Machine Learning to Analyze the Prognostic Value of Current Imaging Biomarkers in Neovascular Age-Related Macular Degeneration

Ursula Schmidt-Erfurth, MD,1 Hrvoje Bogunovic, PhD,1 Amir Sadeghipour, PhD,1 Thomas Schlegl, MSc,1 Georg Langs, PhD,2 Bianca S. Gerendas, MD,1 Aaron Osborne, MBBS,3 Sebastian M. Waldstein, MD, PhD1

**Purpose:** To evaluate the potential of machine learning to predict best-corrected visual acuity (BCVA) outcomes from structural and functional assessments during the initiation phase in patients receiving standardized ranibizumab therapy for neovascular age-related macular degeneration (AMD).

**Design:** Post hoc analysis of a randomized, prospective clinical trial.

**Participants:** Data of 614 evaluable patients receiving intravitreal ranibizumab monthly or pro re nata according to protocol-specified criteria in the HARBOR trial.

**Methods:** Monthly spectral-domain (SD) OCT volume scans were processed by validated, fully automated computational image analysis. This system performs spatially resolved 3-dimensional segmentation of retinal layers, intraretinal cystoid fluid (IRF), subretinal fluid (SRF), and pigment epithelial detachments (PED). The extracted quantitative OCT biomarkers and BCVA measurements at baseline and months 1, 2, and 3 were used to predict BCVA at 12 months using random forest machine learning. This approach was also used to correlate OCT morphology to BCVA at baseline (structure-function correlation).

**Main Outcome Measures:** Accuracy (R²) of the prediction models; ranking of input variables.

**Results:** Computational image analysis enabled fully automated quantitative characterization of neovascular lesions in a large-scale clinical SD-OCT data set. At baseline, OCT features and BCVA were correlated with R² = 0.21. The most relevant biomarker for BCVA was the horizontal extension of IRF in the foveal region, whereas SRF and PED ranked low. In predicting functional outcomes, the model’s accuracy increased in a linear fashion with each month. If only the baseline visit was considered, the accuracy was R² = 0.34. At month 3, final visual acuity outcomes could be predicted with an accuracy of R² = 0.70. The strongest predictive factor for functional outcomes at 1 year was consistently the individual BCVA level during the initiation phase.

**Conclusions:** In this large-scale study based on a wide spectrum of morphologic and functional features, baseline BCVA correlated modestly with baseline SD-OCT, whereas functional outcomes were determined by BCVA levels during the initiation phase with a minor influence of fluid-related features. This finding suggests a re-evaluation of current diagnostic imaging features and a search for novel imaging approaches, where machine learning is a promising approach. Ophthalmology Retina 2018;2:24-30 © 2017 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplemental material available at www.ophthalmologyretina.org.

Intravitreal anti–vascular endothelial growth factor (anti-VEGF) therapy is the current standard of care for neovascular age-related macular degeneration (AMD).1 In the pivotal clinical trials, patients receiving anti-VEGF treatment gained on average 1 to 2 lines in best-corrected visual acuity (BCVA) from baseline at 1 year of therapy.2,3 However, the functional response to treatment on an individual patient level is markedly heterogeneous and difficult to predict clinically. For instance, in the large-scale Comparison of AMD Treatments Trials, at year 1, roughly 30% of patients showed a BCVA gain of 3 lines or more, whereas about 10% of patients experienced a BCVA loss of 1 line or more.4 Therefore, to counsel patients appropriately, and also to provide more reliable end points for clinical trials, identification of precise and robust methods to predict BCVA outcomes on an individual patient level represents an important goal of research.

Extensive research efforts have been directed at the discovery of structural parameters (“imaging biomarkers”) that would allow a more accurate functional prognosis in the management of neovascular AMD.5 In general, the BCVA level at baseline has become an established prognostic factor for functional gains and final BCVA outcomes.3 Although patients with higher initial BCVA achieve, on
average, less relative BCVA gain than individuals with pronounced pre-existing BCVA loss (a phenomenon known as the “ceiling effect”), best absolute BCVA outcomes are observed in patients with high baseline BCVA. Concerning imaging biomarkers with relevance for vision outcomes, mainly OCT-based features such as the presence and extent of intraretinal cystoid fluid (IRF) or photoreceptor signal loss have been demonstrated to correlate with BCVA outcomes, along with other markers such as size of the choroidal neovascularization (CNV) lesion on fluorescein angiography.6,7

Machine learning is a subfield of computer science that constructs automated algorithms to empirically recognize pathognomonic and prognostic patterns in large-scale multivariable datasets, rather than relying on predefined hypotheses.8 The spectrum of potential variables extracted from different morphologic and functional data sources is therefore unlimited. Methods from this field of artificial intelligence are currently introduced into ophthalmology in a pioneering effort to predict recurrence of disease or therapeutic needs in anti-VEGF therapy, to predict progression in atrophic AMD, but also to construct realistic segmentation algorithms for morphologic features.9–11

The aim of this study is to introduce machine learning methodology to, first, correlate morphologic OCT parameters at baseline to the corresponding visual function in active neovascular disease; and second, predict final BCVA levels after 1 year of standardized anti-VEGF therapy from functional and structural parameters acquired during the initiation phase in a large-scale randomized clinical trial setting.

Methods

In this post hoc analysis of a comprehensive clinical trial database, prospectively collected BCVA data and spectral-domain (SD) OCT images of patients enrolled in the HARBOR trial were included.12 The study was conducted in compliance with the Declaration of Helsinki. Approval was obtained by the Ethics Committee at the Medical University of Vienna as well as at each participating center for the HARBOR trial. Patients provided written informed consent for inclusion into the HARBOR trial. The HARBOR trial is registered at clinicaltrials.gov (identifier NCT00891735).

Study Design and Inclusion and Exclusion Criteria

The study design and main outcomes of the HARBOR trial have been published previously.12 In brief, patients with treatment-naive subfoveal CNV secondary to AMD as diagnosed by a retina specialist using fluorescein angiography and SD-OCT were included. Eligibility for the study was confirmed by a central reading center. Patients had to be aged 50 years or older and were eligible if BCVA was between 20/40 and 20/320 (Snellen equivalent). At baseline, all patients were randomized 1:1:1:1 to receive intravitreal ranibizumab 0.5 mg monthly, ranibizumab 0.5 mg pro re nata (PRN; after a 3-monthly initiation phase), ranibizumab 2.0 mg monthly, or ranibizumab 2.0 mg PRN. At each monthly visit, patients underwent BCVA testing using Early Treatment Diabetic Retinopathy Study (ETDRS) charts by certified examiners after formal refraction. SD-OCT was performed by certified operators using the Cirrus HD-OCT III instrument (Carl Zeiss Meditec, Dublin, CA), having 512 × 128 × 1024 voxels with a size of 11.7 × 47.2 × 2.0 μm3, covering a volume of 6 × 6 × 2 mm3.

Computational Image Analysis

Of the HARBOR dataset (n = 1095), 70% were randomly selected for analysis. The remaining 30% were kept for future studies. All SD-OCT images from baseline to month 3 underwent a standardized analysis for imaging biomarkers at a certified reading center (Vienna Reading Center, Vienna, Austria). A validated, fully automated computational image analysis
pipeline was used to process SD-OCT data (Fig 1). First, SD-OCT volumes were preprocessed using motion artifact removal to reduce image artifacts caused by saccadic and anterior–posterior eye motion during acquisition. Subsequently, fully automated segmentation algorithms based on graph theory and convolutional neural networks were applied to delineate the retinal layers and the CNV-associated lesion components, IRF, subretinal fluid (SRF), and pigment epithelial detachment (PED) (Schlegl T, Waldstein SM, Bogunovic H, et al. Fully Automated Detection and Quantification of Macular Fluid in Optical Coherence Tomography using Deep Learning, submitted for publication). Segmentation of fluid-filled layers, including total retinal thickness, IRF, SRF, and PED, resulted in 4 morphologic maps (Fig 2). To define the extension of the feature vector, the macular retina was divided into 9 areas according to the ETDRS grid. Locations included the foveal area and 4 parafoveal as well as 4 perifoveal areas for the nasal, temporal, superior, and inferior quadrants (Fig 2). The resulting wide range of quantitative structural variables (Table 1, available at www.ophthalmologyretina.org) was stored in a modeling database for subsequent machine learning.

Structure–Function Correlation using Machine Learning

To evaluate the relevance of the quantified SD-OCT features for visual function at baseline, where the variability in OCT morphology is greatest, a random forest regression model was trained using SD-OCT imaging biomarkers listed in Table 1 as input features and BCVA at baseline as a target variable. The random forest was grown with 500 trees, for which it was empirically observed that the out-of-bag error had converged. The relative importance of the individual input variables for the prediction was evaluated by permuting the values of each input variable and measuring any decrease in the prediction accuracy of the model. Thus, the input variables are ranked according to their individual value for an accurate prediction.

Prediction of Visual Outcomes using Machine Learning

Supervised machine learning regression using random forests was applied to predict BCVA at 12 months (target variable) from the input variables listed in Table 1. Separate models were constructed for the visits at baseline and at month 1 to month 3. The ranibizumab dose and treatment regimen were included in the model as fixed effects.

Statistical Analysis

This was an exploratory post hoc analysis of a comprehensive and prospective clinical trial database. Hence, no formal hypotheses were made and no P values are reported. The goodness of fit of the predictive models was evaluated by comparing the R^2 of the
models. All models were evaluated using a fivefold cross-validation strategy.

**Results**

**Patient Disposition**

After removal of patients showing incomplete data during the initiation phase and/or segmentation errors, data of 614 eyes were included into the current analysis, comprising a total of 2456 SD-OCT volumes for analysis. Descriptive statistics of individual CNV lesion characteristics as per automated image analysis are provided in Table 2 (available at www.ophthalmologyretina.org).

**Structure—Function Correlation using Machine Learning**

At baseline, the random forest regression model was used to predict the corresponding level of BCVA from the extracted morphologic imaging biomarkers, resulting in an accuracy of $R^2 = 0.21$. A scatterplot of predicted vs. ground truth (i.e., measured BCVA) is shown in Figure 3. The root mean square error was 11.4 letters. The resulting ranking of feature importance was led by the feature IRF and demonstrated that the horizontal extension of IRF in the central 1-mm and 3-mm foveal area, as well as IRF volume in the central 1 mm, conveyed the highest predictive power for concomitant BCVA, followed by total retinal thickness (Fig 3). Importantly, parameters quantifying SRF and PED did not contribute to baseline BCVA in a relevant manner, irrespective of their macular location.

**Prediction of Visual Outcomes using Machine Learning**

At baseline (i.e., before treatment initiation), the regression model including functional and anatomic data predicted BCVA at 12 months with an accuracy of $R^2 = 0.36$ and a root mean square error of 12.9 letters. The most relevant features for outcome prediction included baseline BCVA, followed by IRC area and volume (Fig 4). Scatterplots of predicted vs. ground truth BCVA from month 0 to month 3 are shown in Figure 4. The accuracy of prediction increased with each month into the initiation phase. At month 3, the model demonstrated the highest predictive power with an $R^2 = 0.70$ and a root mean square error of 8.6 letters. For the prediction of functional outcomes, the last measured BCVA during the initiation phase was by far the most important predictive factor, and this was consistent throughout each interval during the initiation phase. Morphologic variables ranked significantly lower at each individual month during the initiation phase, and this also applied to IRF, identified as the leading morphologic feature in structure—function correlation at baseline.

**Discussion**

The ability to transform data into knowledge using machine learning will progressively disrupt clinical medicine. Current prognostic models rely on a handful of qualitative variables restricted to specific morphologic characteristics recognized in previous clinical observations, which may be subject to pre-existing bias. Machine learning, conversely, approaches diagnostic and prognostic problems by extracting rules from data. Starting with patient-level observations, algorithms sift through vast amounts of features, looking for hallmarks that reliably predict outcomes.

Large-scale clinical trials in retinal pharmacotherapy provide structured datasets supporting generation of proof of principle for machine learning methods and their breakthrough potential in modern ophthalmology. With the advent of 3-dimensional raster scanning in SD-OCT the age of “big data” has clearly reached ophthalmic research. Image datasets as large as the one investigated here defy manual inspection and analysis by clinicians and also by specialized reading centers. Automated computational analysis of retinal imaging data promises an efficient and more refined mode of analysis and opens a window of opportunity in biomarker research. The utility of big data analysis covers the most important tasks in medicine overall, which are to anticipate, prevent, and treat disease in an optimal way for each individual patient. The HARBOR trial offers an exceptionally suitable dataset for machine learning because of its large sample size, standardized high-resolution SD-OCT imaging, and well-designed and well-executed treatment protocol.

In the study presented herein, a random forest regression algorithm was used to predict the development of BCVA at 1 year based on the initial therapeutic response known to exert the most substantial impact on disease morphology and visual function. The correlation between predicted and ground-truth 12-month BCVA scores was loose at baseline. At month 3, by contrast, the individual visual acuity levels reached a solid predictive power for BCVA outcomes at month 12. Despite input of a huge volume of precise spatiotemporal morphologic variables vastly exceeding that of previous statistical analyses, our study indicates that the individual longer-term BCVA outcome in neovascular AMD therapy is strongly dependent on the initial BCVA response to treatment. BCVA at month 3 represented by far the strongest predictive factor in the machine learning models and explained 70% of the variability of BCVA development between month 3 and month 12. Morphologic features, despite a comprehensive and disease-specific characterization of imaging variables, were found to be largely irrelevant for BCVA outcome in neovascular disease. Noteworthy, the analyses of retinal features included classifications beyond any feasible diagnostic evaluation in clinical routine.

The study of “imaging biomarkers” has become an important field of research, with several post hoc analyses of large clinical trial datasets providing highly relevant data. Our analysis of the HARBOR dataset aims far beyond previous studies by using innovative computer science technology, which generates knowledge in a data-driven way rather than focusing on pre-existing hypotheses. In the machine learning approach, a multitude of
features are recorded and a comprehensive search for correlations is performed. In our analysis, distinct morphologic features such as total retinal thickness and 3-dimensionally quantified IRF, SRF, and PED were screened in a differential manner, with the exact localization and time course of each feature included as well. Our data are, in general, consistent with previous studies in identifying baseline and/or month 3 BCVA as the most important individual variables for long-term BCVA outcomes based on extensive human expertise and manpower.17 Nevertheless, even throughout this phase of most intensive morphologic responsiveness, imaging biomarkers obtained by high-resolution SD-OCT showed only a limited predictive value, ranking far inferior compared with the predictive power of BCVA as an obvious functional biomarker. The Comparison of AMD Treatments Trials group performed a post hoc analysis of macular morphology and visual acuity and described solid associations between VA and morphologic features such as foveal IRF, abnormally thin retina, and SRF through years 1 and 2.6,21 The group also concluded, in another analysis, that the visual acuity response at week 12 is more predictive of 2-year vision outcomes than either several baseline characteristics (including total foveal thickness and elevation of the retinal pigment epithelium) or week 4 response.22

With respect to the impact of morphologic parameters on visual function from the structure–function correlation at baseline (i.e., before any therapeutic change has occurred), the machine learning methods used in our study are clearly putting the relations in adequate proportion. In treatment-naïve disease, BCVA correlated to SD-OCT parameters with an R² of 0.21. Our data imply that the classical exudative features such as fluid within and underneath the retina, as well as PED, may have limited value in explaining visual function in neovascular AMD, and they may not be informative in modifying the individual visual prognosis. This finding should trigger the search for additional morphologic markers that may not be fluid-related, such as a disruption of the external limiting membrane, the integrity of the inner and outer photoreceptor segments, or subretinal fibrosis and RPE atrophy.23,24 Other imaging modalities complementary to SD-OCT may be required to identify features of relevance, such as fundus autofluorescence or polarization-sensitive OCT.25,26 Interestingly, though, in the machine learning model, the horizontal extension of IRC in the foveal area was the most important predictor of BCVA, which corroborates the findings of a previous study of our group.27 In our previous analysis based on manual segmentation, the predictive importance of IRF on BCVA was larger, which may be a consequence of a smaller sample size.

As baseline BCVA is a critical predictor of later BCVA, early diagnosis and treatment initiation is of overwhelming importance and may matter more than a sophisticated morphologic analysis of established disease. The important impact of disease duration and time to first treatment has been studied globally,28,29 uniformly suggesting better outcomes with earlier detection and better initial BCVA.30 As the incidence of fellow-eye conversion toward neovascular AMD is as high as 20% in 2 years, second-eye screening is mandatory to preserve function in a better-seeing eye.31

This study is limited in its retrospective nature, which can cause selection bias. However, the inclusion of a random sample of the complete population of a randomized controlled trial conferred some of the strengths of a prospective study. Several retinal structures were automatically analyzed to provide a comprehensive characterization of CNV lesions. However, some important structures with known relevance for BCVA, such as the integrity of the ellipsoid zone and the external limiting membrane, were not considered because of the lack of broadly available automated segmentation methods for these structural features.32

In conclusion, this study introduces machine learning to evaluate the impact of classic fluid-based morphologic features in neovascular AMD on BCVA and, furthermore, to
predict BCVA after 1 year of anti-VEGF therapy in a large standardized dataset. When functional and structural parameters obtained during the initiation dose were used for outcome prediction, individual BCVA scores show by far the highest predictive values for visual acuity at month 12, explaining as much as 70% of variability. Among morphologic features, the horizontal IRC extension in a foveal location ranked highest in determining visual function in treatment-naïve neovascular AMD, but no currently used fluid-based morphologic feature showed any relevant role for prediction of the therapeutic response. The clinical utility of artificial intelligence methods such as machine learning in screening, anticipating, and treating disease is extremely promising and will largely enhance the understanding of the pathophysiology of retinal disease. Adequate algorithms detecting specific functionally relevant morphologic features, likely unrelated to the secondary leakage phenomena, must be developed to better understand AMD disease and therapeutic outcomes.

References


### Footnotes and Financial Disclosures

Originally received: March 21, 2017.
Final revision: March 30, 2017.
Accepted: March 30, 2017.

1. Christian Doppler Laboratory for Ophthalmic Image Analysis, Vienna Reading Center, Department of Ophthalmology, Medical University of Vienna, Vienna, Austria.
2. Christian Doppler Laboratory for Ophthalmic Image Analysis, Computational Imaging Research Lab, Department of Radiology and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria.

Financial Disclosures:
U.S.E.: Consultancy — Alcon, Bayer, Böhringer Ingelheim, and Novartis.
S.M.W.: Consultancy — Bayer and Novartis; research support — Genentech. All other authors have no financial disclosures.

Financial Support:
This study was supported in part by the Christian Doppler Research Society (Vienna, Austria) and Genentech (South San Francisco, CA). Genentech participated in review and approval of the manuscript.

Author Contributions:
Research design: Schmidt-Erfurth, Bogunovic, Langs, Gerendas, Osborne, Waldstein
Data analysis and/or interpretation: Schmidt-Erfurth, Bogunovic, Langs, Gerendas, Waldstein
Data acquisition and/or research execution: Bogunovic, Sadeghipour, Schlegl, Osborne
Manuscript preparation: Schmidt-Erfurth, Bogunovic, Osborne, Waldstein
Human Subjects: This study includes human subject/tissues. Study protocol was approved by IRB. Informed consent was obtained from all human subjects. All tenets of the Declaration of Helsinki were followed.

Abbreviations and Acronyms:
AMD = age-related macular degeneration; BCVA = best-corrected visual acuity; CNV = choroidal neovascularization; ETDTRS = Early Treatment Diabetic Retinopathy Study; IRF = intraretinal fluid; PED = pigment epithelial detachment; RPE = retinal pigment epithelium; SD-OCT = spectral-domain OCT; SRF = subretinal fluid; VEGF = vascular endothelial growth factor.

Correspondence: Prof. Dr. Ursula Schmidt-Erfurth, MD, Department of Ophthalmology, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria. E-mail: ursula.schmidt-erfurth@meduniwien.ac.at.