A computer-assisted algorithm for narrow-band-imaging–based tissue characterization in Barrett's esophagus

A computer-assisted algorithm for narrow-band imaging-based tissue characterization in Barrett's esophagus

Maarten R. Struyvenberg, MD,[1,*] Albert J. de Groof, MD,[1,*] Joost van der Putten, MSc,[2] Fons van der Sommen, PhD,[2] Francisco Baldaque-Silva, MD, PhD,[4] Masami Omae, MD,[4] Roos Pouw, MD, PhD,[1] Raf Bisschops, MD, PhD,[5] Michael Vieth, MD, PhD,[6] Erik J. Schoon, MD, PhD,[3] Wouter L. Curvers, MD, PhD,[3] Peter H. de With, PhD,[2] Jacques J. Bergman, MD, PhD[1]

Amsterdam, Eindhoven, the Netherlands; Stockholm, Sweden; Leuven, Belgium; Bayreuth, Germany

Background and Aims: The endoscopic evaluation of narrow-band imaging (NBI) zoom imagery in Barrett's esophagus (BE) is associated with suboptimal diagnostic accuracy and poor interobserver agreement. Computer-aided diagnosis (CAD) systems may assist endoscopists in the characterization of Barrett's mucosa. Our aim was to demonstrate the feasibility of a deep-learning CAD system for tissue characterization of NBI zoom imagery in BE.

Methods: The CAD system was first trained using 494,364 endoscopic images of general endoscopic imagery. Next, 690 neoplastic BE and 557 nondysplastic BE (NDBE) white-light endoscopy overview images were used for refinement training. Subsequently, a third dataset of 112 neoplastic and 71 NDBE NBI zoom images with histologic correlation was used for training and internal validation. Finally, the CAD system was further trained and validated with a fourth, histologically confirmed dataset of 59 neoplastic and 98 NDBE NBI zoom videos. Performance was evaluated using fourfold cross-validation. The primary outcome was the diagnostic performance of the CAD system for classification of neoplasia in NBI zoom videos.

Results: The CAD system demonstrated accuracy, sensitivity, and specificity for detection of BE neoplasia using NBI zoom images of 84%, 88%, and 78%, respectively. In total, 30,021 individual video frames were analyzed by the CAD system. Accuracy, sensitivity, and specificity of the video-based CAD system were 83% (95% confidence interval [CI], 78%-89%), 85% (95% CI, 76%-94%), and 83% (95% CI, 76%-90%), respectively. The mean assessment speed was 38 frames per second.

Conclusion: We have demonstrated promising diagnostic accuracy of predicting the presence/absence of Barrett's neoplasia on histologically confirmed unaltered NBI zoom videos with fast corresponding assessment time. (Gastrointest Endosc 2021;93:89-98.)

Abbreviations: AUC, area under the curve; BE, Barrett's esophagus; CADe, computer-aided detection; CADx, computer-aided diagnosis; CI, confidence interval; EAC, esophageal adenocarcinoma; HGD, high-grade dysplasia; IQR, interquartile range; NBI, narrow-band imaging; NDBE, nondysplastic Barrett's esophagus; ROC, receiver operating characteristic; WLE, white-light endoscopy.

DISCLOSURE: *Dr Bergmann has received research support from NinePoint Medical and speaker fees from Fujifilm. Dr Baldaque-Silva has received research support from Boston Scientific. Dr Bisschops has received research support, consulting fees, and speaker fees from Fujifilm. Dr Vieth has received honoraria for lecturing from Falk, Shire, and Olympus. All other authors disclosed no financial relationships.*

*Drs Struyvenberg and de Groof contributed equally to this article and share first authorship.

0016-5107
https://doi.org/10.1016/j.gie.2020.05.050

Received January 21, 2020. Accepted May 17, 2020.

Current affiliations: Department of Gastroenterology and Hepatology, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands (1); Department of Electrical Engineering, VCA Group, Eindhoven University of Technology, Eindhoven, the Netherlands (2); Department of Gastroenterology and Hepatology, Catharina Hospital Eindhoven, Eindhoven, the Netherlands (3); Department of Gastroenterology and Hepatology, Karolinska University Hospital, Stockholm, Sweden (4); Department of Gastroenterology and Hepatology, University Hospitals Leuven, Leuven, Belgium (5); Institute of Pathology, Bayreuth Clinic, Bayreuth, Germany (6).

Reprint requests: J.J. Bergman, MD, PhD, Professor of Gastrointestinal Endoscopy, Director of Endoscopy, Academic Medical Center Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.

INTRODUCTION

Barrett's esophagus (BE) is a known precursor for esophageal adenocarcinoma (EAC), which is often preceded by the presence of high-grade dysplasia (HGD). Patients with BE undergo regular endoscopic surveillance to detect neoplasia (ie, HGD/EAC) at an early stage. Early neoplasia can be treated endoscopically with preservation of the esophagus and an excellent prognosis.[1-4] However, the current BE surveillance protocol, consisting of inspection with white-light endoscopy (WLE) and random biopsies, is suboptimal. The endoscopic detection of early neoplasia is difficult due to its subtle endoscopic appearance, and random biopsies are associated with sampling error.[5,6]

The endoscopic diagnosis of BE neoplasia is generally a 2-step process of primary detection in WLE in overview, followed by detailed inspection of these visible abnormalities for characterization. This detailed inspection is often performed using narrow-band imaging (NBI; Olympus, Tokyo, Japan) in magnification because of its ability to improve visualization of mucosal and vascular patterns. However, correct characterization in NBI magnified view is challenging for endoscopists performing BE surveillance. Several NBI classification systems have been proposed using a variety of criteria but are still suboptimal.[7-12]

Computer-aided diagnosis (CAD) using deep-learning techniques has shown promising results in different scientific research domains, including GI endoscopy.[13-16] An important distinction is made between computer-aided detection (CADe) and computer-aided diagnosis (CADx). CADe systems are produced to detect pathology (ie, detection of neoplastic lesions), in contrast to CADx systems, which are developed to classify pathology (ie, characterization as nondysplastic or neoplastic). Recently, our group demonstrated a computer-aided detection system that recognized and localized BE neoplasia on white-light overview images with high accuracy, enabling primary detection.[17,18] We speculated that this system might be supplemented by a second CADx algorithm using NBI for tissue characterization. The aim of the current study was to investigate the feasibility of a novel deep-learning CADx system for the characterization of NBI zoom imagery in BE.

METHODS

Setting

This study was performed at the Departments of Gastroenterology and Hepatology of the Amsterdam University Medical Centers (Academic Medical Center) and Karolinska University Hospital (Stockholm), both tertiary referral centers for Barrett's neoplasia, and at the Department of Electrical Engineering of the Eindhoven University of Technology. Official approval for the use of all imagery was obtained from the local Ethics Committee at both medical centers.

CAD system architecture

Deep-learning models in endoscopy are generally pretrained using large natural image databases (eg, ImageNet[19]) consisting of a large variety of classes of "general imagery" such as boats, cars, and dogs. We envisioned that an endoscopic CAD system would gain more discriminative knowledge from specific, endoscopy-driven pretraining. We used the GastroNet database comprising 494,364 labeled endoscopic images obtained from upper- and lower-GI endoscopy to pretrain our deep-learning system to learn informative discriminative patterns.[20] These informative patterns were then exploited by application of transfer learning techniques on our specific BE datasets.

The overall architecture of our CADx system involved a custom-made convolutional neural network (hybrid ResNet-UNet architecture). Via transfer learning techniques, we trained the CADx system in a stepwise manner using 4 endoscopic datasets and finally tested its performance on NBI zoom imagery. All WLE images in datasets 1 and 2 were used solely for pretraining the system. Datasets 3 and 4 contained only NBI imagery that was used to train and test the CADx system. Images or videos obtained from the same patient were always allocated to either the training or the test set to prevent data leakage and thereby risk of overfitting (ie, patient-split analysis). See Table 1 for the stepwise construction of our CADx system architecture using the 4 endoscopic databases.

Datasets used for CAD (pre)training and testing

Dataset 1: GastroNet. The CADx system was pretrained on a dataset of 494,364 labeled endoscopic images from 15,286 patients, named GastroNet.[18] These images contained a variety of endoscopic imagery (eg, colon, stomach, duodenum, esophagus). All images were collected retrospectively, during 2012 to 2018 at the Amsterdam UMC (location AMC) and automatically deidentified after extraction from the database. Olympus HQ190, HQ180, and HQ290 endoscopes, and Fujifilm 700 series endoscopes were used to record all images. Next, each image was labeled by organ system using a separate semi-supervised computer algorithm created for this project.[20]

Dataset 2: WLE overview images. Subsequently, the CADx system was further pretrained and enhanced on a dataset consisting of 690 WLE overview images of BE neoplasia (ie, HGD/EAC) from 198 patients and 557 nondysplastic BE (NDBE) images from 216 patients. The lesions on the neoplastic images were delineated by 2 experts (M.S., A.J.G.), and the NDBE images were reviewed by the same experts for absence of visible neoplasia.

Dataset 3: NBI zoom images. Dataset 3 consisted of 71 NDBE NBI zoom images of 50 patients, and 112 neoplastic (HGD/EAC) NBI zoom images of 50 other

TABLE 1. Overview of the stepwise construction of our CADx system architecture using 4 endoscopic datasets

	Type of imagery	Number of images	Type of labeling	Purpose of dataset
Dataset 1 (GastroNet)	WLE overview (eg, colon, stomach, duodenum, esophagus)	494,364	Hand-labeled subset by experts, followed by automatic pseudo-labeling	Pretraining
Dataset 2	WLE overview	1247		Refinement training
Neoplastic images		690	Delineated by 2 experts	
NDBE images		557	Hand-labeled by 2 experts	
Dataset 3	NBI zoom images	183		Refinement training and internal validation
Neoplastic images		112	Correlating pathology	
NDBE images		71	Correlating pathology	
Dataset 4	NBI zoom videos	157		External validation
Neoplastic videos		59	Correlating pathology	
NDBE videos		98	Correlating pathology	

CADx, Computer-aided diagnosis; *WLE*, white-light endoscopy; *NDBE*, nondysplastic Barrett's esophagus; *NBI*, narrow-band imaging.

patients. If more than 1 region of interest was selected per patient, every NBI image was obtained at a different area in the BE segment. All images were collected in previous studies by our group and had corresponding histology of that specific area confirmed by an expert BE pathologist.[21-23] The CADx system was trained and internally validated on dataset 3, using 4-fold cross-validation methodology.

Dataset 4: NBI zoom videos. We hypothesized that video analysis would enhance CAD performance. Therefore, we further trained and tested the CADx system on 98 nondysplastic NBI zoom videos of 33 patients with BE and 59 neoplastic NBI zoom videos of 17 patients with BE. These were all unaltered NBI zoom videos; that is, no human processing or manual video-frame selection was performed. All NBI zoom videos contained corresponding histopathology of each area and were confirmed by an expert BE pathologist (M.V.). The videos were obtained using the same protocol in the Amsterdam UMC and Karolinska University Hospital by 3 endoscopists with experience in zoom endoscopy (J.B., R.P., and F.B.S.). Only videos from flat-type mucosa were selected (Fig. 1). The average duration of each NBI video was 10 seconds, each consisting of approximately 250 video frames.

Automated video analysis

First, we trained our CADx system using the NBI still images from dataset 3, resulting in an image-based performance. We envisioned increasing this image-based performance using automated video analysis because videos contain multiple frames and thereby additional temporal (and spatial) information in contrast to single still images. In the video analysis, the system automatically created 1 single prediction for each video by calculating the average of all frame predictions, in contrast to calculating the likelihood of neoplasia per individual frame. To assess the incremental value of this approach, we calculated the increase in CADx system performance using automated video analysis against its performance per individual frame. In both analyses, the threshold for a NBI zoom image or video to be scored as neoplastic was 70%. For example, in the still image analysis, if an NBI zoom image contained a likelihood of neoplasia $\geq$70%, this image was considered neoplastic. In the video analysis, this means that a video was scored neoplastic if the average of all frames within a video contained a likelihood of neoplasia $\geq$70%. The optimal neoplasia cutoff value of 70% was determined based on the NBI image dataset, and this value was subsequently used to evaluate the performance metrics on the NBI video dataset. Figure 2 provides 3 examples of NBI zoom videos and their corresponding likelihood of neoplasia prediction and illustrates the temporal information available in the analysis of subsequent video frames.

Comparison of pretraining strategies

To assess the incremental value of different pretraining strategies, we analyzed the same NBI zoom videos using a CADx system with pretraining using GastroNet (ie, endoscopy-driven pretraining), pretraining using ImageNet (ie, generic imagery for pretraining), and a CADx system without pretraining. Architecture design and associated hyper parameters of these different CADx systems were identical to allow for a correct comparison. Performance of the CADx system was evaluated using the receiver operating characteristic (ROC) curve and corresponding area under the curve (AUC).

In a second analysis, we assessed the true confidence estimate of these different CADx systems when assessing

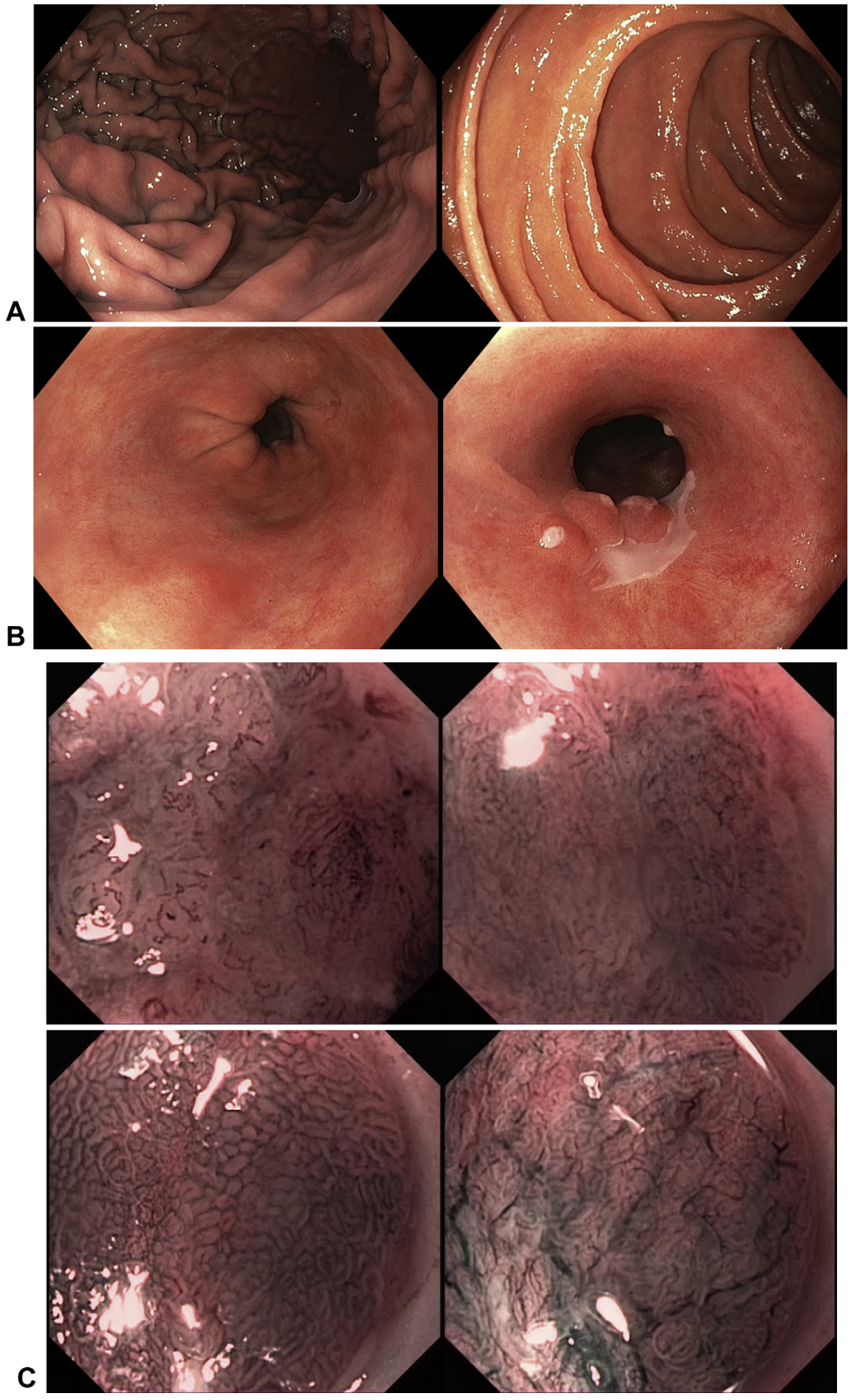

Figure 1. **A**, Colon and gastric images derived from dataset 1. **B**, Nondysplastic and neoplastic Barrett white-light overview images derived from dataset 2. **C**, Nondysplastic and neoplastic Barrett narrow-band imaging zoom images derived from dataset 3.

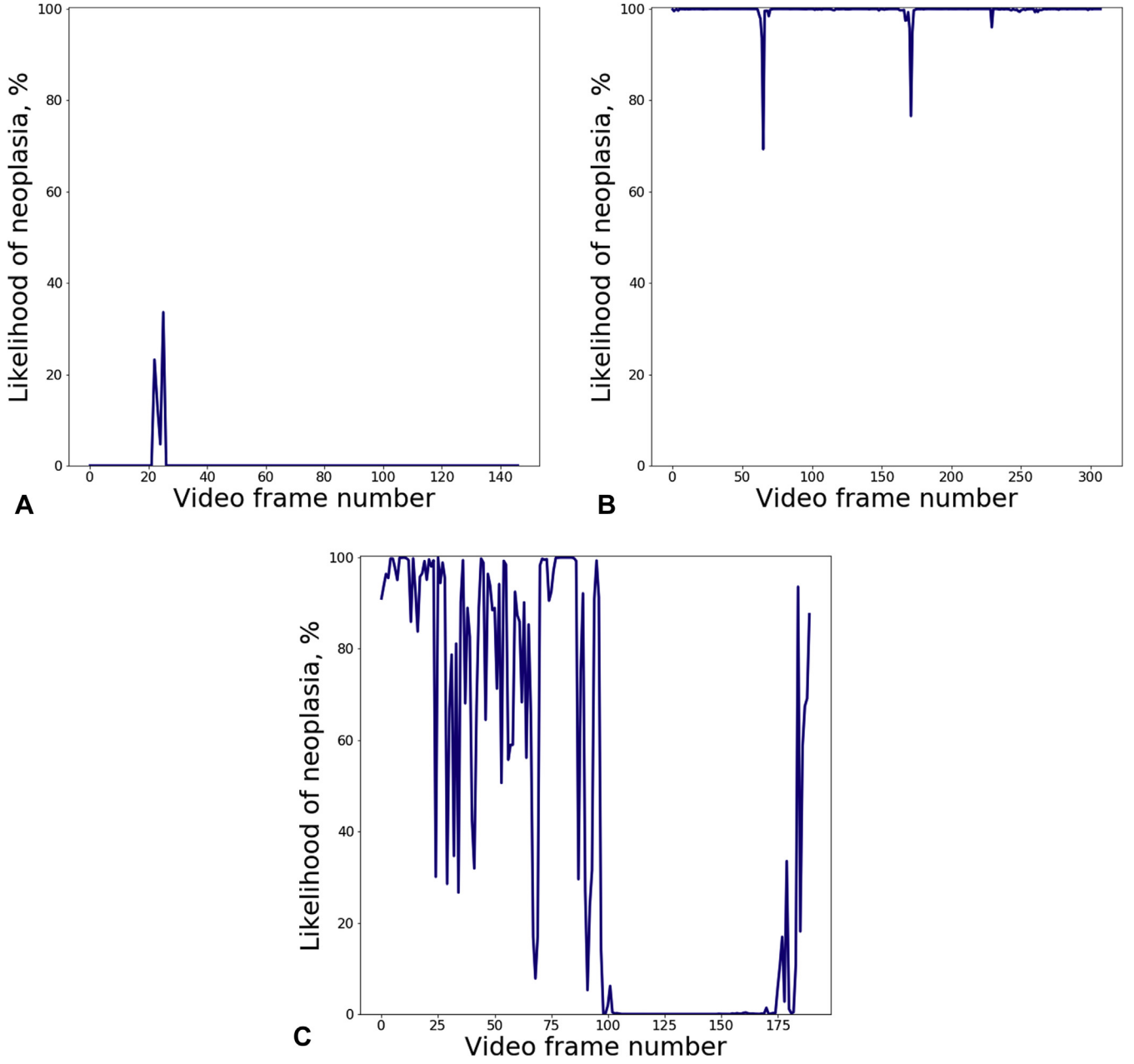

Figure 2. Temporal information on subsequent narrow-band imaging (NBI) zoom video frames and their likelihood of neoplasia prediction. **A**, Nondysplastic NBI video with a correct nondysplastic Barrett's esophagus (NDBE) video diagnosis after application of automated video analysis. Almost all video frames show a correct nondysplastic diagnosis with a 0% likelihood of neoplasia. **B**, Neoplastic NBI video with a correct neoplasia video diagnosis after application of automated video analysis. Almost all video frames show a correct neoplastic diagnosis with a 100% likelihood of neoplasia. **C**, Neoplastic NBI video with an incorrect NDBE video diagnosis. Video frames 100 to 175 showed a large section of the video with an incorrect diagnosis, most likely due to an out-of-focus lens and thus the presence of multiple low-quality non-informative frames.

an NBI video. The confidence estimate was defined by the uncalibrated confidence of the system for a given diagnosis from 0% to 100%. For a neoplastic video, the most correct neoplastic diagnosis was 100%. For a nondysplastic video, the most correct nondysplastic diagnosis was also 100%, calculated by 100, the likelihood for neoplasia. For example, if the system correctly provided a neoplastic video with a high likelihood of neoplasia of 90%, the true confidence estimate was 90%. If the system incorrectly provided a nondysplastic video with a high likelihood of neoplasia of 90%, the true confidence estimate was 100% – 90% = 10%. When this analysis was used, the system's prediction was evaluated in a quantitative manner compared with a binominal analysis of correct versus incorrect based on a cutoff value of 70%.

Outcome measures

Primary outcome measure.

- Diagnostic performance (accuracy, sensitivity, and specificity) of the CADx system for characterization of neoplastic BE on NBI zoom images (dataset 3)
- Diagnostic performance (accuracy, sensitivity, and specificity) of the CADx system for characterization of neoplastic BE on NBI zoom videos (dataset 4), assessed per individual frame and per video

Secondary outcome measures.

- Diagnostic performance on NBI zoom videos using different pretraining strategies (ie, endoscopy-driven pretraining, pretraining using ImageNet, and no pretraining)
- Assessment time of the CADx system for NBI zoom imagery

Statistical analysis

Diagnostic performance of the system per image and per video was calculated in terms of accuracy, sensitivity, specificity using the histology as criterion standard. A fourfold cross-validation methodology was used to assess performance on images and videos from datasets 3 and 4. When this approach is used, data are randomly split into 4 equal parts, after which 1 part is used as a test set and the other 3 parts are used for training. This is repeated a total of 4 times, each time with a different test set. Subsequently, the scores of the 4 experiments are pooled into a point estimate, yielding a result that is more robust against data variation. The Wilcoxon signed-rank test was used to allow for a paired comparison of the diagnostic accuracy of the CADx system using the 3 different pretraining strategies (ie, endoscopy-driven pretraining, pretraining using ImageNet, and no pretraining). Because this was the first study of its kind, developing a deep-learning CADx system for characterization of BE neoplasia using NBI videos, no formal sample size calculation was conducted. The CADx system was developed in the PyTorch framework and statistical tests were performed with the Matlab 2018a software package (Mathworks, Inc, Natick, Mass, USA).

RESULTS

Primary outcome measurements

Diagnostic performance on NBI zoom images (dataset 3). The CADx system was first trained and tested on NBI images from dataset 3, resulting in an accuracy of 84% (95% confidence interval [CI], 81%-88%), with a corresponding sensitivity of 88% (95% CI, 86%-94%) and specificity of 78% (95% CI, 72%-84%) for correct differentiation between NDBE and BE neoplasia.

Diagnostic performance on NBI zoom videos (dataset 4) per individual video frame. In total, 30,021 individual video frames were analyzed by the CADx system. Accuracy, sensitivity, and specificity of the CADx system for characterization of BE neoplasia calculated per individual frame were 85% (25,400 of 30,021), 75% (7698 of 10,265), and 90% (17,702 of 19,756), respectively.

TABLE 2. Diagnostic accuracy of the computer-aided diagnosis system for differentiating between nondysplastic and neoplastic Barrett's esophagus using 157 narrow-band imaging (NIB) zoom videos

	Accuracy, % (n/N)	Sensitivity, % (n/N)	Specificity, % (n/N)
Fold1	92 (23/25)	71 (5/7)	100 (18/18)
Fold2	74 (46/62)	88 (22/25)	65 (24/37)
Fold3	84 (27/32)	79 (11/14)	89 (16/18)
Fold4	92 (35/38)	92 (12/13)	92 (23/25)
Total	83 (131/157)	85 (50/59)	83 (81/98)

Statistical analysis performed using 4-fold cross-validation. NBI zoom videos obtained from the same patient were always allocated to either the training or the test set (eg, patient-split analysis).

Diagnostic performance on NBI zoom videos (dataset 4) per video. In a per patient analysis, diagnostic accuracy, sensitivity, and specificity of the CADx system for characterization of BE neoplasia were 83% (95% CI, 78%-89%), 85% (95% CI, 76%-94%), and 83% (95% CI, 76%-90%), respectively. These results are summarized in Table 2.

Figure 2A and B shows the likelihood of a correct diagnosis of neoplasia for a NBI zoom video using automated video analysis. Figure 2C shows the likelihood of neoplasia for a video with an incorrect NDBE diagnosis.

Secondary outcome measures

Comparison of pretraining strategies. The CADx system without any form of pretraining resulted in an accuracy of 75% (118 of 157), sensitivity of 75% (44 of 59), and specificity of 76% (74 of 98). The CADx system with pretraining using ImageNet demonstrated an accuracy, sensitivity, and specificity of 82% (128 of 157), 80% (47 of 59), and 83% (81 of 98), respectively. Finally, the CADx system with endoscopy-driven pretraining demonstrated an accuracy, sensitivity, and specificity of 83% (131 of 157), 85% (50 of 59), and 83% (81 of 98), respectively. Endoscopy-driven pretraining resulted in an AUC of 0.92 compared with AUC of 0.88 with pretraining using ImageNet and an AUC of 0.87 with no pretraining. See Figure 3 for the corresponding ROC curves.

In the second analysis assessing the true confidence estimate, the CADx system, which received endoscopy-driven pretraining using GastroNet, resulted in a median accuracy of 89% (interquartile range [IQR], 67%-97%) compared with 83% (IQR, 62%-89%) with pretraining using ImageNet and 65% (IQR, 47%-82%) with no pretraining. The CADx system, which received endoscopy-driven pretraining, significantly outperformed both pretraining with ImageNet and no pretraining, as shown in Figure 4 ($P < .001$).

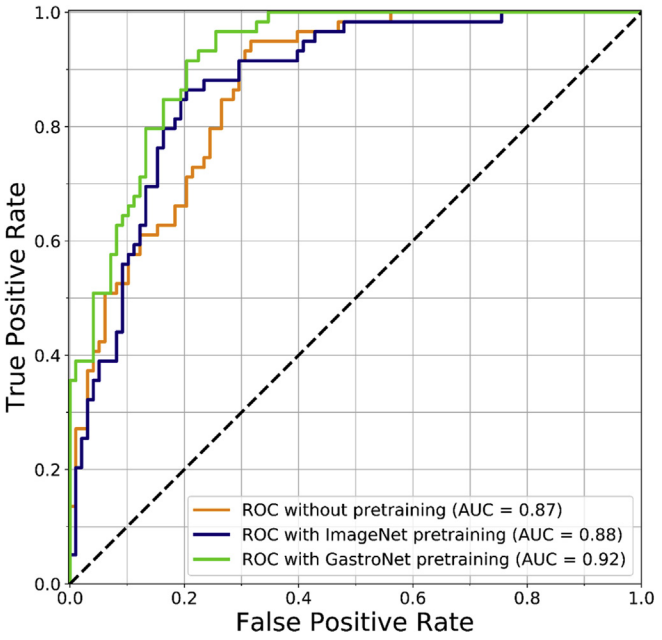

Figure 3. Receiver operating curves (ROC) of 3 different computer-aided diagnosis (CADx) system architectures. A CADx system with endoscopy-driven pretraining was compared with a CADx system with ImageNet pretraining and a CADx system that did not receive pretraining. *AUC*, Area under the curve.

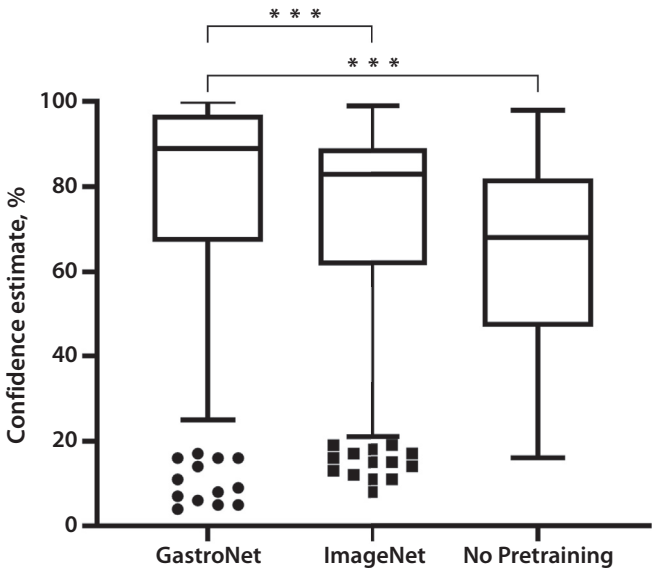

Figure 4. Box and whiskers plot comparing 3 different pretraining strategies. A computer-aided diagnosis (CADx) system architecture with endoscopy-driven pretraining including GastroNet was compared with a CADx system with ImageNet pretraining and with no pretraining. All algorithms were trained and evaluated using the same narrow-band imaging zoom videos. The CADx system that received endoscopy-driven pretraining using GastroNet significantly outperformed both pretraining with ImageNet and no pretraining ($P < .001$, Wilcoxon signed-rank test).

Assessment time. The mean assessment time per NBI zoom video with a duration of 10 seconds was 5.1 seconds (standard deviation, ±2.6 seconds). Mean assessment time per video frame was 0.026 seconds (SD, ±0.0024 seconds). The CADx system was able to process 38 frames per second.

DISCUSSION

Endoscopic diagnosis of early BE neoplasia is generally a 2-step process of primary detection in overview, followed by secondary characterization of any visible abnormalities with NBI used by most endoscopists. Endoscopists may struggle in evaluating NBI zoom imagery for subtle abnormalities, because NBI classifications are either too crude (regular vs irregular) or consist of too many subclassifications; all of them resulting in a subjective assessment. In adjunct to our work on developing a CAD system using WLE overview imagery, further characterization of suspicious areas might benefit from a computer-assisted diagnosis system using NBI zoom imagery.

In this study, diagnostic accuracy of the CADx system on NBI zoom images was reasonable, yet specificity remained suboptimal. Presumably, this was caused by the relatively small number of images available for training. In contrast, video analysis provides a vast increase in the amount of data available for training, displays temporal information available in subsequent video frames, and has the potential advantage of automatically excluding multiple frames of low quality. Our CADx system demonstrated promising diagnostic accuracy, sensitivity, and specificity using unaltered NBI videos. In particular, the specificity of the video-based CADx system increased when compared with the image-based system. Our video-based CADx system correctly classified 81 of 98 NDBE videos, demonstrating a specificity of 83%. High specificity is vital for successful clinical implementation of a CADx system for BE characterization. During surveillance endoscopies, endoscopists will most likely first be alerted to abnormal areas in overview, after which they will interrogate this area in detail for the presence of neoplasia using NBI. A characterization CADx system with high specificity would then be able to decrease the number of false-positive predictions by endoscopists, and thus decrease unnecessary biopsies.

Although the diagnostic performance of the current CADx system was promising, the current results do not yet support its application in clinical practice. A possible explanation may be that the collection of NBI zoom videos was performed in multiple hospitals, by different endoscopists and endoscopes, resulting in different levels of image quality and levels of magnification. In particular, the collection of NBI zoom images and videos was performed using an Olympus ME-NBI endoscope (GIF-Q160Z; maximal magnification, ×115) and a GIF-HQ190 endoscope with near-focus mode. Due to the relatively small sample size of videos used in this pilot study, the system may not have been robust enough against this background

TABLE 3. Overview of the limitations of our study and the corresponding solutions that we will address in future research

Limitations	Solutions
Relatively small sample size of NBI zoom videos	Additional data collection
No separate external validation dataset	The CADx algorithm has to be tested using a separate external validation dataset collected with the new-generation NBI scopes to ascertain the performance
Low-grade dysplasia was not analyzed	Include low-grade dysplasia cases in future studies
No comparison with physician interpretation	Offer an independent external set of NBI images and videos to endoscopists with different levels of endoscopic expertise and relate their performance to that of the CADx system
NBI images and videos were captured with 2 different Olympus endoscopes (Q160Z /HQ190) and thus different levels of magnification. A CADx system may possibly learn to discriminate based on image quality rather than neoplastic features	The CADx system has to be tested using a separate external validation dataset collected with the new-generation NBI scopes to ascertain the performance. Robustness of the CADx system may be improved by incorporating images of different quality and different NBI systems available
Optical zoom techniques, in contrast to NBI zoom (near-focus), is mainly used by the Japanese and would limit its applicability in western countries	Demonstrate the additional value of a fast, accurate, and easy-to-use NBI zoom (near-focus) system when used by general endoscopists in western countries
The diagnostic performance of the NBI-CADx system cannot be compared with CADe systems focusing on primary detection in WLE overview. This system serves a different purpose.	The NBI-CADx system in the current study serves as a characterization tool after primary detection by the endoscopist or a WLE-CADe system, red flagging areas of interest as potentially neoplastic. The post-test likelihood of an NBI image being neoplastic therefore is a result of 2 sequential diagnostic processes
Out-of-focus NBI images were included in the training of the CADx system which may have affected the performance of the algorithm.	In future CAD systems, we will exclude such non-informative frames to optimize overall performance
The cutoff of 70% in the call for neoplasia in the NBI image/video analysis was arbitrary and not validated based on a performance characteristic analysis compared with the criterion standard	Validate the cutoff in the call for neoplasia on a performance characteristic analysis compared with the criterion standard

NBI, Narrow-band imaging; *CADx*, computer-aided diagnosis; *WLE*, white-light endoscopy; *CADx*, computer-aided detection.

noise in our data. We do feel that the introduction of heterogeneous data is positive for the generalizability of the algorithm, because once the system is ready for use in clinical practice, it may be used by general endoscopists from different countries using different endoscopes and levels of magnification. In our ongoing studies, we are focusing on current and next-generation NBI systems in which we try to balance uniformity and heterogeneity of the imagery on one side with performance and robustness of our algorithms on the other. We observed that the diagnostic performance of our system decreased when using a patient-split analysis; that is, taking into account that all videos obtained from the same patient are always allocated to either the training or the test set. This study illustrates the importance of performing a patient-split analysis, which is not always routinely performed in CAD research. Finally, upon review of the videos with an incorrect CADx diagnosis, we noticed that these frequently showed decreased image quality with blurring, specular reflections, and movement of the lens, which may have compromised performance. In future CAD systems, we will therefore exclude such noninformative frames to optimize overall performance. It is expected that additional data collection and future refinement on external datasets may further increase the diagnostic value of CADx systems and ultimately improve their effectiveness for surveillance of BE.

In addition, we calculated the assessment time of the CADx system when analyzing the NBI zoom videos. The CADx system was able to analyze 38 video frames per second. This is approximately 1.5 times faster than required for real-time video processing (ie, 25 frames per second). Therefore, our execution speed enables real-time video analyses. Given this execution speed, the diagnostic performance may be increased in the future when noninformative, low-quality frames are automatically excluded by the CADx system.

We envisioned that specific, endoscopy-driven pretraining would increase the diagnostic performance of the CADx system compared with pretraining with of images of unrelated objects, such as in the ImageNet dataset. Therefore, we compared the performance of the CADx system after endoscopy-driven pretraining using GastroNet with that of a CADx system with pretraining using ImageNet and a CADx system without any form of pretraining. Endoscopy-driven pretraining significantly outperformed both other CADx systems with their associated pretraining strategies. Based on these results, we hypothesize that endoscopy-driven pretraining, using large databases of

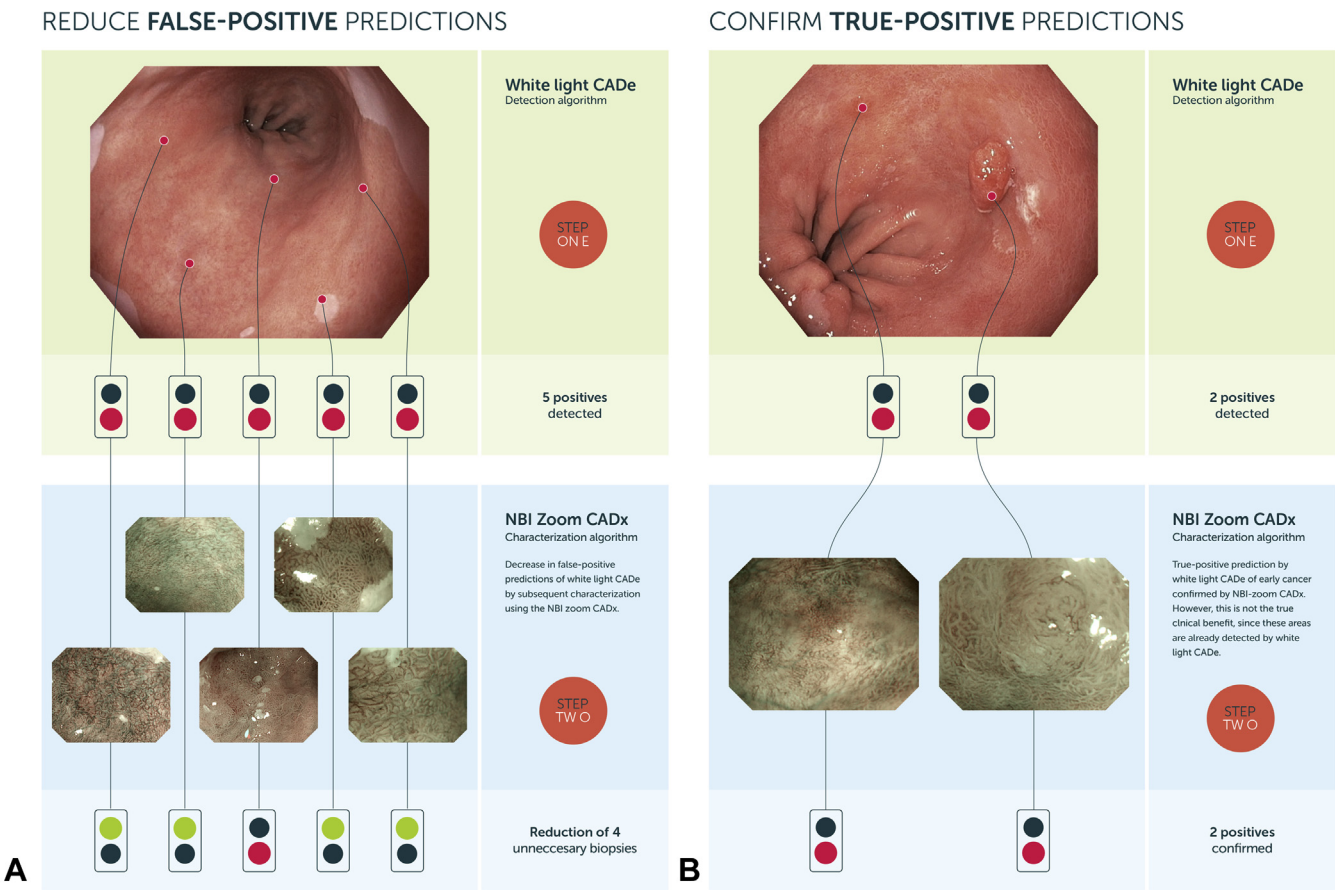

Figure 5. A, Reduction of false-positive predictions using a combination of white-light CADe and subsequent characterization using NBI zoom CADx. **B**, Confirmation of true-positive predictions using a combination of white-light CADe and subsequent characterization using NBI zoom CADx. *CADe*, computer-aided detection; *CADx*, computer-aided diagnosis; *NBI, narrow-band imaging*.

endoscopic imagery, may offer a promising alternative for the development of different deep-learning CAD systems in endoscopy. Future studies should investigate the feasibility of this approach for other endoscopic classification and localization problems and compare this against CAD systems that received nonendoscopy driven pretraining such as ImageNet.

Our study has several strengths. It is the first video-based CADx system for characterization of BE neoplasia using a unique multi-stage setup of endoscopy-driven pretraining (including GastroNet). In addition, all NBI imagery was matched with corresponding histology that was evaluated by expert pathologists. Our study also has potential limitations (Table 3). First, we analyzed nondysplastic and neoplastic NBI zoom imagery only, reflecting the more obvious pathologic cases. Images from low-grade dysplasia were not included in the study. This may hamper the extrinsic validity on larger datasets with a wider variety in pathologic diagnosis. Second, we have not evaluated the system's performance using a separate external validation dataset. Third, analysis was performed on retrospective images and videos. Fourth, no formal sample size calculation was performed given the lack of a performance point estimate. Fifth, out-of-focus NBI images were included in the training of the CADx system, which may have affected the performance of the algorithm. Sixth, the cutoff of 70% in the call for neoplasia in the NBI image/video analysis was arbitrary and not validated based on a performance characteristic analysis compared with the criterion standard. Sixth, we did not compare the performance of our system with the assessment of endoscopists. Future studies should compare CADx system performance with that of expert and nonexpert endoscopists.

Our group has recently published on a WLE-CADe system that assists endoscopists in the primary detection of early Barrett's neoplasia using WLE overview images.[18] The NBI-CADx described in the current study serves a different purpose: it allows endoscopists to characterize areas of interest as either normal or neoplastic. We envision that these 2 CAD systems will work in conjunction: the WLE-CADe system will direct endoscopists to areas of interest, red flagging areas with an abnormal white-light appearance, which can then be further characterized using the NBI-CADx system, as shown in Figure 5. The 2 CAD systems are trained and tested using different types of imagery (ie, WLE overview images for the WLE-CADe

system versus detailed NBI images for the NBI-CADx system), and therefore their performance characteristics cannot be compared. In addition, because the 2 systems serve a different purpose, their performance characteristics should indeed be different: the most relevant performance characteristic of the WLE-CADe is a high sensitivity to avoid missing neoplastic areas; for the NBI-CADx system, the importance is in reducing the number of unnecessary biopsies from false-positive areas red flagged by the WLE-CADe system that are clearly non-neoplastic, hence, a high specificity is the most dominant performance characteristic here.

In conclusion, we have developed a CADx system for NBI-based characterization of areas within a BE, which may work in conjunction with a primary detection CADe system.[18] Future work will focus on optimizing our current CADx system by automatically excluding low-quality frames, validation of our results using a separate prospective NBI zoom dataset containing a larger variety of pathologic diagnoses, and benchmarking results against endoscopists' performance.

REFERENCES

1. Shaheen NJ, Falk GW, Iyer PG, et al. ACG clinical guideline: diagnosis and management of Barrett's esophagus. Am J Gastroenterol 2016;111:30-50.
2. Belghazi K, Bergman J, Pouw RE. Endoscopic resection and radiofrequency ablation for early esophageal neoplasia. Dig Dis 2016;34:469-75.
3. Pech O, May A, Manner H, et al. Long-term efficacy and safety of endoscopic resection for patients with mucosal adenocarcinoma of the esophagus. Gastroenterology 2014;146:652-60.e1.
4. Weusten B, Bisschops R, Coron E, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2017;49:191-8.
5. Gordon LG, Mayne GC, Hirst NG, et al. Cost-effectiveness of endoscopic surveillance of non-dysplastic Barrett's esophagus. Gastrointest Endosc 2014;79:242-56.e6.
6. Tschanz ER. Do 40% of patients resected for barrett esophagus with high-grade dysplasia have unsuspected adenocarcinoma? Arch Pathol Lab Med 2005;129:177-80.
7. Kara MA, Ennahachi M, Fockens P, et al. Detection and classification of the mucosal and vascular patterns (mucosal morphology) in Barrett's esophagus by using narrow band imaging. Gastrointest Endosc 2006;64:155-66.
8. Nogales O, Caballero-Marcos A, Clemente-Sánchez A, et al. Usefulness of non-magnifying narrow band imaging in EVIS EXERA III video systems and high-definition endoscopes to diagnose dysplasia in Barrett's esophagus Using the Barrett International NBI Group (BING) classification. Dig Dis Sci 2017;62:2840-6.
9. Alvarez Herrero L, Curvers WL, Bansal A, et al. Zooming in on Barrett oesophagus using narrow-band imaging: an international observer agreement study. Eur J Gastroenterol Hepatol 2009;21:1068-75.
10. Singh M, Bansal A, Curvers WL, et al. Observer agreement in the assessment of narrowband imaging system surface patterns in Barrett's esophagus: a multicenter study. Endoscopy 2011;43:745-51.
11. Silva FB, Dinis-Ribeiro M, Vieth M, et al. Endoscopic assessment and grading of Barrett's esophagus using magnification endoscopy and narrow-band imaging: accuracy and interobserver agreement of different classification systems (with videos). Gastrointest Endosc 2011;73:7-14.
12. Baldaque-Silva F, Marques M, Lunet N, et al. Endoscopic assessment and grading of Barrett's esophagus using magnification endoscopy and narrow band imaging: impact of structured learning and experience on the accuracy of the Amsterdam classification system. Scand J Gastroenterol 2013;48:160-7.
13. Gross S, Trautwein C, Behrens A, et al. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. Gastrointest Endosc 2011;74:1354-9.
14. Chen P-J, Lin M-C, Lai M-J, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018;154:568-75.
15. Takeda K, Kudo S-E, Mori Y, et al. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. Endoscopy 2017;49:798-802.
16. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018;155:1069-78.e8.
17. de Groof J, van der Sommen F, van der Putten J, 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 Gastroenterol J 2019;7:538-47.
18. de Groof AJ, Struyvenberg MR, van der Putten J, 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 2020;158:915-29.e4.
19. Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009. p, 248-55.
20. van der Putten J, de Groof J, van der Sommen F, et al. Pseudo-labeled bootstrapping and multi-stage transfer learning for the classification and localization of dysplasia in Barrett's esophagus. In: Suk HI, Liu M, Yan P, et al. (editors). Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science, vol. 11861. Cham: Springer; 2019. p. 169-77.
21. Kara MA, Peters FP, Fockens P, et al. Endoscopic video-autofluorescence imaging followed by narrow band imaging for detecting early neoplasia in Barrett's esophagus. Gastrointest Endosc 2006;64:176-85.
22. Curvers WL, Alvarez Herrero L, Wallace MB, et al. Endoscopic tri-modal imaging is more effective than standard endoscopy in identifying early-stage neoplasia in Barrett's esophagus. Gastroenterology 2010;139:1106-14.
23. Curvers W, Baak L, Kiesslich R, et al. Chromoendoscopy and narrow-band imaging compared with high-resolution magnification endoscopy in Barrett's esophagus. Gastroenterology 2008;134:670-9.