer aided detection, deep learning, endoscopy, Barrett's neoplasia

INTRODUCTION

Barrett's esophagus (BE) is a known precursor for esophageal adenocarcinoma (EAC). Therefore, BE patients undergo regular surveillance endoscopies to detect EAC at an early stage[1, 2]. Early Barrett's neoplasia is however easily missed, because early lesions are usually flat with only subtle morphological changes. Furthermore, due to the low progression rate of BE (<1% per patient year), most general endoscopist rarely encounter early Barrett's neoplasia and therefore remain unfamiliar with its

endoscopic appearance[3, 4]. This issue has been addressed in a recent study that describes the development and validation of online training program identification of early neoplasia by endoscopists[5]. The quality of high-definition endoscopes is now so good that nearly all subtle neoplastic lesions are recognized reliably by expert endoscopists specialized in management of BE[6]. General endoscopists would therefore benefit from expert endoscopists' recognition skills.

In recent years, multiple computer aided detection (CAD) systems using deep learning techniques have been developed in the field of endoscopy, showing promising results[7-12]. The ARGOS-consortium envisions to develop a real-time CAD system for Barrett's neoplasia[13, 14]. Recently, a deep learning CAD system has been developed by our consortium that can automatically detect and localize early neoplastic Barrett's lesions on white-light endoscopy (WLE) images. This CAD system was validated and subsequently benchmarked with a large group of international endoscopists. The CAD system detected and localized neoplasia with high accuracy, and outperformed all 53 endoscopist assessors.

The aim of this pilot study is 2-fold: (1) to evaluate the in-vivo performance of our CAD system for primary detection of Barrett's neoplasia during live endoscopic procedures in real-time; and (2) to assess concordance between sequential CAD predictions of the same area in the Barrett's segment.

METHODS

Setting and design

This prospective study was performed at the departments of Gastroenterology and Hepatology of the Amsterdam University Medical Centers, location Academic Medical Center, and the Catharina Hospital Eindhoven, both tertiary referral centers for Barrett's neoplasia. This study was approved by the local medical ethical committee of both hospitals and registered at The Netherlands National Trials Registry (*NTR NL7544*).

Deep learning CAD system design

In this study, we assessed the feasibility and preliminary diagnostic accuracy of an image-based CAD system in a real-time setting in the endoscopy suite. This CAD system has previously been developed by our group and validated extensively in an offline setting on prospectively collected data[15]. This CAD system has several unique features. First, it is pretrained with a large dataset of 494,364 labeled endoscopic images (GastroNet), containing a variety of endoscopic imagery of the gastrointestinal tract using a semisupervised learning approach. Subsequently, the CAD system was trained with multilevel,

Barrett's-specific imagery, containing a total of 1,544 images of Barrett's neoplasia and NDBE. The CAD system was then tested on 2 separate external test datasets, with a total of 160 unique patients. In these test sets, the CAD system showed to detect neoplasia with high accuracy and near-perfect localization performance. Finally, algorithm performance was benchmarked by 53 international endoscopists, which were all outperformed by the CAD system.

Our CAD system has been developed to serve as a primary detection tool, that can assist endoscopists during surveillance endoscopies by highlighting visible abnormalities that might contain neoplasia. When presented with an endoscopic image, the CAD system first classifies an image to be neoplastic or nondysplastic (ND) BE. This classification is quantified in a classification score. This score is defined as the probability of the corresponding image being neoplastic, ranging from 0-100, where a classification score of 100% indicating the highest likelihood of neoplasia, and 0% the highest likelihood of being nondysplastic. The threshold for classifying an image to be neoplastic, was set at 60%, as established in a previous publication by our group. If an image is classified as neoplastic, the CAD system directly produces a delineation, delineating the area most suspicious for neoplasia. Additionally, it indicates the preferred biopsy location within that delineated area. For this study, a graphical user interface was developed, to visualize classification scores, confidence intervals, and segmentation performance (ie, delineations and preferred biopsy location) in a single image (Figure 1).

Model description

For the CAD system, a custom-made, fully residual, hybrid ResNet/U-Net model was used. Although generally classification and segmentation tasks are trained individually, using different models, we developed a hybrid model that can be trained to solve both tasks simultaneously. This property facilitates better learning as each task reveals complementary information to the other task, better guiding the training process. An extensive, technical description of the algorithm model has been published elsewhere[16].

Patients

Patients were eligible for inclusion when referred for endoscopic work-up of early Barrett's neoplasia (n=10), defined as histologically confirmed high-grade dysplasia (HGD) or early adenocarcinoma (EAC), and patients referred for surveillance endoscopy of nondysplastic (ND)BE (n=10), defined as the histologically confirmed absence of dysplasia in previous endoscopies.

Endoscopic procedure

All endoscopic procedures were performed by expert endoscopists (J.B., R.P., E.S., W.C.) with extensive experience in the use of advanced imaging techniques and endoscopic treatment of early Barrett's neoplasia. The procedures were performed from March 2019 to August 2019. All cases were histologically confirmed by expert pathologists.

Image acquisition and CAD analyses

During the endoscopic procedure, the esophagus was cleaned and examined with white-light endoscopy and blue-light imaging (BLI) using the EG-760 endoscope of the ELUXEO system (Fujifilm Corporation, Tokyo, Japan), after which the length of the Barrett's segment was recorded according to the Prague-classification system[17]. Then, the endoscope was positioned ± 2 centimeters proximal of the gastroesophageal junction (level 1) and an endoscopic image of the Barrett's segment in overview was made in WLE. During image acquisition, the endoscope was kept in a centralized position, without focusing on any lesion (if present). This WLE image was automatically transferred to the CAD system via the high-definition video output of the endoscopy system, after which the CAD system directly analyzed the image and produced a classification score between 0 and 100, as previously discussed. In this way 3 sequential images were obtained at this level. Subsequently, the endoscope was pulled back 2 centimeters (level 2), and again 3 corresponding endoscopic images of the Barrett's segment in overview were captured, and analyzed separately by the CAD system. This workflow continued every 2 centimeters until the proximal maximum extend of the Barrett's segment was reached. On each level, before CAD analyses, the endoscopist had to indicate whether a visible lesion was present.

CAD-guided tissue sampling

After image acquisition, on each level the CAD output of the 3 endoscopic images was reviewed. In case a lesion was identified by the CAD system in ≥ 1 of the 3 images, a targeted biopsy specimen was obtained from the preferred biopsy location (as indicated by the CAD system) or, when the prior assessment of the expert endoscopist corresponded with the CAD diagnosis, the supposedly neoplastic lesion was resected according to standard protocol[1, 2]. In case a lesion was missed by the CAD system, yet identified by the endoscopist, the lesion was removed by endoscopic resection. In case no lesion was identified by the CAD system, and the endoscopist had indicated that indeed no visible abnormalities were present, only random biopsy specimens were taken at that level, according to standard protocol. Yet, in case the latter resulted in a histopathological diagnosis of low-grade dysplasia or indefinite for

dysplasia, the level was still considered as NDBE because our CAD system is not developed to detect non-visible dysplasia.

The rationale to obtain 3 images per level was 2-fold: (1) to test concordance of separate CAD predictions on still images obtained at the same level; and (2) to assess whether combined multi-image information would lead to increased performance of the CAD system.

Outcome measures

We evaluated performance of the CAD system in a per-image, per-level and per-patient analyses.

Outcomes of per-image analysis

- Diagnostic accuracy of the CAD system in a per-image analyses;
- Diagnostic accuracy of the CAD system in a per-image analyses, when only including high confidence predictions (ie, classification scores of $\leq 20\%$ or $\geq 80\%$) are included.

Outcomes of per-level and patient analysis using combined information of multiple images per-level

Because 3 images were obtained on each level, we were able to create 2 definitions of a “positive” CAD prediction on a level (Figure 2): (1) a “minority voting analysis,” in which the CAD prediction was considered positive when at least 1/3 of predictions surpassed the neoplasia threshold of 60%; and 2) a “majority voting analysis,” in which the CAD prediction was considered positive when at least 2/3 of predictions surpassed the neoplasia threshold of 60%. These 2 outcome measures were first analyzed per level, followed by a per patient analyses.

Diagnostic performance was thereby defined as the number of levels/patients with visible lesions, as identified by the expert as potentially neoplastic, detected by the CAD system (sensitivity) and the number of levels/patients without visible endoscopic lesions, confirmed by normal random biopsies, correctly identified by the CAD system as normal (specificity). Outcomes are reported as follows:

- Per-level analysis of diagnostic accuracy of the CAD system, for the 2 definitions of CAD predictions: the minority voting- and majority voting analyses;
- Per-patient analysis of diagnostic accuracy of the CAD system, for the 2 definitions of CAD predictions. In both outcomes, the per-patient CAD prediction was considered positive when *at least* 1 level was identified as positive. The per-patient CAD prediction was considered negative when none of the levels were identified as positive;

- Concordance of predictions by the CAD system, defined as the number of levels where the CAD system was consistent in its classification for all 3 images (ie, 3x NDBE or 3x neoplasia).
- Computation time of the CAD system: Time required for analyzing an image and producing a classification and delineation and indicating the preferred biopsy location.

Statistical analysis

Statistical analysis was performed using SPSS Statistical software package for Windows (version 25, SPSS Inc., Chicago, Ill, USA). For descriptive statistics, normally distributed data was shown as mean ($\pm$ standard deviation [SD]) and variables with skewed distribution were shown as median (interquartile range [IQR]).

RESULTS

In total, 23 patients were included in this study. The first 3 patients were excluded from the analysis because a coding bug was discovered in the algorithm software resulting in aberrant results. After patient 3, this coding bug was resolved. The remaining twenty consecutive patients were analyzed (13 males; mean age 67 years [SD ± 10]). Mean circumferential extent of the Barrett's segment was 5 centimeters (SD ± 4), mean maximum extent was 7 centimeters (SD ± 3). In these 20 patients, a total of 48 levels were analyzed by the CAD system. Eleven levels were considered neoplastic as the endoscopist identified a potentially neoplastic lesion which was subsequently resected in the same session. Thirty-seven levels were considered as NDBE, as the endoscopist did not identify a potentially neoplastic lesion and corresponding 4-quadrant biopsies yielded no neoplasia (Table 1).

Per image analyses

In total, 25 out of 33 neoplastic images and 96 out of 111 NDBE images were correctly predicted by the CAD system, resulting in an image-based accuracy, sensitivity, and specificity of 84%, 76%, and 86%, respectively.

In 96 out of 144 predictions (67%), the CAD system predicted an image with high confidence (ie, classification score of $<20\%$ or $\geq 80\%$). Diagnostic accuracy of the high confidence predictions was 91% for accuracy, sensitivity and specificity, respectively (Table 2).

Per level- and per patient analysis using combined information of multiple images per level

Per level analyses

In total, 48 levels of interest were analyzed, out of which 11 contained a visible lesion and 37 had no visible abnormalities.

In both the minority voting- and majority voting analysis, 10 out of 11 neoplastic levels (91%) were correctly identified by the CAD system. In the minority voting analysis, 28 out of 37 levels (76%) were correctly identified as NDBE. When using majority voting, a total of 33 out of 37 (89%) NDBE levels were correctly identified (Table 2).

Per patient analyses

Again, in both the minority and majority voting analysis, in 9 out of 10 neoplastic patients the lesion was correctly identified by the CAD system, resulting in a sensitivity of 90% for both analyses.

In the minority voting analysis, in 7 out of 10 NDBE patients, none of the predictions surpassed the neoplasia threshold, resulting in a per patient specificity of 70%. In the other 3 patients, at least one prediction was false positive. In the majority voting analysis, in 9 out of 10 NDBE patients, only ≤ 1 predictions surpassed the neoplasia threshold, resulting in a specificity of 90%. In only 1 patient, 2 or more predictions per level were false-positive (Table 1 and 2).

The single lesion not detected by CAD in both the per level- and per patient analyses, showed NDBE in the endoscopic resection specimen.

Concordance of CAD predictions

In 36 out of 48 levels (75%), all 3 predictions were similar (ie, 3x NDBE or 3x neoplastic). In 12 levels, one prediction was not concordant with the other 2 predictions.

Computation time

The mean time for classification per image was 0.191 seconds (SD 0.001). The mean time for segmentation (ie, producing a delineation and biopsy location) was 0.328 seconds (SD 0.002).

DISCUSSION

This paper describes a unique in vivo study evaluating a computer aided detection system for primary detection of early Barrett's neoplasia.

Early Barrett's lesions are easily missed during surveillance endoscopies [5, 6]. Our group has developed a computer aided detection system that may aid endoscopists in the primary detection of these lesions. In an ex vivo study using endoscopic images, this CAD system detected early Barrett's neoplasia with an accuracy of 88% and with a sensitivity and specificity of 93% and 83%, respectively. When benchmarked

against 53 nonexpert endoscopists, the CAD system outperformed all assessors in terms of accuracy and sensitivity[15].

In the current study, we used the CAD system in vivo during endoscopic procedures. The aim was to assess the preliminary diagnostic performance of our CAD system and its concordance when evaluating more or less identical consecutive images obtained at the same level within the Barrett's segment.

We designed a stepwise CAD workflow that resembles the envisioned application of our image-based CAD system in clinical practice. Similar to the current Seattle biopsy protocol, where multiple random biopsies are obtained at every 2 centimeters, we obtained multiple image-based CAD predictions at 2 centimeters intervals[1, 2]. At each level, 3 consecutive images were obtained without significantly moving the endoscope. These multiple CAD assessments per level were performed for 2 reasons: (1) to test the robustness of the system: a robust CAD system is expected to provide similar predictions for slightly different images of the same region; (2) to allow post-hoc analysis of how such consecutive CAD predictions can best be combined to improve CAD performance.

In a per-image analyses, accuracy, sensitivity, and specificity of the CAD system was 84%, 76%, and 86%, respectively. The diagnostic accuracy increased considerably when only the high-confidence predictions (ie, classification score of $<20\%$ or $\geq 80\%$) were analyzed, resulting in an accuracy, sensitivity, and specificity of 91%, respectively. This indicates a correlation between high confidence of the CAD system and likelihood of presence of visible abnormalities. This suggests that the performance of our CAD system can be improved by integrating a filter that only includes high confidence predictions.

In a per level and per patient analyses, using the combined predictions of 3 consecutive still images, our CAD system detected neoplasia in vivo with high accuracy. Nine out of 10 neoplastic lesions were detected by the CAD system, regardless of the post-hoc analysis approach. All lesions detected by the CAD system were endoscopically resected with histology showing HGD or EAC. The single lesion identified by the endoscopist as potentially neoplastic yet not detected by the CAD system (Figure 3) did not show dysplasia upon histological examination. In retrospect, this visible abnormality probably represented a recent biopsy scar rather than a true neoplastic lesion. Per our definition, however, areas identified by the expert endoscopist as potentially neoplastic are the target areas for CAD detection and therefore this lesion was considered as being missed by our CAD-system. Vice versa, none of the NDBE levels contained a visible abnormality, as assessed by the endoscopist. In 5 levels, random biopsies,

however, resulted in the diagnosis of indefinite for dysplasia or low-grade dysplasia. Yet, given the absence of any visible abnormalities, these areas were considered as true negative CAD predictions.

In 9 out of 10 NDBE patients, all evaluated NDBE levels were correctly classified as such by the CAD system, when following a majority voting approach. In only 4 out of 37 NDBE levels, originating from 1 NDBE patient and 2 neoplastic patients, the CAD system produced false positive predictions. This resulted in the acquisition of only 4 redundant biopsy specimens in a total of 20 patients.

In the minority voting approach, in 7 out of 10 patients all 28 NDBE levels were correctly classified. The CAD system misclassified 9 NDBE levels, in 3 NDBE patients (6 levels) and 2 neoplastic patients (3 levels).

These results indicate that the minority voting approach (ie, a positive CAD prediction on a *single* still image is considered as abnormal) yields a relatively large number of false-positive detections, although a per level specificity of 76% and per patient specificity of 70% might still be considered acceptable: in an average length Barrett's segment, this would lead to a maximum of a single additional targeted biopsy. It does, however, appear logical to use the CAD predictions of multiple still images for decision making, in particular because CAD predictions showed to be concordant to a high extent, thereby indicating robustness of the system.

Incorporating the CAD predictions of 3 consecutive still images obtained in more or less the same scope position increased specificity of the system whereas the 91% sensitivity remained unaltered. When the CAD system correctly classified a level as neoplastic, it did so in at least 2 out of 3 images. However, when the CAD system incorrectly classified a level as neoplastic, generally only a single of the 3 images yielded this false-positive prediction. Therefore, by using a majority voting approach (2 positive predictions are required to label an area as neoplastic) a significant number of false positives were eliminated while maintaining the high sensitivity of the CAD system.

Using multiple sequential still images for a combined CAD prediction is logical because performance of a CAD system generally increases when it is allowed to predict multiple similar data points, generating a more balanced overall prediction, and thus being more robust against outliers. This multi-frame approach is thereby comparable to video-based CAD analyses, where a prediction is produced over multiple adjacent frames. A video-based CAD system could also incorporate additional filters that exclude non-informative frames (eg, by the presence of blur or suboptimal expansion of the esophagus) and low-confidence predictions. This will most likely increase CAD performance even further.

What would be the envisioned approach of our image-based CAD system in daily practice? Based on the results of this pilot study, a majority voting approach should be considered as the preferred endoscopic workflow for our image-based CAD system. The CAD system could be used in its current form.

Endoscopists performing a surveillance endoscopy would obtain 2 corresponding images every 2 centimeters. In case both images are classified as neoplastic by the CAD system, a targeted biopsy specimen is obtained. In case both images are labelled as NDBE, the endoscopist continues proximal to the next level. In case the CAD prediction is inconclusive (ie, 1 NDBE and 1 neoplastic), a third endoscopic image is obtained. The outcome of this third image determines whether or not a targeted biopsy is warranted. The low false-positive rate of the CAD system will likely enable a swift and user-friendly workflow, in which the endoscopist is not constantly alerted by spurious detections. Additionally, a filter could be integrated in the CAD system that automatically excludes low confidence predictions from its analyses, as this will increase CAD performance. It should be noted that the acquisition of similar images in a Barrett's segment in overview may be cumbersome for less-experienced endoscopists because esophageal insufflation and timing of image acquisition can be challenging.

The computational speed of our CAD system was approximately 0.5 seconds per image. In the current implementation, an ensemble of models is used and each model is loaded into the memory sequentially before making a prediction. To increase speed, this process can be enhanced significantly by preloading all models into the GPU memory and performing inference in parallel. Additionally, postprocessing is currently performed on a CPU with nonoptimized functions, which limits computational speed. In the current implementation, the CAD system was however able to provide a complete prediction (ie, classification and segmentation) within half a second after taking the image, which is feasible for an image-based approach.

Our CAD system is designed to detect neoplasia in overview with high-definition white-light endoscopy because recent guidelines recommend the use of high-definition white-light endoscopy as the technique of choice for primary detection of Barrett's neoplasia[2, 15]. In our opinion, optical chromoscopy techniques are more useful for detailed characterization after detection. A dedicated computer aided diagnosis (CADx) system for detailed characterization is under development by our group.

This study has several strengths. First, this is one of the first in vivo studies to test a specific CAD system for primary detection of Barrett's neoplasia. Ebigbo et al. have recently published a short report describing the in vivo use of an image-based CAD system for Barrett's neoplasia in a limited number of

cases. The details of the algorithm remain unknown yet the training was performed on only 129 images without independent external validation and the system appears to be designed for classification of neoplasia in magnification, rather than primary detection in overview [18, 19].

Second, in our study we used a custom made, state-of-the art deep learning architecture. The CAD system used in this pilot study has already been extensively interrogated in an offline setting, including benchmark performance by international endoscopists, providing a solid basis for its intrinsic validity.

Third, this CAD system is easy-to-use in daily endoscopic practice, providing easy interpretable and direct feedback that can be summoned by the convenience of the endoscopist, without continuous distractive output during the endoscopic procedure.

This study has several limitations. First, this was a pilot study with only a limited number of patients. The diagnostic accuracy reported in this study can therefore only be considered as a rough estimate of its performance in daily practice. Second, all procedures were performed by expert endoscopists. Therefore, image quality, as well as esophageal cleaning and expansion, and endoscope positioning might have been more optimal than is to be expected in a general surveillance setting. Third, the first 3 patients in this study were excluded from analysis, due to an unfortunate software bug that resulted in groundless predictions. In short, during the first 3 cases of this study, the coding bug caused the network to use the running averages of a single sample at a time (the single images captured during the study) instead of the running averages calculated during training. This resulted in predictions that were not based on the knowledge the CAD system had acquired during training. Fourth, this CAD system has been trained with and tested on Fujifilm Eluxeo imagery. Its performance when used in conjunction with other endoscopy systems is therefore unknown. With the current trend of steady growth in vision quality of endoscopic vision, we foresee that the applied CAD system, when trained with sufficient specific data, will also work on alternative endoscopic vision systems, in particular after making some input video adaptations.

In conclusion, we report the in vivo testing of a CAD system for primary detection of Barrett's neoplasia. The CAD system detected neoplastic lesions with high sensitivity, with only a small number of false-positive predictions. The results of this pilot study indicate that our CAD system is user-friendly and well-suited for usage in clinical practice. Future multicenter, randomized studies should be performed in a surveillance setting to assess the true additive value of the CAD system in daily practice.

REFERENCES

- [1] Spechler SJ, Sharma P, Souza RF, Inadomi JM, Shaheen NJ. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology*. 2011;140:1084-91.
- [2] Weusten B, Bisschops R, Coron E, Dinis-Ribeiro M, Dumonceau JM, Esteban JM, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. 2017;49:191-8.
- [3] Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med*. 2011;365:1375-83.
- [4] Sikkema M, de Jonge PJ, Steyerberg EW, Kuipers EJ. Risk of esophageal adenocarcinoma and mortality in patients with Barrett's esophagus: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2010;8:235-44.
- [5] Bergman J, de Groof AJ, Pech O, Ragunath K, Armstrong D, Mostafavi N, et al. An Interactive Web-Based Educational Tool Improves Detection and Delineation of Barrett's Esophagus-Related Neoplasia. *Gastroenterology*. 2019;156:1299-308.
- [6] Scholvinck DW, van der Meulen K, Bergman JJ, Weusten BL. Detection of lesions in dysplastic Barrett's esophagus by community and expert endoscopists. *Endoscopy*. 2017;49:113-20.
- [7] Misawa M, Kudo SE, Mori Y, Nakamura H, Kataoka S, Maeda Y, et al. Characterization of Colorectal Lesions Using a Computer-Aided Diagnostic System for Narrow-Band Imaging Endocytoscopy. *Gastroenterology*. 2016;150:1531-2.
- [8] Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, 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;83:643-9.
- [9] Takemura Y, Yoshida S, Tanaka S, Kawase R, Onji K, Oka S, et al. Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). *Gastrointest Endosc*. 2012;75:179-85.
- [10] Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology*. 2018;155:1069-78.
- [11] Horie Y, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, Ishiyama A, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc*. 2019;89:25-32.
- [12] Byrne MF, Chapados N, Soudan F, Oertel C, Linares Perez M, Kelly R, 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.
- [13] Groof Jd, van der Sommen F, van der Putten J, Struyvenberg MR, Zinger S, Curvers WL, et al. The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterology Journal*. 2019:538-47.
- [14] van der Sommen F, Zinger S, Curvers WL, Bisschops R, Pech O, Weusten BL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy*. 2016;48:617-24.
- [15] de Groof AJ, Struyvenberg MR, van der Putten J, van der Sommen F, Fockens KN, Curvers WL, et al. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multi-Step Training and Validation Study with Benchmarking. *Gastroenterology*. 2019:Epub ahead of print.
- [16] van der Putten J, de Groof J, van der Sommen F, Struyvenberg M, Zinger S, Curvers W, et al. Pseudo-labeled Bootstrapping and Multi-stage Transfer Learning for the Classification and Localization of Dysplasia in Barrett's Esophagus. In: Suk H-I, Liu M, Yan P, Lian C, editors. *Machine Learning in Medical Imaging*. Cham: Springer International Publishing; 2019. p. 169-77.

[17] Sharma P, Dent J, Armstrong D, Bergman JJ, Gossner L, Hoshihara Y, et al. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. *Gastroenterology*. 2006;131:1392-9.

[18] Ebigbo A, Mendel R, Probst A, Manzeneder J, Prinz F, de Souza LA, Jr., et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut*. 2019:Epub ahead of print.

[19] Ebigbo A, Mendel R, Probst A, Manzeneder J, Souza LA, Jr., Papa JP, et al. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut*. 2018:1143-5.

TABLES AND FIGURES

Table 1. Characteristics of included patients, and CAD performance displayed as true positives, true negatives, false positives, and false negatives, based on the “majority voting analyses”.

	Prague classification	Ground truth diagnosis on patient level	Paris classification	Number of levels analyzed	True positive levels	True negative levels	False positive levels	False negative levels
Patient 1	C13M14	NDBE	NA	5	0	5	0	0
Patient 2	C2M5	NDBE	NA	2	0	2	0	0
Patient 3	C9M10	NDBE	NA	4	0	4	0	0
Patient 4	C11M12	NDBE	NA	5	0	3	2	0
Patient 5	C2M5	NDBE	NA	2	0	2	0	0
Patient 6	C8M10	NDBE	NA	3	0	3	0	0
Patient 7	C5M5	Neoplasia	II-b	3	0	1	1	1
Patient 8	C0M3	Neoplasia	II-b/c	1	1	0	0	0
Patient 9	C0M5	Neoplasia	II-b	2	1	1	0	0
Patient 10	C4M5	Neoplasia	II-a	2	1	1	0	0
Patient 11	C0M2	Neoplasia	I-s	1	1	0	0	0
Patient 12	C5M6	Neoplasia	III	2	1	1	0	0
Patient 13	C2M5	Neoplasia	II-a/c	1	1	0	0	0
Patient 14	C6M8	Neoplasia	II-a/b	2	1	0	1	0
Patient 15	C7M8	Neoplasia	II-a/b	2	2	0	0	0
Patient 16	C5M6	NDBE	NA	2	0	2	0	0
Patient 17	C8M10	NDBE	NA	3	0	3	0	0
Patient 18	C9M10	NDBE	NA	3	0	3	0	0
Patient 19	C6M8	NDBE	NA	2	0	2	0	0
Patient 20	C0M3	Neoplasia	II-a/b	1	1	0	0	0
Total	NA	NA	NA	48	10	33	4	1

Table 2. Diagnostic accuracy of the CAD system in a per level-, per patient-, and per image analyses.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Per level analyses			
Minority voting analyses	79.2 (38/48)	90.9 (10/11)	75.7 (28/37)
Majority voting analyses	89.6 (43/48)	90.9 (10/11)	89.2 (33/37)
Per patient analyses			
Minority voting analyses	80.0 (16/20)	90.0 (9/10)	70.0 (7/10)
Majority voting analyses	90.0 (18/20)	90.0 (9/10)	90.0 (9/10)
Per image analyses			
Overall analyses	84.0 (121/144)	75.8 (25/33)	86.5 (96/111)
High confidence predictions	90.6 (87/96)	90.9 (20/22)	90.5 (67/74)

Figure 1. Visualization of the graphical user interface of the CAD system, providing real-time feedback to the endoscopist. In the 2 NDBE images (*upper row*), the CAD system did not detect any abnormalities, resulting in low classification scores. In the 2 neoplastic images (*lower row*), the CAD system detected a lesion, resulting in high classification scores and corresponding visualization of the delineation and preferred biopsy location.

Figure 2. Visualization of diagnosis per level, based on 3 individual predictions using the minority voting- (MiV) and the majority voting (MaV) approach.

Level 1: MiV NDBE, MaV NDBE;

Level 2: MiV neoplasia, MaV NDBE;

Level 3: MiV neoplasia, MaV neoplasia;

Level 4: MiV neoplasia, MaV neoplasia.

Figure 3. Subtle abnormality at 12 o'clock that was not detected in any of the images by the CAD system, yet assessed to be neoplastic by the endoscopist. Histopathological evaluation of the resection specimen showed the lesion only contained NDBE.

List of acronyms

CAD:	Computer aided detection
NDBE:	Non-dysplastic Barrett's esophagus
BE:	Barrett's Esophagus
WLE:	White light endoscopy
HGD:	High grade dysplasia
EAC:	Esophageal adenocarcinoma
BLI:	Blue Light Imaging
SD:	Standard deviation
IQR:	Interquartile range