




ARTICLE OPEN



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

Ben Asani^{1,5}, Olle Holmberg^{2,5}, Johannes B. Schiefelbein^{1,5}, Michael Hafner¹, Tina Herold¹, Hannah Spitzer², Jakob Siedlecki¹, Christoph Kern¹, Karsten U. Kortuem³, Amit Frishberg², Fabian J. Theis^{2,4,5} and Siegfried G. Priglinger^{1,5}

© The Author(s) 2024

OBJECTIVES: To determine real-life quantitative changes in OCT biomarkers in a large set of treatment naïve patients in a real-life setting undergoing anti-VEGF therapy. For this purpose, we devised a novel deep learning based semantic segmentation algorithm providing the first benchmark results for automatic segmentation of 11 OCT features including biomarkers for neovascular age-related macular degeneration (nAMD).

METHODS: Training of a Deep U-net based semantic segmentation ensemble algorithm for state-of-the-art semantic segmentation performance which was used to analyze OCT features prior to, after 3 and 12 months of anti-VEGF therapy.

RESULTS: 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 naïve 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 provokes most potent changes in the first 3 months while a gradual loss of RPE hints at a progressing decline of visual acuity. Additional research is required to understand how these accurate OCT predictions can be leveraged for a personalized therapy regimen.

Eye (2024) 38:3180–3186; <https://doi.org/10.1038/s41433-024-03264-1>

INTRODUCTION

Ocular coherence tomography (OCT) imaging of the retina and the identification of several disease biomarkers has become key in creating treatment protocols and tracking of disease progression in neovascular age-related macular degeneration (nAMD) [1]. Today, decision for treatment is mainly guided by the findings in OCT images of nAMD patients where several biomarkers have been detected to guide therapeutic decisions [2–5]. That has, in the past, led to an overall high therapy intensity with monthly to bimonthly injections over several years, posing a major challenge for both patients and the healthcare system: A recent study by Chopra et al. shows an 11-fold increase in annual intravitreal injections from 2009 to 2019 and is projected to continue to rise [6]. Newer studies, however, have shown that ignoring subretinal fluid (SRF) ($<200 \mu\text{m}$ at the foveal center) does not change outcome in visual acuity for patients but can lessen their treatment burden [7]. Additionally, a very stable fibrovascular pigment epithelial detachment (fPED) may correlate with a protection for the development of macular atrophy [8].

Consequently, to understand the prognostic value of different biomarkers it is important to assess their overall distribution in patients in a real-life setting under optimal therapeutic adherence. However, manually analyzing these and their changes over time becomes difficult as well as 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 and heterogeneity of the scan [9]. This generates high-volume data making the manual segmentation process close to impossible.

In this project, we propose a deep learning-based semantic segmentation algorithm trained with 458 manually annotated macular OCT scans, to allow automatic segmentation of clinical features of a large series of treatment-naïve patients eyes suffering from neovascular AMD and undergoing anti-VEGF (vascular endothelial growth factor) treatment.

We additionally show a validation of the algorithm on independent test sets of previously unseen patients combined with detailed analysis of the inter-annotator variance for ambiguous and hard to annotate features. It furthermore allows us to give a one-of-a-kind extensive description on the

¹Department of Ophthalmology, Ludwig-Maximilians-University, Munich, Germany. ²Institute of Computational Biology, Helmholtz Centre Munich, Munich, Germany. ³Department of Ophthalmology, Ulm University, Ulm, Germany. ⁴Department of Mathematics, TU Munich, Munich, Germany. ⁵These authors contributed equally: Ben Asani, Olle Holmberg, Johannes B. Schiefelbein, Fabian J. Theis, Siegfried G. Priglinger. ✉email: ben.asani@med.uni-muenchen.de

Received: 11 August 2023 Revised: 28 June 2024 Accepted: 16 July 2024

Published online: 27 July 2024

distribution of morphologic features as well as disease biomarkers of that patient group found in a real-life setting.

METHODS

This case series included patients from our clinic with treatment-naïve nAMD in the study eye and had a follow-up observational period of at least three and twelve months. Treatment naïve was defined as never having had any form of intravitreal injection, with the first injection 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. 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 anti-VEGF (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 [10].

Patient identification

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 (BCVA) using standard Snellen Chart, intraocular pressure using non-contact tonometry, dilated indirect funduscopy 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 et al. and Schulze-Bonsel et al. [11, 12]. 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 fellows in medical retina (B.A., J.B.S., and M.H.) using the annotation tool LabelMe [13]. 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 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 [14]. 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), SRF, Subretinal hyperreflective material (SHRM), Drusenoid pigment epithelial detachment (dPED), fPED, Fibrosis, Choroid, Posterior hyaloid membrane (PHM). For annotation examples as well as exact class distribution statistics see supplement. The central retinal thickness C0 was defined according to the ETDRS grid [15].

An additional data set of 30 scans were 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. Afterwards, a consensus annotation was produced by a panel including the retinal experts producing in total four annotations for each of the 30 scans.

Model architecture

The algorithm used for segmenting the retinal OCT scans is a deep convolutional neural network [16] of a U-net type architecture [17].

Specifically, the network consists of eleven convolution layers followed by batch normalization [18] and relu activation functions [19]. 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 [20] 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. After filtering and max pooling operations 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 hyperparameter tuning on a validation set. 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 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 18,522 OCT scans from 378 eyes enabling the statistical analysis of morphological OCT features including nAMD biomarker distribution on treatment naïve 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 [21]. Normality of data was assumed due to sufficiently large sample sizes well above 30 as specified by the central limit theorem [22]. 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 clinical features were segmented with a top performance of 0.98 F1-score for features NR and SRF. The lowest F1-scores were observed for features ERM, dPED, and SHRM (see Fig. 1a). Test set examples of the segmented features can be seen in Fig. 1b, showing the variety of features accurately segmented in unseen

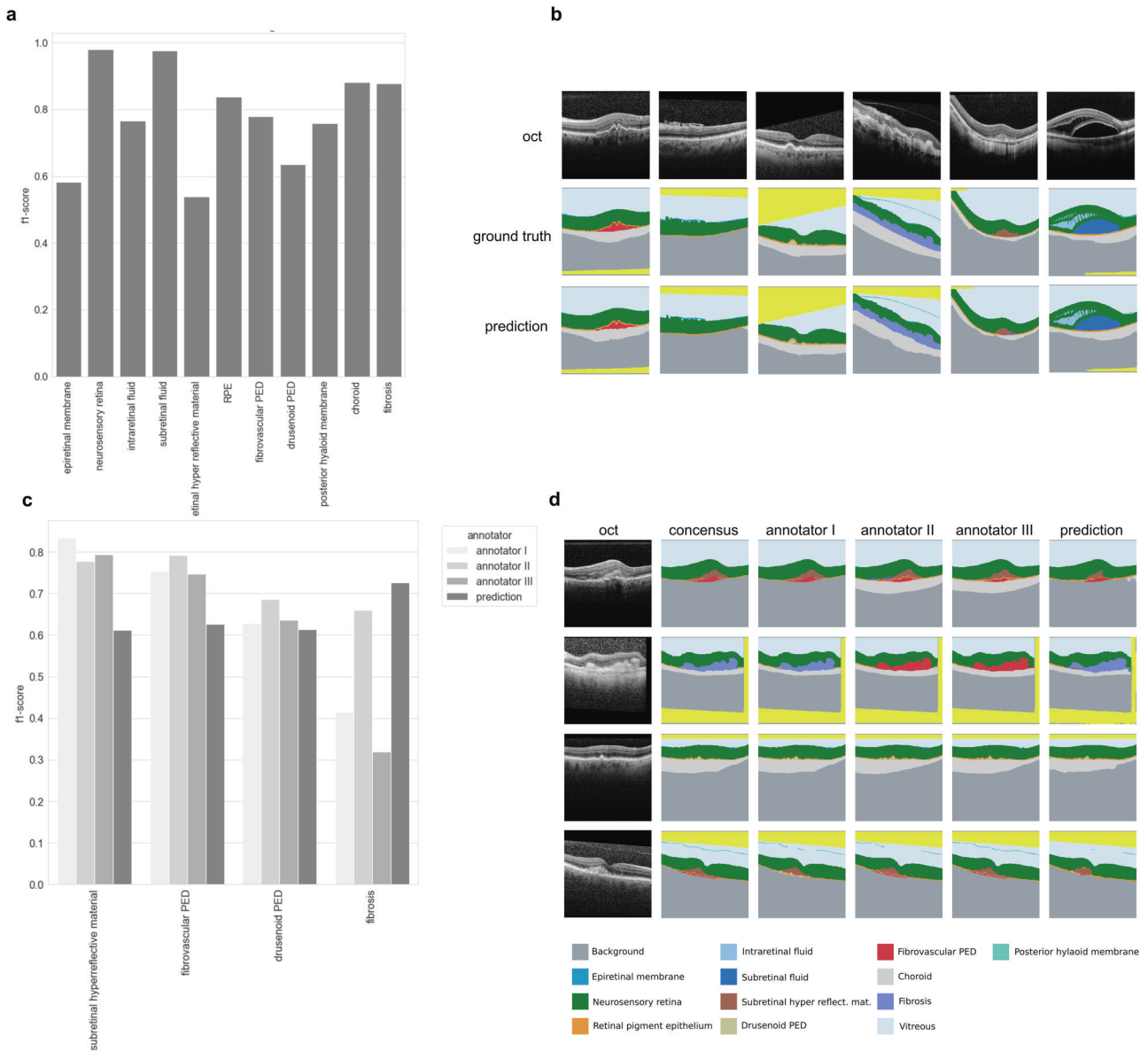


Fig. 1 Deep learning segments clinical features on par with human experts from retinal OCT images. a F1 scores for 11 clinical features segmented on a test set from 37 to the algorithm previously unseen patients. **b** Example of OCT images selected to illustrate various segmentation classes with ground 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. Yellowish color was not evaluated and either stands for a crop of the image or vitreous cavity.

patients. The annotators, on the other end, highly agreed on SHRM, fPED, and dPED, while largely disagreeing in the case of fibrosis (see Fig. 1c). Overall, the segmentation algorithm performed worse than the annotators on SHRM and fPED, on par with respect to dPED and better than all annotators when segmenting fibrosis (see Fig. 1c). In Fig. 1d we see examples of consensus, annotator and segmentation algorithm predicted segmentations for example OCT images.

Most morphological changes of treatment naive patients occur during the first three months of anti-VEGF therapy

After filtering for the inclusion criteria a total of 378 eyes consisting of 18,522 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 ± 8 years. Features in treatment-naive patients that were most prominent on initial

presentation were by far fPED (Mean: 0.12 μm³, SD: 0.19 μm³), followed by SRF (Mean: 0.08 μm³, SD: 0.26 μm³), IRF (Mean 0.03 μm³, SD 0.08 μm³) and SHRM (Mean: 0.02 μm³, SD: 0.05 μm³, see Fig. 2). Mean number of injections was 3.8 ± 1.5 after 3 months and 8.3 ± 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 fPED from 0.12 μm down to 0.09 μm (SD: 0.15 μm, *p*-value = 0.02) but no difference afterwards. 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 fPED but seems to halt further growth.

Difference in means as tested by t-tests showed a statistical significant change in IRF, SRF, SRHM, and CRT C0 after 3 months

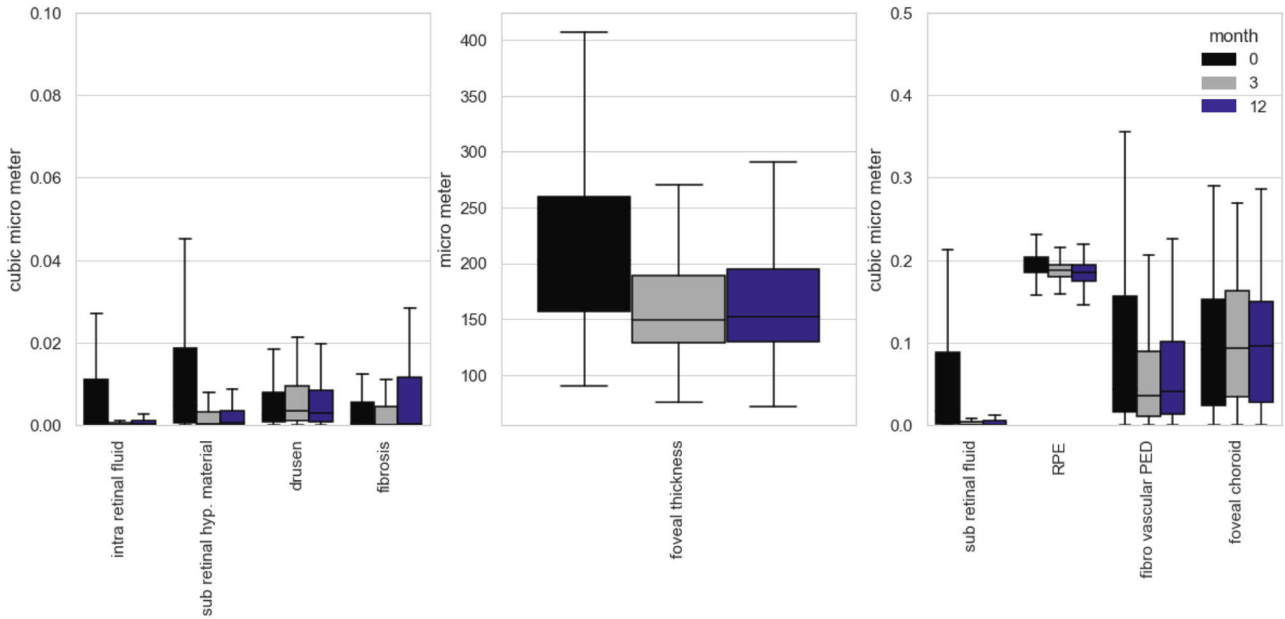


Fig. 2 Volumetric changes of treatment naive OCT biomarkers under anti-VEGF treatment. Significant reduction was seen in IRF, SRF, SHRM, RPE and central foveal thickness.

Table 1. Mean volumetric values before (mean 1), three months (mean 3) and twelve months after initiation of anti-VEGF therapy.

	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
<i>p</i> -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*
<i>p</i> -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*
<i>p</i> -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

**p* < 0.001.
All values are in μm^3 .

of anti-VEGF therapy but no significant improvement or change until month 12 occurred, meaning setbacks under therapy were uncommon. 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 Fig. 2 and Table 1. The RPE showed a significant decline throughout the complete observation period therefore hinting at a steady increase in atrophy.

We additionally used linear regression to model the inter-dependence between the different biomarkers and visual acuity outcome in 12 months. There was a negative correlation between visual acuity and intraretinal fluid, ERM, fPED, and fibrosis. Since most of these features are most prominent in the active disease state (excluding ERM) and bring immense structural changes to the retina, this finding does not seem very surprising. Interestingly, SRF 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.

DISCUSSION

Structural changes in OCT images of nAMD patients have been extensively studied in the past [5, 23] however, not to the extent that is made possible by a high-performing segmentation algorithm. In this work, we were able to realistically describe not only the morphologic characteristics of the patients but also actually quantify those volumetrically including the disease biomarkers in their respective scans. This was leveraged to give a one-of-a-kind description of a large number of treatment-naive nAMD patients in a real-life setting.

Contribution to the application of deep learning algorithms in understanding the retina

Semantic segmentation of OCT images has been studied before [24–27]. A similar segmentation algorithm was developed by De Fauw and colleagues [28] which also uses a 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 secondly, 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. This entails evaluating uncertainties surrounding feature

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

Model predictors	<i>b</i>	SE <i>b</i>	<i>t</i>	<i>p</i>
Intercept	0.64*	0.23	2.82	0.01
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
fvped	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^2	0.189			
<i>F</i>	8.08***			

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

ambiguities by examining how various retinal experts interpret different image features through annotation comparisons and then testing the algorithm's performance against these experts to assess its accuracy. 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 fellows 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 SHRM (see Fig. 2).

These difficulties specifically present in transitions from fPED 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 [29]. 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.

Another particular issue with algorithms trained on a specific training set is the transferability to different devices with possibly other axial resolutions. However, as shown in a previous study, we demonstrated that this is indeed possible [30]. SVDNA, a minimal method for unsupervised domain adaptation, shows the transferability of this algorithm to other devices such as Dri-OCT (Topcon, Tokyo, Japan), Cirrus OCT (Carl Zeiss AG, Oberkochen, Germany), and Bioptigen-OCT (Leica, Wetzlar, Germany). We show that it performs at least en par or even outperforms the more complex state-of-the-art UDA methods while considerably reducing training complexity [30].

Distribution of biomarkers and their overall plasticity on treatment-naive patients

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 seems to overwhelm injection function 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 fPED.

We showed that of those markers only IRF and fPED were predictive of a generally worse visual acuity outcome after 12 months. Although it seems logical that higher amounts 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 fPED for which there was prior evidence for example as described by Boyer et al. [31]. 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 as delay to treatment is a significant factor for a poorer outcome [32]. Other than these, fibrosis and ERM correlated negatively with visual acuity which is in direct contradiction to earlier findings by Alkin et al., where they concluded comparable results (with and without ERM) after treatment with bevacizumab [33].

Our findings also clearly suggest a gradual loss of RPE throughout 12 months. This might suggest a relationship between intravitreal anti-VEGF and retinal atrophy as it has been proposed in the past. Direct proof is still missing [34] however, and the loss was happening unaffected by change in injection intensity and favors the thesis of a natural disease progression.

Similar descriptive studies have been done in the past but with a much smaller case series, manual segmentation and less biomarkers which in general led to various findings [29, 31, 35, 36].

Lai et al. observed treatment response for intraretinal cysts (IRC), SRF, PED, and their correlation with BCVA changes for a time period of a year whilst setting similar time points as in our study (months 1, 3, 6, and 12) [35]. 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 [35]. In correlation with that we saw a statistically significant and prominent volumetric reduction in IRF and SRF, whilst the decrease in fPED was not significant over a period of 12 months. Other observations by Golbaz et al. demonstrated that IRF and SRF immediately responded to anti-VEGF treatment albeit the overall plasticity of the morphological changes declined over time which corroborates with our findings. Interestingly the sub-RPE compartments showed the least or no changes to anti VEGF therapy [36]. 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 [37]. 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 [38]. 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.

None of the aforementioned studies documented any changes to the RPE layer nor did we find any other studies observing a steady loss in RPE with nAMD patients under anti-VEGF therapy, at least not in a large cohort as in the treatment-naive dataset presented here.

Our study is mainly limited by its retrospective nature. 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 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 of these features smoothly.

To conclude, our study shows the feasibility of a high-performing segmentation algorithm 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-naïve 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 besides a gradual loss of the RPE layer which seems to continue at least until the end of the observed time frame. For future work, this algorithm and findings could be used to analyze larger time series data sets and possibly predict the best-fitting therapy for each individual patient leading the way to a more personalized approach in treatment regimens.

SUMMARY

What was known before

- Limitations in Algorithmic Research Quality and Biomarker Inclusion: Existing segmentation algorithms to date present substantial quality ambiguities and lack consistent up-to-date biomarkers. Furthermore, a validation on independent test sets seems to be absent, which represents a notable gap in the current state of OCT segmentation research, raising questions about their reliability and generalizability.
- Shortcomings in characterization of treatment naïve AMD While existing literature, encompassing trials like MARINA, VIEW, or HORIZON, provide a comprehensive understanding of age-related macular degeneration (AMD) as a whole, the characterization of treatment-naïve AMD patients, including detailed volumetric analysis of retinal morphology and disease biomarkers, remains absent. Studies focusing on treatment-naïve patients typically exhibited limitations such as small sample sizes, a constrained array of biomarkers, and incomplete volumetric data, which hindered the ability to gain a holistic understanding of this patient group.

What this study adds

- Improving existing segmentation algorithms for big data analysis: This study improves segmentation algorithms in a safe and high quality manner as it is tested on independent test sets which then can be leveraged for big data analysis to define risk factors in the process. Furthermore, it follows the Consensus Nomenclature for Reporting Neovascular Age-Related Macular Degeneration of the AAO for disease biomarkers.
- Extensive descriptive analysis of a large amount of treatment-naïve nAMD patients: The analysis of a large cohort of treatment-naïve patients in a real-world setting as well as analysis of changes in the various OCT biomarkers over time and under therapy gives us a more precise anatomical insight to the disease while defining risk factors in the process. This can be leveraged to move towards a more personalized therapeutic regimen in the future

CODE AVAILABILITY

The code for the models and training procedures as well as result analysis will be made available through the public Github repository upon publication. Additional information and other data related to this study is available upon request to the corresponding author. Sensitive patient data is carefully and thoroughly anonymized.

REFERENCES

1. Keane PA, Patel PJ, Liakopoulos S, Heussen FM, Sadda SR, Tufail A. Evaluation of age-related macular degeneration with optical coherence tomography. *Surv Ophthalmol* 2012;57:389–414.
2. von der Burchard C, Treumer F, Ehlken C, Koinzer S, Purtskhvanidze K, Tode J, et al. Retinal volume change is a reliable OCT biomarker for disease activity in neovascular AMD. *Graefes Arch Clin Exp Ophthalmol* 2018;256:1623–9.
3. Bogunovic H, Montuoro A, Baratsits M, Karantonis MG, Waldstein SM, Schlanitz F, et al. Machine learning of the progression of intermediate age-related macular degeneration based on OCT imaging. *Invest Ophthalmol Vis Sci* 2017;58:10141–10150.
4. Schmidt-Erfurth U, Waldstein SM, Klimscha S, Sadeghipour A, Hu X, Gerendas BS, et al. Prediction of individual disease conversion in early AMD using artificial intelligence. *Invest Ophthalmol Vis Sci* 2018;59:3199–208.
5. de Sistiernes L, Simon N, Tibshirani R, Leng T, Rubin DL. Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progression. *Invest Ophthalmol Vis Sci* 2014;55:7093–103.
6. Chopra R, Preston GC, Keenan TDL, Mulholland P, Patel PJ, Balaskas K, et al. Intravitreal injections: past trends and future projections within a UK tertiary hospital. *Eye*. 2021. <https://doi.org/10.1038/s41433-021-01646-3>.
7. Guymer RH, Markey CM, McAllister IL, Gillies MC, Hunyor AP, Arnold JJ, 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;126:723–34.
8. 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 Retin Vitreous*. 2015;1:15.
9. Joeres S, Tsong JW, Updike PG, Collins AT, Dustin L, Walsh AC, et al. Reproducibility of quantitative optical coherence tomography subanalysis in neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci* 2007;48:4300–7.
10. 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.
11. Lange C, Feltgen N, Junker B, Schulze-Bonsel K, Bach M. Resolving the clinical acuity categories ‘hand motion’ and ‘counting fingers’ using the Freiburg Visual Acuity Test (FrACT). *Graefes Arch Clin Exp Ophthalmol* 2009;247:137–42.
12. Schulze-Bonsel K, Feltgen N, Burau H, Hansen L, Bach M. Visual acuities ‘hand motion’ and ‘counting fingers’ can be quantified with the Freiburg visual acuity test. *Invest Ophthalmol Vis Sci* 2006;47:1236–40.
13. Wada K. labelme: image polygonal annotation with Python. 2016. <https://github.com/wkentaro/labelme>.
14. Spaide RF, Jaffe GJ, Sarraf D, Freund KB, Sadda SR, Staurengli G, et al. Consensus nomenclature for reporting neovascular age-related macular degeneration data: consensus on neovascular age-related macular degeneration nomenclature study group. *Ophthalmology* 2020;127:616–36.
15. Kinyoun J, Barton F, Fisher M, Hubbard L, Aiello L, Ferris F 3rd. Detection of diabetic macular edema. *Ophthalmology* versus photography-Early Treatment Diabetic Retinopathy Study Report Number 5. The ETDRS Research Group. *Ophthalmology*. 1989;96:746–50.
16. Anon. Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award. Available at: <https://www.acm.org/media-center/2019/march/turing-award-2018> [Accessed July 22, 2023]
17. 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. p. 234–41
18. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Bach F, Blei D, editors. *Proceedings of the 32nd International Conference on Machine Learning*. Proceedings of Machine Learning Research. Lille, France: PMLR; 2015. p. 448–56
19. Weng L, Zhang H, Chen H, Song Z. Towards fast computation of certified robustness for relu networks. *International*. 2018. https://proceedings.mlr.press/v80/weng18a.html?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news.

20. Kumar SK. On weight initialization in deep neural networks. arXiv [cs.LG]. 2017. Available at: <http://arxiv.org/abs/1704.08863>.
21. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–72.
22. Hoeffding W, Robbins H. The central limit theorem for dependent random variables. *DMJ* 1948;15:773–80.
23. Metrangolo C, Donati S, Mazzola M, Fontanel L, Messina W, D'alterio G, et al. OCT biomarkers in neovascular age-related macular degeneration: a narrative review. *J Ophthalmol*. 2021;2021:9994098.
24. Guo Y, Hormel TT, Xiong H, Wang J, Hwang TS, Jia Y. Automated segmentation of retinal fluid volumes from structural and angiographic optical coherence tomography using deep learning. *Transl Vis Sci Technol* 2020;9:54.
25. Trucco E, MacGillivray T, Xu Y. Computational retinal image analysis: tools, applications and perspectives. Elsevier; 2019
26. Mishra Z, Ganegoda A, Selicha J, Wang Z, Sadda SR, Hu Z. Automated retinal layer segmentation using graph-based algorithm incorporating deep-learning-derived information. *Sci Rep*. 2020;10:9541.
27. Kugelmann J, Alonso-Caneiro D, Read SA, Hamwood J, Vincent SJ, Chen FK, et al. Automatic choroidal segmentation in OCT images using supervised deep learning methods. *Sci Rep*. 2019;9:13298.
28. Fauw JD, De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–50. <https://doi.org/10.1038/s41591-018-0107-6>.
29. Ohayon A, De Rosa I, Semoun O, Jung C, Colantuono D, El Ameen A'a, 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;104:1085–8.
30. Koch V, Holmberg O, Spitzer H, Schiefelbein J, Asani B, Hafner M, et al. Noise transfer for unsupervised domain adaptation of retinal OCT images. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*. Switzerland: Springer Nature; 2022. p. 699–708
31. Boyer DS, Antoszyk AN, Awh CC, Bhisitkul RB, Shapiro H, Acharya NR, et al. Subgroup analysis of the MARINA study of ranibizumab in neovascular age-related macular degeneration. *Ophthalmology* 2007;114:246–52.
32. Lim JH, Wickremasinghe SS, Xie J, Chauhan DS, Baird PN, Robman LD, 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.e1–2.
33. Alkin Z, Ozkaya A, Osmanbasoglu OA, Agca A, Karakucuk Y, Yazici AT, et al. The role of epiretinal membrane on treatment of neovascular age-related macular degeneration with intravitreal bevacizumab. *Sci World J* 2013;2013:958724.
34. Gemenetzi M, Lotery AJ, Patel PJ. Risk of geographic atrophy in age-related macular degeneration patients treated with intravitreal anti-VEGF agents. *Eye*. 2017;31:1–9.
35. Lai T-T, Hsieh Y-T, Yang C-M, Ho T-C, Yang C-H. 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.
36. Golbaz I, Ahlers C, Stock G, Schütze C, Schriefel S, Schlanitz F, 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;52:1599–605.
37. Bitzer M. Altersbedingte Makuladegeneration – Makulamorphologie und Visusverbesserung nach VEGF-Therapie. *Klin Monbl Augenheilkd* 2014;231:98–99.
38. Bolz M, Simader C, Ritter M, Ahlers C, Benesch T, Prünte C, et al. Morphological and functional analysis of the loading regimen with intravitreal ranibizumab in neovascular age-related macular degeneration. *Br J Ophthalmol* 2010;94:185–9.

AUTHOR CONTRIBUTIONS

BA: acquisition of data, manual segmentation of oct images, drafting of manuscript, review, and synthesis of literature, analysis and interpretation of data. OH: training of deep segmentation algorithm, drafting of manuscript, review, and synthesis of literature, analysis and interpretation of data. JBS: acquisition of data, manual segmentation of oct images, drafting of manuscript, review, and synthesis of literature, analysis and interpretation of data. MH: acquisition of data, manual segmentation of oct images. JS and TRH: manual segmentation of oct images, critical review, and revision of the manuscript. HS, CK, and KK: critical review and revision of the manuscript, analysis, and interpretation of data. SGP and FJT: conceptualization

of the manuscript, critical review and revision of the manuscript, conception, and design of the work/project, analysis, and interpretation of data. All authors contributed to the article and approved the submitted version.

FUNDING

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]. The sponsor or funding organization had no role in the design or conduct of this research. Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The following authors declare no Competing Non-Financial Interests but the following Competing Financial interests: Ben Asani receives speaker fees and study support from Novartis Pharma GmbH. Olle Holmberg has ownership in Helsing GmbH. Johannes Schiefelbein receives study support from Novartis Pharma GmbH. Tina Herold receives speaker honoraria and study support from Novartis Pharma GmbH and Bayer AG, speaker honoraria from Abbvie, and study support from Roche Holding AG. Christoph Kern has received travel fees from Bayer AG and Novartis Pharma GmbH. Consultant: Bayer AG. Karsten Kortuem has received speaker fees from Novartis Pharma 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. 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; has received research funding from Abbott, Alcon, Hoya, Oculentis, Oculus, Schwind and Zeiss. Fabian Theis consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, and Omniscope Ltd, and has ownership interest in Dermagnostix GmbH and Cellarity. Hannah Spitzer and Michael Hafner declare no Competing Financial or Non-Financial Interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethics committee of the medical faculty of the Ludwig-Maximilians-University of Munich gave ethical approval for this work.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41433-024-03264-1>.

Correspondence and requests for materials should be addressed to Ben Asani.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024