



Article

Prediction of Function in ABCA4-Related Retinopathy Using Ensemble Machine Learning

Philipp L. Müller ^{1,2,3,4,*}, Tim Treis ⁵ , Alexandru Odainic ¹ , Maximilian Pfau ^{1,6} ,
Philipp Herrmann ^{1,2}, Adnan Tufail ^{3,4} and Frank G. Holz ^{1,2}

¹ Department of Ophthalmology, University of Bonn, 53127 Bonn, Germany; S4alodai@uni-bonn.de (A.O.); Maximilian.Pfau@ukbonn.de (M.P.); Philipp.Herrmann@ukbonn.de (P.H.); Frank.Holz@ukbonn.de (F.G.H.)

² Center for Rare Diseases, University of Bonn, 53127 Bonn, Germany

³ Moorfields Eye Hospital NHS Foundation Trust, London EC1V 2PD, UK; Adnan.Tufail@nhs.net

⁴ Institute of Ophthalmology, University College London, London EC1V 9EL, UK

⁵ BioQuant, University of Heidelberg, 69120 Heidelberg, Germany; Tim.Treis@bioquant.uni-heidelberg.de

⁶ Department of Biomedical Data Science, Stanford University, Stanford, CA 94305-5464, USA

* Correspondence: Philipp.Mueller@ukbonn.de

Received: 24 June 2020; Accepted: 28 July 2020; Published: 29 July 2020



Abstract: Full-field electroretinogram (ERG) and best corrected visual acuity (BCVA) measures have been shown to have prognostic value for recessive Stargardt disease (also called “ABCA4-related retinopathy”). These functional tests may serve as a performance-outcome-measure (PerfO) in emerging interventional clinical trials, but utility is limited by variability and patient burden. To address these limitations, an ensemble machine-learning-based approach was evaluated to differentiate patients from controls, and predict disease categories depending on ERG (‘inferred ERG’) and visual impairment (‘inferred visual impairment’) as well as BCVA values (‘inferred BCVA’) based on microstructural imaging (utilizing spectral-domain optical coherence tomography) and patient data. The accuracy for ‘inferred ERG’ and ‘inferred visual impairment’ was up to $99.53 \pm 1.02\%$. Prediction of BCVA values (‘inferred BCVA’) achieved a precision of $\pm 0.3 \text{LogMAR}$ in up to 85.31% of eyes. Analysis of the permutation importance revealed that foveal status was the most important feature for BCVA prediction, while the thickness of outer nuclear layer and photoreceptor inner and outer segments as well as age of onset highly ranked for all predictions. ‘Inferred ERG’, ‘inferred visual impairment’, and ‘inferred BCVA’, herein, represent accurate estimates of differential functional effects of retinal microstructure, and offer quasi-functional parameters with the potential for a refined patient assessment, and investigation of potential future treatment effects or disease progression.

Keywords: retina; Stargardt disease; optical coherence tomography; visual acuity; electroretinogram; hereditary retinal disease; artificial intelligence

1. Introduction

Recessive Stargardt disease (STGD1) or ABCA4-related retinopathy is an important cause of inherited retinal degeneration and loss of vision in early life [1]. As an autosomal-recessive disorder, it is caused by biallelic mutations in the *ATP-binding cassette sub-family A member 4* (ABCA4) gene (Online Mendelian Inheritance in Man # 601691) [2]. It is characterized by a generally increased fundus autofluorescence (AF) intensity followed by distinct patterns of increased and decreased AF as well as thinning of retinal layers in the optical coherence tomography (OCT) associated with loss of function [3–5]. However, the clinical presentation and individual progression rate are highly variable [6–8].

In the view of emerging interventional clinical trials, the ability to identify patients by rate of expected disease progression would be advantageous in order to maximize the potential of an

intervention to demonstrate therapeutic effectiveness and improve the efficiency of future trials [7]. Several morphological and functional classifications have been proposed previously [9,10]. The full-field electroretinogram (ERG) based classification developed by Lois and colleagues has been shown to have prognostic value for disease progression and has been used in various studies to date [7–9,11]. However, the measurement of ERG is burdensome (i.e., time for dark adaptation), requires highly trained personnel, and is uncomfortable (e.g., due to corneal electrodes) [12]. The intensity of such tests makes patients' cooperation problematic and in conflict with the demand for study, protocols to be time- and resource-efficient.

Biological cause and effect relationships including retinal structure–function correlations are typically multifactorial (incl. categorical parameters) and non-linear making a conventional statistical approach problematic. Recently, machine learning (ML) approaches have been introduced into ophthalmology [13–15]. Using algorithms that learn to perform a specific task without being explicitly programmed to use conventional logic rules, ML approaches have the potential to elucidate previously unknown correlations and have been shown to predict retinal function based on structural parameters with great accuracy providing a quasi-functional parameter (so-called 'inferred sensitivity') in age-related macular degeneration (AMD) [14,15]. Accurate quasi-functional parameters based on objective and easily achievable structural data might allow for simplification of patient assessments and to make study protocols less burdensome for patients without losing potentially important information.

To date, ML applications for prediction of function in *ABCA4*-related retinopathy have not been explored. We, therefore, utilized ML approaches to explore the potential to predict ERG results, the grade of visual impairment, and best corrected visual acuity (BCVA) in *ABCA4*-related retinopathy based on patients' demographic factors and structural data of the retina. The structural retinal data were obtained using a non-invasive routinely collected OCT imaging device that is fast to obtain data. By combining multiple learning algorithms with different perspectives (ensemble ML), we aimed for more robust predictions [16]. The accuracy of the resulting 'inferred ERG', 'inferred visual impairment', and 'inferred BCVA' was validated in order to investigate the potential of these novel quasi-functional parameters in future patient assessment and clinical trial design.

2. Experimental Section

This retrospective monocenter cohort study was in adherence with the Declaration of Helsinki. The Institutional Review Board (IRB) of the University of Bonn approved the study (approval ID: #316/11 and #288/17). Written patients' informed consent was obtained after the explanation of the nature and possible consequences of the study.

2.1. Subjects

The study included patients with *ABCA4*-related retinopathy based on the presence of at least one disease-causing mutation in *ABCA4* as well as the absence of mutations in *Peripherin-2* (*PRPH2*) and a phenotype compatible with STGD1 [17,18]. Genetic testing was performed as either Sanger sequencing and multiplex ligation-dependent probe amplification or next-generation sequencing. Verification of identified mutations was carried out by PCR and subsequent Sanger sequencing [19]. Patients were recruited from a clinic dedicated to rare retinal diseases. Exclusion criteria were insufficient pupil dilation, additional retinal pathology, previous vitreoretinal surgery, or other ocular comorbidities substantially affecting visual function (e.g., significant media opacity, amblyopia, or optic nerve disease). Age-matched healthy subjects without ocular disease served as controls.

2.2. Imaging and Functional Testing

All subjects underwent a complete ophthalmologic examination including best corrected visual acuity (BCVA) testing using Early Treatment Diabetic Retinopathy Study (ETDRS) charts, slit lamp examination, indirect ophthalmoscopy, ERG testing in accordance to standards defined by the International Society for Clinical Electrophysiology of Vision (ISCEV) using a Ganzfeld stimulator

and Burian–Allan bipolar corneal electrodes (Toennies Multiliner Vision 1.70, Hochberg Germany), and standardized imaging protocol. Prior to imaging, pupils were dilated using 0.5% tropicamide and 2.5% phenylephrine. The imaging protocol consisted of AF-imaging (Spectralis HRA, Heidelberg Engineering, Heidelberg, Germany), and spectral domain (SD)-OCT (Spectralis HRA-OCT, Heidelberg Engineering, Germany). OCT was performed with single horizontal and vertical line scans centered on the fovea as well as volume scans ($25^\circ \times 30^\circ$, 61 scans) with at least 20 frames per scan averaged.

A proprietary approach for image analysis was implemented as previously described [15]. Briefly, the device-associated software (Spectralis Viewing Module 6.3.2.0, Heidelberg Engineering, Heidelberg, Germany) allowed for an initial automatic segmentation of volumetric OCT data. The segmentation was reviewed and, if indicated, manually corrected by a trained and experienced staff (A.O.). Correct Segmentation was further controlled by a senior grader (P.L.M.). As the automatic segmentation tool of the device-associated software was not programmed or trained for retinal pathologies associated with *ABCA4*-related retinopathy, manual correction was necessary in all patients. In cases with interruption of retinal structures (e.g., in areas of atrophy), the segmentation line was continued on the next outer segmented retinal structure. We defined the thickness between the internal limiting membrane (ILM) and the outer plexiform layer (OPL)-outer nuclear layer (ONL) boundary as ‘inner retinal (IR)’ thickness [20]. Henle fiber layer was counted towards the ‘ONL’. The photoreceptor ‘inner and outer segments (ISOS)’ ranged from band 1 (external limiting membrane, ELM) to band 3, and ‘retinal pigment epithelium (RPE)’ from band 3 to Bruch’s Membrane (BM, Figure 1) [20]. These four volumetric thickness maps for each layer were transferred as a tab-delimited file to ImageJ (U.S. National Institutes of Health, Bethesda, Maryland, USA) and overlaid with an ETDRS-Grid (i.e., circles with a diameter of 1, 3 and 6 mm, with the two eccentric rings separated into four quadrants) in order to measure the thickness of the respective layer in each of the nine subfields (Figure 1). The addition of all four retinal layers accounted for ‘full retinal (FR)’ thickness (i.e., thickness between ILM and BM).

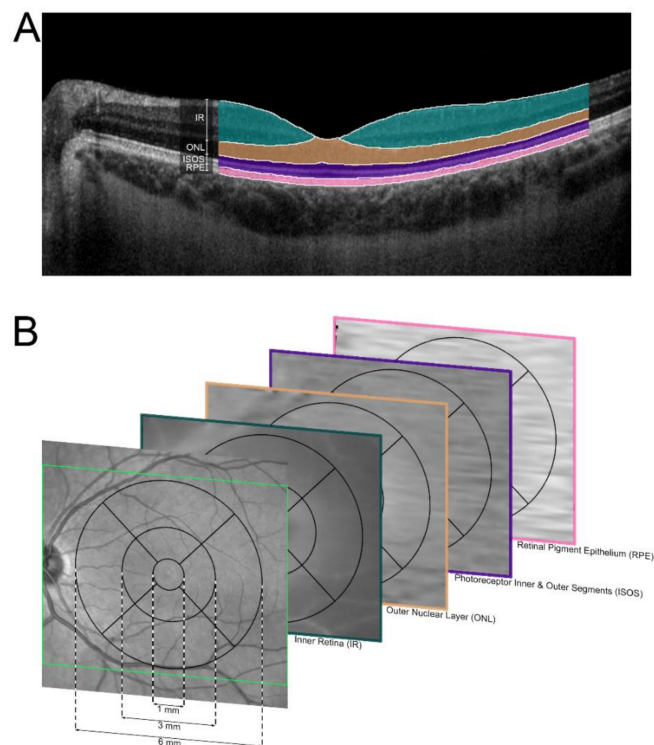


Figure 1. Image annotation and analysis. (A) Spectral-domain optical coherence tomography was

semi-automatically annotated. The herein used segmentations for the inner retina (IR, green overlay), outer nuclear layer (ONL, yellow overlay), photoreceptor inner and outer segments (ISOS, purple overlay), retinal pigment epithelium (RPE, pink overlay) are highlighted color-coded. (B) An ETDRS grid was registered to the thickness maps of the respective OCT layers. For each subsegment, the mean thickness was extracted.

2.3. Classification

Based on full-field ERG measurements, patients were categorized according to the classification developed by Lois and colleagues [9]: Group 1 contained eyes with normal scotopic and photopic responses, group 2 eyes with normal scotopic responses but reduced (over 2 standard deviations) photopic B-wave and 30-Hz flicker amplitudes, and group 3 eyes with impairment of both rod- and cone-driven responses. Using a multimodal assessment (AF and OCT), eyes were further graded according to the presence of foveal involving RPE atrophy by two independent experienced readers blinded to the results of each other, as central retinal thickness might directly be affected [21].

2.4. Machine Learning Analysis

Analysis was performed with Python (version 3.7.6; Python Software Foundation, Wilmington, Delaware, USA) using common data analysis packages including SciPy 1.0, Scikit-learn, XGBoost, and Eli5 [22–25]. In line with previous studies [7,26], BCVA was categorized into ‘no impairment’ (≤ 0.1 logMAR, $\geq 20/25$ Snellen equivalent), ‘mild impairment’ (0.2–0.5 logMAR, 20/32–20/60), ‘moderate impairment’ (0.6–0.9 logMAR (20/80–20/160), and ‘severe impairment’ (≥ 1.0 logMAR, $\leq 20/200$). Prior to modelling, categorical features were one-hot encoded (e.g., foveal status) and numerical features were normalized using a z-score transformation. For the BCVA and ERG classification as well as the BCVA regression, different ML algorithms were combined into soft voting ensembles [27–29]: For the prediction of the functional classifications, we used k-nearest-neighbor, random forest, support vector machine with radial basis function (rbf) kernel, and eXtreme Gradient Boosting (XGBoost) algorithm. For the functional regression, k-nearest-neighbor, kernel ridge, support vector machine with rbf kernel, and XGBoost. The algorithms were chosen based on previously demonstrated performance in the field of disease prediction through supervised learning [30,31]. To improve on their individual performance, the described sets of individual learners were combined into a stacked ensemble for more robust predictions [32]. These were then trained on three subsets of the available features: A) retinal layer thickness, B) full imaging feature set comprising retinal layer thickness and status of the fovea, and C) additional demographic features including age, refractive error (spherical equivalent), age of onset, and disease duration. The feature sets A and B included patients’ and controls’ data, while analysis with feature set C only used patients’ data as many demographic features (e.g., disease duration) was not defined for controls. We used the term ‘inferred ERG’ and ‘inferred visual impairment’ to describe the prediction of ERG and visual impairment category (classification models), and ‘inferred BCVA’ for BCVA value prediction (regression model) based on clinically feasible OCT imaging. For tuning the hyperparameters during model fitting, the F1-score (for classification) and the Root-Mean-Squared-Error (for regression) were optimized through nested cross-validation (for details see Supplementary Figure S1) as it significantly reduces bias when working with small datasets [33]. For both the inner and outer loop, a k_{fold} of 10 was chosen, resulting in an 81:9:10 split. The hyperparameters of the models used for the ensemble were optimized in the inner loop. Then, these optimized models were used to predict the holdout set of the outer loop. The importance of the used features was analyzed by employing permutation importance. Instead of removing the currently investigated feature and retraining the model, it replaces the feature with values randomly drawn from the distribution of the feature. The loss of information then indicates the importance of the relevant feature.

2.5. Statistical Analysis

Statistical analysis was also performed using Python (version 3.7.6; Python Software Foundation, Wilmington, Delaware, USA). For significance testing, different tests were employed depending on the nature of the tested data. For comparing two sets of data, either Welch's t-test or Mann–Whitney U-test was used, depending on the nature of the data. For comparing multiple sets of data, the Friedmann test was used, followed up with post-hoc Nemenyi testing. A significance threshold of $\alpha = 0.05$ was chosen for all tests.

3. Results

3.1. Cohort Characteristics

The study included a total of 311 eyes of 156 patients (101 female) with ABCA4-related retinopathy (mean age 38.77 years, total range 9.43–86.12 years) and 108 eyes of 54 controls (35 female, mean age 39.24 years, total range 13.14–77.02 years). One eye of one patient was excluded due to status post retinal detachment. The distributions of age at baseline and sex did not differ between patients and controls (Table 1 and Supplementary Figure S2). Individual demographic and genetic data can be found in Supplementary Table S1.

Table 1. Demographic and functional data of all included subject.

Patient-Level Data	Controls	All Patients	Group 1	Group 2	Group 3
Patients (n)	54	156	71	55	30
Sex (F/M)	35/19	101/55	46/25	36/19	19/11
Age at Exam (Y) ^A	39.24 ± 16.00	38.77 ± 18.23	33.76 ± 15.17	45.03 ± 19.521	39.15 ± 18.79
Age of Onset (Y) ^A	—	26.39 ± 17.16	24.75 ± 15.48	32.91 ± 18.93	18.30 ± 12.59
Disease Duration (Y) ^A	—	12.39 ± 12.45	9.01 ± 9.55	12.12 ± 12.00	20.85 ± 15.07
Eye-Level Data					
Eyes (n)	108	311	142	109	60
BCVA (LogMAR) ^A	0.00 ± 0.00	0.68 ± 0.47	0.57 ± 0.44	0.62 ± 0.43	1.06 ± 0.4
Refractive Error (Dpt) ^A	−0.80 ± 2.05	−1.15 ± 1.99	−1.23 ± 2.05	−0.64 ± 1.95	−1.88 ± 1.67
Foveal Status (FI/FNI)	—	197/114	79/63	65/44	53/7

^A Values indicate mean ± standard deviation. BCVA: best corrected visual acuity, dpt: diopters, f: female, FI: foveal involvement, FNI: foveal non-involvement, LogMAR: logarithm of the minimum angle of resolution, M: Male, Y: Years.

Based on the full-field ERG classification [9], 71 patients were assigned to group 1, 55 patients to group 2, and 30 patients to group 3. Both eyes of an individual patient always met the same group classification. The classification closely correlated with the phenotypes [4,8,10]: Group 1 eyes typically revealed localized foveal alterations surrounded by single flecks and a homogeneous background, group 2 eyes showed atrophic changes at the posterior pole with a more heterogeneous background caused by numerous flecks and group 3 eyes were associated multiple widespread atrophic areas at the macula with a heterogeneous background. The distribution of sex between the groups was similar. Group 2 patients revealed the highest age at exam and age of onset, while group 1 and 3 patients had the shortest and longest disease duration, respectively. Group 3 eyes were more myopic compared to eyes assigned to both other groups and showed foveal non-involvement of RPE atrophy much less frequently. The latter was in line with the significantly worse visual acuity in group 3 eyes compared to group 1 and 2 eyes (Table 1 and Supplementary Figure S2).

3.2. Retinal Layer Thickness

Overall, patients with ABCA4-related retinopathy revealed reduced macular full retinal thickness compared to controls (whole ETDRS grid [mean ± standard deviation]: patients, 247.92 ± 36.18 μm ; controls, 305.63 ± 13.81 μm ; $p < 0.001$; Supplementary Figure S3). This difference mainly derived from a thinner ONL (patients, 33.46 ± 13.83 μm ; controls, 61.88 ± 7.09 μm ; $p < 0.001$). ISOS (patients, 30.48 ± 13.60 μm ; controls, 44.15 ± 11.30 μm ; $p < 0.001$) and IR (patients, 149.67 ± 14.37 μm ; controls, 163.63 ± 11.30 μm ;

$p < 0.001$) also accounted for the thinning, while RPE (patients, $34.31 \pm 5.63 \mu\text{m}$; controls, $35.97 \pm 2.18 \mu\text{m}$; $p = 0.089$) did not show significant differences between patients and controls.

The differences in retinal thickness between patients and controls were most pronounced in the central ETDRS subfield (FR: patients, $156.10 \pm 55.04 \mu\text{m}$; controls, $285.11 \pm 25.05 \mu\text{m}$; $p < 0.001$), but still significant in the intermediate (FR: patients, $243.23 \pm 40.55 \mu\text{m}$; controls, $338.01 \pm 17.98 \mu\text{m}$; $p < 0.001$) and external (FR: patients, $252.71 \pm 37.48 \mu\text{m}$; controls, $296.80 \pm 14.46 \mu\text{m}$; $p < 0.001$) ETDRS subfields (Figure 2 and Supplementary Figure S3). The central retinal thinning was distinctly less expressed in eyes with foveal non-involvement (FR: foveal involvement, $131.28 \pm 36.19 \mu\text{m}$; foveal non-involvement, $201.67 \pm 54.57 \mu\text{m}$; $p < 0.001$), while the foveal status was of minor importance in the intermediate (FR: foveal involvement, $230.99 \pm 33.99 \mu\text{m}$; foveal non-involvement, $265.70 \pm 41.98 \mu\text{m}$; $p < 0.001$) and external (FR: foveal involvement, $246.24 \pm 37.94 \mu\text{m}$; foveal non-involvement, $264.58 \pm 33.53 \mu\text{m}$; $p < 0.001$) ETDRS subfields (Figure 2).

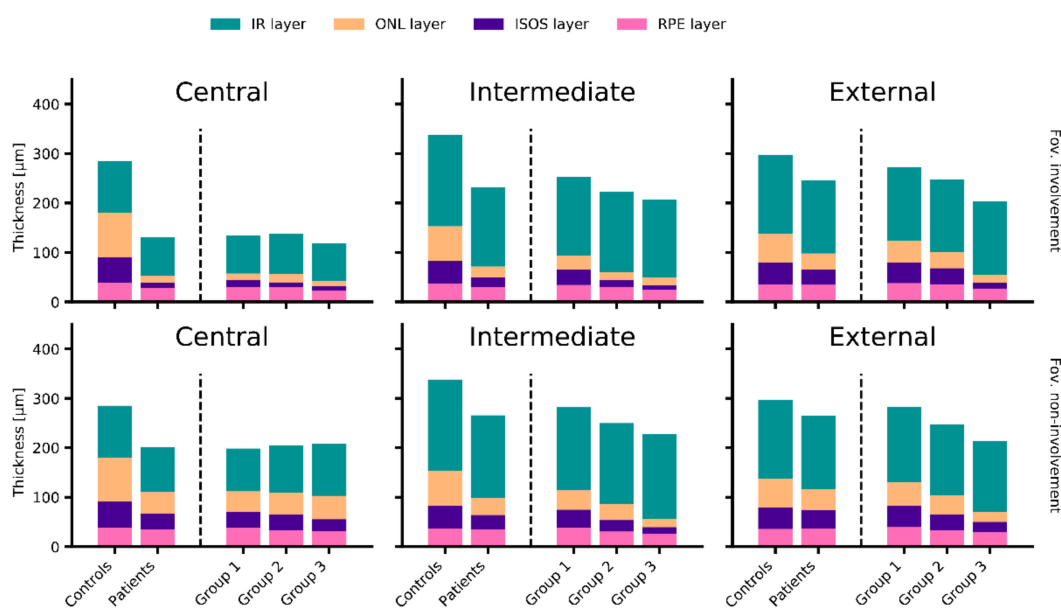


Figure 2. Retinal layer thickness measurements, eccentricity and foveal status. The bar plots demonstrate the mean retinal layer thickness measurements in the central subfield (first column), intermediate subfields (second column) and external subfields (third column) of the ETDRS, for controls’ and patients’ eyes depending on the presence (first row) or absence (second row) of foveal involvement. Patients’ eyes were further divided according to full-field electroretinogram based group classifications. Compared to controls, patients’ eyes revealed reduced retinal thickness in each eccentricity, especially in the outer nuclear layer (ONL, yellow) and the photoreceptor inner and outer segment layers (ISOS, purple). In the central subfield, the retinal thinning was expressed in eyes assigned to all groups of patients, while in the more eccentrically subfields, it was increasingly expressed from group 1 over group 2 to group 3 eyes. Concerning the foveal status, the most obvious difference was noticeable in the central subfield, where the retinal thinning was distinctly more pronounced in eyes with foveal involvement. The bar plots and exact numbers of measurements according to each ETDRS subfield can be found in Supplementary Figure S3 and in Supplementary Table S2, respectively. IR: inner retina (green), RPE: retinal pigment epithelium (pink).

Comparisons between the three ERG groups revealed that group 3 eyes showed the thinnest retina in all analyzed eccentricities ($p < 0.001$) with measures of $130.59 \pm 44.65 \mu\text{m}$, $209.89 \pm 29.95 \mu\text{m}$ and $204.76 \pm 27.38 \mu\text{m}$ for full retina in the central, intermediate and external ETDRS subfields. While group 1 and 2 revealed similar retinal thickness in the central subfield (full retina: $161.13 \pm 53.33 \mu\text{m}$ and $163.50 \pm 53.33 \mu\text{m}$ for group 1 and 2, respectively; $p = 0.426$), there was a significant difference in the intermediate (full retina: $265.31 \pm 33.86 \mu\text{m}$ and $232.91 \pm 36.78 \mu\text{m}$; $p < 0.001$) and external

ETDRS subfields (full retina: $276.76 \pm 20.02 \mu\text{m}$ and $247.81 \pm 32.15 \mu\text{m}$; $p < 0.001$; Figure 2 and Supplementary Figure S3).

In order to evaluate potential directional preferences of disease manifestation, we analyzed the retinal thinning relative to controls in the superior, inferior, nasal, and temporal subfields, which was overall significantly different ($p < 0.001$): The nasal ETDRS subfields revealed the highest and the superior subfields the lowest difference between patients and controls, respectively (Supplementary Figure S4).

Further layer thickness measurements for each ETDRS subfield can be found in Supplementary Figure S3 and Supplementary Table S2.

3.3. Prediction of Panretinal Function

Using our classifier ensemble model, panretinal function in terms of full-field ERG results (groups 1 to 3, as proposed by Lois and colleagues) [9] could accurately be predicted based on each feature set without much difference: With feature set A (i.e., pure objective retinal layer thickness data), the accuracy of the prediction was $99.68 \pm 0.64\%$ to differentiate controls and patients, as well as $94.63 \pm 2.57\%$, $93.52 \pm 2.49\%$, and $96.99 \pm 2.51\%$ to categorize patients into groups 1, 2, and 3, respectively (Figure 3A,D). Using additional features did not lead to significantly more accuracy in this context: For feature set B (i.e., all structural data), the model correctly differentiated controls from patients with $99.53 \pm 1.02\%$ accuracy, and patients into groups 1, 2 and 3 with accuracies of $93.84 \pm 2.18\%$, $93.36 \pm 2.22\%$ and $96.84 \pm 2.36\%$, respectively (Figure 3B,E). For feature set C (i.e., structural and demographic data), the accuracy of the prediction revealed values of $94.79 \pm 1.86\%$ (group 1), $95.11 \pm 1.92\%$ (group 2) and $97.47 \pm 2.03\%$ (group 3, Figure 3C,F).

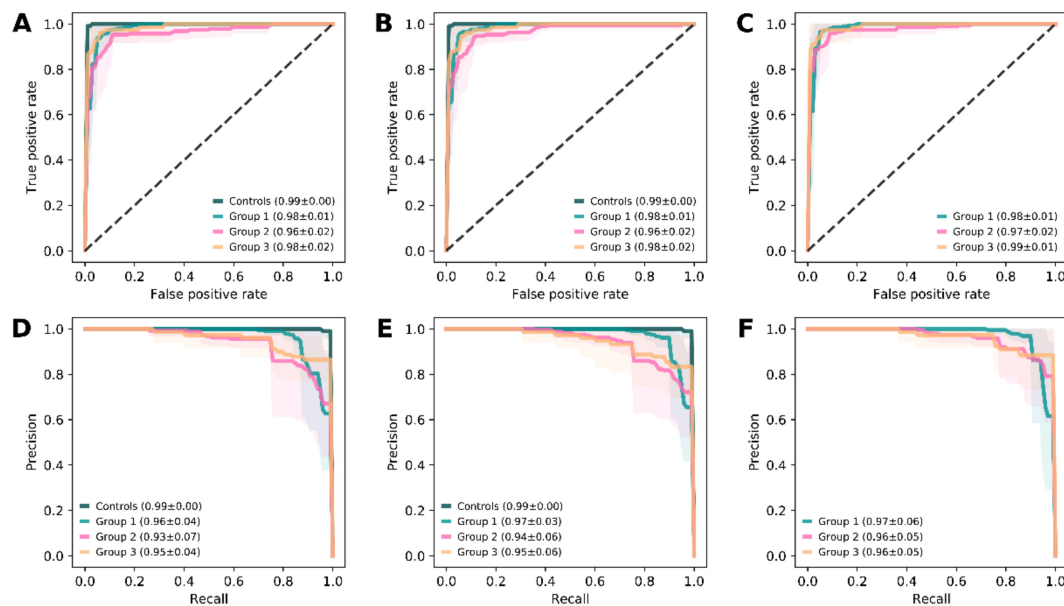


Figure 3. Model performance for the prediction of panretinal function. Receiver operating characteristic (ROC) curves that describe the true positive rate versus the false positive rate (top row, A–C), and graphs that plot the precision against the recall (bottom row, D–F) are demonstrated for the models predicting panretinal function based on feature set A (first column, A and D), B (second column, B and E), and C (third column, C and F). The lines represent the mean (\pm standard deviation, shaded area) of all folds for each group category according to full-field electroretinogram results, and the given values are the respective area under the curve (mean \pm standard deviation). The plots and respective values demonstrate the high robustness and power of the models to predict the panretinal function. The use of pure retinal thickness data (feature set A, first column) already reached the performance level of the models using additional features (feature set B, second column; feature set C, third column).

The weights that the model has assigned to each feature was again assessed using permutation importance. For pure retinal thickness features (feature set A; Supplementary Figure S5A) as well as all imaging features (feature set B; Supplementary Figure S5B), the thickness of the ONL, ISOS, and IR had distinctly more impact for the correct prediction of panretinal function than RPE and foveal status. All three eccentricities (Central, intermediate, and external) were represented among the most important features. Including additional demographic features (feature set C; Supplementary Figure S5C), age of onset and refractive error/spherical equivalent also showed high relevance, while age at exam and disease duration ranked within the second quarter of important factors for the model.

3.4. Prediction of Visual Impairment

Similar to panretinal function, the classifier ensemble models were trained to predict the grades of visual impairment: Using features set A, prediction of no, mild, moderate and severe visual impairment reached accuracies of $92.25 \pm 1.52\%$, $86.10 \pm 3.78\%$, $89.27 \pm 5.65\%$, and $88.64 \pm 4.01\%$, respectively (Figure 4A,D). With features set B, prediction showed accuracies at the same level with values of $93.68 \pm 2.25\%$, $89.10 \pm 3.27\%$, $90.23 \pm 5.62\%$, and $88.17 \pm 4.39\%$, respectively (Figure 4B,E). The same was true for feature set C, where the accuracies of the predictions were $91.44 \pm 3.42\%$, $86.84 \pm 3.75\%$, $89.57 \pm 3.83\%$, and $87.26 \pm 3.91\%$, respectively (Figure 4C,F).

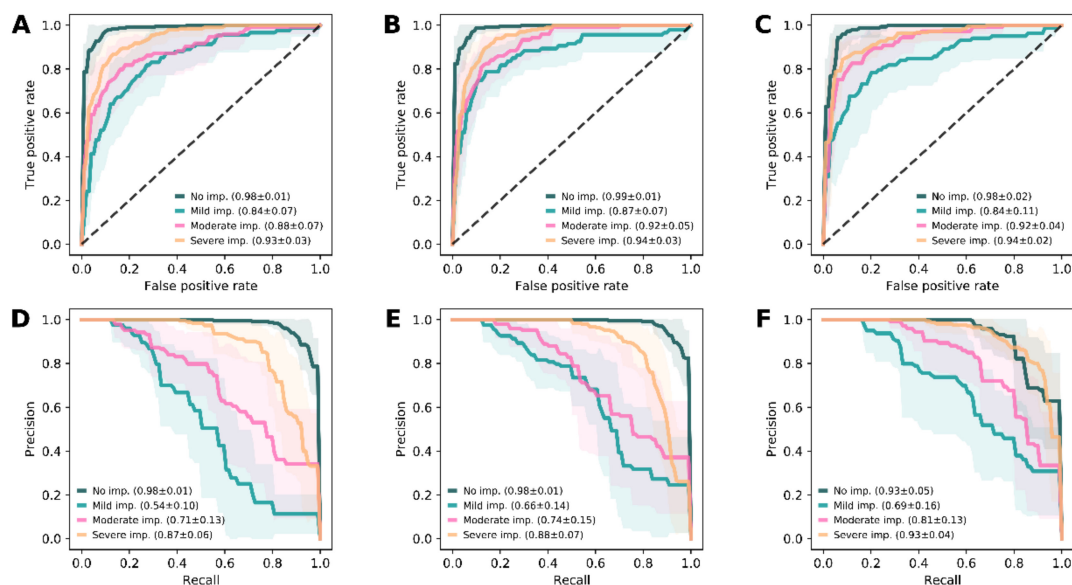


Figure 4. Model performance for the prediction of visual impairment. Receiver operating characteristic (ROC) curves that describe the true positive rate versus the false positive rate (top row, A–C), and graphs that plot the precision against the recall (bottom row, D–F), are demonstrated for the models predicting the grade of visual impairment (imp.) based on feature set A (first column, A and D), B (second column, B and E), and C (third column, C and F). The lines represent the mean (\pm standard deviation, shaded area) of all folds for category according to visual impairment, and the given values are the respective area under the curve (mean \pm standard deviation). The robustness and power of the models to predict the grade of visual impairment was best for correct prediction of the extreme values (no impairment, dark green; severe impairment, yellow). The use of pure retinal thickness data (feature set A, first column) already reached the performance level of the models using additional features (feature set B, second column; feature set C, third column).

Analysis of the individual structural feature importance (feature set A and B) revealed that the foveal status was the most important feature for predictions, followed by layer thickness measures (especially, ONL, ISOS, and IR) in the central ETDRS subfield. Measures in the intermediate and external

ETDRS subfields were of less impact. For feature set C, all demographic parameters were within the first twelve ranks but did not reach the feature importance of foveal status (Supplementary Figure S5D,F).

3.5. Prediction of Best Corrected Visual Acuity

Apart from the grades of visual impairment, the actual visual acuity values were predicted with ensemble ML algorithms. Compared to the pure retinal layer thickness model (feature set A, Figure 5A,D), inclusion of foveal status resulted in significantly improved accuracy of the prediction as shown by values for feature set B (Figure 5B,E) or feature set C (Figure 5C,F): for example, the percentage of BCVA prediction within ± 0.3 LogMAR (15 Letters, three lines on the LogMAR chart) of the ground truth was 72.67%, 85.31%, and 83.92% for the feature sets A, B and C, respectively (Table 2). In this context, the feature set B model revealed even less deviation from the ground truth compared to feature set C, which could be explained by the inclusion of controls with uniform good visual acuity in feature sets A and B. Irrespective of the model, the deviation between prediction and true value increased towards the extreme BCVA values (Supplementary Figure S6).

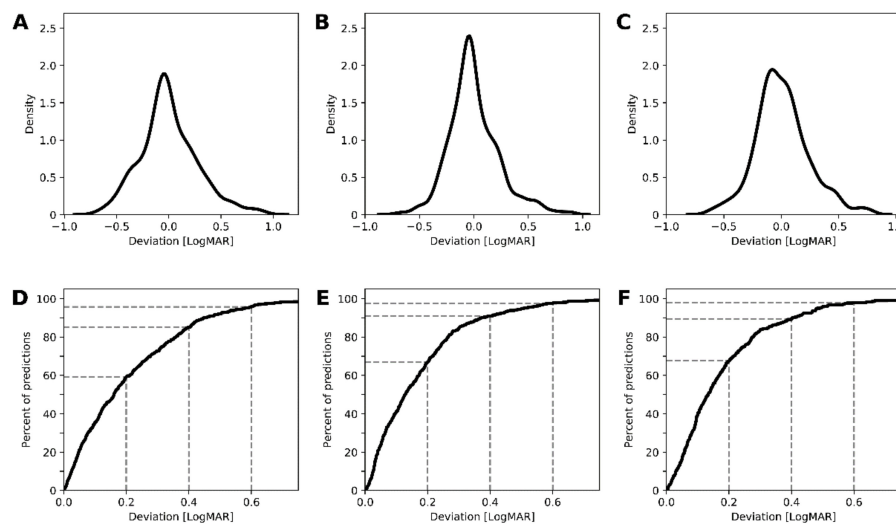


Figure 5. Model performance for the prediction of best corrected visual acuity (BCVA). The density plots that demonstrate the deviation of the predictions from the measured BCVA values (top row, A–C), and graphs that plot the cumulative true positive rate against the deviation (bottom row, D–F), are demonstrated for the models based on feature set A (first column, A and D), B (second column, B and E), and C (third column, C and F). The inclusion of foveal status with (feature set C, third column) or without (feature set B, second column) demographic data improved the performance of the ensemble machine learning algorithms compared to pure retinal thickness data (feature set A, first column). LogMAR: Logarithm of the Minimum Angle of Resolution.

Table 2. Accuracy of predicting best corrected visual acuity (BCVA).

Maximal Deviation from Truth [LogMAR]	Feature Sets [%] ^A		
	A	B	C
–0.5 To 0.5	92.10	96.37	95.82
–0.4 To 0.4	85.15	92.42	89.35
–0.3 To 0.3	72.67	85.31	83.92
–0.2 To 0.2	59.24	75.36	67.64
–0.1 To 0.1	35.39	53.55	39.87

^A Values indicate the percentages of predicted BCVA values that fall within the interval of deviation from the truth. LogMAR: logarithm of the minimum angle of resolution.

Analysis of the feature permutation importance values revealed that for prediction of BCVA, foveal status carried the most weight. This was followed by OCT layer thickness measures in the central ETDRS subfield, especially for ONL and ISOS. Including demographic factors, age of onset and disease

duration ranked at the level of RPE and IR measures in the central ETDRS subfield (i.e., behind foveal status and ONL and ISOS in the central subfield). Age at exam was of medium relevance and spherical equivalent carried very low importance (Supplementary Figure S5G–I).

4. Discussion

Based on a distinct structure-function correlation between OCT layer thickness and panretinal (ERG) as well as central retinal (BCVA) function, the possibility to predict retinal function in *ABCA4*-related retinopathy by imaging data and patients' demographic parameters was demonstrated by this study. These predictions of inferred functional categories ('inferred ERG' and 'inferred visual impairment') obtained by ensemble ML algorithms reached an accuracy of up to $97.47 \pm 2.03\%$ and $93.68 \pm 2.25\%$, respectively. The pure objective data from retinal layer thickness measures (feature set A) already achieved these levels of exactness. Including further simply determinable features (feature set B and C, especially foveal status) into the models, absolute BCVA values could also be predicted ('inferred BCVA') with high precision (Table 2). Apart from substantiating the functional relevance of structural alterations in *ABCA4*-related retinopathy, these novel calculated parameters may serve as quasi-functional surrogate markers in patient assessment and future clinical trials.

The term 'ML' covers different approaches to artificial intelligence that have been targeted on computers to learn without being explicitly programmed since 1959 [34]. In the last decade, ML techniques have entered visual science, including analysis of retinal images [13]. Due to demographic changes, the number of affected individuals with retinopathies, including diabetic retinopathy and AMD, has been increasing. The ML techniques were, therefore, mostly used for automatic imaging analysis in order to detect and classify pathologic features, where the power of conventional logic rules and linear approaches is quite restricted [35]. More recently, ML approaches have been used to predict retinal function based on structural parameters similar to the inferred retinal function ML model presented in this study that has shown to provide a highly sensitive tool. Rohm and colleagues predicted future course of visual acuity in treated neovascular AMD by database parameters and OCT data achieving comparable result to VA measurement reliability for three months [14]. Also, for neovascular AMD, von der Emde and colleagues achieved high accuracy by using ML analysis of OCT data to estimate fundus controlled perimetry results (also called 'inferred sensitivity'), giving a quasi-functional surrogate marker and an additional spatial dimension [15]. To date, only one publication showed the possible application of artificial intelligence in *ABCA4*-related retinopathy showing automatic cone photoreceptor localization for adaptive optics imaging in Stargardt disease, achromatopsia, and retinitis pigmentosa [36]. Previous studies for *ABCA4*-related retinopathy that investigated the evidence for structure–function correlations between retinal sensitivity and multimodal imaging only used a limited number of narrowly selected predictors and/or application of linear models [37,38]. As biological relationships including retinal structure–function correlations are typically multifactorial (including categorical parameters such as foveal status) and non-linear, correlations found in these studies have to be interpreted carefully. Especially categorical predictions (i.e., ERG-based groups or grade of visual impairment) might be difficult to assess (manual thresholds for each included factor). In contrast, our model was able to identify relationships between variables without manual intervention. Using an ensemble ML approach, different statistical perspectives were combined to increase the robustness of the demography- and morphology-based functional prediction (inferred retinal function) in *ABCA4*-related retinopathy [16]. Furthermore, this approach further enabled us to evaluate the effect and importance of individual predicting parameters.

The fact that (apart from foveal status) ONL, which includes light-sensitive photoreceptor cells bodies, and the ISOS, which includes the photoreceptor inner and outer segments (i.e., the location of phototransduction), consistently were among most important imaging-based predictors of retinal function in our model is biologically plausible. In a previous study of AMD, ONL exhibited the highest feature importance for inferred sensitivity [39]. Importance of features of the model may not necessarily indicate a biological correlation. Of note, IR was significantly thicker than other layers and, therefore,

less prone to grading errors. Together with the slight but uniform thinning of this layer, it possibly explains the relatively high feature importance of IR. The correlation of IR thinning with measures of the outer retinal layers might suggest some form of functional syncytium through the whole retina. This could explain previously described inner retinal changes in *ABCA4*-related retinopathy and retinal remodeling processes that accompany photoreceptor loss [40]. Further evidence of the biological plausibility of our model is given by the differential effect of location on central retinal and panretinal function: While the imaging feature importance ranking for BCVA prediction (category and value) was dominated by measurements within the central ETDRS subfield that covers the central 3° of the retina, measurements in the intermediate and external ETDRS subfields had more weight in predicting ERG results (i.e., panretinal function). The distribution of demographic factors between the different ERG-based groups explained their relatively high feature importance: for example, group 3 eyes had earliest age of onset, most negative median refractive error (spherical equivalent), and the longest disease duration (Table 1). Interestingly, a tendency towards negative refractive errors has been previously reported in the setting of *ABCA4*-related retinopathy [41].

Optimizing clinical trial design, including the selection of suitable patients and endpoints, constitutes a prerequisite towards an accurate assessment of potential therapeutic effects. Several approaches have been brought forward to predict progression in patients with *ABCA4*-related retinopathy. Some of those use genetic information, others rely on various imaging parameters [6,17]. The full-field electroretinogram (ERG) based classification developed by Lois and colleagues was shown to have a distinct prognostic value for disease progression and has, therefore, been used by several recent clinical studies [7,9,11]. However, the measurement of ERG is often described as uncomfortable and burdensome (i.e., time for dark adaptation), which is problematic when delivering clinical studies and time-efficient patient appointments [12]. We demonstrated the possibility to infer retinal function based on routinely obtained structural imaging and patients' demographic data using ensemble ML algorithms with high safety levels. The use of 'inferred BCVA', 'inferred visual impairment', and 'inferred ERG' as quasi-functional surrogate markers and endpoints could provide distinct advantages over the respective conventional functional measurements: (I) OCT has become a ubiquitously available imaging device and so commonly used to assess morphologic outcome measure in both routine examination and clinical trials [42]. Therefore, the utility of inferred retinal function could facilitate more time-efficient research studies, while offering comparable benefits to direct functional testing, including prediction of disease progression and objective differential testing of cone and rod function. (II) Digital imaging offers a higher spatial resolution and less variability than psychophysical testing. (III) The application of our models allows for simultaneous calculation of 'inferred BCVA', 'inferred visual impairment' and 'inferred ERG'. (IV) Structural data are faster to obtain and even applicable in patient unsuitable for psychophysical testing. (V) 'Inferred BCVA', 'inferred visual impairment' and 'inferred ERG' could be compared across diseases to potentially facilitate objective cost-benefit analysis. Giving these advantages, patient assessment, selection and study visits might become more time- and resource-effective by using inferred retinal function.

The International Conference on Harmonization Guidelines on Statistical Principles for Clinical Trials has established that the "evidence for surrogacy depends upon (I) the biological plausibility of the relationship, (II) the demonstration in epidemiologic studies of the prognostic value of the surrogate for the clinical outcome, and (III) evidence from clinical trials that treatment effects on the surrogate correspond to effects on the clinical outcome" [43]. The biological plausibility of the inferred retinal function has been shown as stated above. In contrast to conventional surrogate markers that are not directly related to function, 'inferred BCVA', 'inferred visual impairment', and 'inferred ERG' represent quasi-functional endpoints minimizing the relevance of the second and third criterion. Further, the use of nested cross-validation underlines the high accuracy of the model for unknown data. Nevertheless, ML models are primarily limited by the training data, which is the basis of the predictions. Other disease entities or atypical clinical presentations (as possible in interventional clinical trials) would necessitate further training data in order to define the appropriate applicability domain.

Of note, the complexity of the ensemble ML approach might theoretically require a slightly increased number of training examples compared to simple (i.e., non-ensemble) artificial intelligence approaches. Nevertheless, our results using ensemble ML algorithms revealed enhanced prediction accuracy for inferred function compared to published simple ML models for retinal diseases [14]. High accuracy is a prerequisite for future application of this quasi-functional surrogate marker. Apart from the utilization of a sophisticated ML model, the enhanced prediction accuracy in our cohort might be accounted for by looking at a simple monogenetic disease entity together with strict inclusion, as well as exclusion criteria in this study. We excluded other retinal diseases or comorbidities that could substantially affecting visual function including optic nerve diseases and cataract. This might result in a limitation of our study as additional retinal pathologies with functional relevance (e.g., changes in the inner retinal layers due to glaucoma) could be underrepresented in our model. However, clinical trials usually use as least as stringent inclusion criteria in order to generate a uniform study cohort, indicating that the restrictive inclusion and exclusion criteria might only be a theoretical limitation. Similar to recent studies (e.g., multicenter PROGSTAR study) [17,18], the diagnosis of *ABCA4*-related retinopathy was based on a compatible phenotype and the presence of at least one disease-causing mutation in *ABCA4* as well as the absence of mutations in *Peripherin-2* (*PRPH2*). Even if a more stringent definition (i.e., only patients with proven biallelic pathogenic mutation) would be more certain, the commonly accepted and here-used definition of *ABCA4*-related retinopathy is currently the best tradeoff between cohort size and limitations of genetic testing [44]. By using mean retinal layer thickness within the ETDRS subfields, focal alterations might not be detected and the predictive value it holds might not be attributed. For example, local thinning and thickening of RPE (due to flecks) might be balanced within the subfield. However, as group 3 eyes (that also showed most impaired BCVA) revealed consistently reduced RPE (Supplementary Table S2), this specific retinal layer, nevertheless, was attributed distinct feature importance (Supplementary Figure S5). A parameter that was not considered in the study was the area of peripapillary sparing, which can be assessed quantitatively and has been shown to correlate with electroretinographic findings in this disease entity [45,46]. Future models for ‘inferred ERG’ might include this parameter to further improve the predictions. To safely utilize such an ensemble ML model in future studies, the longitudinal accuracy of the model on the respective study cohort needs to be tested on a subset of subjects prior to inferring the functional data for the remaining participants. It has been established that BCVA (as well as ‘inferred BCVA’) does not constitute an optimal endpoint in *ABCA4*-related retinopathy due to its high interindividual variability and the phenomenon of foveal non-involvement [8]. The future evaluation of ‘inferred sensitivity’ based on fundus-controlled perimetry (as described for AMD [15]) might, therefore, be warranted as the high-resolution mapping of localized functional impairment might overcome this limitation and provide an additional dimension [47]. As this specific AMD study was also restricted to a simple ML model, the application of our ensemble ML approach (i.e., using different statistical perspectives) might also allow for improved accuracy in this context.

In conclusion, a consistent structure–function relationship in *ABCA4*-related retinopathy could be demonstrated. Based on this correlation, we have introduced the artificial intelligence-based analysis strategy of ‘inferred BCVA’, ‘inferred visual impairment’, and ‘inferred ERG’ to estimate differential effects of retinal structural alterations on central retinal and panretinal as well as cone and rod function in *ABCA4*-related retinopathy. This method potentially constitutes a valuable tool to predict retinal function without the need of psychophysical examinations enabling fast routine patients’ assessment and efficient study protocols without loss of information. The conception of ‘inferred BCVA’, ‘inferred visual impairment’, and ‘inferred ERG’ as quasi-functional outcome measures are not restricted to *ABCA4*-related retinopathy but might be further applied to other retinal diseases.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2077-0383/9/8/2428/s1>, Supplementary Figure S1: Visualization of the nested cross-validation, Supplementary Figure S2: Cohort characteristics of included participants, Supplementary Figure S3: Retinal layer thickness measurements, Supplementary Figure S4: Directional differences in retinal thickness, Supplementary Figure: Feature importance for the machine learning algorithms, Supplementary Figure S6: Deviation of the best corrected visual acuity (BCVA) predictions,

Supplementary Table S1: Individual demographic and genetic data of included patients with ABCA4-related retinopathy, and Supplementary Table S2: Retinal layer thickness measures for each layer and subfield.

Author Contributions: Conceptualization, P.L.M. and T.T.; methodology, P.L.M., T.T., M.P. and A.T.; formal analysis, P.L.M., T.T. and A.O.; investigation, P.L.M., A.O., P.H. and F.G.H.; resources, P.L.M., T.T., A.T., F.G.H.; data curation, P.L.M.; writing—original draft preparation, P.L.M. and T.T.; writing—review and editing, A.O., M.P., P.H., A.T. and F.G.H.; visualization, P.L.M. and T.T.; supervision, P.L.M., A.T. and F.G.H.; project administration, P.L.M.; funding acquisition, P.L.M., M.P., A.T. and F.G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Research Foundation (DFG, grant # MU4279/2-1 to PLM and PF950/1-1 to MP), and the Department of Health's NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital and UCL Institute of Ophthalmology (funding to AT).

Acknowledgments: The views expressed are those of the authors. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Conflicts of Interest: P.L.M., T.T., A.O. and M.P. declare no conflict of interest. P.H. has received non-financial support by Heidelberg Engineering, Carl Zeiss Meditec AG, and Optos. A.T. has received personal fees by Heidelberg Engineering, Novartis, Bayer, Genentech/Roche, Acucela, and Allergan, and grants by Novartis, Bayer, Genentech/Roche, Acucela, and Allergan. F.G.H. has received non-financial support by Heidelberg Engineering, CenterVue, Carl Zeiss Meditec AG, and Optos, and personal fees by Heidelberg Engineering, Novartis, Bayer, Genentech/Roche, Acucela, Boehringer Ingelheim, Alcon, Optos, and Allergan, and grants by Heidelberg Engineering, Novartis, Bayer, Genentech/Roche, Acucela, Alcon, Optos, and Allergan. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Hanany, M.; Rivolta, C.; Sharon, D. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2710–2716. [[CrossRef](#)] [[PubMed](#)]
- Koenekoop, R.K. The gene for Stargardt disease, ABCA4, is a major retinal gene: A mini-review. *Ophthalmic Genet.* **2003**, *24*, 75–80. [[CrossRef](#)] [[PubMed](#)]
- Müller, P.L.; Gliem, M.; Mangold, E.; Bolz, H.J.; Finger, R.P.; McGuinness, M.; Betz, C.; Jiang, Z.; Weber, B.H.F.; MacLaren, R.E.; et al. Monoallelic ABCA4 Mutations Appear Insufficient to Cause Retinopathy: A Quantitative Autofluorescence Study. *Invest. Ophthalmol. Vis. Sci.* **2015**, *56*, 8179–8186. [[CrossRef](#)] [[PubMed](#)]
- Müller, P.L.; Birtel, J.; Herrmann, P.; Holz, F.G.; Charbel Issa, P.; Gliem, M. Functional Relevance and Structural Correlates of Near Infrared and Short Wavelength Fundus Autofluorescence Imaging in ABCA4 -Related Retinopathy. *Transl. Vis. Sci. Technol.* **2019**, *8*, 46. [[CrossRef](#)]
- Ho, A.; Kuehlewein, L.; Hariri, A.; Wolfson, Y.; Strauss, R.W.; Scholl, H.P.; Sadda, S.R. ProgSTAR Study Group Quantitative Characteristics of Spectral-Domain Optical Coherence Tomography (SDOCT) in Corresponding Areas of Decreased Autofluorescence in Patients with Stargardt Disease. *Investig. Ophthalmol. Vis. Sci.* **2015**, *56*, 5924.
- Klevering, B.J.; Deutman, A.F.; Maugeri, A.; Cremers, F.P.M.; Hoyng, C.B. The spectrum of retinal phenotypes caused by mutations in the ABCA4 gene. *Graefes Arch. Clin. Exp. Ophthalmol.* **2005**, *243*, 90–100. [[CrossRef](#)]
- Müller, P.L.; Pfau, M.; Treis, T.; Pascual-Camps, I.; Birtel, J.; Lindner, M.; Herrmann, P.; Holz, F.G. Progression of abca4-related retinopathy-prognostic value of demographic, functional, genetic, and imaging parameters. *Retina* **2020**, *1*. [[CrossRef](#)] [[PubMed](#)]
- Müller, P.L.; Pfau, M.; Mauschitz, M.M.; Möller, P.T.; Birtel, J.; Chang, P.; Gliem, M.; Schmitz-Valckenberg, S.; Fleckenstein, M.; Holz, F.G.; et al. Comparison of Green Versus Blue Fundus Autofluorescence in ABCA4 -Related Retinopathy. *Transl. Vis. Sci. Technol.* **2018**, *7*, 13. [[CrossRef](#)] [[PubMed](#)]
- Lois, N.; Holder, G.E.; Bunce, C.; Fitzke, F.W.; Bird, A.C. Phenotypic subtypes of Stargardt macular dystrophy-fundus flavimaculatus. *Arch. Ophthalmol.* **2001**, *119*, 359–369. [[CrossRef](#)] [[PubMed](#)]
- Fujinami, K.; Lois, N.; Mukherjee, R.; McBain, V.A.; Tsunoda, K.; Tsubota, K.; Stone, E.M.; Fitzke, F.W.; Bunce, C.; Moore, A.T.; et al. A Longitudinal Study of Stargardt Disease: Quantitative Assessment of Fundus Autofluorescence, Progression, and Genotype Correlations. *Investig. Ophthalmol. Vis. Sci.* **2013**, *54*, 8181–8190. [[CrossRef](#)] [[PubMed](#)]
- Zahid, S.; Jayasundera, T.; Rhoades, W.; Branham, K.; Khan, N.; Niziol, L.M.; Musch, D.C.; Heckenlively, J.R. Clinical phenotypes and prognostic full-field electroretinographic findings in Stargardt disease. *Am. J. Ophthalmol.* **2013**, *155*, 465–473.e3. [[CrossRef](#)] [[PubMed](#)]

12. McCulloch, D.L.; Marmor, M.F.; Brigell, M.G.; Hamilton, R.; Holder, G.E.; Tzekov, R.; Bach, M. ISCEV Standard for full-field clinical electroretinography (2015 update). *Doc. Ophthalmol.* **2015**, *130*, 1–12. [[CrossRef](#)] [[PubMed](#)]
13. Caixinha, M.; Nunes, S. Machine Learning Techniques in Clinical Vision Sciences. *Curr. Eye Res.* **2017**, *42*, 1–15. [[CrossRef](#)] [[PubMed](#)]
14. Rohm, M.; Tresp, V.; Müller, M.; Kern, C.; Manakov, I.; Weiss, M.; Sim, D.A.; Priglinger, S.; Keane, P.A.; Kortuem, K. Predicting Visual Acuity by Using Machine Learning in Patients Treated for Neovascular Age-Related Macular Degeneration. *Ophthalmology* **2018**, *125*, 1028–1036. [[CrossRef](#)] [[PubMed](#)]
15. Von der Emde, L.; Pfau, M.; Dysli, C.; Thiele, S.; Möller, P.T.; Lindner, M.; Schmid, M.; Fleckenstein, M.; Holz, F.G.; Schmitz-Valckenberg, S. Artificial intelligence for morphology-based function prediction in neovascular age-related macular degeneration. *Sci. Rep.* **2019**, *9*, 11132. [[CrossRef](#)] [[PubMed](#)]
16. Džeroski, S.; Ženko, B. Is combining classifiers with stacking better than selecting the best one? *Mach. Learn.* **2004**, *54*, 255–273. [[CrossRef](#)]
17. Strauss, R.W.; Ho, A.; Muñoz, B.; Cideciyan, A.V.; Sahel, J.-A.; Sunness, J.S.; Birch, D.G.; Bernstein, P.S.; Michaelides, M.; Traboulsi, E.I.; et al. The Natural History of the Progression of Atrophy Secondary to Stargardt Disease (ProgStar) Studies. *Ophthalmology* **2016**, *123*, 817–828. [[CrossRef](#)]
18. Müller, P.L.; Pfau, M.; Möller, P.T.; Nadal, J.; Schmid, M.; Lindner, M.; de Sistiernes, L.; Stöhr, H.; Weber, B.H.F.; Neuhaus, C.; et al. Choroidal Flow Signal in Late-Onset Stargardt Disease and Age-Related Macular Degeneration: An OCT-Angiography Study. *Invest. Ophthalmol. Vis. Sci.* **2018**, *59*, AMD122–AMD131. [[CrossRef](#)]
19. Birtel, J.; Eisenberger, T.; Gliem, M.; Müller, P.L.; Herrmann, P.; Betz, C.; Zahnleiter, D.; Neuhaus, C.; Lenzner, S.; Holz, F.G.; et al. Clinical and genetic characteristics of 251 consecutive patients with macular and cone/cone-rod dystrophy. *Sci. Rep.* **2018**, *8*, 4824. [[CrossRef](#)]
20. Staurengi, G.; Sada, S.; Chakravarthy, U.; Spaide, R.F. International Nomenclature for Optical Coherence Tomography (IN•OCT) Panel Proposed Lexicon for Anatomic Landmarks in Normal Posterior Segment Spectral-Domain Optical Coherence Tomography. *Ophthalmology* **2014**, *121*, 1572–1578. [[CrossRef](#)]
21. Sayegh, R.G.; Sacu, S.; Dunavölgyi, R.; Kroh, M.E.; Roberts, P.; Mitsch, C.; Montuoro, A.; Ehrenmüller, M.; Schmidt-Erfurth, U. Geographic Atrophy and Foveal-Sparing Changes Related to Visual Acuity in Patients With Dry Age-Related Macular Degeneration Over Time. *Am. J. Ophthalmol.* **2017**. [[CrossRef](#)] [[PubMed](#)]
22. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. Author Correction: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 352. [[PubMed](#)]
23. Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn. *GetMobile Mob. Comput. Commun.* **2015**, *19*, 29–33. [[CrossRef](#)]
24. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
25. MIT ELI5. Available online: <https://eli5.readthedocs.io/en/latest/index.html> (accessed on 23 March 2020).
26. Müller, P.L.; Treis, T.; Pfau, M.; Esposti, S.D.; Alsaedi, A.; Maloca, P.; Balaskas, K.; Webster, A.; Egan, C.; Tufail, A. Progression of Retinopathy Secondary to Maternally Inherited Diabetes and Deafness—Evaluation of Predicting Parameters. *Am. J. Ophthalmol.* **2020**, *213*, 134–144. [[CrossRef](#)]
27. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [[CrossRef](#)]
28. Kotsiantis, S.B.; Zaharakis, I.D.; Pintelas, P.E. Machine learning: A review of classification and combining techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190. [[CrossRef](#)]
29. Dey, A. Machine Learning Algorithms: A Review. *Int. J. Comput. Sci. Inf. Technol.* **2016**, *7*, 1174–1179.
30. Uddin, S.; Khan, A.; Hossain, M.E.; Moni, M.A. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 1–16. [[CrossRef](#)]
31. Azar, A.T.; El-Metwally, S.M. Decision tree classifiers for automated medical diagnosis. *Neural Comput. Appl.* **2013**, *23*, 2387–2403. [[CrossRef](#)]
32. Sesmero, M.P.; Ledezma, A.I.; Sanchis, A. Generating ensembles of heterogeneous classifiers using Stacked Generalization. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2015**, *5*, 21–34. [[CrossRef](#)]
33. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **2006**, *7*, 1–8. [[CrossRef](#)] [[PubMed](#)]

34. Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **1959**, *3*, 210–229. [[CrossRef](#)]
35. De Fauw, J.; Keane, P.; Tomasev, N.; Visentin, D.; van den Driessche, G.; Johnson, M.; Hughes, C.O.; Chu, C.; Ledsam, J.; Back, T.; et al. Automated analysis of retinal imaging using machine learning techniques for computer vision. *F1000Research* **2016**, *5*, 1573. [[CrossRef](#)] [[PubMed](#)]
36. Davidson, B.; Kalitzeos, A.; Carroll, J.; Dubra, A.; Ourselin, S.; Michaelides, M.; Bergeles, C. Automatic Cone Photoreceptor Localisation in Healthy and Stargardt Afflicted Retinas Using Deep Learning. *Sci. Rep.* **2018**, *8*, 7911. [[CrossRef](#)]
37. Parodi, M.B.; Iacono, P.; Triolo, G.; La Spina, C.; Zucchiatti, I.; Cicinelli, M.V.; Borrelli, E.; Manitto, M.P.; Martina, E.; Bandello, F. Morpho-functional correlation of fundus autofluorescence in Stargardt disease. *Br. J. Ophthalmol.* **2015**, *99*, 1354–1359. [[CrossRef](#)]
38. Ergun, E.; Hermann, B.; Wirtitsch, M.; Unterhuber, A.; Ko, T.H.; Sattmann, H.; Scholda, C.; Fujimoto, J.G.; Stur, M.; Drexler, W. Assessment of central visual function in Stargardt’s disease/fundus flavimaculatus with ultrahigh-resolution optical coherence tomography. *Invest. Ophthalmol. Vis. Sci.* **2005**, *46*, 310–316. [[CrossRef](#)]
39. Von der Emde, L.; Pfau, M.; Thiele, S.; Moller, P.T.; Hassenrik, R.; Fleckenstein, M.; Holz, F.G.; Schmitz-Valckenberg, S. Mesopic and Dark-Adapted Two-Color Fundus-Controlled Perimetry in Choroidal Neovascularization Secondary to Age-Related Macular Degeneration. *Transl. Vis. Sci. Technol.* **2019**, *8*, 7. [[CrossRef](#)]
40. Huang, W.C.; Cideciyan, A.V.; Roman, A.J.; Sumaroka, A.; Sheplock, R.; Schwartz, S.B.; Stone, E.M.; Jacobson, S.G. Inner and Outer Retinal Changes in Retinal Degenerations Associated With ABCA4 Mutations. *Investig. Ophthalmol. Vis. Sci.* **2014**, *55*, 1810. [[CrossRef](#)]
41. Hendriks, M.; Verhoeven, V.J.M.; Buitendijk, G.H.S.; Polling, J.R.; Meester-Smoor, M.A.; Hofman, A.; Kamermans, M.; Ingeborgh van den Born, L.; Klaver, C.C.W.; van Huet, R.A.; et al. Development of Refractive Errors—What Can We Learn From Inherited Retinal Dystrophies? *Am. J. Ophthalmol.* **2017**, *182*, 81–89. [[CrossRef](#)]
42. Müller, P.L.; Wolf, S.; Dolz-Marco, R.; Tafreshi, A.; Schmitz-valckenberg, S.; Holz, F.G. Ophthalmic Diagnostic Imaging: Retina. In *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics*; Bille, J.F., Ed.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 87–106. ISBN 978-3-030-16637-3.
43. International Conference on Harmonisation E9 Expert Working Group ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. *Stat. Med.* **1999**, *18*, 1905–1942.
44. Zernant, J.; Schubert, C.; Im, K.M.; Burke, T.; Brown, C.M.; Fishman, G.A.; Tsang, S.H.; Gouras, P.; Dean, M.; Allikmets, R. Analysis of the ABCA4 gene by next-generation sequencing. *Invest. Ophthalmol. Vis. Sci.* **2011**, *52*, 8479–8487. [[CrossRef](#)] [[PubMed](#)]
45. Burke, T.R.; Rhee, D.W.; Smith, R.T.; Tsang, S.H.; Allikmets, R.; Chang, S.; Lazow, M.A.; Hood, D.C.; Greenstein, V.C. Quantification of peripapillary sparing and macular involvement in Stargardt disease (STGD1). *Invest. Ophthalmol. Vis. Sci.* **2011**, *52*, 8006–8015. [[CrossRef](#)] [[PubMed](#)]
46. Nassisi, M.; Mohand-Saïd, S.; Andrieu, C.; Antonio, A.; Condroyer, C.; Méjécase, C.; Dhaenens, C.M.; Sahel, J.A.; Zeitz, C.; Audo, I. Peripapillary sparing with near infrared autofluorescence correlates with electroretinographic findings in patients with Stargardt disease. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, 4951–4957. [[CrossRef](#)] [[PubMed](#)]
47. Pfau, M.; Holz, F.G.; Müller, P.L. Retinal light sensitivity as outcome measure in recessive Stargardt disease. *Br. J. Ophthalmol.* **2020**, *4*. [[CrossRef](#)] [[PubMed](#)]

