ECRONICON OPEN ACCESS

EC OPHTHALMOLOGY
Research Article

Automated Deep Learning-Based Multi-Class Fluid Segmentation in Swept-Source Optical Coherence Tomography Images

Jonathan D Oakley[1]*, Simrat K Sodhi[2], Daniel B Russakoff[1] and Netan Choudhry[3,4,5]

[1]*Voxeleron LLC, San Francisco, CA, USA*

[2]*University of Cambridge, Cambridge, UK*

[3]*Vitreous Retina Macula Specialists of Toronto, Etobicoke, ON, Canada*

[4]*Department of Ophthalmology and Visual Sciences, University of Toronto, Toronto, ON, Canada*

[5]*Cleveland Clinic Canada, Toronto, ON, Canada*

***Corresponding Author:** Jonathan D Oakley, Voxeleron LLC, San Francisco, CA, USA.

Received: July 04, 2021; **Published:** July 31, 2021

Abstract

Background: To evaluate the performance of a deep learning-based, automated, multi-class, macular fluid segmentation algorithm relative to expert annotations in a heterogeneous population of confirmed wet age-related macular degeneration (wAMD) subjects.

Methods: Twenty-two swept-source optical coherence tomography (SS-OCT) volumes of the macula, comprising 5,632 images from 20 subjects with wAMD, were manually annotated by two expert graders. Results were compared using cross-validation (CV) to automated segmentations using a deep learning-based algorithm encoding spatial information about retinal tissue as an additional input to the network. The algorithm delineates fluid regions in the OCT data, differentiating between intra- and sub-retinal fluid (IRF, SRF), as well as fluid resulting from in serous pigment epithelial detachments (PED). Standard metrics for fluid detection and quantification were used to evaluate performance.

Results: The per slice receiver operating characteristic (ROC) area under the curves (AUCs) for each fluid type were 0.900, 0.945 and 0.939 for IRF, SRF and PED, respectively (95% confidence intervals: 0.886 - 0.912, 0.939 - 0.950 and 0.928 - 0.948). Per volume results were 0.944 and 0.876 for IRF and PED (95% confidence intervals: 0.714 - 1.000 and 0.574 - 1.000); SRF was present in all cases. The correlations (R^2) of fluid volume between expert graders and the algorithm were 0.992 for IRF, 0.986 for SRF and 0.820 for PED.

Conclusion: Automated, deep learning-based segmentation can accurately quantify different macular fluid types in SS-OCT data and in strong agreement with expert graders.

Keywords: *Age-Related Macular Degeneration; Optical Coherence Tomography; Computer Vision; Deep Learning*

Abbreviations

AMD: Age-Related Macular Degeneration; AUC: Area Under the Curve; CCE: Categorical Cross Entropy; CNN: Convolutional Neural Network; CV: Cross-Validation; DME: Diabetic Macular Edema; DSC: Dice Similarity Coefficient; FN: False Negatives; FP: False Positives; ILR: Initial Learning Rate; IRF: Intra-Retinal Fluid; OCT: Optical Coherence Tomography; PED: Pigment Epithelial Detachment; ROC: Receiver

Operating Characteristic; RPE: Retinal Pigment Epithelium layer; RVO: Retinal Vein Occlusion; SRF: Sub-Retinal Fluid; SD-OCT: Spectral Domain OCT; SS-OCT: Swept-Source OCT; TN: True Negatives; TP: True Positives; VA: Visual Acuity; wAMD: wet-AMD

Introduction

Age-related macular degeneration (AMD) is a leading cause of blindness for people over the age of 50, a demographic that is an increasing percentage of the world's population. Declining fertility rates and increasing longevity are shifting the age distribution in industrialized countries towards older age groups [1]. Approximately 170 million individuals are affected with AMD globally and prevalence of AMD is expected to increase to 288 million by the year 2040 [2-5]. The prevalence of this and other ocular diseases places a tremendous burden on the healthcare system, and there is a critical need to streamline diagnosis and improve management of these patients.

Wet age-related macular degeneration (wAMD), the advanced, exudative form of AMD, can be managed using therapeutic agents, but relies on the assessment of fluid amount and type. Optical coherence tomography (OCT) has revolutionized the management of AMD and has been used in investigative trials to associate structure to function in the form of visual acuity (VA) [6]. This ability to measure thickness as a correlate of fluid amount, provides a quantitative value with which to gauge treatment response. There have, however, been discrepancies between this value and reported functional outcomes in the literature [7]. In the absence of automation, identifying the type and location of fluid is a qualitative clinical approach to determining disease activity and then individualizing treatment [8]. Manually assessing the OCT data in this subjective way relies on the clinician/grader's skill and is too time consuming to be feasible in a clinical setting.

The field of computer vision offers a more efficient approach to analysis tasks using automated segmentation algorithms to assess biomarkers such as retinal thickness. These types of software packages are bundled with the OCT devices and remain instrument-specific. While they are able to accurately segment in normal patients, they tend to perform poorly in cases of pathology, particularly so in advanced disease states, where the need for this evaluation is paramount [9,10]. Furthermore, retinal thickness measurements are only a surrogate measure of fluid volume, and explicit, automated fluid detection, quantification and classification are not presently in the armamentarium of analysis tools available clinically.

Deep learning has dramatically changed computer vision making complex tasks, such as automated fluid segmentation, feasible. One of the most powerful aspects of deep learning is that the framework automatically learns both the classifier as well as the salient features that feed into it. Prior to this, feature extraction typically required meticulous manual design and implementation. We apply this approach in this study using convolutional neural networks (CNNs). CNNs are capable of building rich, layered (deep) representations of the data which are then classified through additional layers of representation and learned associations. Such significant technical improvements have advanced semantic segmentation to levels matching expert human graders, taking seconds to perform tasks that would otherwise take hours. This functionality is of significant interest, particularly in ophthalmology where powerful imaging devices are relatively inexpensive, the device software limited, and image data is key to clinical management. However, ahead of clinical deployment these algorithms require extensive evaluation supportive of final regulatory approval.

In the following we report on the performance of macular fluid segmentation in a heterogeneous, wAMD population. Previously reported studies include Schlgel., *et al*. who report on fluid segmentation in AMD, diabetic macular edema (DME) and retinal vein occlusion (RVO) cases, using two different OCT devices [11]. Mehta., *et al*. similarly report on an approach for intra-retinal fluid (IRF) segmentation that showed an existing deep learning model's improvement relative to manual graders after transfer learning [12]. More recently, and underlining the interest of this particular application, an on-line challenge was organized stemming from a medical imaging conference that assimilated and compared a number of different deep learning approaches for segmenting three retinal fluid types in OCT data; intra- and sub-retinal fluid (SRF) and pigment epithelial detachment (PED) [13]. This task maps nicely to the work we present in the following and is returned to in the discussion section of this paper.

Methods

This study was Institutional Review Board of Advarra approved and followed the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants and was in accordance with current ICH/GCP guidelines, section 4.8.10.

The data was collected using the Topcon Triton SS-OCT device (Tokyo, Japan), using the macular protocol, the scan was centered on the fovea and covered a lateral area of $7mm^2$, with a depth of 2.58mm. This study involved 23 subject eyes from 21 patients. One OCT volume was removed due to excessive motion and poor image quality. The resulting 22 volumes were acquired from 20 patients where each volume comprised 256 B-scans, with each B-scan comprising 512 A-scans each of 992 pixels. Two patients had both eyes included in the study. The key inclusion criteria were: (1) patients had to be ≥ 25 years old; (2) each subject eye diagnosed with wAMD and (3) wAMD treatment naïve.

Manual segmentations

Two expert graders, SS and JO separately annotated every fourth B-scan using custom image labeling software, but at different starting B-scans so as not to segment the same B-scan. Combined, this created volumes with every 2^{nd} B-scan manually labelled. Adopting a similar bootstrapping approach as input to generating additional labeling data from De Fauw., *et al*. the intermediate B-scans were segmented using the same semantic segmentation algorithm reported on for the final result, where the training data was every 2^{nd} B-scan from the same volume (Figure 1) [14]. The final correction of the automated labeling was done manually by both readers, ultimately ensuring that all 22 volumes (5,632 B-scans) used in the study were fully labeled and could be used as ground truth segmentations. Consensus segmentation results were arrived at in final consultation and adjudication with a senior retina specialist (NC).

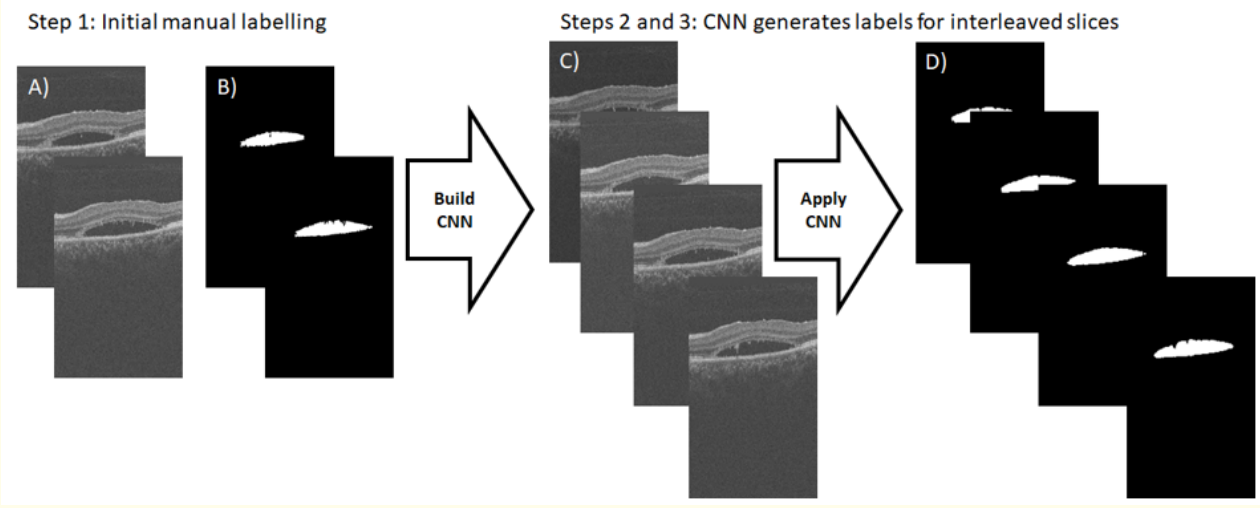

Figure 1: *Manual labeling of the image data. A bootstrapping approach to generation of labeled data uses a CNN to automatically generate labeled images that were then subsequently verified manually. For manually labeled images in the training set (Step 1), a CNN is capable of accurately labeling the neighboring images when presented in the test set (Steps 2 and 3). Example input images (A) are used with associated manual labels (B) to create an initial CNN. Unlabeled images (C) from neighboring B-scans within the same volume are segmented by the CNN. The output labels (D) are then manually verified such that they may themselves be used as ground truth.*

Pre-processing

Each input volume's retinal layers were automatically segmented by Orion™ version 3.0 (Voxeleron, San Francisco, USA), research software for OCT analysis. This segmentation information was then encoded into a separate channel that was entered into the supervised learning process. Existing literature has used, separately, layer information as features in support of the classification of fluid pockets [15], leveraging the fact that retinal fluid exhibits layer dependent properties and thus is used by the model to help determine fluid type; IRF typically being higher up in the volume, SRF close to and above the retinal pigment epithelium layer (RPE), PEDs below the RPE, but not below the retina itself. In figure 2, the images on the left represent two example B-scans within the volume whose segmentations are encoded as image data shown colored on the right. Lastly, each B-scan and the associated label images were sub-sampled to 256-by-256 pixels in size.

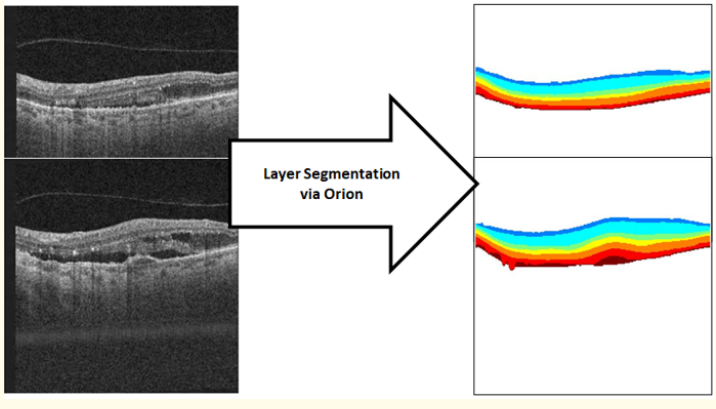

Figure 2: *Manual labeling of the image data. A bootstrapping approach to generation of labeled data uses a CNN to automatically generate labeled images that were then subsequently verified manually. For manually labeled images in the training set (Step 1), a CNN is capable of accurately labeling the neighboring images when presented in the test set (Steps 2 and 3). Example input images (A) are used with associated manual labels (B) to create an initial CNN. Unlabeled images (C) from neighboring B-scans within the same volume are segmented by the CNN. The output labels (D) are then manually verified such that they may themselves be used as ground truth.*

Deep learning

Given 22 volumes, the study comprised a total of 5,632 B-scans. For the task of supervised learning, a cross validation (CV) approach was adopted for training and testing, where data was very strictly stratified by subject.

This approach splits the data such that any image data from a given subject is not used in training the model and testing the result: i.e. of a subject's 256 or 512 B-scans, all or none are in one set or another.

Training the model involves a further split of the data to create a validation set that tells us when the learning processed has converged. In support of more generalizable solutions, we apply the same subject-based stratification of the training/validation split.

Given the limited number of patients, we used 10-fold CV, so on each fold a model was built for all but two or three subject eyes. This meant that, for two folds, the test set comprised solely from the patients who had both eyes included in the study. During the learning process, the model was assessed based on a validation set that randomly chose 5 subjects from the training data (1,280 B-scans). Therefore, on average, training data comprised data from 15 subjects and validation data came from the remaining 5 subjects. This strict stratifica-

tion to subject eye means that, while the learning uses only training set, we stop this process only when the validation error is clearly increasing. This is a standard technique used in supervised learning to prevent over-fitting.

The training data was augmented using random scaling, rotations and shifting such that, for every input B-scan, six additional B-scans were generated. We did not augment the validation data.

Implementation

The deep learning architecture used for first the training and then the testing was that of the U-Net, which is a version of the auto-encoder that uses skip connections to better maintain detail across different scales [16,17]. The implementation uses Tensorflow 2.0 (Google, Mountain View, USA) and training and testing were performed on a Dell Precision 7920 workstation, running NVIDIA Quadro GV100 GPUs.

The network takes as input the OCT image, their encoded segmentations and associated labels (See figure 3). It uses three encoding/decoding levels, as shown in figure 4. Learning is based on minimizing the model's loss, where the loss function uses categorical cross entropy (CCE) and a weighted dice similarity coefficient (DSC). The DSC, for two sets is twice the area of their overlap divided by the total number of elements in both sets. CCE is used for multi-class classification where the labels use one-hot encoding and defines a measure of dissimilarity between the true value and that which is predicted. Due to the one-hot encoding, we have four classes of interest: background, IRF, SRF and sub-RPE. In our application, non-fluid, or background regions will dominate this cost function that is to be minimized, requiring appropriate weighting of the fluid labels. The weighting of the DSC can then be set to diminish influence of the background pixels on the overall dice score. Empirically we look for loss values to decrease with increasing epochs in a fairly monotonic way: i.e. to ensure that, on the unseen validation data, the network continued learning without over-fitting. This combination of CCE and weighted dice loss works well and enables a fairly continuous reduction in validation loss with epochs. There are, however, other hyper-parameters are at play: drop-out was set to 0.25, batch normalization was applied, and the batch size was 8. The optimizer used was Adadelta [18].

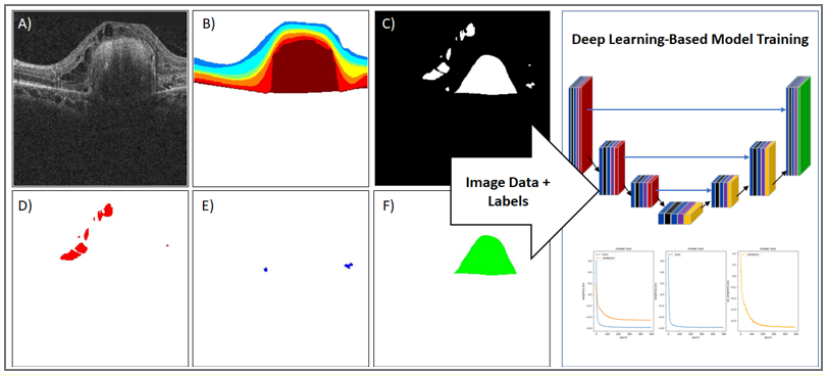

Figure 3: *Deep learning-based supervised learning. The supervised learning approach using a deep convolutional neural network architecture. The input image data are the B-scans (A) and their associated encoded segmentation results (B). The input labels are one-hot encoded versions of each fluid type. This example has one of each: background (C - black), IRF (D - red), SRF (E - blue) and PED (F - green). Typically, one or more of the fluid label images will be entirely blank, meaning that fluid type was not present in that B-scan. Portrayed on the right is the deep learning architecture that takes in the data and labels and minimizes the loss between the two for the training data. An example plot of loss value decreasing over epochs is shown, where the training loss is shown in blue and the validation loss in orange.*

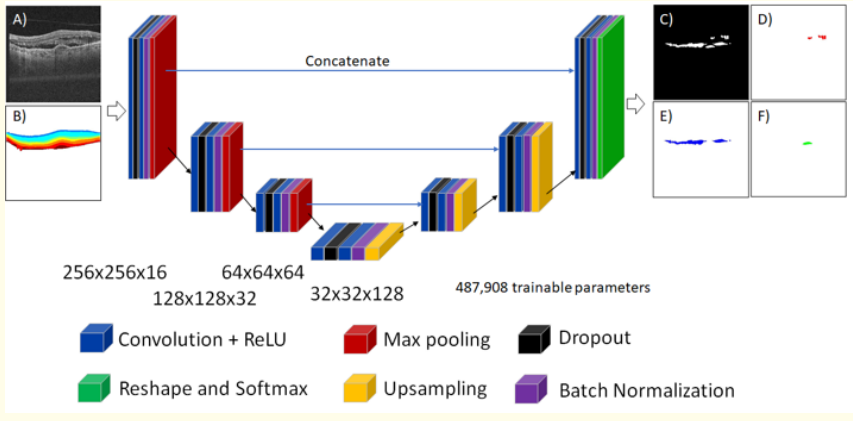

Figure 4: *The deep learning CNN architecture. The CNN architecture used in these analyses. For the two-channel input image (A, B), it outputs four images giving the probability score, at each pixel, of belonging to one of the four classes: background (C - black), IRF (D - red), SRF (E - blue) and PED (F - green). Dropout was set to 0.25 and the final softmax layer ensures these are normalized and can thus be interpreted as probabilities.*

The DSC for a single label is defined as:

$$DSC = \frac{2|y_{true} \cap y_{pred}|}{area(y_{true}) + area(y_{pred})}$$

Where y_{true} and y_{pred} are the labels and the current prediction of the labels, respectively. The weighted dice term, DW, is defined as:

$$DW = -\sum_{l=1}^{4} w_l DSC_l$$

Where l denotes the labels - background, PED, SRF and IRF - and w_l denotes the weights applied to each of these. The final loss function then becomes:

$$Loss = CCE + (1-\alpha)DW$$

Where α controls the influence of the weighted dice term and is set to 0.2. The weights used are w = [0.05, 0.5, 0.225, 0.225], which was set *a priori* based on an estimate of relative proportions in wAMD and the four weights apply to background, IRF, SRF and PED, respectively. In doing this, we are capturing, for example, an estimate that the occurrence of background pixels is ten times that of IRF by setting the weight of IRF's (0.5) contribution to the overall term, DW, as ten times that of the background weight (0.05). It should therefore be noted that if the first weight term is too high relative to the others, background will tend to dominate the segmentation. The exact distribution of the weights is not critical to the method, but the background's influence on the final dice score must be suppressed.

The optimizer's initial learning rate (ILR) was set to 0.5, which changed every n^{th} iteration such that, at epoch i the learning rate used is given as:

$$LR_i = ILR * d^{floor(i+1)/n)}$$

With d set to 0.95, the drop fraction, this resulted in an exponential decay of the initial learning rate that stabilized to a minimum near 600 epochs. Training stops when the loss score applied to the validation data set stops diminishing for 50 epochs.

The training data comprised 15 subject eyes containing 3,840 B-scans, resulting in 26,880 B-scans after augmentation. With a batch size of 8, this means a single epoch takes 3,360 iterations to work through the entire data set.

Results

Fluid detection

Fluid detection is reported on a per B-scan and per volume basis using receiver operating characteristic (ROC) curves, with the summary statistic being the area under the curve (AUC). For a given B-scan, the network outputs a probability score for each class, facilitating the plot of ROCs and the measured performance. In evaluating on a per volume basis, we were only able to report IRF and PED performance as all 22 wAMD subject eyes had at least some SRF as reported by the manual labelling, meaning no negative examples were available. These are given in figure 5, with precision and recall values in table 1.

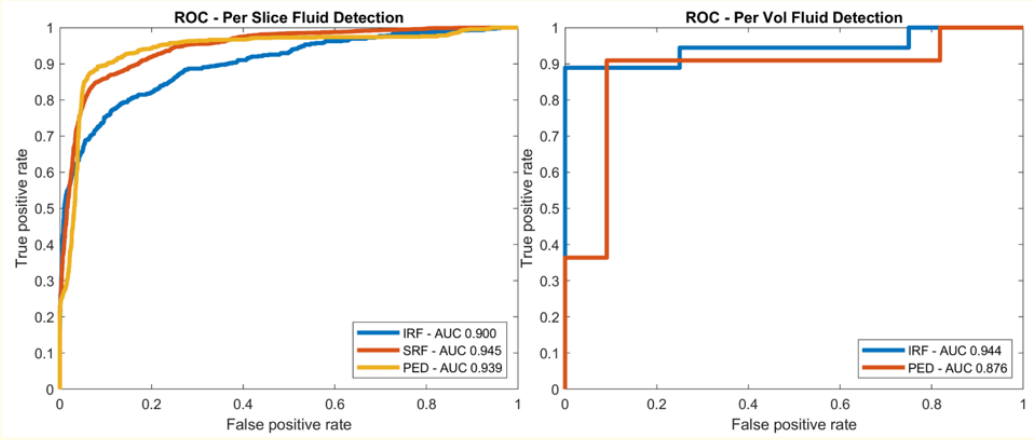

Figure 5: *Detection performance of each fluid type. Per B-scan AUCs were 0.900, 0.945 and 0.939 for IRF, SRF and PED, respectively (95% confidence intervals: 0.886 - 0.912, 0.939 - 0.950, and 0.928 - 0.948). Per volume results were 0.944 and 0.876 for IRF and PED (95% confidence intervals: 0.714 - 1.000 and 0.574 - 1.000).*

	Per B-scan			Per volume		
	IRF	**SRF**	**PED**	**IRF**	**SRF**	**PED**
Precision	0.503	0.749	0.729	1	1	0.978
Recall	0.769	0.827	0.842	0.823	0.921	0.898

Table 1: *Precision and recall proportions for each fluid group by B-scan and volume. Precision measures the proportion of positively labeled cases that were correct, and recall measures the proportion of positive cases that were identified correctly. Precision example: when the algorithm says a volume contains PED, it is correct 97.8% of the time. Recall example: our algorithm correctly identifies 92.1% of volumes with SRF.*

Precision is defined as the number of true positives (TP) divided by the total number of reported positive cases, true and false (FP): TP/(FP+TP). Recall is TP divided by the total number of actual positive cases: TP/(TP+FN). The values are calculated at the Youden index of the ROC curve.

Fluid quantification

Given manual segmentations of all B-scans in all volumes, we report the ability to quantify fluid by directly correlating the total volume of fluid for each of the 22 subject eyes that was manually delineated with that automatically determined. The coefficient of determination (R^2) is used as a summary statistic for correlation, and both the correlation and Bland-Altman plots are shown in figure 6. Furthermore, the area of overlap, as quantified by the DSC, between graders and the algorithm was examined (Figure 7). Finally, example result images are given in figure 8 and 9.

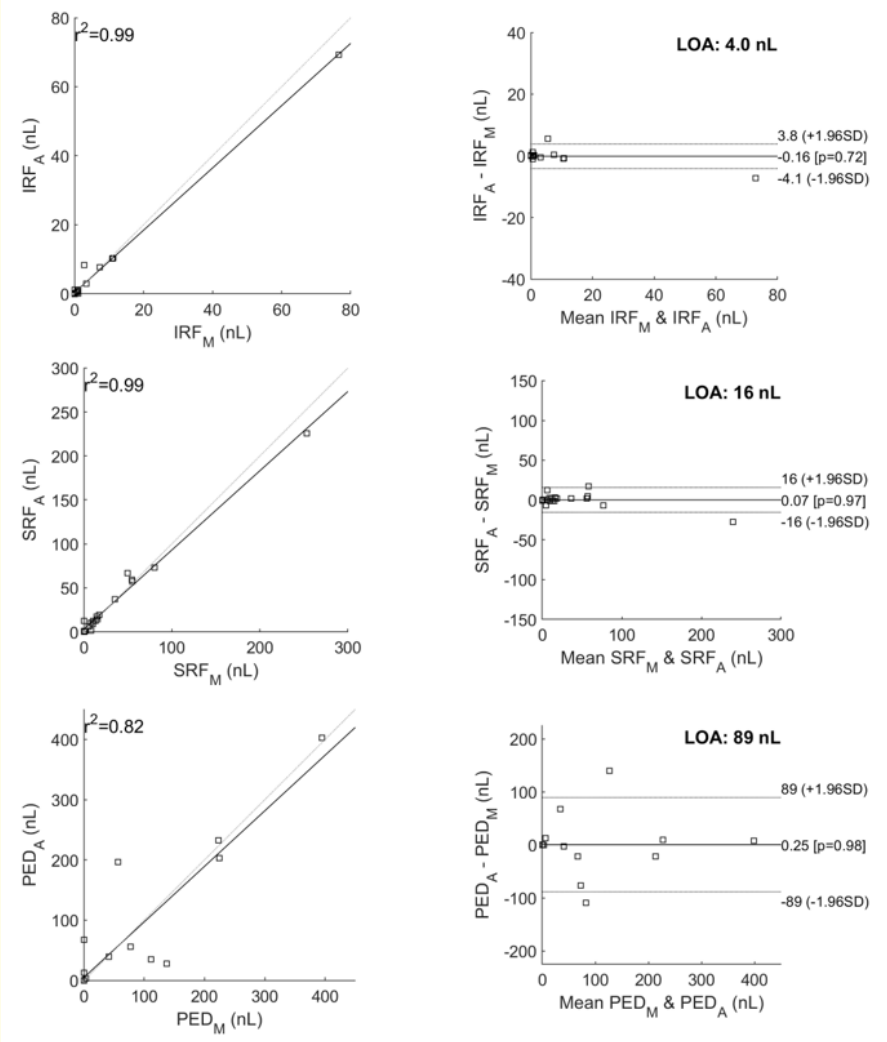

Figure 6: *Manual versus automated fluid quantification. The above shows the manual versus automated reported total areas for each fluid type across all 22 volumes (left) and the Bland-Altman plots on the right. The correlation scores are 0.992, 0.986 and 0.820 for IRF, SRF and PED, respectively (top to bottom). For the Bland-Altman plots, the manual values are denoted with '_M' and the automate values with '_A'. Narrow limits of agreement are shown for IRF and SRF, but are larger for PED, as we address in the discussion.*

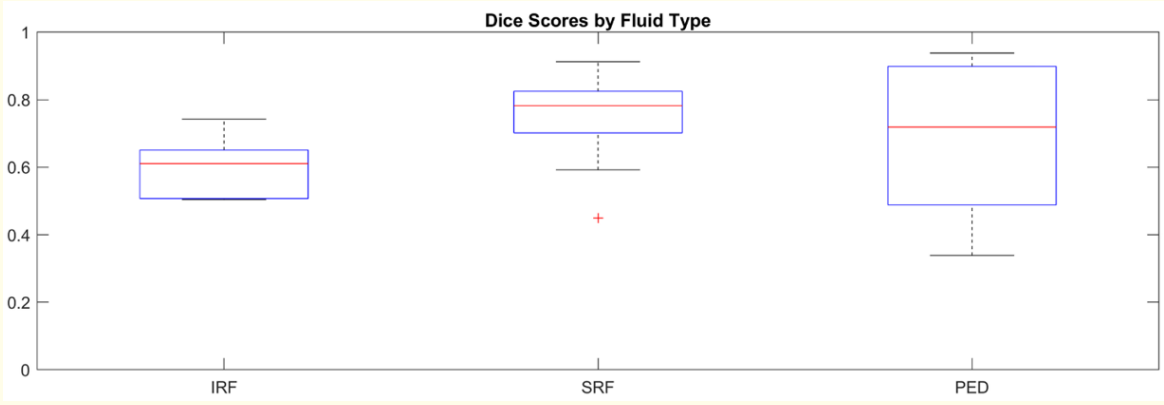

Figure 7: *Distribution of dice scores for each fluid type. Dice coefficients for all cases where a given fluid type was manually labeled as being above 1nL.*

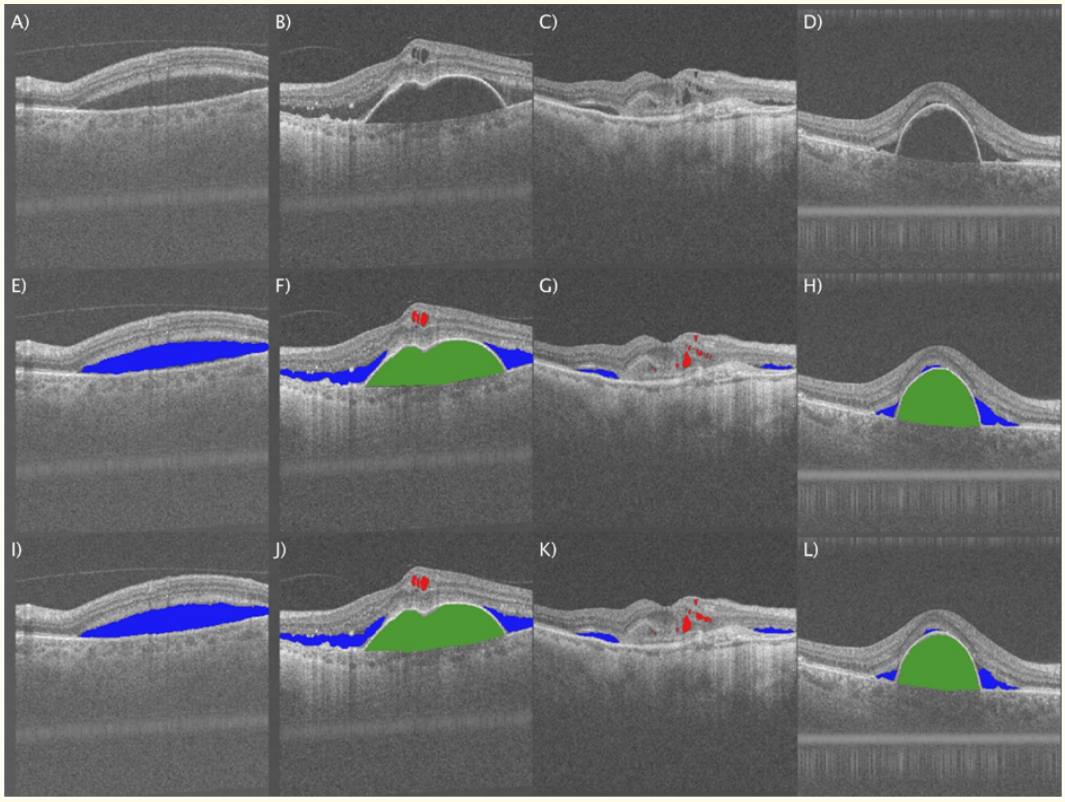

Figure 8: *Example correct segmentations. Input B-scans (A, B, C, D), manual segmentations (E, F, G, H), and the automated segmentations (I, J, K, L). IRF is labeled red, SRF in blue and green constitutes PED.*

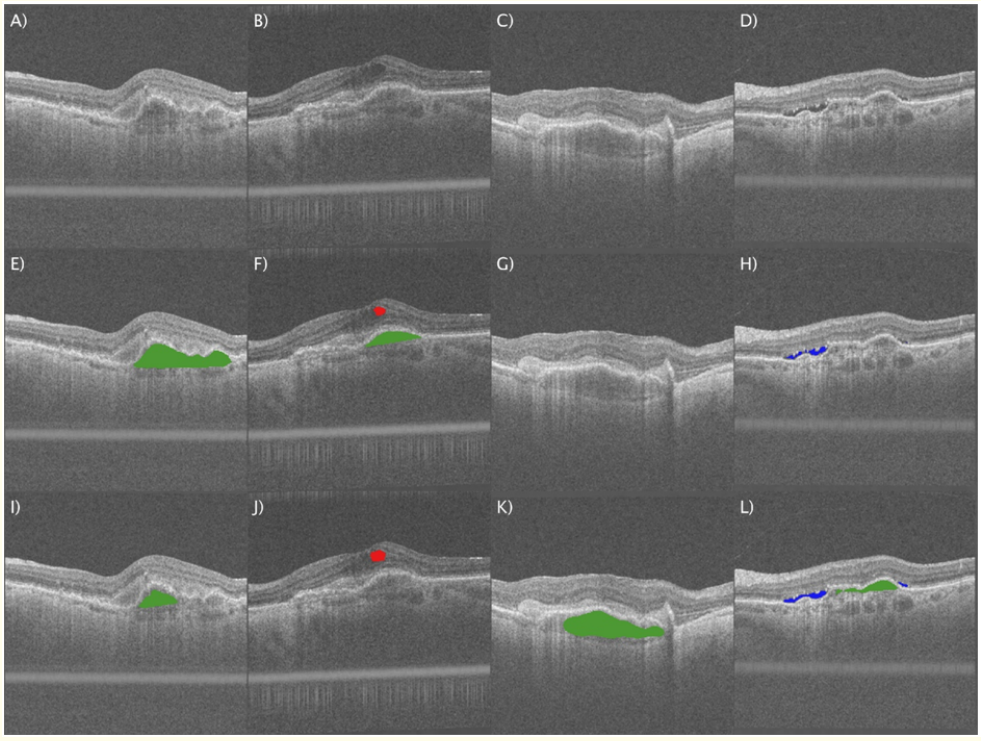

Figure 9: *Example incorrect segmentations. The above shows example input B-scans exhibiting fibrovascular PEDs (A, B, C, D). The manual labelings below (E, F, G, H) decided the only (E) and (F) contained fluid and (G) and (H) were mostly fibrous tissue. The automated segmentations show that cases (I), (K) and (L) labeled as PEDs, but not (F). IRF is labeled red, SRF in blue and green constitutes PED.*

Discussion

We have shown that the performance of automated segmentation of three fluid types in SS-OCT is in very strong agreement with that done manually by expert graders. To the best of our knowledge, this is the first study to report on such a technique using SS-OCT. Previous work using deep learning has, for example, reported on the segmentation of the retina and choroid and also disease detection in SS-OCT, but not retinal fluid segmentation [19,20]. As an imaging modality, SS-OCT is becoming clinically more relevant than the more conventional and established spectral domain OCT (SD-OCT) due to a number of distinct advantages. A first is the speed of acquisition that is typically four times as fast as SD-OCT devices - the Triton device used here acquires 100,000 A-scans/sec - which allows for more data to be captured over wider fields of view in a shorter amount of time. Prototype devices now operate at up to 400,000 A-scans/sec [21]. Speed of acquisition minimizes patient chin rest time and reduces the amount of motion in the scan. A second advantage is the longer wavelength that allows light to penetrate deeper into the retina, supportive of better visualization of the choroid and structure below the RPE. And while SS-OCT has slightly reduced axial resolution relative to similar SD-OCT devices, the overall signal to noise ratio tends to be higher.

Deep learning, however, is the enabling technology of this and other similar studies. In comparison, previous solutions are quite varied. One such example is from Bogunović., *et al*. who assimilate a number of computer vision-based techniques to achieve fluid segmentation

[15]. Their approach begins with layer segmentation to constrain the solution and to add positional information that, along with textural features, is used to assign probabilities of voxels, as "fluid" or "no fluid". This assignment uses an artificial neural network (ANN) where supervised learning is based on a leave-one out (LOO) CV approach. This soft segmentation is hardened in a final step that uses fully 3D geodesic graph cuts to isolate the fluid pockets producing binary labels.

Similar methods exist where layer segmentation is initially used to constrain the problem and then a number of techniques such as region growing, thresholding, boundary tracing or clustering are used to arrive at the final result [22-25].

An important early deep learning reference is the work of Roy., *et al*. who compare a variety of U-Net implementations on 110 labeled images from 10 subjects [26]. Other previously reported studies include Schlgel., *et al*. who report on fluid segmentation in AMD, diabetic macular edema (DME) and retinal vein occlusion (RVO) cases [11]. This was performed using two different OCT devices and has subsequently become an important reference. For comparison, and considering the AMD cases only, two fluid types (IRF and SRF) were segmented using a deep CNN and results reported based on both 4- and 10-fold CV. The number of volumes were 70 and 257, respectively for the two different instruments. Performance is impressive, with volume-based detection of fluid having AUCs of 0.95 (IRF) and 0.98 (SRF) on the Cirrus device (Carl Zeiss Meditec, Dublin, USA); and 0.93 (IRF) and 0.90 (SRF) with the Spectralis data (Heidelberg Engineering, Heidelberg, Germany). Mehta., *et al*. similarly report on a deep learning approach for IRF segmentation [12]. Their reporting mechanism was to show that the DSC between the automated algorithm and the manual graders improved after transfer learning. The model used, from Lee at al. was created based on Spectralis data and then further trained for Cirrus data using 400 OCT B-scans (2D images) from 58 patients [27]. It was deployed on a test set comprising of 70 scans from 70 different patients. The architecture was necessarily that developed by the original authors as transfer learning was applied.

The work of Schlgel is of particular importance given its recent deployment in a retrospective clinical analysis that was able to correlate fluid amounts to clinical outcomes [28]. A key difference in our study, however, is the additional segmentation of PED fluid within the same network. This becomes feasible with this data set, because of the method of spatial injection that uses the layer segmentation software (Orion™). This we see is also the case in the analysis of three fluid labels in the RETOUCH challenge where, in the best performing instances, the approach is to first segment the data, second train a CNN with image data and segmentation as input, and third refine the final result using an additional classifier [13,29]. We see a benefit of having the segmentation element using non-trained data and all of the fluid segmentation within a single CNN, as is reported here. But consensus exists that encoding spatial information about the retinal tissue as an additional channel to the input image data is shown to improve the overall performance of the segmentation algorithm. While fluid segmentation alone can be accomplished using just the input image data, the additional classification of each fluid type benefits significantly from this technique.

This pilot study is not without its limitations. Firstly, we have a restricted number of subjects, meaning that, in CV analysis, there are only limited examples of a given fluid type in the training data when testing against a given subject. This is further confounded by our approach of splitting subject IDs across training, validating, and testing sets to ensure the network is learning but in a way that will generalize to unseen cases by not overfitting the training data. The populations are similar, however, to Bogunović., *et al*. but as they have three volumes per subject, their LOO CV analysis might be unduly influenced by having subject data in both the train and test sets [15]. More recent work has involved extremely large data sets - 3000 subject eyes - to report on anatomical features that include IRF, SRF, PEDs and also hyperreflective foci [30]. While impressive, it is important to note that the algorithm used has not to date been validated for this task, only as a pre-processing step to an inference engine for triaging patients based on severity of disease [14].

While the lack of data may be a limitation of the analysis, it is not necessarily a limitation of the method. The deep learning paradigm thrives on data; the more data we have, the larger the learning space covered, meaning it is more likely we have relevant examples to unseen data and the more natural the regularization of the resulting network. In this respect, we again look to the work of Schlegl., *et al*.

where the subject data is far larger [11]. In considering just their population of wet-AMD patients, however, we see similar ratios of train/test splits, but they do go on to apply this at scale. We consider this reference as a current standard, and as such have looked to contrast our results directly.

Whenever supervised learning is applied, one must ensure that the ground-truth is correct. This can be challenging as the manual quantification of fluid presence, amount and type is difficult, extremely time consuming and highly variable across graders. As illustrated in figure 9, we found a consistent definition of sub-retinal pigment epithelial (RPE) fluid in serous PED difficult, and this is a limitation as this can additionally lead to questioning the ground truth data used in 1) the training of the models using supervised learning and 2) the validation based again on manual labeling. Repeatability between manual graders has been previously reported as low, and in the context of segmentation tasks has been highlighted by Bogunović., *et al*. where their inter-observer analysis showed poor overlap between graders [15,31]. We have attempted to minimize the variability across graders by not reading blinded to each other and instead labeling intermittent slices of the same volume. This allowed each grader to read the images in the context of the other's labelling. PED cases, however, can manifest as either serous, fibrovascular or drusenoid; or indeed a mix of these. Our target labeling was serous, which is sharply demarcated given their large, dark (hyporeflective), fluid regions and dome-like shape. Fibrovascular PEDs on the other hand contain tissue, which refracts the OCT signal and shows, therefore, as lighter than fluid. Drusenoid PED are smaller, in general, creating an undulating RPE surface, and contain hyaline protein deposits, denser tissue that results again in a brighter appearance. Consistent labeling of our target detachment type, serous, was important, but it is still challenging to find a consistent manual "threshold" when some fluid only cases show clear signs of fibrovascular tissue; or vice versa. This is particularly evident in figure 9 where fluid has been manually labeled in (E) and (F), but not in (G) and (H). Consequently, the algorithm too has difficulty in finding this difference. Ideally, one might iterate to find a consensus labeling in conjunction with the algorithm results, but that is also time consuming and while such a consensus would ultimately improve the results it could also inappropriately bias them. With enough data, however, the differentiation of PED subtypes is undoubtedly possible as is indicated in previous work [14].

Conclusion

In conclusion, our study demonstrates that automated fluid detection and segmentation correlates highly with manual grading performed by experts. Combined segmentations of IRF and SRF with sub-RPE fluid in serous PEDs, offer a near complete automated solution for retinal fluid in wAMD. An automated approach to multi-layer fluid segmentation offers an advantage to physicians managing patient care amidst a busy clinical setting, particularly with the growing disease burden of wAMD. Furthermore, applying this deep-learning approach in clinical trials for therapeutic drug development could accelerate trial end points and the achievement of novel therapeutics within commercial timelines.

Conflicts of Interest

Authors JDO and DBR are salaried employees of Voxeleron LLC, San Francisco.

NC consultant for Topcon, Optos, Bayer, Allergan, Novartis, Carl Zeiss Meditec, research equipment from Topcon, Optos and Carl Zeiss Meditec.

Bibliography

1. Anderson GF and Hussey PS. "Population Aging: A Comparison Among Industrialized Countries: Populations around the world are growing older, but the trends are not cause for despair". *Health Affairs* 19.3 (2000): 191-203.

2. Mitchell P., *et al*. "Age-related macular degeneration". *The Lancet* 392.10153 (2018): 1147-1159.

3. Wong WL., *et al*. "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis". *The Lancet Global Health* 2.2 (2014): e106-116.

4. Zhao S., *et al*. "Protocol of global incidence and progression of age-related macular degeneration: a systematic review". *Medicine* 98.10 (2019).

5. Pennington KL and DeAngelis MM. "Epidemiology of age-related macular degeneration (AMD): associations with cardiovascular disease phenotypes and lipid factors". *Eye and Vision* 3.1 (2016): 34.

6. Lalwani GA., *et al*. "A variable-dosing regimen with intravitreal ranibizumab for neovascular age-related macular degeneration: year 2 of the PrONTO Study". *American Journal of Ophthalmology* 148.1 (2009): 43-58.

7. El-Emam S., *et al*. "Correlation of spectral domain optical coherence tomography characteristics with visual acuity in eyes with subfoveal scarring after treatment for wet age-related macular degeneration". *Retina* 33.6 (2013): 1249-1257.

8. Guymer R and Wu Z. "Age-related macular degeneration (AMD): More than meets the eye. The role of multimodal imaging in today's management of AMD". *Clinical and Experimental Ophthalmology* (2020).

9. Asrani S., *et al*. "Artifacts in spectral-domain optical coherence tomography measurements in glaucoma". *JAMA Ophthalmology* 132.4 (2014): 396-402.

10. Chen JJ and Kardon RH. "Avoiding clinical misinterpretation and artifacts of optical coherence tomography analysis of the optic nerve, retinal nerve fiber layer, and ganglion cell layer". *Journal of Neuro-Ophthalmology* 36.4 (2016): 417.

11. Schlegl T., *et al*. "Fully automated detection and quantification of macular fluid in OCT using deep learning". *Ophthalmology* 125.4 (2018): 549-558.

12. Mehta N., *et al*. "Model-to-Data Approach for Deep Learning in Optical Coherence Tomography Intraretinal Fluid Segmentation". *JAMA Ophthalmology* (2020).

13. Bogunović H., *et al*. "RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge". *IEEE Transactions on Medical Imaging* 38.8 (2019): 1858-1874.

14. De Fauw J., *et al*. "Clinically applicable deep learning for diagnosis and referral in retinal disease". *Nature Medicine* 24.9 (2018): 1342-1350.

15. Bogunović H., *et al*. "Geodesic graph cut based retinal fluid segmentation in optical coherence tomography". *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA 2015, Held in Conjunction with MICCAI 2015, Munich, Germany* 9 (2015): 49-56.

16. Ronneberger O., *et al*. "U-net: Convolutional networks for biomedical image segmentation". *In International Conference on Medical Image Computing and Computer-Assisted Intervention* 5 (2015): 234-241.

17. Badrinarayanan V., *et al*. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (2017): 2481-2495.

18. Zeiler MD. "Adadelta: an adaptive learning rate method". *arXiv preprint arXiv* (2012).

19. Maloca PM., *et al*. "Validation of automated artificial intelligence segmentation of optical coherence tomography images". *PloS one* 14.8 (2019): e0220063.

20. Sogawa T., *et al*. "Accuracy of a deep convolutional neural network in the detection of myopic macular diseases using swept-source optical coherence tomography". *Plos one* 15.4 (2020): e0227240.

21. Potsaid B., *et al*. "Ultrahigh Speed and Multiscale Volumetric 1050nm Ophthalmic OCT Imaging at 100,000-400,000 Axial Scans per Second". *Investigative Ophthalmology and Visual Science* 52.14 (2011): 1319-1319.

22. Rashno A., *et al*. "Fully-automated segmentation of fluid regions in exudative age-related macular degeneration subjects: Kernel graph cut in neutrosophic domain". *PloS one* 12.10 (2017): e0186949.

23. González A., *et al*. "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* 20 (2013): 397-400.

24. Wilkins GR., *et al*. "Automated segmentation of intraretinal cystoid fluid in optical coherence tomography". *IEEE Transactions on Biomedical Engineering* 59.4 (2012): 1109-1114.

25. Wang J., *et al*. "Automated volumetric segmentation of retinal fluid on optical coherence tomography". *Biomedical Optics Express* 7.4 (2016): 1577-1589.

26. Roy AG., *et al*. "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks". *Biomedical Optics Express* 8.8 (2017): 3627-3642.

27. Lee CS., *et al*. "Deep-learning based, automated segmentation of macular edema in optical coherence tomography". *Biomedical Optics Express* 8.7 (2017): 3440-3448.

28. Roberts PK., *et al*. "Quantification of Fluid Resolution and Visual Acuity Gain in Patients With Diabetic Macular Edema Using Deep Learning: A Post Hoc Analysis of a Randomized Clinical Trial". *JAMA Ophthalmology* (2020).

29. Lu D., *et al*. "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network". *Medical Image Analysis* 54 (2019): 100-110.

30. Moraes G., *et al*. "Quantitative Analysis of OCT for Neovascular Age-Related Macular Degeneration Using Deep Learning". *Ophthalmology* (2020).

31. Toth CA., *et al*. "Identification of fluid on optical coherence tomography by treating ophthalmologists versus a reading center in the comparison of age-related macular degeneration treatments trials (CATT)". *Retina* 35.7 (2015): 1303.

Volume 12 Issue 8 August 2021