

Defining disease endophenotypes in neovascular AMD by unsupervised machine learning of large-scale OCT data

Philipp Seeböck^{1,2}, Sebastian M. Waldstein¹, René Donner², Bianca S. Gerendas¹, Amir Sadeghipour¹, Aaron Osborne³, Georg Langs^{1,2}, Ursula Schmidt-Erfurth¹

¹Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology and Optometry, Medical University of Vienna

²Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna

³Genentech, Inc, South San Francisco, CA, United States

56 - A0044

Financial disclosures:
Waldstein: C (Bayer Healthcare AG; Novartis Pharma AG);
Gerendas: C (Roche)
Osborne: E (Genentech, Inc)
Schmidt-Erfurth: C (Bayer Healthcare AG; Boehringer Ingelheim GmbH; Novartis Pharma AG)
all other authors: N



Introduction

The morphologic features of choroidal neovascularization (CNV) in patients with neovascular AMD show substantial inter-individual variability. However, existing systems to sub-classify CNV rely on predefined hypotheses, which may represent a bias and may hinder the discovery of novel biomarkers.

Our aim was to use an unbiased representation of spectral-domain OCT images to sub-classify AMD and to identify endophenotypes purely based on image information.

Data and Methodology

As a representative sample of treatment-naive neovascular AMD eyes, we trained our model on >54,000 Cirrus SD-OCT scans (512x128x1024) of 1096 patients enrolled in the HARBOR trial¹.

We propose to learn OCT volume representations in a completely unsupervised manner.

- In the first level, we used an auto-encoder to obtain low dimensional **embeddings of A-scans**.
- In the second level, the resulting coefficients for each volume were embedded with a second auto-encoder model to obtain **volume level embedding** coefficients.
- The low dimensional volume embeddings were **clustered using K-means**, where the number of clusters was selected automatically using the Davies-Bouldin (DB) index.

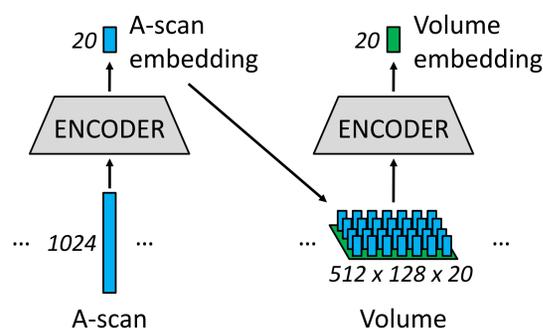


Fig. 1: Basic architecture of the embedding pipeline

Results: A-scan Embedding

The A-scan embedding captures several known disease characteristics such as intraretinal and subretinal fluid, pigment epithelial detachment and subretinal hyperreflective material, in addition to other heretofore unknown biomarkers represented in en-face maps (Fig. 2).

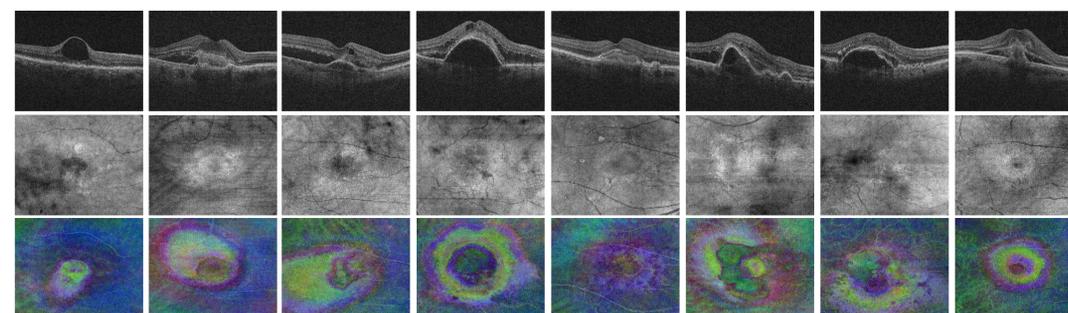


Fig. 2: Baseline volumes of eight different patients are illustrated. The first row shows the central B-Scan of the volume, while en-face view of the volumes are shown in the second (constructed as the sum over the A-scans) and third row (random coefficients used as RGB dimensions).

Results: Volume Embedding

In order to verify the meaningfulness of the individual dimensions of the volume embedding, we computed the Spearman correlation coefficient between the feature-activations and best corrected visual acuity (BCVA) for baseline data (Fig. 3a). Visualizations are provided to illustrate the learned volume embedding (Fig. 3b, 3c).

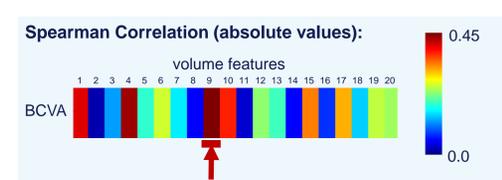


Fig. 3a: The correlation coefficient between feature activations and BCVA.



Fig. 3b: Plot of the volume embedding using t-SNE, a dimensionality reduction technique for visualization.

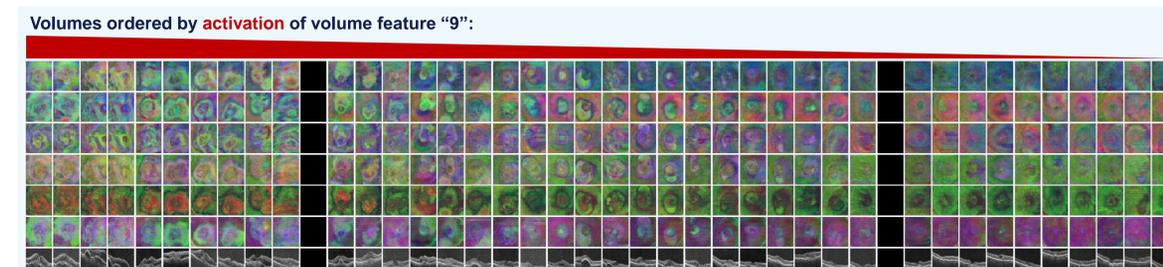


Fig. 3c: This plot shows the meaning of the feature with the highest correlation (see Fig. 3a). Each column corresponds to one volume. Volumes are ordered by activation, where the ten leftmost (rightmost) columns show volumes with highest (lowest) feature activation. While the first six rows show the en-face view of the embedded a-scan coefficients, the last row contains the central B-scan of the volume.

Results: Endophenotypes

The volume embedding enabled a compact and well separated clustering of patients indicated by a low Davies Bouldin index. As can be seen in Fig. 4, the clusters were visually plausible (e.g. C3 shows flat retina).

The detected endophenotypes showed clinically meaningful separation, e.g. in the distribution of lesion types. Of 34 tested SNPs, one gene showed an association with the clusters, indicated by a p-value of less than 0.05.

Variable	C1 (n=239)	C2 (n=234)	C3 (n=247)	C4 (n=154)	p-value
BCVA (BSL)	56±12	53±13	59±11	47±14	8.8 x 10⁻¹⁷
BCVA change at M3	8.6±10	7.8±12	6.1±11	7.2±11	0.22
BCVA change at M12	9.9±14	8.1±14	8.1±13	9.1±14	0.37
BCVA change at M24	8.9±16	7.7±16	8.0±15	7.8±16	0.75
Age	78±8	79±8	78±8	80±8	0.22
Gender (F)	61%	57%	58%	56%	0.76
Lesion type					
Occult	43%	30%	39%	26%	7.3 x 10⁻⁶
Pred. Classic	46%	50%	50%	46%	
Min. Classic	11%	20%	11%	28%	
"rs943080"(VEGFA)					
0 minor alleles	29%	26%	28%	22%	0.02
1 minor allele	46%	49%	54%	64%	
2 minor alleles	25%	25%	18%	14%	

Tab. 1: For all clusters (C1...C4), mean±SD and the p-value of Kruskal-Wallis test are reported for continuous variables, and relative frequencies and the p-value of chi-square test for categorical variables (no correction for multiple testing).

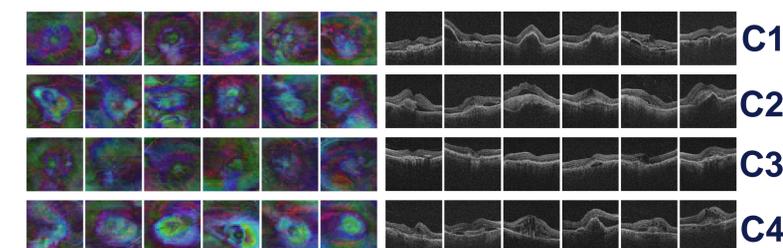


Fig. 4: The nearest neighbors of the cluster centers are shown using the en-face view (left) and the central B-scan (right).

Conclusion

Unsupervised machine learning based purely on image data enabled a low-dimensional representation of OCT volumes. A subsequent clustering allowed identification of clinically meaningful novel disease endophenotypes in neovascular AMD.

Acknowledgements

This work was funded by the Austrian Federal Ministry of Science, Research and Economy, the FWF (I2714-B31), and Genentech. A Tesla K40 used for this research was donated by the NVIDIA Corporation.

philipp.seeböck@meduniwien.ac.at