RetCAD 1.3.0: White paper

18 July 2018

thir○na

Computer aided detection for Age-related Macular Degeneration and Diabetic Retinopathy

About this white paper

This white paper applies to RetCAD 1.3.0 and describes the general principles of the RetCAD software, explains matters like ROC curve, sensitivity, specificity, threshold values and how to interpret the results. Additionally, validations of the RetCAD software are included where the performance of the RetCAD software is compared to that of human expert observers.

Important notices

Thirona does not assume responsibility for the misuse of the RetCAD software. Since Thirona cannot control the use of the RetCAD software, it shall not be held responsible for any direct or consequential personal injury or damage.

Contents

Introduction

RetCAD was developed by Thirona in collaboration with the Radboud University Medical Center in Nijmegen, the Netherlands. RetCAD is a class IIa CE-certified software product using Artificial Intelligence that analyses color fundus images for the presence of Age-related Macular Degeneration (AMD) and Diabetic Retinopathy (DR).

RetCAD takes a CF (color fundus) image as input and produces several outputs. These output includes a quality assessment of the input image, heat maps indicating possible abnormal areas for both Age-related Macular Degeneration (AMD) and Diabetic Retinopathy (DR), and a score for both these retinal diseases that is monotonically related to the likelihood that these diseases are present in the image.

Users can take these outputs into account in their clinical work: they can decide if a new image should be acquired, in case the quality assessment indicates suboptimal image quality; they can decide to refer a patient for further testing for the presence of AMD, DR or other retinal abnormalities, in case the heat maps display suspicious regions that are verified by a human operator or when the scores are above certain thresholds.

ROC analysis

Definitions:

- **Sensitivity**: proportion of positive images (i.e. having an abnormality) correctly labelled as positive.
- **Specificity**: proportion of negative images (i.e. not having an abnormality) correctly labelled as negative.
- **ROC curve**: This curve is created by plotting the True Positive rate (sensitivity) against the False Positive Rate (1-specificity) at various threshold settings.
- **Az**: area under the ROC curve. This number estimates the probability of correct ranking positive/negative. It is bound between 0 and 1: the closest to 1, the better the system performance. 1 means a perfect classification.
- **T**: Threshold value. Different threshold values correspond to different points on the ROC curve. The point on the ROC curve closest to perfect classification (the upper left corner) is not necessarily the optimal threshold for the most cost-effective screening. **T** is the value set by the user to determine which images are labelled abnormal/normal.
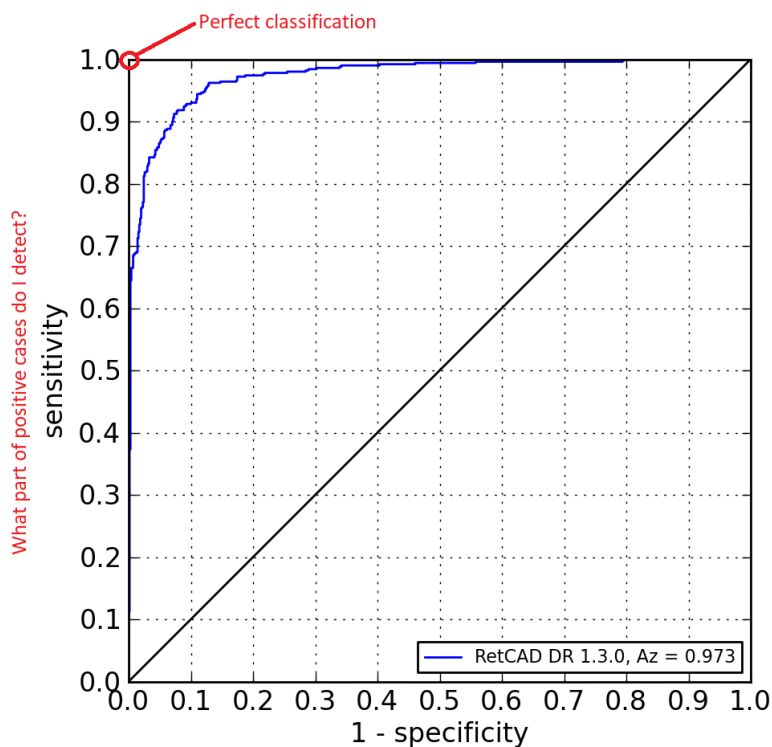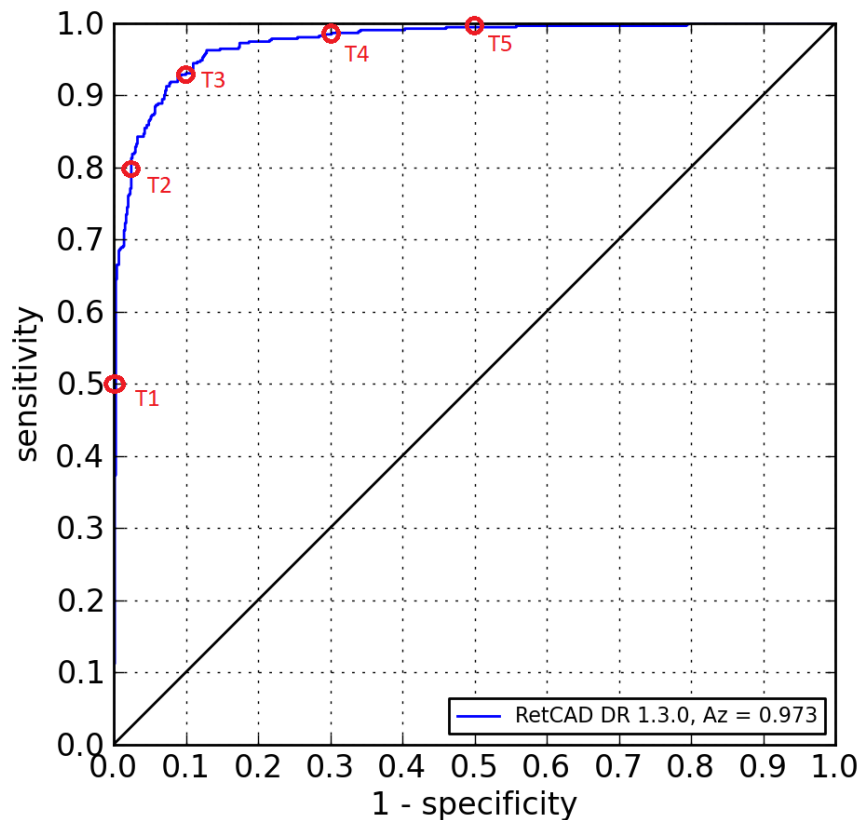
A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system:

- 0.90-1 = excellent (A)
- 0.80-0.90 = good (B)
- 0.70-0.80 = fair (C)
- 0.60-0.70 = poor (D)
- 0.50-0.60 = fail (F)

In this particular example, RetCAD DR Az is 0.973 which according to the above classification would be considered excellent (A).

Table 1: Different RetCAD threshold values with corresponding sensitivity and specificity levels.		
Threshold	True Positive Rate (sensitivity)	True Negative Rate (specificity)
T1	50 %	100 %
T2	80 %	97 %
T3	93 %	90 %
T4	98 %	70 %
T5	100 %	50 %

This can be presented graphically into an ROC curve

A relatively low threshold value of the software corresponds with a higher sensitivity, but at the cost of a lower specificity. A relatively high threshold value of the software corresponds with a higher specificity, but at the cost of a lower sensitivity. Hence, the threshold value is trade-off between sensitivity and specificity.

To summarize: An ROC curve demonstrates several things/characteristics of the test:
✓ It shows the trade-off between sensitivity and specificity (any increase in sensitivity will often be accompanied by a decrease in specificity).
✓ The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
✓ The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
✓ The area under the curve is a measure of test accuracy.

thir⦿na

RetCAD: How does it work?

RetCAD is software based on convolutional neural networks, a state-of-the-art technique in machine learning. In the process of analysing the input CF image, it compares regions in the image with regions extracted from normal and abnormal CF images. These latter images form the training data set of the software. A basic principle in machine learning is that the training data set is properly representative of the test data, otherwise the results may not be reliable.

CF cameras from different manufacturers produce different qualities of images because of hardware differences. In addition, image acquisition protocols can vary across acquisition sites, for example: the illumination, angular resolution (field of view) and the resolution of the image can vary. Furthermore, the patients imaged may originate from different populations in which the appearance of the retina, such as color and pigmentation, may vary.

Specific algorithms to normalize CF images prior to analysis have been included in the RetCAD software, but these algorithms can only reduce these differences to a certain extent and have limitations.

The input of RetCAD is a color fundus image in the RGB color space taken with white flash photography. The angular resolution should be between 30 and 55 degrees and the image should have even illumination across the retina. Furthermore, the image should have a minimal resolution of 1024x1024 pixels and the macula should be visible in the image. The input image should be in one of the following file formats: .dcm, .jpg, .jpeg, .tif, or .png.

RetCAD: Performance evaluation

The RetCAD software has been evaluated in several datasets. The images in these datasets were acquired using different types of CF cameras at different resolution. The performance of the RetCAD software is directly compared with that of (a) human expert(s). Each of the following sections describes an evaluation of a (publicly available) dataset. Firsts the details of the dataset are described after which the performance of the RetCAD software on this dataset is described and compared with human experts.

Messidor

The Messidor database consists of 1200 CF images which were acquired by three ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinography with a 45 degree field of view. The images were captured using 8 bits per color plane at 1440x960, 2240x1488 or 2304x1536 pixels. 800 images were acquired with pupil dilation (one drop of Tropicamide at 0.5%) and 400 without dilation. More information about the database can be found following the website link[1].

The images in the database were provided with DR grade by the medical experts. Four stages have been identified: No DR, mild DR, moderate DR and severe DR.

The RetCAD software was applied to each of the 1200 images in the dataset and the DR component of the RetCAD software was evaluated by comparing the RetCAD DR score with the DR severity grade as set by the medical experts.
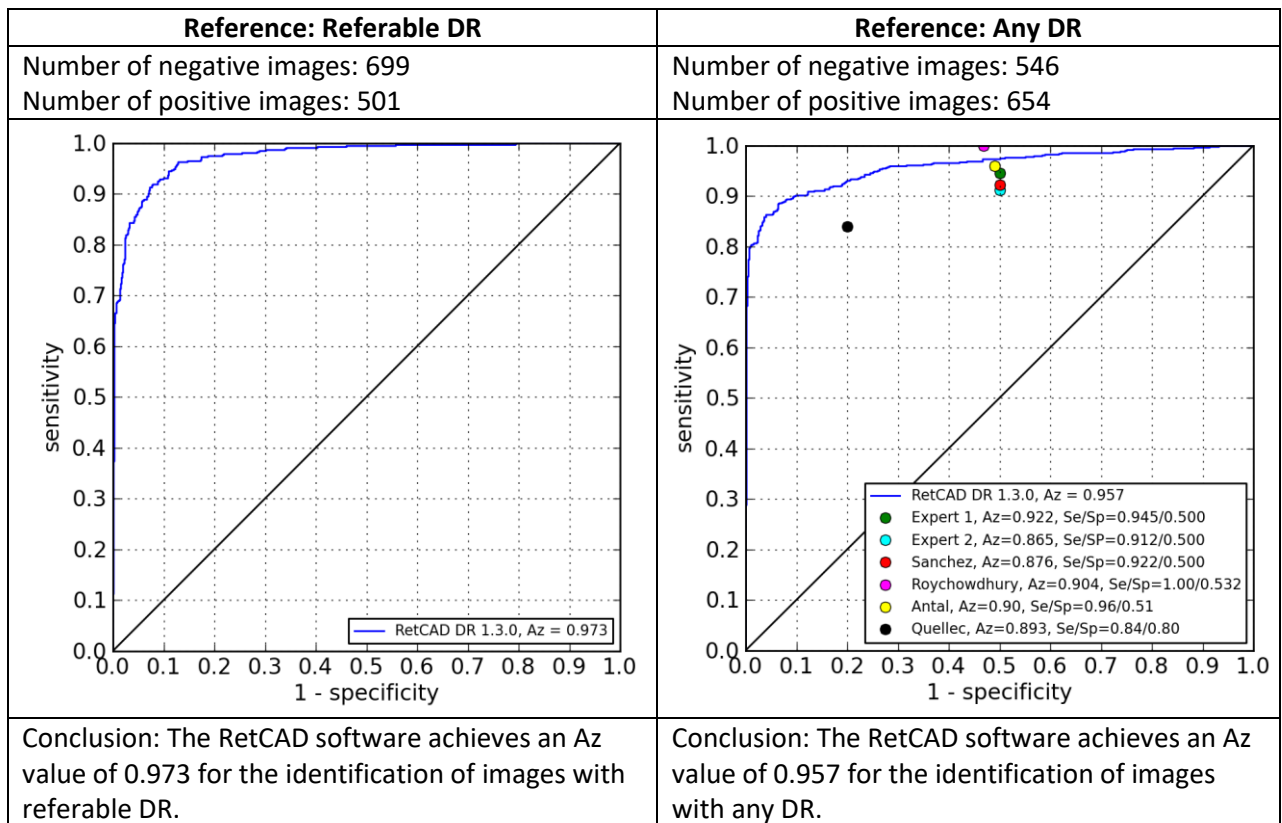
Two types of evaluations were performed:
1. The image was deemed positive if the grade was referable DR, i.e. severity level of moderate or severe DR.
2. The image was deemed positive if the grade was any DR, i.e. mild, moderate or severe DR.

The results of the evaluation are summarized in an ROC graph. In this ROC graph, the operating point of human experts is added whenever this was available.

* Note that the RetCAD software was not trained with any of the images that are part of the Messidor data set.

---

[1] *Kindly provided by the Messidor program partners
(see http://www.adcis.net/en/DownloadThirdParty/Messidor.html).*

Reference: Referable DR	Reference: Any DR
Number of negative images: 699 Number of positive images: 501	Number of negative images: 546 Number of positive images: 654

Conclusion: The RetCAD software achieves an Az value of 0.973 for the identification of images with referable DR.	Conclusion: The RetCAD software achieves an Az value of 0.957 for the identification of images with any DR.

Operating points for RetCAD DR

In Table 2, sensitivity and specificity values of RetCAD for DR detection are given at several threshold values for this specific dataset. The user can decide which threshold value suits his/her operating conditions best. See section "Definitions" for a more detailed explanation of the operating points.

Table 2: Operating points of RetCAD for any DR detection

Threshold	True Positive Rate (sensitivity)	True Negative Rate (specificity)
58	51 %	100 %
47	80 %	98 %
37	93 %	91 %
21	98 %	70 %
17	100 %	51 %

Comparison with other systems and human experts

Performance of other state-of-the-art DR detection systems and human experts on the Messidor database have been reported. The studies all use the criteria of "any DR" for positive cases, i.e. mild, moderate and severe DR are considered as the positive class. The table below reports their performances.

Table 3: Performance of software packages and human experts on Messidor

Author	Az value	Se/Sp	Year	Link
RetCAD DR	0.957	0.93/0.91	2018	-
Antal et al.	0.90	0.96/0.51	2012	http://arxiv.org/abs/1410.8577
Quellec et al.	0.893	0.84/0.80	2016	http://www.ncbi.nlm.nih.gov/pubmed/26774796
Roychowdhury et al.	0.904	1.00/0.53	2014	http://www.ncbi.nlm.nih.gov/pubmed/25192577
Sánchez et al.	0.876	0.92/0.50	2011	http://www.ncbi.nlm.nih.gov/pubmed/21527381
Expert 1	0.922	0.95/0.50	2011	http://www.ncbi.nlm.nih.gov/pubmed/21527381
Expert 2	0.865	0.91/0.50	2011	http://www.ncbi.nlm.nih.gov/pubmed/21527381

thirona

Messidor2

The Messidor-2 dataset is a collection of Diabetic Retinopathy (DR) examinations, each consisting of two macula-centered eye fundus images (one per eye). Part of the dataset (*Messidor-Original*) was kindly provided by the Messidor program partners (see http://messidor.crihan.fr). The remainder (*Messidor-Extension*) consists of never-before-published examinations from Brest University Hospital.

In the original Messidor dataset, some fundus images came in pairs, some others were single. *Messidor-Original* consists of all image pairs from the original Messidor dataset, that is 529 examinations (1058 images).

In order to populate *Messidor-Extension*, diabetic patients were recruited in the Ophthalmology department of Brest University Hospital (France) between October 16, 2009 and September 6, 2010. Eye fundi were imaged, without pharmacological dilation, using a Topcon TRC NW6 non-mydriatic fundus camera with a 45 degree field of view. Only macula-centered images were included in the dataset. *Messidor-Extension* contains 345 examinations (690 images).

Overall, Messidor-2 contains 874 examinations (1748 images). All patients in the database are graded for the presence of referable DR, i.e. moderate or more DR, by experts in the field. In total, 190 patients were graded as having referable DR, while 684 were graded as non-referable DR.

More information about the database can be found following the website link[2].

The RetCAD software was applied to each of the 1748 images in the dataset and the DR component of the RetCAD software was evaluated. In our evaluation, we took the highest score of the two images of a patient to be the patient based score for DR. This score, we compared with the provided reference scores as set by the medical experts and released by Abramoff et al. (https://www.ncbi.nlm.nih.gov/pubmed/27701631)

The results of the evaluation are summarized in an ROC graph. In this ROC graph, the operating point of human experts is added whenever this was available.

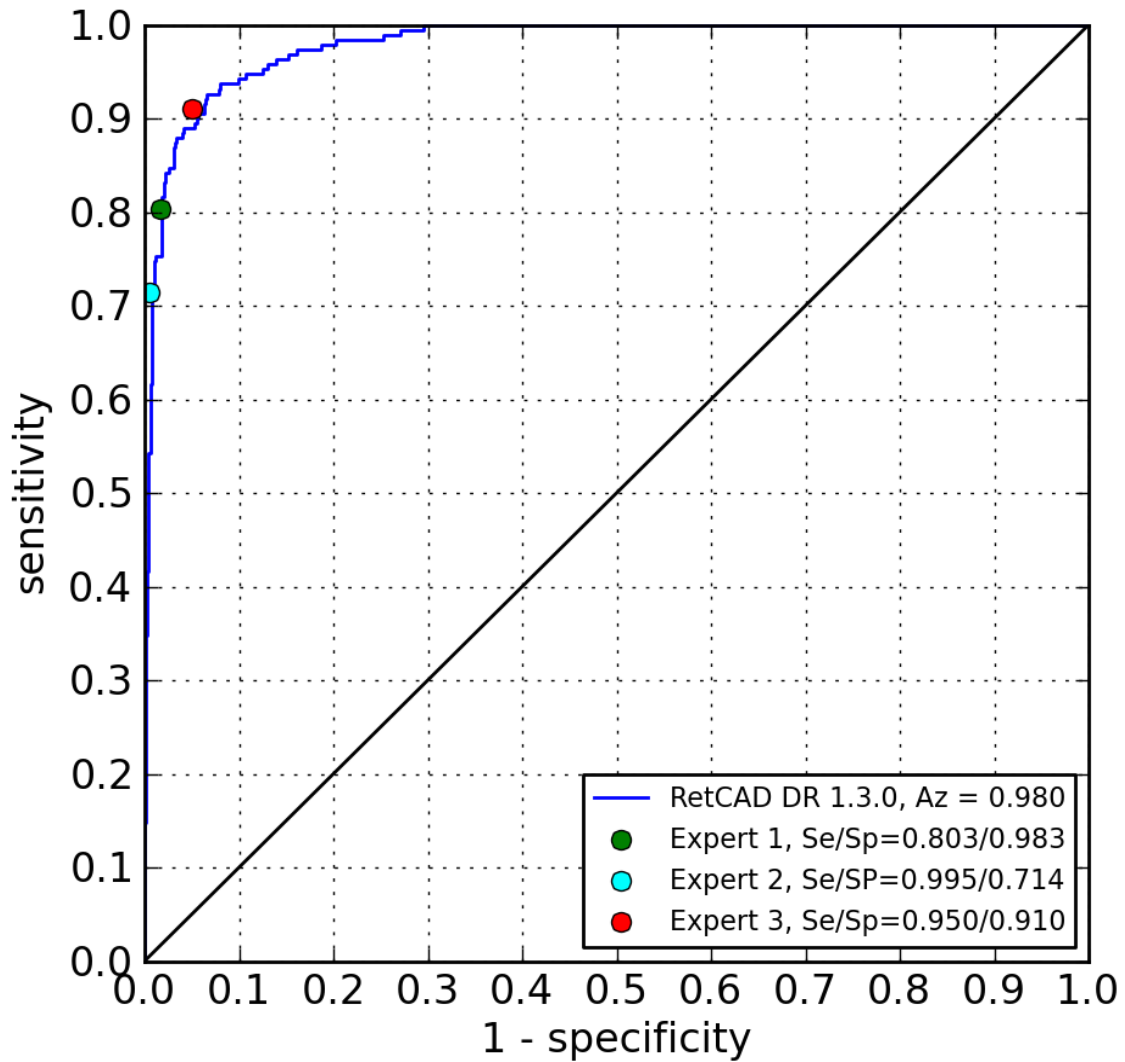* Note that the RetCAD software was not trained with any of the images that are part of the Messidor2 data set.

---

[2] Kindly provided by the LaTIM laboratory (see http://latim.univ-brest.fr/) and the Messidor program partners (see http://messidor.crihan.fr/)

Reference: referable DR
Number of negative patients: 684
Number of positive patients: 190

Conclusion: the RetCAD software achieves an Az value of 0.980 for the identification of patients with referable DR.

Operating points for RetCAD DR

In Table 4, sensitivity and specificity values of RetCAD for DR detection are given at several threshold values for this specific dataset. The user can decide which threshold value suits his/her operating conditions best. See section "Definitions" for a more detailed explanation of the operating points.

Table 4: Operating points of RetCAD for any DR detection		
Threshold	True Positive Rate (sensitivity)	True Negative Rate (specificity)
64	51 %	100 %
57	80 %	98 %
52	93 %	93 %
33	100 %	70 %
23	100 %	50 %

thirona

Comparison with other systems and human experts

Performance of other state-of-the-art DR detection systems on the Messidor2 database have been reported. Additionally, the performance of human graders were reported in one of these publications (Abramoff et al, 2013) and were added.

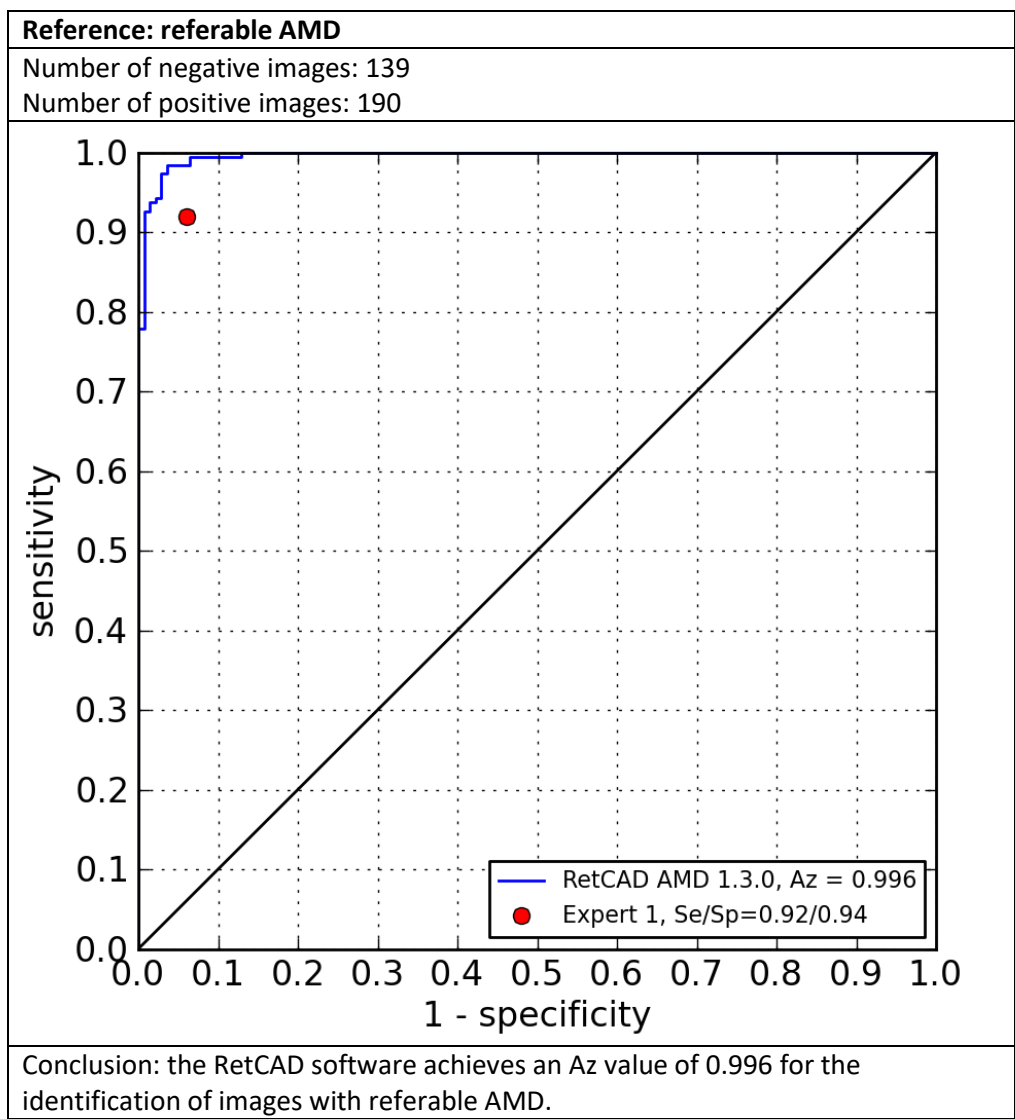| Table 5: Performance of other software packages on Messidor2				
Author	Az value	Se/Sp	Year	Link
RetCAD DR	0.980	0.93/0.93	2018	-
Abramoff et al.	0.937	0.97/0.59	2013	https://www.ncbi.nlm.nih.gov/pubmed/23494039
Abramoff et al.	0.980	0.97/0.87	2016	https://www.ncbi.nlm.nih.gov/pubmed/27701631
Expert 1	-	0.80/0.98	2013	https://www.ncbi.nlm.nih.gov/pubmed/23494039
Expert 2	-	1.00/0.71	2013	https://www.ncbi.nlm.nih.gov/pubmed/23494039
Expert 3	-	0.95/0.91	2013	https://www.ncbi.nlm.nih.gov/pubmed/23494039

Private1 in house dataset

The Private1 dataset is a dataset consisting of 329 macula centered images that were acquired at an ophthalmologic department using either a Topcon TRC 501X model digital fundus camera at 50 degree field of view or with a Canon CR-DGi model non-mydriatic retinal camera at 45 degree field of view. Pupil dilation was achieved with 1.0% tropicamide and 2.5% phenylephrine. All images were macula centered and image resolution varied between 1360x1024 to 3504x2336 pixels.

The images in the database were graded for presence of referable AMD by an experienced person with over 5 years of experience in grading fundus photographs. Referable AMD is defined as having at least 15 small drusen (>63μm) or more than one intermediate sized drusen (>126μm) or any sign of advanced AMD.

The RetCAD software was applied to each of the 329 images in the dataset and the AMD component of the RetCAD software was evaluated by comparing the RetCAD AMD score with the reference as set by the medical expert.

The results of the evaluation are summarized in an ROC graph. In this ROC graph, the operating point of a human expert (over 5 years of experience in grading fundus images) is added for comparison.

* Note that the RetCAD software was not trained with any of the images that are part of this data set.

Reference: referable AMD
Number of negative images: 139 Number of positive images: 190

Conclusion: the RetCAD software achieves an Az value of 0.996 for the identification of images with referable AMD.

Operating points for RetCAD AMD

In Table 6, sensitivity and specificity values of RetCAD for AMD detection are given at several threshold values for this specific dataset. The user can decide which threshold value suits his/her operating conditions best. See section "Definitions" for a more detailed explanation of the operating points.

Table 6: Operating points of RetCAD for AMD detection		
Threshold	True Positive Rate (sensitivity)	True Negative Rate (specificity)
67	50 %	100 %
43	80 %	99 %
23	98 %	96 %
20	100 %	84 %
17	100 %	47 %

Comparison with other systems and human experts

Table 7: Performance of other software packages on Messidor2				
Author	Az value	Se/Sp	Year	Link
RetCAD AMD	0.996	0.97/0.97	2018	-
Expert 1	-	0.92/0.94	2018	-

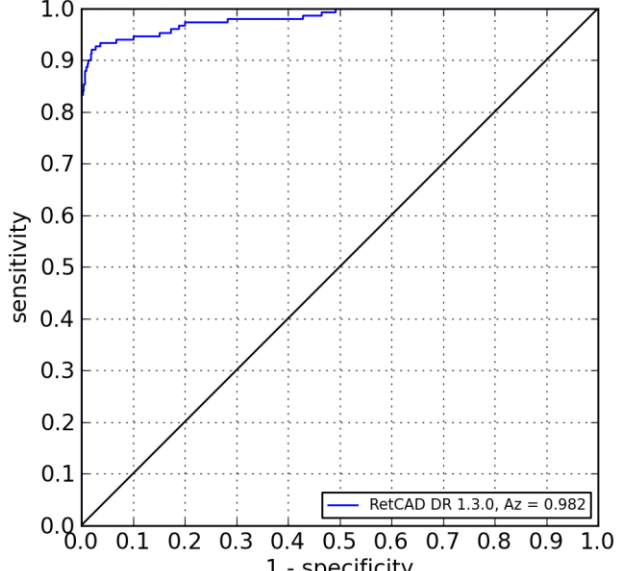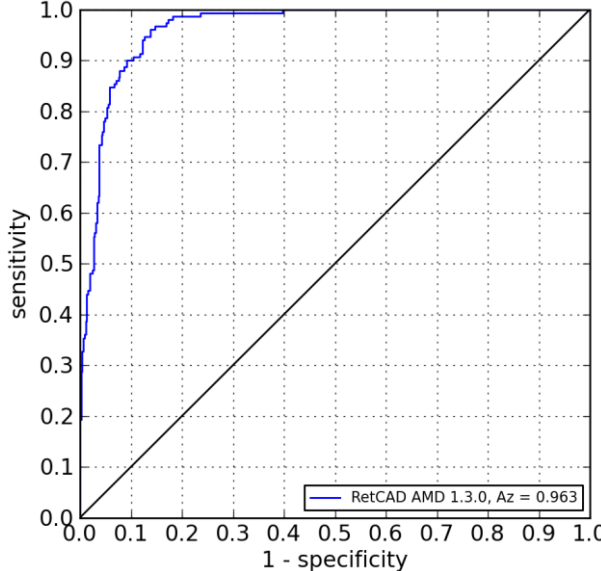thir⬡na

Private2 in house dataset

The Private2 dataset is a dataset consisting of 600 images that were acquired at an ophthalmologic department using a Canon CR-2PlusAF digital fundus camera at 45 degree field of view. No pupil dilating eye-drops were administered. Iimage resolution varied between 2376x1584 to 3456x5184 pixels. The patients that were images had either signs of AMD, DR or were healthy control subjects.

All images were scored for having referrable AMD and/or referrable DR by an human grader with 6 years of experience. Grading criteria were based on the ICDR and AREDS classifications. In total 150 images were graded as having referable AMD, 150 were graded as referable DR and 300 were graded as neither referable AMD nor referable DR.

The RetCAD software was applied to each of the 600 images in the dataset and both the AMD and DR component of the RetCAD software were evaluated by comparing the RetCAD scores with the reference as set by the medical expert.

The results of the evaluation are summarized in an ROC graphs, one for AMD and one for DR.

* Note that the RetCAD software was not trained with any of the images that are part of this data set.

Reference: Referable DR	Reference: AMD
Number of DR negative images: 450 Number of DR positive images: 150	Number of AMD negative images: 450 Number of DR positive images: 150
Conclusion: The RetCAD software achieves an Az value of 0.982 for the identification of images with referable DR.	Conclusion: The RetCAD software achieves an Az value of 0.963 for the identification of images with referable AMD.

Operating points for RetCAD AMD

In Table 8, sensitivity and specificity values of RetCAD for AMD and DR detection are given at several threshold values for this specific dataset. The user can decide which threshold value suits his/her operating conditions best. See section "Definitions" for a more detailed explanation of the operating points.

Table 8: Operating points of RetCAD for AMD and DR detection				
Threshold	True Positive Rate AMD (sensitivity)	True Negative Rate AMD (specificity)	True Positive Rate AMD (sensitivity)	True Negative Rate AMD (specificity)
55	71 %	96 %	48 %	100 %
45	84 %	94 %	67 %	100 %
35	90 %	90 %	80%	100 %
20	99 %	73 %	93 %	96 %
10	100 %	50 %	100 %	50 %

Comparison with other systems and human experts

No comparison has been made with other systems or human experts on this dataset.

thirona

RetCAD Requirements

Specific requirements for the images that are processed:
- Image format: .png, .jpg, .jpeg, .tif, .dcm
- Degree field of view: at least 30 degree with a maximum of 55 degree
- Digital images (scanned films are not usable/acceptable)

Information to be provided:
- Contact name and address
- Brand and type of color fundus camera
- Imaging protocol used to acquire the images
- Description of the settings in which the images are acquired (screening setting, prevalence survey, etc)

Contact

For questions and setting up a data transfer procedure, please contact:

- Dr. Mark van Grinsven (info@thirona.eu)