

# Evaluation of OCT biomarker changes in treatment-naive neovascular AMD using a deep semantic segmentation algorithm

Ben Asani<sup>2\*</sup>, Olle Holmberg<sup>1\*</sup>, Johannes B Schiefelbein<sup>2\*</sup>, Michael Hafner<sup>2</sup>, Tina Herold<sup>2</sup>, Hannah Spitzer<sup>1</sup>, Jakob Siedlecki<sup>2</sup>, Christoph Kern<sup>2</sup>, Karsten U. Kortuem<sup>4</sup>, Amit Frishberg<sup>1</sup>, Fabian J. Theis<sup>1,3\*</sup>, Siegfried G Priglinger<sup>2\*</sup>

(1) Institute of Computational Biology, Helmholtz Zentrum München

(2) Department of Ophthalmology, Ludwig-Maximilians-University, Munich, Germany

(3) Department of Mathematics, TU Munich, Germany

(4) Department of Ophthalmology, Ulm University, Ulm, Germany

**Word Count:** 5486

**Tables:** 2

**Financial support:**

This study was funded by a research grant from the German Federal Ministry of Education and Research (BMBF grant# 031L0210A) and by the Helmholtz Association Initiative and Networking Fund through Helmholtz AI [grant number: ZT-I-PF-5-01].

**Conflict of interest:**

Ben Asani has received speaker fees from Novartis Pharma GmbH. Olle Holmberg has ownership in Helsing GmbH. Christoph Kern has received travel fees from Bayer AG and Novartis GmbH. Consultant: Bayer AG. Karsten Kortuem has received speaker fees from Novartis GmbH, Bayer AG and Allergan. Jakob Siedlecki receives speaker honoraria and travel reimbursement from Carl Zeiss Meditec AG, Novartis Pharma GmbH, Bayer AG, Pharm-Allergan GmbH, Oculentis OSD Medical GmbH. Consultant: Bayer AG, Novartis Pharma GmbH, Pharm-Allergan GmbH. Travel reimbursement: Oertli AG. Fabian Theis consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, and Omniscope Ltd, and has ownership interest in Dermagnostix GmbH and Cellarity. Siegfried Priglinger receives speaker and consultant honoraria from and has served on advisory boards for Abbott, Alcon, Geuder, Oculus, Schwind, STAAR, TearLab, Thieme Compliance, Ziemer, Zeiss and research funding from Abbott, Alcon, Hoya, Oculentis, Oculus, Schwind and Zeiss.

**\*Authors contributed equally to this work**

**Corresponding Authors:**

Fabian Theis ([fabian.theis@helmholtz-muenchen.de](mailto:fabian.theis@helmholtz-muenchen.de)) & Siegfried G. Priglinger, MD, FEBO  
Department of Ophthalmology  
Ludwig-Maximilians-University Munich  
Mathildenstrasse 8  
80336 Munich, Germany  
E-Mail: [s.priglinger@med.uni-muenchen.de](mailto:s.priglinger@med.uni-muenchen.de)

**Keywords:**

neovascular AMD, age-related macular degeneration, OCT, anti-VEGF, artificial intelligence, machine learning, deep learning

**Running Head: (51 out of 60 allowed characters)**

Novel segmentation algorithm shows changes in biomarkers

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## **Abstract**

**Purpose:** To determine real life quantitative changes in OCT biomarkers in a large set of treatment naive patients undergoing anti-VEGF therapy. For this purpose, we devised a novel deep learning based semantic segmentation algorithm providing, to the best of our knowledge, the first benchmark results for automatic segmentation of 11 OCT features including biomarkers that are in line with the latest consensus nomenclature of the AAO for age-related macular degeneration (AMD).

**Design:** Retrospective study.

**Participants:** Segmentation algorithm training set of 458 volume scans as well as single scans from 363 treatment naive patients for the analysis.

**Methods:** Training of a Deep U-net based semantic segmentation ensemble algorithm leveraging multiple deep convolutional neural networks for state of the art semantic segmentation performance as well as analyzing OCT features prior to, after 3 and 12 months of anti-VEGF therapy.

**Main outcome measures:** F1 score for the segmentation efficiency and the quantified volumes of 11 OCT features.

**Results:** The segmentation algorithm achieved high F1 scores of almost 1.0 for neurosensory retina and subretinal fluid on a separate hold out test set with unseen patients. The algorithm performed worse for subretinal hyperreflective material and fibrovascular PED, on par with drusenoid PED and better in segmenting fibrosis. In the evaluation of treatment naive OCT scans, significant changes occurred for intraretinal fluid (mean:  $0.03\mu\text{m}^3$  to  $0.01\mu\text{m}^3$ ,  $p<0.001$ ), subretinal fluid ( $0.08\mu\text{m}^3$  to  $0.01\mu\text{m}^3$ ,  $p<0.001$ ), subretinal hyperreflective material ( $0.02\mu\text{m}^3$  to  $0.01\mu\text{m}^3$ ,  $p<0.001$ ), fibrovascular PED ( $0.12\mu\text{m}^3$  to  $0.09\mu\text{m}^3$ ,  $p=0.02$ ) and central retinal thickness C0 ( $225.78\mu\text{m}^3$  to  $169.40\mu\text{m}^3$ ). The amounts of intraretinal fluid, fibrovascular PED and ERM were predictive of poor outcome.

**Conclusions:** The segmentation algorithm allows efficient volumetric analysis of OCT scans. Anti-VEGF therapy provokes most potent changes in the first 3 months and afterwards only acts as a stabilizing agent. Furthermore, a gradual loss of RPE hints at a progressing decline of visual acuity even beyond month 12. Additional research is required to understand how these accurate OCT predictions can be leveraged for a personalized therapy regimen.

**Précis** (35 words max): Novel high performance segmentation algorithm shows most volumetric changes under anti-VEGF therapy in oct biomarkers occur in the first 3 months. Afterwards the injections seem only to serve as a stabilizing agent.

## **Introduction**

Vascular retinal diseases represent a heterogeneous field with multiple etiologies. In many countries age related macular degeneration (AMD) and diabetic retinopathy are the leading causes of blindness and are therefore of high socioeconomic importance<sup>1,2,3</sup>. Western lifestyle and the aging population are expected to contribute to a strong increase in the incidence of these diseases<sup>4</sup>.

While there are only limited therapeutic options for the dry form of AMD (dAMD), the pharmacological inhibition of the vascular endothelial growth factor (VEGF) has revolutionized treatment of neovascular AMD (nAMD) in the past two decades and has therefore become the gold standard in treatment. The intravitreal injection of anti-VEGF agents such as Ranibizumab (Lucentis®), Aflibercept (Eylea®), Bevacizumab (Avastin®) and the just newly approved Brolucizumab (Bevou®) targets the VEGF-guided growth of pathological choroidal neovascularisations that would normally lead to excessive intraretinal and subretinal fluid or excessive subretinal hemorrhages and lastly to the development of fibrotic tissue or geographic atrophy leading to high visual defects and central scotomas. While large pivotal studies for Lucentis® (MARINA-Study) and Eylea® (VIEW-Study) showed a visual acuity gain of 7.2 to 9.3 EDTRS letters after one to two years<sup>5,6</sup>, follow up studies observed a mean loss in visual acuity of 8.6 ETDRS letters after 7 years<sup>7</sup>. The major cause for this seems to be the development of retinal fibrosis in the context of a wound healing reaction that is shown to occur in approximately 50% of all patients after two years<sup>8</sup>. Apparently, a big part of the scar consists of fibrotically changed vascular membranes due to anti-VEGF therapy<sup>9,10</sup>. In addition to the limitations mentioned above, the high therapy intensity of the established anti-VEGF agents with monthly to bimonthly injections over several years poses a major challenge for both patients and the health care system. A recent study by Chopa et al shows an 11-fold increase in annual intravitreal injection from 2009 to 2019 and is projected to continue to rise<sup>11</sup>.

Besides fibrosis there are several other OCT biomarkers that seem to play an important role in determining the course of disease, the progression into advanced stages and response to treatment<sup>12–15</sup>. Generally, the influence or the importance of OCT biomarkers for visual acuity progression or progression into fibrosis in neovascular AMD is not fully understood. For example, recent studies have shown that ignoring subretinal fluid (<200µm at the foveal center) does not change outcome in visual acuity for patients under anti-VEGF therapy but can lessen their treatment burden<sup>16</sup>. Other studies by Christenbury et al and Dhrami-Gavazi et al have shown that a relatively stable fibrovascular pigment epithelial detachment may correlate with protection for the development of macular atrophy<sup>17,18</sup>. This shows that new findings in further analyzing morphologic features in these patients can give new insights to “older” treatment regimens. To understand changes in these features a basic descriptive understanding of biomarkers in treatment naive patients in a real life setting is crucial. Study conditions can lead to “selection bias” (i.e. through inclusion criteria) and can not offer the same insight into the variety of patients or their compliance as do real field studies. Manually analyzing these treatment naive biomarkers and their changes over time becomes difficult and time consuming once we want to study a greater set of patients. Additionally, the variety of morphologic features in AMD add much to the complexity of the scan and generate high volume data making the manual process close to impossible. Hence, the use of automatic segmentation can offer a great deal of help in analyzing, describing and understanding the changes of the OCT biomarkers as well as their influence on the overall therapeutic outcome. In this respect, deep learning algorithms have already been shown to successfully automate tasks such as disease classification of color fundus photographs and retinal OCT images, a key non-invasive imaging modality in Ophthalmology, as well as support medical decision making in many other medical fields<sup>19–23</sup>. Further, in the field of deep semantic segmentation, the task of classifying individual pixels in an image with numerous applications in medicine, many advances have been made in both natural scenes as well as biomedical applications<sup>24–27</sup>. In this project we propose a deep learning based semantic segmentation algorithm trained with 458 manually annotated macular OCT scans, to allow automatic segmentation, and thus automatic labeling, and volumetric analysis of clinical features of a large series of treatment

naive patients eyes undergoing anti-VEGF treatment. It furthermore allows us to give a one-of-a-kind extensive description on the distribution of morphologic features as well as disease biomarkers of that patient group in a real life setting. To the best of our knowledge, we are the first to present an extensive AMD segmentation algorithm including disease biomarkers and other common morphological OCT features as well as validate this on independent test sets of previously unseen patients combined with detailed analysis of the inter-annotator variance for ambiguous and hard to annotate features.

## **Methods**

This case series included patients with newly diagnosed treatment naive neovascular age related macular degeneration in the study eye and had a follow up observational period of at least three and twelve months. Treatment naive was defined as never having had any form of intravitreal injection (anti-VEGF or other) in our institution or elsewhere. The first intravitreal injection was received at time of diagnosis. Exclusion criteria were comorbidities such as central retinal vein occlusion, retinal branch occlusion, diabetic macular edema, uveitis and other conditions that can lead to the development of intra- or subretinal fluid such as macular telangiectasia, central serous chorioretinopathy and intravitreal injection of any other drug than anti-VEGF. The study was approved by the institutional review board of our institution and adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained from each participant prior to the intervention and all testing outlined herein.

### *Treatment regimen*

Patients received an upload of three monthly injections of any of the anti-VEGF agents (Ranibizumab, Aflibercept or Bevacizumab) and were then treated according to the Treat and Extend regimen: They were either extended for two weeks or continued on a monthly injection routine<sup>28</sup>.

### *Patient identification*

To identify patients suffering from a neovascular AMD, we queried our data warehouse for all patients receiving intravitreal Injections of anti-VEGF between 2013/03/11 and 2020/07/09. Diagnosis of neovascular AMD was confirmed after proof of choroidal neovascularization in initial Fluorescein angiography. We interpolated the data set to 3 measurement points: Start date is the time of the first intravitreal injection. Next monitoring point is after 3 months and lastly after 12 months of treatment.

### *Preoperative examinations*

Examinations before intravitreal injections included testing best corrected visual acuity using standard ETDRS chart at testing distance of 4 meters, intraocular pressure using non contact tonometry, dilated indirect fundoscopy as well as spectral-domain optical coherence tomography of the macula (Spectralis; Heidelberg Engineering GmbH, Heidelberg, Germany). The metrics “counting fingers”, “hand movement” were converted to 1.98 and 2.28 logMar respectively as previously described by Lange C et al and Schulze-Bonsel et al<sup>29,30</sup>. There were no patients included with the metric “light perception”. All visual acuity values in this study are given in logMAR units.

### *Segmentation data sets*

To create the segmentation algorithm, a set of 458 macular OCT scans, each from a different patient, were annotated by ophthalmological residents with a fellowship in medical retina (B.A., J.B.S. and M.H.) using the annotation tool LabelMe<sup>31</sup>. They were then validated by three retinal

experts (J.S., C.K. and T.H.) and re-labeled in case of any discrepancies. Each fellow was assigned his or her own set, roughly a third of the 458 scans, randomly. In this process, we followed the Consensus Nomenclature for Reporting Neovascular Age-Related Macular Degeneration of the AAO (American Academy of Ophthalmology) for disease biomarkers and used 11 different OCT labels<sup>32</sup> also including other common morphological features giving us the possibility to analyze the scan as a whole and not exclusively for the biomarkers. The annotation was made pixel wise, i.e. each pixel in the image was assigned one of the 11 classes: Epiretinal Membrane (ERM), Neurosensory Retina (NR), Retinal pigment epithelium (RPE), Intraretinal fluid (IRF), Subretinal fluid (SRF), Subretinal hyperreflective material (SHRM), Drusenoid pigment epithelial detachment (DPED or Drusenoid PED), Fibrovascular pigment epithelial detachment (FPED or Fibrovascular PED), Fibrosis, Choroid, Posterior hyaloid membrane (PHM). For annotation examples as well as exact class distribution statistics see supplement. Further the class camera effect and the vitreous cavity were automatically labeled. For camera effect this was derived as completely black regions in the OCT images and for vitreous cavity as background above the neurosensory retina. The central retinal thickness C0 was defined as the quantification (in  $\mu\text{m}^3$ ) of the labeled neurosensory retina in that region according to the ETDRS grid<sup>33</sup>.

An additional data set of 30 scans were sampled and then annotated to measure the inter annotator variation between the three annotators for a selection of ambiguous features. To quantify feature ambiguity as well as establishing an upper bound for how well these features can be expected to be segmented, the ophthalmological fellows all annotated the same scans. After doing so, a consensus annotation was produced by a panel including the retinal experts by either annotating a new scan or selecting one of the annotated scans from the annotators. In total, each of the 30 scans in the second data set had three annotations from the three fellows in medical retina as well as a fourth annotation produced by consensus voting.

### *Model architecture*

The model used is a deep ensemble<sup>34</sup> of networks, for increased generalization in the presence of limited number of samples<sup>35</sup>, for segmentation. Each network in the ensemble is the same, just trained with different random weight initializations. The architecture used for the models in the ensemble is described below.

The algorithm used for segmenting the retinal OCT scans is a deep convolutional neural network (CNN)<sup>36</sup> of a U-net type architecture<sup>24</sup>. Specifically the network consists of eleven convolution layers followed by batch normalization<sup>37</sup> and relu activation functions<sup>38</sup>. The convolutional layers use padding so as to not alter the dimensions of the feature maps and have kernel size set to three throughout the network. Each convolutional layer is initialized using the He normal initialization<sup>39</sup> at the start of the training. In the encoder, every two convolutional layers are followed by a max pooling operation making a total of five max poolings, reducing the resolution size of the input from 256 to 8 for the feature maps. Here, the original images are linearly resized from 512 to 256 pixels height and width. The first convolutional layer is set to have 64 filters and this number is doubled after every max pooling layer yielding a maximum of 1024 filters in the bottleneck of the architecture. The number of filters are then halved after every transposed convolution in the decoder. Between the encoder and the decoder a dropout layer with probability 0.2 is applied for regularization. In the decoder, transposed convolutions as well as two layered convolutional blocks as described above are applied consecutively until the original input dimension is reached. Finally, a convolutional layer with kernel size one and a softmax activation function is applied to achieve the final output of the network.



### *Model training*

The models were trained using the Adam optimizer using the categorical cross entropy loss, with an initial learning rate of 0.001, found to be optimal through hyper parameter tuning on a validation set. In total, each network was trained for 2000 epochs. During training, random flipping, rotation, brightness and gaussian noise, with a scale of (2.5, 12.75), were used to augment the data. Each with a probability of 0.2 of being applied to an image or image and mask, except for rotation which was assigned a probability of 0.5 of occurring. The images and annotation masks were split into a train, validation and test set consisting of 338, 84 and 36 images respectively. Further, an ensemble of networks was created for each image in the 36 image test set using a leave-one-out validation scheme and adding the remaining test images to the training data set. In total five models were trained for each test image resulting in 180 different models. At inference time, the softmax outputs for each class and pixel, from all five models, were averaged to obtain an ensemble prediction. The class with the highest average softmax score yielded the final prediction. For the inter doctor variance data set the 10 models, out of 180, with the best validation scores were selected to form an ensemble from which the predictions were obtained in the same way as for the test set.

### *Model evaluation*

The segmentation model is evaluated using the F1 score, i.e. the harmonic mean between precision and recall, a standard evaluation metric for semantic segmentation tasks. The score is then presented for each class. Further the inter doctor variance is presented as the F1 score between each annotator retinal expert, also called annotator and the 10 model ensemble against the consensus annotation. As all pixels are concatenated for all images, as typically done in semantic segmentation tasks, no standard deviation metrics between images are provided. The statistically evaluated model was then used for automatic segmentation of 18522 OCT scans from 378 eyes enabling the statistical analysis of morphological OCT features including nAMD biomarker distribution on treatment naive patients and how they are affected by anti-VEGF injections.

### *Statistical analysis*

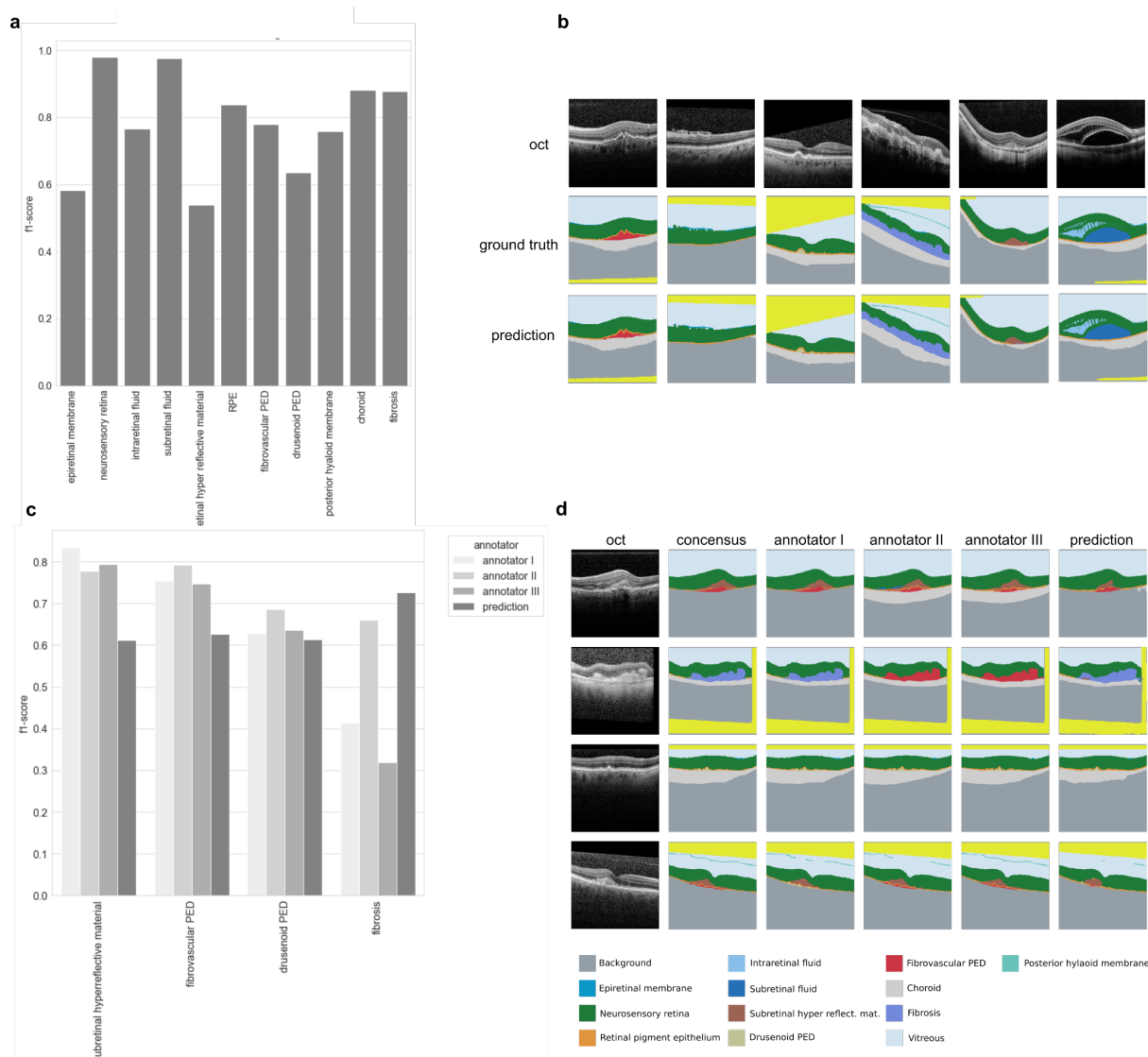
All statistical analysis was performed using the python programming language with the scipy stats software package<sup>40</sup>. Normality of data was assumed due to sufficiently large sample sizes well above 30 as specified by the central limit theorem<sup>41</sup>. We applied the independent samples t-test for parametric comparisons. The level of statistical significance was defined as  $p < 0.05$ .

The code for the models and training procedures as well as result analysis will be made available through the public Github repository upon publication.

## **Results**

**Deep learning segmentations accurately quantifies presence of clinical features in retinal OCT images.** The segmentation algorithm was trained using 338 doctor-annotated OCT scans from AMD patients to infer the volumes of 11 different clinical retinal features (See **Methods**). The segmentation algorithm segmented the clinical features with a top performance of 0.98 F1-score for features neurosensory retina and subretinal fluid. The lowest F1-scores were observed for features epiretinal membrane, drusenoid PED and subretinal hyperreflective material (see Fig. 1a). Test set examples of the segmented features can be seen in figure 1b, showing the variety of features accurately segmented in unseen patients. The annotators, on the other end, highly agreed on subretinal hyperreflective material, fibrovascular PED and drusenoid PED, while largely disagreeing in the case of fibrosis (see Fig. 1c). Overall, the segmentation algorithm performed worse than the annotators on

subretinal hyperreflective material and fibrovascular PED, on par with respect to drusenoid PED and better than all annotators when segmenting fibrosis (see Fig. 1c). In figure 1d we see examples of consensus, annotator and segmentation algorithm predicted segmentations for example OCT images. One can clearly see how fibrosis is annotated as fibrovascular PED by annotators II and III and while detected, subretinal hyperreflective material is overall segmented differently by the algorithm than the annotators and consensus segmentations.



**Figure 1: Deep learning segments clinical features on par with human experts from retinal OCT images.** (a) F1 scores for 11 clinical features segmented on a test set from 37 to the algorithm previously unseen patients. (b) Example of OCT images selected to illustrate various segmentation classes with ground truth and predicted segmentation maps (c) F1 scores for subretinal hyperreflective material, fibrovascular PED, drusenoid PED as well as fibrosis on 30 challenging test patients containing these features. (d) Example OCT images with consensus ground truth, annotations from three different annotators as well as predicted segmentation maps displaying segmentation of multiple features.

**Most morphological changes of treatment naive patients occur during the first three months of anti-VEGF therapy.** Through our data warehouse we identified 592 eyes of 536 patients, out of 1656 AMD patients in total, that were treatment naive and had the necessary follow up visits to fit the study. After filtering for the inclusion criteria as described in the

methods sections (i.e. follow-up time) a total of 378 eyes consisting of 18522 OCT scans from 339 different patients were segmented and analyzed. Of those patients 144 were male and 234 female with an average age of  $82 \pm 8$  years. Pathological biomarkers in treatment naive patients that were most prominent on initial presentation were by far fibrovascular pigment epithelial detachment (Mean:  $0.12 \mu\text{m}^3$ , SD:  $0.19 \mu\text{m}^3$ ), followed by subretinal fluid (Mean:  $0.08 \mu\text{m}^3$ , SD:  $0.26 \mu\text{m}^3$ ), intraretinal fluid (Mean  $0.03 \mu\text{m}^3$ , SD  $0.08 \mu\text{m}^3$ ) and subretinal hyperreflective material (Mean:  $0.02 \mu\text{m}^3$ , SD:  $0.05 \mu\text{m}^3$ , see Fig 2). The mean central retinal thickness C0 (CRT) was  $225.78 \mu\text{m}$  with SD  $90.36 \mu\text{m}$  in the cohort of treatment naive patients. After 3 and 12 months of intravitreal treatment the CRT decreased compared to baseline (3 months: mean  $169.40 \mu\text{m}$ , SD  $64.50 \mu\text{m}$ , p-value  $< 0.001$ ; 12 months: mean  $169.42 \mu\text{m}$ , SD  $59.03 \mu\text{m}$ , p-value  $< 0.001$ ) which is mostly related to the resolution of fluids. Mean number of injections was  $3.8 \pm 1.5$  after 3 months and  $8.3 \pm 3.5$  after 12 months, meaning that on average the therapy regimen was extended at some point during the observed time frame.

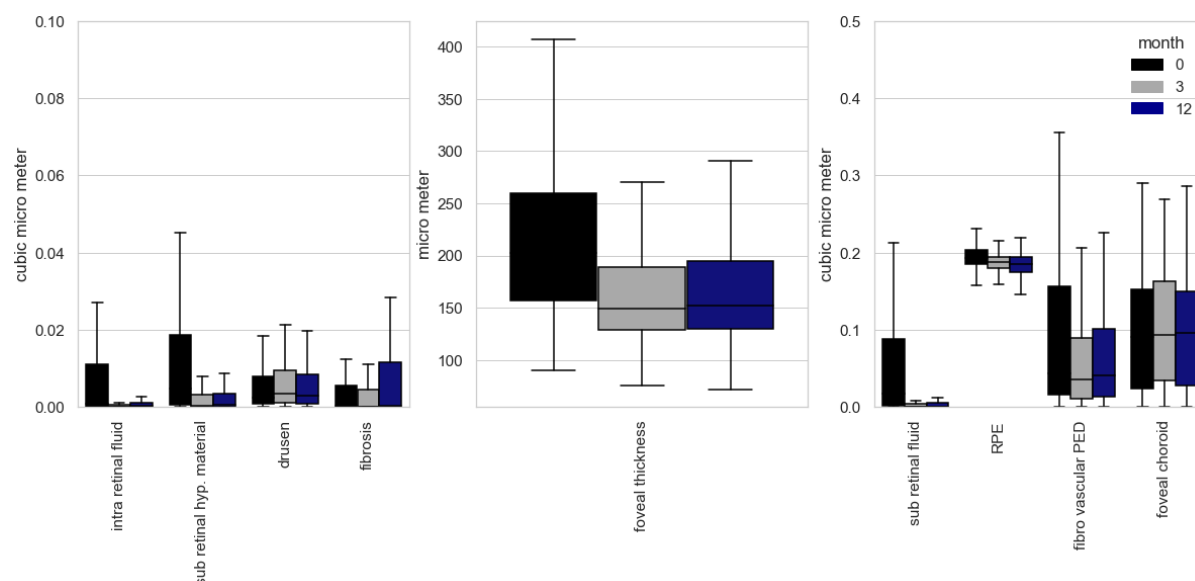
After 3 months of intravitreal treatment there was little but significant change in fibrovascular PED from  $0.12 \mu\text{m}$  down to  $0.09 \mu\text{m}$  (p-value = 0.02) but no difference afterwards (mean 12 months:  $0.10$ , p-value from month 3-9: 0.62). The overall change of this marker over the course of twelve months was barely not statistically significant (p-value: 0.05) meaning that therapy in general did not lead to real regression of the fibrovascular PED over a long term treatment regimen but rather contributes to a stabilization of this biomarker.

Difference in means as tested by t-tests showed a statistical significant change in IRF ( $p < 0.001$ ), SRF ( $p < 0.001$ ), SRHM ( $p < 0.001$ ) and CRT C0 ( $p < 0.001$ ) after 3 months of anti-VEGF therapy but no significant improvement or change after that initial upload of 3 injections until month 12 occurred, meaning setbacks under therapy were uncommon and no further improvement could be reached. This was on par with the observed changes in visual acuity. Significant improvement was achieved from before treatment during the first 3 months (mean VA on month 1: 0.59 logMAR, mean VA on month 3: 0.50 logMAR; p-value: 0.003) but no significant change occurred afterwards up until month 12. The distribution of these and the other morphological features including disease biomarker amongst treatment naive patients is summarized in Figure 2 and Table 1. The RPE feature changed from initially  $0.192 \mu\text{m}^3$  to  $0.185 \mu\text{m}^3$  after 3 months ( $p < 0.01$ ) and lastly  $0.181 \mu\text{m}^3$  after 12 months (p-value month 3-12: 0.01) meaning there was a significant loss of RPE therefore hinting at a steady increase in atrophy. However, the cause-relationship between intravitreal injections and atrophy remains unclear as the resulting loss of RPE could be attributed to a natural disease progression. There was no change in choroidal thickness in 12 months even though there was a gradual loss in RPE.

**Table 1.** Mean volumetric values before (mean 1), three months (mean 3) and twelve months after initiation of anti-VEGF therapy. All values are in  $\mu\text{m}^3$ .

	erm	irf	srf	srhm	rpe	fvped	drusen	phm	choroid	fibrosis	crt
mean 1	0.01	0.03	0.08	0.02	0.19	0.12	0.01	0.02	0.93	0.02	223.82
mean 3	0.01	0.01	0.01	0.01	0.19	0.09	0.01	0.02	0.96	0.02	169.65
mean 12	0.01	0.01	0.01	0.00	0.18	0.09	0.01	0.02	0.95	0.03	172.62
sd 1	0.02	0.08	0.25	0.05	0.02	0.18	0.01	0.03	0.41	0.06	88.29
sd 3	0.02	0.07	0.09	0.02	0.02	0.15	0.01	0.03	0.41	0.06	64.11
sd 12	0.02	0.04	0.04	0.01	0.02	0.15	0.01	0.03	0.41	0.08	63.76
p-value (1-3)	0.62	0.00	0.00	0.00	0.00	0.02	0.44	0.39	0.33	0.59	0.00
p-value (1-12)	0.38	0.00	0.00	0.00	0.00	0.05	0.82	0.24	0.50	0.15	0.00
p-value (3-12)	0.70	0.94	0.86	0.32	0.01	0.59	0.30	0.74	0.76	0.05	0.52





**Figure 2.** Volumetric changes of treatment naive OCT biomarkers under anti-VEGF treatment. Significant reduction was seen in IRF, SRF, SHRM, RPE and central foveal thickness.

We additionally used linear regression to model the interdependence between the different biomarkers and visual acuity outcome in 12 months. In treatment naive patients there was a negative correlation (meaning more of the feature was correlated with poor visual acuity outcome) between visual acuity and intraretinal fluid ( $p < 0.001$ ), epiretinal membrane ( $p = 0.013$ ), fibrovascular PED ( $p = 0.009$ ) and fibrosis ( $p = 0.049$ ). Since most of these features are most prominent in the active disease state (excluding epiretinal membrane) and bring immense structural changes to the retina, this finding does not seem very surprising. Interestingly, subretinal fluid had no correlation with a worse outcome in visual acuity, supporting the thesis of being protective against retinal atrophy and further decline of vision. Other findings are summarized in Table 2.

Table 2. Results from multiple regression analysis of oct markers and their effect on 12 month visual acuity ( $n=336$ ).

Model predictors	b	SE b	t	p
intercept	0.64*	.23	2.82	.01
erm	2.5**	0.90	2.51	0.01
irf	1.24***	0.29	4.32	0.00
srf	-0.15	0.09	-1.73	0.08
srhm	0.64	0.47	1.35	0.18
rpe	-1.58	1.25	-1.26	0.21
fvpd	0.37*	0.14	2.64	0.01
drusen	-1.19	2.06	-0.58	0.56
phm	-0.36	0.57	-0.63	0.53
choroid	0.02	0.05	0.55	0.59
fibrosis	1.03	0.52	0.05	0.05
crt	0.17	0.38	0.65	0.65
Adj. R <sup>2</sup>	.189			
F	8.08***			

Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

## **Discussion**

Structural changes and understanding of OCT images have been part of many major studies in the past. OCT imaging has changed AMD diagnostics and shaped the way patients are introduced into their treatment plans. To our knowledge, there is still missing an extensive morphologic description including disease biomarkers of never before treated nAMD patients. Most studies lack a full volumetric/quantitative interpretation of the images as manual segmentation seems unfeasible with high volume data. Therefore, in this study we introduced and showed the feasibility of a novel and high performing segmentation algorithm to present a thorough descriptive analysis of biomarkers in treatment naive nAMD patients in a real life setting and their overall plasticity after anti-VEGF treatment. Furthermore we identified several biomarkers that seem to contribute to a reduced prognosis in best corrected visual acuity after 12 months giving us a glance at predicting disease outcome.

### **Contribution to the application of deep learning algorithms in understanding the Retina.**

Semantic segmentation of fluids, PEDs as well as retinal layers in retinal OCT images has been studied before<sup>42–45</sup>. To the best of our knowledge, this is the first project aiming to segment features as defined in the latest Consensus Nomenclature for Reporting Neovascular Age-Related Macular Degeneration of the AAO (American Academy of Ophthalmology) and uses 11 different OCT labels<sup>32</sup>. The only similar segmentation algorithm to our knowledge was developed by De Fauw and colleagues<sup>22</sup> which used a similar segmentation network as a pre-processing method for predicting retinal disease. Our work differs with regards to two main aspects. First, we include the latest consensus nomenclature as specified by the AAO, and second, we provide an in-depth analysis of feature ambiguity by inter annotator variation analysis and test performance metrics reporting on a large number of unseen test patients. We show that segmentation of the above features is indeed possible to a high degree of accuracy (see Fig. 1a). While the algorithm segments at a similar proficiency as the annotators and even outperforms the retinal experts at segmenting fibrosis, we also clearly show the current limitation of segmentation of the retina by quantifying the ambiguity of the features fibrosis, drusen and subretinal hyperreflective material (see Fig. 2).

These features, which mainly occur in the subretinal region, can often, due to lower image quality, be hard to distinguish as they often share hyperreflective parts that can be mistaken for one another. These difficulties specifically present in transitions from fibrovascular PED to fibrosis as they would appear very ambiguous on OCT scans. For example, Ohayon et al segmented fibrovascular PED into three layers with a more hyperreflective layer 2 that did not respond as efficiently as the other layers following anti-VEGF treatment being suggestive of a fibrotic component in the PED<sup>46</sup>. Interestingly in accordance to that, the segmentation algorithm presented in this study occasionally segmented the fibrovascular PED into a fibrotic sub-compartment as well, making the segmentation process in this case potentially more accurate than the training set provided by the retinal fellows as they had to make a definite decision on which biomarker was present for the whole structure. However, this warrants further investigation. Lastly, due to the lower quality in some of the scans, it seemed difficult to precisely isolate the borders between the different OCT classes on a pixel-wise level in this OCT region. Quality issues in OCT-scans can have a variety of reasons and are a very common phenomena as already described by Domalpally et al<sup>47</sup>. For the purpose of this segmentation algorithm, it was crucial to be able to handle difficult images (quality-wise) as well, so we did not exclude scans with a lower image quality in the training sets.

Future work should aim to expand the number of features, such as adding Serous PED, segmented by the algorithm by collecting and annotating more data. Further, to deal with the difficulty of hard to annotate examples and annotator disagreement, future data sets should contain even more annotators and retinal experts annotating all the images in order to model the feature uncertainty.

**Distribution of Biomarkers and their overall plasticity on treatment naive patients.** After proving the efficacy of the automatic deep segmentation algorithm, we described and

examined the quantitative volumetric change over 12 months of therapy in OCT biomarkers in a large series of treatment naive nAMD patients after anti-VEGF therapy in a real life setting down to the microscopic level. We show that the most prominent change was to be seen in the first three months suggesting the initial upload phase of three injections as being the most potent for morphological changes to the retina. Interestingly, almost no changes could be determined afterwards until month 12, revealing that any further injection past the initial monthly upload phase only functioned as a stabilizing agent. The biomarkers most sensitive to change were usually the ones that have been most prominent at the beginning: IRF, SRF, SHRM, RPE and fibrovascular PED alongside CRT.

We showed that of those markers only intraretinal fluid and fibrovascular PED were predictive of a generally worse visual acuity outcome after 12 months. Although this seems logical, as higher amounts of those markers would indicate higher disease activity, we couldn't find any prior evidence that would necessarily link the actual volumetric amount of fluid to a different outcome in contrast to fibrovascular PED for which there was prior evidence for example as described by Boyer et al<sup>48</sup>. One reason would be that patients with an abundance of these markers might have generally presented later after the first onset of symptoms and thus natural progression of the disease would have led to more irreversible retinal damages. Delay to treatment is a significant factor for poorer outcomes as previously described by Lim et al<sup>49</sup>. Other than these, fibrosis significantly contributes to a worse outcome (which is again a marker of a progressed disease state) and interestingly epiretinal membrane. This is in direct contradiction to earlier findings by Alkin et al, where they concluded comparable results between patients with ERM and without ERM that were treated with bevacizumab<sup>50</sup>. However, reasons for our different findings could be a much higher number in investigated patients and the actual volumetric quantification of the observed epiretinal membranes in our oct scans. An increase of ERM did not necessarily mean more traction but it seems as simply a volumetric increase of this feature was enough to be predictive of a poorer outcome. It remains unclear why this would be the case. A possible explanation, at least in part, could be an increased traction over time (although in this work not possible to quantify) and thus more damage to the rpe and photoreceptor cells. However a missing increase in intraretinal fluid would contradict this. In additional contrast to Alkin et al. we did not solely include patients treated with bevacizumab, but rather conglomerated the different anti-VEGF injections altogether. Although highly unlikely, this might be part of the reason for different findings in our work.

Similar descriptive studies have been done in the past but with a much smaller case series, manual segmentation and only some of the biomarkers segmented in this study which in general led to various findings<sup>46,48,51,52</sup>.

Lai et al observed treatment response for intraretinal cysts (IRC), subretinal fluid, pigment epithelial detachment and their correlation with BCVA changes for a time period of a year whilst setting similar time points as in our study (month 1, 3, 6 and 12)<sup>51</sup>. In 126 eyes only 33.3% showed a resolution of their PED whereas IRC resolved in 53.8% and SRF in 51.6% of cases<sup>51</sup>. In correlation with that we saw a statistically significant and prominent volumetric reduction in IRF and SRF, whilst the decrease in Fibrovascular PED was barely significant. This supports the findings of an actual lower resolution of PEDs that were observed in the aforementioned study. This could mean that while anti-VEGF may be sufficient in controlling the PED, it lacks strong efficacy in decreasing its size or making it disappear. Other observations by Golbaz et al demonstrated that intraretinal fluid and subretinal fluid immediately responded to anti-VEGF treatment albeit the overall plasticity of the morphological changes declined over time which surely corroborates with our findings. Interestingly the sub-RPE compartments showed the least or no changes to anti VEGF therapy<sup>52</sup>. The high immediate effect of anti-VEGF on especially IRF or SRF is all in all well documented in several other studies in the literature. Bolz et al for example showed a significant effect of anti-VEGF on retinal fluid compartments as early as one week after the injection whilst there was no significant change after the second and third injection<sup>53</sup>. While this might be suggestive of cutting the loading dose to just one injection, the small number of

cases (n=29) and missing long term observation of relapsing cases in that study warrants caution.

Our study is mainly limited by its retrospective nature. A true randomized, double blind prospective study would give us much more reliable information to what effect anti-VEGF actually influences all the OCT biomarkers but it would not be feasible to include controls groups as it warrants many ethical concerns. However, given the high number of real treatment naive patients in this study, the findings can more or less be very suggestive on the effect of anti-VEGF treatment. Moreover, it gives us a general idea on a broad population of patients, and what morphological properties they present upon first presentation. Other limitations obviously include the discussed feature ambiguities and the lower F1 scores for some of the biomarkers. While this could lead to some changes being misinterpreted, the F1 numbers are still quite high in general. In some cases they can be artificially lower as for certain features (especially membranes) it is difficult for manual annotators to segment correctly down to a pixel level with the precision that an automatic segmentation algorithm would. These micro-differences negatively impact F1 scores, albeit the algorithm being able to handle real segmentation perfectly.

Another limitation is that a steady decrease of RPE is not a real marker of “atrophy” or “geographic atrophy” and just gives us the idea or a hint towards such a feature.

An adequate definition of atrophy in OCT volume scans is provided by Sadda et al where 3 criteria have to be met<sup>54</sup>: choroidal hyper transmission, attenuation of the RPE band, collapse or thinning of the outer retinal layer. A new training set for atrophy would be required that includes all three previously described properties to safely detect it as a biomarker for age-related macular degeneration.

To conclude, our study shows the feasibility of a high-performing segmentation algorithm, trained with a fixed set of 458 macular OCT scans, to quantitatively segment the whole region of a much larger number of OCT volume scans and thereby enable precise determination of volumetric changes down to a microscopic level. With that we were able to present a general description of the volumetric distribution of biomarkers in treatment naive patients and confirm that the most significant plasticity in biomarkers happens during the first 3 months of therapy with changes coming to a halt afterwards. For future work, this algorithm and findings could be used to analyze larger time series data sets and possibly predict the best possible therapy for each individual patient leading the way to a more personalized approach in treatment regimens.

## **References**

1. Friedman DS, O'Colmain BJ, Muñoz B, et al. Prevalence of age-related macular degeneration in the United States. *Arch Ophthalmol* 2004;122:564–572.
2. Yau JWY, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012;35:556–564.
3. Resnikoff S, Pascolini D, Etya'ale D, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ* 2004;82:844–851.
4. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis.

Lancet Glob Health 2014;2:e106–16.

5. Rosenfeld PJ, Brown DM, Heier JS, et al. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med* 2006;355:1419–1431.

6. Heier JS, Brown DM, Chong V, et al. Intravitreal aflibercept (VEGF trap-eye) in wet age-related macular degeneration. *Ophthalmology* 2012;119:2537–2548.

7. Rofagha S, Bhisitkul RB, Boyer DS, et al. Seven-year outcomes in ranibizumab-treated patients in ANCHOR, MARINA, and HORIZON: a multicenter cohort study (SEVEN-UP). *Ophthalmology* 2013;120:2292–2299.

8. Daniel E, Toth CA, Grunwald JE, et al. Risk of scar in the comparison of age-related macular degeneration treatments trials. *Ophthalmology* 2014;121:656–666.

9. Ishikawa K, Kannan R, Hinton DR. Molecular mechanisms of subretinal fibrosis in age-related macular degeneration. *Exp Eye Res* 2016;142:19–25.

10. Grossniklaus HE, Green WR. Histopathologic and ultrastructural findings of surgically excised choroidal neovascularization. Submacular Surgery Trials Research Group. *Arch Ophthalmol* 1998;116:745–749.

11. Chopra R, Preston GC, Keenan TDL, et al. Intravitreal injections: past trends and future projections within a UK tertiary hospital. *Eye* 2021. Available at: <http://dx.doi.org/10.1038/s41433-021-01646-3>.

12. von der Burchard C, Treumer F, Ehlken C, et al. Retinal volume change is a reliable OCT biomarker for disease activity in neovascular AMD. *Graefes Arch Clin Exp Ophthalmol* 2018;256:1623–1629.

13. Bogunovic H, Montuoro A, Baratsits M, et al. Machine Learning of the Progression of Intermediate Age-Related Macular Degeneration Based on OCT Imaging. *Invest Ophthalmol Vis Sci* 2017;58:BIO141–BIO150.

14. Schmidt-Erfurth U, Waldstein SM, Klmscha S, et al. Prediction of Individual Disease Conversion in Early AMD Using Artificial Intelligence. *Invest Ophthalmol Vis Sci* 2018;59:3199–3208.

15. de Sisternes L, Simon N, Tibshirani R, et al. Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progression. *Invest Ophthalmol Vis Sci* 2014;55:7093–7103.

16. Guymer RH, Markey CM, McAllister IL, et al. Tolerating Subretinal Fluid in Neovascular Age-Related Macular Degeneration Treated with Ranibizumab Using a Treat-and-Extend Regimen: FLUID Study 24-Month Results. *Ophthalmology* 2019;126:723–734.

17. Christenbury JG, Phasukkijwatana N, Gilani F, et al. PROGRESSION OF MACULAR ATROPHY IN EYES WITH TYPE 1 NEOVASCULARIZATION AND AGE-RELATED MACULAR DEGENERATION RECEIVING LONG-TERM INTRAVITREAL ANTI-VASCULAR ENDOTHELIAL GROWTH FACTOR THERAPY: An Optical Coherence Tomographic Angiography Analysis. *Retina* 2018;38:1276–1288.

18. Dhrami-Gavazi E, Balaratnasingam C, Lee W, Freund KB. Type 1 neovascularization may confer resistance to geographic atrophy amongst eyes treated for neovascular age-related macular degeneration. *Int J Retina Vitreous* 2015;1:15.

19. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina.



Prog Retin Eye Res 2018;67:1–29.

20. Burlina PM, Joshi N, Pekala M, et al. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol* 2017;135:1170–1176.

21. Ting DSW, Cheung CY-L, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* 2017;318:2211–2223.

22. Fauw JD, De Fauw J, Ledsam JR, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine* 2018;24:1342–1350. Available at: <http://dx.doi.org/10.1038/s41591-018-0107-6>.

23. Holmberg OG, Köhler ND, Martins T, et al. Self-supervised retinal thickness prediction enables deep learning from unlabeled data to boost classification of diabetic retinopathy. Available at: <http://dx.doi.org/10.1101/861757>.

24. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing; 2015:234–241.

25. Cordts M, Omran M, Ramos S. The cityscapes dataset for semantic urban scene understanding. Proceedings of the 2016. Available at: [http://openaccess.thecvf.com/content\\_cvpr\\_2016/html/Cordts\\_The\\_Cityscapes\\_Dataset\\_CVPR\\_2016\\_paper.html](http://openaccess.thecvf.com/content_cvpr_2016/html/Cordts_The_Cityscapes_Dataset_CVPR_2016_paper.html).

26. Chen L-C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 2018;40:834–848.

27. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning. *Ophthalmology* 2018;125:549–558.

28. German Society of Ophthalmology (Deutsche Ophthalmologische Gesellschaft, DOG), German Retina Society (Retinologische Gesellschaft e. V. , RG), Professional Association of German Ophthalmologists (Berufsverband der Augenärzte Deutschlands e. V. , BVA). Statement of the German Ophthalmological Society (DOG), the German Retina Society (GRS), and the Professional Association of German Ophthalmologists (BVA) on anti-VEGF treatment in neovascular age-related macular degeneration. *Ophthalmologie* 2021;118:31–39.

29. Lange C, Feltgen N, Junker B, et al. Resolving the clinical acuity categories “hand motion” and “counting fingers” using the Freiburg Visual Acuity Test (FrACT). *Graefes Arch Clin Exp Ophthalmol* 2009;247:137–142.

30. Schulze-Bonsel K, Feltgen N, Burau H, et al. Visual acuities “hand motion” and “counting fingers” can be quantified with the freiburg visual acuity test. *Invest Ophthalmol Vis Sci* 2006;47:1236–1240.

31. Wada K. labelme: Image Polygonal Annotation with Python. 2016. Available at: <https://github.com/wkentaro/labelme>.

32. Spaide RF, Jaffe GJ, Sarraf D, et al. Consensus Nomenclature for Reporting Neovascular Age-Related Macular Degeneration Data: Consensus on Neovascular Age-

Related Macular Degeneration Nomenclature Study Group. *Ophthalmology* 2020;127:616–636.

33. Kinyoun J, Barton F, Fisher M, et al. Detection of diabetic macular edema. Ophthalmoscopy versus photography--Early Treatment Diabetic Retinopathy Study Report Number 5. The ETDRS Research Group. *Ophthalmology* 1989;96:746–50; discussion 750–1.

34. Zahavy T, Kang B, Sivak A, et al. Ensemble Robustness and Generalization of Stochastic Deep Learning Algorithms. *arXiv [csLG]* 2016. Available at: <http://arxiv.org/abs/1602.02389>.

35. Lakshminarayanan B, Pritzel A, Blundell C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *arXiv [statML]* 2016. Available at: <http://arxiv.org/abs/1612.01474>.

36. Anon. Geoffrey Hinton, Yoshua Bengio, and Yann LeCun: Recipients of ACM's A. M. Turing Award. 2019.

37. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Bach F, Blei D, eds. *Proceedings of the 32nd International Conference on Machine Learning*. Vol 37. Proceedings of Machine Learning Research. Lille, France: PMLR; 2015:448–456.

38. Weng L, Zhang H, Chen H, Song Z. Towards fast computation of certified robustness for relu networks. *International* 2018. Available at: [https://proceedings.mlr.press/v80/weng18a.html?utm\\_source=miragenews&utm\\_medium=miragenews&utm\\_campaign=news](https://proceedings.mlr.press/v80/weng18a.html?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news).

39. Kumar SK. On weight initialization in deep neural networks. *arXiv [csLG]* 2017. Available at: <http://arxiv.org/abs/1704.08863>.

40. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–272.

41. Hoeffding W, Robbins H. The central limit theorem for dependent random variables. *dmj* 1948;15:773–780. Available at: [Accessed July 26, 2021].

42. Guo Y, Hormel TT, Xiong H, et al. Automated Segmentation of Retinal Fluid Volumes From Structural and Angiographic Optical Coherence Tomography Using Deep Learning. *Transl Vis Sci Technol* 2020;9:54.

43. Trucco E, MacGillivray T, Xu Y. *Computational Retinal Image Analysis: Tools, Applications and Perspectives*. Elsevier; 2019.

44. Mishra Z, Ganegoda A, Selicha J, et al. Automated Retinal Layer Segmentation Using Graph-based Algorithm Incorporating Deep-learning-derived Information. *Sci Rep* 2020;10:9541.

45. Kugelman J, Alonso-Caneiro D, Read SA, et al. Automatic choroidal segmentation in OCT images using supervised deep learning methods. *Sci Rep* 2019;9:13298.

46. Ohayon A, De Rosa I, Semoun O, et al. Subretinal pigment epithelium fibrotic tissue morphological changes after a single anti-vascular endothelial growth factor injection in age-related macular degeneration. *Br J Ophthalmol* 2020;104:1085–1088.

47. Domalpally A, Danis RP, Zhang B, et al. Quality issues in interpretation of optical

coherence tomograms in macular diseases. *Retina* 2009;29:775–781.

48. Boyer DS, Antoszyk AN, Awh CC, et al. Subgroup analysis of the MARINA study of ranibizumab in neovascular age-related macular degeneration. *Ophthalmology* 2007;114:246–252.

49. Lim JH, Wickremasinghe SS, Xie J, et al. Delay to treatment and visual outcomes in patients treated with anti-vascular endothelial growth factor for age-related macular degeneration. *Am J Ophthalmol* 2012;153:678–86, 686.e1–2.

50. Alkin Z, Ozkaya A, Osmanbasoglu OA, et al. The role of epiretinal membrane on treatment of neovascular age-related macular degeneration with intravitreal bevacizumab. *ScientificWorldJournal* 2013;2013:958724.

51. Lai T-T, Hsieh Y-T, Yang C-M, et al. Biomarkers of optical coherence tomography in evaluating the treatment outcomes of neovascular age-related macular degeneration: a real-world study. *Sci Rep* 2019;9:529.

52. Golbaz I, Ahlers C, Stock G, et al. Quantification of the therapeutic response of intraretinal, subretinal, and subpigment epithelial compartments in exudative AMD during anti-VEGF therapy. *Invest Ophthalmol Vis Sci* 2011;52:1599–1605.

53. Bolz M, Simader C, Ritter M, et al. Morphological and functional analysis of the loading regimen with intravitreal ranibizumab in neovascular age-related macular degeneration. *Br J Ophthalmol* 2010;94:185–189.

54. Sadda SR, Abdelfattah NS, Lei J, et al. Spectral-Domain OCT Analysis of Risk Factors for Macular Atrophy Development in the HARBOR Study for Neovascular Age-Related Macular Degeneration. *Ophthalmology* 2020;127:1360–1370.