Performance of an Artificial Intelligence Decision Support Tool, the Vienna Fluid Monitor in Neovascular Age-Related Macular Degeneration

Gregor S. Reiter[1]; Christian Unterrainer[2], Adriadne Whitby[2], Bilal Haj Najeeb[1,2], Thomas Grundnig[2], Amir Sadeghipour[2], Ursula Schmidt-Erfurth[1]

1. Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology and Optometry, Medical University of Vienna, Vienna, Austria
2. RetInSight, RetInSight GmbH, Vienna, Austria

Corresponding Author (also address for reprints):
Prof. Ursula Schmidt-Erfurth, MD
Department of Ophthalmology and Optometry
Medical University of Vienna
Währinger Gürtel 18-20
1090 Vienna, Austria
ursula.schmidt-erfurth@meduniwien.ac.at
Tel: +43 01 40400 79310
Fax: +43 01 40400 79320

Conflict of interest: GSR received research funds from RetInSight. US-E is a scientific consultant for Genentech, Heidelberg Engineering, Kodiak, Novartis, RetInSight and Roche. CU, AW, BHN, TG, AG and AS are employees of RetInSight.

Purpose

Real-life treatment outcomes in neovascular age-related macular degeneration (nAMD) are far from satisfactory compared to randomized clinical trials (RCT). In RCT retinal imaging is often assessed by certified readers in a central reading center. However, real-world interpretations of retinal images are lacking precision and repeatability. This study was conducted to report on the performance of the Vienna Fluid Monitor, a clinical artificial intelligence (AI)-based decision support tool for the quantification of macular fluid compartments (subretinal fluid (SRF), intraretinal fluid (IRF) and pigment epithelium detachment (PED)) in nAMD.

Setting/Venue

Optical coherence tomography (OCT) data from a real-world outpatient clinic in a tertiary referral center (Department of Ophthalmology at the Medical University of Vienna, Austria).

Methods

A multi-class deep learning-based method was developed to identify regions of SRF, IRF and PED in nAMD subjects. In this framework, dilated convolutions are used to detect fluid regions at multiple features and ensembling methods were used to increase confidence in the final pixel label. The dataset was comprised of 219 OCT scans from Heidelberg Spectralis HRA+OCT devices (Heidelberg Engineering, Heidelberg, Germany), with 6210 b-scans from 155 different nAMD patients. The dataset included data from patients with type 1, 2 and 3 nAMD from both treatment-naïve and post-treatment stage. Extensive data augmentation was used to address variable image quality, and several types of OCT specific artifacts found in real world data. These include hyperreflective foci (HRF), large blood vessels which cause an underlying hyporeflective shadow (mimicking IRF), outer nuclear layer (ONL) thickening and posterior vitreous detachment. A combination of custom-built augmentations tools for shadowing, vignetting, noise, blur and affine transformation were utilised for this purpose. The performance of the algorithm was evaluated on a pixel-level basis for IRF, SRF and PED against ground truth of manual reading made by a

reading expert. The dice score is used as a primary metric. Confidence intervals were calculated using Wilson confidence interval.

Results

One hundred and fifty-five patients were split into three sets: 99 for training, 28 to validate model selection and 28 for the test set used to report performance. The training set contained 136 OCT scans with 3620 annotated b-scans, the validation set contained 40 OCT scans with 1314 annotated b-scans, and the test set contained 43 OCT scans with 1276 annotated b-scans. The validation and testing sets were selected randomly and supplemented by data with confounding morphological characteristics. There was no patient overlap between the separate sets. For a random subset (n=13) the test set was fully annotated: The Pearson correlations in the central 1mm of the ETDRS-grid for each quantitative fluid volume showed strong correlations between the algorithm and the expert gradings (IRF: $r=0.9998$, $p<0.001$; SRF: $r=0.9940$, $p<0.001$, and PED: $r=0.9848$, $p<0.001$). Fluid detection was consistently well performed by the algorithm for each compartment on the b-scan level: IRF (dice: 0.9207, sensitivity: 0.8960, precision: 0.9468), SRF: (dice: 0.9660, sensitivity 0.9733, precision 0.9589), and PED (dice: 0.9153, sensitivity 0.9450, precision 0.8873). Sensitivity and precision on a pixel level were high for each fluid type: IRF (sensitivity: 0.73, precision: 0.81), SRF (sensitivity: 0.92, precision: 0.88), and PED (sensitivity: 0.92, precision: 0.85).

Conclusion

Disease activity in nAMD can be measured in a precise and repeatable manner using AI-based decision support tools such as the Vienna Fluid Monitor. Quantifications of macular fluids enable clinicians to base their re-treatment criteria on objective measurements that can be implemented in real-time in a busy outpatient clinic. With the use of accurate AI-based fluid quantification every clinician has the possibility to access reading center-like precision in determination of disease activity and help narrow the gap between real-world and clinical trials. Future investigations need to focus on the implementation of AI-quantifications in prospective trials and real-world clinical settings to optimize the current qualitative

assessment of OCT images. Going beyond the current practice will introduce true precision medicine and personalized treatment regimens into daily routine along with easing the burden on health-care providers, treating clinicians and patients.