



# Prospective development and validation of a volumetric laser endomicroscopy computer algorithm for detection of Barrett's neoplasia

Maarten R. Struyvenberg, MD,<sup>1</sup> Albert J. de Groof, MD,<sup>1</sup> Roger Fonollà, MSc,<sup>2</sup> Fons van der Sommen, PhD,<sup>2</sup> Peter H. N. de With, PhD,<sup>2</sup> Erik J. Schoon, MD, PhD,<sup>3</sup> Bas L. A. M. Weusten, MD, PhD,<sup>4</sup> Cadman L. Leggett, MD, PhD,<sup>5</sup> Allon Kahn, MD,<sup>6</sup> Arvind J. Trindade, MD,<sup>7</sup> Eric K. Ganguly, MD,<sup>8</sup> Vani J. A. Konda, MD,<sup>9</sup> Charles J. Lightdale, MD,<sup>10</sup> Douglas K. Pleskow, MD,<sup>11</sup> Amrita Sethi, MD,<sup>12</sup> Michael S. Smith, MD, MBA,<sup>13</sup> Michael B. Wallace, MD, MPH,<sup>14</sup> Herbert C. Wolfsen, MD,<sup>14</sup> Gary J. Tearney, MD, PhD,<sup>15</sup> Sybren L. Meijer, MD, PhD,<sup>16</sup> Michael Vieth, MD, PhD,<sup>17</sup> Roos E. Pouw, MD, PhD,<sup>1</sup> Wouter L. Curvers, MD, PhD,<sup>3</sup> Jacques J. Bergman, MD, PhD<sup>1</sup>

Amsterdam, Eindhoven, Nieuwegein, the Netherlands; Rochester, Minnesota; Scottsdale, Arizona; New Hyde Park, New York; Burlington, Vermont; Dallas, Texas; New York, New York; Boston, Massachusetts; Jacksonville, Florida, USA; Bayreuth, Germany

**Background and Aims:** Volumetric laser endomicroscopy (VLE) is an advanced imaging modality used to detect Barrett's esophagus (BE) dysplasia. However, real-time interpretation of VLE scans is complex and time-consuming. Computer-aided detection (CAD) may help in the process of VLE image interpretation. Our aim was to train and validate a CAD algorithm for VLE-based detection of BE neoplasia.

**Methods:** The multicenter, VLE PREDICT study, prospectively enrolled 47 patients with BE. In total, 229 nondysplastic BE and 89 neoplastic (high-grade dysplasia/esophageal adenocarcinoma) targets were laser marked under VLE guidance and subsequently underwent a biopsy for histologic diagnosis. Deep convolutional neural networks were used to construct a CAD algorithm for differentiation between nondysplastic and neoplastic BE tissue. The CAD algorithm was trained on a set consisting of the first 22 patients (134 nondysplastic BE and 38 neoplastic targets) and validated on a separate test set from patients 23 to 47 (95 nondysplastic BE and 51 neoplastic targets). The performance of the algorithm was benchmarked against the performance of 10 VLE experts.

**Results:** Using the training set to construct the algorithm resulted in an accuracy of 92%, sensitivity of 95%, and specificity of 92%. When performance was assessed on the test set, accuracy, sensitivity, and specificity were 85%, 91%, and 82%, respectively. The algorithm outperformed all 10 VLE experts, who demonstrated an overall accuracy of 77%, sensitivity of 70%, and specificity of 81%.

**Conclusions:** We developed, validated, and benchmarked a VLE CAD algorithm for detection of BE neoplasia using prospectively collected and biopsy-correlated VLE targets. The algorithm detected neoplasia with high accuracy and outperformed 10 VLE experts. (The Netherlands National Trials Registry (NTR) number: NTR 6728.) (Gastrointest Endosc 2021;93:871-9.)

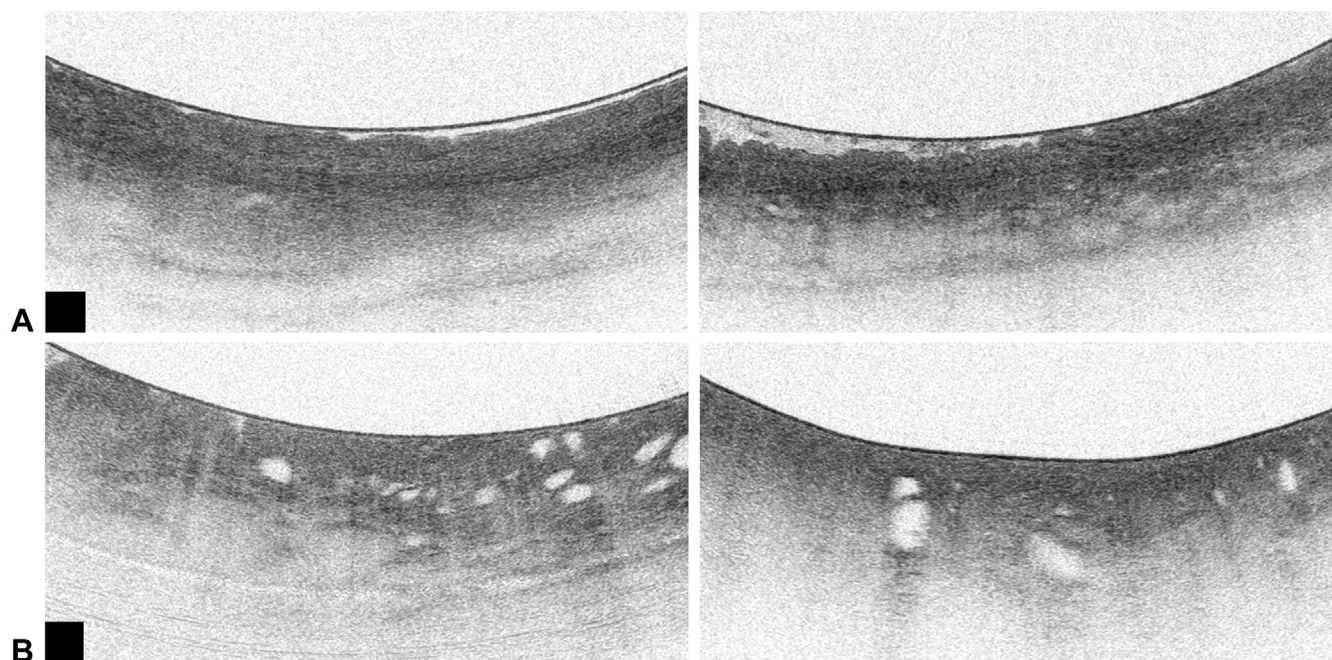
(footnotes appear on last page of article)

## INTRODUCTION

Patients with Barrett's esophagus (BE) are at risk of developing esophageal adenocarcinoma and require endoscopic surveillance for detection and treatment of early neoplasia, including high-grade dysplasia and intramucosal adenocarcinoma.<sup>1,2</sup> However, current BE surveillance practices are suboptimal because early neoplasia can be missed due to its subtle

endoscopic appearance and sampling error of random biopsies.<sup>3,4</sup>

Volumetric laser endomicroscopy (VLE) has the potential to identify and mark areas suspicious for early BE neoplasia not appreciated under high-definition white-light endoscopy. VLE uses second-generation optical coherence tomography technology to create cross-sectional images of tissue, based on differences in optical scattering of different tissue structures. In 90 seconds,



**Figure 1.** Volumetric laser endomicroscopy images of different tissue types. **A**, Nondysplastic BE with clear mucosal layering and irregular surface. **B**, High-grade dysplasia showing multiple irregular glands and lack of layering; esophageal adenocarcinoma with irregular glands and high surface signal intensity compared with subsurface intensity.

this balloon-based system makes a scan to circumferentially visualize surface and subsurface esophageal layers in microscopic resolution (Fig. 1). The system incorporates a VLE-guided laser marking tool that allows targeted histologic sampling of areas of interest. VLE scans are composed of 1200 sequential gray-scale images encompassing a longitudinal span of 6 cm. Due to subtle imaging features, VLE interpretation is complex and time-consuming, thus limiting application of VLE during endoscopy. The addition of computer-aided detection (CAD) may overcome this limitation and may enhance the potential of this imaging technology to improve BE surveillance.

Recent ex vivo VLE studies have focused on CAD algorithms for detection of BE neoplasia, based on single images or small image stacks of transverse regions of interest (ROIs), showing reasonable accuracy.<sup>5,6</sup> In contrast, human VLE users have shown varying degrees of interobserver agreement and diagnostic accuracy in differentiating early neoplasia from nondysplastic BE.<sup>7–11</sup> These studies have been hampered by the lack of a criterion standard diagnosis for the VLE images: VLE-guided laser marking was not available and direct corresponding histopathology was therefore lacking. For VLE to be successfully implemented into clinical practice, VLE interpretation should be fast and rely on high diagnostic accuracy for detection of BE neoplasia. The aim of this study was, therefore, to create a CAD algorithm for VLE-guided identification of early BE neoplasia using in vivo, biopsy-correlated

VLE targets and to compare CAD performance with the performance of 10 recognized VLE experts.

## METHODS

### Setting

This prospective multicenter study for the PREDICTion of BE neoplasia (PREDICT study) was performed at the departments of gastroenterology and hepatology of the Amsterdam University Medical Centers, location Academic Medical Center, Catharina Hospital Eindhoven and St. Antonius Hospital Nieuwegein, all tertiary referral centers for Barrett's neoplasia in the Netherlands, and at the department of Electrical Engineering of the Eindhoven University of Technology, the Netherlands. The study was approved by the Institutional Review Board of the Amsterdam University Medical Centers (protocol NTR 6728, registered at <http://www.trialregister.nl>) and written informed consent was obtained from all patients.

### Patient and VLE image database

Patients undergoing surveillance of nondysplastic Barrett's esophagus (NDBE), or referred for work-up and treatment of early BE neoplasia (high-grade dysplasia and/or esophageal adenocarcinoma), were eligible for this study. Patients enrolled in this study did not have a history of endoscopic eradication therapy for BE. Patients with erosive esophagitis, significant stenosis of the esophagus, and esophageal tears or ulcers were excluded.

## Endoscopic and VLE procedure

Endoscopic procedures were performed with the Olympus HQ190 endoscope by 4 endoscopists (J.B., R.P., W.C., B.W.) with extensive experience in the use of advanced imaging modalities and endoscopic treatment of early BE neoplasia. Length of the BE segment was recorded according to the Prague C&M classification.<sup>12</sup> In cases with a visible lesion, overview and detailed images were obtained. Location was measured as distance from the incisors and clockwise orientation, size was estimated by lesion diameter, and lesions were categorized according to the Paris classification.<sup>13,14</sup>

Subsequently, the balloon-based optical probe was introduced through the working channel of the endoscope and positioned in the distal esophagus. The spatial orientation in the VLE scan was provided by the gastroesophageal junction, cautery marks placed at the 12 o'clock position at 2 cm intervals, and the balloon registration line (black line, which is visible endoscopically on the balloon and on the VLE scan). After initial imaging scans, the BE segment was systematically laser marked every 2 cm at 3, 6, 9, and 12 o'clock, similar to the Seattle protocol for random biopsies. These superficial marks (ie, laser marks) were placed at the position of the Seattle protocol biopsies and were not targeted at suspicious VLE areas. This was done to safeguard objective collection of VLE targets to train the CAD algorithm.

After the systematic laser marking procedure, a final VLE scan was performed, which was used to train the CAD algorithm. The location of every laser-marked target was systematically recorded on the case record form at the time of the procedure by consensus of 2 VLE research fellows (M.S., J.G.) and 1 expert endoscopist. After completion of the VLE procedure, targeted biopsy specimens were individually obtained across the laser-marked areas (Fig. 2) and were submitted for histopathologic interpretation in independent jars. The histopathologic correlation to the VLE scan was assumed to apply over 2.5 mm<sup>2</sup> of the VLE scan, comprising 25 cross-sectional images proximal and 25 cross-sectional images distal to the laser marking frame, resulting in 51 images per VLE target. All histology slides were interpreted by consensus of 2 expert BE pathologists (S.M. and M.V.).

## Design of the computer-aided detection algorithm

In our study, we used an ensemble of deep convolutional neural networks, based on the VGG16 architecture.<sup>15</sup> A total of 3 deep convolutional neural networks were used to build the ensemble. Each network was pretrained with ImageNet to learn basic discriminative image features.<sup>16</sup> Subsequently, these networks were fine-tuned using the PREDICT training set by means of transfer learning<sup>17</sup> to differentiate between nondysplastic and neoplastic BE tissue. A 3-fold cross-validation approach

was used to create a diverse ensemble and estimate its performance on the training data. For data augmentation during training, we chose to apply a combination of horizontal flip, motion blur, and optical distortion, which best represent the behavior of VLE images. For exact details of the construction of the deep learning CAD algorithm, refer to the technical publication by our group.<sup>18</sup> A schematic representation of our CAD algorithm is shown in Figure 3.

The training dataset of the CAD algorithm consisted of VLE targets of the first 22 patients, imaged between October 2017 and May 2018. During this phase we trained the algorithm and optimized the hyperparameters until convergence was achieved. Subsequently, we used VLE targets of the second 25 patients, imaged between June 2018 and January 2019, as a separate test set. This validation step was performed to evaluate the algorithm's performance using a separate and unseen test dataset.

## Benchmark performance by VLE experts

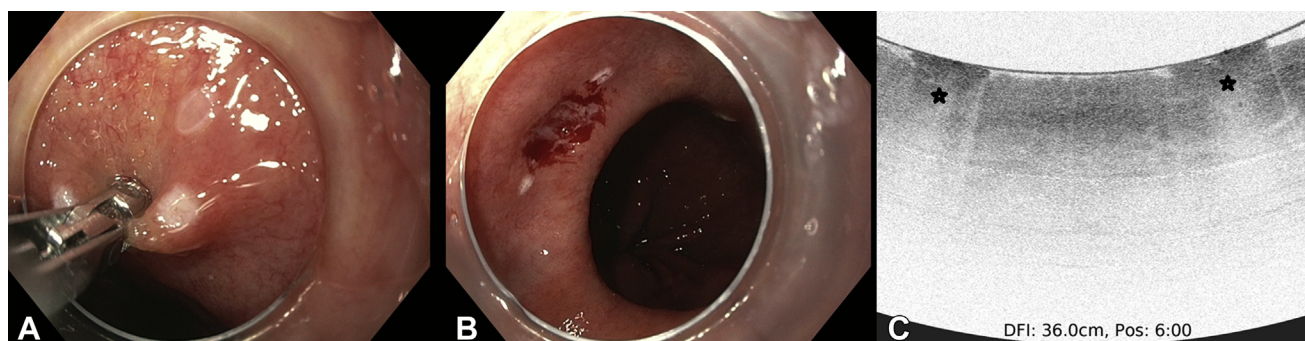
A randomly selected subset of the test set was evaluated by the VLE experts to benchmark the performance of the algorithm. Experts evaluated a randomly selected subset of the entire test set to balance out the nondysplastic versus neoplastic cases and to allow for a feasible amount of VLE ROIs to be evaluated by 10 assessors. These VLE targets were assessed using a web-based scoring module specifically developed to assess the diagnostic performance of the VLE experts for BE neoplasia detection. In total, 10 U.S. VLE experts completed the assessment blinded to histology. All VLE experts had participated in multiple VLE inter-observer studies, prospective VLE trials, and had more than 5 years of experience using the technology.<sup>8,9,19</sup>

During the assessment, VLE experts were asked to label each VLE target as NDBE or neoplastic. VLE raters were also asked to provide a level of confidence (high, moderate, low) for each score. VLE targets were locked after being scored to prevent assessors from going back to revise their earlier evaluations. The order in which the images were presented for evaluation was randomized among assessors. The diagnostic performance of each VLE expert was calculated and their mean diagnostic performance was compared with the performance of the CAD algorithm.

## Correlation between level of confidence and diagnostic performance

The algorithm provided a VLE target with a dysplasia probability estimate between 0% (nondysplastic) to 100% (neoplastic). In a post hoc analysis, if the VLE algorithm's dysplasia probability estimate was between 40% and 60%, the algorithm was considered to have insufficient confidence to make a prediction. These VLE targets were excluded, mimicking a low confidence interpretation by an endoscopist. Subsequently, only VLE targets with a high level of confidence were assessed in this post hoc analysis. In addition, the correlation between VLE expert level of confidence (high, moderate, low) and diagnostic





**Figure 2.** These images demonstrate the volumetric laser endomicroscopy (VLE) laser marking process that enabled correlation with the histopathology. These laser marks were targeted using VLE to allow for correlation with biopsy. During the endoscopy, laser marks are appreciated as white superficial cautery marks. A biopsy specimen is obtained across the laser mark to provide VLE-histology correlation for that specific region of interest. In the VLE image, the 2 laser marks are visible as small areas of high surface signal intensity (indicated by the stars).

performance for BE neoplasia detection was also evaluated.

### Primary outcome measures

- Diagnostic performance of the CAD algorithm for correct differentiation between nondysplastic and neoplastic BE using the first 22 patients (training set).
- Diagnostic performance of the CAD algorithm for correct differentiation between nondysplastic and neoplastic BE using the second 25 patients (test set).
- Benchmarking to VLE experts: the performance of the CAD algorithm was compared with that of 10 VLE experts for correct differentiation between nondysplastic and neoplastic BE using a subset of the unseen test set.

### Secondary outcome measures

- Correlation between level of confidence and diagnostic performance
- Assessment time

### Statistical analysis

Performance of the CAD algorithm was evaluated using the receiver operating characteristic curve. Corresponding accuracy, sensitivity, and specificity were calculated based on internal cross-validation on the training set and a separate analysis on the unseen test set. A CAD algorithm prediction was created for every esophageal VLE target. A VLE target was considered neoplastic when the average neoplasia probability estimate was  $>60\%$ . This neoplasia threshold of 60% was determined based on the training dataset, and this value was subsequently used to evaluate the performance metrics on the test set. In addition, VLE targets obtained from the same patient were always allocated to either the training or the internal validation set to prevent data leakage and thereby risk of overfitting. Because this was the first study to evaluate the performance of the CAD algorithm in vivo, no formal sample size calculation was conducted.  $P \leq .05$  was considered significant.

All statistical analyses were performed using Matlab 2016a (Mathworks Inc, Natick, Mass, USA).

## RESULTS

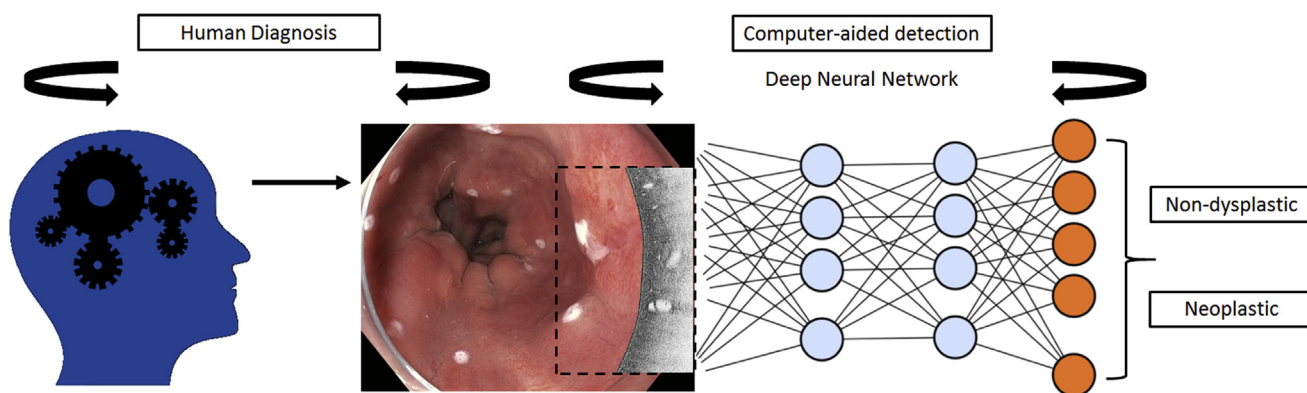
### Patient and VLE characteristics

VLE imaging was performed on 50 patients with BE; 3 patients were excluded because of technical failures, including balloon leakage after the balloon's black registration line was hit by the laser marking system, and no clear visual endoscopic appearance of both laser marks assuring adequate correlation between VLE and histology. Thus, 47 patients were eligible for inclusion into the study (Fig. 4). The mean age was 66 years (standard deviation, 8 years) and 40 (80%) were men. The median circumferential extent of the BE segment was 6 cm (interquartile range [IQR], 4-9 cm) and the maximum extent was 9 cm (IQR, 6-11 cm).

In total, 340 VLE targets were obtained from 47 patients with BE. Of these, 9 patients with low-grade dysplasia, 7 with indefinite for dysplasia, and 6 targets with inadequate image quality were excluded. Next, 229 NDBE and 89 neoplastic targets were selected, giving a total of 16,218 VLE images. The training set of the algorithm consisted of the first 22 patients with 172 targets (134 NDBE and 38 neoplastic). The separate test set consisted of the subsequent 25 patients with 146 targets (95 NDBE and 51 neoplastic).

### Primary outcome measurements

**Diagnostic performance of the CAD algorithm using the first 22 patients (training set).** Internal cross-validation accuracy of the CAD algorithm for correct differentiation between nondysplastic and neoplastic BE was 92% (95% confidence interval [CI], 89%-96%), sensitivity was 95% (95% CI, 91%-99%), and specificity was 92% (95% CI, 88%-96%). See Figure 5 for a representation of a VLE CAD algorithm prediction in BE.



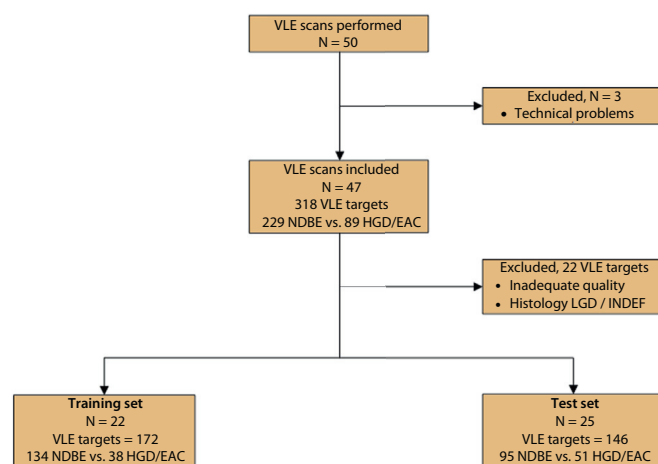
**Figure 3.** Schematic representation of the architecture of the computer-aided detection algorithm and the prediction of nondysplastic or neoplastic Barrett's esophagus using volumetric laser endomicroscopy images.

**Diagnostic performance of the CAD algorithm using the second 25 patients (test set).** Accuracy, sensitivity, and specificity for BE neoplasia detection were 85% (95% CI, 79%-91%), 91% (95% CI, 86%-96%), and 82% (95% CI, 76%-86%), respectively. The diagnostic performance of the CAD algorithm and that of the VLE experts is summarized in Table 1.

**Benchmarking of algorithm performance to that of VLE experts.** The VLE experts scored a randomly selected subset of the total dataset used to develop the CAD algorithm. This subset included 112 targets (73 NDBE and 39 neoplastic) from the test dataset. The accuracy, sensitivity, and specificity of VLE experts for BE neoplasia detection were 77% (95% CI, 73%-80%), 70% (95% CI, 64%-76%), and 81% (95% CI, 77%-85%), respectively. Figure 6 provides a summary of the diagnostic performance of individual VLE users in comparison with the CAD algorithm on the test set. The interobserver agreement between assessors for the diagnosis of BE neoplasia, defined by the median kappa (IQR), was 0.29 (0.18-0.37).

**Correlation between level of confidence and diagnostic performance.** In a post hoc analysis of algorithm performance on the test set, 117 of 146 ROIs (80%) were scored by the algorithm with a high level of confidence. Twenty-nine of 146 VLE targets (20%) were scored with a low level of confidence. When a VLE target was assessed by the CAD algorithm with a high level of confidence, accuracy, sensitivity, and specificity were 92% (7% increase), 91% (no increase), and 92% (10% increase), respectively. When a VLE target was assessed by the CAD algorithm with a low level of confidence, accuracy, sensitivity, and specificity were 55%, 88%, and 35%, respectively.

In a post hoc analysis of VLE expert performance on the test set, 67 of 112 (60%) of the targets contained a high level of confidence prediction and were included, resulting in an accuracy of 86% (9% increase), with a corresponding sensitivity of 80% (10% increase) and specificity of 89% (8% increase).

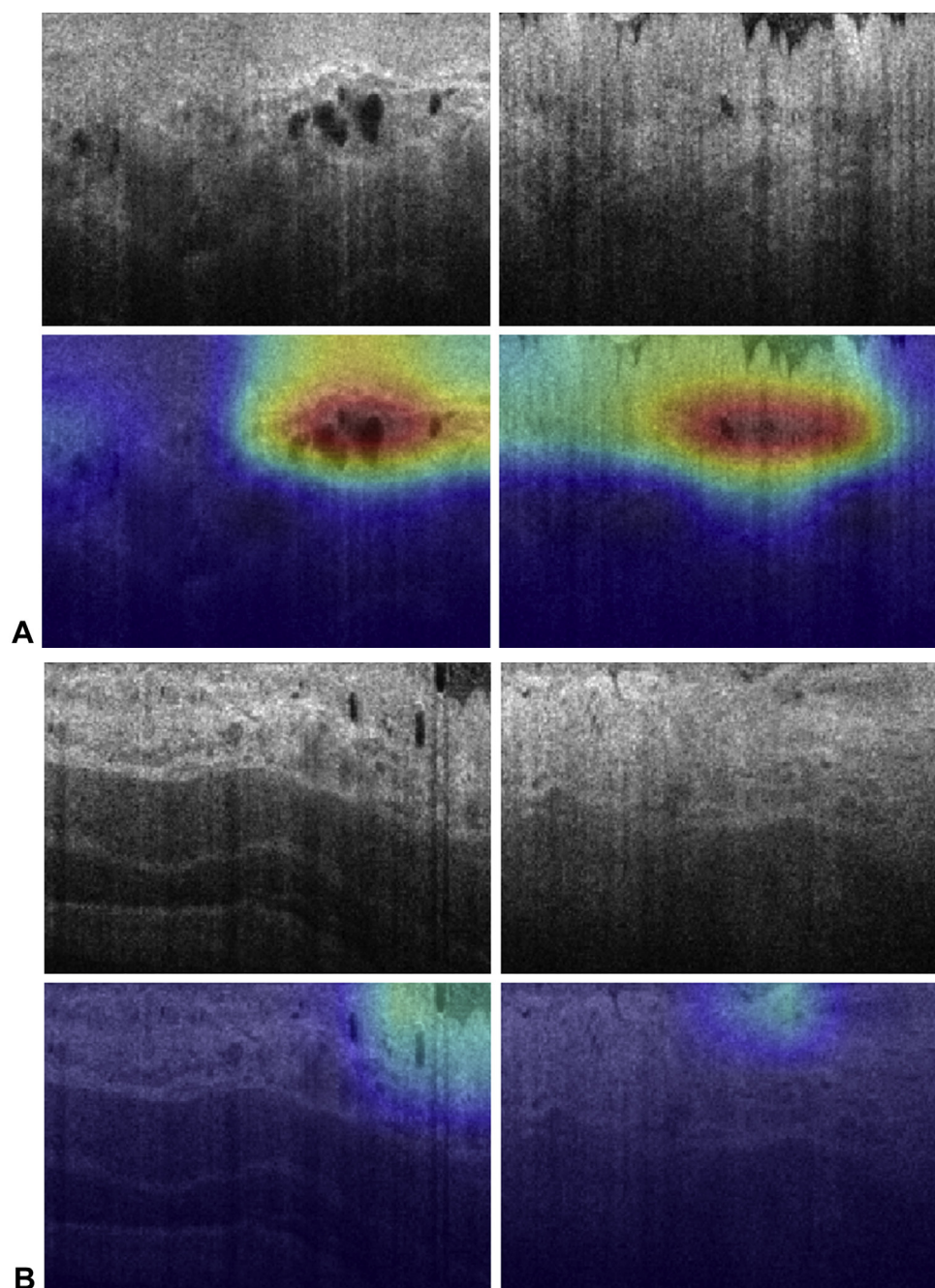


**Figure 4.** Flowchart outlining the patient inclusion scheme for the VLE PREDICT study. EAC, Esophageal adenocarcinoma; HGD, high-grade dysplasia; INDEF, indefinite for dysplasia; LGD, low-grade dysplasia, NDBE, nondysplastic Barrett's esophagus; VLE, volumetric laser endomicroscopy.

**Assessment time of the CAD algorithm and VLE experts.** The mean assessment time per VLE target (51 frames) of the CAD algorithm was 0.91 seconds (standard deviation, 0.033). The mean assessment time per VLE target of the VLE experts was 42 seconds (range, 8-299 seconds).

## DISCUSSION

VLE has the potential to improve detection of early BE neoplasia. However, interpretation of a large volume of gray-scale VLE images is complex and time-consuming for the human brain. In this study, we aimed to train and validate a CAD algorithm for VLE-based detection of BE neoplasia. Our study is the first to use prospectively collected and biopsy-correlated VLE targets to train and validate a CAD algorithm and to subsequently compare its performance with that of 10 recognized VLE experts.



**Figure 5.** **A**, Neoplastic volumetric laser endomicroscopy images of high-grade dysplasia with the corresponding dysplasia heatmaps. The heatmap visualization of the computer algorithm marks the most abnormal area in *red* and the nonsuspicious areas in *blue*. **B**, Nondysplastic volumetric laser endomicroscopy images with the corresponding dysplasia heatmaps. The heatmap visualization of the computer algorithm marks the most abnormal area in *red* and the nonsuspicious areas in *blue*.

We report high diagnostic accuracy and sensitivity of the CAD algorithm on histopathology-correlated targets from the VLE PREDICT database, both on the training set and a separate test set. The performance of the algorithm was subsequently benchmarked against the performance of recognized VLE experts, and it outperformed all 10 VLE experts. In addition, interobserver agreement between VLE experts for the identification of BE neoplasia was only fair (kappa, 0.29). Using a CAD algorithm would eliminate

variations in intra- and interobserver agreement between VLE users evaluating VLE images. This analysis provides evidence of the value of a CAD algorithm for assessment of VLE in BE and suggests that CAD analysis may improve the detection of early BE neoplasia, guiding targeted biopsies, and therefore reducing procedure time and costs.

Based on the separate test set, 9 of 9 neoplastic patients were correctly identified by the CAD algorithm. In 9 of 16 nondysplastic patients, 16 VLE ROIs were scored as



**TABLE 1. Diagnostic performance of the computer-aided detection algorithm and volumetric laser endomicroscopy experts for the identification of Barrett's neoplasia**

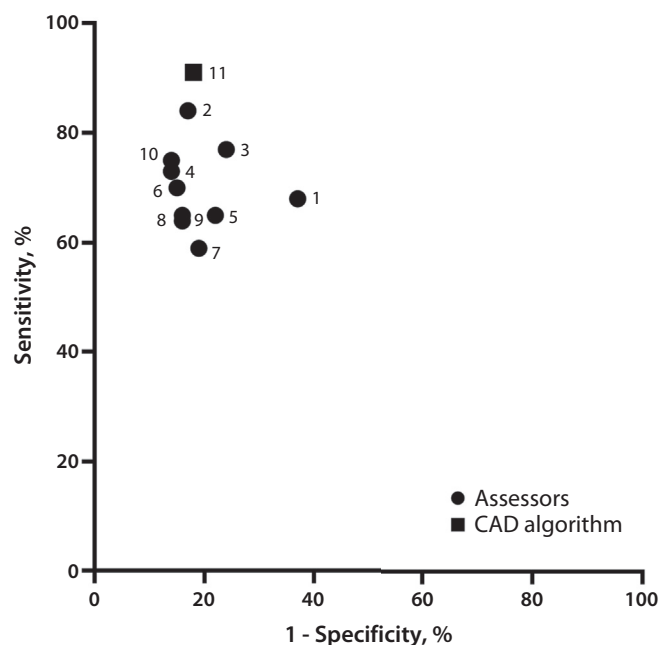
Dataset	Accuracy, % (95% confidence interval)	Sensitivity, % (95% confidence interval)	Specificity, % (95% confidence interval)	Area under the curve, % (95% confidence interval)
CAD on training set	92 (89-96)	95 (91-99)	92 (88-96)	98 (96-99)
CAD on test set	85 (79-91)	91 (86-96)	82 (76-86)	95 (91-98)
VLE experts on test set	77 (73-80)	70 (64-76)	81 (77-85)	75 (71-80)

CAD, Computer-aided detection; VLE, volumetric laser endoscopy.

neoplastic by the CAD algorithm. In clinical practice, this would mean that 16 biopsy specimens would have to be obtained due to a false-positive prediction by the CAD algorithm. In addition, when BE neoplasia was identified by either the CAD algorithm or the VLE experts with a high level of confidence, diagnostic accuracy, sensitivity, and specificity increased. Again, the CAD algorithm outperformed the VLE experts even when only predictions with a high level of confidence were analyzed. In total, 20% of all VLE targets evaluated by the CAD algorithm contained a low level of confidence prediction, whereas 40% of the VLE targets were labeled with low level of confidence by the experts. Thus, CAD has the potential to limit nondiagnostic interpretations.

The interobserver agreement between the best-scoring VLE expert and the CAD algorithm was relatively good (kappa, 0.7). The algorithm reached a sensitivity of 91% compared with 70% of experts, thereby making the algorithm much more sensitive to detect neoplasia in comparison with the experts. In the future, the CAD algorithm may make the experts more sensitive if they were used in conjunction. In contrast, the algorithm reached a specificity of 82% compared with 81% for the experts. Interestingly, multiple nondysplastic VLE areas that were scored incorrectly by the algorithm were scored correctly by multiple experts. After carefully reviewing these specific VLE areas in a post hoc analysis, many of these VLE areas were easily scored as nondysplastic using the existing VLE scoring criteria. Therefore, in the future, we envision that VLE experts may lower the number of false-positive predictions by the CAD algorithm during clinical practice. However, this will need to be evaluated in a separate prospective study when the combined performance of the CAD algorithm and experts is assessed.

The assessment speed of the CAD algorithm was 0.91 seconds per VLE target consisting of 51 frames. As expected, the speed of the CAD algorithm outperformed that of the VLE experts with a corresponding mean assessment time of 42 seconds per VLE target. A VLE target comprised 60 degrees of the circumferential VLE frame over a longitudinal extent of 51 VLE frames. Therefore, the assessment time of the algorithm is not yet amenable to real-time inference of 13 circumferential VLE frames



**Figure 6.** Diagnostic performance per assessor for the identification of Barrett's neoplasia using volumetric laser endomicroscopy regions of interest compared with the computer-aided detection (CAD) algorithm. The CAD algorithm outperformed all 10 volumetric laser endomicroscopy experts in terms of diagnostic accuracy and sensitivity.

per second. In its current form, a CAD prediction would be available in near real-time with a delay of 0.032 seconds per frame. Given the current pace of technological advancements, we anticipate that real-time performance is achievable in the near future using an optimal parallel graphics processing unit and faster network architectures.

The diagnostic performance of the VLE CAD algorithm was slightly lower than in previous work by our group. Previously, sensitivity and specificity were 90% and 93%, based on 30 NDBE and 30 neoplastic ex vivo VLE images<sup>5</sup> compared with 91% and 82% in the present study. The difference may be explained by the fact that we used in vivo, biopsy-correlated ROIs in contrast to the analysis of only a limited number of ex vivo VLE images. The current study provides a more realistic estimation of a VLE

CAD algorithm in clinical practice given the in vivo application and careful histology correlation.

The strengths of this study are that the data were recorded in a standardized prospective method in multiple hospitals. We followed a systematic VLE laser marking process allowing for the assessment of a large sample of histopathology-correlated VLE targets. Our study is the first to benchmark the performance of our CAD algorithm against the performance of a group of recognized VLE experts. This dataset may be used for benchmarking the performance of other CAD algorithms and/or the web-based assessment module may be used for creating new benchmarking sets. Finally, we analyzed larger VLE ROIs consisting of 51 neighboring frames, reflecting a biopsy sample, compared with most VLE studies with single-frame interpretation, generally lacking histopathologic correlation. This is more representative of VLE in clinical practice and has the advantage of increasing the training data to boost the performance of the CAD algorithm.

There are some limitations to this study. First, selection bias might have occurred, given that only high-quality images were used, which were recorded by expert endoscopists. This may limit the generalizability. Second, we did not evaluate the performance of the algorithm using a separate externally collected dataset from a different study outside our group. Third, we analyzed NDBE and neoplastic VLE images only, reflecting the more obvious pathologic cases. This may hamper the extrinsic validity on larger datasets with a wider variety in pathologic diagnosis, such as low-grade dysplasia. Fourth, we analyzed VLE ROIs only, because these contained precise histologic confirmation. In clinical practice, full scan interpretation of the entire BE segment will be required. Fifth, the web-based module used by the experts was not identical to the clinical VLE viewing station; however it did allow for a high-quality assessment of VLE targets. Sixth, multiple VLE targets per patient were analyzed, resulting in the possibility of statistical dependencies, yet this was corrected for in the statistical analysis. Seventh, we have used VLE images, including the laser markings, to train and test the performance of the algorithm. These markings may have influenced algorithm training. To address this theory, class activation maps (heatmaps) were used in a subanalysis (Fig. 5).<sup>20</sup> This allowed us to look at the prediction of the CAD algorithm to see which VLE features and their corresponding locations were predicted as abnormal. During this analysis, we saw that the laser markings were not predicted as abnormal/neoplastic in the class activation maps of the VLE images, thereby most likely limiting this effect. Finally, VLE expert performance aided by the CAD algorithm was not evaluated in this study. Further development and validation of the CAD algorithm addressing these limitations are underway.

In conclusion, we have developed and validated a VLE CAD algorithm for detection of BE neoplasia using pro-

spectively collected VLE targets that were laser marked and histologically sampled. The high diagnostic accuracy and corresponding assessment speed of the CAD algorithm were validated on a prospectively collected test dataset, showing promising performance. The performance of this CAD algorithm was subsequently compared with the performance of recognized VLE experts and outperformed all 10 experts. Future work will focus on improving the performance of the current CAD algorithm and its generalizability, and real-time VLE interpretation of the CAD algorithm.

## ACKNOWLEDGMENTS

The collaboration project is financed by the Ministry of Economic Affairs of the Netherlands by means of the PPP Allowance made available by the Top Sector Life Sciences & Health to Academic Medical Center Amsterdam to stimulate public-private partnerships.

## 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. 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.
4. 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.
5. Swager AF, van der Sommen F, Klomp SR, et al. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017;86:839-46.
6. Struyvenberg MR, van der Sommen F, Swager AF, et al. Improved Barrett's neoplasia detection using computer-assisted multiframe analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2020;33:doz065.
7. Swager AF, Tearney GJ, Leggett CL, et al. Identification of volumetric laser endomicroscopy features predictive for early neoplasia in Barrett's esophagus using high-quality histological correlation. *Gastrointest Endosc* 2017;85:918-26.e7.
8. Leggett CL, Gorospe EC, Chan DK, et al. Comparative diagnostic performance of volumetric laser endomicroscopy and confocal laser endomicroscopy in the detection of dysplasia associated with Barrett's esophagus. *Gastrointest Endosc* 2016;83:880-8.e2.
9. Trindade AJ, Inamdar S, Smith MS, et al. Volumetric laser endomicroscopy in Barrett's esophagus: interobserver agreement for interpretation of Barrett's esophagus and associated neoplasia among high-frequency users. *Gastrointest Endosc* 2017;86:133-9.
10. Alshelleh M, Inamdar S, McKinley M, et al. Incremental yield of dysplasia detection in Barrett's esophagus using volumetric laser endomicroscopy with and without laser marking compared with a standardized random biopsy protocol. *Gastrointest Endosc* 2018;88:35-42.
11. Smith MS, Cash B, Konda V, et al. Volumetric laser endomicroscopy and its application to Barrett's esophagus: results from a 1,000 patient registry. *Dis Esophagus* 2019;32:doz029.



12. Alvarez Herrero L, Curvers WL, van Vilsteren FGI, et al. Validation of the Prague C&M classification of Barrett's esophagus in clinical practice. *Endoscopy* 2013;45:876-82.
13. Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy* 2005;37:570-8.
14. Participants in the Paris Workshop; Aiko T, Sasako M, et al. 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.
15. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv 2015 arXiv:1409.1556. Available at: <https://arxiv.org/abs/1409.1556>. Accessed September 11, 2020.
16. 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 2009:248-55.
17. Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst* 2014;4:3320-8.
18. Fonollà R, Scheeve T, Struyvenberg MR, et al. Ensemble of deep convolutional neural networks for classification of early Barrett's neoplasia using volumetric laser endomicroscopy. *Appl Sci* 2019;9:2183.
19. Trindade AJ, Inamdar S, Smith MS, et al. Learning curve and competence for volumetric laser endomicroscopy in Barrett's esophagus using cumulative sum analysis. *Endoscopy* 2018;50:471-8.
20. Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization 2017 IEEE International Conference on Computer Vision (ICCV), Venice 2017:618-26. Available at: <https://ieeexplore.ieee.org/document/8237336>. Accessed September 11, 2020.

*Abbreviations:* BE, Barrett's esophagus; CAD, computer-aided detection; CI, confidence interval; IQR, interquartile range; NDBE, nondysplastic Barrett's esophagus; ROI, region of interest; VLE, volumetric laser endomicroscopy.

*DISCLOSURE:* Dr Weusten has received research support and speakers fees from Pentax Medical. Dr Leggett has received indirect research support from NinePoint Medical. Dr Kahn received an unrestricted travel grant and research equipment from NinePoint Medical. Dr Trindade has received research support NinePoint Medical and consultant fees from Olympus America and Pentax Medical. Dr Ganguly has received consultant fees from Boston Scientific. Dr Konda has received a research grant from Pentax. Dr Pleskow has received consultant fees from NinePoint Medical, Boston Scientific, Olympus, Fuji, Medtronic, and CSA. Dr Sethi has received consultant fees from Boston Scientific, Olympus, and Fujifilm. Dr Smith has received consultant fees from NinePoint Medical. Dr Wallace has received research grants from NinePoint Medical, Fujifilm, Boston Scientific, Olympus, Medtronic, NinePoint

Medical, Cosmo/Aries Pharmaceuticals; consultant fees from Virgo Inc, Cosmo/Aries Pharmaceuticals, Anx Robotica (2019), Covidien, and GI Supply; he holds stock options in Virgo Inc and has consulted on behalf of Mayo Clinic for GI Supply (2018), Endokey, Endostart, Boston Scientific, and Microtek; he has received general payments from Synergy Pharmaceuticals, Boston Scientific, and Cook Medical. Dr Tearney has received consultant fees and royalties from NinePoint Medical and research support from Boston Scientific, iLumen, and CN USA Biotech Holdings. Dr Vieth has received honoraria for lecturing from Falk, Shire, and Olympus. Dr Bergman has received research support from NinePoint Medical and speaker fees from Fujifilm. All other authors disclosed no financial relationships.

Copyright © 2021 by the American Society for Gastrointestinal Endoscopy. Published by Elsevier, Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

0016-5107

<https://doi.org/10.1016/j.gie.2020.07.052>

Received April 6, 2020. Accepted July 21, 2020.

Current affiliations: Department of Gastroenterology and Hepatology, Amsterdam UMC, location AMC, Amsterdam (1); Department of Electrical Engineering, VCA group, Eindhoven University of Technology, Eindhoven (2); Department of Gastroenterology and Hepatology, Catharina Hospital, Eindhoven (3); Department of Gastroenterology and Hepatology, St. Antonius Hospital, Nieuwegein, the Netherlands (4); Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota (5); Division of Gastroenterology and Hepatology, Mayo Clinic, Scottsdale, Arizona (6); Division of Gastroenterology and Hepatology, Zucker School of Medicine at Hofstra/Northwell, Long Island Jewish Medical Center, New Hyde Park, New York (7); Department of Gastroenterology and Hepatology, University of Vermont Medical Center, Burlington, Vermont (8); Department of Gastroenterology and Hepatology, Baylor University Medical Center at Dallas, Dallas, Texas (9); Division of Gastroenterology and Hepatology, New York-Presbyterian Hospital, New York, New York (10); Department of Gastroenterology and Hepatology, Beth Israel Deaconess Medical Center, Boston, Massachusetts (11); Department of Gastroenterology and Hepatology, Columbia University Medical Center, New York, New York (12); Division of Gastroenterology and Hepatology, Mount Sinai West & Mount Sinai St. Luke's Hospitals, New York, New York (13); Division of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, Florida (14); Department of Pathology, Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, Massachusetts, USA (15); Department of Pathology, Amsterdam UMC, location AMC, Amsterdam, the Netherlands (16); Institute of Pathology, Bayreuth Clinic, Bayreuth, Germany (17).

Reprint requests: Prof. dr. J.J.G.H.M. Bergman, Department of Gastroenterology and Hepatology, Amsterdam University Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.