

OPTICAL COHERENCE TOMOGRAPHY BIOMARKERS IN WET AGE-RELATED MACULAR DEGENERATION

PhD Thesis

Attila Kovács MD



Szeged

2021

OPTICAL COHERENCE TOMOGRAPHY BIOMARKERS IN WET AGE-RELATED MACULAR DEGENERATION

PhD Thesis

Attila Kovács MD

Supervised by: Rózsa Dégi MD, PhD

University of Szeged, Faculty of Medicine
Clinical Medical Sciences Doctoral School

PhD Program:

Clinical and Experimental Research for Reconstructive and Organ-sparing Surgery
Program Director: Prof. György Lázár MD, PhD

Szeged

2021

PUBLICATIONS DIRECTLY RELATED TO SUBJECT OF THE THESIS

- I. **Kovacs A, Kiss T, Rarosi F, Somfai GM, Facsko A, Degi R. The effect of ranibizumab and aflibercept treatment on the prevalence of outer retinal tubulation and its influence on retreatment in neovascular age-related macular degeneration. *BMC Ophthalmol.* 2018 Nov 14;18(1):298.**
IF (2018): 1.431
- II. Varga L, **Kovacs A**, Grosz T, Thury G, Hadarits F, Degi R, Dombi J. **Automatic segmentation of hyperreflective foci in OCT images.** *Comput Methods Programs Biomed.* 2019 Sep; 178:91-103.
IF (2019): 3.632
- III. Katona M, **Kovacs A**, Degi R, Nyul LG. **Automatic Detection of Subretinal Fluid and Cyst in Retinal Images.** In: Battiato S., Gallo G., Schettini R., Stanco F. (eds) *Image Analysis and Processing - ICIAP 2017.* Lecture Notes in Computer Science, vol 10484. pp.: 606-616, Springer, Cham.
- IV. Katona M, **Kovacs A**, Varga L, Grosz T, Dombi J, Degi R, Nyul LG. **Automatic Detection and Characterization of Biomarkers in OCT Images.** In: Campilho A., Karray F., ter Haar Romeny B. (eds) *Image Analysis and Recognition. ICIAR 2018.* Lecture Notes in Computer Science, vol 10882. pp.: 706-714, Springer, Cham.
- V. Katona M, **Kovacs A**, Degi R, Nyul LG. **Segmentation of Subretinal Hyperreflective Material and Pigment Epithelial Detachment Using Kernel Graph Cut.** In: Burduk R., Kurzynski M., Wozniak M. (eds) *Progress in Computer Recognition Systems. CORES 2019.* Advances in Intelligent Systems and Computing, vol 977. pp.: 98-106, Springer, Cham.

TABLE OF CONTENTS

PUBLICATIONS DIRECTLY RELATED TO SUBJECT OF THE THESIS	2
LIST OF ABBREVIATIONS	4
INTRODUCTION.....	5
PART I: OUTER RETINAL TUBULATION	7
BACKGROUND.....	7
METHODS	8
<i>Statistical methods</i>	<i>10</i>
RESULTS	11
DISCUSSION	13
PART II: AUTOMATIC IDENTIFICATION OF OCT BIOMARKERS	16
BACKGROUND.....	16
METHODS	18
DELINEATION OF BOUNDARY LAYERS: ILM/RPE.....	21
<i>Algorithm based on vertically projected data</i>	<i>22</i>
<i>Graph-cut based approach.....</i>	<i>23</i>
<i>Evaluation and results.....</i>	<i>24</i>
MEASUREMENT OF RETINA LAYER DISTORTION (D)	25
SRF/IRF	26
<i>Evaluation and results.....</i>	<i>27</i>
PED.....	29
SHRM	30
<i>Classification of PED vs. SHRM</i>	<i>30</i>
<i>Evaluation and results.....</i>	<i>31</i>
ORT	32
<i>Detection using classical image processing operators.....</i>	<i>33</i>
<i>ORT localization with neural network – i.e., machine learning method</i>	<i>33</i>
<i>Evaluation and results.....</i>	<i>34</i>
HYPERREFLECTIVE FOCI.....	35
<i>Data preparation</i>	<i>37</i>
<i>Image preprocessing and feature generation.....</i>	<i>39</i>
<i>Training the nets</i>	<i>40</i>
<i>Evaluation.....</i>	<i>41</i>
<i>Results and discussion.....</i>	<i>41</i>
<i>Conclusions</i>	<i>44</i>
SUMMARY AND CONCLUSIONS	46
ACKNOWLEDGMENTS.....	49
REFERENCES.....	50

LIST OF ABBREVIATIONS

AI:	artificial intelligence
AMD:	age-related macular degeneration
ANN:	artificial neural networks
AUC:	area under the curve
BCVA:	best-corrected visual acuity
CNN:	convolutional neural network
D:	layer distortion
DNN:	deep neural network
DRN:	deep rectifier neural networks
ELM:	external limiting membrane
ETDRS:	early treatment diabetic retinopathy study
FCN:	fully convolutional neural networks
HF:	hyperreflective foci
ILM:	internal limiting membrane
IRF:	intraretinal fluid
LoG:	Laplacian of Gaussian
MITK:	Medical Imaging Interaction Toolkit
OCT/SD-OCT:	spectral domain optical coherence tomography
ORT:	outer retinal tubulation
PED:	pigment epithelial detachment
PRN:	pro re nata
RPE:	retinal pigment epithelium
SHRM:	subretinal hyperreflective material
SRF:	subretinal fluid
VEGF:	vascular endothelial growth factor

INTRODUCTION

The globally experienced growth of life expectancy has led to an increase in the number of age-related diseases, thus age-related macular degeneration (AMD) has become the leading cause of vision loss in the Western World, and a health problem worldwide. The large prevalence of AMD causes inestimable burden from the point of both the patients' and the health care system, therefore proper management is crucial [1-4].

AMD can be divided into two categories, namely the dry (nonexudative) form that accounts for approximately 90% of all cases and the wet (exudative, neovascular) form that accounts for the remaining, causing more severe and rapid cases of vision loss [5, 6].

Although the exact pathogenesis of the disease is not yet fully understood, vascular endothelial growth factor (VEGF) plays a central role in the development of this latter, more dangerous wet form. An imbalance unfolds between proangiogenic and antiangiogenic factors, leading to the overproduction of VEGF. VEGF induces endothelial cell proliferation followed by the formation of choroidal neovascular membrane, new vessels that leak fluid into or under the layers of the retina. This results in accumulation of subretinal or intraretinal fluid (SRF/IRF) or pigment epithelial detachment (PED), but bleeding can appear as well, bringing forth an advanced and acute vision loss [2].

The first choice of treatment for neovascular AMD is anti-VEGF intravitreal injection, which has revolutionized the treatment of wet AMD [7]. As a chronic condition, AMD requires repeated administration of anti-VEGF injections along with frequent patient follow-ups [8]. Early diagnosis and rigorous monitoring are essential in the management of AMD, and for both, utilization of optical coherence tomography (OCT) is fundamental. With the help of OCT, it became possible to detect in vivo the layers of the retina and the characteristic features of the disease, so-called OCT biomarkers, such as SRF, IRF, PED, outer retinal tubulation (ORT), hyperreflective foci (HF) or subretinal hyperreflective material (SHRM) (see, e.g., Figure 1) and their response to anti-VEGF treatment. Assessment of these markers is the key to make the adequate decision for treating, re-treating or observing a patient [9,10].

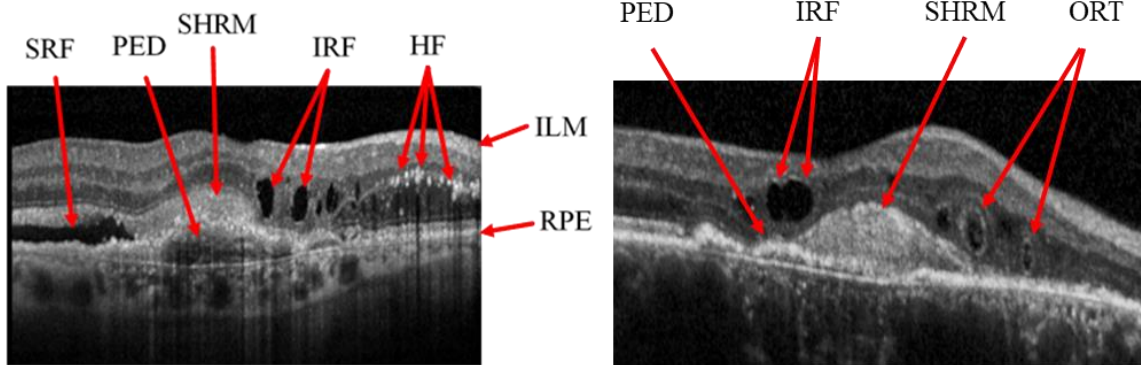


Figure 1. OCT images in wet AMD. The internal limiting membrane (ILM) and the retinal pigment epithelium (RPE) are marked together with biomarkers of AMD such as: SRF, IRF; PED; SHRM; ORT; and HF.

We have to emphasize that from this point, under the term IRF the intraretinal cystoid fluid is meant alone in this thesis.

Despite the extensive knowledge about OCT biomarkers provided by literature, there are still unanswered questions in this field. Our work basically concentrated on these biomarkers, from a clinical point of view (i.e. their response to therapy, their effect on therapy, their effect on each other). At first our attention was focused on a then relatively novel biomarker, the outer retinal tubulation, intended to investigate its characteristics and potential role in the treatment prognosis.

Our observations based on ORT (listed in the first part of the PhD thesis) spurred further thoughts, since we soon realized that in order to explore farther correlations between the OCT biomarkers and their response/effect to therapy, it is essential to use faster and more precise measurements, thereby development of automated algorithms with automatic identification and quantification of OCT biomarkers (HF in particular, as well as ORT, SRF, IRF, PED and SHRM) became our next goal (second part of the thesis).

PART I: OUTER RETINAL TUBULATION

BACKGROUND

Outer retinal tubulation (ORT) as per Zweifel et al.'s definition is a hyporeflective, branching tubular structure with hyperreflective borders within the outer nuclear layer of the retina [10, 11]. The „en face" OCT technique can help map these branching networks [12]. ORTs have been observed in many retinal diseases, including wet age-related macular degeneration [11]. Based on histological reports, the border of the outer retinal tubulation consists of photoreceptor inner segment mitochondria and external limiting membrane (ELM), with fluid and photoreceptor outer segments being potentially present in the ovoid hyporeflective lumen of the ORT [13-15]. Adaptive optics scanning laser ophthalmoscopy findings are in correlation with histology reports and show lack of ORT cone reflectivity, which can be due to the loss of cone outer segments and subsequent retinal remodeling [16].

Schaal et al. classified outer retinal tubulations as either open (incomplete closure with curving external limiting membrane at the ends, horizontally elongated shape in cross-section) or closed (completely encircled, oval shape in cross-section) ORTs [13].

Though pathogenesis of ORT formation is still not entirely clear, recent reports facilitated understanding the process. Dolz-Marco et al. called the attention on the role of Müller cells in the pathomechanism of ORT development, namely the progressive photoreceptor damage that can result in Müller cell activation, which thereby starts to produce glial fibrillary acidic protein, facilitating the formation of ORT [17]. Based on histological examinations, it seems that the evolution of ORT begins with ELM and ellipsoid zone disruption [11]. ELM starts to scroll inward at its free edges, representing an initial form of ORT, leading to the development of a formed open ORT. With time the large, open ORTs split, their margins beginning to scroll ending in multiple smaller closed ORTs. During the scrolling process a downward displacement of adjacent inner nuclear layer and outer plexiform layer happens, separating each ORT and causing the appearance of microcystic lesions in the inner nuclear layer. The downward displacement of these layers might

be due to the involvement of Müller cells in this scrolling/dragging process as Müller cells are contributing to the constitution of ELM with the inner segments of photoreceptors [13, 17-19].

ORT can be mistaken for intraretinal cysts or subretinal fluid. In contrast to ORT, detection of these fluid compartments on OCT scans is the main criterion of the retreatment process. With the recognition of the hyperreflective border of the ORT and its special occurrence in the outer nuclear layer, these mistakes can be reduced, leading to a reduction of anti-VEGF overtreatment in wet AMD [11].

The ORT prevalence in wet AMD is low at the time of first diagnosis, but over time, during anti-VEGF therapy its prevalence increases [20, 21]. The importance of ORT as an OCT biomarker for photoreceptor degeneration is due to its connection with reduced visual acuity [10, 20-22].

It has also been reported that ORTs develop above areas of subretinal hyperreflective material or atrophy [20, 21]. SHRM is a medium- to hyperreflective mass between the neurosensory retinal layers and retinal pigment epithelium on OCT [23]. It usually represents either a type II choroidal neovascular complex or is the consequence of an active choroidal neovascularization, including subretinal haemorrhage and lipid or fluid exudation [10, 23, 24].

The aim of our study was:

- to investigate the prevalence of ORTs in eyes with neovascular AMD undergoing anti-VEGF treatment either with ranibizumab or aflibercept
- to examine the changes in the frequency of injections before and after ORT appearance
- to further assess the presence of subretinal hyperreflective material and its relationship with ORT

METHODS

This retrospective study was performed at the Medical Retina Unit of the Department of Ophthalmology, University of Szeged, Hungary. The study was approved by the Institutional Review Board of University of Szeged Albert Szent-Györgyi Clinical Centre and was in accordance with the ethic standards of the Declaration of Helsinki.

Treatment-naïve wet AMD patients were enrolled in the study. For the ranibizumab group, enrollment took place between October 2014 and April 2016, while patients in the aflibercept group were enrolled between April 2015 and April 2016.

All patients were over 50 years of age, mean follow-up period was 16.3 months and 9.2 months (ranged from 6-24 months and 6-12 months) in the ranibizumab and aflibercept groups, respectively.

During each visit a comprehensive ophthalmic examination was carried out including best-corrected visual acuity (BCVA, Early Treatment Diabetic Retinopathy Study (ETDRS) score) assessment, slit-lamp biomicroscopy, dilated funduscopy and SD-OCT examination of the retina (Heidelberg Spectralis, Heidelberg Engineering, Heidelberg, Germany). Eyes with poor quality SD-OCT scans (Q index below 20) or with poor compliance were excluded from the study (14 eyes from the ranibizumab and 3 eyes from the aflibercept group).

Treatment regimen for both ranibizumab (0.5 mg) and aflibercept (2 mg) started with 3 monthly injections. After initiation phase the ranibizumab group was treated by a pro re nata (PRN) regimen with follow-up visits scheduled monthly. The retreatment criteria for ranibizumab patients consisted of any subretinal or intraretinal fluid on OCT, or new haemorrhage on funduscopy. In the aflibercept group follow-up after the loading phase was scheduled every two months, treatment was given at each follow-up. The above regimens were in accordance with the available treatment guidelines in Hungary at the time of the study.

For SD-OCT imaging a pattern size of 5.8 x 5.8 mm, 20° x 20° was applied with 25 B-scans, using the „follow-up” mode. By manual review of the scan volumes the first appearance of the ORT was determined in both groups. Presence of ORT was assessed at fixed time points at baseline, month 6 and 12 in both groups and at months 18 and 24 in the ranibizumab group. Two independent retina specialists evaluated the images. In case of incongruity, the images were referred to a third retina specialist. During the evaluation of OCT scans, no distinction was realized between the above described open (incomplete hyperreflective ring) and closed (complete hyperreflective ring) forms of ORT, according to Schaal [13]. Thus, both types of ORT detected on the images were considered an ORT positive case. The criterion of ORT was a hyperreflective ovoid-elongated structure with lower reflective content in the outer nuclear layer of the retina. (Figure 2.)

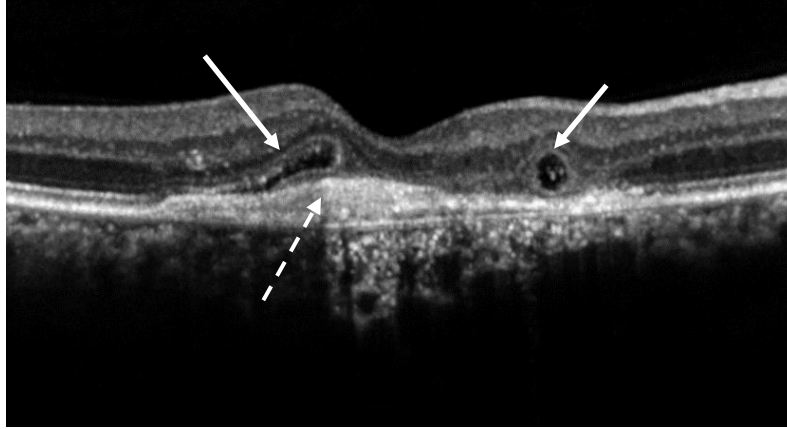


Figure 2.: Outer retinal tubulations and subretinal hyperreflective material on an SD-OCT single B-scan: Open and closed ORTs in cross section (left and right solid arrows) above subretinal hyperreflective material (dash arrow).

The graders also identified the presence or absence of subretinal hyperreflective material on SD-OCT images at treatment initiation. The criterion for SHRM was a medium- to hyperreflective mass between the neurosensory retinal layers and the retinal pigment epithelium, as described by Keane et al [23]. (Figure 2.)

Statistical methods

The BCVA was compared across the two groups using the Mann-Whitney U-test. The survival analysis for ORT development was analyzed by a Cox proportional hazard model. We analyzed the correlation between the presence of SHRM at treatment initiation with the development of ORT by Chi-square test and calculated relative risks. Where zeros were involved for the computation of relative risk, 0.5 was added to all cells, according to the previous suggestions [25].

The injection rate was calculated only in the PRN treated ranibizumab group due to the fixed 2 month therapeutic regimen of aflibercept. The injection rate was assessed before versus after the appearance of outer retinal tubulation and was compared using the Mann-Whitney U-test. In order to correct bias rising from the unequal follow-up time (some patients had a higher number of injections due to longer follow-up), monthly injections were calculated so that follow-up time was divided with the number of injections.

A p-value of $p < 0.05$ was taken as statistically significant. For the analyses the IBM SPSS Software (Version 22) was used.

RESULTS

In the ranibizumab group 184 eyes of 179 patients were evaluated, with a median age of 74 years (range 51 to 88), while in the aflibercept group there were 52 eyes of 51 patients with a median age of 75 years (range 58 to 87).

The mean baseline best corrected visual acuities in the two groups were (mean \pm SD) 59.16 \pm 13.9 (median 61) and 53.96 \pm 13.54 (median 55.5) ETDRS letters in the ranibizumab and aflibercept group, respectively. There was no significant difference between the two groups (Mann-Whitney U-test $p=0.083$). The BCVA at the end of the follow-up was 57.19 \pm 20.19 (median 63) and 59.46 \pm 15.54 (median 64) ETDRS letters in the ranibizumab and aflibercept group, respectively. There was no significant difference between the two groups (Mann-Whitney U-test $p=0.69$).

Table 1 shows the number of eyes during the follow-up in the two groups. The number of eyes was reduced over time due to gradual enrollment in the study, thereby not every patient reached the same follow-up time. In the ranibizumab group outer retinal tubulation was observed in 17.4% of cases at baseline, in 33.7% of cases at month 6, in 45.3% of cases at month 12, and in 55.3% and in 60.8% of cases at months 18 and 24, respectively. The ORT prevalence in the aflibercept group was 23.1% at baseline, 40.4% at month 6, and 50% at month 12.

Time point	ranibizumab n (eyes)	aflibercept n (eyes)
Baseline	184	52
at 6 months	184	52
at 12 months	161	28
at 18 months	103	0
at 24 months	51	0

Table 1.: Number of eyes reaching the follow-up in the ranibizumab and aflibercept treated groups The column with „n” corresponds to the number of eyes reaching the follow-up.

The survival analysis showed no significant difference between the ranibizumab and aflibercept treated groups in terms of ORT development. ($p=0.79$, hazard ratio 0.92, 95% confidence interval 0.500-1.693). (Figure 3)

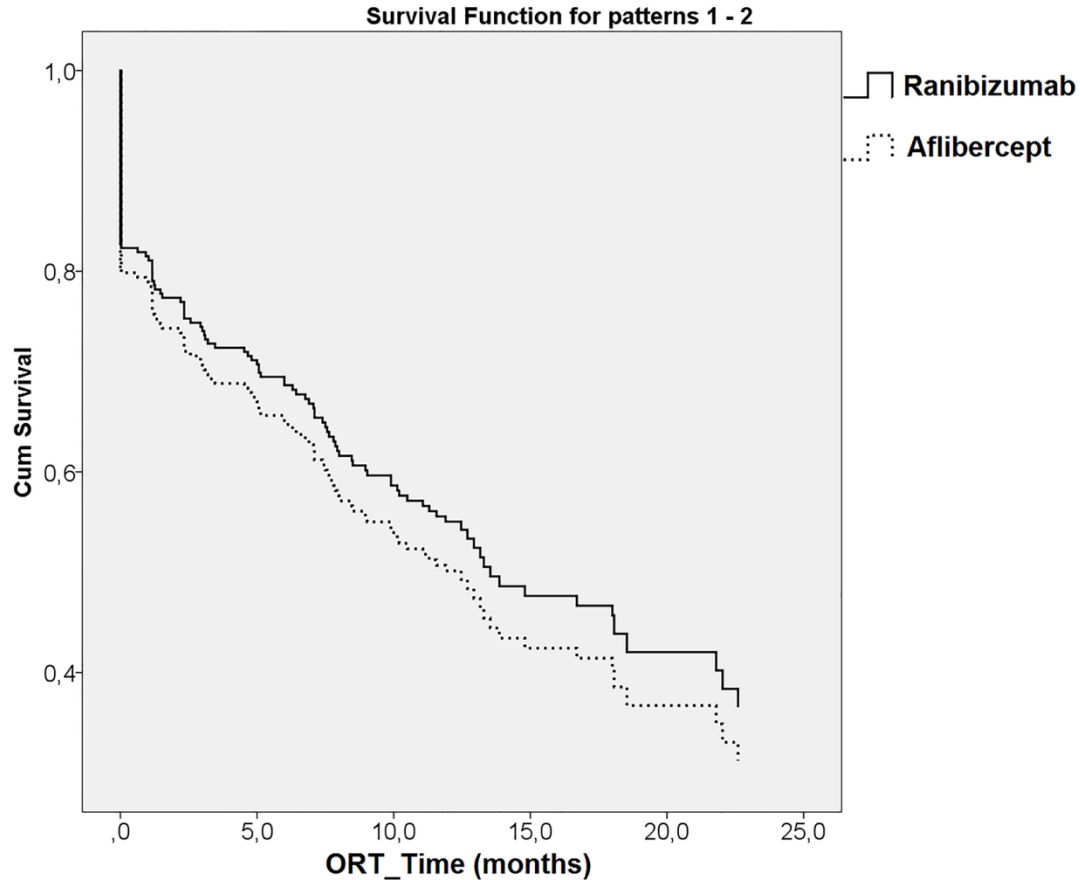


Figure 3.: Cox proportional hazard model of ORT survival in the ranibizumab and aflibercept groups

As per the injection rate, mean injection number per month before ORT appearance was 0.37 ± 0.17 , while after ORT development decreased to 0.21 ± 0.17 (Mann-Whitney U-test $p=0.004$).

The presence of subretinal hyperreflective material at treatment initiation in the two subgroups was 75.5% in the ranibizumab, and 80.8% in the aflibercept group. In the ranibizumab treated group ORT developed in 85 eyes of 139 eyes with SHRM (61.15%), while without SHRM (45 eyes) ORTs were found merely in 10 eyes (22.2%) corresponding to a relative risk of 2.75. ($p<0.01$). In the aflibercept treated group 55.81% of eyes with SHRM developed ORT (24 eyes of

43). No ORT developed in the eyes without SHRM (out of 9 eyes), consistent with a relative risk of 11.14 ($p < 0.01$).

DISCUSSION

Zweifel first described outer retinal tubulation in 2009 [11]. Later the authors reported ORTs in various retinal disorders like neovascular AMD, geographic atrophy, polypoidal choroidal vasculopathy, non-AMD associated choroidal neovascularization, and other degenerative retinal disorders (e.g. retinitis pigmentosa, gyrate atrophy, choroideremia, Stargardt disease, pattern dystrophy) [11, 12, 18, 26-30].

Most of the studies in the field focused on wet AMD and its relationship with ORT. From these reports it is known that ORT is an SD-OCT biomarker, the prevalence of which increases with time and is associated with decreased visual acuity [10, 20, 21]. It has also been reported that ORTs develop adjacent to areas of subretinal hyperreflective material or atrophy [20, 21]. The differential diagnostic importance of outer retinal tubulation lies on its correct identification in OCT scans due to its similarity to intraretinal cysts, thus avoiding the possible overtreatment of wet-AMD [11, 20].

In the present study, we evaluated treatment-naïve wet AMD patients treated with ranibizumab and aflibercept regarding the presence of outer retinal tubulation. Altogether 236 eyes were followed in both groups with no statistical difference between the baseline characteristics of the two groups considering age and BCVA.

The prevalence of ORT continuously increased during the follow-up period, in both groups. In the ranibizumab group its prevalence almost quadrupled at the 24-month follow-up, while there was a doubling in the aflibercept group in 12 months. Necessary to note, baseline prevalence was lower in the ranibizumab group (17.4% versus 23.1%). Dirani et al. found a similar increasing trend in their study starting with 2.5% at presentation, reaching 41.6% at 4 years of follow-up [20]. In our study the baseline ORT prevalence was higher compared to other reports [20, 21]. The reason behind this could be the more advanced disease state at the time of presentation (due to the real-life nature of our retrospective study) and a relative delay in therapy initiation due to country-

specific financial difficulties. The relatively poor baseline BCVA in both groups supports this idea as well.

There is one previous article known, describing ORT development in 24 non-treatment-naïve eyes, treated with aflibercept only after receiving at least 6 ranibizumab injection, reporting an initial 97% ORT prevalence which later decreased to 75% [31].

To our knowledge, our study is the first to report results in treatment naïve patients treated with aflibercept and its connection with ORT development. The Cox proportional hazard model analysis suggested that there was no difference between the two in-label therapies ranibizumab and aflibercept regarding the prevalence of outer retinal tubulation. Lee et al. in the Comparison of AMD Treatment Trials study group evaluated the prevalence of ORTs in ranibizumab and bevacizumab treated neovascular AMD patients, and found no difference between the two drugs related to the prevalence of outer retinal tubulation [21].

In the present study we found a statistically significant difference in the monthly injection rate before and after the appearance of outer retinal tubulation in the ranibizumab treated group. These results suggest that a decrease in the retreatment rate can be expected at patients developing ORT, which may be a critical clinical marker. Although a PRN regimen was mandatory to follow in the treatment procedure according to the Hungarian guidelines at the time, Lee et al. found no difference between the fixed monthly regimen versus PRN regimen in regard to ORT development in patients treated either with ranibizumab or bevacizumab [21].

We found a statistically significant connection between ORT development and the presence of subretinal hyperreflective material at treatment initiation. When subretinal hyperreflective material was present the chance of developing ORT was 2.75 and 11.14 higher in the ranibizumab and aflibercept groups, respectively, in accordance with the results of Lee et al. in ranibizumab and bevacizumab treated patients [21].

Our findings, in concordance with the above mentioned study results suggest that ORT is independent of the chosen anti-VEGF drug or the dosing regimen of intravitreal anti-VEGF treatment. The appearance of ORT suggests that the clinicians can expect a decrease in the number of injections when following a pro re nata ranibizumab regimen. Our fixed bimonthly treatment with aflibercept did not allow us to analyze the injection rate before and after ORT development

in this group. Our study also supports the previously reported higher prevalence of ORT development in the presence of subretinal hyperreflective material at treatment initiation [21].

There is a number of limitations of our study. Namely, the relatively small sample size in the aflibercept group compared to the ranibizumab group, along with the bimonthly follow-up in the aflibercept group. We believe that the number of subjects involved in both groups is comparable with other studies published in the field, while the bimonthly treatment regimen with aflibercept was fixed due to the country-specific guideline regulations. The decreasing number of eyes during the follow-up could also bias the analysis by including patients with increasing disease severity. However, we believe these factors were similar to those in similar studies available in the field. The strength of this report is the comparison of present in-label therapies, ranibizumab and aflibercept in wet AMD patients in relation to ORT besides the evaluation of the injection rate in association with outer retinal tubulation. Real-life data were used, that makes the study more relevant in the daily clinical practice.

Further studies are needed to explore the nature and development of ORTs employing a comparable dosing and follow-up regimen of all three currently available anti-VEGF drugs ranibizumab, bevacizumab and aflibercept.

PART II: AUTOMATIC IDENTIFICATION OF OCT BIOMARKERS

BACKGROUND

The importance of artificial intelligence (AI) and deep learning have already been proven in many medical fields such as radiology, pathology and dermatology. With the development of machine learning, a branch of AI, it is now possible to automatically recognize anatomical structures or lesions in images (segmentation); to assign an image into different categories (classification); and also to predict the outcome of a process (prediction). Their disadvantage is the large amount of labeled data required and a relatively longer training time. In the contrary, classic image processing approaches are well applicable for various detection and segmentation tasks as well, without requiring a large set of image data and by that their development can be faster. In ophthalmology, thanks to diagnostic innovations of the 21st century, for example optical coherence tomography, retinal structures can be visualized with an unprecedented resolution. With the high resolution of SD-OCT scans, we can now develop classical image processing and AI methods in order to reduce diagnostic and therapeutic errors and promote personalized medicine [32].

As mentioned before, retinal disorders are among the main causes of severe vision loss and blindness worldwide, particularly age-related macular degeneration [4]. Concerning the vast prevalence of AMD, development of convenient management strategies is essential.

With the help of OCT, several disease relevant biomarkers have already been identified in wet AMD. Since the activity of the disease is indicated by the presence or absence, or even the amount of these biomarkers, retina specialists decide whether to treat or to observe a patient relying on the information provided by OCT machines.

Main criterion of the retreatment process is based on the recurrence of fluid (SRF, IRF) detected on OCT scans, offering an easy-to-apply decision-making mechanism. However, consideration of other biomarkers could be expedient [32]. For example, reports revealed that any growth in size of PED, a separation of the RPE from the underlying Bruch membrane, indicates urgent anti-VEGF therapy and therefore should be monitored precisely, hence preventing sustained vision

disturbances [32]. The importance of SHRM as a biomarker lies in its connection with visual acuity, as increased thickness of SHRM and subsequent scar development is associated with decreased visual acuity and/or long-term visual loss [32]. ORT is believed to be a specific pattern of photoreceptor rearrangement in response to damage, and a decrease in retreatments can be expected upon its appearance on OCT scans [33].

In 2009, Coscas et al. first reported a new OCT biomarker in wet AMD, the hyperreflective dots [9], also known as hyperreflective foci (HF). Their exact pathomorphological origin remains controversial, since no histological data is available in the current literature. Reports demonstrated that HF response rapidly to anti-VEGF therapy, and were also found to be the first detectable change for each clinical recurrence, even before fluid accumulation. Since the main criteria of the retreatment process in wet AMD is based on the recurrence of fluid (SRF, IRF) detected on OCT scans, the determination of the number of HF can be a potentially more sensitive biomarker than the fluid reappearance, which may facilitate earlier decision making, and achieve greater visual acuity protection [32, 34–39]. It's important to note, that in the above-mentioned studies counting was conducted manually often using a single B-scan of the retina, where the authors depicted that an automated software can help reduce human error [34–39].

Thus, it is comprehensible that relying alone on the presence or absence of fluid upon when deciding whether to treat a patient is no longer adequate, and other biomarkers should be considered as well. Still, evaluation of all the biomarkers supplied by OCT and integration of this data from each scans of the retina of every patient treated in the Retina Departments is no longer feasible for an ophthalmologist, and can only be realized with the help of image processing and artificial intelligence. Segmentation and quantification carried out by image processing and AI could provide the objective and precise tracking of biomarkers on OCT scans.

Hence our aim was to create algorithms to automatically identify biomarkers, namely:

- **SRF/IRF**
- **PED**
- **SHRM**
- **ORT**
- **HF**

METHODS

Our work was based on an intense and constant cooperation with a group of image processing computer scientists from the Department of Image Processing and Computer Graphics, University of Szeged.

The first step of the process was to gather the clinical data for the project. Our studies were carried out in accordance with the principles of the Declaration of Helsinki, and it was performed with the ethical approval of the Institutional Review Board of University of Szeged, Albert Szent-Györgyi Clinical Center.

During the 4 years course of the study data of 28 eyes of 28 patients diagnosed with wet AMD were retrospectively selected at the Department of Ophthalmology, University of Szeged. These patients were either treatment-naïve or treated with anti-VEGF injections and had biomarkers on the OCT scans including SRF, IRF, PED, SHRM, ORT and HF. No other retinal diseases were diagnosed. SD-OCT volume scans with a quality score above 16 were acquired using a Spectralis OCT scanner (Heidelberg Engineering, Heidelberg, Germany, version 6.5.2.0).

One SD-OCT sequence from 22 patients, two consecutive SD-OCT sequences from 4 patients, and three consecutive SD-OCT sequences from 2 patients were collected, hence altogether 36 SD-OCT sequences were used in the studies. In case of patients with multiple captures, the timespan between each recording was at least one month.

As the dataset was acquired from everyday clinical practice, two slightly different parameter sets were used. This is why 14 SD-OCT sequences contained 49 B-scan slices with a distance of 122 μm between consecutive slices, and 22 sequences contained 25 B-scan slices with a distance of 251 μm , giving 1236 B-scans altogether. The data processing was performed on a slice-by-slice basis, so the slice distance did not influence the results. The dimensions of each SD-OCT image were the following: 6 by 6 mm of the macular region, and the slices had a resolution of 512 by 496 pixels with pixel sizes of 11.45 and 3.87 μm , vertically and horizontally.

The 4 year long process of the research began with images of merely 11 patients for the identification of SRF/IRF, but along with other biomarkers joining the research and fine tuning of the algorithms, a 28 patients data bank was reached eventually.

With the help of computer scientists, a framework was designed where biomarkers on OCT scans were identified and annotated by graders. The software used for annotation was the Medical Imaging Interaction Toolkit (MITK). MITK is a free open-source software for development of interactive medical image processing software.

Images were assessed by two independent graders (ophthalmology trainees), prior educated by two retina specialists using particular definitions for each OCT biomarker examined. The definitions were as follows:

SRF: a hyporeflective space between the neurosensory retinal layers and retinal pigment epithelium

IRF: a round or oval hyporeflective space without hyperreflective walls in the neurosensory retina

PED: a separation of the retinal pigment epithelium from the underlying Bruch membrane

SHRM: a medium- to hyperreflective mass between the neurosensory retinal layers and retinal pigment epithelium [23]. In case a transitional zone between an SRF and an SHRM appeared as a gradual change in reflectivity, making it difficult to determine whether the region was an SRF or an SHRM, the zone was marked as SHRM [40].

PED+SHRM: a joint complex of PED and SHRM, when they were no longer safely separable (the RPE is no more identifiable) [40].

ORT: a hyperreflective ovoid-elongated structure in the outer nuclear layer of the retina with lower reflective content [11].

HF: bright spots with a reflectivity higher than the RPE band, with a diameter between 20 and 40 μm [35].

For the annotation of HF, the graders used the pre-sized Paint option (between 3-6 pixels) from the MITK 2D Segmentation tool kit, for all the other biomarkers the Add, Subtract, Correction and Erase tool were used, semi-automated freehand drawing tools in the MITK software. (Figure 4.)

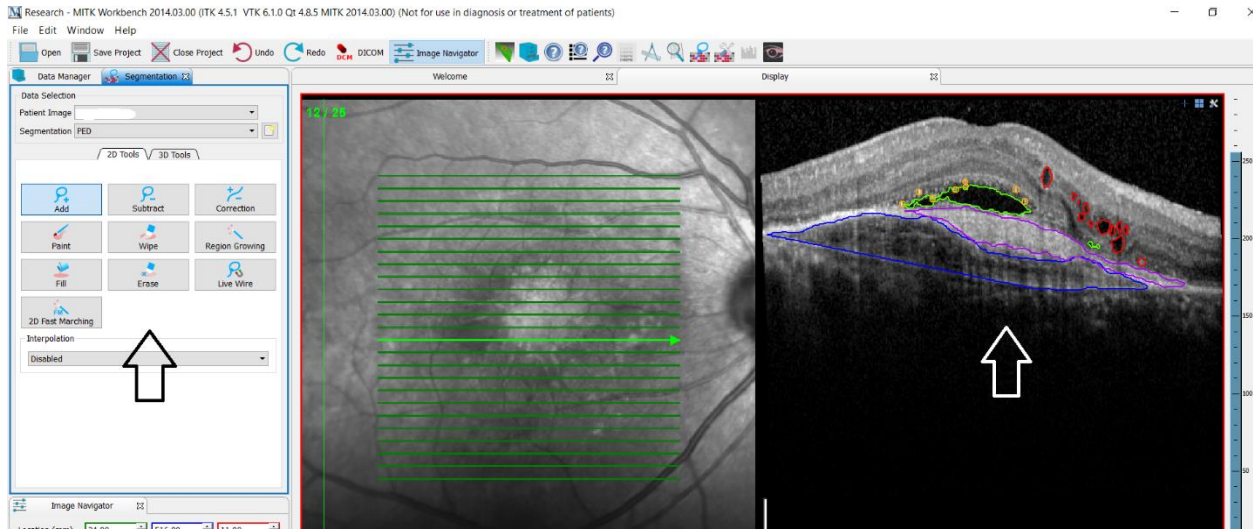


Figure 4.: The user interface of the software MITK. Segmentation tools (black arrow), annotated OCT image (white arrow)

After grader annotation, computer scientists transformed the images using various image processing techniques in order to extract features which best characterize these biomarkers; the feature extraction being determined and guided by the medical team. The image processing methods presented in the dissertation are based on both classical image processing operations and deep learning approaches, involving new ideas and combining them with generally known techniques.

The biomarker automatic analysis started with the annotation and identification of SRF/IRF, followed by PED, SHRM, ORT and HF.

An OCT scan of the retina does not solely include the retina, but also some parts of the vitreous located anteriorly to the retina and also some parts of the choroid posterior to it. In order to separate the retina itself on the B-scan image, its boundaries had to be delineated, namely the internal limiting membrane (ILM) and retinal pigment epithelium (RPE). From the location of these two parameters the thickness of the retina could also be calculated, which was also used. To help identify the biomarkers, another non biomarker abnormality was introduced as well, namely the layer distortion (D), which is caused by either accumulation of SRF or the elevation of PED or SHRM, or all these together.

The structure of our approaches is shown in Table 2. A stair-like frame emerges, where the column of ILM/RPE is without mark, every other biomarker/abnormality is built upon previous knowledge developed. To deal with our biomarkers, first we had to delineate the boundary layers of the retina then search for an approach for measuring the layer distortion.

	ILM/RPE	D	SRF/IRF	PED/SHRM	ORT	HF
Boundary layers: ILM/RPE		X	X	X	X	X
Layer distortion (D)			X	X	X	
SRF/IRF				X	X	X

Table 2.: The structure of our approaches. Each biomarker and layer boundary/distortion column contains the information necessary for its detection.

Statistical analysis contained visual comparison with manual annotations or with re-implemented literature methods, Dice coefficient and Recall (Sensitivity) metrics detailed in each biomarker identification section.

In the next sections a brief summary is given about the methodology of the algorithm development, concentrating on the key elements and presenting our results.

Delineation of boundary layers: ILM/RPE

In the last decade, numerous approaches have been described for the detection of ILM/RPE layers based on distinct techniques. Classical image processing techniques, such as graph theory or active contour method have relatively low tolerance to noise, which is a general characteristic of the images, yet may give a more reliable result than a simple clustering algorithm [41-45]. Machine learning has been widely used recently, also for retinal layer analysis [46, 47]. Despite the fact that methods based on active contour or machine learning provide effective solutions, one should keep in mind their time-consuming nature.

We developed two methods to delineate the boundary layers of the retina.

1. Algorithm based on vertically projected data

The quality of the OCT images is usually poor for image processing, so they often require pre-processing to improve contrast and normalize intensity levels. In favor of reducing the noise of the B-scans, noise filtering and contrast enhancement were applied by using a fuzzy operator [48]. This operation can highlight major retinal layers with the ILM and RPE enhanced as well (Figure 5).

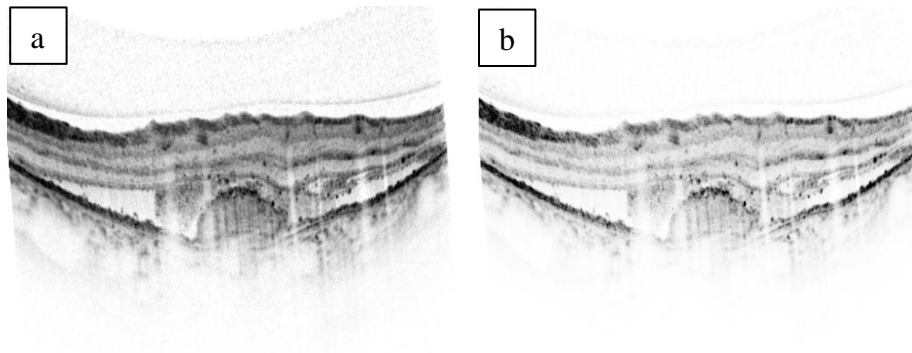


Figure 5.: OCT image before (a) and after (b) applying the fuzzy operator. The filter suppressed the noise and highlighted boundary layers.

The basis of our method is the vertical profile calculation of the image; however, OCT images are affected by retinal vessel shadows that may yield to false detections. According to our observations, the maximal width of detected blood vessel shadows were 10 pixels, thus this problem could be eliminated by dividing the image into columns with fixed width size of 10 pixels. The next major step of our proposed method is to calculate and analyze vertical profiles of these 10 pixel width columns, because large intensity steps in pixel density are assumed to correspond to change of tissue. To further smoothen the signal, the Savitzky-Golay filter was applied [49]. which effectively preserves the adequate high frequency components (Figure 6.). The determination of the RPE is a harder problem than the ILM, because of the reflective choroidal layers underneath. Change of intensity (reflectivity) in these regions results in several peaks in the projections. Fortunately, in most cases, these peaks are not prominent thus do not cause problem in choosing the right location. The algorithm chooses the most important local minimum in the top

and bottom of that given image column (in our case the ILM and RPE) from the projected data to identify the possible inner and outer boundaries. In the next step, with the bars of image columns again next to each other the outliers were filtered, and a curve was fitted to the remaining points drawing the possible ILM and RPE layers.

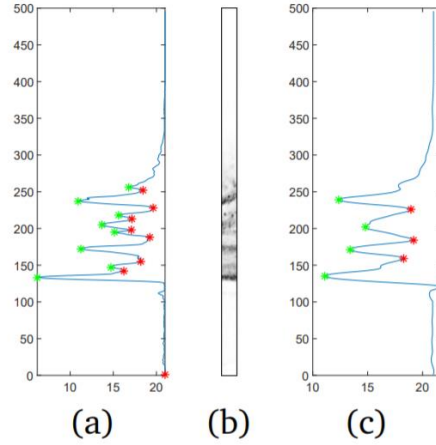


Figure 6.: Calculated projection for vertical profile analysis. Original data (a) of the image column (b) and the Savitzky-Golay filtered data (c)

2. Graph-cut based approach

The vertically projected data approach showed some inappropriate results in layer detection, especially when distortion due to PED or SHRM was present, thus we upgraded it, with keeping some steps of the previous idea, resulting in better outcomes. No preprocessing of the input data was performed, although in many cases the image was remarkably noisy. Graph cut is a semi-automated method that requires seed points [50]. It works as a connection network and with the help of clustering technique, in our case the kernel k-means, the seed points were automatically determined. The clusters of pixels are based mainly upon their intensity and proximity from each other. In better quality images 5 clusters were isolated empirically, while in poorer quality scans the cluster number was increased. With the help of graph cut [51] partitioning was optimized resulting in a better classification (Figure 7.). The ILM can be determined using an Otsu thresholding [52], because the foreground and background (the vitreous and retina) can be surely

distinguished in the clustered image. The designation of RPE layer is more difficult, so similarly to the previous method, vertical projection was calculated in every column and local minimum was chosen from the projected data. Prior to fitting a curve to the remaining points, outlier points found in our results were needed to be extracted. A shape-preserving piecewise cubic spline interpolation was used to determine the RPE layer.

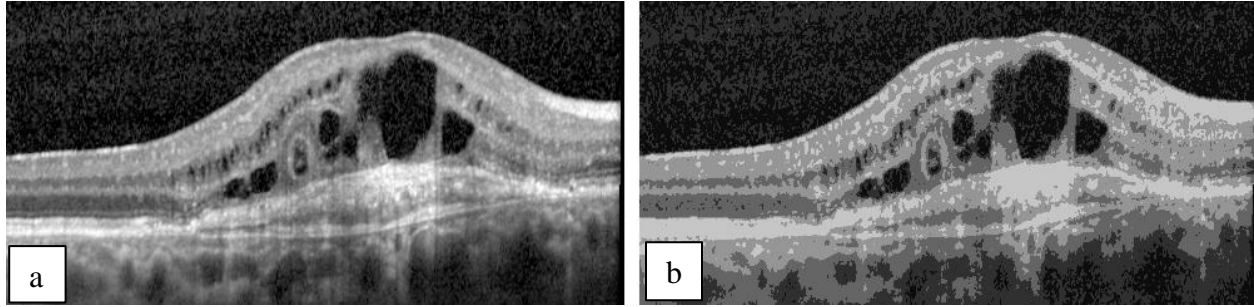


Figure 7.: (a) original and (b) clustered B-scan image using graph cut with 5 clusters

Evaluation and results

The algorithm was evaluated and validated during comparison of its results against the manual delineations of the graders (7 annotated sequences, each of them consisted of 49 slices, 343 B-scans altogether), calculating the mean, maximum and standard deviation of boundary errors for each surface and producing an average result of 1,98, 16,57 and 0,94 pixel respectively (Table 3.).

	Mean	Standard deviation	Maximum
Seq_02	2.01	1.56	17
Seq_03	2.10	0.69	15
Seq_04	1.44	0.65	15
Seq_05	1.96	0.80	19
Seq_06	2.39	1.63	18
Seq_07	2.17	0.64	15
Seq_08	1.83	0.65	17
All	1.98	0.94	16.57

Table 3.: Mean, standard deviation, and maximum error (in pixels) between manually segmented and automatically identified boundary layers in the 7 annotated OCT recordings.

The maximal distance between manually segmented and automatically detected layer boundaries was 19 pixels (ca. $73.5\ \mu\text{m}$). This deviation originates from considerable jumping between B-scans and layer distortions caused by merging of PED+SHRM (Figure 8. image c), with which we deal later in SHRM section. In most cases, the mean errors were under 2 pixels and deviations were insignificant between layers delineated by the annotators and the boundaries determined automatically. Figure 8. shows some examples of the automatically segmented and the manually annotated layers.

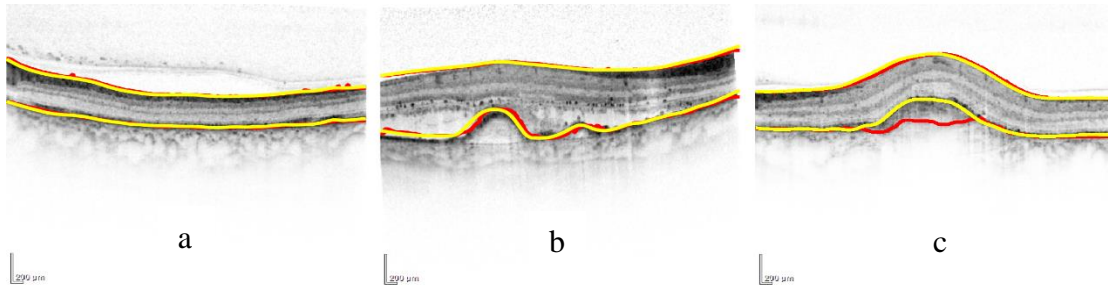


Figure 8.: Visual comparison of grader annotated (red line) and automatically detected (yellow line) layers. B-scan without (a) and with distortion (b, c). Failure of automatic detection due to the layer distortions caused by merging of PED+SHRM (c)

These evaluations were conducted for the vertical projection design. Although there was no measurement comparison between our two techniques after the newly developed graph-cut based method was introduced, but graph-cut based approach was a more advanced solution. It utilized identical steps for RPE delineation as the vertical projection algorithm, with both an upgraded RPE detection in layer distorted cases and a developed ILM determination thanks to the clustering technique. All in all, graph-cut based approach was applied in further biomarker detection.

The most prominent advantage of our method is simplicity, using plain image processing operations that can be parallelized, without relying on many parameters that are difficult to tune.

Measurement of retina layer distortion (D)

The layer distortion (D) is caused by either accumulation of SRF or the elevation of PED or SHRM, or all these together. The detection and determination of start and end points of distorted regions is a useful prior information for biomarker identification. From the previously defined ILM layer points the highest point was selected, i.e. the smallest y coordinate (the (0,0) coordinate at the top

left of the image). Using this point, the image was splitted into two parts, left and right retinal segment. The $\frac{1}{4}$ of either side of the image was disregarded, because they are deformed in most cases and layer information is not reliable due to image registration process during image acquisition. Lowest ILM points were also determined, i.e. the highest y coordinates in the right and left retinal segment and with lines fitted to these side points and the previously defined outer boundary layer (RPE), the starting and end point of the distortion were defined (Figure 9.).

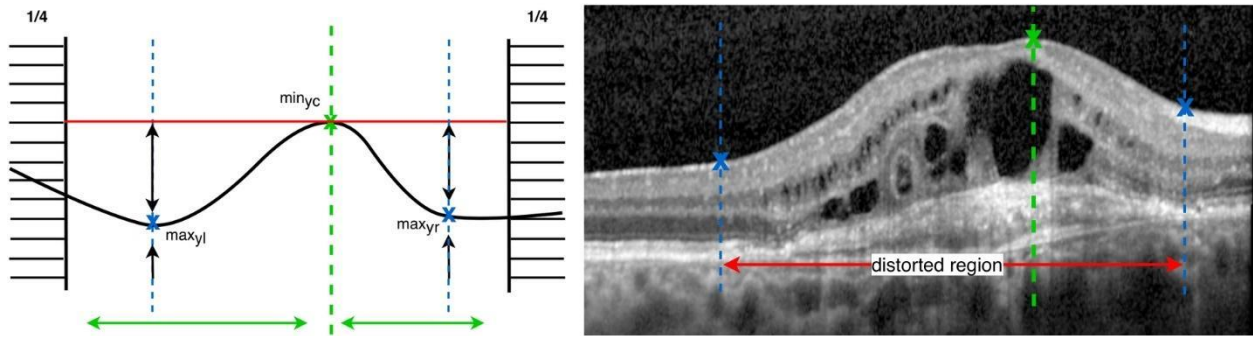


Figure 9.: Schematic view of the determination of distortion. \min_{yc} represents the smallest y coordinate (highest ILM point); \max_{yr} and \max_{yl} are the highest y coordinates (lowest ILM points) in the right and left retinal segments.

SRF/IRF

Many researches focus on the automatic detection of SRF, using classical image processing methods [53-56] and machine learning methods as well [57-59].

Based on prior information, we have developed a method that can simultaneously segment and differentiate fluid compartments, i.e. SRF and IRF. First, the retina was delineated for processing using the graph-cut based method to highlight ILM and RPE layers. An edge-preserving **anisotropic diffusion filter** [60] was used to eliminate various effects of artifacts and image noise. With **quantization** the grayscale image was classified into five intensity levels. This operation facilitated the separation of retinal layers from fluid compartments. During the **binarization** step, the brightest image points were kept, which corresponded to fluid regions. In the next phase an edge-based smoothening with **active contour** process was utilized (Figure 10.).

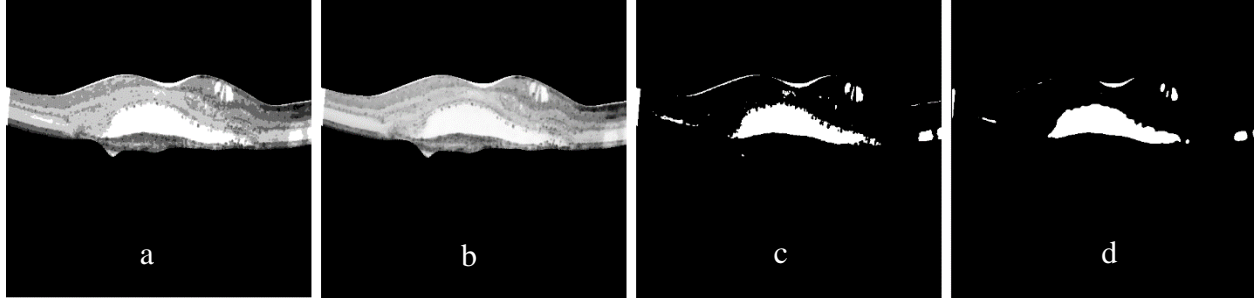


Figure 10.: Intermediate stages of the algorithm for SRF/IRF detection. (a) anisotropic diffusion, (b) quantization, (c) binarization, (d) result of active contour.

With these steps, possible important objects like SRF and IRF were identified, but false segments likewise, which needed to be separated from each other. The classification of the filtered objects was based on three criteria at the given target level:

- the location of the object within the retina [next to the hyperreflective outer boundary (RPE or SHRM)] \rightarrow SRF, *OR* further from the RPE in the neuroretina \rightarrow IRF
- whether the layer is distorted \rightarrow SRF, or not \rightarrow IRF
- what is the extent and the shape of the object (smaller oval shape \rightarrow IRF)

False segments may also appear in the image, commonly caused by retinal vessel shadowing, but considering the fact that these are usually small objects, they can be removed with an area-based filtering. The key stages of the procedure explained are summarized in Figure 11.

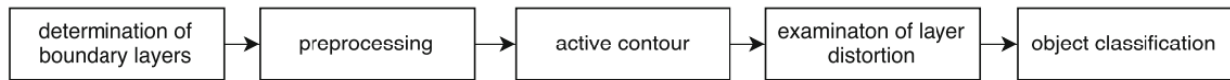


Figure 11. Flowchart of the proposed

Evaluation and results

The annotations were not yet available at this stage of the research, so we did visual comparison of our results with the outcome of two different methods from the literature (Wieclawek's et al [61] and Wilkins et al [62]) (Figure 12). These two existing methods, both segmenting IRF, were re-implemented by our research group according to the original papers. Both literature algorithms

and our approach applied various **simple** classical image processing solutions and started the biomarker segmentation with defining the boundary layers, but this was the only similarity between them.

The segmentation results of the method developed by Wilkins were almost the same as ours, but in many cases, it kept false objects. The other method from literature by Wieclawek detected fewer possible IRF regions. The reason for their inaccuracies could be the fact that they apply a pre-specified threshold. In contrast, our method uses dynamic requirements based on prior information, thereby these mistakes could be eliminated. The re-implemented methods were tested on images containing subretinal and intraretinal fluid as well. The literature approaches could detect SRF, because its reflectivity is similar to IRF, but our algorithm could distinguish these two types of fluid compartments from each other marking them with a different color (Figure 12.).

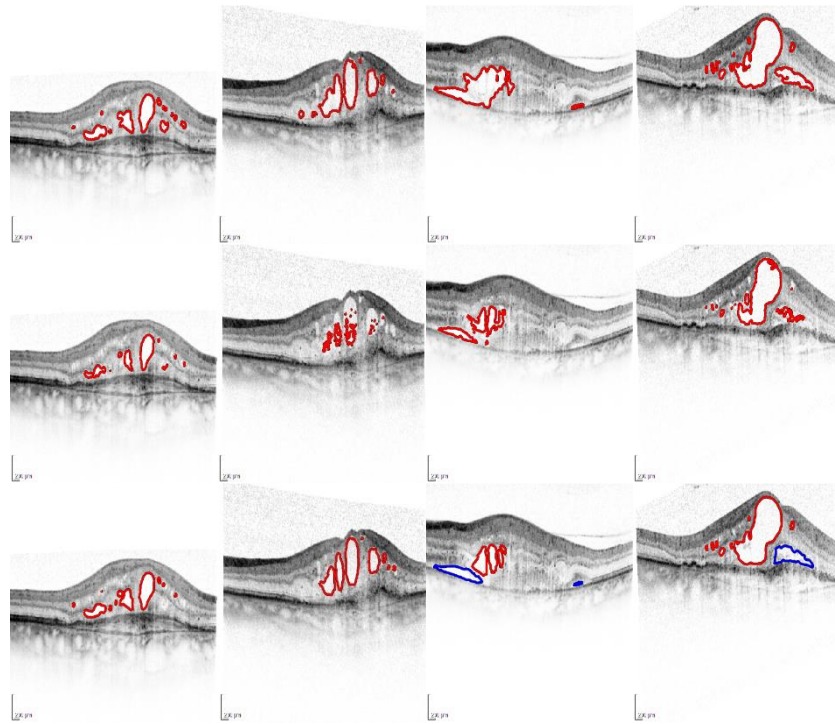


Figure 12.: Illustration of detected intraretinal (red) and subretinal fluid (blue) by three different algorithms. Rows (from top to bottom): Wilkins et al., Wieclawek et al., our method. The first two columns contain only IRF, while the last two have SRF as well.

Automatically identified subretinal fluid can be quantified and then visually presented in different forms. In contrary to the ordinary slice-by-slice display of SRF, providing a good depth content

within a slice, the colored overlay provides anatomical content and quantitative parameters (thickness and extent of the SRF). To create the color overlay the detected SRF segments were summarized in each slice and using this information a new image was generated overlaying the original scanning laser ophthalmoscope infrared image (Figure 13.).

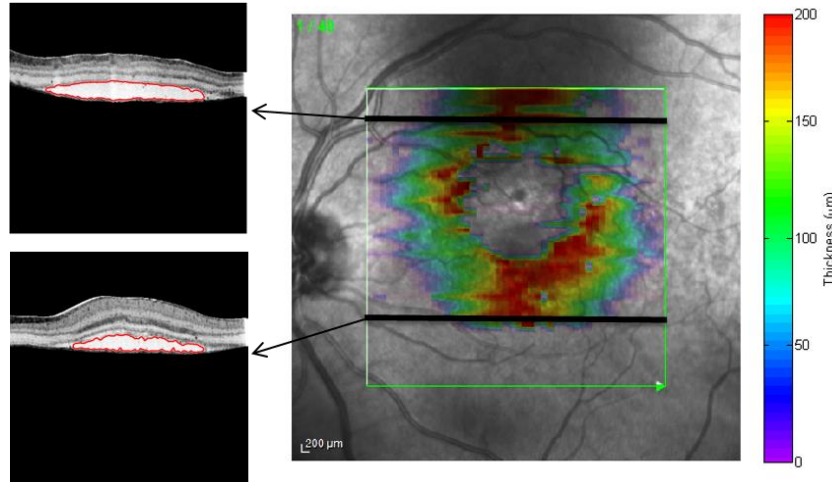


Figure 13.: SRF volume thickness color overlay. Left: Red marks denote the SRF in each slice. Right: The color scale describes the SRF thickness.

PED

PED is a separation of the RPE from the underlying Bruch membrane. With the help of our graph cut based approach the boundary layers were determined (the elevated RPE was found) and using the layer distortion measurement method the start and end point of distortion (in this case the RPE elevation) was located. In order to identify a PED, we needed to define the original position of the RPE in the distorted region. The points of the RPE layer were taken until the previously detected starting point of the distortion was reached in both sides and a quadratic curve was fitted to these data. Since it cannot be assumed that the retina was visible on the whole image, the first two detected points on both sides of the image was neglected during fitting.

The height of the PED can be defined as the distance between the maximum point of the distorted zone (maximum point of PED) and the pertaining minimum point of the potential normal layer. The extent of the PED can be calculated from the enclosed area (Figure 14.).

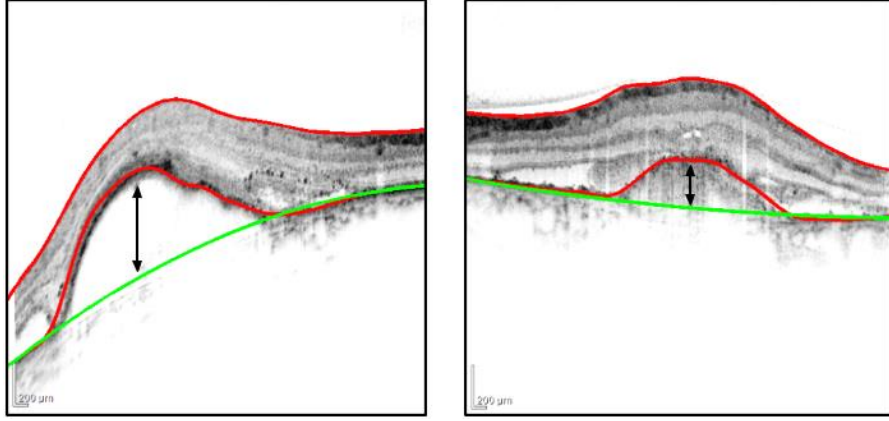


Figure 14.: Example of our PED detection. Red curve demonstrates the detected boundaries (ILM/RPE) and green indicates the original RPE position/Bruch membrane.

SHRM

The retinal morphology and reflectivity may change resulting from changes in the neovascular tissue. A good example for that is subretinal hyperreflective material, since it is composed of many elements, including fibrin, vessels, blood. The content of SHRM alters over time, thus the reflectivity is heterogeneous and changing. In early stages of the disease the border (RPE) of the PED is clearly distinguishable from the SHRM and can be separated from each other. Nevertheless, due to fibrosis of the neovascular tissue, the SHRM and the vascularized PED become isorefective and the location and presence of RPE will be unclear. In these cases, the abnormalities are so inseparable that they are managed as one structure (PED+SHRM).

Just a few literature methods are available for identification of PED/SHRM and most of them are applying deep neural network (DNN) and convolutional neural network (CNN) [40, 63]. Our method is based on classical image processing techniques, and opposite to deep learning methods it requires substantially less data and training time.

Classification of PED vs. SHRM

To localize these biomarkers, a graph cut technique with clustering was used. The pixel-intensities of PED and SHRM are close to each other, but with proper clustering they can be separated.

Clusters are sampled from the distorted region areas (determined utilizing our layer distortion approach), because both PED and SHRM are located around the RPE layer, along or near the distortion. To reduce the clusters to the sought segments only, our prior technique was applied to detect SRF/IRF and exclude them. We investigated in a specific range around the RPE layer, since SHRM is located above the RPE in the subretinal space. Therefore, a threshold was determined using the thickness of the retina:

$$T = \begin{cases} \max(CF_{(x, \max(y))}), & \text{if } \max(CF_{(x, \max(y))}) > 0 \\ RT_x \cdot 0.2, & \end{cases}$$

where CF represents the binary image, x is the actual x position, $\max(y)$ is the actual maximum y position and RT denotes the calculated retinal thickness (from ILM and RPE layers). The clusters above the RPE are separated as SHRM, and underneath it as PED.

Evaluation and results

The evaluation dataset consisted of 18 recordings with 25 and 49 slices (594 B-scans altogether). The results of our method were compared with the annotations produced by the graders, using the Dice coefficient and Recall metrics. The location of the possible SHRM is not clearly detectable in many cases, due to fibrotic changes and merging of PED and SHRM. Most of the errors committed by our algorithm were in these parts, which were annotated as PED+SHRM by the graders. Even so, the average Dice coefficients were above 0.75 in both cases and the Recall was 0.93 for PED and 0.77 for SHRM (Figure 15.). In Figure 16. illustration of our detection are demonstrated. The CNN-based method mentioned earlier [40] also determined these biomarkers with similar efficiency. This means that in order to segment PED and SHRM a complex system with a large image database is not definitely necessary.

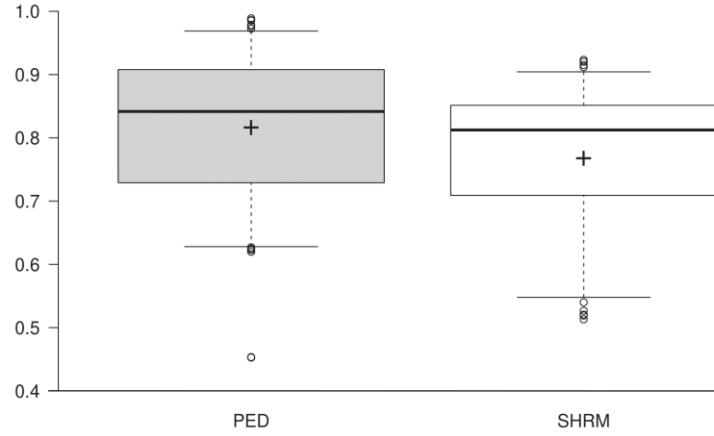


Figure 15: Box plots representing the Dice coefficients of PED and SHRM detection by our developed algorithm. In both cases, the mean Dice coefficients were greater than 0.75. Black cross = mean Dice coefficient

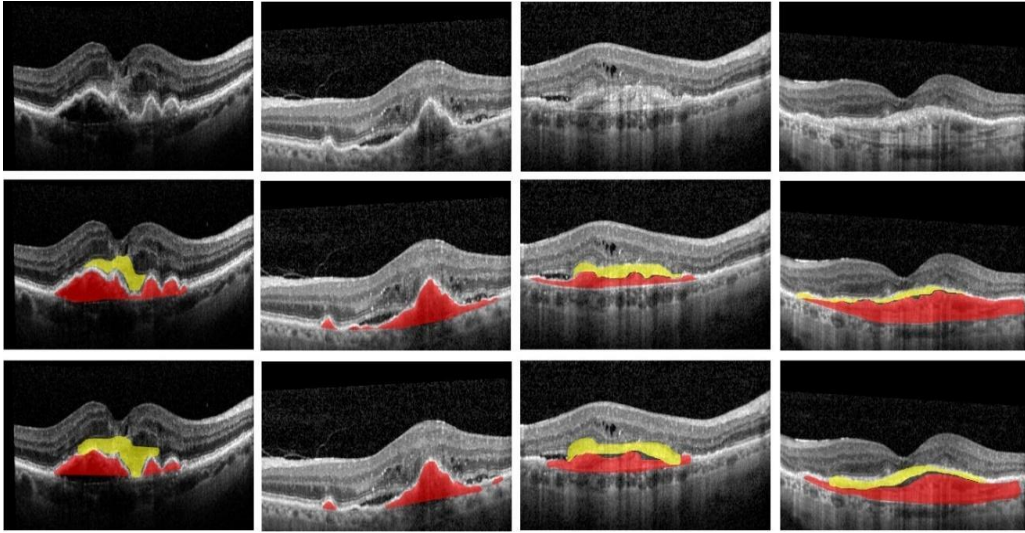


Figure 16.: Examples of grader annotated (middle row) and automatically detected (bottom row) PED and SHRM, marked with red and yellow, respectively.

ORT

To our knowledge, our study was the first to report methods of automatic identification for ORT segmentation. Two sorts of approaches were tried: a classical image processing technique and a machine learning method.

Detection using classical image processing operators

The ORT has hyperreflective border with a hyporefective content, thus our procedure is based on finding the hyperreflective points of its border. Using Wiener filter the image noise was reduced, then using a Hessian detector the reflective points were localized [64]. Since ORT is in the outer nuclear layer within the distorted retina region or its neighborhood, we used our prior method for distortion location and calculated the retina thickness (knowing the boundaries of the retina with our graph cut method) keeping those points only located in the outer third of the retina. Biomarker areas previously identified such as SRF/IRF were excluded using their detection methods. To distinguish between the hyperreflective wall of the ORT and its surroundings, adaptive histogram equalization was performed with hysteresis thresholding. After this step, points that were part of an object were retained, and their distance map [65] was computed. The map was thresholded and finally the convex hull of the objects was computed (Figure 17.).

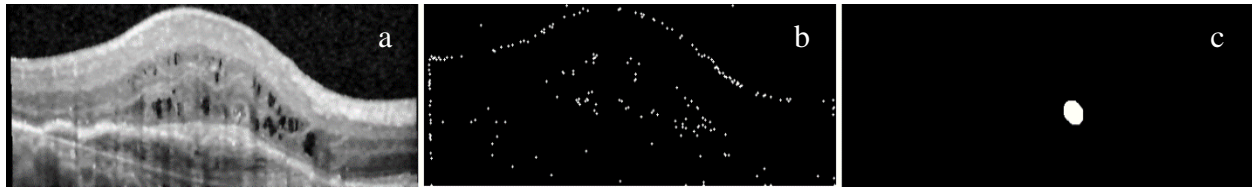


Figure 17.: Some stages of the algorithm based on classical image processing operations for ORT identification. (a) Wiener filter, (b) Hessian detector, (c) segmented ORT.

ORT localization with neural network – i.e., machine learning method

The previously presented method based on classical image processing operators was overly meticulous against more diverse images, thus we have developed a more general approach, a variant of deep neural network, namely a convolutional network (CNN). CNNs are one of the most suitable deep learning architectures for imaging data, providing new solutions day by day. It is well-known that the more training data available for deep neural networks, the more precise results achieved. In our case 8 sequences containing 25 or 49 B-scans (320 B-scan all in all) were annotated by each grader, but for the network training only the ORT positive images were used, ergo 132 scans altogether. Since the available amount of training data was small, real time

augmentation techniques were administered to raise the data, thus reducing overfitting. The techniques for augmentation were: rotation, horizontal mirroring, shearing, vertical and horizontal shifting. To save computation time, the images were resized (220×256 pixels), and the intensities were normalized to be in range [0, 1]. and then standardized. Nested cross-validation was applied for hyperparameter tuning and for measuring the test error. The original version of the U-Net architecture [66] (the name of it originates from the U shape of the architecture), a type of CNN, was used, but slightly modified:

- reducing the number of layers - to accelerate training time,
- doubling the number of filters in each layer - to supply larger capacity to the model
- administering Dropout [67] between all convolutional layers but the last two - to prevent overfitting

Stochastic gradient descent was used for optimizing the loss function.

Evaluation and results

Merely 0.3% of the whole dataset was segmented as ORT, a biomarker that varies in number and presents frequently small both in size and in number in the images. For statistical analysis only those images were considered which contained segmentation using global and object Dice coefficients, since there were images with multiple ORTs. Dice coefficient measures how precise a segmentation is; however, our main goal was not a perfect segmentation of ORT, but the cardinality of them objects. To fulfill this, the object level Recall was measured, which describes how many annotated ORT objects were found in the segmentation. To create a union of annotations for the images graded by the two experts, the masks of the graders' annotated image sets were combined using a logical operator. This served as a basis for statistical analysis. Table 3. shows the average of achieved Dice and Recall values. The object level Dice coefficient was better than the global one, suggesting that if ORTs were present the model could segment them properly. The object level Recall reached the performance of graders with 0.847, meaning that our model can assist ophthalmologists to identify ORT in OCT images.

	Recall (object)	Dice (global)	Dice (object)
MD_1	0.908	0.682	0.665
MD_2	0.812	0.625	0.590
modified U-Net	0.847	0.579	0.583

Table 3.: Performance of the graders (MD_1 and MD_2) and the proposed convolutional network. The first two rows represent the graders results against each other. In the third row the results of our modified U-Net architecture against the union of the graders are listed.

Taking a brief look at the segmentation examples seen in Figure 18., images demonstrate a reasonably good performance of the model both in precise localization and finding each object in most of the cases. The reason for mistakes were usually elongated open-type ORTs as shown in the third column.

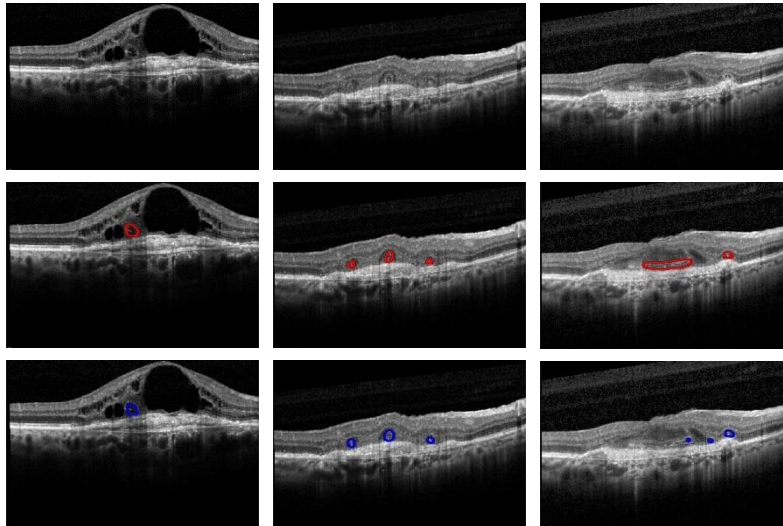


Figure 18.: Examples of annotated (middle row) and detected (bottom row) ORT biomarkers.

The first row contains the original input images. The reason for mistakes were usually elongated open-type ORTs as shown in the third column.

HYPERREFLECTIVE FOCI

For the automatic identification of HF, machine learning techniques were utilized. 8 different neural networks were developed and compared with the annotations of the two graders.

The 8 different neural networks consisted of 4 basic types:

- Artificial Neural Networks (ANN),
- Deep Rectifier Neural Networks (DRNs),
- Convolutional Neural Networks (CNN), and
- Fully Convolutional Neural Networks (FCN),

and 4 modified types of neural networks, such as:

- Pixel ANN and Pixel DRN (just using raw pixel data to validate the quality of the extracted features),

- Split DRN (it had a split first hidden layer; half of the neurons being connected to the pixel input and the other half being connected to the feature vectors)

- Dice DRN (trained with the same error measure that was used in the final evaluation).

When constructing our automatic segmentation framework, data processing was divided into consecutive stages. These were data preparation, feature extraction, training, and evaluation performed on different models. It should be noted that the 8 different methods required different pre- and post-processing steps and they have different data flowcharts. The flow chart of the data preparation can be seen in Figure 19., while the flowcharts for the segmentation methods are depicted in Figure 20.

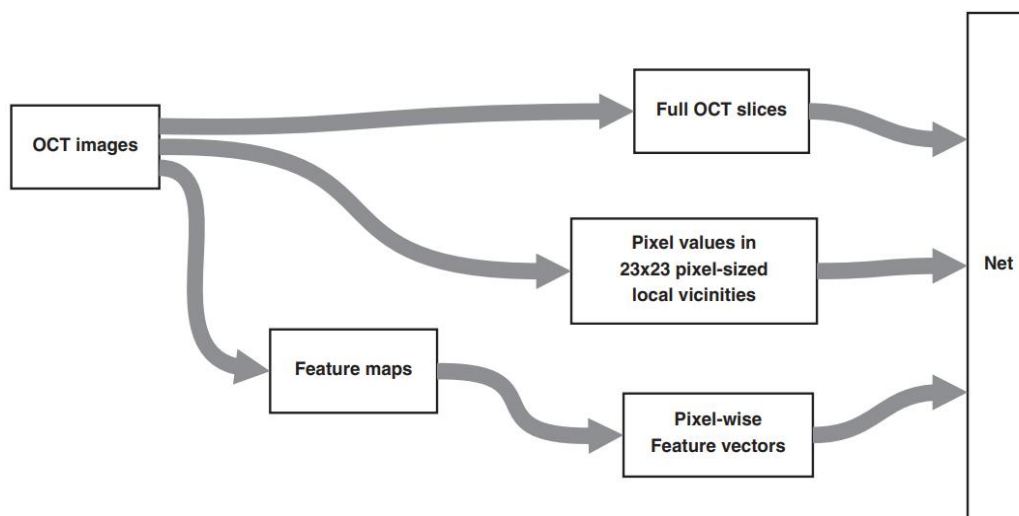


Figure 19.: Steps and output data of the data preparation

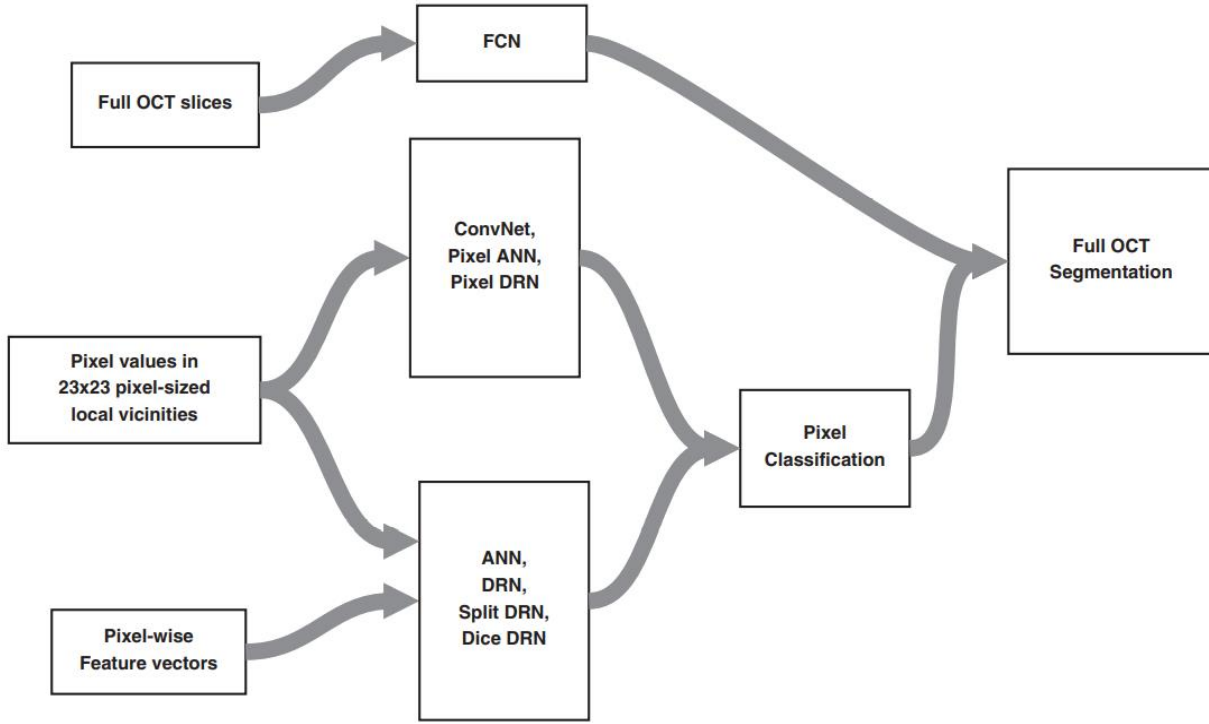


Figure 20. An outline of the data flow used in the image segmentation process

Data preparation

Two sets of OCT data were used for our study. The **first data set** was used for training and testing the methods, and for selecting the best network structure. The purpose of the **second dataset** was to further validate the best performing network structure. The first data set was collected from 16 eyes of 16 patients diagnosed with wet AMD, as per follows: 1 SD-OCT sequence from 11 patients, 2 consecutive SD-OCT sequences from 3 patients, and 3 consecutive SD-OCT sequences from 2 patients, thus altogether 23 SD-OCT sequences (14 SD-OCT sequences with 49 B-scans, and 9 sequences with 25 B-scans, giving 911 B-scans altogether).

The 911 B-scans, after being converted from white-on-black to black-on-white images, were annotated separately by the two graders for training and evaluation purposes. The first dataset was divided into two parts. Here, 19 out of 23 sequences were used for the preparation of the methods, i.e., designing the features and training the networks. 4 sequences (which we called the test dataset), each containing 49 slices, were kept for evaluation purposes. These recordings originated

from 4 different patients, whose data was not included in any way in the training and data preparation process. The data flow of the procedure is shown in Figure 21. Images of the test dataset were selected according to the criterion of holding various amounts of HF.

The second dataset consisted of 8 sequences of 7 patients. The data of the selected patients was not included in the first dataset in any form. Each sequence of the second dataset contained 25 slices, and – just like the first test dataset – also had various HF counts. This data set – which we called secondary test dataset – was entirely used for further validation of the best performing network.

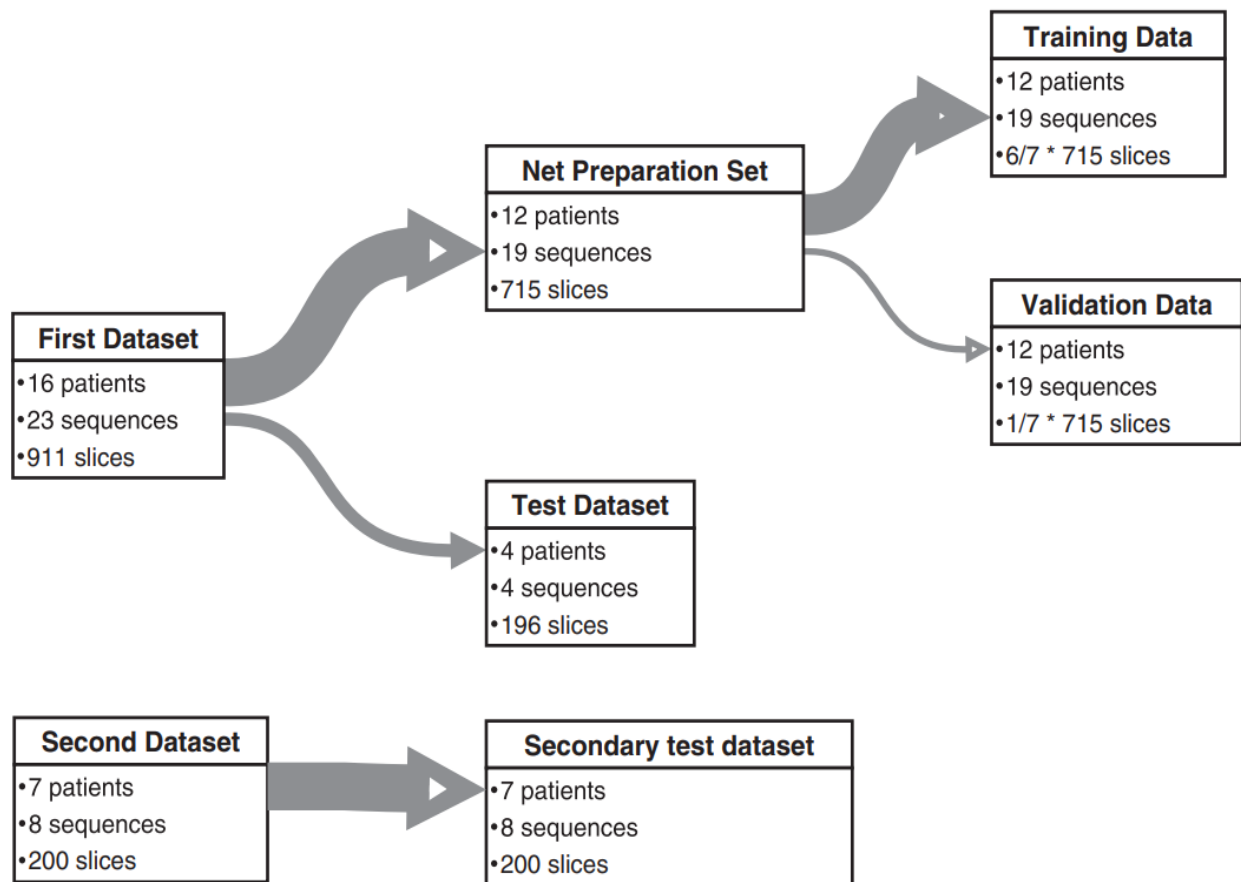


Figure 21. Illustration of how the data was partitioned

The input of the Convolutional and Fully Convolutional networks were the images themselves, while the input of other Neural Networks was created on a pixel-by-pixel basis. A feature vector was generated for each pixel that served as a basis for classification. These features characterize

the indicators of HF, and they were extracted by applying image processing techniques described below.

Image preprocessing and feature generation

The feature extraction for the DRNs and the ANN was performed after taking into account the suggestions made by retina specialists. According to the literature, our relevant targets are small bright spots with a reflectivity higher than the RPE band, and have a diameter between 20 and 40 μm [35]. They are also known to occur near sub-retinal fluid [34–37]. HF are similar to pigment particles and blood vessels in the inner retinal layers, so in order to clearly distinguish HF from them, different types of features were used.

First, the feature vectors contained raw pixel information. For each pixel, the first feature was its intensity, and the 23×23 vicinity and local information was provided concerning the intensity of the pixels for the neural networks (see Figs. 19. and 20.).

Furthermore, other pixel-wise feature vectors were generated by using various image processing techniques, described below.

Eleven feature values were obtained by convolving the images with different Laplacian of Gaussian (LoG) filters, providing a basis for bright spot detection. The spatial properties of the HF were also exploited, i.e., commonly located in the outer retinal layers, and/or around pockets of fluid accumulation [34–37]. For this reason, the ILM and RPE layer was determined. Subretinal fluid was also detected using our methods (described earlier in the thesis). Having detected the boundaries and the fluid, distance maps were generated from the specified regions, and for each pixel, three distance values were included (the distance between the processed pixel and the ILM, RPE and SRF) in the feature vector. These values are depicted in Figure 22e–g. Next, difference between HF, pigment particles and blood vessels were quantified. Pigment fragments and blood vessels cast shadows on OCT images, which appear as a 20–40 pixel long vertical dark area under the bright spot, whereas HF do not generate such shadow. This is why differentiating between HF, pigments and blood vessels was achievable by calculating the average of the pixel values below the pixel coordinate shown in Figure 22h.

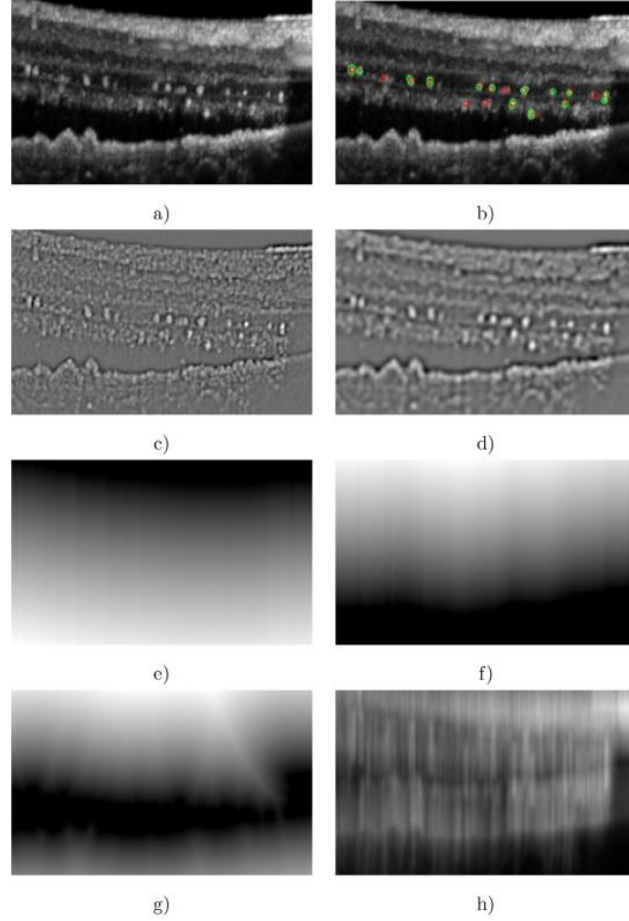


Figure 22. Feature maps generated by image preprocessing: a) the original image; b) location of HF according to the two experts (red, green); c–d) the results of LoG filters; e–g): a distance map from ILM, RPE and SRF, respectively; h) a convolution with the vertical mask.

Training the nets

Using the training dataset and the extracted features, several networks were trained for the task to classify the pixels as either HF pixels or as part of the background.

As only a small percentage of the training data pixels belonged to the HF (less than 1 thousandths of the pixels were HFs), we chose to re-sample the data during training using **probabilistic sampling**. One should know that in case one of the classes is significantly overrepresented during training, it might bias the Neural Network and favor the most frequent class on its output. In our case only a few pixels formed part of the HF, so this problem created a serious issue during

training. To overcome this problem, it is necessary to balance the class distribution by presenting more examples taken from the rarer classes to the model. Usually, we have no way of generating additional samples from a rare class, so balancing can be achieved by either reducing the number of the examples belonging to the most common classes (downsampling) or by presenting the rare examples more frequently (upsampling). Probabilistic sampling [68] offers a third option where one combines the two previous sampling methods at the same time.

Evaluation

After the training phase, segmentations of the test data were performed with the trained nets. The retina specialists of the research group evaluated the output images, and also numerical measurements were performed using the Dice coefficient as an accuracy metric. For a systematic evaluation, the scores were compared with the annotations of the two graders. Afterwards, to get a baseline for the desired accuracy, annotations of the graders were compared with each other. The annotated data were compared with the segmentation results of the neural networks on a pixel basis, using the original and a modified version of the Dice formula. Although the annotation of the data was performed using a pre-sized Paint tool in MITK, it included small (1–2 pixel) misalignments at the edges of the HF due to the imperfect sizing and placement by the graders (see, e.g, the manual segmentations in Figure 23.). Therefore the modified Dice formula included a small tolerance, i.e., a positive pixel was accepted – of either the reference or the segmentation – as a true positive, if the other image had a positive label in a given vicinity.

Results and discussion

Examples of the segmentations produced by the nets can be seen in Figure 23. The numerical results of the various comparisons are listed in Table 4. The values in the table that were nearly as accurate (they had a Dice coefficient higher than 95%) as the overlap between the two annotating doctors have been highlighted. Observing the results, it is clear that the performance of different networks varied greatly. The best results were produced by the FCN. In most cases its score was close to – or even higher than – the baseline between the annotators. Apart from the FCN, other networks performed quite well, especially the DRN using all the features, and the Dice DRN. The good performance of the Dice DRN can be accounted for by its loss function, as this network was

trained with the same error measure that was used in the final evaluation, hence the training process could be improved. The DRN with the full feature set, however, had additional information extracted from the image, and helped to find a better model.

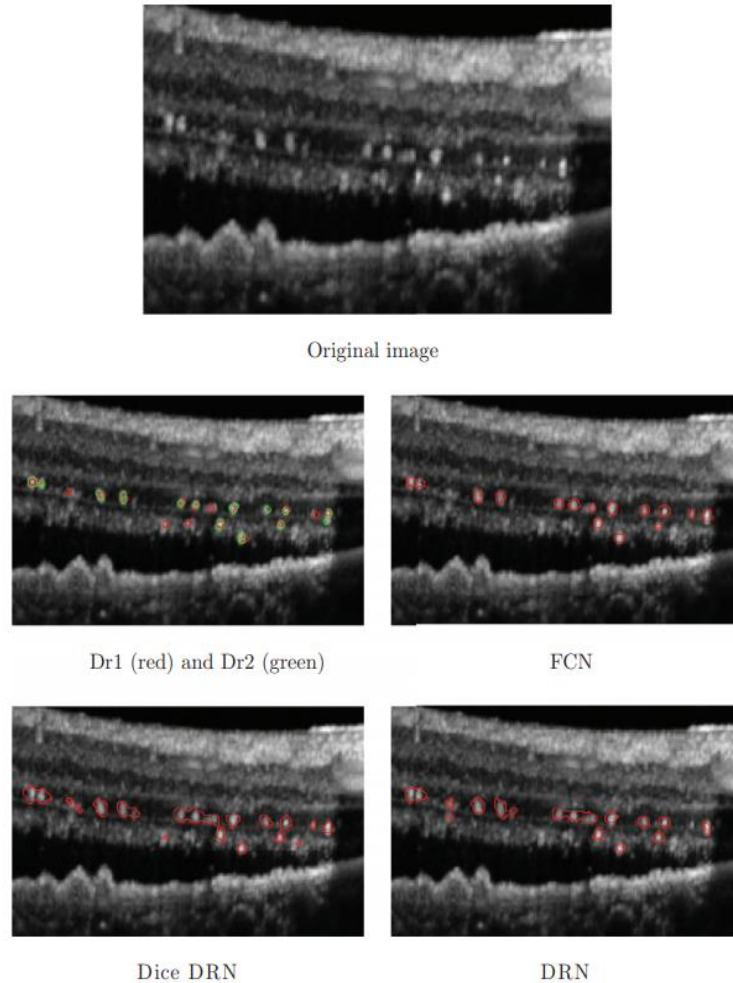


Figure 23. Segmentation examples of an OCT image by manual annotation (Dr1 and Dr2 representing the two graders), and by the three best neural networks: FCN, Dice DRN and DRN

Based on these results, it can be concluded that FCN, Dice DRN and Full DRN could provide accurate segmentations of HF on OCT images, which in terms of accuracy are close to the correlation values gained between manual segmentations.

	Tolerance: 0 pixel				Tolerance: 1 pixel			
	Seq_1	Seq_2	Seq_3	Seq_4	Seq_1	Seq_2	Seq_3	Seq_4
Dr1 ↔ Dr2	0.463	0.420	0.378	0.401	0.680	0.619	0.578	0.594
Dr1 ↔ FCN	0.480	0.329	0.386	0.310	0.695	0.516	0.573	0.493
Dr1 ↔ ConvNet	0.229	0.088	0.115	0.094	0.374	0.161	0.196	0.174
Dr1 ↔ ANN	0.408	0.188	0.181	0.164	0.553	0.263	0.267	0.238
Dr1 ↔ DRN	0.468	0.283	0.358	0.302	0.610	0.384	0.494	0.412
Dr1 ↔ Pixel ANN	0.408	0.199	0.185	0.199	0.546	0.277	0.268	0.275
Dr1 ↔ Pixel DRN	0.455	0.289	0.293	0.306	0.603	0.389	0.420	0.403
Dr1 ↔ Split DRN	0.470	0.289	0.353	0.285	0.626	0.391	0.482	0.395
Dr1 ↔ Dice DRN	0.478	0.282	0.371	0.277	0.623	0.381	0.495	0.381
Dr2 ↔ FCN	0.548	0.420	0.438	0.402	0.747	0.589	0.629	0.579
Dr2 ↔ ConvNet	0.241	0.109	0.117	0.119	0.382	0.187	0.204	0.211
Dr2 ↔ ANN	0.432	0.161	0.229	0.174	0.568	0.229	0.320	0.244
Dr2 ↔ DRN	0.520	0.288	0.408	0.352	0.657	0.383	0.515	0.461
Dr2 ↔ Pixel ANN	0.435	0.208	0.229	0.225	0.572	0.278	0.307	0.306
Dr2 ↔ Pixel DRN	0.500	0.309	0.344	0.354	0.642	0.411	0.455	0.462
Dr2 ↔ Split DRN	0.511	0.307	0.391	0.339	0.663	0.415	0.506	0.448
Dr2 ↔ Dice DRN	0.524	0.311	0.399	0.334	0.668	0.409	0.498	0.435
	Tolerance: 2 pixel				Tolerance: 3 pixel			
	Seq_1	Seq_2	Seq_3	Seq_4	Seq_1	Seq_2	Seq_3	Seq_4
Dr1 ↔ Dr2	0.779	0.710	0.660	0.673	0.829	0.744	0.707	0.715
Dr1 ↔ FCN	0.801	0.614	0.673	0.586	0.858	0.669	0.730	0.639
Dr1 ↔ ConvNet	0.494	0.224	0.266	0.244	0.611	0.397	0.339	0.316
Dr1 ↔ ANN	0.733	0.407	0.369	0.372	0.807	0.462	0.427	0.430
Dr1 ↔ DRN	0.793	0.561	0.631	0.577	0.856	0.626	0.697	0.635
Dr1 ↔ Pixel ANN	0.739	0.441	0.355	0.427	0.809	0.509	0.407	0.488
Dr1 ↔ Pixel DRN	0.781	0.576	0.540	0.556	0.845	0.642	0.614	0.608
Dr1 ↔ Split DRN	0.796	0.559	0.608	0.526	0.864	0.618	0.660	0.587
Dr1 ↔ Dice DRN	0.802	0.557	0.633	0.527	0.866	0.622	0.693	0.579
Dr2 ↔ FCN	0.825	0.649	0.713	0.656	0.858	0.671	0.740	0.701
Dr2 ↔ ConvNet	0.494	0.254	0.279	0.285	0.601	0.326	0.347	0.364
Dr2 ↔ ANN	0.715	0.339	0.433	0.361	0.768	0.378	0.478	0.403
Dr2 ↔ DRN	0.802	0.538	0.641	0.626	0.842	0.585	0.688	0.681
Dr2 ↔ Pixel ANN	0.727	0.410	0.414	0.436	0.774	0.461	0.446	0.483
Dr2 ↔ Pixel DRN	0.784	0.559	0.596	0.614	0.823	0.599	0.634	0.654
Dr2 ↔ Split DRN	0.796	0.551	0.621	0.574	0.837	0.590	0.655	0.614
Dr2 ↔ Dice DRN	0.803	0.553	0.663	0.577	0.841	0.595	0.697	0.618

Table 4.: Dice coefficients got for the various types of neural networks with different pixel tolerances. (Values above 95% of the baseline are highlighted in bold.).

As a further validation of the results, we decided to perform tests using the 8 new sequences of the secondary test dataset and evaluated the best performing net, the FCN. These results can be seen in Table 5.

In the new tests, the network was assessed on the first and second dataset separately. The assessment was extended by including the Precision, Recall, and AUC values. Looking closely at the results, one can see a slight shift in the balance towards the Recall value. It seems that the networks tend to segment a few more HF than the human annotators. This may be due to the uncertainty of the annotations in the test dataset. We also found that the results for first and second

datasets were comparable, and in both cases, the FCN produced results close to the baseline in the Dice, Precision, Recall and AUC values.

Tolerance: 0 pixel					
Test		Dice	Precision	Recall	AUC
	Dr1 ↔ Dr2	0.443	0.418	0.473	0.706
	Dr1 ↔ FCN	0.431	0.429	0.433	0.762
	Dr2 ↔ FCN	0.504	0.536	0.477	0.803
Secondary test					
	Dr1 ↔ Dr2	0.386	0.342	0.442	0.676
	Dr1 ↔ FCN	0.344	0.302	0.401	0.704
	Dr2 ↔ FCN	0.442	0.415	0.428	0.776
Tolerance: 1 pixel					
Test		Dice	Precision	Recall	AUC
	Dr1 ↔ Dr2	0.654	0.615	0.699	n/a
	Dr1 ↔ FCN	0.638	0.621	0.656	n/a
	Dr2 ↔ FCN	0.696	0.712	0.680	n/a
Secondary test					
	Dr1 ↔ Dr2	0.589	0.528	0.666	n/a
	Dr1 ↔ FCN	0.554	0.492	0.634	n/a
	Dr2 ↔ FCN	0.604	0.593	0.617	n/a
Tolerance: 2 pixel					
Test		Dice	Precision	Recall	AUC
	Dr1 ↔ Dr2	0.748	0.701	0.802	n/a
	Dr1 ↔ FCN	0.741	0.712	0.773	n/a
	Dr2 ↔ FCN	0.771	0.783	0.760	n/a
Secondary test					
	Dr1 ↔ Dr2	0.690	0.622	0.776	n/a
	Dr1 ↔ FCN	0.673	0.605	0.757	n/a
	Dr2 ↔ FCN	0.681	0.673	0.689	n/a
Tolerance: 3 pixel					
Test		Dice	Precision	Recall	AUC
	Dr1 ↔ Dr2	0.795	0.742	0.856	n/a
	Dr1 ↔ FCN	0.798	0.764	0.834	n/a
	Dr2 ↔ FCN	0.804	0.820	0.789	n/a
Secondary test					
	Dr1 ↔ Dr2	0.751	0.699	0.863	n/a
	Dr1 ↔ FCN	0.752	0.724	0.854	n/a
	Dr2 ↔ FCN	0.722	0.748	0.732	n/a

Table 5.: Evaluations of the FCN on the Test and Secondary test datasets. The table contains statistics that compare the annotations of the two graders with each other and with the FCN.

The columns give Dice, Precision, Recall and AUC values evaluated on the complete datasets. The sub-tables give statistics for different tolerances in the vicinity of positive pixels. (Values above 95% of the baseline are highlighted in bold.).

Conclusions

Previous research on the topic mainly focused on the assessment of HF in manually annotated datasets counting in a single B-scan [34–39] without automated evaluation, while our study focused on the automatic segmentation of HF in whole OCT sequences. One parallel study about automatic detection of HF was published [69] on arXiv (a peer-review free system), where the authors applied deep learning based methods using variants of the U-Net [66]. The approach

described by our group differs from the previous study involving training nets from two parallel sets of uncertain data. Moreover, two independent evaluations were performed of our models by comparing the network outputs with the annotations of the two graders.

The novelty of our study lies in testing various neural networks for the automatic segmentation of HF. Unlike previous studies, neural networks were trained not only by relying on raw pixel information, but also on features extracted by image preprocessing steps. The annotations of two different graders were compared with each other to determine a baseline accuracy for HF segmentation as a reference for our methods. Our methods were evaluated on two test datasets, and we found that our methods led to Dice Coefficients values comparable with the similarity score between manual annotations performed by the physicians of the research group. Lastly, we demonstrated that in contrast to current trends, small networks can also provide reasonably good results, and they do not need enormous amounts of training. Our given methods are robust in the sense that they were trained using a small amount of uncertain data, and this allows one to adapt them to new environments with a minimal burden on the valuable time of medical staff.

SUMMARY AND CONCLUSIONS

Technological and scientific progress experienced in the field of retina is unlike any other subfields of ophthalmology lately. Optical coherence tomography (OCT) reinterpreted the definition of imaging by providing an unprecedented resolution of ophthalmic structures. Tremendous amount of data arised both from this constantly developing imaging technology and the knowledge it generated. As the layers of the retina can be now visualized almost on a histological level, numerous details and features, typically appearing in retinal disorders, were discovered and described recently, namely the OCT biomarkers.

Albeit improvement shall support the work of the ophthalmologists, recognition, quantification and processing of all these details does not appear viable in the routine work of a physician. This antilogy could only be resolved in an interdisciplinary approach, combining the expertise of clinicians and modern computational tools used in the fields of image processing and artificial intelligence (AI) [10]. Applicability of AI has already been proven in several field of medicine, and presented clinically acceptable diagnostic performance in detecting many retinal diseases, such as diabetic retinopathy, glaucoma, retinopathy of prematurity and AMD [70].

Both the prevalence and the significance of age-related diseases show an increasing tendency due to longer life expectancy of populations [3]. Retinal disorders are among the main causes of severe vision loss and blindness worldwide, particularly age-related macular degeneration [4]. Because of the large prevalence of AMD, the proper management is crucial. With the help of OCT, several disease relevant biomarkers have been identified in AMD, but their exact recognition and manual quantification is not just time consuming, but also impractical. Fine changes of these biomarkers determine whether to treat a wet AMD patient, and considering the price of a single anti-VEGF injection, there is a need for a diversified and profound knowledge with an increasingly precise decision making in order to maximize the effects of the therapy, but also to reduce unnecessary costs.

The PhD thesis summarized the work of improving our comprehension of OCT biomarkers in wet AMD. First, outer retinal tubulations were investigated in a retrospective real-life study.

1.) Our study found no significant difference between the ranibizumab and aflibercept treated groups according to ORT development, meaning ORT is independent of the chosen anti-VEGF drug or the dosing regimen of intravitreal anti-VEGF treatment. To our best knowledge this was the first publication, which presented results in treatment naïve patients treated with aflibercept and its connection with ORT development.

2.) There was a statistically significant reduction in the frequency of injections in the pro re nata ranibizumab treated group before and after ORT appearance, meaning that the clinician can expect a decrease in the number of injections after ORT development. Our fixed bimonthly treatment with aflibercept did not allow us to analyze the injection rate before and after ORT development in this group. Our study was the first to report results according to injection rate.

3.) The presence of SHRM at treatment initiation as a biomarker had a statistically significant correlation with the development of ORT in both groups. When SHRM was present the chance of developing ORT was 2.75 and 11.14 higher in the ranibizumab and aflibercept groups, respectively. No previous reports were found, which presented results in treatment naïve patients treated with aflibercept in this topic.

In the second part of the thesis our collaboration with computer scientists was presented that resulted in novel algorithms, which can automatically identify and quantify OCT biomarkers (such as SRF, IRF, PED, SHRM, ORT, HF) using classical image processing and artificial intelligence. The developed toolset is essential for further investigations of medically relevant questions in relation with OCT.

4.) I acted as the key medical expert in the development of an OCT image analysis toolset. My contributions were: raising the problem; selecting the adequate OCT images; performing and supervising biomarker annotations required for training and validation; explaining the properties and thus the characteristic features of biomarkers; evaluation of the results from a medical point of view.

In further investigations we would like to use our biomarker identification algorithms on a bigger number of real-life wet AMD patients, to look for possible relationships and correlations in their appearance and behavior. More specifically we would like to study the fluid compartments (SRF/IRF) and HF changes with the precise help of AI, whether HF is really a more sensitive

marker than the fluid reappearance. With farther help of computer scientists, we try to develop decision-aiding mechanisms to classify the different subtypes of wet AMD and to help in treatment decisions. We also plan to extend the software with including other ophthalmologic conditions, such as diabetic macular oedema and retinal vein occlusion.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Dr. Rózsa Dégi, for her patient guidance, constant support ever since I met her as a 5th year old medical student. Without her I am not sure I would even choose ophthalmology, for what I am forever grateful.

I would like to express my gratitude to Dr. Edit Tóth Molnár, for giving me the opportunity to carry out my doctoral research in the Department of Ophthalmology at University of Szeged.

I am thankful for the amazing people I have the opportunity to work with in our image processing research group from the Department of Ophthalmology: Flóra Hadarits, Géza Thury and from the Department of Image Processing and Computer Graphics: László Varga, Melinda Katona, Tamás Grósz, Prof. László Nyúl and Prof. József Dombi.

Last but not the least, I would like to thank my wife, my parents and family for their support and encouragement without which I would not come this far.

REFERENCES

1. Klein R, Klein BE, Linton KL. Prevalence of age-related maculopathy. The Beaver Dam Eye Study. *Ophthalmology*. 1992 Jun;99(6):933-43.
2. Velez-Montoya R, Oliver SC, Olson JL, Fine SL, Mandava N, Quiroz-Mercado H. Current knowledge and trends in age-related macular degeneration: today's and future treatments. *Retina*. 2013 Sep;33(8):1487-502.
3. Jaul E, Barron J. Age-Related Diseases and Clinical and Public Health Implications for the 85 Years Old and Over Population. *Front Public Health*. 2017 Dec 11;5:335.
4. Jonas JB, Bourne RR, White RA, Flaxman SR, Keeffe J, Leasher J, Naidoo K, Pesudovs K, Price H, Wong TY, Resnikoff S, Taylor HR; Vision Loss Expert Group of the Global Burden of Disease Study. Visual impairment and blindness due to macular diseases globally: a systematic review and meta-analysis. *Am J Ophthalmol*. 2014 Oct;158(4):808-15.
5. Richer S, Stiles W, Statkute L, Pulido J, Frankowski J, Rudy D, Pei K, Tsipursky M, Nyland J. Double-masked, placebo-controlled, randomized trial of lutein and antioxidant supplementation in the intervention of atrophic age-related macular degeneration: the Veterans LAST study (Lutein Antioxidant Supplementation Trial). *Optometry*. 2004 Apr;75(4):216-30.
6. Jager RD, Mieler WF, Miller JW. Age-related macular degeneration. *N Engl J Med*. 2008 Jun 12;358(24):2606-17.
7. Rosenfeld PJ, Brown DM, Heier JS, Boyer DS, Kaiser PK, Chung CY, Kim RY; MARINA Study Group. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2006 Oct 5;355(14):1419-31.
8. Wykoff CC, Ou WC, Brown DM, Croft DE, Wang R, Payne JF, Clark WL, Abdelfattah NS, Sadda SR; TREX-AMD Study Group. Randomized Trial of Treat-and-Extend versus Monthly Dosing for Neovascular Age-Related Macular Degeneration: 2-Year Results of the TREX-AMD Study. *Ophthalmol Retina*. 2017 Jul-Aug;1(4):314-321.
9. Coscas G et al., Clinical features and natural history of AMD on OCT, in: G. Coscas, et al. (Eds.), *Optical Coherence Tomography in Age-Related Macular Degeneration*, Springer, Heidelberg, 2009, pp. 195–274.
10. Schmidt-Erfurth U, Waldstein SM. A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration. *Prog Retin Eye Res*. 2016 Jan;50:1-24.
11. Zweifel SA, Engelbert M, Laud K, Margolis R, Spaide RF, Freund KB. Outer retinal tubulation: a novel optical coherence tomography finding. *Arch Ophthalmol*. 2009;127(12):1596–1602.

12. Wolff B, Matet A, Vasseur V, Sahel JA, Mauget-Faÿsse M. En face OCT imaging for the diagnosis of outer retinal tubulations in age-related macular degeneration. *J Ophthalmol*. 2012;2012:542417.
13. Schaal KB, Freund KB, Litts KM, Zhang Y, Messinger JD, Curcio CA. Outer retinal tubulation in advanced age-related macular degeneration: optical coherence tomographic findings correspond to histology. *Retina*. 2015;35(7):1339–1350.
14. Litts KM, Messinger JD, Freund KB, Zhang Y, Curcio CA. Inner segment remodeling and mitochondrial translocation in cone photoreceptors in age-related macular degeneration with outer retinal tubulation. *Invest Ophthalmol Vis Sci*. 2015;56(4):2243–2253.
15. Litts KM, Messinger JD, Dellatorre K, Yannuzzi LA, Freund KB, Curcio CA. Clinicopathological correlation of outer retinal tubulation in age-related macular degeneration. *JAMA Ophthalmol*. 2015;133(5):609–612.
16. Litts KM, Wang X, Clark ME, Owsley C, Freund KB, Curcio CA, Zhang Y. Exploring photoreceptor reflectivity through multimodal imaging of outer retinal tubulation in advanced age-related macular degeneration. *Retina*. 2017;37(5):978–988.
17. Dolz-Marco R, Litts KM, Tan ACS, Freund KB, Curcio CA. The evolution of outer retinal tubulation, a neurodegeneration and gliosis prominent in macular diseases. *Ophthalmology*. 2017;124(9):1353–1367.
18. Hua R, Liu L, Hu Y, Chen L. The occurrence and progression of outer retinal tubulation in Chinese patients after intravitreal injections of ranibizumab. *Sci Rep*. 2015;7(5):7661.
19. Preti RC, Govetto A, Filho RGA, Cabral Zacharias L, Gianotti Pimentel S, Takahashi WY, Monteiro MLR, Hubschman JP, Sarraf D. Optical coherence tomography analysis of outer retinal tubulations: sequential evolution and pathophysiological insights. *Retina*. 2018;38(8):1518–1525.
20. Dirani A, Gianniou C, Marchionno L, Decugis D, Mantel I. Incidence of outer retinal tubulation in ranibizumab-treated age-related macular degeneration. *Retina*. 2015;35(6):1166–1172.
21. Lee JY, Folgar FA, Maguire MG, Ying GS, Toth CA, Martin DF, Jaffe GJ, CATT Research Group. Outer retinal tubulation in the comparison of age-related macular degeneration treatments trials (CATT). *Ophthalmology*. 2014;121(12):2423–2431.
22. Faria-Correia F, Barros-Pereira R, Queirós-Mendanha L, Fonseca S, Mendonça L, Falcão MS, Brandão E, Falcão-Reis F, Carneiro AM. Characterization of neovascular age-related macular degeneration patients with outer retinal tubulations. *Ophthalmologica*. 2013;229(3):147–151.
23. 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(5):389–414.

24. Shah VP, Shah SA, Mrejen S, Freund KB. Subretinal hyperreflective exudation associated with neovascular age-related macular degeneration. *Retina*. 2014;34(7):1281–1288.
25. Agresti A. *Categorical data analysis*. 2. Gainesville: Wiley; 2002.
26. Hariri A, Nittala MG, Sadda SR. Outer retinal tubulation as a predictor of the enlargement amount of geographic atrophy in age-related macular degeneration. *Ophthalmology*. 2015;122(2):407–413.
27. Goldberg NR, Greenberg JP, Laud K, Tsang S, Freund KB. Outer retinal tubulation in degenerative retinal disorders. *Retina*. 2013;33(9):1871–1876.
28. Iriyama A, Aihara Y, Yanagi Y. Outer retinal tubulation in inherited retinal degenerative disease. *Retina*. 2013;33(7):1462–1465.
29. Giachetti Filho RG, Zacharias LC, Monteiro TV, Preti RC, Pimentel SG. Prevalence of outer retinal tubulation in eyes with choroidal neovascularization. *Int J Retina Vitreous*. 2016;2(1):6.
30. Jung JJ, Freund KB. Long-term follow-up of outer retinal tubulation documented by eye-tracked and en face spectral-domain optical coherence tomography. *Arch Ophthalmol*. 2012;130(12):1618–1619.
31. Massamba N, Dirani A, Butel N, Fardeau C, Bodaghi B, Ingram A, Lehoang P. Evaluation of outer retinal tubulations in eyes switched from intravitreal ranibizumab to aflibercept for treatment of exudative age-related macular degeneration. *Graefes Arch Clin Exp Ophthalmol*. 2017;255(1):61–67.
32. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. *Prog Retin Eye Res*. 2018 Nov;67:1-29.
33. Kovacs A, Kiss T, Rarosi F, Somfai GM, Facsco A, Degi R. The effect of ranibizumab and aflibercept treatment on the prevalence of outer retinal tubulation and its influence on retreatment in neovascular age-related macular degeneration. *BMC Ophthalmol*. 2018 Nov 14;18(1):298.
34. Coscas G, De Benedetto U, Coscas F, Li Calzi CI, Vismara S, Roudot-Thoraval F, Bandello F, Souied E. Hyperreflective dots: a new spectral-domain optical coherence tomography entity for follow-up and prognosis in exudative age-related macular degeneration. *Ophthalmologica*. 2013;229(1):32-7.
35. Lee H, Ji B, Chung H, Kim HC. CORRELATION BETWEEN OPTICAL COHERENCE TOMOGRAPHIC HYPERREFLECTIVE FOCI AND VISUAL OUTCOMES AFTER ANTI-VEGF TREATMENT IN NEOVASCULAR AGE-RELATED MACULAR DEGENERATION AND POLYPOIDAL CHOROIDAL VASCULOPATHY. *Retina*. 2016 Mar;36(3):465-75.

36. Parodi MB, Arrigo A, Romano F, Aragona E, Marchese A, Cicinelli MV, Mercuri S, Bandello F. Hyperreflective Foci Number Correlates with Choroidal Neovascularization Activity in Angioid Streaks. *Invest Ophthalmol Vis Sci*. 2018 Jul 2;59(8):3314-3319.
37. Segal O, Barayev E, Nemet AY, Geffen N, Vainer I, Mimouni M. PROGNOSTIC VALUE OF HYPERREFLECTIVE FOCI IN NEOVASCULAR AGE-RELATED MACULAR DEGENERATION TREATED WITH BEVACIZUMAB. *Retina*. 2016 Nov;36(11):2175-2182.
38. Christenbury JG, Folgar FA, O'Connell RV, Chiu SJ, Farsiu S, Toth CA; Age-related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study Group. Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci. *Ophthalmology*. 2013 May;120(5):1038-45.
39. Korot E, Comer G, Steffens T, Antonetti DA. Algorithm for the Measure of Vitreous Hyperreflective Foci in Optical Coherence Tomographic Scans of Patients With Diabetic Macular Edema. *JAMA Ophthalmol*. 2016 Jan;134(1):15-20.
40. Lee H, Kang KE, Chung H, Kim HC. Automated Segmentation of Lesions Including Subretinal Hyperreflective Material in Neovascular Age-related Macular Degeneration. *Am J Ophthalmol*. 2018 Jul;191:64-75.
41. Abhishek AM, Berendschot TTJM, Rao SV, Dabir S. Segmentation and analysis of retinal layers (ILM & RPE) in Optical Coherence Tomography images with Edema. In *Biomedical Engineering and Sciences (IECBES)*, 2014 IEEE Conference on, pages 204–209, 2014.
42. Chiu SJ, Li XT, Nicholas P, Toth CA, Izatt JA, Farsiu S. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt Express*. 2010 Aug 30;18(18):19413-28.
43. Garvin MK, Abramoff MD, Wu X, Russell SR, Burns TL, Sonka M. Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans Med Imaging*. 2009 Sep;28(9):1436-47.
44. Lu S, Cheung CY, Liu J, Lim JH, Leung CK, Wong TY. Automated layer segmentation of optical coherence tomography images. *IEEE Trans Biomed Eng*. 2010 Oct;57(10):2605-8.
45. Yazdanpanah A, Hamarneh G, Smith B, Sarunic M. Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach, pages 649–656. Springer Berlin Heidelberg, 2009.
46. Lang A, Carass A, Hauser M, Sotirchos ES, Calabresi PA, Ying HS, Prince JL. Retinal layer segmentation of macular OCT images using boundary classification. *Biomed Opt Express*. 2013 Jun 14;4(7):1133-52.

47. Ngo L, Cha J, Han JH. Deep Neural Network Regression for Automated Retinal Layer Segmentation in Optical Coherence Tomography Images. *IEEE Trans Image Process*. 2019 Aug 1.
48. Dombi J. Modalities, pages 53–65. Springer Berlin Heidelberg, 2012.
49. Schafer RW. What Is a Savitzky-Golay Filter? *IEEE Signal Processing Magazine*, 28(4):111–117, 2011.
50. Dodo BI, Li Y, Eltayef K, Liu X. Graph-Cut Segmentation of Retinal Layers from OCT Images. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING*, pages 35–42, 2018.
51. Ben Salah M, Mitiche A, Ben Ayed I. Multiregion image segmentation by parametric kernel graph cuts. *IEEE Trans Image Process*. 2011 Feb;20(2):545-57.
52. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
53. Goldstein T, Bresson X, Osher S. Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction. *Journal of Scientific Computing*, 45(1):272–293, 2010.
54. Novosel J, Wang Z, de Jong H, van Velthoven M, Vermeer KA, van Vliet LJ. Locally-adaptive loosely-coupled level sets for retinal layer and fluid segmentation in subjects with central serous retinopathy. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 702–705, 2016.
55. Xiayu Xu, Kyungmoo Lee, Li Zhang, Sonka M, Abramoff MD. Stratified Sampling Voxel Classification for Segmentation of Intraretinal and Subretinal Fluid in Longitudinal Clinical OCT Data. *IEEE Trans Med Imaging*. 2015 Jul;34(7):1616-1623.
56. Wang J, Zhang M, Pechauer AD, Liu L, Hwang TS, Wilson DJ, Li D, Jia Y. Automated volumetric segmentation of retinal fluid on optical coherence tomography. *Biomed Opt Express*. 2016 Mar 30;7(4):1577-89.
57. Ding W, Young M, Bourgault S, Lee S, Albiani DA, Kirker AW, Forooghian F, Sarunic MV, Merkur AB, Beg MF. Automatic detection of subretinal fluid and sub-retinal pigment epithelium fluid in optical coherence tomography images. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7388– 7391, 2013.
58. Lee CS, Tying AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017 Jun 23;8(7):3440-3448.

59. Zheng Y, Sahni J, Campa C, Stangos AN, Raj A, Harding SP. Computerized assessment of intraretinal and subretinal fluid regions in spectral-domain optical coherence tomography images of the retina. *Am J Ophthalmol*. 2013 Feb;155(2):277-286.e1.
60. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7): 629–639, 1990.
61. Wieclawek W. Automatic cysts detection in optical coherence tomography images. In *Mixed Design of Integrated Circuits Systems (MIXDES)*, 2015 22nd International Conference, pages 79–82, 2015.
62. Wilkins GR, Houghton OM, Oldenburg AL. Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Trans Biomed Eng*. 2012 Apr;59(4):1109-14.
63. Xu Y, Yan K, Kim J, Wang X, Li C, Su L, Yu S, Xu X, Feng DD. Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy. *Biomed Opt Express*. 2017 Aug 10;8(9):4061-4076.
64. Beaudet P. Rotationally invariant image operators. In: *International Joint Conference on Pattern Recognition*, Kyoto, Japan. pp. 579–583 (1978)
65. Danielsson PE. Euclidean distance mapping. *Computer Graphics and Image Processing*, 14(3):227–248, 1980.
66. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*. MICCAI. Lecture Notes in Computer Science, vol 9351, Springer, Cham, 2015.
67. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.
68. Toth L, Kocsor A. Training HMM/ANN hybrid speech recognizers by probabilistic sampling, in: *Artificial Neural Networks: Biological Inspirations ICANN 2005*, LNCS, 3696, 2005, pp. 597–603.
69. Schlegl T, Bogunovic H, Klimesch S, Seebeck P, Sadeghipour A, Gerendas B, Waldstein SM, Langs G, Schmidt-Erfurth U. Fully automated segmentation of hyperreflective foci in optical coherence tomography images. *ArXiv:1805.03278*.
70. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer L, Keane PA, Wong TY. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019 Feb;103(2):167-175.

I.

RESEARCH ARTICLE

Open Access



The effect of ranibizumab and aflibercept treatment on the prevalence of outer retinal tubulation and its influence on retreatment in neovascular age-related macular degeneration

Attila Kovacs¹, Timea Kiss¹, Ferenc Rarosi², Gabor M. Somfai^{3,4}, Andrea Facsco¹ and Rozsa Degi^{1*}

Abstract

Background: We aimed to analyze the differences in the prevalence of outer retinal tubulation (ORT) in neovascular age-related macular degeneration (AMD) treated with anti-vascular endothelial growth factor (anti-VEGF) agents, either aflibercept or ranibizumab. Our further aim was to examine the changes in the frequency of injections of ranibizumab before and after ORT appearance.

Methods: Two hundred thirty six eyes of 230 patients were included in the study (184 eyes treated with ranibizumab by pro re nata regimen (PRN), 52 eyes with aflibercept bimonthly) and followed for 6–24 months. Using optical coherence tomography (OCT), the first appearance of ORT was documented, and fixed time point evaluations were also made every six months to determine the existence of ORT. The number of injections, the presence or absence of subretinal hyperreflective material (SHRM) at treatment initiation and visual acuity were also noted.

Results: The survival analysis with Cox proportional hazard model showed no significant difference between the ranibizumab and aflibercept groups in relation to the development of ORT ($p = 0.79$, hazard ratio 0.92). In the PRN treated ranibizumab group the number of injections showed significant decrease after ORT development ($p = 0.004$). When SHRM was present at treatment initiation the chance of developing ORT was 2.75 and 11.14 times higher in the ranibizumab and aflibercept groups, respectively.

Conclusions: The prevalence of ORT increased over time independently from the chosen anti-VEGF drug. Our results suggest that upon the appearance of ORT a decrease in retreatments can be expected.

Keywords: Outer retinal tubulation, Prevalence, Anti-VEGF, Aflibercept, Retreatment, Subretinal hyperreflective material

Background

Outer retinal tubulation (ORT) is a spectral-domain optical coherence tomography (SD-OCT) biomarker [1], first described by Zweifel et al. [2]. They defined ORTs as hyporeflective, branching tubular structures with hyperreflective borders within the outer nuclear layer of the retina [2]. The “en face” OCT technique can help map these

branching networks [3]. ORTs have been observed in many retinal diseases, including exudative age-related macular degeneration (AMD) [2]. Based on histological reports, the border of the outer retinal tubulation consists of photoreceptor inner segment mitochondria and external limiting membrane (ELM), with fluid and photoreceptor outer segments being potentially present in the ovoid hyporeflective lumen of the ORT [4–6]. Adaptive optics scanning laser ophthalmoscopy findings are in correlation with histology reports and show lack of ORT cone

* Correspondence: degirozsa57@gmail.com

¹Department of Ophthalmology, Faculty of Medicine, University of Szeged, 10-11 Koranyi fasor, Szeged 6720, Hungary

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

reflectivity which can be due to the loss of cone outer segments and subsequent retinal remodeling [7].

Schaal et al. classified outer retinal tubulations as either open (incomplete closure with curving external limiting membrane at the ends, horizontally elongated shape in cross-section) or closed (completely encircled, oval shape in cross-section) ORTs [4].

ORT can be mistaken for intraretinal cysts, or subretinal fluid but with the recognition of its hyperreflective border and special occurrence in the outer nuclear layer these mistakes can be reduced, leading to a reduction in the rate of anti-VEGF overtreatment in exudative AMD [2].

The ORT prevalence in exudative AMD is low at the time of first diagnosis but over time during anti-VEGF therapy its prevalence increases [8, 9]. The importance of ORT as an OCT biomarker for photoreceptor degeneration is due to its connection with reduced visual acuity [1, 8–10].

It has been also reported that ORTs develop above areas of subretinal hyperreflective material (SHRM) or atrophy [8, 9]. SHRM is a medium- to hyperreflective mass between the neurosensory retinal layers and retinal pigment epithelium on OCT [11]. It usually represents either a type II choroidal neovascular complex or is the consequence of an active choroidal neovascularisation, including subretinal haemorrhage and lipid or fluid exudation [1, 11, 12].

The aim of the present study was to investigate the prevalence of ORTs in eyes with neovascular AMD undergoing treatment either with ranibizumab or aflibercept. Our further aim was to examine the changes in the frequency of injections before and after ORT appearance. We also assessed the presence of subretinal hyperreflective material and its relationship with ORT.

Methods

Ethics, consent

This retrospective study was performed at the Medical Retina Unit of the Ophthalmology Department of University of Szeged, in Hungary. The study was approved by the Institutional Review Board of University of Szeged Albert Szent-Györgyi Clinical Centre (reference number: 3650) and was in accordance with the ethical standards of the Declaration of Helsinki. Since this was a retrospective review of patient data, informed consent was not required. The need for a consent was formally waived by the ethics committee, and this was also in line with the national regulations.

Patients

Treatment-naïve exudative AMD patients were enrolled in the study. For the ranibizumab group enrollment took place between October 2014 to April 2016 while patients in the aflibercept group were enrolled between April 2015 to April 2016.

All patients were over 50 years of age, the mean follow-up period was 16.3 months and 9.2 months (range 6–24 months and 6–12 months) in the ranibizumab and aflibercept groups, respectively.

During each visit a comprehensive ophthalmic examination was carried out including best-corrected visual acuity (BCVA, Early Treatment Diabetic Retinopathy Study (ETDRS) score) assessment, slit-lamp biomicroscopy, dilated funduscopy and SD-OCT examination of the retina (Heidelberg Spectralis, Heidelberg Engineering, Heidelberg, Germany). Eyes with poor quality SD-OCT scans (Q index below 20) or with poor compliance were excluded from the study (14 eyes from the ranibizumab and 3 eyes from the aflibercept group).

Treatment regimen for both ranibizumab (0.5 mg) and aflibercept (2 mg) started with 3 monthly injections. After this initiation phase the ranibizumab group was treated by a pro re nata (PRN) regimen with follow-up visits scheduled monthly. The retreatment criteria for ranibizumab patients consisted of any subretinal or intraretinal fluid on OCT, or new haemorrhage on funduscopy. In the aflibercept group follow-up after the loading phase was scheduled every two months, treatment was given at each follow-up. The above regimens were in accordance with the available treatment guidelines in Hungary at the time of the study.

For SD-OCT imaging a pattern size of 5.8×5.8 mm, $20^\circ \times 20^\circ$ was applied with 25 B-scans, using the “follow-up” mode. By manual review of the scan volumes we determined the first appearance of the ORT in both groups. We also assessed the presence of ORT at fixed time points at baseline, month 6 and 12 in both groups and at months 18 and 24 in the ranibizumab group. Images were assessed by two independent retina specialists, in case of incongruity the images were referred to a third retina specialist to make a decision. During the evaluation of OCT scans we did not differentiate between the above described open (incomplete hyperreflective ring) and closed (complete hyperreflective ring) forms of ORT according to Schaal [4]. Thus, both types of ORT detected on the images were considered an ORT positive case. The criterion of ORT was a hyperreflective ovoid-elongated structure in the outer nuclear layer of the retina with lower reflective content (Fig. 1).

The readers also identified the presence or absence of subretinal hyperreflective material on SD-OCT images at treatment initiation. The criterion for SHRM was a medium- to hyperreflective mass between the neurosensory retinal layers and retinal pigment epithelium, as described by Keane et al. [11] (Fig. 1).

Statistical methods

The BCVA was compared across the two groups using the Mann-Whitney U-test. The survival analysis for

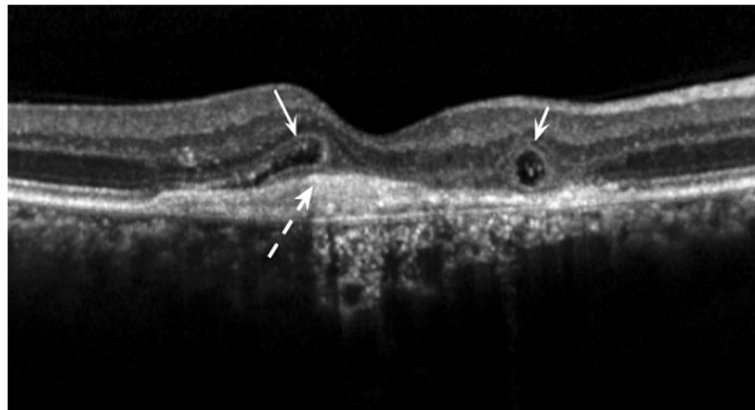


Fig. 1 Outer retinal tubulations and subretinal hyperreflective material on an SD-OCT single B-scan. Open and closed ORTs in cross section (left and right solid arrows) above subretinal hyperreflective material (dash arrow). The definition of ORT was a hyperreflective ovoid-elongated structure in the outer nuclear layer of the retina with lower reflective content

ORT development was analyzed by a Cox proportional hazard model. We analyzed the correlation between the presence of SHRM at treatment initiation with the development of ORT by Chi-square test and calculated relative risks. Where zeros were involved for the computation of relative risk, 0.5 was added to all cells, according to the previous suggestions [13].

The injection rate was calculated only in the PRN treated ranibizumab group due to the fixed 2 month therapeutic regimen of aflibercept. We assessed the injection rate only before and after the appearance of outer retinal tubulation and compared using the Mann-Whitney U-test. In order to correct bias due to the unequal follow-up time (some patients had a higher number of injections due to the longer follow-up), we divided the follow-up time with the number of injections and calculated with monthly injections.

A p -value of $p < 0.05$ was taken as statistically significant. For the analyses the IBM SPSS Software (Version 22) was used.

Results

In the ranibizumab group we evaluated 184 eyes of 179 patients, with a median age of 74 years (range 51 to 88), while in the aflibercept group there were 52 eyes of 51 patients with a median age of 75 years (range 58 to 87).

The mean baseline best corrected visual acuities in the two groups were (mean \pm SD) 59.16 ± 13.9 (median 61) and 53.96 ± 13.54 (median 55.5) ETDRS letters in the ranibizumab and aflibercept group, respectively. There was no significant difference between the two groups (Mann-Whitney U-test $p = 0.083$). The BCVA at the end of the follow-up was 57.19 ± 20.19 (median 63) and 59.46 ± 15.54 (median 64) ETDRS letters in the ranibizumab and aflibercept group, respectively. There

was no significant difference between the two groups (Mann-Whitney U-test $p = 0.69$).

Table 1 shows the number of eyes during the follow-up in the two groups. The number of eyes was reduced over time due to gradual enrollment in the study, thereby not every patient reached the same follow-up time. In the ranibizumab group outer retinal tubulation was observed in 17.4% of cases at baseline, in 33.7% of cases at month 6, in 45.3% of cases at month 12, and in 55.3% and in 60.8% of cases at months 18 and 24, respectively. The ORT prevalence in the aflibercept group was 23.1% at baseline, 40.4% at month 6, and 50% at month 12.

The survival analysis showed no significant difference between the ranibizumab and aflibercept treated groups in terms of ORT development. ($p = 0.79$, hazard ratio 0.92, 95% confidence interval 0.500–1.693) (Fig. 2).

The injection rates showed that the mean injection number per month before the ORT appearance was 0.37 ± 0.17 while after the ORT development it decreased to 0.21 ± 0.17 (Mann-Whitney U-test $p = 0.004$).

The presence of subretinal hyperreflective material at treatment initiation in the two subgroups was 75.5% in the ranibizumab, and 80.8% in the aflibercept group. In

Table 1 Number of eyes reaching the follow-up in the ranibizumab and aflibercept treated groups

	Ranibizumab <i>n</i> (eyes)	Aflibercept <i>n</i> (eyes)
Time point		
Baseline	184	52
at 6 months	184	52
at 12 months	161	28
at 18 months	103	0
at 24 months	51	0

Table legend: The column with “n” corresponds to the number of eyes reaching the follow-up

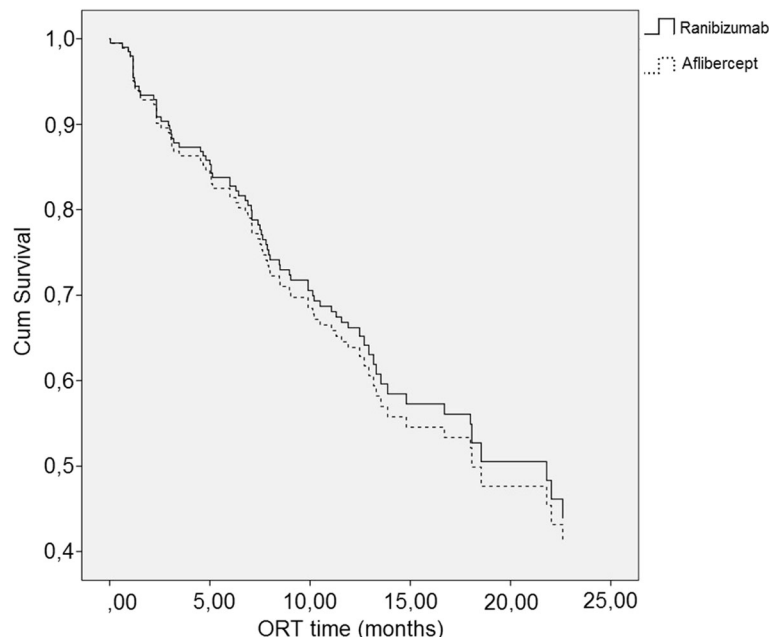


Fig. 2 Cox proportional hazard model of ORT survival in the ranibizumab and aflibercept groups

the ranibizumab treated group ORT developed in 85 eyes of 139 eyes with SHRM (61.15%), while without SHRM (45 eyes) ORTs were found merely in 10 eyes (22.2%) corresponding to a relative risk of 2.75. ($p < 0.01$). In the aflibercept treated group 55.81% of eyes with SHRM developed ORT (24 eyes of 43). No ORT developed in the eyes without SHRM (out of 9 eyes), consistent with a relative risk of 11.14 ($p < 0.01$).

Discussion

Our study found no significant difference between the ranibizumab and aflibercept treated groups according to ORT development. There was a statistically significant reduction in the frequency of injections in the ranibizumab treated group before and after ORT appearance. The presence of SHRM at treatment initiation as a biomarker had a statistically significant correlation with the development of ORT in both groups.

Zweifel first described outer retinal tubulation in 2009 [2]. Later the authors reported ORTs in various retinal disorders like neovascular AMD, geographic atrophy, polypoidal choroidal vasculopathy, non-AMD associated choroidal neovascularisation, and other degenerative retinal disorders (e.g. retinitis pigmentosa, gyrate atrophy, choroideremia, Stargardt disease, pattern dystrophy) [2, 3, 14–19].

The pathogenesis of ORT formation is still not completely clear, though recent reports helped a lot exploring it. Dolz-Marco et al. called the attention on the role of Müller cells in the pathomechanism of ORT development, namely the progressive photoreceptor damage

that can result in Müller cell activation which thereby starts to produce glial fibrillary acidic protein, facilitating the formation of ORT [20]. Based on histological examinations it seems that the evolution of ORT starts with ELM and ellipsoid zone disruption [4]. ELM starts to scroll inward at its free edges, representing an initial form of ORT, leading to the development of a formed open ORT. With time the large, open ORTs split, their margins beginning to scroll ending in multiple smaller closed ORTs. During the scrolling process a downward displacement of adjacent inner nuclear layer and outer plexiform layer happens separating each ORT, and causing the appearance of microcystic lesions in the inner nuclear layer. The downward displacement of these layers might be due to the involvement of Müller cells in this scrolling/dragging process as Müller cells are contributing to the constitution of ELM with the inner segments of photoreceptors [4, 15, 20, 21].

Most of the studies in the field focused on exudative AMD and its relationship with ORT. From these reports it is known that ORT is an SD-OCT biomarker, the prevalence of which increases with time and is associated with decreased visual acuity [1, 8, 9]. It has been also reported that ORTs develop adjacent to areas of subretinal hyperreflective material or atrophy [8, 9]. The differential diagnostic importance of outer retinal tubulation lies in the avoidance of overtreatment due to its similarity to intraretinal fluid [2, 8].

In the present study, we evaluated treatment-naïve exudative AMD patients treated with ranibizumab and

aflibercept regarding the presence of outer retinal tubulation. Altogether 236 eyes were followed in both groups with no statistical difference between the baseline characteristics of the two groups considering age and BCVA.

The prevalence of ORT continuously increased during the follow-up period, in both groups. In the ranibizumab group its prevalence almost quadrupled at the 24-month follow-up, while there was a doubling in the aflibercept group in 12 months. It is important to note, that in the ranibizumab group the baseline prevalence was lower (17.4% versus 23.1%). Dirani et al. found a similar increasing trend in their study starting with 2.5% at presentation, reaching 41.6% at 4 years of follow-up [8]. In our study the baseline ORT prevalence was higher compared to other reports [8, 9]. The reason behind this could be the more advanced disease state at the time of presentation (due to the real life nature of our retrospective study) and a relative delay in therapy initiation due to country-specific financial difficulties. The poorer baseline BCVA in both groups also supports this idea.

There is one previous article known, describing ORT development in 24 non-treatment-naïve eyes, treated with aflibercept only after receiving at least 6 ranibizumab injection, reporting an initial 97% ORT prevalence which later decreased to 75% [22].

To our knowledge, our study is the first to report results in treatment naïve patients treated with aflibercept and its connection with ORT development. The Cox proportional hazard model analysis suggested that there was no difference between the two in-label therapies ranibizumab and aflibercept in regard to the prevalence of outer retinal tubulation. Lee et al. in the Comparison of AMD Treatment Trials (CATT) study group evaluated the prevalence of ORTs in ranibizumab and bevacizumab treated neovascular AMD patients, and found no difference between the two drugs related to the prevalence of outer retinal tubulation [9].

In the present study we found a statistically significant difference in the monthly injection rate before and after the appearance of outer retinal tubulation in the ranibizumab treated group. Our results suggest that in patients who develop ORT a decrease in the retreatment rate can be expected which may be a very important clinical marker. Although we had a PRN regimen according to the Hungarian guidelines, Lee et al. found no difference between the fixed monthly regimen versus PRN regimen in regards to ORT development in patients treated either with ranibizumab or bevacizumab [9].

We found a statistically significant connection between ORT development and the presence of subretinal hyperreflective material at treatment initiation. When subretinal hyperreflective material was present the chance of developing ORT was 2.75., and 11.14 higher in the ranibizumab and aflibercept groups, respectively, in

accordance with the results of Lee et al. in ranibizumab and bevacizumab treated patients [9].

Our findings, in concordance with the above mentioned study results suggest that ORT is independent of the chosen anti-VEGF drug or the dosing regimen of intravitreal anti-VEGF treatment. The appearance of ORT suggests that the clinicians can expect a decrease in the number of injections when following a pro re nata ranibizumab regimen. Our fixed bimonthly treatment with aflibercept did not allow us to analyze the injection rate before and after ORT development in this group. Our study also supports the previously reported higher prevalence of ORT development in the presence of subretinal hyperreflective material at treatment initiation [9].

There is a number of limitations of our study. Namely, the relatively small sample size in the aflibercept group compared to the ranibizumab group, along with the bimonthly follow-up in the aflibercept group. We believe that the number of subjects involved in both groups is comparable with other studies published in the field, while the bimonthly treatment regimen with aflibercept was fixed due to the country-specific guideline regulations. The decreasing number of eyes during the follow-up could also bias the analysis by including patients with increasing disease severity. However, we believe these factors were similar to those in similar studies available in the field. The strength of this report is the comparison of present in-label therapies, ranibizumab and aflibercept in exudative AMD patients in relation to ORT besides the evaluation of the injection rate in association with outer retinal tubulation. We used real life data that makes the study more relevant in the daily clinical practice.

Conclusions

The development of ORT could be a potential biomarker for the treatment prognosis in patients with wet AMD. Its clinical significance lies also in its similarity to activity-related intraretinal fluid. Further studies are needed to explore the nature and development of ORTs employing a comparable dosing and follow-up regimen of all three currently available anti-VEGF drugs ranibizumab, bevacizumab and aflibercept.

Abbreviations

AMD: Age-related macular degeneration; anti-VEGF: Anti-vascular endothelial growth factor; BCVA: Best-corrected visual acuity; CATT: Comparison of AMD Treatment Trials; CNV: Choroidal neovascularisation; ELM: External limiting membrane; ETDRS: Early Treatment Diabetic Retinopathy Study; OCT: Optical coherence tomography; ORT: Outer retinal tubulation; PRN: Pro re nata; SD-OCT: Spectral-domain optical coherence tomography; SHRM: Subretinal hyperreflective material

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Funding

No funding was received for this research.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Authors' contributions

AK, RD made the conception and design of the study and analyzed the OCT images, in case of incongruity the final decision was made by AF. AK and TK collected and arranged the data (demographics, visual acuity, injection number and date). FR made the statistical analyzes and also the interpretation of data together with GMS. AK, TK drafted the manuscript. GMS was a major contributor editing it. AF, GMS and RD revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was approved by the the Human Investigation Review Board of the University of Szeged Albert Szent-Györgyi Clinical Centre (reference number: 3650) and was performed in accordance with the tenets of the Declaration of Helsinki. In the study we used a retrospective review of patient data; therefore consent was not collected from participants. The need for a consent was formally waived by the ethics committee, and this was also in line with the national regulations (a reference to the relevant legislation: 23/2002. (V. 9.) Eüm 20/Q. §).

Consent for publication

The attached OCT image is entirely unidentifiable and there are no details on individuals reported within the manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Ophthalmology, Faculty of Medicine, University of Szeged, 10-11 Koranyi fasor, Szeged 6720, Hungary. ²Department of Medical Physics and Informatics, Faculty of Medicine, University of Szeged, Szeged, Hungary. ³Augenzentrum Pallas Kliniken, Olten, Switzerland. ⁴Department of Ophthalmology, Faculty of Medicine, Semmelweis University, Budapest, Hungary.

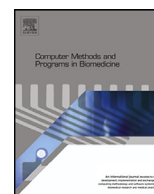
Received: 2 November 2017 Accepted: 26 October 2018

Published online: 14 November 2018

References

- Schmidt-Erfurth U, Waldstein SM. A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration. *Prog Retin Eye Res*. 2016;50:1–24. <https://doi.org/10.1016/j.preteyeres.2015.07.007>.
- Zweifel SA, Engelbert M, Laud K, Margolis R, Spaide RF, Freund KB. Outer retinal tubulation: a novel optical coherence tomography finding. *Arch Ophthalmol*. 2009;127(12):1596–602. <https://doi.org/10.1001/archophthol.2009.326>.
- Wolff B, Matet A, Vasseur V, Sahel JA, Mauget-Fajssse M. En face OCT imaging for the diagnosis of outer retinal tubulations in age-related macular degeneration. *J Ophthalmol*. 2012;2012:542417. <https://doi.org/10.1155/2012/542417>.
- Schaal KB, Freund KB, Litts KM, Zhang Y, Messinger JD, Curcio CA. Outer retinal tubulation in advanced age-related macular degeneration: optical coherence tomographic findings correspond to histology. *Retina*. 2015;35(7):1339–50. <https://doi.org/10.1097/IAE.0000000000000471>.
- Litts KM, Messinger JD, Freund KB, Zhang Y, Curcio CA. Inner segment remodeling and mitochondrial translocation in cone photoreceptors in age-related macular degeneration with outer retinal Tubulation. *Invest Ophthalmol Vis Sci*. 2015;56(4):2243–53. <https://doi.org/10.1167/iov.14-15838>.
- Litts KM, Messinger JD, Dellatorre K, Yannuzzi LA, Freund KB, Curcio CA. Clinicopathological correlation of outer retinal tubulation in age-related macular degeneration. *JAMA Ophthalmol*. 2015;133(5):609–12. <https://doi.org/10.1001/jamaophthol.2015.126>.
- Litts KM, Wang X, Clark ME, Owsley C, Freund KB, Curcio CA, Zhang Y. Exploring photoreceptor reflectivity through multimodal imaging of outer retinal tubulation in advanced age-related macular degeneration. *Retina*. 2017;37(5):978–88. <https://doi.org/10.1097/IAE.0000000000001265>.
- Dirani A, Gianniou C, Marchionno L, Decugis D, Mantel I. Incidence of outer retinal tubulation in ranibizumab-treated age-related macular degeneration. *Retina*. 2015;35(6):1166–72. <https://doi.org/10.1097/IAE.0000000000000439>.
- Lee JY, Folgar FA, Maguire MG, Ying GS, Toth CA, Martin DF, Jaffe GJ, CATT Research Group. Outer retinal tubulation in the comparison of age-related macular degeneration treatments trials (CATT). *Ophthalmology*. 2014;121(12):2423–31. <https://doi.org/10.1016/j.ophtha.2014.06.013>.
- Faria-Correia F, Barros-Pereira R, Queirós-Mendanha L, Fonseca S, Mendonça L, Falcão MS, Brandão E, Falcão-Reis F, Carneiro AM. Characterization of neovascular age-related macular degeneration patients with outer retinal tubulations. *Ophthalmologica*. 2013;229(3):147–51. <https://doi.org/10.1159/000346854>.
- 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(5):389–414. <https://doi.org/10.1016/j.survophthal.2012.01.006>.
- Shah VP, Shah SA, Mrejen S, Freund KB. Subretinal hyperreflective exudation associated with neovascular age-related macular degeneration. *Retina*. 2014;34(7):1281–8. <https://doi.org/10.1097/IAE.0000000000000166>.
- Agresti A. *Categorical data analysis*. 2nd ed. Gainesville: Wiley; 2002.
- Hariri A, Nittala MG, Sadda SR. Outer retinal tubulation as a predictor of the enlargement amount of geographic atrophy in age-related macular degeneration. *Ophthalmology*. 2015;122(2):407–13. <https://doi.org/10.1016/j.ophtha.2014.08.035>.
- Hua R, Liu L, Hu Y, Chen L. The occurrence and progression of outer retinal tubulation in Chinese patients after intravitreal injections of ranibizumab. *Sci Rep*. 2015;7(5):7661. <https://doi.org/10.1038/srep07661>.
- Goldberg NR, Greenberg JP, Laud K, Tsang S, Freund KB. Outer retinal tubulation in degenerative retinal disorders. *Retina*. 2013;33(9):1871–6. <https://doi.org/10.1097/IAE.0b013e318296b12f>.
- Iriyama A, Aihara Y, Yanagi Y. Outer retinal tubulation in inherited retinal degenerative disease. *Retina*. 2013;33(7):1462–5. <https://doi.org/10.1097/IAE.0b013e31828221ae>.
- Giachetti Filho RG, Zacharias LC, Monteiro TV, Preti RC, Pimentel SG. Prevalence of outer retinal tubulation in eyes with choroidal neovascularization. *Int J Retina Vitreous*. 2016;2(1):6. <https://doi.org/10.1186/s40942-016-0029-8>.
- Jung JJ, Freund KB. Long-term follow-up of outer retinal tubulation documented by eye-tracked and en face spectral-domain optical coherence tomography. *Arch Ophthalmol*. 2012;130(12):1618–9. <https://doi.org/10.1001/archophthol.2012.1902>.
- Dolz-Marco R, Litts KM, Tan ACS, Freund KB, Curcio CA. The evolution of outer retinal tubulation, a neurodegeneration and gliosis prominent in macular diseases. *Ophthalmology*. 2017;124(9):1353–67. <https://doi.org/10.1016/j.ophtha.2017.03.043>.
- Preti RC, Govetto A, Filho RGA, Cabral Zacharias L, Gianotti Pimentel S, Takahashi WY, Monteiro MLR, Hubschman JP, Sarraf D. Optical coherence tomography analysis of outer retinal tubulations: sequential evolution and pathophysiological insights. *Retina*. 2018;38(8):1518–25. <https://doi.org/10.1097/IAE.0000000000001810>.
- Massamba N, Dirani A, Butel N, Fardeau C, Bodaghi B, Ingram A, Lehoang P. Evaluation of outer retinal tubulations in eyes switched from intravitreal ranibizumab to aflibercept for treatment of exudative age-related macular degeneration. *Graefes Arch Clin Exp Ophthalmol*. 2017;255(1):61–7. <https://doi.org/10.1007/s00417-016-3423-x>.

III.



Automatic segmentation of hyperreflective foci in OCT images

László Varga*, Attila Kovács, Tamás Grósz, Géza Thury, Flora Hadarits, Rózse Dégi, József Dombi

University of Szeged, Interdisciplinary Excellence Centre, Hungary

ARTICLE INFO

Article history:

Received 19 December 2018

Revised 26 May 2019

Accepted 16 June 2019

Keywords:

OCT

HF segmentation

Deep neural network

Convolutional network

Image processing

GPGPU

ABSTRACT

Background and Objective: The leading cause of vision loss in the Western World is Age-related Macular Degeneration (AMD), but together with modern medicines, tracking the number of Hyperreflective Foci (HF) on Optical Coherence Tomography (OCT) images should assist the treatment of patients. Here, we developed a framework based on deep learning for the automatic segmentation of HF in OCT images.

Methods: We collected OCT images and annotated them, then these images underwent image preprocessing, and feature extraction steps. Using the prepared data we trained different types of Conventional-, Deep- and Convolutional Neural Networks to perform the task of the automatic segmentation of HF.

Results: We evaluated the various Neural Networks, by performing HF segmentation of clinical data belonging to patients, whose data were excluded from the training process. The results suggest that our systems can achieve reasonably high Dice Coefficient values, and they are comparable with (i.e., in most cases above 95%) the similarity between manual annotations performed by different physicians.

Conclusion: From the results, it can be concluded that neural networks can be used to accurately segment HF in OCT images. The results are sufficiently accurate for us to incorporate them into the next phase of the research, building a decision support system for everyday clinical practice.

© 2019 Published by Elsevier B.V.

1. Introduction

The importance of artificial intelligence (AI) and Deep learning have already been proven in many medical fields such as radiology, pathology, and dermatology. With the development of Machine Learning (ML), a branch of AI, it is now possible to automatically recognize anatomical structures or lesions in medical images (segmentation); to place an image into different categories (classification); and also to predict the outcome of a process (prediction). In ophthalmology, thanks to 21st century diagnostic innovations, the Optical Coherence Tomography (OCT), retinal structures can be visualized with an unprecedented resolution. With these high resolution OCT scans, we can now develop deep learning methods and utilize machine learning algorithms in order to reduce diagnostic and therapeutic errors, and promote personalized medicine [1]. ML has already provided clinically acceptable diagnostic performance in detecting many retinal diseases, such as diabetic retinopathy (DR), glaucoma, retinopathy of prematurity (ROP) and Age-related Macular Degeneration (AMD) [2].

The global increase in life expectancy has led to an increase in the number of age-related diseases, and AMD has become the leading cause of vision loss in the Western World [3], and a health problem worldwide [4]. Because of the large prevalence of AMD, the proper management of the disease is crucial.

With the help of OCT, several disease relevant biomarkers have been recently identified in AMD, such as Hyperreflective Foci (HF), which could provide a sensitive marker for the treatment decision process. However their manual quantification is quite challenging [5,6]. Therefore, we addressed the task in an interdisciplinary fashion, combining the expertise of clinicians with modern computational tools taken from the fields of image processing and artificial intelligence.

In our study, a framework was developed where the ophthalmologists could mark HF on OCT scans. Then, the marked images were transformed using various image processing techniques in order to extract features which best characterise HF. Quite recently, Deep Neural Networks have achieved excellent results in medical data analysis [7–9], and in the next step, image data sets were used for training Artificial Neural Networks (ANN), Deep Neural Network (DNN), and Convolutional Neural Networks (CNN). These networks were later used in the automatic detection of HF. Our methods were also validated on clinical data, gathered from

* Corresponding author.

E-mail address: vargalg@inf.u-szeged.hu (L. Varga).

patients of the Department of Ophthalmology at the Clinical Center at the University of Szeged.

A full summary of related scientific work can be found in Section 2. Previous research on the topic mainly focused on the automated evaluation of HF in manually annotated datasets [10–15] without automated evaluation, while our study focuses on the automatic segmentation of HF. One parallel study on this topic is available [16] on arxiv, where the authors applied deep learning-based methods using variants of the U-Net [9] for the automatic segmentation of HF. The approach described by our group also differs from the previous study involving training nets from two parallel sets of uncertain data. Moreover, we performed two independent evaluations of our models by comparing the network outputs with the annotations of two clinical experts.

In another study, the authors of [17] applied similar methods to ours for the quantization of HF in OCT images, but they did not perform a pixel-wise evaluation of the results. In their study, they used a smaller test dataset, and now we propose new methods that perform better.

The novelty of our study lies in testing various neural networks for the automatic segmentation of HF. Unlike previous studies, neural networks were trained not by only relying on raw pixel information, but also on features extracted by image preprocessing steps. We compared the annotations of two different clinical doctors with each other to determine a baseline accuracy for HF segmentation as a reference for our methods. We evaluated our methods on two test datasets, and we found that our methods led to Dice Coefficients values, which are comparable with the similarity score between manual annotations performed by physicians. Lastly, we demonstrated that in contrast to current trends, small networks can also provide reasonably good results, and they do not need enormous amounts of training. Our given methods are robust in the sense that they were trained using a small amount of uncertain data, and this allows one to adapt them to new environments with a minimal burden on the valuable time of medical staff.

2. Related work and background

When considering the task of HF segmentation, one has to incorporate knowledge from medical science and computer science. In this section, we give an overview of the related work in the literature and the technological background of our results.

2.1. Medical context

AMD can be divided into two categories, namely the dry (non-exudative) form that accounts for around 90% of all cases and the wet (exudative, neovascular) form that accounts for the remaining, but the more severe and rapid cases of vision loss [4,18].

Although the exact pathogenesis of the disease is not yet fully understood, vascular endothelial growth factor (VEGF) plays a crucial role in the development of the more dangerous wet form. There is an imbalance between proangiogenic and antiangiogenic factors, which lead to the overproduction of VEGF. Owing to VEGF, endothelial cell proliferation occurs and this forms a choroidal neovascular membrane (CNV). From the developed CNV, fluid leakage leads to the accumulation of sub- or intraretinal fluid, or RPE detachment, but bleeding can also occur and lead to advanced and acute sight loss [4].

The first choice of treatment for neovascular AMD is anti-VEGF intravitreal injections, which has revolutionized the treatment of wet AMD [19]. Since wet AMD has a chronic course with recurrences of fluid accumulation, repeated anti-VEGF injections and rigorous patient follow-up are required in order to prevent the deterioration of visual acuity. As the result of the innovations of the last decade, OCT has become the most useful device for an

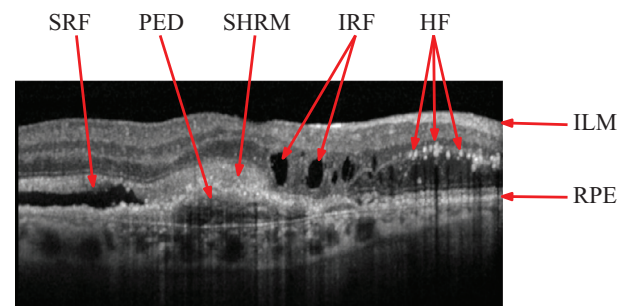


Fig. 1. Example of an OCT image. The Internal Limiting Membrane (ILM) and the Retinal Pigment Epithelium (RPE) are marked together with biomarkers of AMD such as: subretinal/intraretinal fluid accumulation (SRF/IRF); pigment epithelial detachment (PED); subretinal hyperreflective material (SHRM); outer retinal tubulation (ORT); and hyperreflective foci (HF).

ophthalmologist when diagnosing and monitoring AMD and its response to anti-VEGF treatment. With the help of OCT, it became possible to detect the layers of the retina and the effects of the disease, the so-called OCT biomarkers, such as subretinal/intraretinal fluid accumulation (SRF/IRF), pigment epithelial detachment (PED), subretinal hyperreflective material (SHRM), outer retinal tubulation (ORT) (see, e.g., Fig. 1). Assessment of these markers is the key to make the adequate decision for treating, re-treating or observing patients [5,6].

In 2009, Coscas et al. first reported a new OCT finding in exudative AMD, the Hyper-Reflective Dots (HRD) [5], also known as HF (Fig. 1). Since then HF were identified in additional retinal diseases, including diabetic macular oedema and retinal vein occlusion, [20,21]; but their exact pathomorphological origin remains controversial, since no histological data is available in the current literature. Reports demonstrated that HF have a rapid response to anti-VEGF therapy, and were also found to be the first detectable change for each clinical recurrence, even before fluid accumulation. In wet AMD the main criteria of the retreatment process is based on the recurrence of fluid (SRF, IRF), detected on OCT scans. The determination of the number of HF, can be a potentially more sensitive biomarker than the fluid reappearance, which may facilitate earlier decision making, and achieve greater visual acuity protection [1,10–15].

The objective quantization of the HF is extremely hard and time consuming to carry out manually, but recording slight changes in the amount of HF may be a crucial part of treatment planning. In the above-mentioned studies counting was conducted manually often using a single B-scan of the retina, and automated software can help reduce human error [10–15]. Therefore, our aim was to develop a system for the automatic segmentation and quantification of Hyperreflective Foci of wet AMD patients, which would provide a method for the objective tracking of HF on OCT scans.

2.2. Mathematical context

Artificial neural networks are based on the computational model of biological neural networks, their key component being the artificial neuron, also called a perceptron. Each artificial neuron works in the following way: once the input values have been given to them, they calculate a weighted sum of their inputs, then they transform the value with an activation function. The weights associated with the inputs are the learnable parameters of the neuron, and they determine the networks performance.

Deep learning techniques are widely used in image- [22–24] and medical-image processing [7–9,25,26] to solve a variety of problems.

2.2.1. Artificial neural networks

In Artificial Neural Networks, the neurons are organized into layers. Traditionally, the task of the hidden layer is to extract useful information from the input, while the output layer combines knowledge of the hidden layers to make a prediction. By training the weights of the neurons (e.g., with the stochastic gradient descent algorithm) one can create networks which can perform a variety of tasks.

In most cases, one single layer of hidden neurons cannot handle the complexity of harder tasks. This is why Deep Neural Networks were invented, which have many hidden layers, each layer receiving its input from a previous hidden layer. These networks can handle much more complex tasks, but to properly train them we need to modify the training algorithm or the network itself.

2.2.2. Deep rectifier neural networks

Deep Rectifier Neural Networks (DRNs) were proposed by Glorot et al. [27] as one of the first deep learning methods, which do not require any tedious pre-training steps to produce good results. DRNs modify just the hidden neurons in the network by using rectified linear units (ReLU). These rectified units differ from standard neurons only in their activation function, namely the rectifier activation function ($\max(0, x)$).

The rectifier function has two main properties, these being its hard saturation at 0 and its linear behaviour for positive input. Due to the linear behavior of the active units, there is no “vanishing gradient” effect [27]. This linear behaviour also means that the computational cost will be smaller, as there is no need to compute the exponential function when evaluating the activation function, and the sparsity of the outputs can also be exploited. Unfortunately, there is a difficulty arising from this linearity, namely the “exploding gradient” effect, which means that the gradients can grow without limit. This effect could be readily alleviated by applying regularization techniques such as L_2 weight normalization [28]. The hard saturation for negative inputs means that only a subset of neurons will be active in each hidden layer. In theory, this could harm optimization by blocking gradient backpropagation, but experimental results do not support this, and they tell us that DRNs can be trained as long as the gradient can propagate along some path [27].

2.2.3. Convolutional networks

The revolution of deep learning in image processing came with the development of convolutional networks. Convolutional networks are processing tools designed especially for image and signal processing. Despite the similarity to neural networks, convolutional layers are introduced where classical neurons are replaced by a set of convolutional kernels. CNNs can have a variety of structures. The first were classification nets, and they typically use a few convolutional layers followed by fully connected layers, and they predict class labels for a whole image.

One of the first widely used convolutional network was called AlexNet, which is a structure designed for image classification, and it won the first prize of the 2012 ImageNet ILSVRC. Since then, many new structures have evolved for image processing such as the ResNet [23] and VGG16 [24], which are highly accurate and versatile tools for image classification.

The next big revolution in image processing came with the invention of fully convolutional neural networks. These networks contain only convolutional and local image processing steps and their output is an image, which is usually the same size as the input image. Using fully convolutional networks direct segmentations of images can be carried out, which perform better, and yield higher accuracy scores, than when using classification networks for image segmentation tasks. Here one of the most commonly used networks is the U-Net [9] structure. Variations of this networks

have been applied in various medical image segmentation tasks [26,29].

2.2.4. Probabilistic sampling

It is well known that if one of the classes is significantly over-represented during training, it might bias the Neural Network and favor the most frequent class on its output. In our case only a few pixels formed part of the HF, so this problem created a serious issue during training. To overcome this problem, it is necessary to balance the class distribution by presenting more examples taken from the rarer classes to the model. Usually, we have no way of generating additional samples from a rare class, so balancing can be achieved by either reducing the number of the examples belonging to the most common classes (downsampling) or by presenting the rare examples more frequently (upsampling).

Probabilistic sampling [30] offers a third option where one combines the two previous sampling methods at the same time. It applies a simple two-step sampling scheme; namely, we first select a class, then pick a training sample belonging to this class. Before the first step, we assign a probability to each class, which will determine how frequently it will be selected:

$$P(c_k) = \lambda \frac{1}{K} + (1 - \lambda) \text{Prior}(c_k), \quad (1)$$

where $\text{Prior}(c_k)$ is the prior possibility of class c_k , K is the number of classes and $\lambda \in [0, 1]$ is a parameter. If λ is 1, then we get a uniform distribution over the classes; and with $\lambda = 0$ we get the original class distribution. Using intermediate λ values leads to a linear combination of these two distributions.

3. Methods

When constructing our automatic segmentation framework, the processing of data was divided into consecutive stages. These were data preparation, feature extraction, training, and evaluation performed on the different models. It should be added that 8 different methods were tested for the segmentation, which required different pre- and post-processing steps and they have different data flowcharts. The flow chart of the data preparation can be seen in Fig. 2, while the flowcharts for the segmentation methods are depicted in Fig. 3.

3.1. Data preparation

The first step of the process was to gather the clinical data for the project.

Of course, this study was carried out in accordance with the principles of the Declaration of Helsinki, and it was performed with the ethical approval of the Institutional Review Board of University of Szeged, Albert Szent-Györgyi Clinical Center. Here, we used two sets of OCT data for our study.

The first data set was used for training and testing the methods, and for selecting the best network structure. The data of 16 eyes of 16 patients, diagnosed with wet AMD were retrospectively selected at the Department of Ophthalmology, University of Szeged. These patients were either treatment-naïve or treated with anti-VEGF and they had biomarkers on the OCT scans including HF, SRF, PED, SHRM and IRF. No other retinal diseases were diagnosed among them. SD-OCT volume scans with a quality score above 16 were acquired using a Spectralis OCT scanner (Heidelberg Engineering, Heidelberg, Germany, version 6.5.2.0).

1 SD-OCT sequence from 11 patients, 2 consecutive SD-OCT sequences from 3 patients, and 3 consecutive SD-OCT sequences from 2 patients were collected, hence altogether 23 SD-OCT sequences were used in the study. In the case of patients with multiple captures, the timespan between each recording was at least one month.

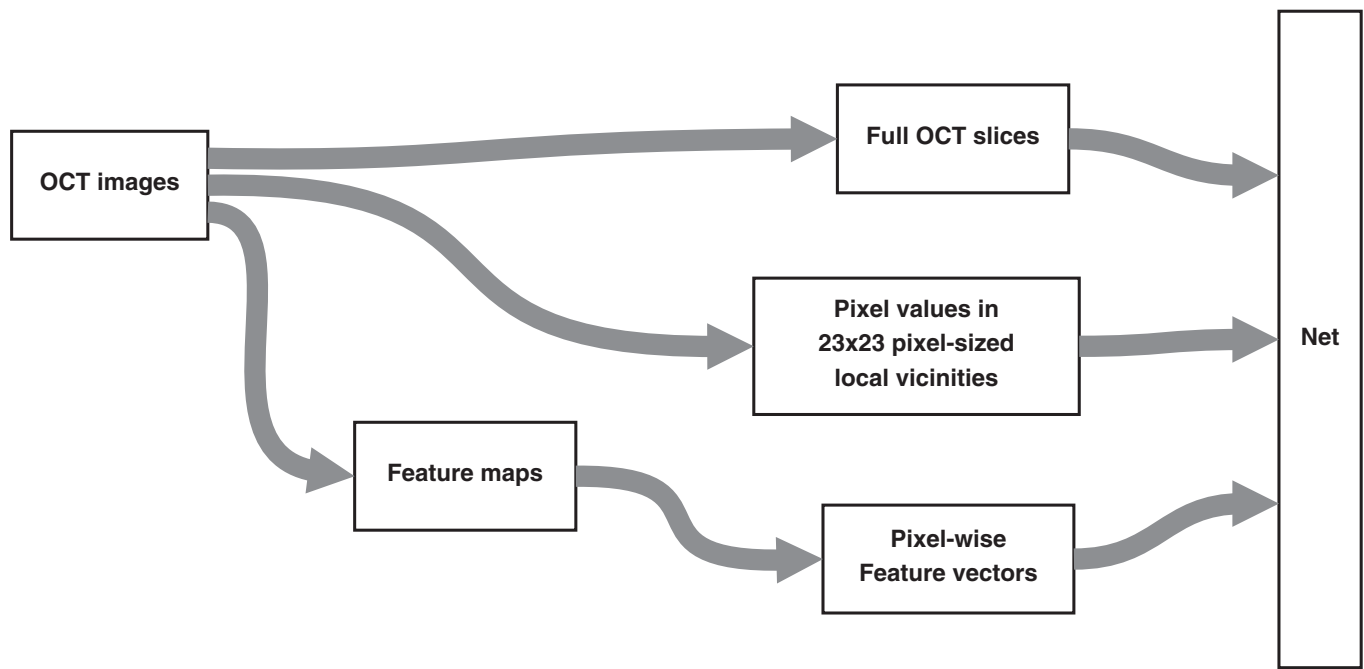


Fig. 2. Steps and output data of the data preparation procedure.

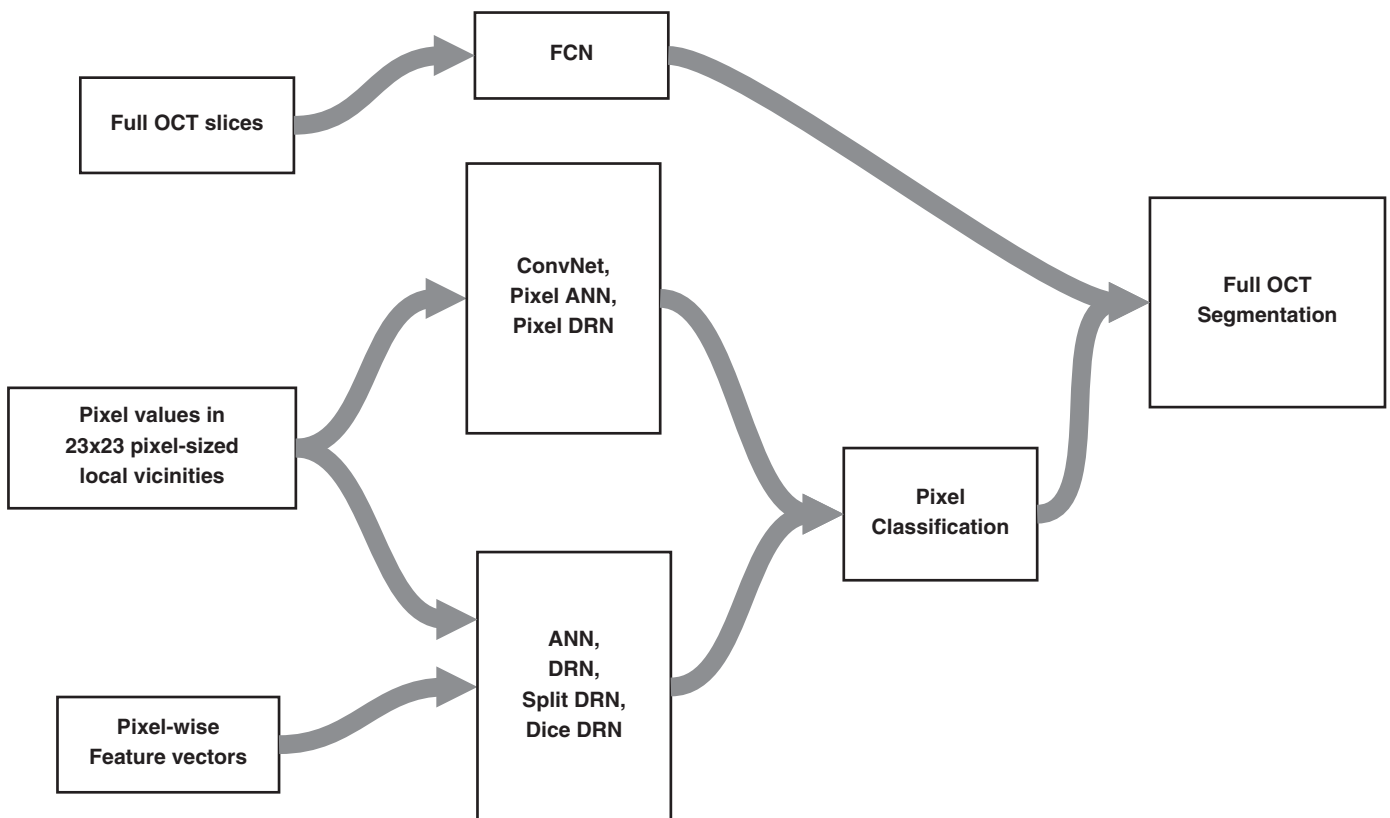


Fig. 3. An outline of the data flow used in the image segmentation process.

As the dataset was acquired from everyday clinical practice, two slightly different parameter sets were used. This is why 14 SD-OCT sequences contained 49 B-scan slices with a distance of 122 μm between consecutive slices, and 9 sequences contained 25 B-scan slices with a distance of 251 μm , giving 911 B-scans altogether. The data processing was performed on a slice-by-slice basis, so the

slice distance did not influence the results. The dimensions of each SD-OCT image were the following: 6 by 6 mm of the macular region, and the slices had a resolution of 512 by 496 pixels with pixel sizes of 11.45 and 3.87 μm , vertically and horizontally.

The 911 B-scans, after being converted from white-on-black to black-on-white images were annotated separately for training and

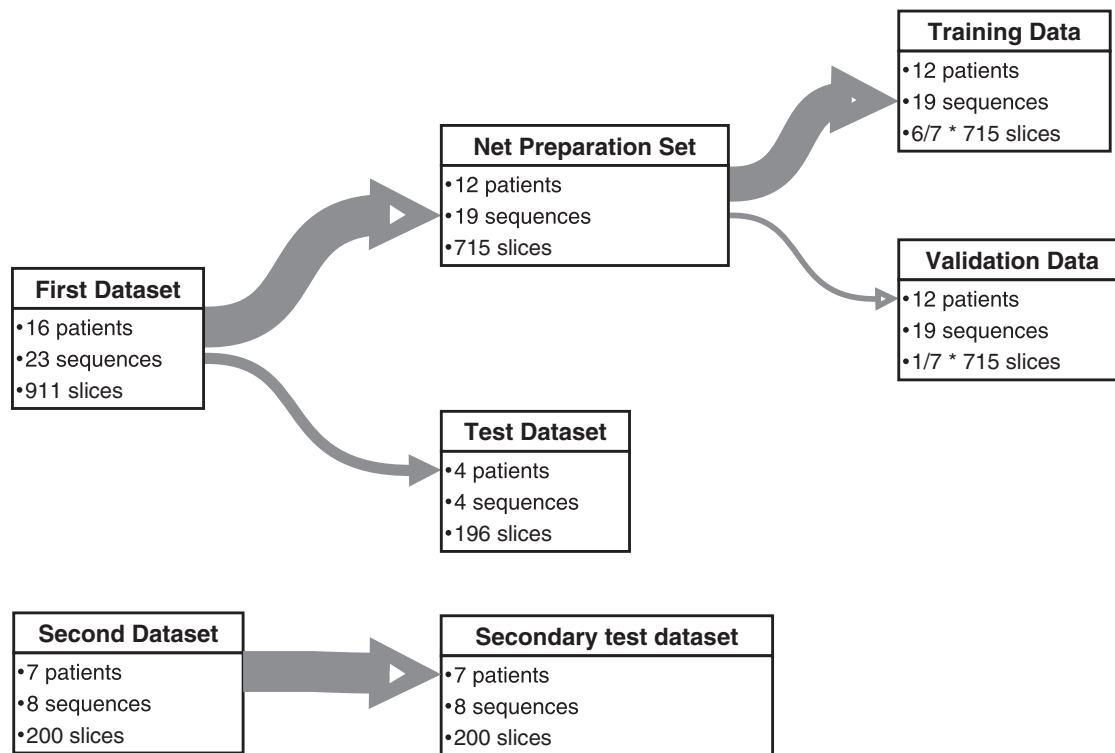


Fig. 4. Illustration of how the data was partitioned.

evaluation purposes by two clinical ophthalmologists. Before annotation both graders were given guidelines describing HF, hence the marked targets were bright spots with a reflectivity higher than the RPE band, and they had a diameter between 20 and 40 μm [11].

The first dataset was divided into two parts. Here, 19 out of 23 sequences were used for the preparation of the methods, i.e., designing the features and training the networks. The other 4 sequences (which we will call the test dataset), each containing 49 slices, were kept for evaluation purposes. These recordings came from 4 different patients, whose data were not included in any way in the training and data preparation process. The data flow of the procedure is shown in Fig. 4.

We chose the images of the test dataset in such a way that they could hold various amounts of HF. Table 1 lists the statistics of HF in the chosen sequences of our test set, and it shows both the number of marked HF and their area.

In addition, we collected a second dataset to further validate the best performing FCN network structure. The new data set consisted of 8 sequences of 7 patients. The data of selected patients were not included in the first dataset at all. This second dataset was collected and annotated 8 months after making the first one, with the same parameters, except that each sequence contained 25 slices, and – just like the first test dataset – it also had various HF counts listed in Table 1. This data set – which we will also call

Table 1

Amount and area of HF in the given sequences of the test dataset based on the two manual annotations.

	Dr1		Dr2		Total voxel count	Volume dimensions
	HF count	HF area	HF count	HF area		
Test dataset						
Seq_1	1060	21,195	870	18,117	12,443,648	$512 \times 496 \times 49$
Seq_2	279	4717	245	4866	12,443,648	$512 \times 496 \times 49$
Seq_3	52	1095	65	1424	12,443,648	$512 \times 496 \times 49$
Seq_4	196	4023	187	3917	12,443,648	$512 \times 496 \times 49$
Secondary test dataset						
Seq_5	163	3313	93	1800	6,348,800	$512 \times 496 \times 25$
Seq_6	422	8599	60	1278	6,348,800	$512 \times 496 \times 25$
Seq_7	235	4821	181	3104	6,348,800	$512 \times 496 \times 25$
Seq_8	114	2652	138	2774	6,348,800	$512 \times 496 \times 25$
Seq_9	196	4085	163	3139	6,348,800	$512 \times 496 \times 25$
Seq_10	252	5407	387	7424	6,348,800	$512 \times 496 \times 25$
Seq_11	430	8789	400	7664	6,348,800	$512 \times 496 \times 25$
Seq_12	308	6439	350	6967	6,348,800	$512 \times 496 \times 25$

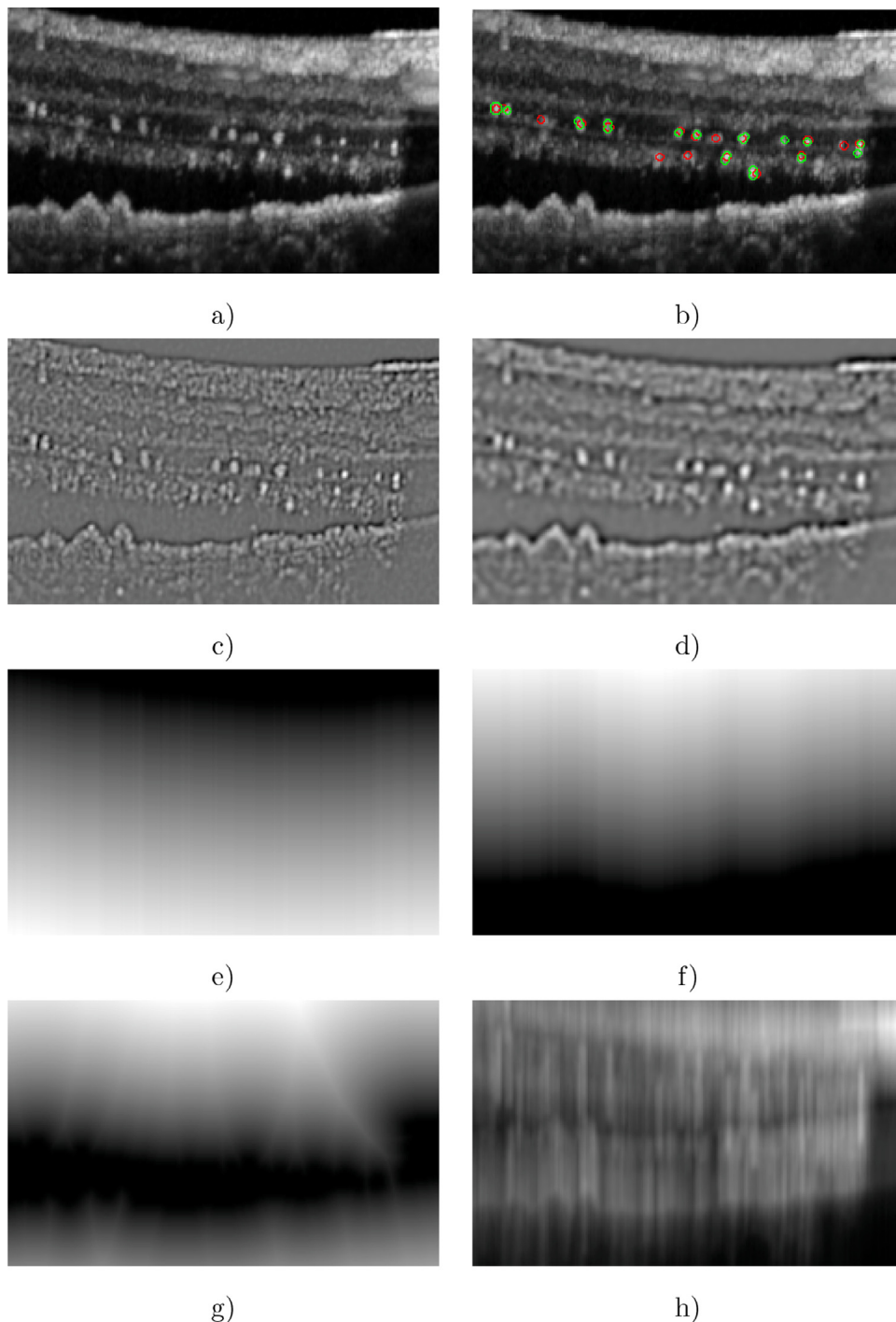


Fig. 5. Feature maps generated by image preprocessing: a) the original image; b) the location of HF according to the two experts (red, green); c–d) the results of LoG filters; e–g): a distance map from ILM, RPE and IRF, respectively; h) a convolution with the vertical mask.

secondary test dataset – was used entirely for the testing of the methods.

We used the two datasets for the training and evaluation of variants of four basic types of neural networks that were trained to segment the images. These types were Artificial Neural Networks (ANN), Deep Rectifier Neural Networks [27] (DRNs), Convolutional Neural Networks (CNN), and Fully Convolutional Neural Networks (FCN). The input of the Convolutional and Fully Convolutional networks were the images themselves, while the input of the other Neural Networks was created on a pixel-by-pixel basis. A feature

vector was generated for each pixel that served as a basis for classification. These features characterize the indicators of HF, and they were extracted by applying image processing techniques described next.

3.2. Image preprocessing and feature generation

The feature extraction for the DRNs and the ANN was performed after taking into account the suggestions made by medical experts. According to the literature, our relevant targets are small

Table 2

Summary of the network structures used in this study. In the Type column: “Conv” – convolutional layers; FC – Fully Connected layers. The “pixel input” column tells us whether raw pixel values in the 23x23 sized vicinity were used, while “feature input” tells us whether the net got feature vectors obtained via image processing steps.

Name	Type	Pixel input	Feature input	# layers	# params.
FCN	Conv	yes	–	4	69.7K
ConvNet	Conv + FC	yes	–	4	79.4K
ANN	FC	yes	yes	1	4.7M
DRN	FC	yes	yes	5	4.7M
Pixel ANN	FC	yes	–	1	4.4M
Pixel DRN	FC	yes	–	5	4.6M
Split DRN	FC	yes	yes	5	4.3M
Dice DRN	FC	yes	yes	5	4.7M

Table 3

Hyperparameters used for the training of the neural networks. The column names are: Loss–Loss function; optimizer–optimizer method; batch–batch size; epochs – number of epochs used in training; parameters–parameters of the optimization method. In the parameters column: lr. – learning rate; m. – momentum, wd. – weight decay.

Name	Loss	Optim.	Batch	Epochs	Parameters
FCN	Dice	Adam	10	70	lr.: 0.001; $\epsilon = 10^{-8}$
ConvNet	Logistic	SGD	128	4	β_1 : 0.9; β_2 : 0.999; lr.: 0.001; m.: 0.9; wd.: 0.004
ANN	CrossEntropy	SGD	100	max 25	lr.: 0.001
DRN	CrossEntropy	SGD	100	max 25	lr.: 0.001
Pixel ANN	CrossEntropy	SGD	100	max 25	lr.: 0.001
Pixel DRN	CrossEntropy	SGD	100	max 25	lr.: 0.001
Split DRN	CrossEntropy	SGD	100	max 25	lr.: 0.001
Dice DRN	Dice	SGD	100	max 25	lr.: 0.001

bright spots with a reflectivity higher than the RPE band, and a width of 3–6 pixels. They are also known to occur near sub-retinal fluid [10–13].

HF are similar to pigment particles, and blood vessels in the inner retinal layers, so in order to clearly distinguish HF from these, different types of features were used.

First, the feature vectors contained raw pixel information. For each pixel the first feature was the intensity of the pixel and the 23×23 vicinity and we provided local information concerning the intensity of the pixels for the neural networks (see Figs. 2 and 3).

Furthermore, other pixel-wise feature vectors were generated by using various image processing techniques, described below.

Eleven feature values were obtained by convolving the images different Laplacian of Gaussian (LoG) filters, which provided a basis for bright spot detection. The filters had different σ dispersal parameter values and they highlighted different sizes of spots in the images. Here, the σ values ranged from $3.87 \mu\text{m}$ to $7.74 \mu\text{m}$ with a step size of $0.387 \mu\text{m}$. Two examples of these features are shown in Fig. 5c and d.

We also exploited the spatial properties of the microglia, i.e., they should be mainly located in the outer retinal layers, and/or around pockets of fluid accumulation [10–13]. For this reason the internal limiting membrane (ILM) and the RPE layer was determined. Subretinal fluid was also detected using the methods outlined in [31]. Having detected the boundaries and the fluid, distance maps were generated from the specified regions, and for each pixel, the three distance values were included (the distance between the processed pixel and the ILM, RPE and IRF) in the feature vector. These values are depicted in Fig. 5e–g.

Next, the difference between HF, and pigment particles and blood vessels, was quantified. Pigment fragments and blood vessels cast shadows on the OCT images, which appear as a 20–40 pixel long vertical dark area under the bright spot, whereas HF do not generate this kind of shadow. This is why differentiating between HF, and pigments and blood vessels, was achievable by calculating

the average of the pixel values below the pixel coordinate as

$$A(i, j) = \frac{\sum_{k=1}^{40} I(i, j + k)}{40}, \quad (2)$$

assuming that I is the original OCT image, with index (i, j) as x and y coordinates, respectively, ascending from left to right and top to bottom. This feature is shown in Fig. 5h.

After calculating the line averages, we added the value set for each pixel

$$S(i, j) = \{A(i + k, j) \mid k \in \{-12, \dots, 12\}\}, \quad (3)$$

to the feature vectors. This set is a sequence of consecutive line averages under the pixels, which should enable the Neural Networks to detect shadows caused by pigment cells and blood vessels.

3.3. Training the nets

Using the training dataset, and the extracted features we trained several networks for the task of classifying the pixels as either HF pixels or as part of the background.

In order to carry out this task, different network types and network structures were applied. The eight most significant network structures that we tried out are described below. A summary of the structures and the training parameters can be found in Tables 2 and 3, respectively.

3.3.1. DRNs and ANNs

Following preliminary experiments, we decided to utilize a very simple network structure. Namely, we trained DRNs with 5 fully connected hidden layers, each containing 1000 neurons. As a comparison we also trained conventional artificial neural networks (ANNs) with one large hidden layer, and these ANNs had the same amount of trainable parameters as the deep ones. The last layer had 2 neurons with a SoftMax activation. Here, we utilized the Cross Entropy loss function.

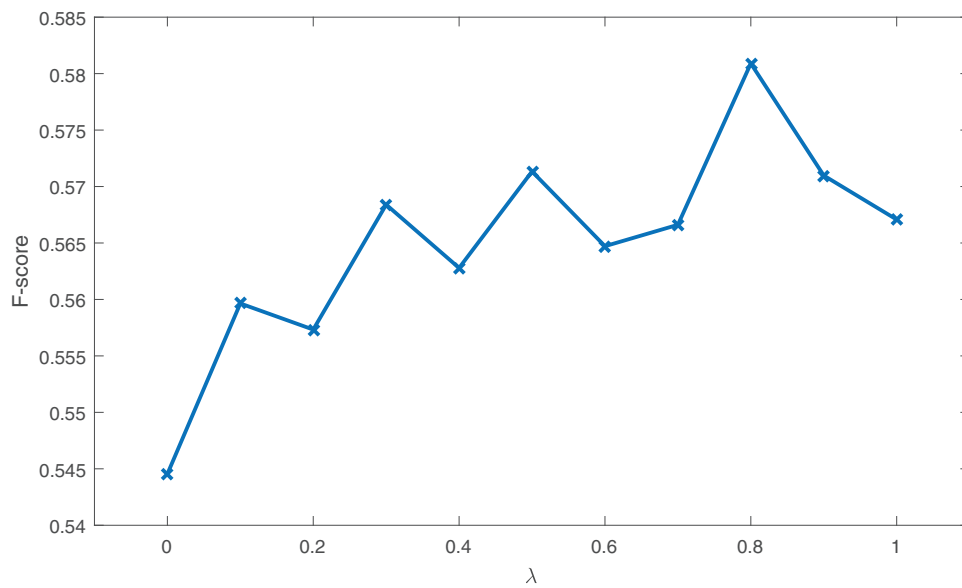
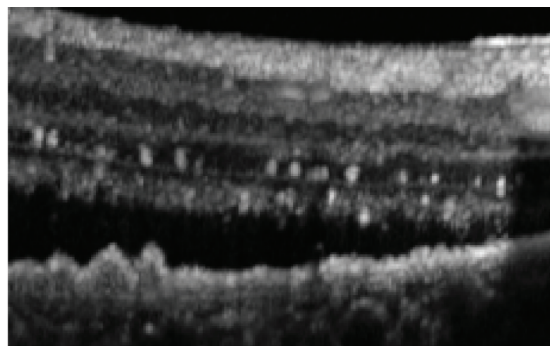
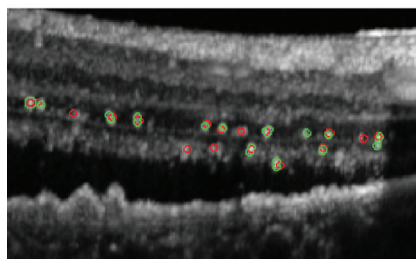


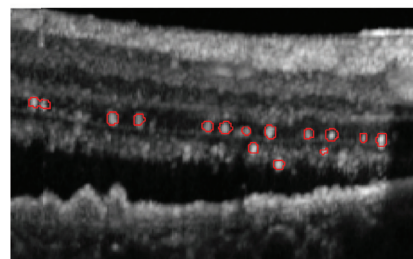
Fig. 6. Pixel level F -scores got on the validation set using probabilistic sampling.



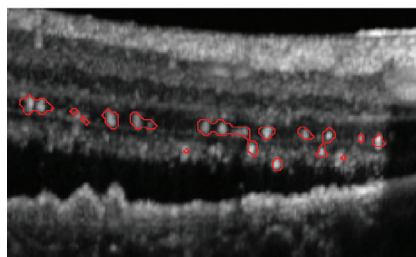
Original image



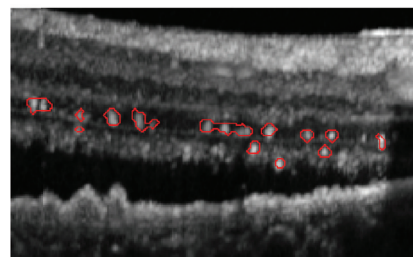
Dr1 (red) and Dr2 (green)



FCN



Dice DRN



DRN

Fig. 7. Segmentations of an OCT image got by manual annotation (Dr1 and Dr2), and by the three best Neural Networks.

Table 4

Dice coefficients got for the various types of neural networks with different pixel tolerances. (Values above 95% of the baseline are highlighted in bold.)

	Tolerance: 0 pixel				Tolerance: 1 pixel			
	Seq_1	Seq_2	Seq_3	Seq_4	Seq_1	Seq_2	Seq_3	Seq_4
Dr1 ↔ Dr2	0.463	0.420	0.378	0.401	0.680	0.619	0.578	0.594
Dr1 ↔ FCN	0.480	0.329	0.386	0.310	0.695	0.516	0.573	0.493
Dr1 ↔ ConvNet	0.229	0.088	0.115	0.094	0.374	0.161	0.196	0.174
Dr1 ↔ ANN	0.408	0.188	0.181	0.164	0.553	0.263	0.267	0.238
Dr1 ↔ DRN	0.468	0.283	0.358	0.302	0.610	0.384	0.494	0.412
Dr1 ↔ Pixel ANN	0.408	0.199	0.185	0.199	0.546	0.277	0.268	0.275
Dr1 ↔ Pixel DRN	0.455	0.289	0.293	0.306	0.603	0.389	0.420	0.403
Dr1 ↔ Split DRN	0.470	0.289	0.353	0.285	0.626	0.391	0.482	0.395
Dr1 ↔ Dice DRN	0.478	0.282	0.371	0.277	0.623	0.381	0.495	0.381
Dr2 ↔ FCN	0.548	0.420	0.438	0.402	0.747	0.589	0.629	0.579
Dr2 ↔ ConvNet	0.241	0.109	0.117	0.119	0.382	0.187	0.204	0.211
Dr2 ↔ ANN	0.432	0.161	0.229	0.174	0.568	0.229	0.320	0.244
Dr2 ↔ DRN	0.520	0.288	0.408	0.352	0.657	0.383	0.515	0.461
Dr2 ↔ Pixel ANN	0.435	0.208	0.229	0.225	0.572	0.278	0.307	0.306
Dr2 ↔ Pixel DRN	0.500	0.309	0.344	0.354	0.642	0.411	0.455	0.462
Dr2 ↔ Split DRN	0.511	0.307	0.391	0.339	0.663	0.415	0.506	0.448
Dr2 ↔ Dice DRN	0.524	0.311	0.399	0.334	0.668	0.409	0.498	0.435
	Tolerance: 2 pixel				Tolerance: 3 pixel			
	Seq_1	Seq_2	Seq_3	Seq_4	Seq_1	Seq_2	Seq_3	Seq_4
Dr1 ↔ Dr2	0.779	0.710	0.660	0.673	0.829	0.744	0.707	0.715
Dr1 ↔ FCN	0.801	0.614	0.673	0.586	0.858	0.669	0.730	0.639
Dr1 ↔ ConvNet	0.494	0.224	0.266	0.244	0.611	0.397	0.339	0.316
Dr1 ↔ ANN	0.733	0.407	0.369	0.372	0.807	0.462	0.427	0.430
Dr1 ↔ DRN	0.793	0.561	0.631	0.577	0.856	0.626	0.697	0.635
Dr1 ↔ Pixel ANN	0.739	0.441	0.355	0.427	0.809	0.509	0.407	0.488
Dr1 ↔ Pixel DRN	0.781	0.576	0.540	0.556	0.845	0.642	0.614	0.608
Dr1 ↔ Split DRN	0.796	0.559	0.608	0.526	0.864	0.618	0.660	0.587
Dr1 ↔ Dice DRN	0.802	0.557	0.633	0.527	0.866	0.622	0.693	0.579
Dr2 ↔ FCN	0.825	0.649	0.713	0.656	0.858	0.671	0.740	0.701
Dr2 ↔ ConvNet	0.494	0.254	0.279	0.285	0.601	0.326	0.347	0.364
Dr2 ↔ ANN	0.715	0.339	0.433	0.361	0.768	0.378	0.478	0.403
Dr2 ↔ DRN	0.802	0.538	0.641	0.626	0.842	0.585	0.688	0.681
Dr2 ↔ Pixel ANN	0.727	0.410	0.414	0.436	0.774	0.461	0.446	0.483
Dr2 ↔ Pixel DRN	0.784	0.559	0.596	0.614	0.823	0.599	0.634	0.654
Dr2 ↔ Split DRN	0.796	0.551	0.621	0.574	0.837	0.590	0.655	0.614
Dr2 ↔ Dice DRN	0.803	0.553	0.663	0.577	0.841	0.595	0.697	0.618

To validate the quality of the extracted features we also trained networks just using the raw pixel data (Pixel ANN/DRN). Then, we used a DRN that had a split first hidden layer (denoted by Split DRN), half of the neurons being connected to the pixel input and the other half being connected to the other features described Section 3.2.

To train the neural networks we applied stochastic gradient descent (i.e. backpropagation) training with a mini-batch size of 100. The initial learn rate was set to 0.001, which was halved after each iteration when the F-score on the validation set did not increase. As only a small percentage of the training data actually belonged to the HF, we chose to re-sample the data during training using probabilistic sampling. Fig. 6 shows the results we got by using the re-sampling method described earlier. The best performance at the pixel-level on the validation set was achieved with $\lambda = 0.8$, which is close to that of uniform re-sampling. Training of the ANNs and DRNs took 6–25 epochs with our custom implementation.

3.3.2. CNNs and FCNs

The CNNs had an input of a 23×23 pixel-sized image. To be able to use classification networks for the segmentation, we extracted 23×23 pixel-sized windows of the OCT images. The extracted parts were then selected around the given pixel, and we obtained extracts using all the coordinates of the OCT images. The expected output of the net was the label of the center pixel of the 23×23 pixel-sized local window of the image. Hence, the CNNs

themselves learned to classify pixels in the OCT image based on their surroundings. By classifying all the pixels, a segmentation of an OCT image can be assembled.

The CNN had three convolutional layers, each followed by a ReLU, and a 2×2 max pooling layer. The convolutional layers had 32, 32 and 64 neurons, respectively. The final pooling layer was followed by a fully connected layer of 2 neurons that had a softmax activation function on its output, and we used the summed absolute difference of the label values as the loss function. The training was carried out with probabilistic sampling with a $\lambda = 0.8$ value, and the network optimization converged after 4 epochs using the stochastic gradient descent method. Additional parameters can be found in Table 3.

Along with this network, which we will call CNN, we also trained the AlexNet structure, which gave us less accurate results. This is why we decided to exclude it from any further evaluation.

The network we call FCN had a simple structure containing 4 layers with $11 \times 11 \times [\text{input channel num}]$ -sized kernels. The first three layers had 16 kernels and they were followed by a ReLU activation, while the final layer only had one output channel and a sigmoid activation function. The loss function was defined as the Dice coefficient of the segmentation compared to the ground truth. We trained the FCN with the Adam optimizer [32], using the parameters given in Table 3.

It should be mentioned that we experimented with other network structures as well, of which the currently applied one sup-

Table 5
Evaluations of the FCN on the Test and Secondary test datasets. The tables contain statistics that compare the annotations of the two medical experts with each other and with the FCN. The columns give Dice (F1-score), Precision, Recall and AUC values evaluated on the complete datasets. The sub-tables give statistics for different tolerances in the vicinity of positive pixels. (Values above 95% of the baseline are highlighted in bold.)

Tolerance: 0 pixel					
		Dice	Precision	Recall	AUC
Test	Dr1 ↔ Dr2	0.443	0.418	0.473	0.706
	Dr1 ↔ FCN	0.431	0.429	0.433	0.762
	Dr2 ↔ FCN	0.504	0.536	0.477	0.803
Secondary test	Dr1 ↔ Dr2	0.386	0.342	0.442	0.676
	Dr1 ↔ FCN	0.344	0.302	0.401	0.704
	Dr2 ↔ FCN	0.442	0.415	0.428	0.776
Tolerance: 1 pixel					
		Dice	Precision	Recall	AUC
Test	Dr1 ↔ Dr2	0.654	0.615	0.699	n/a
	Dr1 ↔ FCN	0.638	0.621	0.656	n/a
	Dr2 ↔ FCN	0.696	0.712	0.680	n/a
Secondary test	Dr1 ↔ Dr2	0.589	0.528	0.666	n/a
	Dr1 ↔ FCN	0.554	0.492	0.634	n/a
	Dr2 ↔ FCN	0.604	0.593	0.617	n/a
Tolerance: 2 pixel					
		Dice	Precision	Recall	AUC
Test	Dr1 ↔ Dr2	0.748	0.701	0.802	n/a
	Dr1 ↔ FCN	0.741	0.712	0.773	n/a
	Dr2 ↔ FCN	0.771	0.783	0.760	n/a
Secondary test	Dr1 ↔ Dr2	0.690	0.622	0.776	n/a
	Dr1 ↔ FCN	0.673	0.605	0.757	n/a
	Dr2 ↔ FCN	0.681	0.673	0.689	n/a
Tolerance: 3 pixel					
		Dice	Precision	Recall	AUC
Test	Dr1 ↔ Dr2	0.795	0.742	0.856	n/a
	Dr1 ↔ FCN	0.798	0.764	0.834	n/a
	Dr2 ↔ FCN	0.804	0.820	0.789	n/a
Secondary test	Dr1 ↔ Dr2	0.751	0.699	0.863	n/a
	Dr1 ↔ FCN	0.752	0.724	0.854	n/a
	Dr2 ↔ FCN	0.722	0.748	0.732	n/a

plied the best results. We also tried using the V-net [25] and U-Net [9] structures, but neither of these networks was able to give reasonable results within 70 epochs using various optimization techniques. Hence, we will not go into more detail here. The training difficulties are probably due to the small area of positive HF-s in the images and to overfitting.

In the case of fully convolutional networks, they were implemented using the Tensorflow [33] framework.

3.4. Evaluation

After the training phase, we performed segmentations with the trained nets on the test data.

Our medical experts evaluated the output images, and also numerical measurements were performed using the Dice coefficient (which is equivalent to the f1-score) as an accuracy metric. For a systematic evaluation, we compared the scores with the annotations of two doctors. Afterwards, to get a baseline for the desired accuracy the annotations of the doctors were compared with each other.

The annotated data were compared with the segmentation results of the neural networks on a pixel basis, using the original and a modified version of the Dice formula. The annotation of the data was performed manually, including small (1–2 pixel) misalignments at the edges of the HF (see, e.g, the manual segmentations in Fig. 7). This is why, the modified Dice formula included a small tolerance, i.e., a positive pixel was accepted – of either the reference or the segmentation – as a true positive, if the other image had a positive label in a given vicinity.

4. Results and discussion

Some examples of the segmentations produced by the nets can be seen in Fig. 7. The segmentations were approved by the medical experts, and based on our numerical results, we can now proceed to the large-scale clinical evaluation of the methods, and the further development of a decision aiding process.

The numerical results of the numerous comparison are listed in Table 4. The values in the table that were nearly as accurate (they had a Dice coefficient higher than 95%) as the overlap between the two annotating doctors have been highlighted.

Observing the results, it is clear that the performance of different networks varied greatly. The best results were produced by the FCN. In most cases its score was close to – or even higher than – the baseline between the annotators.

Apart from the FCN, some of the other networks performed quite well, especially the DRN using all the features, and the Dice DRN. The good performance of the Dice DRN can be accounted for by its loss function, as this network was trained with the same error measure that was used in the final evaluation, hence the training process could be improved. The DRN with the full feature set, however, had additional information extracted from the image, and help it to find a better model.

As for the other neural networks, the Split DRN, and the only Pixel-based DRN gave slightly worse results, and the ANNs could not attain the results of the best networks.

Surprisingly, the CNN gave outstandingly bad results, which might be due to several reasons. First of all, the network performed worse than the FCN because of its small batch size. The

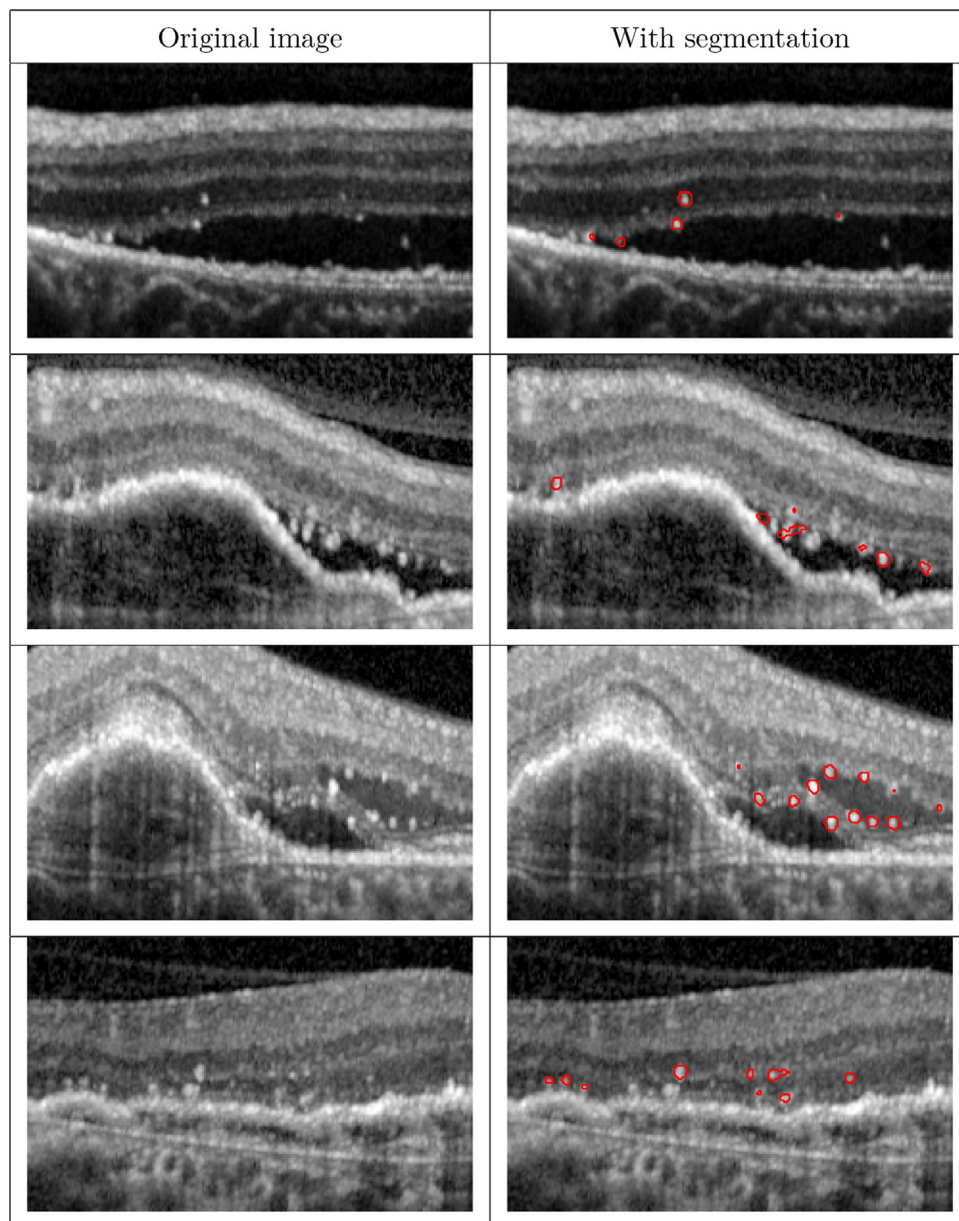


Fig. 8. Additional segmentations of the FCN model.

training data sets are really biased towards negative labels, and there are many batches in the training process that contain just a few positive pixels. The probabilistic sampling can alleviate this problem, although changing the distribution of the samples too much can bias the network output towards positive labels. The FCN is not influenced too much by this, since it gets 10 full images in one batch, i.e., one batch in practice contains over 2 million samples, as opposed to the 128 samples when training the CNN. The large amount of processed pixels gives the FCN sufficient positive samples for a good convergence, and together with the Dice loss function it can handle the underrepresentation of positive labels.

Another disadvantage of the CNN lies in its structure. The CNN contains convolutions on local regions followed by pooling layers. The pooling layers decrease the resolution of the images, and this degrades the spatial accuracy that might be vital for the accurate recognition of small – 3–6 pixel-sized – objects like HF. Note that the best performing FCN does not contain any pooling layers.

Comparing the performance of the CNN with the non-convolutional networks (i.e., the variants of the DRN and the ANN), we see that the DRN and ANN have a different advantage. The CNN contains convolutional kernels which perform the same operation on the whole image. The DRN and ANN, however, calculate a weighted sum, but they can learn spatially different weights. Since the input images are quite specific around positive samples (e.g., they have a bright spot in the middle, and a grey area around them) and spatially differentiated operations can have an advantage.

Based on these results, we can say that the FCN, the Dice DRN and the Full DRN can provide accurate segmentations of HF on OCT images, which in terms of accuracy are close to the correlation values got between manual segmentations.

As a further validation of the results, we decided to perform tests using the 8 new sequences of the secondary test dataset, and evaluated the best performing net which we called FCN. These results can be seen in Table 5, and some other examples of segmentation are shown in Fig. 8.

In the new tests, we assessed the networks on the first and second dataset separately. We also extended the assessment by including the Precision, Recall, and AUC values.

Looking closely at the results, one can see a slight shift in the balance towards the Recall value. It seems that the networks tend to segment a few more HF than the human annotators. This may be due to the uncertainty of the annotations in the test dataset.

We also found that the results for first and second datasets were comparable, and in both cases, the FCN produced results close to the baseline in the Dice value. This was also the case for the Precision, Recall and AUC values.

5. Availability of the software

The software package implemented for the study is available on request. The interested reader should contact us by email at vargalg@inf.u-szeged.hu.

6. Conclusions and future plans

Due to the rapid development of OCT machines, several biomarkers detectable on OCT were identified in the examination of chorioretinal diseases [6]. In a real-life situation where making the decision of whether to treat a patient or not, the ophthalmologist needs to take into consideration a number of factors, including the biomarkers on each slice of the scans.

The number and the localization of HF was found to correlate with the progression of AMD and the activity of choroidal neovascularization [1,6,11–13]. Our aim was to create software tools that could recognize HF with great accuracy, and hence facilitate the highly challenging assessment of this biomarker and assist the ophthalmologists.

In this study, we devised a framework for the automatic segmentation of HF in OCT images. We applied image processing methods and ANNs, DRNs and CNNs to learn from the annotation of medical doctors and carry out the segmentation.

Two possible limitations of the study is the relatively small size of population that provided the base for the dataset and the fact that OCT scans were obtained from the imaging system of just one manufacturer. In further investigations we would like to use a bigger number of real-life patients with wet AMD, and to extend the software to include other ophthalmologic conditions, such as diabetic macular oedema and retinal vein occlusion. Moreover, we intend to automatically segment other biomarkers and look for possible relationships and correlations in their appearance and behavior. Finally, we are also planning to compare our methods with other methods provided by a parallel study in [16].

Our results confirm that with the right software, we can support the clinician in analyzing the OCT scans and help reduce time-consuming examinations in healthcare. The results of the automatic segmentation methods were subsequently investigated by retina specialists and the results were deemed sufficiently accurate to track the changes in HF. Hence, it may reasonably be concluded that neural networks can be used to segment HF in OCT images to a high accuracy.

Ethical statement

The study was carried out in accordance with the principles of the Declaration of Helsinki, and it was carried out with the ethical approval of the Institutional Review Board of University of Szeged, Albert Szent-Györgyi Clinical Center (reference number: 3650).

Declaration of Competing Interest

This research was supported by the project "Integrated program for training new generation of scientists in the fields of computer

science", no EFOP-3.6.3-VEKOP-16-2017-0002. The project has been supported by the European Union and co-funded by the European Social Fund. Ministry of Human Capacities, Hungary grant 20391-3/2018/FEKUSTRAT is acknowledged. Tamás Grósz was supported by the National Research, Development and Innovation Office of Hungary through the Artificial Intelligence National Excellence Program (grant no.: 2018-1.2.1-NKP-2018-00008).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.cmpb.2019.06.019](https://doi.org/10.1016/j.cmpb.2019.06.019).

References

- [1] U. Schmidt-Erfurth, A. Sadeghipour, B.S. Gerendas, S.M. Waldstein, H. Bogunovi, Artificial intelligence in retina, *Prog. Retin. Eye Res.* (2018), doi:[10.1016/j.preteyeres.2018.07.004](https://doi.org/10.1016/j.preteyeres.2018.07.004). Pii: S1350-9462(18)30011-9. [Epub ahead of print]
- [2] D.S.W. Ting, L.R. Pasquale, L. Peng, Artificial intelligence and deep learning, *Ophthalmol. Br. J. Ophthalmol.* 103 (2019) 167–175.
- [3] R. Klein, et al., Prevalence of age-related maculopathy. The Beaver Dam Study, *Ophthalmology* 99 (1992) 933–943.
- [4] R. Velez-Montoya, et al., Current knowledge and trends in age-related macular degeneration. Today's and future treatments, *Retina* 33 (2012) 1487–1502.
- [5] G. Coscas, et al., Clinical features and natural history of AMD on OCT, in: G. Coscas, et al. (Eds.), *Optical Coherence Tomography in Age-Related Macular Degeneration*, Springer, Heidelberg, 2009, pp. 195–274.
- [6] U. Schmidt-Erfurth, S.M. Waldstein, A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration, *Prog. Retin. Eye Res.* 50 (2016) 1–24.
- [7] A.I. Shahin, Y. Guo, K.M. Amin, A.A. Sharawi, White blood cells identification system based on convolutional deep neural learning networks, *Comput. Methods Programs Biomed.* (2017).
- [8] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: a review, *Comput. Methods Programs Biomed.* 161 (2018) 1–13.
- [9] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*. MICCAI. Lecture Notes in Computer Science, vol 9351, Springer, Cham, 2015.
- [10] G. Coscas, et al., Hyperreflective dots: a new spectral-domain optical coherence tomography entity for follow-up and prognosis in exudative age-related macular degeneration, *Ophthalmologica* 229 (2013) 32–37.
- [11] H. Lee, B. Ji, H. Chung, H.C. Kim, Correlation between optical coherence tomographic hyperreflective foci and visual outcomes after anti-VEGF treatment in neovascular age-related macular degeneration and polypoidal choroidal vasculopathy, *Retina* 36 (3) (2016) 465–475, doi:[10.1097/IAE.0000000000000645](https://doi.org/10.1097/IAE.0000000000000645).
- [12] M.B. Parodi, A. Arrigo, F. Romano, E. Aragona, A. Marchese, M.V. Cicinelli, S. Mercuri, F. Bandello, Hyperreflective foci number correlates with choroidal neovascularization activity in angioid streaks, *Investig. Ophthalmol. Vis. Sci.* 59 (8) (2018) 3314–3319, doi:[10.1167/jovs.18-24291](https://doi.org/10.1167/jovs.18-24291).
- [13] O. Segal, E. Barayev, A.Y. Nemet, N. Geffen, I. Vainer, M. Mimouni, Prognostic value of hyperreflective foci in neovascular age-related macular degeneration treated with bevacizumab, *Retina* 36 (11) (2016) 2175–2182.
- [14] J.G. Christenbury, F.A. Folgar, R.V. O'Connell, S.J. Chiu, S. Farsiu, C.A. Toth, Progression of intermediate age-related macular degeneration with proliferation and inner retinal migration of hyperreflective foci, *Ophthalmology* 120 (5) (2013) 1038–1045.
- [15] E. Korot, G. Comer, T. Steffens, D.A. Antonetti, Algorithm for the measure of vitreous hyperreflective foci in optical coherence tomographic scans of patients with diabetic macular edema, *JAMA Ophthalmol.* 134 (1) (2016) 15–20.
- [16] T. Schlegel, H. Bogunovic, S. Klimscha, P. Seebck, A. Sadeghipour, B. Gerendas, S.M. Waldstein, G. Langs, U. Schmidt-Erfurth, Fully automated segmentation of hyperreflective foci in optical coherence tomography images. *ArXiv:1805.03278*.
- [17] M. Katona, A. Kovács, L. Varga, T. Grósz, J. Dombi, R. Dégi, L.G. Nyl, Automatic detection and characterization of biomarkers in OCT images, *Lect. Notes Comput. Sci.* 10882 (2018) 706–714.
- [18] S. Richer, et al., Double-masked, placebo-controlled, randomized trial of lutein and antioxidant supplementation in the intervention of atrophic age-related macular degeneration: the veterans LAST study, *Optometry* 75 (4) (2004) 216–230.
- [19] P.J. Rosenfeld, et al., Ranibizumab for neovascular age-related macular degeneration, *N. Engl. J. Med.* 355 (2006) 1419–1431.
- [20] M. Bolz, U. Schmidt-Erfurth, G. Deak, G. Mylonas, K. Kriechbaum, C. Scholda, Diabetic retinopathy research group vienna. optical coherence tomographic hyperreflective foci: a morphologic sign of lipid extravasation in diabetic macular edema, *Ophthalmology* 116 (5) (2009) 914–920, doi:[10.1016/j.ophtha.2008.12.039](https://doi.org/10.1016/j.ophtha.2008.12.039).
- [21] K. Ogino, T. Murakami, A. Tsujikawa, K. Miyamoto, A. Sakamoto, M. Ota, N. Yoshimura, Characteristics of optical coherence tomographic hyperreflective foci in retinal vein occlusion, *Retina* 32 (1) (2012) 77–85, doi:[10.1097/IAE.0b013e318217ffc7](https://doi.org/10.1097/IAE.0b013e318217ffc7).

- [22] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Neural Inf. Process. Syst.* (2012). 25. [10.1145/3065386](https://doi.org/10.1145/3065386).
- [23] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), NV, Las Vegas, 2016, pp. 770–778, doi:[10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [24] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. ArXiv:[1409.1556](https://arxiv.org/abs/1409.1556).
- [25] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation. ArXiv:[1606.04797](https://arxiv.org/abs/1606.04797).
- [26] A.G. Roy, S. Conjeti, S. Karri, D. Sheet, A. Katouzian, C. Wachinger, N. Navab, Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks, *Biomed. Opt. Express* 8 (8) (2017) 3627–3642, doi:[10.1364/BOE.8.003627](https://doi.org/10.1364/BOE.8.003627).
- [27] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier networks, in: *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, JMLR W&CP, 2011, pp. 315–323. 15
- [28] T. Salimans, D.P. Kingma, Weight normalization: a simple reparameterization to accelerate training of deep neural networks, 2016. ArXiv:[1602.07868](https://arxiv.org/abs/1602.07868) [cs.LG], url: [1602.07868](https://arxiv.org/abs/1602.07868).
- [29] L. Geert, K. Thijs, B.B. Ehteshami, S.A.A. Adiyoso, C. Francesco, G. Mohsen, V.D.L.J. Awm, V.G. Bram, I. Sánchez Clara, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88. Elsevier
- [30] L. Tóth, A. Kocsor, Training HMM/ANN hybrid speech recognizers by probabilistic sampling, in: *Artificial Neural Networks: Biological Inspirations ICANN 2005*, LNCS, 3696, 2005, pp. 597–603.
- [31] M. Katona, A. Kovács, R. Dégi, L.G. Nyúl, Automatic detection of subretinal fluid and cyst in retinal images, in: *Image Analysis and Processing – ICIAP 2017 – 19th International Conference*, Catania, Italy, 2017, pp. 606–616. September 11–15, *Proceedings, Part I*
- [32] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization. ArXiv:[1412.6980](https://arxiv.org/abs/1412.6980).
- [33] M. Abadi, et al., Tensorflow: large-scale machine learning on heterogeneous systems, Software Available from tensorflow.org, 2015.

III.

Automatic Detection of Subretinal Fluid and Cyst in Retinal Images

Melinda Katona¹, Attila Kovács², Rózsa Dégi², and László G. Nyúl¹(✉)

¹ Department of Image Processing and Computer Graphics, University of Szeged,
Árpád tér 2, Szeged 6720, Hungary
{mkatona,nyul}@inf.u-szeged.hu

² Department of Ophthalmology, University of Szeged, Korányi fasor 10-11,
Szeged 6720, Hungary
{kovacs.attila,degi.rozsa}@med.u-szeged.hu

Abstract. A modern tool for age-related macular degeneration (AMD) investigation is Optical Coherence Tomography (OCT) that can produce high resolution cross-sectional images about retinal layers. AMD is one of the most frequent reasons of blindness in economically advanced countries. AMD means degeneration of the macula which is responsible for central vision. Since AMD affects only this specific part of the retina, unattended patients lose their fine shape- and face recognition, reading ability, and central vision. We present a novel algorithm to localize subretinal fluid and cyst segments and extract quantitative measures thereof. Since, these algorithms are fully automated, the doctor does not need to perform extremely time-consuming manual contouring and human inaccuracies can be also eliminated.

Keywords: Optical Coherence Tomography · SD-OCT · Age-related macular degeneration · AMD · Subretinal fluid · Cyst

1 Introduction

Age-related macular degeneration is one of the most frequent reasons of acquired blindness in economically advanced countries. The constant growing of AMD patient population is more and more challenging. AMD means degeneration of the macula which is the region of the retina responsible for central vision. Since AMD affects only this specific part of the retina, unattended patients lose their fine shape- and face recognition, reading ability, and central vision [9].

Basically, AMD has two forms: dry and wet form, and the latter causes rapid and serious visual impairment in 10% of the cases [13]. In this type of the disease, abnormal angiogenesis starts from the choroid under the macula. Fluid and blood leak out of the neovascularized membrane into retina layers that ruins the photoreceptors.

Experiments have demonstrated that the vascular endothelial growth factor (VEGF) plays a vital role in the formation of choroidal neovascularization [4].

Currently, the most common and effective clinical treatment for wet AMD is anti-VEGF therapy, which is a periodic intravitreal (into the eye) injection [11].

In the last decade, Optical Coherence Tomography (OCT) has been widely used in the diagnosis of AMD and follow-up therapy. Spectral domain OCT (SD-OCT) produces 3D volumes of data, which have been useful in clinical practice. Existing OCT systems are partially suited to monitoring the progress of the disease, but OCT shows many features about AMD such as hyper-reflective dots (HRD), subretinal fluid and cysts. Figure 1 illustrates an SD-OCT B-scan with biomarkers of AMD.

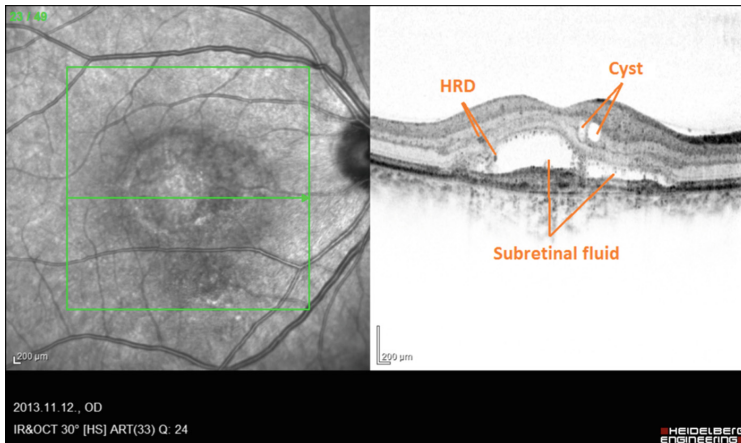


Fig. 1. Original Optical Coherence Tomography (SD-OCT) image with biomarkers of AMD in inverted display (as our medical experts use it in daily routine).

A large number of publications in the scientific literature deal with the problem of detecting retinal layers based on various techniques. One approach is the automatic segmentation procedure using graph theory [1, 2, 6]. In this approach, the graph nodes usually relate to image pixels, the graph edges are assigned to pairs of pixels, the edge weights depend on the intensity differences between the node pixels, and also may depend on the spatial distance between the pixels. Image segmentation then becomes a graph cutting problem, which can be solved via dynamic programming. These approaches are less tolerant to noise, that is a disadvantage, because real images are often very noisy. Another idea relies on the well-known energy-minimizing active contour method which, unfortunately, also has problems in handling low contrast and noise. Yazdanpanah et al. [18] suggested a multi-phase framework with a circular shape prior in order to model the boundaries of retinal layers and estimate the shape parameters. They used a contextual scheme to balance the weight of different terms in the energy functional. Machine learning is widely used in recent years, also for retinal image analysis. Lang et al. [12] used random forest classifier to segment retinal layers.

The random forest classifier learns the boundary pixels between layers and produces an accurate probability map for each boundary, which is further processed to finalize boundaries. Procedures based on active contour or machine learning provide effective solution, but these methods are too time-consuming. Hassan et al. [8] used a structure tensor approach combined with a nonlinear diffusion process for layer detection. A structure tensor is a second-moment matrix that shows similarities and prominent orientations of the image gradient. Some other approaches use optimized boundary tracking [5] or polynomial smoothing [14]. These algorithms are rather complex. We have developed an algorithm which uses simple operations to localize subretinal areas. It is based on vertical profile analysis [10].

Relatively few publications deal with the problem of automatically detecting cysts. Gonzalez et al. [7] described a method based on watershed segmentation and different machine learning classifiers. They focused on feature extraction which can help to eliminate false regions. Other approaches can also be found in the literature, two of which are discussed in Sect. 2.1.

In this study, we deal with the automatic localization of subretinal fluid areas and cysts and also analyze major retinal layers, since layer information can help localizing and distinguishing fluid and cyst regions. We present an algorithm that automatically delineates the ILM (inner-limiting membrane) and RPE (retinal pigment epithelium) retinal layers. We also describe a method to detect subretinal fluid and cyst segments and distinguish them from each other. We compare our results with some other algorithms from literature.

2 Methods

In this section, we present several algorithms that use different approaches to determine cyst and subretinal fluid. First, we briefly describe two existing approaches that we re-implemented according to the original papers for comparison. Then, we describe our novel approach in more detail. The procedures first delineate the inner and outer boundary layers (ILM and RPE, resp.) for easier determination of the important areas.

2.1 Literature Procedures

Firstly, we describe Wieclawek's [16] algorithm. First, the input image is normalized to the $[0\ 1]$ interval, because images can be made with different settings, so that their intensity range may vary. The OCT images are affected by distortions like noise, so the authors used a non-linear filtering to reduce this effect. Next, they applied a spatial averaging filtering technique which is based on the real product of complex diffusion. The tools of mathematical morphology was used to delineate specific cystic areas, based on the observation that cysts appear as darker segments in the images. Among other operations, they used H-minima transform to highlight important regions. The single control parameter is a threshold value. This value has been fixed experimentally to 30% of the

maximum brightness in the image. The next step was the binarization of the obtained image with a given threshold value. Since the result may still contain false regions, they filtered all objects that are above ILM and below RPE layers.

Wilkins et al. [17] investigated the problem from another point of view. They discarded color information in the first step of the algorithm and determined the major layer boundaries. To improve image quality and filter noise, they used the combination of a median filter and a bilateral filter during preprocessing. After binarization, the method determined the boundaries of the remaining possible cyst segments and they defined three conditions based on empirical studies. They investigated the extent of the objects, the degree of scattering between the intensity of the pixels in the segment, and whether the object is located between the ILM and RPE layers.

2.2 Proposed Method

OCT images are affected by distortions like “shadowing” by blood vessels, that may yield to false detections. In the first step of our proposed algorithm, we improve the image quality by noise filtering and contrast enhancement using a fuzzy operator [3]. This operation can highlight major retinal layers. We analyze vertical profiles of the filtered image and large intensity steps in pixel density are assumed to correspond to the change of tissue. The used fuzzy function is defined as

$$\kappa_{\nu}^* = \frac{1}{1 + \frac{1-\nu}{\nu} \left(\frac{1-x}{x} \frac{\nu}{1-\nu} \right)^{\lambda}}, \quad (1)$$

where ν is a threshold, x is pixel intensity and λ denotes the sharpness of the filtering. The function κ_{ν}^* can highlight boundary layers while suppressing noise. We determine dynamically the input parameter ν in a simple way. We sample from the top range of the image and calculated average intensity for this ROI. The parameter λ was set to 3 empirically. Figure 2 shows an example of applying the function κ_{ν}^* .

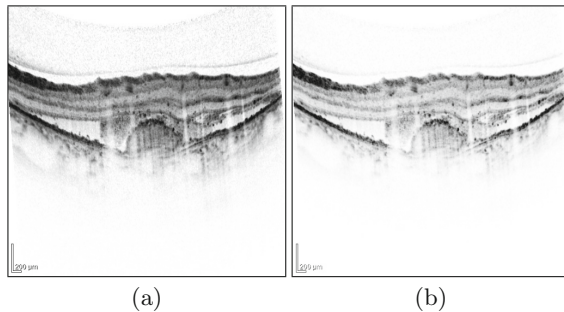


Fig. 2. Sample OCT image before (a) and after (b) applying the function κ_{ν}^* . It filters out the noise outside the retina and highlights boundary layers, so it is easier to delineate them.

After filtering, we divide the image into bars with fixed width. A bar consists of 10 contiguous pixel columns and we calculate horizontal projections of each bar to determine boundaries. One of the main steps of our proposed method is to analyze the vertical profiles. This signal is usually noisy, so there is a need for filtering the data. We use the Savitzky-Golay filter [15] which is a smoothing digital filter. This filter is effective in preserving the relevant high frequency components of the signal, which is an important aspect for our detection method.

Determining the outer layer boundary is harder than that of the inner boundary, because Choriocapillaris and Chorodoidal vessels are located under the RPE layer. The intensity of these regions vary, so several peaks appear in the projections. Fortunately, in most cases, these minimum points are not prominent, and do not cause problem in choosing the right locations. The algorithm chooses the most important local minimum from the projected data to identify the possible inner and outer layer. In the next step, we filter out the outliers and we fit a curve to the remaining points.

After we determined the boundary layers, the next step is the segmentation of fluid and cyst areas. It can be observed that these regions appear as spots with brighter intensity in the image. For processing, we use the inverse of the signals, because our medical colleagues used the inverted presentation of images for visual assessment and also exported the image data for us in this format. The zones of the disease and the intensity of the vitreous body of the eye are almost within the same range (if distortions are not considered). Anisotropic diffusion is used to eliminate various errors from imaging or blood shadows. Using the filter, it is more apparent that some parts of the retina are within a given intensity range, so we quantize the grayscale image into five intensity levels. Our observations showed that the layers of the retina are only in some intensity ranges, so this operation facilitates the separation of the 8–10 main retinal layers. During the binarization, we keep the brightest points, because we know that the reflectivity of the cyst similar to that of the vitreous body. By this step, we create a mask for the active contour process. To achieve the appropriate segmentation result, the input parameters of the model were given based on empirical studies. After that, there may be holes in some objects, so we use hole filling. Figure 3 illustrates these steps.

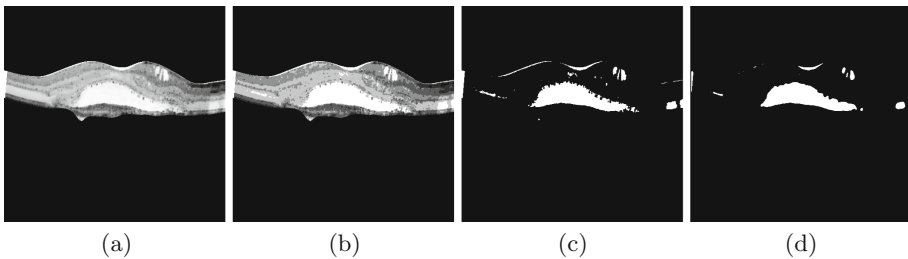


Fig. 3. Intermediate stages of the proposed algorithm. (a) Anisotropic diffusion, (b) Quantization, (c) Binarization, (d) Result of active contour.

So far we have identified possible important segments. Next, we need to separate cyst, fluid and false segments from each other. This step may be omitted in some cases, when there is no object in the picture that would be detected as a pathological mutation. We developed a condition set for filtering cysts and fluid regions. We test the fulfillment of the four criteria for classification at the object level.

The greater distortion of layers was observed in those parts where these symptoms appear, so firstly, we examine whether the actual layer is distorted or not. This plays an important role in distinguishing between cysts and fluid areas. To determine whether the layers are creased or not, we use the top layer boundary points. We calculate the minimal y coordinate of the top points. For this, the method does not take into account the left and right 25% of the image. The sides of the image may not contain information, because of the image registration, and large distortions also may appear in these parts of the images. In Fig. 4 we illustrate these mentioned effects. Next, we search the minimal y coordinate of the top points, divide this image into two parts along the established peak, and we investigate the maximal y point on both sides. We estimate the degree of creasing. Sometimes there is no change in the middle of the image, so we determined a threshold to decide if there is any crease in the slice or not. The threshold for the minimum y point and the given maximum y point was defined in 5 pixels experimentally.

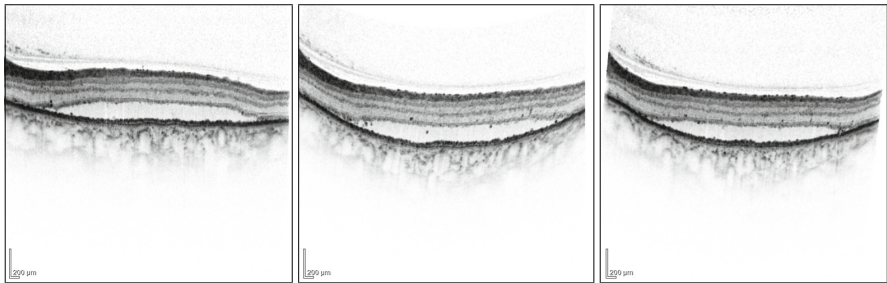


Fig. 4. Example of distortion in consecutive slices.

Various a priori information can be used to distinguish between cysts and fluid areas, and to filter out the false segments. We investigated a condition system and we considered where the object is located within the retina, what is the extent and the shape of the object, and whether the layer is distorted or not. Fluid areas have larger extent and they are located close to the bottom layer boundary or may appear in the distorted area. In the case of higher distortion of the retina, they may also appear on the left or right side. When determining fluid regions, it is also important to examine the cases where there is no creasing in the layer. In the case of cysts, we need to find objects with oval shapes and the observations show that these segments are in the increased zone. Contrary

to fluid regions, cysts are found in higher layers. We distinguish the symptoms based on these characteristics. False segments may also appear in the image, but these are small objects, so we can remove them easily with an area-based filtering. The key stages of the procedure are summarized in Fig. 5.

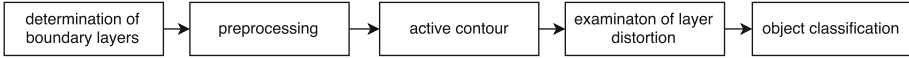


Fig. 5. Flowchart of the proposed algorithm.

3 Evaluation

3.1 Image Data

Our evaluation dataset contained 11 Heidelberg Spectralis OCT scans of wet age related macular degeneration patients treated with anti-VEGF intravitreal injections. The scanning parameters are: 49 scan pattern, pattern size: 5.8×5.8 mm, distance between B-scans: $121 \mu\text{m}$, size X: 512 pixel, size Z: 496 pixel, the pixel size is $11.44 \mu\text{m}$ and $3.87 \mu\text{m}$ in X and Z direction, respectively.

Manual ILM and RPE layer segmentation was performed by ophthalmologists for 7 image sequences. This was considered as a ground truth for evaluating the boundary layer detection method on these 7 image volumes.

3.2 Results and Discussion

We implemented our proposed method in MATLAB, using the Image Processing Toolbox. We evaluated our retinal layer detection algorithm in two different ways. We compared the results of our algorithm against the manual delineations and we also compared the proposed method for automatic detection of subretinal fluid and cyst with some published methods from the literature.

Firstly, we consider the result of localization of major layers. We calculated the mean, maximum and standard deviation of boundary errors for every surface. The 7 curves shown in Fig. 6 depict the error histogram for those OCT volumes where manual annotation was available. Each curve aggregates the boundary errors in the 49 scans (slices) of a study. It can be observed that the largest error is between 1 and 4 pixels in most cases and Table 1 asserts to this statement. As presented in Table 1, the maximal distance between manually segmented and automatically detected layer boundary is 19 pixels (ca. $73.5 \mu\text{m}$). This deflection comes from two sources, namely, the substantial jumping between B-scans and layer distortions due to the disease. Unfortunately, we could not exploit 3D information directly to segment the retina layers, because there are some anomalies among slices of the OCT volume, due to the image acquisition and registration process (within the device's software).

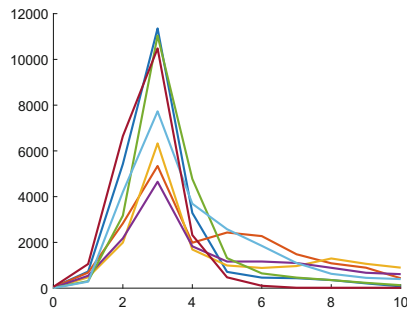


Fig. 6. Error histogram of 7 image sequences.

Table 1. Summary of mean, standard deviation and maximum error (in pixels) between manually segmented and automatically detected layers in 7 annotated OCT image sequence.

		Seq_02	Seq_03	Seq_04	Seq_05	Seq_06	Seq_07	Seq_08	All
Mean	Wilkins	1.90	1.86	2.90	1.88	9.00	1.22	10.91	4.24
	Wieclawek	6.86	10.34	11.20	17.82	6.88	5.42	7.11	9.38
	Proposed	2.01	2.10	1.44	1.96	2.39	2.17	1.83	1.98
Std. dev.	Wilkins	3.34	2.05	2.67	1.62	2.73	0.18	2.93	2.28
	Wieclawek	7.92	5.59	5.92	6.31	5.97	4.78	5.39	5.98
	Proposed	1.56	0.69	0.65	0.80	1.63	0.64	0.65	0.94
Maximum	Wilkins	26	19	17	17	25	11	25	20
	Wieclawek	35	31	24	22	21	20	19	24.57
	Proposed	17	15	15	19	18	15	17	16.57

In Sect. 2, we have presented two methods from the literature for the segmentation of cysts, as well as our proposed method, which is also suitable for delineating fluid areas. Unfortunately, expert annotation was not yet available to evaluate segmentation results, so we compare visually the outputs of the algorithms.

Figure 7 illustrates segmentation results by the algorithms in some slices. The method developed by Wieclawek detected fewer possible cyst regions, which may be due to the fact that the given threshold only keeps the actual light points. The disadvantage of this is, that important areas may be lost during processing. The other method from literature by Wilkins yields almost the same segmentation results, but in many cases it holds false objects, because the thresholds are not dynamically defined. In contrary, our method uses dynamic requirements based on a priori information.

We tested the re-implemented earlier published methods on images in which cysts and fluids may also appear. They can also detect these regions because these segments are also lighter object in the image. Our algorithm, however, as it can be seen in Fig. 7(c), can also distinguish these two types of structures from each other.

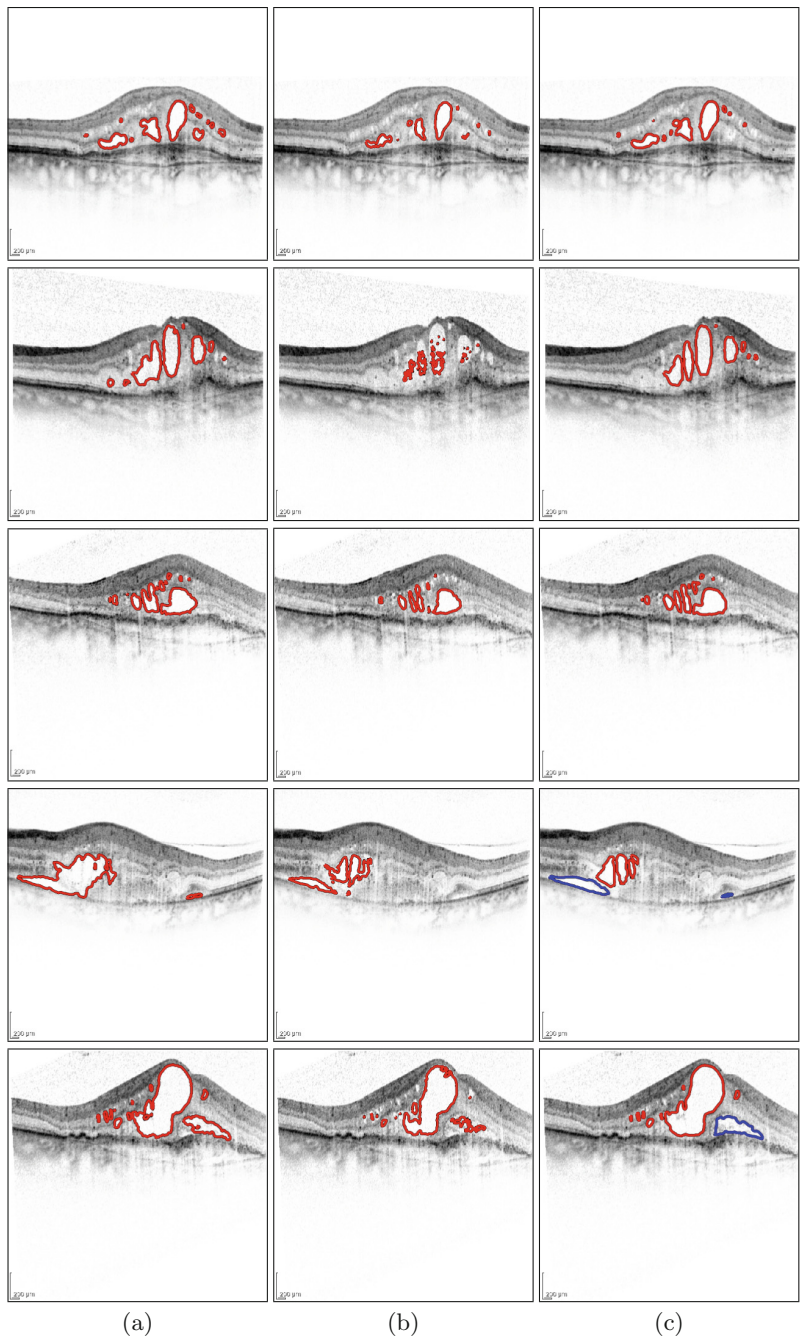


Fig. 7. Illustration of detected cyst (red curve) and subretinal fluid (blue curve) regions by the described algorithms. Columns: (a) Wilkins et al., (b) Wieclawek, (c) Proposed method. Top three rows contain only cysts, while the bottom two scans also have fluid areas. (Color figure online)

4 Conclusion

We presented a novel algorithm for the detection of subretinal fluid areas and cysts and we compared it with two methods from the literature for cyst localization. After having seen the results, the medical colleagues believe that digital image processing can help the quantitative assessment of the OCT features of AMD by providing automatic tools to detect abnormalities and to describe by objective metrics the current state and longitudinal changes during disease evolution and treatment. Using SD-OCT to follow up changes of subretinal fluid and cysts volume will become a useful tool in detecting subtle changes during the treatment process. Further studies are planned to evaluate these new tools in a cohort of AMD patients.

Acknowledgments. This work was supported by the NKFIH OTKA [grant number K112998]. The authors thank Dr. József Dombi for suggesting the use of fuzzy operators for the image preprocessing step.

References

1. Abhishek, A.M., Berendschot, T., Rao, S.V., Dabir, S.: Segmentation and analysis of retinal layers (ILM & RPE) in Optical Coherence Tomography images with Edema. In: 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES), pp. 204–209 (2014)
2. Chiu, S.J., Li, X.T., Nicholas, P., Toth, C.A., Izatt, J.A., Farsiu, S.: Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt. Express* **18**(18), 19413–19428 (2010)
3. Dombi, J.: Modalities. In: Melo-Pinto, P., Couto, P., Serôdio, C., Fodor, J., De Baets, B. (eds.) *Eurofuse 2011. Advances in Intelligent and Soft Computing*, vol. 107, pp. 53–65. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24001-0_7](https://doi.org/10.1007/978-3-642-24001-0_7)
4. Ferrara, N.: Vascular endothelial growth factor: basic science and clinical progress. *Endocr. Rev.* **25**(4), 581–611 (2004)
5. Fu, D., Tong, H., Luo, L., Gao, F.: Retinal automatic segmentation method based on prior information and optimized boundary tracking algorithm. In: *Proceedings of SPIE*, vol. 10033, p. 100331C: 1–100331C: 6 (2016)
6. Garvin, M.K., Abramoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans. Med. Imaging* **28**(9), 1436–1447 (2009)
7. González, A., Remeseiro, B., Ortega, M., Penedo, M.G., Charlón, P.: Automatic cyst detection in oct retinal images combining region flooding and texture analysis. In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pp. 397–400 (2013)
8. Hassan, B., Raja, G., Hassan, T., Akram, M.U.: Structure tensor based automated detection of macular edema and central serous retinopathy using optical coherence tomography images. *J. Opt. Soc. Am.* **33**(4), 455–463 (2016)
9. Hee, M.R., Bauman, C.R., Puliafito, C.A., Duker, J.S., Reichel, E., Wilkins, J.R., Coker, J.G., Schuman, J.S., Swanson, E.A., Fujimoto, J.G.: Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. *Ophthalmology* **103**(8), 1260–1270 (1996)

10. Katona, M., Nyúl, L.G.: An approach to the quantitative assessment of retinal layer distortions and subretinal fluid in SD-OCT images. *Acta Cybern.* Accepted for publication (2017)
11. Kovach, J.L., Schwartz, S.G., Flynn Jr., H.W., Scott, I.U.: Anti-VEGF treatment strategies for wet AMD. *J. Ophthalmol.* **22**, 786870:1–786870:7 (2012)
12. Lang, A., Carass, A., Hauser, M., Sotirchos, E.S., Calabresi, P.A., Ying, H.S., Prince, J.L.: Retinal layer segmentation of macular OCT images using boundary classification. *Biomed. Opt. Express* **4**(7), 1133–1152 (2016)
13. Lim, J.: *Age-Related Macular Degeneration*. CRC Press, Boca Raton (2012)
14. Lu, S., Cheung, C.Y.L., Liu, J., Lim, J.H., Leung, C.K.S., Wong, T.Y.: Automated layer segmentation of optical coherence tomography images. *IEEE Trans. Biomed. Eng.* **57**(10), 2605–2608 (2010)
15. Schafer, R.W.: What is a Savitzky-Golay filter? *IEEE Sig. Process. Mag.* **28**(4), 111–117 (2011)
16. Wieclawek, W.: Automatic cysts detection in optical coherence tomography images. In: 2015 22nd International Conference on Mixed Design of Integrated Circuits Systems (MIXDES), pp. 79–82 (2015)
17. Wilkins, G.R., Houghton, O.M., Oldenburg, A.L.: Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Trans. Biomed. Eng.* **59**(4), 1109–1114 (2012)
18. Yazdanpanah, A., Hamarneh, G., Smith, B., Sarunic, M.: Intra-retinal layer segmentation in optical coherence tomography using an active contour approach. In: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 649–656. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-04271-3_79](https://doi.org/10.1007/978-3-642-04271-3_79)

IV.



Automatic Detection and Characterization of Biomarkers in OCT Images

Melinda Katona¹, Attila Kovács⁴, László Varga¹, Tamás Grósz²,
József Dombi³, Rózsa Dégi⁴, and László G. Nyúl¹(✉)

¹ Department of Image Processing and Computer Graphics,
University of Szeged, Szeged, Hungary
{mkatona, vargalg, nyul}@inf.u-szeged.hu

² MTA-SZTE Research Group on Artificial Intelligence,
University of Szeged, Szeged, Hungary
groszt@inf.u-szeged.hu

³ Department of Computer Algorithms and Artificial Intelligence,
University of Szeged, Árpád tér 2, Szeged 6720, Hungary
dombi@inf.u-szeged.hu

⁴ Department of Ophthalmology, University of Szeged,
Korányi fasor 10-11, Szeged 6720, Hungary
{kovacs.attila, degi.rozsa}@med.u-szeged.hu

Abstract. Optical Coherence Tomography (OCT) is one of the most advanced, non-invasive method of eye examination. Age-related macular degeneration (AMD) is one of the most frequent reasons of acquired blindness. Our aim is to develop automatic methods that can accurately identify and characterize biomarkers in OCT images, related to AMD. We present methods for quantizing hyperreflective foci (HRF) with deep learning. We also describe an algorithm for determining pigmentepithelial detachment (PED) and localizing outer retinal tubulation (ORT) that appears between the layers of the retina.

Keywords: Age-related macular degeneration · Biomarker
Pigmentepithelial detachment · Hyperreflective foci
Outer retinal tubulation · Optical coherence tomography

1 Introduction

Age-related Macular Degeneration (AMD) is a health problem worldwide, that is the leading cause of vision loss in the Western World. While symptoms are rare in patients below 50 years of age, an increasing prevalence of AMD can be detected in the elderly population. AMD means degeneration of the macula, the region of the retina responsible for central vision. Since only this specific part of the retina is affected by AMD, untreated patients lose their fine shape- and face recognition, reading ability, and central vision [4]. AMD can be divided

into two subtypes; the dry (non-exudative) and the wet (exudative, neovascular) form. The latter one causes rapid and serious visual impairment and accounts for 10% of the cases. In this more acute, neovascular type of the disease, abnormal angiogenesis causes fluid and blood leakage into the retinal layers thus resulting in photoreceptor lesion. Albeit the exact pathomechanism of the disease is still unclear, it is known that the vascular endothelial growth factor (VEGF) plays crucial role in the pathogenesis [10]. The first choice of treatment in neovascular AMD is anti-VEGF intravitreal injection, a periodic injection into the eye.

During the last decade, optical coherence tomography (OCT) has become a basic tool in diagnosing and monitoring neovascular AMD and its response to anti-VEGF treatment. With the help of OCT, we are capable of detecting the layers of the retina, and also the effects of the disease, the so-called OCT biomarkers, such as subretinal/intraretinal fluid accumulation, pigmentepithelial detachment (PED), outer retinal tubulation (ORT) or hyperreflective foci (HRF) (Fig. 1). These markers help the clinical decision-making process for observing/treating/re-treating a patient. To improve the treatment procedure, there is a need for more precise measurements, hence our aim was to create algorithms which can automatically identify and quantify some of the above mentioned biomarkers, namely PED, ORT and HRF.

A large number of publications in the scientific literature deal with cysts, subretinal fluid detection and retinal layer segmentation, however only a small number of papers are available on PED. Haq et al. [3] defined 12 retinal layer with a multi resolution graph-search method and the PED was calculated from the relative position of the lower two layers. The algorithm of Shi et al. [9] applied machine learning for PED detection. We have not come across any papers about automatic segmentation of ORT. In case of the HRFs there are some simple techniques for the detection (see, e.g., [8]), but we have not found any methods in use related to AMD.

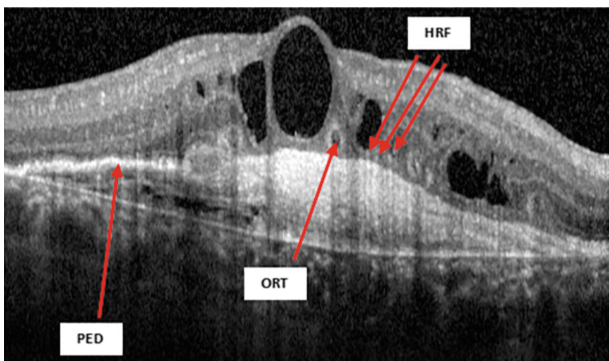


Fig. 1. Optical Coherence Tomography (SD-OCT) image of the retina with biomarkers

2 Materials and Methods

In this section, we present several algorithms to detect HRF, PED and ORT. The OCT images were acquired using Heidelberg Spectralis (Spectralis, Heidelberg Engineering, Heidelberg, Germany) OCT scans on wet age-related macular degeneration. Patients were either treated naively or with anti-VEGF intravitreal injections. The annotated images contained many biomarkers, such as subretinal fluid, PED, subretinal and hyperreflective material. The currently used data consist of images of different patients taken at different times. All image sequences consisted of 49 slices taken with a 6×6 mm pattern size and $122 \mu\text{m}$ slice distance. The slices were generated by averaging 30 frames on each B-scan. Slices had a resolution of 512×496 pixels with pixel sizes 11.45 and $3.87 \mu\text{m}$ and a quality score above 16 dB .

We evaluated our PED and ORT segmentation method by comparing the results to manual segmentations of ophthalmologists for 2 image sequences. The database of HRF consisted of 11 image sequences taken from 7 clinical patients. The annotation was performed by two clinical doctors, independently marking the hyperreflective foci.

2.1 PED Detection and ORT Localization

Pigment epithelium detachment (PED) can be an important medical feature of the disease. PED estimation can be made relatively easily after the retinal pigment epithelium (RPE) layer is detected. Determining the outer boundaries, such as the internal limiting membrane (ILM) and RPE layers are defined in our previous work [5]. We calculated a possible normal layer boundary using the known RPE boundaries to characterize the detachment. We took 1% of the points from both sides of the image and we fitted a smoothed cubic spline to these points, giving the possible boundary.

In many cases, even for doctors, it is difficult to see the ORTs, they can only guess their place and extent. Consequently, an approximate segmentation result can also help.

The ORT has hyperreflective contour and contains hyper- hyporefective points. The procedure is based on finding hyperreflective points. As we can see in Fig. 1, the input image is noisy, so we used a Wiener filter with a 3×5 pixel kernel. Our input image size was 509×496 , so a smaller kernel was sufficient. Reflective points were localized using a Hessian detector [1]. Then, we performed a non-maximum suppression and considered the 100 highest points. The result image may also contain a several points that are not relevant for ORT localization.

For filtering the false points, we used some prior information about the biomarkers. Firstly, we calculated the retina thickness and we kept only those points which are located in the lower third of the retina. Since we know that ORT is close to the RPE layer and in the distorted retina region or nearby surroundings, by estimating the beginning and end of the distortion on the slices additional points can be removed. We used our published algorithm [5] to detect

the extent and location of the distortion of the retina and the detection of cyst and liquid areas. Limiting the specific extent of the ORT in many cases is very difficult, because there is no clear distinction between the hyperreflective wall and its surroundings. We performed adaptive histogram equalization in the image so that at least a part of the possible contours became separable by hysteresis thresholding. The two threshold values are given as the lower and upper third of the maximum intensity value in the image. At the end of the last filtering, we kept only those points which were part of an object in the binary image. Then, we calculated distance map for the points, thresholded and finally computed the convex hull of the objects. The key stages of the procedure are summarized in Fig. 2.

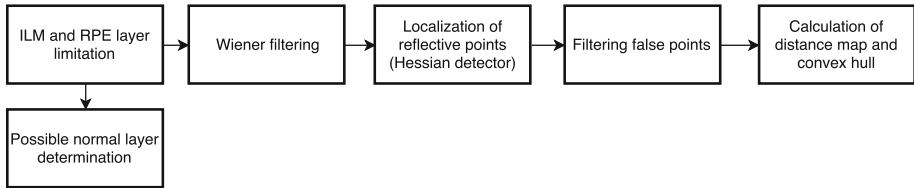


Fig. 2. Flowchart of the proposed PED detection and ORT localization algorithm.

2.2 Quantifying Change in HRFs Using Deep Neural Networks

One of the aims of this work was to quantify the number of HRF pixels in the OCT slices. Although this included the segmentation of HRF pixels, our goal was not a detection, but rather the tracking slight changes in the amount of HRFs. This can help the doctors to objectively track the progress of the disease and aid the treatment planning. For this task, we used Deep Neural Networks (DNNs), which were trained using annotated images by ophthalmologists.

We applied several types of networks, namely the standard Artificial Neural Networks with one hidden layer, Deep Rectifier Neural Networks (DRNs) [2] and Convolution Neural Networks (CNNs) [7], which were successfully used in many previous medical studies [6]. The goal of these networks was to classify one single pixel from a given feature vector. The full image classification was performed by classifying each pixel of the image, separately. We gave the networks two types of input data. The first one was the raw pixel data, which consisted of raw pixel intensities in 25×25 pixel vicinity of the pixel to classify. The other type of input consisted of feature vectors extracted from the OCT images. These features were:

- Weighted sum of pixel intensities in the neighborhood weighted with a Laplacian of Gaussian (LoG) kernels of different σ values. LoG filter σ value ranged from 1 to 2 with a step of 0.1;
- Distance of the pixel from the ILM and RPE layers;
- Distance from subretinal fluid;

- To help the networks recognize shadows of veins we added 25 average intensity values of 40 pixel long vertical strips under the pixel in question, i.e., $\left\{ \sum_{j=1}^{40} I(x-i, y-j), \quad i \in \{-12, -11, \dots, 11, 12\} \right\}$, where $I(x, y)$ is the pixel of question with x and y coordinates.

The marked targets by ophthalmologist were the bright spots of the images having equal or higher reflectivity than the RPE band, and a diameter of approximately $20 - 40 \mu\text{m}$ (2–4 pixels). The annotation consisted of the delineation the HRF pixels of the slices. Before the training phase, we separated the dataset into two partitions, 7 out of 11 sequences were used for training the networks. The other 4 sequences were kept for testing. The images in the test set were taken from 4 different patients. The data of these patients were not included in the training dataset in any way (i.e., the training data set was taken from other patients). Within the training dataset, we used 1 out of every 7 slices for development purposes (i.e. hyperparameter tuning).

We trained an ANN and a DRN using only the pixel data. Furthermore, we also changed the network structure by splitting the first hidden layer of the DRN; half the neurons were connected to the raw pixel input and the other half was connected only to the extracted features.

The structure of the networks were determined empirically by seeking the structures giving the best results. The ANNs had only one hidden layer and 7000 hidden neurons, the DRNs had 5 hidden layers each having 1000 rectifier neurons. To train the ANNs and the DRNs we applied stochastic gradient descent (i.e. backpropagation) training with a mini-batch size of 100. The initial learn rate was set to 0.001, which was halved after each iteration if the performance on the validation set did not improve. During the preliminary experiments we found that the optimal value of the sampling parameter (λ) was 0.8.

The CNN had 3 convolutional layers having a kernel size of 5×5 and output size of 32, 32 and 64, respectively, followed by a fully connected layer of 2 neurons, and a softmax layer. The CNN was trained with backpropagation method with a fixed learning rate of 0.001, momentum of 0.9, and weight decay of 0.004. The batch size was 128. Before training and evaluation we normalized the data to the interval $[-1, 1]$. The net was initialized with random values of uniform distribution, and we did not use any pre-training.

3 Evaluation and Results

3.1 PED and ORT Detection

For calculating accuracy, we used the Jaccard coefficient of similarity, which measures the overlap of the annotated segment and the detected biomarker region. The result of the pigmentepithel detachment localization depends on the location of the pre-determined RPE layer. It may also appear as subretinal hyperreflective material, which is not distinguished from the detachment, so this appears as a false detected region in evaluation. Figure 3 represent the Jaccard values for

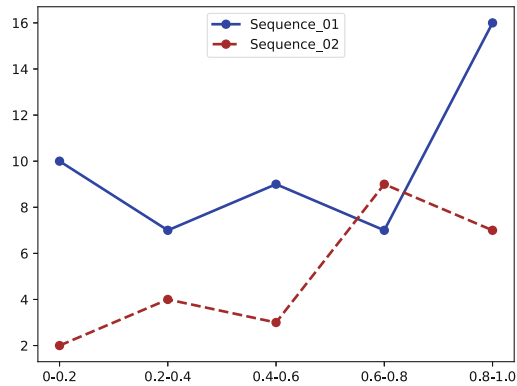


Fig. 3. Jaccard histogram of 2 image sequences. Horizontal axis represents Jaccard values and vertical axis is the number of slices with Jaccard index falling into the given ranges.

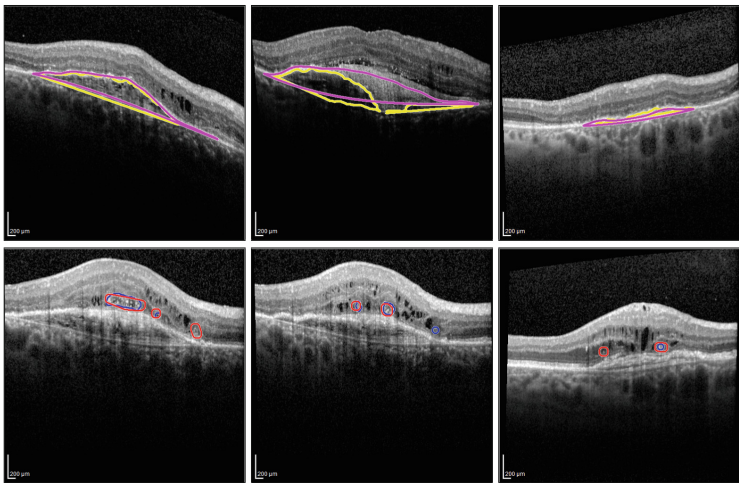


Fig. 4. Illustration of detected and annotated PED and ORT. Upper images show PED regions (magenta - detected, yellow - annotated) and lower pictures represent ORT (red - detected, blue - annotated). (Color figure online)

the two annotated sequences. It can be concluded that the procedure in most cases closely approximates the PED area designated by the ophthalmologist.

In the case of ORT, we analyse the result in two different ways, because not only the localization is important for the ophthalmologist, but sometimes it is also appropriate to determine, if ORT appears on images or not. We achieved a Jaccard value of 0.530 for the two sequences, but we filtered images which do not contain ORT with 97.6% success. Figure 4 illustrates the results of the proposed methods.

Running time of the method was 0.401 ms, PED detection was 0.006 ms on average and ORT localization took 0.091 ms per slice.

3.2 Tracking the Amount of HRFs

The small size and the limited amount of the HRFs made the use of simple (F1 score, or Dice coefficient-based) pixel-wise evaluation metrics meaningless, since only a 1 pixel misalignment in the segmentation would show significant error on the 2–4 pixel sized HRFs while the result is still useful. Also, segmenting HRF-s was not the goal of this study as the doctors wanted a method to track the changes in the HRF amounts. For the comparison, we evaluated the performance of the nets by calculating the Pearson’s correlation coefficient between the number of HRF pixels on the automatic and manual segmentations. Using 196 slices, we pairwise compared the pixel counts got by the automatic segmentations to those of the gold standard annotation. As a baseline for the study, we also calculated the correlation between the manual annotations of the two doctors. The results can be seen in Table 1.

Table 1. Pearsons correlation coefficient between the automatic and manual segmentations. In columns one can see the comparison of segmentations sources (NN - Neural network; MD.# - physician).

Data (optimizer goal)	NN \Leftrightarrow MD.1	NN \Leftrightarrow MD.2	MD.1 \Leftrightarrow MD.2
ANN - pixel (accuracy)	0.778	0.789	0.812
ANN - pixel & feature (accuracy)	0.698	0.775	
DRN - pixel (accuracy)	0.812	0.783	
DRN - pixel & feature (accuracy)	0.796	0.782	
split DRN - pixel & feature (accuracy)	0.812	0.790	
DRN - pixel & feature (dice)	0.802	0.788	
CNN - pixel (accuracy)	0.845	0.862	

We argue, that if a neural network can reach a higher correlation than the one between the annotations of the doctors (0.812), then its outputs are useful, since they reflect the number of HRFs on a slice as well as the annotation of a medical expert would. In most of the cases, the networks were able to provide good results as almost all of them achieved a correlation value above 0.78. Interestingly, using the extracted features did not improve the quality of the segmentation, suggesting that the nets learned to extract those informations from the raw data. Furthermore, switching the loss function to the Dice loss was neither beneficial.

The best results were achieved by the CNN, with correlation coefficients over 0.845. This indicates that it can produce an acceptable, and reliable quantization. We should also note that the hand made full delineation of HRFs in an OCT

sequence leading to an accurate quantization is time consuming, and our method can give a quick automatic estimation of the amount of HRFs.

Based on the above results, we can say that our methods are capable of tracking the number of HRFs in OCT images, and they could aid the doctors during the planning of the treatment, by speeding up the decision making process.

4 Conclusions

We have described procedures for analysing some OCT features of AMD patients. The methods include detection of PED, localization of ORT between retinal layers, and the quantization of HRF. We determined the possible normal layer boundary for the characterization of PED and we introduced a method to localize ORT. We compared our results with the annotated data by medical colleagues. Furthermore, automatic detection of these specified biomarkers can be used not only in AMD patients, so it can help the doctor during patient examination.

Our HRF quantization method achieved a correlation coefficient as good as the one between the annotations by the ophthalmologists. Hence, we argue that they are suitable for aiding the diagnosis and treatment planning process.

Acknowledgements. We would like to thank the NVIDIA Corporation for the donation of the Tesla K40 GPU used for this research. Tamás Grósz was supported by the ÚNKP-17-3 New National Excellence Programme of the Ministry of Human Capacities.

References

1. Beaudet, P.: Rotationally invariant image operators. In: International Joint Conference on Pattern Recognition, Kyoto, Japan, pp. 579–583 (1978)
2. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, vol. 15, pp. 315–323. PMLR (2011)
3. Haq, A., Wilk, S.: Detection of wet age-related macular degeneration in OCT images: a case study. *Innovations in Biomedical Engineering*. AISC, vol. 623, pp. 43–51. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-70063-2_5
4. Hee, M.R., et al.: Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. *Ophthalmology* **103**(8), 1260–1270 (1996)
5. Katona, M., Kovács, A., Dégi, R., Nyúl, L.G.: Automatic detection of subretinal fluid and cyst in retinal images. In: Battiato, S., Gallo, G., Schettini, R., Stanco, F. (eds.) ICIAP 2017. LNCS, vol. 10484, pp. 606–616. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68560-1_54
6. Kayalibay, B., Jensen, G., van der Smagt, P.: CNN-based Segmentation of Medical Imaging Data. CoRR abs/1701.03056 (2017)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012)
8. Mokhtari, M., Kamasi, Z.G., Rabbani, H.: Automatic detection of Hyperreflective foci in optical coherence tomography b-scans using morphological component analysis. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1497–1500 (July 2017)

9. Shi, F., et al.: Automated 3-D retinal layer segmentation of macular optical coherence tomography images with serous pigment epithelial detachments. *IEEE Trans. Med. Imaging* **34**(2), 441–452 (2015)
10. Velez-Montoya, R., et al.: Current knowledge and trends in age-related macular degeneration: today's and future treatments. *Retina* **33**4, 1487–1502 (2013)

V.



Segmentation of Subretinal Hyperreflective Material and Pigment Epithelial Detachment Using Kernel Graph Cut

Melinda Katona¹, Attila Kovács², Rózsa Dégi², and László G. Nyúl¹(✉)

¹ Interdisciplinary Excellence Centre,
Department of Image Processing and Computer Graphics,
University of Szeged, Árpád tér 2, Szeged 6720, Hungary
{mkatona,nyul}@inf.u-szeged.hu

² Department of Ophthalmology, University of Szeged,
Korányi fasor 10-11, Szeged 6720, Hungary
{kovacs.attila,degi.rozsa}@med.u-szeged.hu

Abstract. Optical Coherence Tomography (OCT) is one of the most advanced, non-invasive method of eye examination. Age-related macular degeneration (AMD) is one of the most frequent reasons of acquired blindness and it has two forms. Our aim is to develop automatic methods that can accurately identify and characterize biomarkers in SD-OCT images, related to wet AMD. Detection of biomarkers can be challenging because of their variable shape, size, location and reflectivity. In this paper, we present an automatic method to localize subretinal hyperreflective material (SHRM) and pigment epithelial detachment (PED) via kernel graph cut. The proposed method is evaluated using an annotated dataset by ophthalmologists. The Dice coefficient was 0.81 (± 0.11) in the case of PED and 0.77 (± 0.11) for SHRM. In many cases, the ophthalmologist cannot clearly determine the exact location and extent of the biomarkers, so our achieved results are promising.

Keywords: Age-related macular degeneration ·
Subretinal hyperreflective material · Pigment epithelial detachment ·
Optical coherence tomography

1 Introduction

Age-related Macular Degeneration (AMD) is a common eye condition in people above 50 years worldwide, currently affecting 170 million people globally. AMD means degeneration of the central part of the retina, responsible for fine vision. Untreated patients lose their shape- and face recognition, reading ability, and central vision [3]. Two subtypes of AMD are known; the dry (non-exudative) and the wet (exudative, neovascular) form. From these two subtypes the wet form

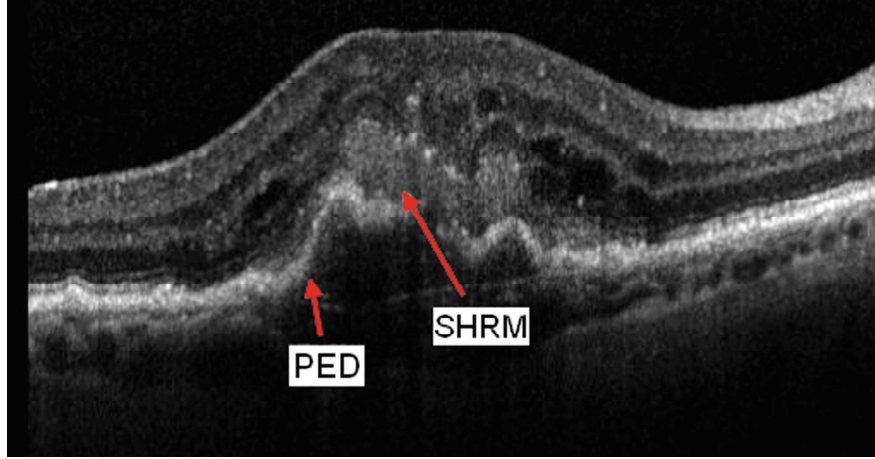


Fig. 1. Optical Coherence Tomography (SD-OCT) image of the retina with PED and SHRM biomarkers. Our medical experts use it in inverted display in daily routine.

causes rapid and serious visual impairment accounting for 10% of the cases. In this neovascular type of the disease, fluid and blood leakage occurs from abnormal angiogenesis from the choroid, leading to photoreceptor damage. Vascular endothelial growth factor (VEGF) plays crucial role in the pathogenesis of the wet form [12], leading to a treatment option by blocking it with periodically repeated anti-VEGF intravitreal injections, the first choice of treatment in neovascular AMD [8]. Nowadays, image processing tasks are solved by training different type of neural networks. Lee et al. [6] concerned with segmentation of PED and SHRM simultaneously. In this approach, CNN-based deep neural network was trained.

In the last decade, optical coherence tomography (OCT) has become a basic imaging technique in diagnosing and following neovascular AMD and its response to treatment, since it provides high-resolution cross-sectional images of the retina. With OCT we can detect not only the layers of the retina, but we are also capable of identifying the detailed morphology and the effects of the disease, the so-called OCT biomarkers, such as subretinal/intraretinal fluid accumulation (SRF/IRF), pigment epithelial detachment (PED), subretinal hyperreflective material (SHRM), outer retinal tubulation (ORT) or hyperreflective foci (HRF) (see, e.g., Fig. 1). These markers help the clinical decision-making process for treating/re-treating/observing a patient [1]. In our previous studies [4, 5], we have already discussed IRF, SRF, ORT, and HRF, the present analysis focussed on PED and SHRM.

More reports have dealt with PED, which is an elevated lesion below the retinal pigment epithelium (RPE). Reports showed that PED should be monitored precisely, and any growth in size urgently treated with anti-VEGF injections, thus preventing sustained vision loss. SHRM can be detected on OCT as a medium- to hyperreflective mass between the neurosensory retinal layers and RPE.

To improve the treatment procedure, there is a need for more precise measurements, hence our aim was to create an algorithm which automatically identifies and characterizes two of the above-mentioned biomarkers, namely the PED and the SHRM.

A large number of publications in the scientific literature deal with the analysis of retinal images. Nowadays, most papers use machine learning methods to solve various image processing problems. Using these approaches, it is possible to classify the retinal disease by using some biomarkers. Srinivasan et al. [10] developed a method which can determine RPE layer after some preprocessing steps to improve image quality. To classify the individual eye diseases, they were calculated HOG descriptors and they used these features to train an SVM. In addition, to differentiate AMD and Diabetic Macular Edema (DME), Sugmk et al. [11] have applied a binary classification based on different threshold results. Some of the procedures work with color fundus images. Mohaimin et al. [7] localized the disease regions via color and boundary based segmentation method. In these cases, the determination of the RPE layer and/or the localization of PED was the target for classification.

In many cases, it is not sufficient for a medical doctor to decide what illness may affect the patient's eyes, but also the extent, numeracy, etc. of the biomarkers. We have already worked on the topic of localization of outer retinal layers [4] and detection of PED [5]. In this paper, we deal with segregation of subretinal hyperreflective material and pigment epithelial detachment. For this, we introduce a graph cut based approach to determine biomarkers. The achieved results were compared with annotated data by ophthalmologists.

2 Materials and Methods

The OCT images were acquired using Heidelberg Spectralis OCT (Spectralis, Heidelberg Engineering, Heidelberg, Germany) on wet age-related macular degeneration. Patients were either treated naively or with anti-VEGF intravitreal injections. The annotated images contained several biomarkers, such as subretinal fluid, PED, Cyst, SHRM. In this study, we used only PED and SHRM annotation and 18 image sequences were available. These are recordings of 12 different patients. These sequences consisted of 49 or 25 slices taken with a 6×6 mm pattern size and $251 \mu\text{m}$ slice distance. Slices had a resolution of 512×496 pixels with pixel sizes 11.74 and $3.87 \mu\text{m}$ and a quality score above 23 dB .

2.1 Graph Cut Based ILM and RPE Layer Segmentation

We worked with 2D images because we could not exploit 3D information directly to segment the retina layers, since there are some anomalies among slices of the OCT volume, due to the image acquisition and registration process (within the device's software). In the literature, a number of procedures are based on graph cut [2] since it is a robust method against other techniques. As a result, no preprocessing of the input data has been performed, although in many cases the

image is very noisy. Graph cut is a semi-automatic method that requires seed points. There are several existing approaches to automatically assigning them. We used kernel k-means to automatically calculate these points. We map data points in the input space onto a high-dimensional feature space using Gaussian radial basis kernel function:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right),$$

where $k(x, y)$ is a point in higher dimensional space. The number of clusters is determined by how noisy the picture is. So, we calculated no-reference image quality score for the image using the Naturalness Image Quality Evaluator (NIQE). This model based on a set of local features extraction from an image then fits the feature vectors to a multivariate Gaussian. In the higher quality images, 5 clusters were isolated empirically, while in the other cases this number increased. In cluster image, value 0 represents the darkest region and $k - 1$ is the brightest, where k is the number of clusters.

We used graph cut [9] to optimize partitioning. This gives a better classification. PED and SHRM are located near the RPE layer, so its detection helps later classification. The ILM elevation can be determined after a simple Otsu thresholding because the foreground and background can be clearly distinguished from that part of the retina in the clustered image. RPE layer has higher reflectivity in the original image, so it has also higher cluster number in the clustered image. Using a method in our previous publication [4], we calculated vertical projection in every 10th column to determine boundary and we chose local minimum from the projected data.

To determine all pixels which can produce the layer, we fitted a curve to the resampled points. Our data probably include outliers, because Choriocapillaris and Choroidal vessels are located under the RPE layer, so The intensity of these regions vary. These salient points may mislead the fitting, so eliminating them is important. The distribution of the points is not a normal, due to the distortions, therefore outliers are defined as elements more than 1.5 interquartile ranges above the upper quartile (75%) or below the lower quartile (25%). We fitted shape-preserving piecewise cubic spline interpolation to determine RPE layer.

2.2 Classification of PED and SHRM

Subretinal Hyperreflective Material, as its name suggests, is likely composed of many components, including fluid, fibrin, blood, etc., and its composition changes over time. So, the reflectivity of SHRM is heterogeneous. By contrast, Pigment Epithelial Detachment has lower intensity, so they can be separated from each other. Nevertheless, in many cases, their location and presence are unclear.

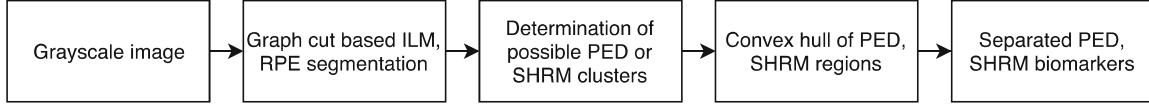


Fig. 2. Flowchart of the proposed method.

The pixels of the PED and the SHRM regions are roughly in the same intensity range, and therefore the results of a proper clustering can form a separate region. These biomarkers are usually located around the RPE layer, along or near the distortion, so clusters were sampled from these areas only. Patients with other lesions may also be observed, so we have also taken into account whether the patient has a cyst or fluid area. We also used our work referred to in Subsect. 2.1 to determine the location of distortion and to detect other biomarkers (cysts, fluid areas). We only investigated in a specific range over the RPE layer. For this, the threshold was determined as follows:

$$T = \begin{cases} \max(CF_{(x, \max(y))}), & \text{if } \max(CF_{(x, \max(y))}) > 0 \\ RT_x * 0.2, & \text{otherwise} \end{cases},$$

where CF represents the binary image with detected cyst and subretinal fluid, x is the actual x position, $\max(y)$ is the actual maximum y position and RT denotes the calculated retinal thickness (from ILM and RPE layer).

The collected clusters produce SHRM regions. Those cluster(s) will be PED area which defined as a non-SHRM segment below the RPE layer. They may also have amorphous shapes because they are made up of different sub-areas, so we have defined their convex shapes to accomplish the final result. The key stages of the procedure are summarized in Fig. 2.

3 Results

We evaluated segmentation ability in terms of overlap between each segmented area by graders. We considered data obtained from medical doctors as ground truth.

3.1 Evaluation Metrics

We used Dice coefficients to measure the ratio of overlap.

$$\text{Dice coefficient} = \frac{2 \cdot |X \cap Y|}{|X| + |Y|},$$

where $|X|$ and $|Y|$ are the numbers of pixels representing the segmented regions, while $|X \cap Y|$ are the overlapping regions in pixels.

In addition, for the further characterization of the procedure, the sensitivity was also calculated.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

where TP denotes the numbers of overlapping pixels, FP represents the numbers of segmented, but not overlapping pixels.

3.2 Localization of PED and SHRM

We implemented our proposed method in MATLAB, using the Image Processing Toolbox, and compared the results with annotated data by ophthalmologists. The location of the possible SHRM has played an important role in determining PED. In many cases, SHRM is not clearly detectable. In spite of all this, as you can see in Fig. 3, in both cases, the average Dice coefficients are above 0.75 and the sensitivity is 0.93 for PED and 0.77 for SHRM. The detection of SHRM is not yet widely investigated in the literature. In contrast, several articles deal with the delimitation of PED. The CNN-based method mentioned in Sect. 1 also determines the biomarkers with similar efficiency. This means that it is not necessary to have a complex system with large image database to segment the affected regions. Most errors are in parts where the ophthalmologist cannot clearly define the area of the given biomarker.

The kernel k-means more computationally than the basic k-means which separates linearly, so the running time is also increasing. This means that the running time of the process is 30s on average. In the Fig. 4, and you can see some examples of automatic results and also the manual annotation for a visual comparison.

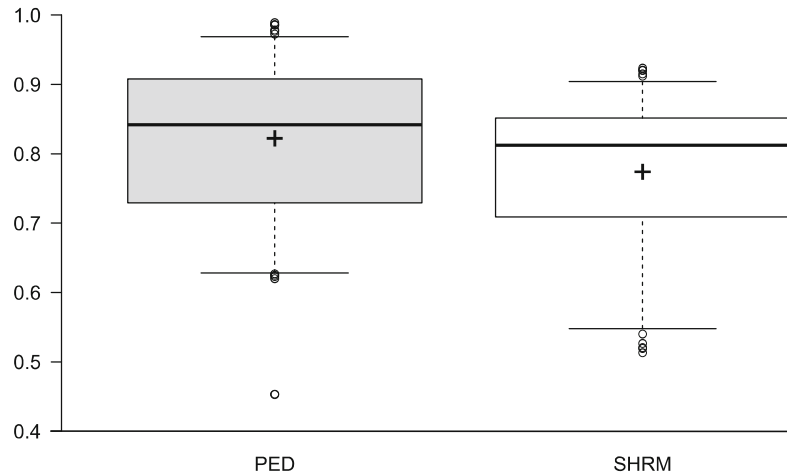


Fig. 3. Box plots representing the Dice coefficients of PED and SHRM generated by proposed method. In both cases, the mean Dice coefficients were greater than 0.7. Black square = mean Dice coefficient; empty circles = outliers; SHRM = subretinal hyperreflective material; PED = pigment epithelial detachment.

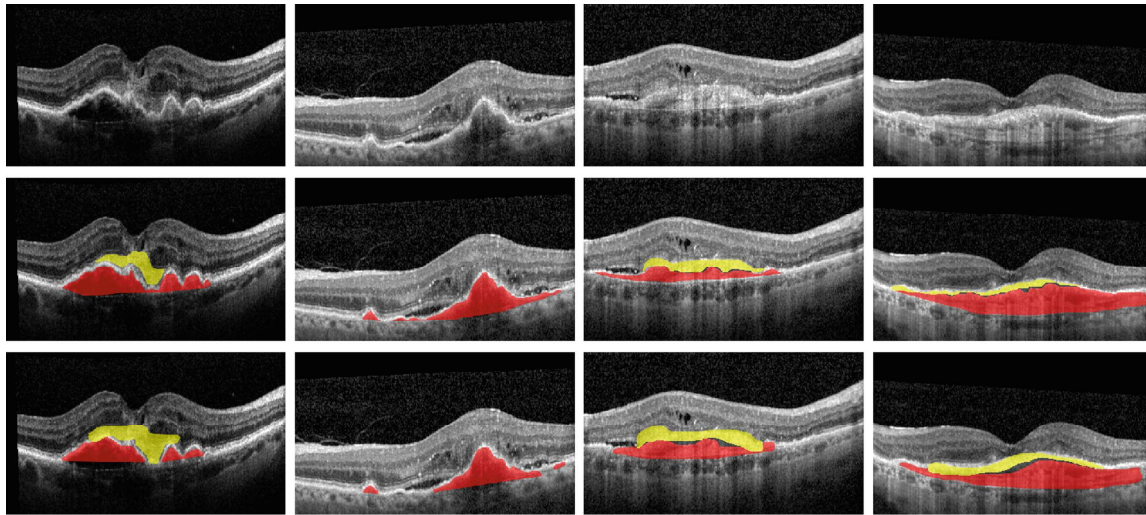


Fig. 4. Illustration of annotated (in middle row) and detected (bottom row) PED and SHRM. Red regions show PED yellow represents SHRM.

4 Conclusion

We introduced a graph cut based approach to detect PED and SHRM regions on SD-OCT B-scans. The examined biomarkers have different reflectivity, but in a significant part of the cases, their extent cannot be clearly defined. In the process, ILM and RPE layers were also determined, which served as the basis for our main goal. The results were compared with data annotated by medical doctors. We calculated measures to quantify features of OCT in point of wet AMD patients. This automated localization method can help the quantitative assessment of the OCT biomarkers by providing automatic tools to detect abnormalities and to describe by objective metrics the current state and longitudinal changes during disease evolution and treatment.

Acknowledgements. Melinda Katona was supported by the UNKP-18-3 New National Excellence Program of the Ministry of Human Capacities.

Ministry of Human Capacities, Hungary grant 20391-3/2018/FEKUSTRAT is acknowledged.

Melinda Katona and László G. Nyúl were supported by the project “Integrated program for training new generation of scientists in the fields of computer science”, No. EFOP-3.6.3-VEKOP-16-2017-0002. The project has been supported by the European Union and co-funded by the European Social Fund.

References

1. Coscas G et al (2009) Clinical features and natural history of AMD on OCT, pp 195–274. Berlin, Heidelberg
2. Dodo BI, Li Y, Eltayef K, Liu X (2018) Graph-cut segmentation of retinal layers from OCT images. In: Proceedings of the 11th international joint conference on biomedical engineering systems and technologies, BIOIMAGING, vol 2. INSTICC, SciTePress, pp 35–42. <https://doi.org/10.5220/0006580600350042>

3. Hee MR et al (1996) Optical coherence tomography of age-related macular degeneration and choroidal neovascularization. *Ophthalmology* 103(8):1260–1270
4. Katona M, Kovács A, Dégi R, Nyúl LG (2017) Automatic detection of subretinal fluid and cyst in retinal images. In: *Image analysis and processing - ICIAP 2017, Proceedings, Part I*, pp 606–616
5. Katona M, Kovács A, Varga L, Grósz T, Dombi J, Dégi R, Nyúl LG (2018) Automatic detection and characterization of biomarkers in OCT images. In: *Image analysis and recognition*, pp 706–714
6. Lee H, Kang KE, Chung H, Kim HC (2018) Automated segmentation of lesions including subretinal hyperreflective material in neovascular age-related macular degeneration. *Am. J. Ophthalmol.* 191:64–75. <https://doi.org/10.1016/j.ajo.2018.04.007>. <http://www.sciencedirect.com/science/article/pii/S0002939418301673>
7. Mohaimin SM, Saha SK, Khan AM, Arif ASM, Kanagasingam Y (2018) Automated method for the detection and segmentation of drusen in colour fundus image for the diagnosis of age-related macular degeneration. *IET Image Process.* 12(6):919–927
8. Rosenfeld PJ, Brown DM, Heier JS, Boyer DS, Kaiser PK, Chung CY, Kim RY (2006) Ranibizumab for neovascular age-related macular degeneration. *New Engl. J. Med.* 355(14):1419–1431
9. Salah MB, Mitiche A, Ayed IB (2011) Multiregion image segmentation by parametric kernel graph cuts. *IEEE Trans. Image Process.* 20(2):545–557
10. Srinivasan PP, Kim LA, Mettu PS, Cousins SW, Comer GM, Izatt JA, Farsiu S (2014) Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* 5(10):3568–3577
11. Sugmk J, Kiattisin S, Leelasantitham A (2014) Automated classification between age-related macular degeneration and diabetic macular edema in OCT image using image segmentation. In: *The 7th 2014 biomedical engineering international conference*, pp 1–4
12. Velez-Montoya R et al (2013) Current knowledge and trends in age-related macular degeneration: today's and future treatments. *Retina* 334:1487–1502