

Automated Staging of Age-Related Macular Degeneration Using Optical Coherence Tomography

Freerk G. Venhuizen,^{1,2} Bram van Ginneken,¹ Freekje van Asten,² Mark J. J. P. van Grinsven,^{1,2} Sascha Fauser,^{3,4} Carel B. Hoyng,² Thomas Theelen,² and Clara I. Sánchez^{1,2}

¹Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands

²Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands

³Cologne University Eye Clinic, Cologne, Germany

⁴Roche Pharma Research and Early Development, F. Hoffmann-La Roche Ltd, Basel, Switzerland

Correspondence: Freerk G. Venhuizen, Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Postbus 9101, 6500 HB Nijmegen, The Netherlands; freerk.venhuizen@radboudumc.nl

Submitted: August 16, 2016

Accepted: March 24, 2017

Citation: Venhuizen FG, van Ginneken B, van Asten F, et al. Automated staging of age-related macular degeneration using optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2017;58:2318–2328. DOI:10.1167/iops.16-20541

PURPOSE. To evaluate a machine learning algorithm that automatically grades age-related macular degeneration (AMD) severity stages from optical coherence tomography (OCT) scans.

METHODS. A total of 3265 OCT scans from 1016 patients with either no signs of AMD or with signs of early, intermediate, or advanced AMD were randomly selected from a large European multicenter database. A machine learning system was developed to automatically grade unseen OCT scans into different AMD severity stages without requiring retinal layer segmentation. The ability of the system to identify high-risk AMD stages and to assign the correct severity stage was determined by using receiver operator characteristic (ROC) analysis and Cohen's κ statistics (κ), respectively. The results were compared to those of two human observers. Reproducibility was assessed in an independent, publicly available data set of 384 OCT scans.

RESULTS. The system achieved an area under the ROC curve of 0.980 with a sensitivity of 98.2% at a specificity of 91.2%. This compares favorably with the performance of human observers who achieved sensitivities of 97.0% and 99.4% at specificities of 89.7% and 87.2%, respectively. A good level of agreement with the reference was obtained ($\kappa = 0.713$) and was in concordance with the human observers ($\kappa = 0.775$ and $\kappa = 0.755$, respectively).

CONCLUSIONS. A machine learning system capable of automatically grading OCT scans into AMD severity stages was developed and showed similar performance as human observers. The proposed automatic system allows for a quick and reliable grading of large quantities of OCT scans, which could increase the efficiency of large-scale AMD studies and pave the way for AMD screening using OCT.

Keywords: retinal image analysis, automated grading, OCT, AMD classification, machine learning

Age-related macular degeneration (AMD) is the primary cause of legal blindness among elderly people in developed countries.^{1,2} AMD affects the central field of vision, slowly progressing from early to intermediate stages, with no or only subtle visual changes, to an advanced stage, where severe loss of central vision may occur rapidly. Current therapy allows for halting or reversing aspects of vision loss resulting from AMD,^{3,4} but there is still a large percentage of nonresponders to available treatments.⁵

Accurate grading of the AMD severity stage is important for identification of patients at risk of progression who may benefit most from therapy. In addition to patient stratification for high risk, accurate staging of AMD in its progressive subtypes is of great importance in the analysis of large data sets of patients with early and intermediate stages of AMD in order to learn more about the effect of phenotype on risk for progression to late-stage AMD by detecting minor divergences between phenotypes, including novel biomarkers.^{6,7}

Optical coherence tomography (OCT) is becoming a standard in clinical trials as well as in clinical practice for the

diagnosis and follow-up of patients with AMD.^{8,9} Current spectral-domain OCT allows for a noninvasive, three-dimensional visualization of the retina with high resolution. OCT is capable of accurately characterizing the three-dimensional shape and extent of drusen and their change over time in early and intermediate AMD.¹⁰ Also, atrophic areas and signs of neovascularization in advanced stages may be identified by modern OCT technology.^{11–13} However, the manual analysis of OCT volumes for AMD staging is time-consuming and prone to errors owing to the necessity to review multiple scans to identify distinguishing features associated with the different stages. This problem cannot be solved without computer aided detection to guarantee high quality of analysis at low costs and time burden. Indeed, machine learning for automated retinal image analysis has been acknowledged to be of major value in risk screening for retinal diseases.¹⁴

In the past years, computer-based algorithms have demonstrated their potential in the automatic analysis of retinal images and, particularly, in OCT volumes.¹⁵ Previously proposed works for OCT analysis have mostly focused on the development of



automated retinal layer segmentation algorithms.^{7,16} Furthermore, machine learning has been applied to detect textural properties for assessing changes in the structure of retinal tissue composition,¹⁷ to detect and segment retinal vessels,¹⁸ and to detect various retinal lesions such as intraretinal cysts or subretinal fluid.¹⁹ In recent years, deep learning methods have gained popularity in the field of computer vision and are now also entering the field of retinal image analysis, for example, the problem of cyst segmentation has been tried with convolutional neural networks with promising results.²⁰

Although machine learning—and more recently, deep learning—has made its mark in OCT analysis, it is still minor compared to the considerable effort that has been devoted for the automatic analysis of color fundus images.¹⁵ Only few machine learning algorithms have been published that automatically analyze OCT scans for AMD classification and grading.^{21–26} Owing to the large variability of the pathologic changes of AMD in OCT, these studies have mainly focused on identifying only single severity stages of AMD. Specifically, as the changes are particularly minimal in the earlier stages, most studies have focused on discriminating patients with neovascular AMD from normal patients or from other macular pathologies not related to AMD.^{24–26} Different thickness biometrics, as recently used to distinguish intermediate AMD from normal subjects, depend strongly on accurate layer segmentation algorithms.²³ These algorithms have a tendency to fail when evaluating heavily affected retinas and require manual corrections to avoid misleading outcomes.^{27,28}

In a previous work²⁹ we have presented a method to distinguish intermediate AMD from normal subjects, based on a publicly available data set.²³ To our knowledge, there is currently no method available for the automated identification of the different AMD severity stages, using OCT volumes.

In our current study we therefore extended and improved upon our previous work by developing and evaluating a machine learning algorithm that automatically grades four AMD severity stages and distinguishes them from healthy controls, based on OCT scans, without the need for an accurate presegmentation of the retinal layers.

METHODS

Data

For this study a total of 3265 OCT volumes obtained from 1016 patients were randomly selected from the European Genetic Database (EUGENDA; <http://eugenda.org>, in the public domain), a large multicenter database for clinical and molecular analysis of AMD.^{30,31} Written informed consent was obtained before enrolling patients in EUGENDA. The EUGENDA study was performed according to the tenets set forth in the Declaration of Helsinki, and Institutional Review Board approval was obtained.

OCT volumes were acquired with a Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany) at a wavelength of 870 nm, a transversal resolution ranging from 5.5 to 14 μm , and an axial resolution of up to 3.9 μm . The dimension in the axial resolution was 496 pixels; in the transversal direction the dimensions varied between 512 and 1536 pixels. The number of slices, that is, the number of B-scans, varied from 19 to 60, corresponding to a B-scan spacing ranging from ~ 320 up to ~ 110 μm , respectively. Before processing, to remove the variability in resolution, all B-scans from an OCT volume were resampled to a constant pixel size of 5.5 $\mu\text{m} \times 3.9$ μm corresponding to the lowest resolution present in the data set. This resampling scale was selected so as not to generate new information due to upsampling.

TABLE 1. Criteria for Grading AMD on Color Fundus Imaging According to the CIRCL

Grade	AMD	
	Severity Stage	Criteria
1	No AMD	No drusen or small, hard drusen only.
2	Early AMD	>10 small (<63 μm), hard drusen + pigmentary changes or 1–15 intermediate (63–124 μm) drusen.
3	Intermediate AMD	>15 intermediate (63–124 μm) drusen or any large (>125 μm) drusen or GA not in the central circle of the ETDRS grading grid.
4	Advanced AMD GA	Presence of GA.
5	Advanced AMD CNV	Evidence of active or previous CNV lesion.
6	CNV without signs for AMD	Chosen if CNV is present but no drusen of any size are present within field 2 of the ETDRS grid.
7	Cannot grade	Image is regarded as not gradable.

ETDRS, Early Treatment Diabetic Retinopathy Study.

For each OCT volume in the EUGENDA database, the AMD severity stage was assessed by the Cologne Image Reading Center and Laboratory (CIRCL). These stages or grades were assigned from the assessment of a color fundus image acquired at the same time of the OCT scan, following the AMD classification criteria shown in Table 1. For this study, all available OCT scans from a random subset of 1016 patients were extracted. Scans with grade 6 or 7, that is, choroidal neovascularization (CNV) without signs of AMD or ungradable, respectively, were excluded from this study.

Example OCT scans from EUGENDA with different AMD severity stages are shown in Figure 1. The data were randomly divided into two sets (80/20 split on patient level): a training set, consisting of 2884 OCT scans from 814 patients, for the development and optimization of the machine learning algorithm; and a test set, consisting of 381 OCT scans from 202 patients, for the evaluation of the algorithm. Scans from the same patients were kept in the same set. When multiple OCT volumes from the same eye were present, a single volume was selected randomly to be included in the test set. Table 2 shows the distribution of OCT volumes for the different AMD severity stages within both sets. The number of eyes in the respective subgroup is denoted in parentheses.

To assess the generalizability of the proposed algorithm an external set was used from a publicly available database²³ containing 384 OCT volumes of which 269 show intermediate AMD and 115 are controls. In this data set, intermediate AMD was defined as having large drusen (>125 μm) in both eyes or large drusen in one eligible eye and advanced AMD in the fellow eye. OCT volumes were acquired by using a Bioptigen SD-OCT imaging system (Bioptigen, Inc., Research Triangle Park, NC, USA) with 1000 A-scans per B-scan and 100 B-scans per volume in a 6.7 mm \times 6.7 mm region surrounding the fovea. For further details and information concerning the inclusion criteria, see article describing the data set.²³

Machine Learning Algorithm

The proposed algorithm automatically analyzed a whole OCT volume and indicated the corresponding AMD severity stage from a general representation of the OCT content. The algorithm was built around the Bag of Words (BoW) approach, a computer-based model firstly introduced to perform text categorization and further adapted for image classification.³² In this approach, a “dictionary” is created by using representative

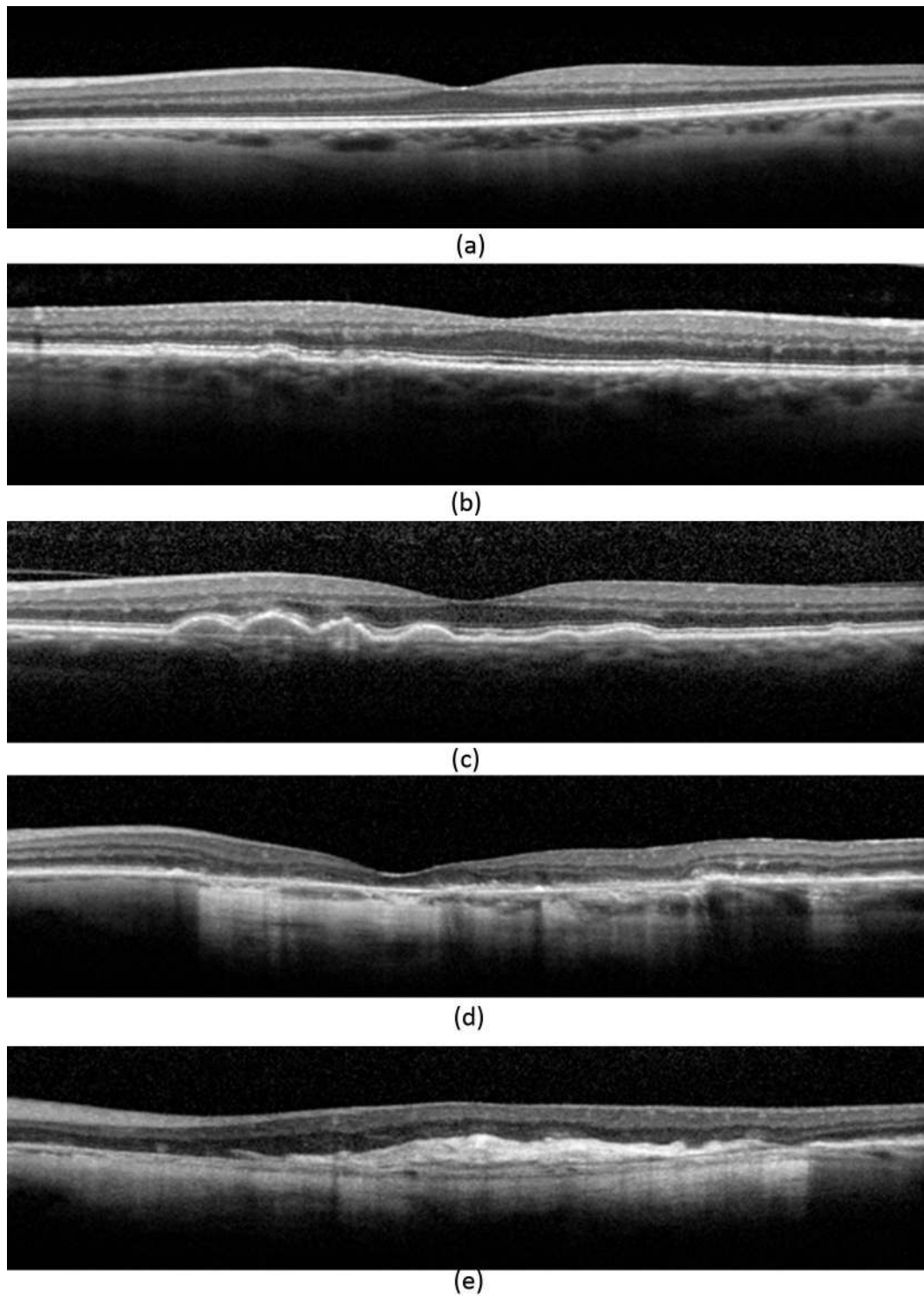


FIGURE 1. Examples of B-scans showing the different severity stages of AMD as defined by the CIRCL grading criteria shown in Table 1: (a) No AMD, (b) early AMD, (c) intermediate AMD, (d) advanced AMD with GA, and (e) advanced AMD with CNV.

visual words, where a visual word is usually defined through local image patches showing a localized view of the image content. Based on this dictionary, an image can then be represented as a frequency vector (histogram) of visual word occurrences. This general representation can be used to compare and classify images and their content.³³ To apply the BoW approach for the analysis of OCT scans, the proposed algorithm followed several steps (visualized in Fig. 2).

Salient Patch Detection. To define the visual words of the dictionary, image patches were extracted from different locations in the B-scans of the OCT volumes. Although these patches can be randomly sampled from any region of the image, only patches from regions mainly affected by AMD, that is, the outer retinal layers,³⁴ were considered in order to create a dictionary with a higher information content for AMD classification. To automatically identify these regions, a simple

TABLE 2. Distribution of OCT Volumes in the Training Set and Test Set Given the AMD Severity Stage*

Grade	AMD Severity Stage	Training Set	Test Set	Total
1	No AMD	870 (731)	174 (174)	1044 (905)
2	Early AMD	299 (176)	40 (40)	339 (216)
3	Intermediate AMD	483 (222)	57 (57)	540 (279)
4	Advanced AMD GA	115 (67)	13 (13)	128 (80)
5	Advanced AMD CNV	1117 (383)	97 (97)	1214 (480)
Total		2884 (1579)	381 (381)	3265 (1960)

* The number of eyes in the respective subgroup is denoted in parentheses.

and coarse layer segmentation method suffices. First, the absolute value of the Gaussian derivative³⁵ of the OCT reflectivity values along the axial direction was calculated for each B-scan (see Figs. 3c, 4c). Next we thresholded the resulting gradient image at the 90th percentile of the ordered intensity values of the gradient image in order to identify regions with large contrast changes, such as the boundaries of high reflective layers (see Figs. 3d, 4d). To focus in the outer retina, only points with an axial coordinate higher than the average axial position of the detected regions were selected. Image patches of size $n \times n$ were then randomly sampled from the detected regions and normalized to zero mean and unit variance in order to reduce variance and to enhance contrast. Figures 3b and 4b show examples of selected locations for patch extraction, while Figures 3e through 3h and Figures 4e through 4h show examples of patches extracted from these salient locations.

Dictionary Generation. A dictionary of representative visual words for AMD classification was then created by using the training set. M patches from the detected salient regions were randomly sampled from each training OCT volume. These patches were grouped into five sets by the AMD severity stage from the OCT scan they belonged to. Each set of patches was further partitioned into k subsets or clusters by using the k -means-clustering algorithm, in which each patch belongs to the cluster with the nearest mean or cluster centroid.³⁶ Each cluster centroid acts as a representative (or visual word) of the patches belonging to each cluster. A dictionary of $5k$ visual words was created by using the $5k$ cluster centroids calculated from the training set. Having visual words from each of the AMD severity stages provided a better representation of the different stages, especially because one of the stages, namely, stage 4, was slightly underrepresented in the training set.³⁷ To reduce computational complexity during the clustering

process, principal component analysis was applied to the patches to lower the dimensional space,³⁸ keeping the first p principal components for further processing.

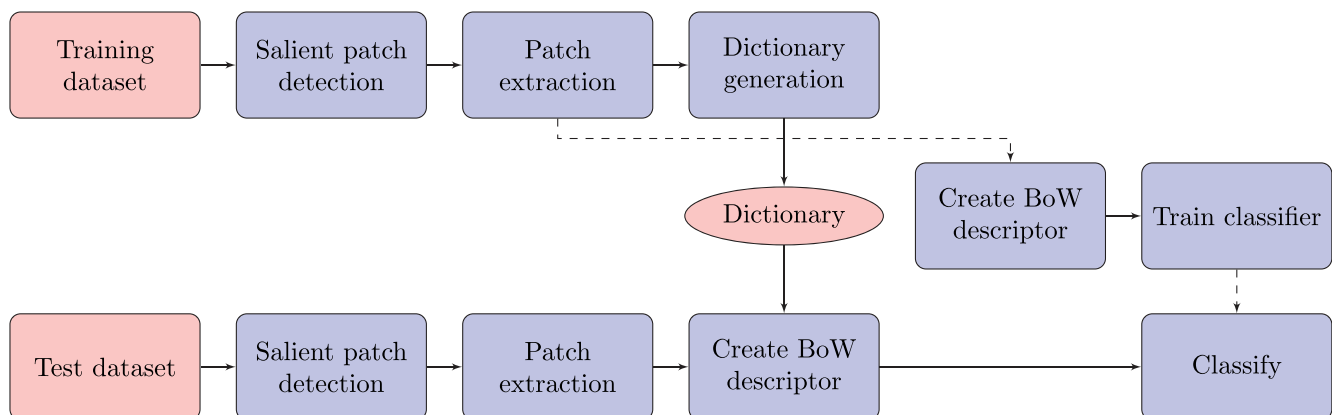
OCT Representation. Once the dictionary was created, a given OCT volume could now be represented as a “bag of visual words.” First, M patches were extracted from the detected salient regions as described in the previous subsection. Each patch was then assigned to the nearest visual word from the dictionary by using k -nearest neighbor search.³⁹ Finally, the OCT content was represented as a histogram of the visual word occurrences, where each bin of the histogram counts how many times each of the visual words occurs in the OCT volume.³³ Figure 5 shows examples of the bag of visual words representations for each AMD severity stage corresponding to the images in Figure 1. For this example, a small dictionary of 100 visual words was used to create the histogram representations. Given this representation of the OCT content as input, a multiclass random forest classifier was then trained to identify the AMD severity stage.⁴⁰ This classifier was trained on the training set by using a “one-versus-all” approach.⁴¹ The output of this classifier is a vector of five probabilities that indicate the likelihood of the OCT scan belonging to each of the AMD severity stages. The class with the highest probability was selected as the final classification output. The processing time required to predict the AMD severity stage for a single OCT volume is in the order of 2 to 5 seconds, depending on the scan density.

Observer Study

To compare the performance of the proposed machine learning algorithm to that of human observers, two retinal specialists, with 12 and 4 years of OCT reading experience, manually analyzed the OCT volumes. Solely on the basis of OCT information, the specialists were asked to assign an AMD severity stage, following the criteria in Table 1. OCT scans marked as stage 6 or 7 by at least one of the observers were excluded from the statistical analysis. The volumes were graded in different sessions depending on the observers’ availability. Scans were visualized on an LCD screen with a custom retinal image analysis workstation. The software allows for a uniform vendor-independent visualization with the possibility of scrolling/zooming/panning.

Data Analysis

The performance of the machine learning algorithm and the two human observers was compared separately to the reference standard by using Cohen’s κ agreement and receiver

**FIGURE 2.** Overview of the proposed algorithm for the identification of AMD severity stages, based on OCT images.

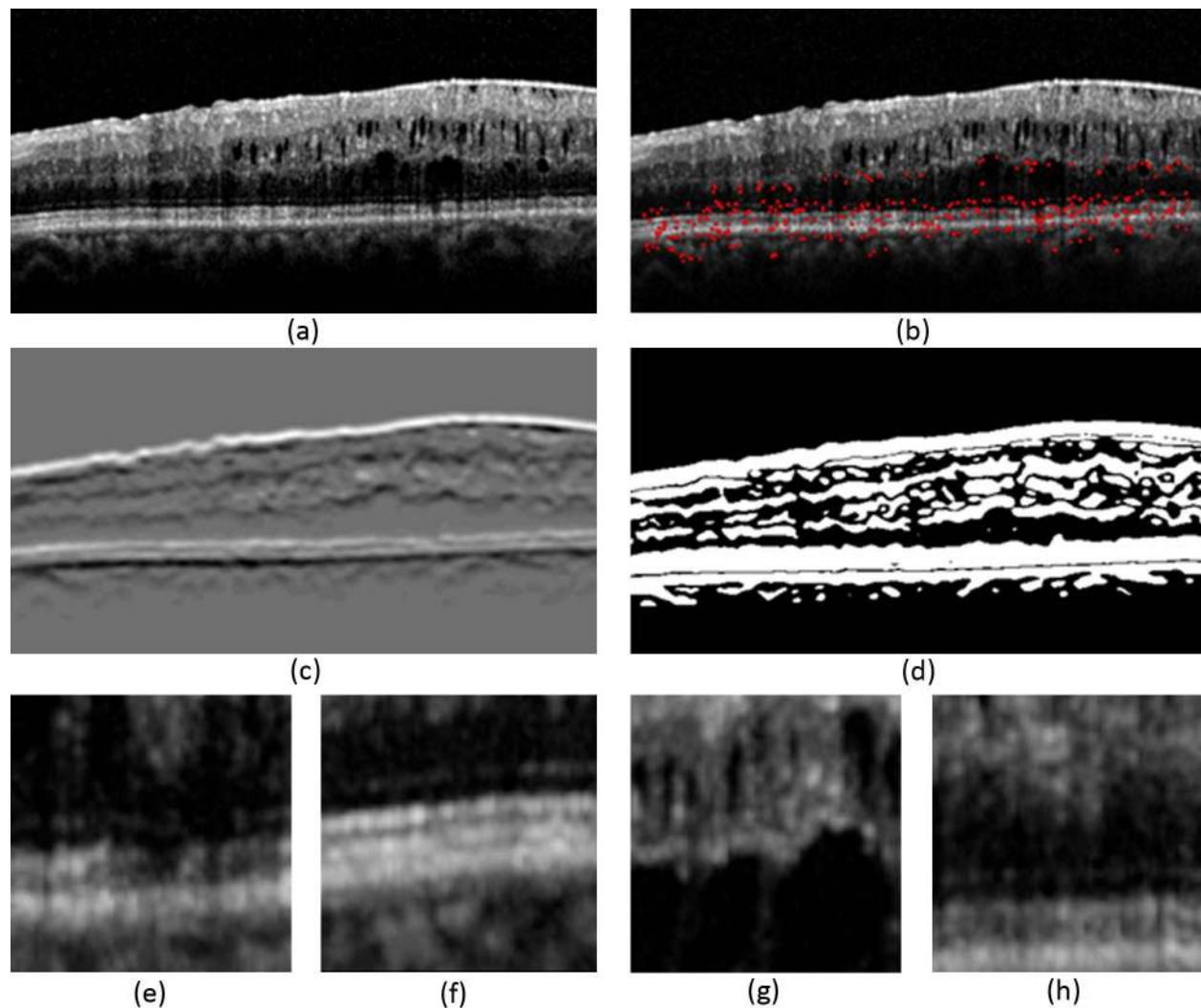


FIGURE 3. Example showing the steps for salient patch detection: (a) original image, (b) selected saliency locations shown in red, (c) Gaussian derivative along the axial direction, (d) output after thresholding the derivative image shown in (c) at the 90th percentile, and (e–h) examples of extracted salient patches.

operating characteristic (ROC) analysis. Bootstrap analysis⁴² was performed to obtain the mean ROC curve and the 95% confidence intervals. We performed two different experiments to evaluate the performance of the algorithm in different scenarios: (1) AMD grading into five severity stages, as shown in Table 1, in the test set; and (2) AMD high-risk level identification in the test set and the external set. Instead of a detailed AMD staging, in experiment 2 the algorithm was retrained for the binary task of identifying high-risk patients for progression to AMD, by grouping severity grades 1 to 2 into low risk and 2 to 5 into high risk. This experiment allowed comparison with previous works²⁹ and the assessment of the generalizability of the algorithm to data from a different source. The area (Az) under the ROC curve and sensitivity/specificity values were used as a performance measure for experiment 2. For experiment 1, overall agreement between the reference standard and the algorithm output and the observers' opinion was calculated by using κ statistics (SPSS, v20.0.0; IBM Corp., Armonk, NY, USA). The parameters of the algorithm, namely, the number M of patches per OCT volume, the patch size n , the number p of principal components, and the number k of visual words per AMD stage, were optimized by using one-eighth of the training set. The parameter M has to be set high

enough to accurately capture the characteristics of the OCT volume; it was set to 10,000 patches. A higher number of patches had no effect on the performance. A grid search was performed over the remaining three parameters. The parameter n was varied between 11 and 61 pixels, k was varied between 50 and 2500 visual words, and p was varied from 10 to 150 components. The optimal values were identified as $n = 61$ pixels, $k = 2500$ visual words, and $p = 100$ components.

RESULTS

Experiment 1: AMD Staging

The confusion matrix comparing the output of the machine learning algorithm to the reference standard is shown in Table 3. As 14 images (3.7%) were deemed ungradable by at least one of the two human observers, these images were excluded, leaving a total of 367 OCT volume scans for statistical analysis. Quantitatively, a κ -value of 0.713 was obtained for the automated grading of AMD into five severity stages. Tables 4 and 5 show the agreement between the human observers and the reference standard, with a κ -value of 0.775 and 0.755 for

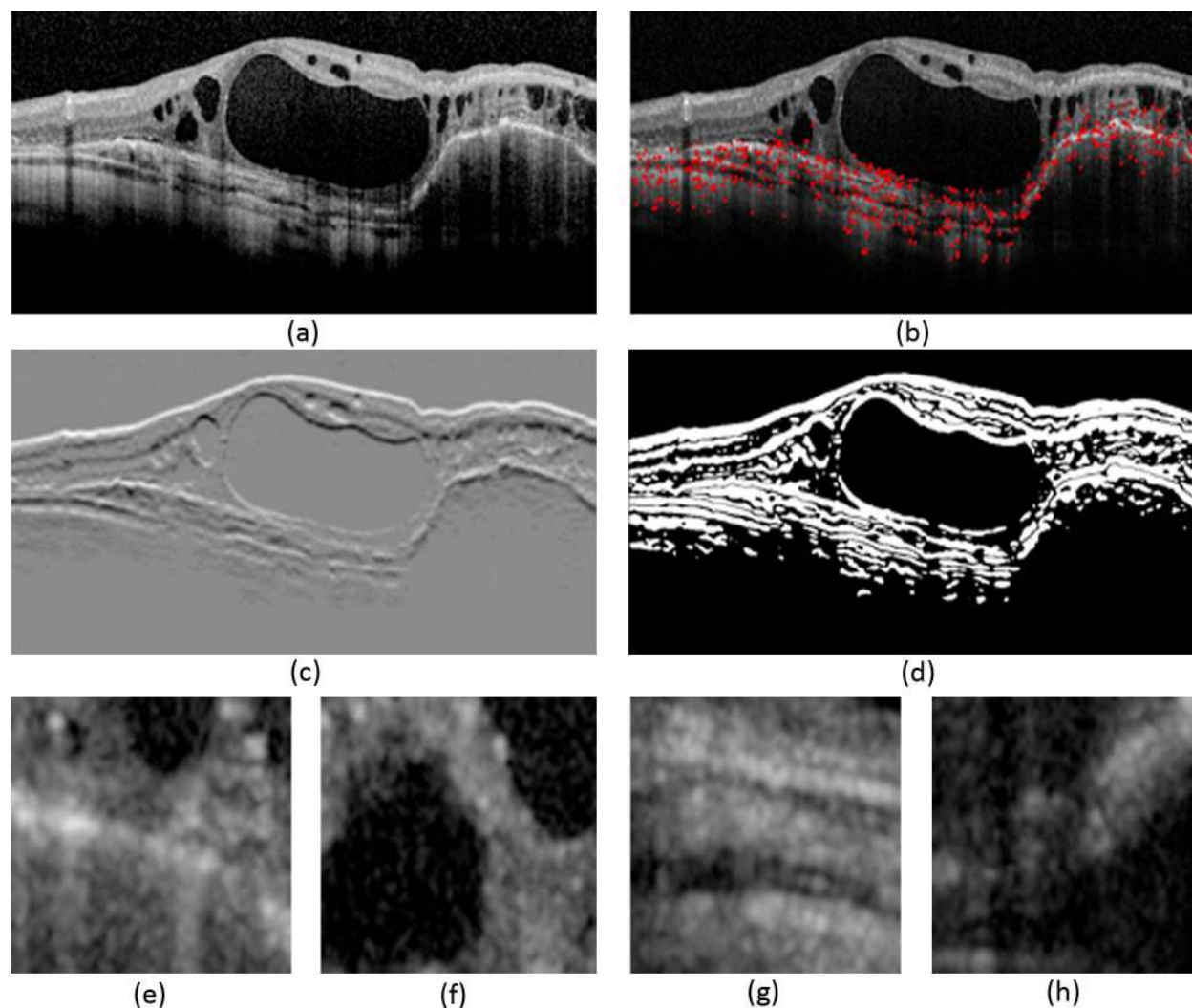


FIGURE 4. Example showing the steps for salient patch detection: (a) original image, (b) selected saliency locations shown in red, (c) Gaussian derivative along the axial direction, (d) output after thresholding the derivative image shown in (c) at the 90th percentile, and (e-h) examples of extracted salient patches.

observer 1 and observer 2, respectively. The interobserver agreement was 0.796.

Experiment 2: AMD High-Risk Identification

The ROC curve for classifying patients as either being at low risk or at high risk for developing advanced AMD is shown in Figure 6a. An area under the ROC curve of 0.980 and a maximum accuracy of 0.942 can be observed for classifying OCT volumes in the test set. At the point of maximum accuracy on the ROC curve, that is, the point closest to the top left corner, a specificity of 0.912 and sensitivity of 0.982 were obtained.

Figure 6a also shows the performance of the two human observers in the test set. Observer 1 achieved a sensitivity of 0.970 and a specificity of 0.897; observer 2 obtained a sensitivity of 0.994 and a specificity of 0.872. No significant difference was observed between the observers and the proposed machine learning algorithm. Figure 6b shows the ROC curve of the proposed algorithm on the external set, achieving an A_z of 0.978. In our previous work, an A_z of 0.993 has been achieved when training and evaluating the system using the external set.²⁹

DISCUSSION

In this study we assessed the performance of a machine learning algorithm for the grading of AMD severity stages, using OCT scans. A novelty of the developed system was the low requirements with regard to retinal layer segmentation; a simple and fast algorithm was sufficient to obtain high performance.

In our large consecutive set of 3265 OCT scans from 1016 patients we demonstrated that the performance of our system approached human performance in grading different AMD stages. Quantitatively, a κ -value of 0.713 was obtained, which approaches the performance of both human observers. As shown in the confusion matrix there was good agreement between the proposed algorithm and the reference standard, especially for grades 1, 4, and 5. These classes typically have clear distinct visual characteristics, which are successfully captured by the automated method. When considering class 2, that is, early AMD, 21 cases were wrongly classified as belonging to the control group. Similar errors are also shown in the confusion matrices for both human observers in Tables 4 and 5. This might be caused by the minor visual changes that

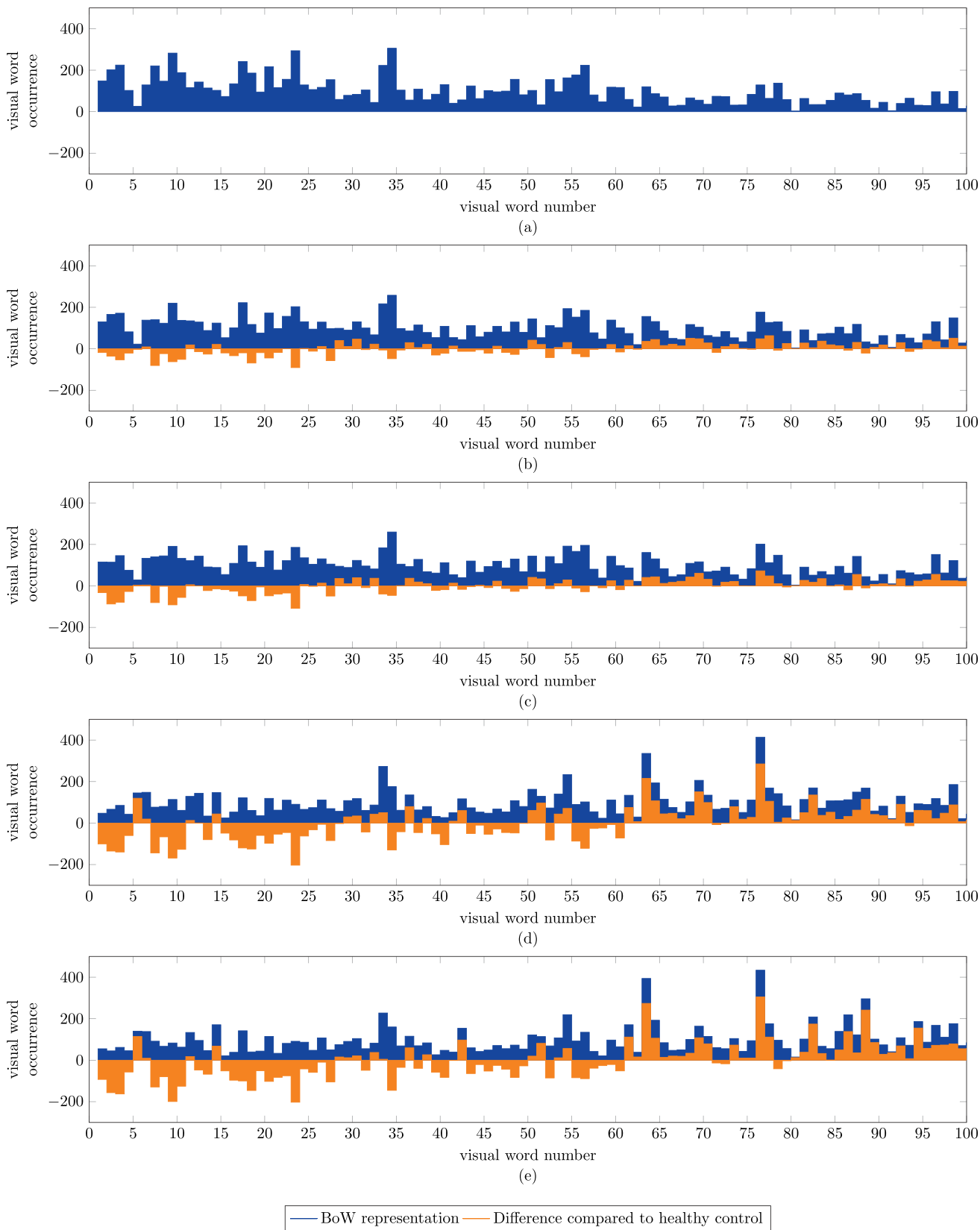


FIGURE 5. Example of BoW representations based on a dictionary of 100 visual words and 10,000 patches corresponding to the images in Figure 1: (a) no AMD, (b) early AMD, (c) intermediate AMD, (d) advanced AMD with GA, and (e) advanced AMD with CNV. Visual inspection already reveals distinct differences between lower and higher stages.

TABLE 3. Confusion Matrices for the Staging of AMD Into the Five Severity Stages Defined in Table 1 Between the Machine Learning Algorithm and the Reference Standard

Reference Standard	Machine Learning Algorithm					
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Total
Grade 1	154	5	4	0	2	165
Grade 2	21	8	6	1	2	38
Grade 3	5	3	41	4	4	57
Grade 4	0	1	0	10	1	12
Grade 5	0	1	7	5	82	95
Total	180	18	58	20	91	367

characterize early AMD, which can be easily overlooked by even experienced graders. This type of error might also be caused by the definition of the reference standard, as the assessment of the severity stage was based on color fundus images instead of OCT scans. Small drusen that are visible on color fundus imaging might be missed by an OCT volume with a large spacing between B-scans, introducing possible labeling errors.

On the other hand, nine early cases were wrongly classified by the algorithm with a higher severity stage. The observers each made this error in 14 and 20 cases, respectively. AMD signs, such as nascent geographic atrophy (GA), which are not visible on color fundus imaging, might be visible on OCT imaging and could therefore be correlated with different severity stages.^{11,12} Of the 107 advanced cases of AMD, only two cases were misclassified as early AMD. After visual inspection of these misclassified cases, signs of advanced AMD were present, although to a small extent, which are prone to be missed by the algorithm. A single global histogram is created for an entire OCT volume, and a small localized lesion might not contribute enough to the histogram for it to be classified in a higher severity stage. Figure 7 shows an example of an underestimated case containing a small GA lesion. Advanced AMD with GA was underrepresented in our data set; adding more samples from this severity stage might further improve sensitivity and performance. For OCT volumes outside of the specified range of AMD subcategories, the system will attempt to assign the most fitting AMD category. The system is trained to link certain structural changes in an OCT image to a certain stage of AMD. If those structural changes are similar, the method will wrongly classify the scan as the most structurally similar AMD class. Adding training samples from that class in the training set, as well as adding other types of clinical information, might help to discriminate between these categories.

For the identification of high-risk AMD stages, the system achieved an A_z of 0.980 with a sensitivity of 0.982 at a

TABLE 4. Confusion Matrices for the Staging of AMD Into the Five Severity Stages Defined in Table 1 Between Observer 1 and the Reference Standard

Reference Standard	Human Observer 1					
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Total
Grade 1	149	9	3	2	2	165
Grade 2	10	14	12	2	0	38
Grade 3	1	4	48	3	1	57
Grade 4	0	0	0	11	1	12
Grade 5	0	0	1	7	87	95
Total	160	27	64	25	91	367

TABLE 5. Confusion Matrices for the Staging of AMD Into the Five Severity Stages Defined in Table 1 Between Observer 2 and the Reference Standard

Reference Standard	Human Observer 2					
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Total
Grade 1	145	14	5	0	1	165
Grade 2	10	8	14	2	4	38
Grade 3	0	1	48	5	3	57
Grade 4	0	0	1	10	1	12
Grade 5	0	0	1	1	93	95
Total	155	23	69	18	102	367

specificity of 0.912. This compares favorably with the two human observers who achieved sensitivities of 0.970 and 0.994 at specificities of 0.897 and 0.872, respectively. We also evaluated the performance on the external set containing OCT scans acquired by a different OCT scanner, for which the amount of noise, image quality, and contrast varies strongly.⁴³ The performance of the proposed algorithm for the identification of high-risk cases reached an A_z of 0.978, similar to the performance obtained on the test set. This result showed that our automated algorithm is highly discriminative and generalizes well over different OCT scanners and imaging characteristics. Note that a higher performance can be achieved ($A_z = 0.993$ for the external set) if the algorithm is retrained with scans of similar characteristics and acquired with the same scanner, as we have shown in a previous study.²⁹

A higher scan density allows for a better prediction of the AMD severity grade and may even allow the use of 3D patches. However, scan spacing encountered in OCTs from different clinical settings and/or different scanners varies widely, diminishing the impact of using 3D information. The use of 2D patches provides the method robustness against these spacing changes, which is demonstrated in the similar performance obtained for data sets with different spacing (Fig. 6b).

For the proposed algorithm a rather large patch size is selected (61×61) as compared to the patch size (9×9) used in other classification methods based on BoW descriptors.^{33,44} Small patches are typically advised to allow a patch to function as a common building block. A possible hypothesis for the successful application of larger patches in the proposed algorithm is the homogeneity of the data due to the consistent structure of the retina, allowing larger patches to still be general enough to function as a common building block for a retinal OCT image. It can also be hypothesized that owing to their size, larger patches are better at capturing retinal pathology, which is difficult to capture in smaller image patches.

As noted earlier, small localized lesions might not contribute enough to the global BoW histogram for the algorithm to classify them correctly. For the current implementation, the saliency detector selects the same number of salient locations in every B-scan; modifying the saliency detector to focus more strongly on B-scans with pathology might remove this limitation.

The system has been shown to be highly robust to variations in image quality. This is shown by the performance on both the private dataset and also the external data set. The private data used to train and evaluate the algorithm are part of the EUGENDA consortium and are obtained from multiple institutes with varying imaging protocols, resulting in OCT volumes with varying scan-density, resolution, and noise levels due to different settings for the B-scan averaging parameter

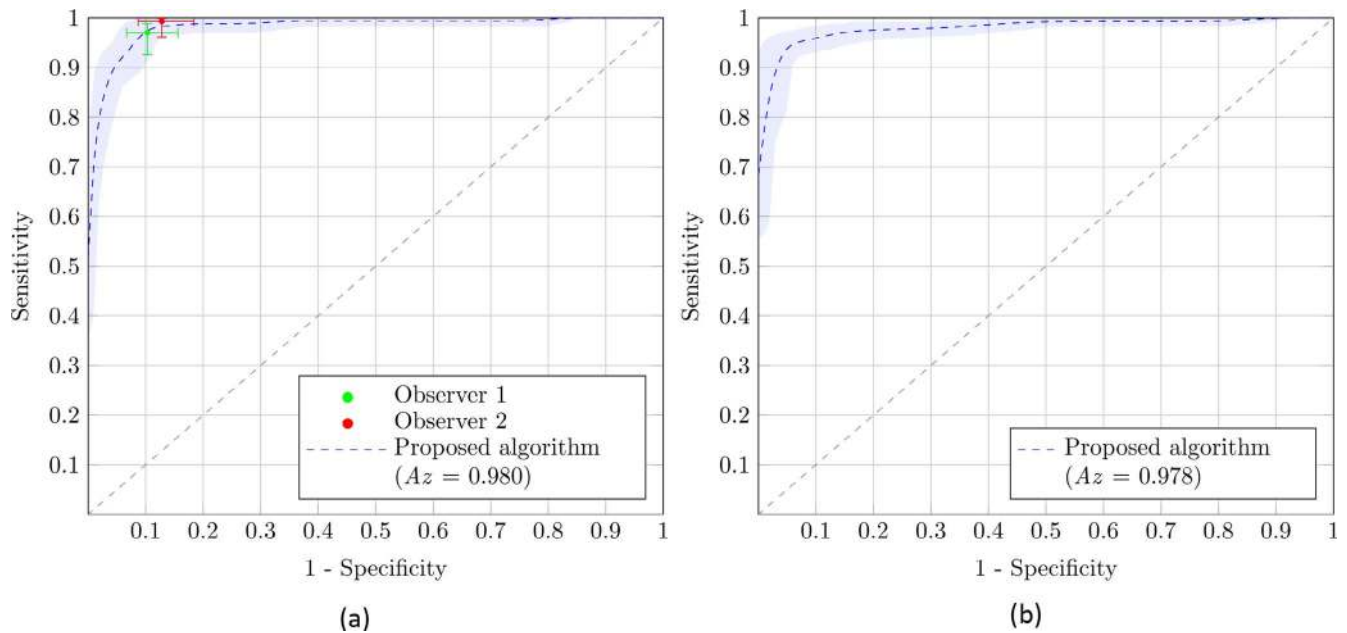


FIGURE 6. ROC curves of the proposed machine learning algorithm for AMD high-risk identification on (a) the test set and (b) the external set. The performance of the human observers in the test set is also included.

used in the Heidelberg Spectralis OCT scanner. The external set is obtained with a Bioptigen OCT scanner that does not implement B-scan averaging, resulting in B-scans with a substantially higher level of noise. The proposed system has been shown to be invariant to these quality variations by achieving a similar classification performance without the need for retraining the algorithm.

Considering the strengths and possible limitations of the developed automated classification algorithm, a few clinical applications are to be considered or are within reach. One such application would be the identification of AMD subgroups in large population studies. To gain more insight into risk factors and disease mechanisms involved in AMD, there is a need for detailed analysis of genotype-phenotype correlations.³¹ Manual identification of AMD subgroups in large studies is time-consuming and prone to error, as human grading can be subjective. An automated system does not suffer from fatigue or state of mind, and is therefore less prone to variability. The proposed system has performance in the range of human graders and can therefore be of major importance in

selecting homogeneous subgroups in such large population studies. Another possible application of the algorithm, based on the results described in “Experiment 2: AMD High-Risk Identification,” is the automated stratification of patients at high risk for AMD in a screening setting based on OCT imaging. An automated system could improve the efficacy of the ophthalmologist by separating out the easy-to-diagnose from the difficult-to-diagnose patients.

In conclusion, we developed a fully automated system to identify four different AMD stages and to discriminate these from healthy status. The system proved to have excellent performance compared to that of expert human observers on data from different OCT vendors in two distinct large data sets. Our data suggest this new algorithm allows for fast and reliable identification of homogeneous AMD subgroups on a large scale, for example, in population studies, multicenter data sets, and screening settings. Our automatic approach should therefore be considered as a reliable and cost-effective alternative for human graders in future AMD research.

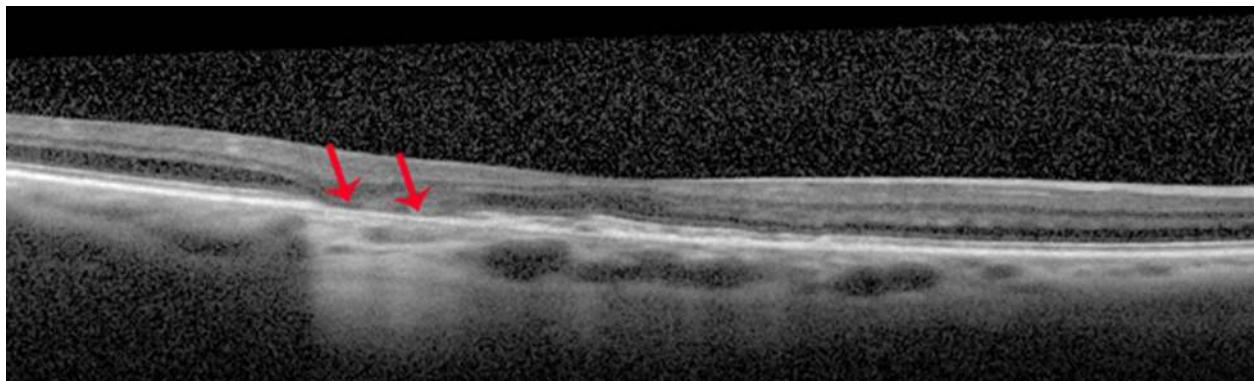


FIGURE 7. Example case that is misclassified by the machine learning algorithm. The OCT volume is classified as being early AMD, while a small but apparent GA lesion (indicated by the red arrows) is present.

Acknowledgments

Supported by the following foundations: Macula Degeneratie (MD) fonds, Landelijke Stichting voor Blinden en Slechtzienden (LSBS) fonds, and Oogfonds that contributed funds through UitZicht. The funding organizations had no role in the design or conduct of this research. They provided unrestricted grants.

Disclosure: **F.G. Venhuizen**, None; **B. van Ginneken**, None; **F. van Asten**, None; **M.J.J.P. van Grinsven**, None; **S. Fauser**, None; **C.B. Hoyng**, None; **T. Theelen**, None; **C.I. Sánchez**, None

References

- Bressler NM. Age-related macular degeneration is the leading cause of blindness. *JAMA*. 2004;291:1900-1901.
- Klaver CC, Wolfs RC, Assink JJ, van Duijn CM, Hofman A, de Jong PT. Genetic risk of age-related maculopathy: population-based familial aggregation study. *Arch Ophthalmol*. 1998;116:1646-1651.
- Rosenfeld PJ, Brown DM, Heier JS, et al. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2006;355:1419-1431.
- Martin DF, Maguire MG, Ying GS, Grunwald JE, Fine SL, Jaffe GJ. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2011;364:1897-1908.
- Enders P, Scholz P, Muether PS, Fauser S. Variability of disease activity in patients treated with ranibizumab for neovascular age-related macular degeneration. *Eye (Lond)*. 2016;30:1072-1076.
- Schmidt-Erfurth U, Waldstein SM. A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration. *Prog Retin Eye Res*. 2016;50:1-24.
- Kanagasigam Y, Bhuiyan A, Abramoff MD, Smith RT, Goldschmidt L, Wong TY. Progress on retinal image analysis for age related macular degeneration. *Prog Retin Eye Res*. 2014;38:20-42.
- Adhi M, Duker JS. Optical coherence tomography—current and future applications. *Curr Opin Ophthalmol*. 2013;24:213-221.
- Yehoshua Z, Rosenfeld PJ, Gregori G, Penha F. Spectral domain optical coherence tomography imaging of dry age-related macular degeneration. *Ophthalmic Surg Lasers Imaging*. 2010;41(suppl):S6-S14.
- van de Ven JP, Boon CJ, Smailhodzic D, et al. Short-term changes of Basal laminar drusen on spectral-domain optical coherence tomography. *Am J Ophthalmol*. 2012;154:560-567.
- Jain N, Farsiu S, Khanifar AA, et al. Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs. *Invest Ophthalmol Vis Sci*. 2010;51:4875-4883.
- Mokwa NF, Ristau T, Keane PA, Kirchhof B, Sadda SR, Liakopoulos S. Grading of age-related macular degeneration: comparison between color fundus photography, fluorescein angiography, and spectral domain optical coherence tomography. *J Ophthalmol*. 2013;2013:385915.
- Schuman SG, Koreishi AF, Farsiu S, Jung SH, Izatt JA, Toth CA. Photoreceptor layer thinning over drusen in eyes with age-related macular degeneration imaged in vivo with spectral-domain optical coherence tomography. *Ophthalmology*. 2009;116:488-496.e2.
- De Fauw J, Keane P, Tomasev N, et al. Automated analysis of retinal imaging using machine learning techniques for computer vision. *F1000Res*. 2016;5:1573.
- Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Trans Med Imaging*. 2010;3:169-208.
- Kafieh R, Rabbani H, Kermani S. A review of algorithms for segmentation of optical coherence tomography from retina. *J Med Signals Sens*. 2013;3:45-60.
- Quellec G, Lee K, Dolejsi M, Garvin MK, Abramoff MD, Sonka M. Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula. *IEEE Trans Med Imaging*. 2010;29:1321-1330.
- Hu Z, Niemeijer M, Abramoff MD, Garvin MK. Multimodal retinal vessel segmentation from spectral-domain optical coherence tomography and fundus photography. *IEEE Trans Med Imaging*. 2012;31:1900-1911.
- Esmacili M, Dehnavi AM, Rabbani H, Hajizadeh F. Three-dimensional segmentation of retinal cysts from spectral-domain optical coherence tomography images by the use of three-dimensional curvelet based K-SVD. *J Med Signals Sens*. 2016;6:166-171.
- Schlegl T, Waldstein SM, Vogl WD, Schmidt-Erfurth U, Langs G. Predicting semantic descriptions from medical images with convolutional neural networks. *Inf Process Med Imaging*. 2015;24:437-448.
- Zhang YG, Zhang BL, Coenen F, Xiao JM, Lu WJ. One-class kernel subspace ensemble for medical image classification. *Eurasip J Adv Sig Pr*. 2014;2014:1-13.
- Srinivasan PP, Kim LA, Mettu PS, et al. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed Opt Express*. 2014;5:3568-3577.
- Farsiu S, Chiu SJ, O'Connell RV, et al. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology*. 2014;121:162-172.
- Serrano-Aguilar P, Abreu R, Anton-Canalis L, et al. Development and validation of a computer-aided diagnostic tool to screen for age-related macular degeneration by optical coherence tomography. *Br J Ophthalmol*. 2012;96:503-507.
- Liu YY, Ishikawa H, Chen M, et al. Computerized macular pathology diagnosis in spectral domain optical coherence tomography scans based on multiscale texture and shape features. *Invest Ophthalmol Vis Sci*. 2011;52:8316-8322.
- Albarrak A, Coenen F, Zheng Y. Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction. *Proc MIUA*. 2013:59-64.
- Lee K, Buitendijk GH, Bogunovic H, et al. Automated segmentability index for layer segmentation of macular SD-OCT images. *Trans Vis Sci Tech*. 2016;5(2):14.
- Waldstein SM, Gerendas BS, Montuoro A, Simader C, Schmidt-Erfurth U. Quantitative comparison of macular segmentation performance using identical retinal regions across multiple spectral-domain optical coherence tomography instruments. *Br J Ophthalmol*. 2015;99:794-800.
- Venhuizen FG, van Ginneken B, Bloemen B, et al. Automated age-related macular degeneration classification in OCT using unsupervised feature learning. *Proc SPIE*. 2015: 941411.
- Fauser S, Smailhodzic D, Caramoy A, et al. Evaluation of serum lipid concentrations and genetic variants at high-density lipoprotein metabolism loci and TIMP3 in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52:5525-5528.
- van de Ven JP, Smailhodzic D, Boon CJ, et al. Association analysis of genetic and environmental risk factors in the cuticular drusen subtype of age-related macular degeneration. *Mol Vis*. 2012;18:2271-2278.
- Sivic J, Zisserman A. Efficient visual search of videos cast as text retrieval. *IEEE Trans Pattern Anal Mach Intell*. 2009;31:591-606.
- Avni U, Greenspan H, Konen E, Sharon M, Goldberger J. X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Trans Med Imaging*. 2011;30:733-746.

34. Leuschen JN, Schuman SG, Winter KP, et al. Spectral-domain optical coherence tomography characteristics of intermediate age-related macular degeneration. *Ophthalmology*. 2013;120:140-150.
35. Romeny MH. *Front-End Vision and Multi-Scale Image Analysis: Multi-scale Computer Vision Theory and Applications, Written in Mathematica*. Dordrecht, The Netherlands: Springer; 2003.
36. Forgy EW. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*. 1965;21:768-769.
37. Aly M, Munich M, Perona P. Multiple dictionaries for Bag of Words large scale image search. *Conf Proc IEEE Int Imag Proc*. 2011:1121-1123.
38. Pearson K. LIII: on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*. 1901;2:559-572.
39. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat*. 1992;46:175-185.
40. Tin Kam H. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell*. 1998;20:832-844.
41. Rifkin R, Klautau A. In defense of one-vs-all classification. *J Mach Learn Res*. 2004;5:101-141.
42. Efron B. 1977 Rietz lecture: bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;7:1-26.
43. Chen CL, Ishikawa H, Ling Y, et al. Signal normalization reduces systematic measurement differences between spectral-domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2013;54:7317-7322.
44. Coates A, Lee H, Ng AY. An analysis of single-layer networks in unsupervised feature learning. *AISTATS*. 2011;15:215-223.