



# AI-based structure-function correlation in age-related macular degeneration

Leon von der Emde<sup>1</sup> · Maximilian Pfau<sup>1,2</sup> · Frank G. Holz<sup>1</sup> · Monika Fleckenstein<sup>3</sup> · Karsten Kortuem<sup>4,5</sup> ·  
Pearse A. Keane<sup>6</sup> · Daniel L. Rubin<sup>2</sup> · Steffen Schmitz-Valckenberg<sup>1,3</sup>

Received: 28 December 2020 / Revised: 24 February 2021 / Accepted: 9 March 2021 / Published online: 25 March 2021  
© The Author(s) 2021. This article is published with open access

## Abstract

Sensitive and robust outcome measures of retinal function are pivotal for clinical trials in age-related macular degeneration (AMD). A recent development is the implementation of artificial intelligence (AI) to infer results of psychophysical examinations based on findings derived from multimodal imaging. We conducted a review of the current literature referenced in PubMed and Web of Science among others with the keywords ‘artificial intelligence’ and ‘machine learning’ in combination with ‘perimetry’, ‘best-corrected visual acuity (BCVA)’, ‘retinal function’ and ‘age-related macular degeneration’. So far AI-based structure-function correlations have been applied to infer conventional visual field, fundus-controlled perimetry, and electroretinography data, as well as BCVA, and patient-reported outcome measures (PROM). In neovascular AMD, inference of BCVA (hereafter termed inferred BCVA) can estimate BCVA results with a root mean squared error of ~7–11 letters, which is comparable to the accuracy of actual visual acuity assessment. Further, AI-based structure-function correlation can successfully infer fundus-controlled perimetry (FCP) results both for mesopic as well as dark-adapted (DA) cyan and red testing (hereafter termed inferred sensitivity). Accuracy of inferred sensitivity can be augmented by adding short FCP examinations and reach mean absolute errors (MAE) of ~3–5 dB for mesopic, DA cyan and DA red testing. Inferred BCVA, and inferred retinal sensitivity, based on multimodal imaging, may be considered as a quasi-functional surrogate endpoint for future interventional clinical trials in the future.

## Introduction

Age-related macular degeneration (AMD) is the leading cause of visual disability among the elderly in industrialised countries [1]. While anti-vascular endothelial growth factor (VEGF) therapy has markedly improved outcomes for macular neovascularisation secondary to AMD, disease-

specific therapy for early and non-exudative manifestations, including geographic atrophy (GA), is lacking [2, 3]. Beyond analysis of structural changes, there is an unmet need to establish meaningful functional endpoints for assessment of visual impairment associated with AMD manifestations. Best-corrected visual acuity (BCVA), the most common used functional endpoint, only measures photopic function of the central retina and is therefore not sensitive to measure therapeutic benefits outside of the fovea [4, 5]. Other measures, including low-luminance visual acuity (LLVA), reading speed, fundus-controlled microperimetry (FCP), and patient-reported outcome measures (PROM), potentially assess impairment of visual function in more detail. However, most of these tests are time consuming and can be particularly demanding for the elderly patient [6–9]. Therefore, alternative analysis strategies would be desirable that are more practicable and easier to obtain, while addressing different facets of visual function.

Artificial intelligence (AI) is creating a paradigm shift in every sector of medicine. Ophthalmology is at the forefront of implementing AI-enabled health care, with a few

✉ Steffen Schmitz-Valckenberg  
steffen.valckenberg@utah.edu

<sup>1</sup> Department of Ophthalmology, University of Bonn, Bonn, Germany

<sup>2</sup> Department of Biomedical Data Science, Radiology, and Medicine, Stanford University, Stanford, CA, USA

<sup>3</sup> John A. Moran Eye Center, University of Utah, Salt Lake City, UT, USA

<sup>4</sup> Augenklinik, Universität Ulm, Ulm, Deutschland

<sup>5</sup> Augenarztpraxis Dres. Kortüm, Ludwigsburg, Deutschland

<sup>6</sup> Moorfields Eye Hospital NHS Foundation Trust, London, UK

commercially available AI tools already available for clinical care. First applications have already crossed the threshold into clinical care [10]. These tools can now assist clinicians in diagnosis of fundus photographs as well as achieve automated annotations of optical coherence tomography (OCT) imaging [11–14]. In AMD, AI has been deployed to estimate the number of anti-VEGF injections needed and to predict GA progression [15, 16]. There is a wide range of potential AI-algorithms used for predictive modelling. In recent years, convolutional neural networks have proven particularly successful for image classification tasks by extracting features from raw data through hierarchies of increasing abstraction which superficially represent visual processing in the brain [17, 18]. These deep-learning (DL) algorithms require large image data sets and results can be hard to interpret. Machine-learning (ML) algorithms using pre-defined, “hand-crafted” (e.g. Random Forest or Lasso regression) can therefore be a viable alternative. A recent development is AI-based predictions of retinal function with the use of multimodal imaging modalities. So far, AI-based structure-function correlations have been applied for the inferring of conventional visual field and FCP data, BCVA, vision-related quality-of-life, and electroretinogram (ERG) characteristics.

This review summarises the progress in this field, compares goodness of fit measures and AI algorithms utilised while focusing on possible implementations in AMD patients.

### AI based structure-function correlation in visual fields

Although the focus of this review is on AMD relevant diagnostic tools, it is important to note that algorithms for inference of function have a long-standing history in glaucoma detection. First automated programs to diagnose visual field deficits date back to the 1980s and some studies now deploy OCT imaging to guide decision making [19, 20]. Newer studies, like Christopher et al. deploy large data sets of almost 10,000 visual field/OCT pairs from over a 1000 participants to train DL algorithms and are able to estimate the mean deviation of the visual field with an accuracy of 2.5 dB ( $R^2$  0.7) [21]. Inference of sectoral visual field loss varied between high accuracy in the inferior-nasal sector ( $R^2$  0.6) and low accuracy in both the central ( $R^2$  0.15) and temporal ( $R^2$  0.12) sectors. This algorithm was trained with retinal nerve fibre layer (RNFL) thickness maps and the results show that accuracy was lowest in areas with a physiological thin RNFL layer (temporal sector) and highest in areas with physiological thick RNFL layer (inferior-nasal sector) [22]. This is in accordance with the notion that pathological RNFL layer thinning is harder to discriminate in areas of decreased physiological thickness.

Overall, progress on visual field Inferences will likely provide deeper insights into structural-functional correlations when large data sets may become available.

### AI based structure-function correlation in electroretinogram

In ABCA4-related retinopathy, a recent study predicted ERG Results through OCT layer thickness with an accuracy of  $97.47 \pm 2.03\%$  [23]. The most relevant OCT-based imaging features in the applied machine learning approach were the outer nuclear layer (ONL) and the inner- and outer segments (IS/OS). The high impact of the ONL and IS/OS layers is biologically plausible as they represent parts of rod and cone photoreceptors. Notably, prediction accuracies may not necessarily translate to AMD as the ABCA4-related retinopathy cohort is a hereditary, monogenetic disease with rather well-defined ERG changes, while ERG findings in AMD are known to be much less specific. However, AI-based structure-function correlations in AMD subjects also demonstrated biological plausibility, showing that ONL thickness changes had the highest predictor importance for functional deficits that can be detected by other functional tests, including FCP (FCP; see below) [24–27].

### AI-based structure-function correlation in BCVA

In the context of AMD, inference of BCVA from OCT images has been proposed predominantly for macular neovascularisation (Table 1) [23, 27–31]. During anti-VEGF treatment, retinal imaging plays a pivotal role in disease management, with patients being regularly monitored by OCT. Beyond qualitative interpretation by the human eye, the automated analysis of these extensive amounts of imaging data may be particularly useful for estimation of BCVA, in addition to routine BCVA measurements, potentially sparing time and allowing for a more consistent assessment.

Until today, goodness of fit measures varied between studies for inference of BCVA based on imaging using AI tools. Two studies reported similar accuracies with a mean absolute error (MAE) of 0.11–0.14 LogMAR and 0.14 LogMAR. Another study documented their accuracy with a Root Mean Squared Error (RMSE) of 8.21 letters. The MAE, as the mean of the absolute values of the individual prediction errors, is an easy interpretable evaluation metric to judge the accuracy of regression models. For an exemplary patient with an BCVA of 0.3 LogMAR (Snellen 6/12), the inferred BCVA would be on average between 0.19 and 0.41 logMAR (~6/9–6/15). In some cases, when

**Table 1** AI-based structure-function in best-corrected visual acuity (BCVA).

Author/Ref.	Title	Disease	Technique	Prediction	Outcome measure	Outcome
Rohm et al. [27]	Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration	Neovascular age-related macular degeneration	<ul style="list-style-type: none"> <li>Five different machine-learning algorithms</li> <li>Best performance by Lasso regression</li> </ul>	<ul style="list-style-type: none"> <li>logMAR visual acuity after 3- and 12 months</li> </ul>	<ul style="list-style-type: none"> <li>Mean Absolute Error (MAE)</li> <li>Root Mean Squared Error (RMSE)</li> </ul>	<ul style="list-style-type: none"> <li>3 Months = MAE: 0.11–0.14/RMSE: 0.18–0.2</li> <li>12 Months = MAE: 0.16–0.2/RMSE: 0.2–0.22</li> </ul>
Schmidt-Erfurth et al. [28]	Machine learning to analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration	Neovascular age-related macular degeneration	<ul style="list-style-type: none"> <li>Random forest</li> </ul>	<ul style="list-style-type: none"> <li>BCVA at Baseline and 3 months follow-up</li> </ul>	<ul style="list-style-type: none"> <li>Accuracy (<math>R^2</math>)</li> </ul>	<ul style="list-style-type: none"> <li><math>R^2 = 0.21</math> baseline</li> <li><math>R^2 = 0.70</math> 3 months</li> </ul>
Gerenda et al. [29]	Computational image analysis for prognosis determination in DME	Diabetic macular edema	<ul style="list-style-type: none"> <li>Random forest</li> </ul>	<ul style="list-style-type: none"> <li>BCVA at Baseline and 1-year follow-up</li> </ul>	<ul style="list-style-type: none"> <li>Accuracy (<math>R^2</math>)</li> </ul>	<ul style="list-style-type: none"> <li><math>R^2 = 0.21</math> baseline</li> <li><math>R^2 = 0.23</math> 1 year</li> </ul>
Aslam et al. [30]	Use of a neural net to model the impact of optical coherence tomography abnormalities on vision in age-related macular degeneration	Neovascular age-related macular degeneration	<ul style="list-style-type: none"> <li>Scaled conjugate gradient backpropagation (supervised learning)</li> </ul>	<ul style="list-style-type: none"> <li>BCVA</li> </ul>	<ul style="list-style-type: none"> <li>Root Mean Squared Error (RMSE)</li> </ul>	<ul style="list-style-type: none"> <li>8.21 Letters</li> </ul>
Pfau et al. [31]	Artificial intelligence in ophthalmology: guideline for physicians for the critical evaluation of studies	Neovascular age-related macular degeneration	<ul style="list-style-type: none"> <li>Nested cross validation</li> </ul>	<ul style="list-style-type: none"> <li>BCVA (LogMAR)</li> </ul>	<ul style="list-style-type: none"> <li>MAE</li> </ul>	<ul style="list-style-type: none"> <li>0.142</li> </ul>
Müller et al. [23]	Prediction of function in ABCA4-related retinopathy using ensemble machine learning	ABCA4-related Retinopathy	<ul style="list-style-type: none"> <li>Ensemble machine learning algorithms</li> <li>Three models               <ol style="list-style-type: none"> <li>Retinal layer</li> <li>All structural data</li> <li>demographic data</li> </ol> </li> </ul>	<ul style="list-style-type: none"> <li>BCVA</li> <li>Divided into four categories from no impairment to severe impairment</li> </ul>	<ul style="list-style-type: none"> <li>Area under the curve (ROC)</li> </ul>	<ul style="list-style-type: none"> <li>(a) 88.64–92.25%</li> <li>(b) 90.23–93.68%</li> <li>(c) 87.26–91.44%</li> </ul>
Sumaroka et al. [45]	Foveal therapy in blue cone monochromacy: predictions of visual potential from artificial intelligence	Blue Cone Monochromacy	<ul style="list-style-type: none"> <li>Random forest</li> <li>Layer thickness</li> <li>Reflectivity</li> </ul>	<ul style="list-style-type: none"> <li>BCVA</li> </ul>	<ul style="list-style-type: none"> <li>Root mean squared error (RMSE)</li> </ul>	<ul style="list-style-type: none"> <li>(a) 0.159</li> <li>(b) 0.167</li> </ul>

not only the average error but also the outliers are of interest, it may be helpful to indicate other evaluation metrics like the RMSE. The RMSE indicates the size of the squared error. As a result, larger errors have a disproportionately larger effect on the RMSE.

Rather than inferring function in a cross-sectional manner, another interesting aspect is to predict BCVA in the future, based on data available at baseline. This prediction might be particularly helpful to better estimate possible treatment effects. As one may expect, studies have reported lower accuracy in predicting BCVA as compared to the inference of BCVA based on imaging data from the same visit. By comparing five different ML algorithms, Rohm et al. [27] reported that cross-validated LASSO regression achieved most precise results with 0.14 logMAR RMSE (equals 7 letters) for the short-term (90 days from baseline) and 0.23 logMAR RMSE (equals 11.5 letters) for long-term predictions. This means that if an exemplary patient would have a measured BCVA of 0.3 logMAR (Snellen 6/12) at the 3 months follow-up visit, the model based on multimodal imaging at baseline would have predicted the BCVA being on average between 0.16 and 0.44 logMAR (~6/9–6/15). This study further demonstrated that not always the most complex AI algorithm achieved most accurate predictions. LASSO regression, the most accurate algorithm in this study, builds on linear regression and is therefore easily computable and easy for the non-AI specialist to interpret.

These algorithms could serve to inform the patient over their individual disease progression and give a prognosis of their driving capability. It should be considered that the 95% limits of agreement (LoA) of repeatability for tested BCVA is already about  $\pm 0.1$  logMAR under perfect conditions for healthy subjects [32, 33]. Even with improved accuracy of the algorithm, it is important to note that inferred BCVA faces the same challenges and limitations as measured BCVA. These are specifically due to the ceiling effect (given the limited by the retinal peak cone density) and the relative focus on foveal (and/or para-foveal) function.

### AI-based structure-function correlation in patient-reported outcome measures

The European Medicines Agency and the Food and Drug Administration increasingly demand the employment of PROM as functional endpoints in clinical trials. During the Phase 2 Mahalo Study for GA, the 25-Item National Eye Institute Visual Function Questionnaire (NEI VFQ-25) has demonstrated to be a valid and reliable measure of patients Vision-Related Quality-of-Life (VRQoL) [34]. A recent study applied predictive modelling (LASSO regression) on both functional and structural biomarkers to project NEI VFQ-25 VRQoL for GA patients [35]. Interestingly, they

found that VRQoL predominantly depended on the better eye. Structural biomarkers only explained up to 22% of variability but in combination with functional parameters like LLVA achieved excellent results in predicting VRQoL. These observations may be used for modelling of function in the clinical trial setting. In early phase clinical trials, a common approach is to rather test a new intervention in the worse as compared to the better eye because of ethical considerations, particularly the uncertainty of the risk profile of a new intervention. In this context, the authors suggested to extract the information from treating the worse eye to infer the expected effect on VRQoL in the better seeing eye [35].

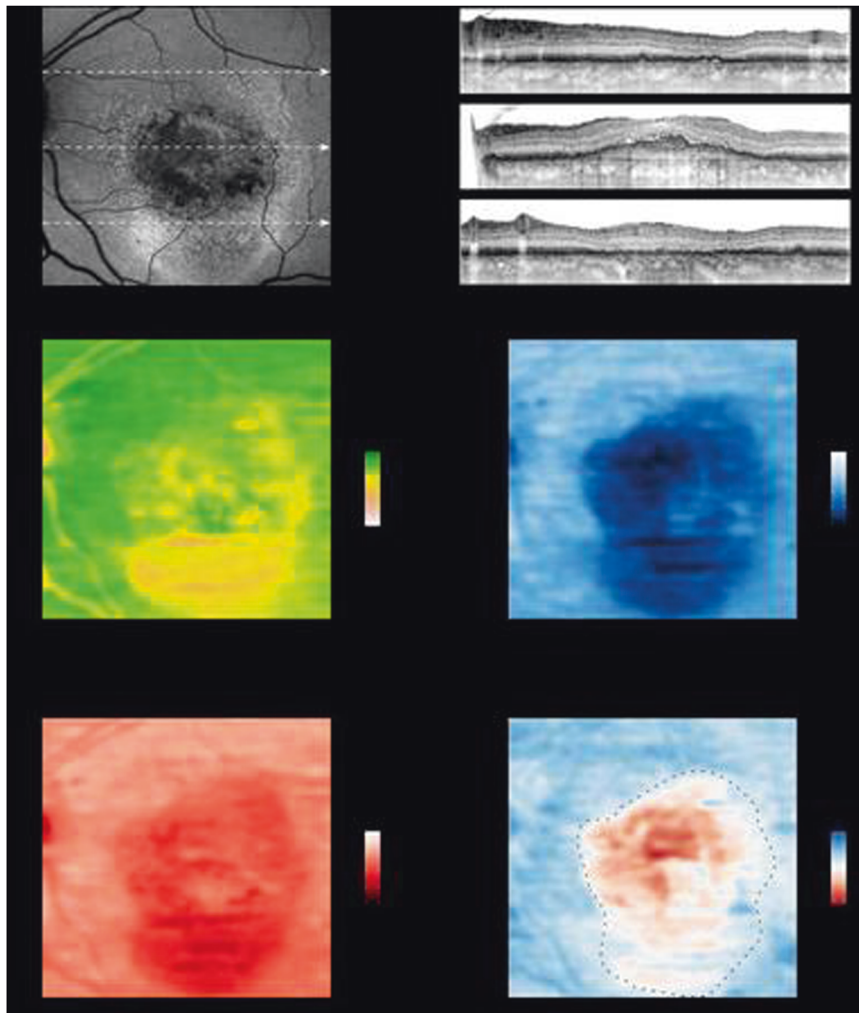
### AI-based structure-function correlation in fundus-controlled perimetry

Beyond BCVA which is limited to the assessment of foveal function, FCP (also termed microperimetry) can detect impairment throughout the central retina while correcting for fixation losses [36]. At the same time, FCP is time-consuming and burdensome for both patients and medical health care professionals. A recent development of FCP is the ability to test two-colour dark-adapted (DA) sensitivity. This setup allows to confirm previous reports, based on histopathology and general clinical assessment, that rod loss exceeds and precedes cone loss in eyes with AMD. FCP may allow for more precise assessment in the clinical setting as compared to more general clinical tests and therefore excel in detecting earliest changes in AMD. Using AI, early data indeed show that the precise probing of rod and cone function opens the door for a more detailed structure-function correlations between microstructural changes seen by multimodal imaging and retinal sensitivity as measured by FCP [37–40]. In this context, one interesting observation is that specific lesion components exhibit a distinct effect on cone (e.g. cuticular drusen with central, pseudo-vitelliform detachment), while other do on rod (e.g. reticular drusen; macular oedema) dysfunction [41, 42]. In neovascular late-stage AMD, FCP revealed a characteristic functional response to pigment epithelial detachments, subretinal- and intraretinal fluid [43]. For example, subretinal fluid appeared to affect rod function to a greater degree than cone function. Intraretinal fluid seemed to impair both cone and rod function to a similar extent. Participants with a shallow pigment epithelial detachment exhibited relative preserved cone and rod function. Given that structure and function are so closely intertwined, AI-based inference of retinal sensitivity through multimodal imaging (herein termed inferred sensitivity) would be tremendously advantageous beyond just replacing a burdensome psychophysical examination (Table 2). Inferred sensitivity has the potential

**Table 2** AI-based structure-function in fundus-controlled perimetry (FCP).

Author/Ref.	Title	Disease	Technique	Prediction	Outcome measure	Outcome
Kihara et al. [44]	Estimating retinal sensitivity using optical coherence tomography with deep-learning algorithms in macular telangiectasia type 2	Macular telangiectasia type 2	1. Million-variable deep-learning model (convolutional neural network) 2. Letnet modell 3. Linear regression	• Fundus-controlled perimetry (FCP) (a) Mesopic	• Mean absolute error (MAE)	1. 3.36 dB 2. 3.66 dB 3. 4.51 dB
von der Emde and Pfau et al. [25]	Artificial intelligence for morphology-based function prediction in neovascular age-related macular degeneration	Neovascular age-related macular degeneration	• Random forest (LOO-CV) • Two scenarios 1. Without FCP 2. With FCP data	• Fundus-controlled perimetry (FCP): (a) Mesopic (b) DA Cyan (c) DA Red	• Mean absolute error (MAE) • Root mean squared error (RMSE)	• Scenario 1 • 3.94 dB (a) 4.89 dB (b) 4.05 dB (c) • Scenario 2 • 2.8 dB (a) 3.7 dB (b) 2.85 dB (c)
Pfau et al. [24]	Determinants of cone- and rod function in geographic atrophy: AI-based structure-function correlation	geographic atrophy secondary to age-related macular degeneration	• Random forest (LOO-CV) • Two scenarios 1. Without FCP 2. With FCP data	• Fundus-controlled perimetry (FCP): (a) mesopic (b) DA Cyan (c) DA Red	• Mean absolute error (MAE) • Root mean squared error (RMSE)	• Scenario 1 • 4.64 dB (a) 4.89 dB (b) 4.4 dB (c) • Scenario 2 • 2.89 dB (a) 2.86 dB (b) 2.77 dB (c)
Sumaroka et al. [46]	Treatment potential for macular cone vision in leber congenital amaurosis due to CEP290 or NPHP5 Mutations: predictions from artificial intelligence	Retinitis pigmentosa (RP); Congenital Amaurosis (LCA; prediction)	• Random forest (LOO-CV) • Two models 1. Thickness and eccentricity 2. Reflectivity	• Dark-adapted static perimetry (a) DA cyan (b) DA red	• 95th Percentile limits of agreement (LOA)	RP 1. 9.6 dB (a) 8.8 dB (b) 2. 11.9 dB (a) 10.8 dB (b) LCA 1. 4.6–17.6 dB
Sumaroka et al. [45]	Foveal therapy in blue cone monochromacy: predictions of visual potential from artificial intelligence	Blue Cone Monochromacy	• Random forest (a) Layer thickness (b) reflectivity	• Fundus-controlled Perimetry (a) Mesopic	• Root mean squared error (RMSE)	(a) 2.91 dB (b) 2.69 dB
Heß et al. [47]	Mesopic and scotopic light sensitivity and its microstructural correlates in pseudoxanthoma elasticum	Pseudoxanthoma Elasticum	• Random forest (LOO-CV)	• Fundus-controlled perimetry: (a) Mesopic (b) DA Cyan (c) DA Red	• Mean absolute error (MAE)	(a) 4.91 dB (b) 5.44 dB (c) 4.99 dB





**Fig. 1 Inferred sensitivity mapping.** Based on the fundus autofluorescence (FAF), infrared reflection (IR, not shown) and spectral-domain optical coherence tomography (SD-OCT), mesopic as well as dark-adapted (DA) cyan and DA red sensitivity may be reliably inferred and topographically mapped. The arrows in the FAF image indicate the position of the SD-OCT B-scans. Multiple lines of evidence further support the accuracy of the inference. For all three types of testing, angioscotoma are adequately predicted. Further, the central rod-free zone is also correctly inferred as indicated by the marked cyan-red sensitivity difference at the fovea (eccentricity of  $0^\circ$ , middle B-scan). Regions exhibiting increased FAF and absence of

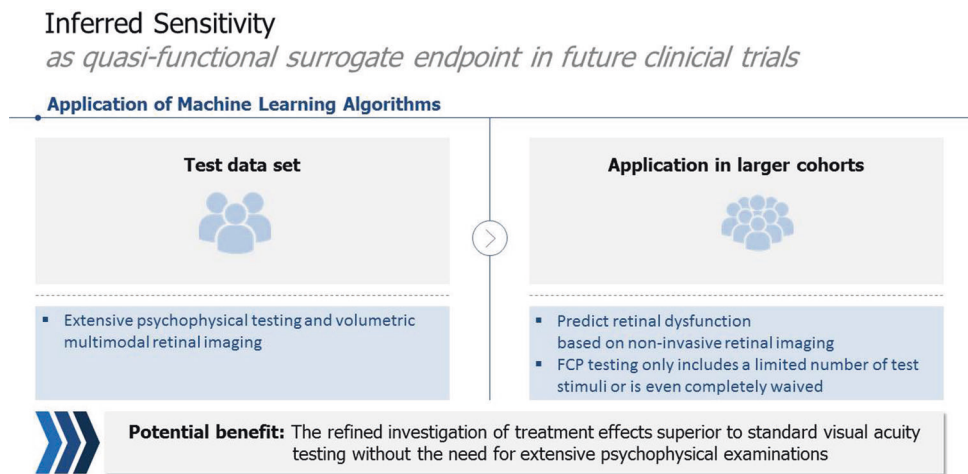
photoreceptor outer and inner-segments (upper and lower SD-OCT scan) show reduced function for all three types of testing. Yet globally the degree of DA cyan dysfunction appears to exceed the degree of DA red dysfunction. Please note, that the inferred cyan-red sensitivity difference in the region of severe cone dysfunction (delimited by the dashed line) is an underestimation of the true cyan-red sensitivity difference due to the floor effects of the perimetry device used in this study that are inevitable reflected by the models. (Reprinted from von der Emde et. al: Artificial intelligence for morphology-based function prediction in neovascular age-related macular degeneration; Scientific reports 9:1132; published [2019] Springer Nature).

to probe larger areas of the retina while still providing a high spatial resolution (Fig. 1). Additionally, with inferred sensitivity patients who are too frail for FCP examinations or have difficulty fixating could now partake in interventional studies. Finally, image acquisition for inferred sensitivity can be performed in centres without the expertise for psychophysical testing. So far inferred sensitivity has been explored in macular telangiectasia type 2, Leber congenital amaurosis (LCA), blue cone monochromasy, pseudoxanthoma elasticum as well as MNV and GA secondary to AMD, demonstrating that

mesopic sensitivity can be inferred with an accuracy of 3.36–4.64 decibel (dB) MAE cross those diseases [24, 25, 44–47].

DA cyan estimates achieved an accuracy of 4.89 dB MAE in late-stage AMD and 8.8 dB 95th percentile LoA in LCA. Similarly, accuracy in DA red testing was 4.05–4.64 dB MAE in late-stage AMD and 9.6 dB LoA in LCA. The results are further underscored by the fact that the 95% coefficient of repeatability for FCP testing was reported with (mean  $\pm$  SD)  $5.99 \pm 1.55$  dB for mesopic,  $6.14 \pm 2.19$  dB for dark-adapted cyan and  $6.06 \pm 1.79$  dB for dark-

**Fig. 2 Inferred sensitivity.** Possible quasi-functional surrogate endpoint in future clinical trial.



adapted red testing in MNV secondary to AMD (with similar results for GA) [6, 7]. Although evaluation metrics are not identical, this shows that the average error of inference based on multimodal imaging differs only marginally from the error resulting from repetitive testing (test-retest reliability) alone.

Studies using ML algorithms further analysed feature importance of inferred sensitivity and demonstrated that layer thicknesses proved more indicative for predictions than layer intensities [24, 25, 45–47]. Hereby, the results of the analysis of retinal thickness seemed biological plausible, as ONL thickness proved to be the most important imaging feature in all types of testing and cross disease stage. Specifically, a thinning of the ONL (indicative of outer retinal atrophy) led to a decrease in sensitivity whereas pathological ONL thickening only negatively affected dark-adapted cyan and red sensitivity [24, 25].

In our studies in late-stage AMD patients, we further tried to boost results by feeding the algorithm the results from a short perimetry testing, since we hypothesised that factors not resolved by OCT imaging (e.g. lenticular opacification) may influence retinal light sensitivity [24, 25]. Indeed, addition of some functional data (in form of a subset of the perimetry results) markedly reduced the MAE up to 1.54 dB in mesopic and DA cyan and red testing [24, 25]. Although not as pronounced an alternative approach using ‘Patient reliability indices’ (e.g. False-positive response rate during FCP testing) also markedly reduced the MAE. These approaches in the context of an interventional study could improve accuracies of inferred sensitivity. Moreover, with additional short perimetry testing potential adverse effects of the agents (e.g. RNFL layer thinning) that are not represented in the training set could still be detected. In summary, we consider inferred sensitivity to be a potential surrogate functional endpoint and a valuable tool for future interventional studies.

## Limitations

Comparison of results among available studies is hampered by non-uniform use of goodness of fit measures. For example, the MAE is easily interpretable but may simulate better results than RMSE as it scrutinises outliers more harshly. Therefore, we suggest that studies reach consensus on measures of goodness of fit or to report additional measures for comparison. Another limitation is the cross-sectional study design of most AI-based function prediction studies so far. To safely utilise AI-based predictive modelling in future interventional studies, the accuracy of longitudinal models needs to be verified. Additionally, most studies were performed with a limited number of participants. This could potentially lead to underrepresenting less frequent manifestations in the training set. In AMD for example, these could be subretinal drusenoid deposits in intermediate AMD, retinal angiomatous proliferation in MNV secondary to AMD or the diffuse-trickling phenotype in GA secondary to AMD.

## Outlook

This review evaluated the current literature on AI-based function inference on a plethora of different psychophysical examinations. In retinal diseases other than AMD, AI-based function inferences serve to accurately forecast visual fields, ERG, BCVA and FCP examinations. In AMD, AI-based function inferences can compute BCVA, PROM and FCP results. We established the term ‘inferred sensitivity’ for multimodal imaging-based estimation of FCP results. Accuracy of inferred sensitivity can be improved by adding short FCP examinations in a subset of patients (Fig. 2). Inferred sensitivity of two-colour DA FCP can also estimate rod function and detect earliest visual impairment in AMD patients. Therefore, we consider

inferred sensitivity to be a quasi-functional surrogate endpoint.

**Author contributions** LVE, MP and SSV were responsible for conception and design of the work. All authors contributed to data collection, interpretation and drafting of the article, the critical revision of the article and finally approved the version to be published.

**Funding** This work was supported by the German Research Foundation (DFG, grant PF950/1-1 to MP), PRO RETINA research stipend to LvdE and in part by an Unrestricted Grant from Research to Prevent Blindness, New York, NY, to the Department of Ophthalmology and Visual Sciences, University of Utah. Open Access funding enabled and organized by Projekt DEAL.

## Compliance with ethical standards

**Conflict of interest** The Department of Ophthalmology, University of Bonn, received nonfinancial support for research material from CenterVue SpA, Padova, Italy, Heidelberg Engineering, Heidelberg, Germany, and Carl Zeiss Meditec, Jena, Germany.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular degeneration. *Lancet*. 2012;379:1728–38.
- Miller JW. Beyond VEGF—the Weisenfeld lecture. *Investig Ophthalmol Vis Sci*. 2016;57:6911–8.
- Holz FG, Sadda SR, Busbee B, Chew EY, Mitchell P, Tufail A, et al. Efficacy and safety of lomalizumab for geographic atrophy due to age-related macular degeneration: chroma and spectri phase 3 randomized clinical trials. *JAMA Ophthalmol*. 2018;136:666–77.
- Amoaku WM, Chakravarthy U, Gale R, Gavin M, Ghanchi F, Gibson J, et al. Defining response to anti-VEGF therapies in neovascular AMD. *Eye*. 2015;29:721–31.
- Lindner M, Nadal J, Mauschitz MM, Luning A, Czauderna J, Pfau M, et al. Combined fundus autofluorescence and near infrared reflectance as prognostic biomarkers for visual acuity in foveal-sparing geographic atrophy. *Investig Ophthalmol Vis Sci*. 2017;58:BIO61–7.
- von der Emde L, Pfau M, Thiele S, Moller PT, Hassenrik R, Fleckenstein M, et al. 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.
- Pfau M, Muller PL, von der Emde L, Lindner M, Moller PT, Fleckenstein M, et al. Mesopic and dark-adapted two-color fundus-controlled perimetry in geographic atrophy secondary to age-related macular degeneration. *Retina*. 2018;40:169–80.
- McGuinness MB, Finger RP, Wu Z, Luu CD, Chen FK, Arnold JJ, et al. Association between patient-reported outcomes and time to late age-related macular degeneration in the laser intervention in early stages of age-related macular degeneration study. *Ophthalmol Retina*. 2020;4:881–8.
- Lindner M, Pfau M, Czauderna J, Goerdts L, Schmitz-Valckenberg S, Holz FG, et al. Determinants of reading performance in eyes with foveal-sparing geographic atrophy. *Ophthalmol Retina*. 2019;3:201–10.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
- Ayhan MS, Kühlewein L, Aliyeva G, Inhoffen W, Ziemssen F, Berens P. Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection. *Med Image Anal*. 2020;64:101724.
- De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–50.
- Maloca PM, Lee AY, de Carvalho ER, Okada M, Fasler K, Leung I, et al. Validation of automated artificial intelligence segmentation of optical coherence tomography images. *PLoS ONE*. 2019;14:e0220063.
- Waldstein SM, Vogl W-D, Bogunovic H, Sadeghipour A, Riedl S, Schmidt-Erfurth U. Characterization of drusen and hyperreflective foci as biomarkers for disease progression in age-related macular degeneration using artificial intelligence in optical coherence tomography. *JAMA Ophthalmol*. 2020;138:740–7.
- Bogunovic H, Waldstein SM, Schlegl T, Langs G, Sadeghipour A, Liu X, et al. Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach. *Investig Ophthalmol Vis Sci*. 2017;58:3240–8.
- Pfau M, Möller PT, Künzel SH, von der Emde L, Lindner M, Thiele S, et al. Type 1 choroidal neovascularization is associated with reduced localized progression of atrophy in age-related macular degeneration. *Ophthalmol Retina*. 2020;4:238–48.
- Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med*. 2018;1:40.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Drasdo N. Patterns and contrasts in ophthalmic investigation. *Ophthalmic Physiol Opt*. 1988;8:3–13.
- Silva FR, Vidotti VG, Cremasco F, Dias M, Gomi ES, Costa VP. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry. *Arq Bras Oftalmol*. 2013;76:170–4.
- Christopher M, Bowd C, Belghith A, Goldbaum MH, Weinreb RN, Fazio MA, et al. Deep learning approaches predict glaucomatous visual field damage from OCT optic nerve head en face images and retinal nerve fiber layer thickness maps. *Ophthalmology*. 2020;127:346–56.
- Jonas JB, Gusek GC, Naumann GO. Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes. *Investig Ophthalmol Vis Sci*. 1988;29:1151–8.
- Müller PL, Treis T, Odainic A, Pfau M, Herrmann P, Tufail A, et al. Prediction of function in ABCA4-related retinopathy using ensemble machine learning. *J Clin Med*. 2020;9:2428.
- Pfau M, von der Emde L, Dysli C, Möller PT, Thiele S, Lindner M, et al. Determinants of cone- and rod-function in geographic



- atrophy: AI-based structure-function correlation. *Am J Ophthalmol.* 2020;217:162–73.
25. von der Emde L, Pfau M, Dysli C, Thiele S, Moller PT, Lindner M, et al. Artificial intelligence for morphology-based function prediction in neovascular age-related macular degeneration. *Sci Rep.* 2019;9:11132.
  26. Saßmannshausen M, Steinberg JS, Fimmers R, Pfau M, Thiele S, Fleckenstein M, et al. Structure-function analysis in patients with intermediate age-related macular degeneration. *Investig Ophthalmol Vis Sci.* 2018;59:1599–608.
  27. Rohm M, Tresp V, Muller M, Kern C, Manakov I, Weiss M, et al. Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration. *Ophthalmology.* 2018;125:1028–36.
  28. Schmidt-Erfurth U, Bogunovic H, Sadeghipour A, Schlegl T, Langs G, Gerendas BS, et al. Machine learning to analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration. *Ophthalmol Retina.* 2018;2:24–30.
  29. Gerendas BS, Bogunovic H, Sadeghipour A, Schlegl T, Langs G, Waldstein SM, et al. Computational image analysis for prognosis determination in DME. *Vision Res.* 2017;139:204–10.
  30. Aslam TM, Zaki HR, Mahmood S, Ali ZC, Ahmad NA, Thorell MR, et al. Use of a neural net to model the impact of optical coherence tomography abnormalities on vision in age-related macular degeneration. *Am J Ophthalmol.* 2018;185:94–100.
  31. Pfau M, Walther G, von der Emde L, Berens P, Faes L, Fleckenstein M, et al. Artificial intelligence in ophthalmology: Guidelines for physicians for the critical evaluation of studies. *Ophthalmologie.* 2020;117:973–88.
  32. Arditì A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Investig Ophthalmol Vis Sci.* 1993;34:120–9.
  33. Vanden Bosch ME, Wall M. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye.* 1997;11:411–7.
  34. Sivaprasad S, Tschosik E, Kapre A, Varma R, Bressler NM, Kimel M, et al. Reliability and construct validity of the NEI VFQ-25 in a subset of patients with geographic atrophy from the phase 2 Mahalo study. *Am J Ophthalmol.* 2018;190:1–8.
  35. Künzel SH, Möller PT, Lindner M, Goerdt L, Nadal J, Schmid M, et al. Determinants of quality of life in geographic atrophy secondary to age-related macular degeneration. *Investig Ophthalmol Vis Sci.* 2020;61:63.
  36. Pfau M, Jolly JK, Wu Z, Denniss J, Lad EM, Guymer RH, et al. Fundus-controlled perimetry (microperimetry): Application as outcome measure in clinical trials. *Prog Retin Eye Res.* 2020;100907. <https://www.sciencedirect.com/science/article/abs/pii/S1350946220300793?via%3Dihub>.
  37. Wu Z, Cunefare D, Chiu E, Luu CD, Ayton LN, Toth CA, et al. Longitudinal associations between microstructural changes and microperimetry in the early stages of age-related macular degeneration. *Investig Ophthalmol Vis Sci.* 2016;57:3714.
  38. Sassmannshausen M, Pfau M, Thiele S, Fimmers R, Steinberg JS, Fleckenstein M, et al. Longitudinal analysis of structural and functional changes in presence of reticular pseudodrusen associated with age-related macular degeneration. *Investig Ophthalmol Vis Sci.* 2020;61:19.
  39. Ponderfer SG, Wintergerst MWM, Gorgi Zadeh S, Schultz T, Heinemann M, Holz FG, et al. Association of visual function measures with drusen volume in early stages of age-related macular degeneration. *Investig Ophthalmol Vis Sci.* 2020;61:55.
  40. Midená E, Vujosevic S. Diagnosing and monitoring diabetic macular edema: structural and functional tests. *Int Ophthalmol.* 2015;35:623–8.
  41. Pfau M, Lindner M, Gliem M, Steinberg JS, Thiele S, Finger RP, et al. Mesopic and dark-adapted two-color fundus-controlled perimetry in patients with cuticular, reticular and soft drusen. *Eye.* 2018;32:1819–30.
  42. Steinberg JS, Fitzke FW, Fimmers R, Fleckenstein M, Holz FG, Schmitz-Valckenberg S. Scotopic and photopic microperimetry in patients with reticular drusen and age-related macular degeneration. *JAMA Ophthalmol.* 2015;133:690–7.
  43. Sulzbacher F, Roberts P, Munk MR, Kaider A, Kroh ME, Sacu S, et al. Relationship of retinal morphology and retinal sensitivity in the treatment of neovascular age-related macular degeneration using aflibercept. *Investig Ophthalmol Vis Sci.* 2014;56:1158–67.
  44. Kihara Y, Heeren TFC, Lee CS, Wu Y, Xiao S, Tzaridis S, et al. Estimating retinal sensitivity using optical coherence tomography with deep-learning algorithms in macular telangiectasia type 2. *JAMA Netw Open.* 2019;2:e188029.
  45. Sumaroka A, Cideciyan AV, Sheplock R, Wu V, Kohl S, Wisinger B, et al. Foveal therapy in blue cone monochromacy: predictions of visual potential from artificial intelligence. *Front Neurosci.* 2020;14:800.
  46. Sumaroka A, Garafalo AV, Semenov EP, Sheplock R, Krishnan AK, Roman AJ, et al. Treatment potential for macular cone vision in leber congenital amaurosis due to CEP290 or NPHP5 mutations: predictions from artificial intelligence. *Investig Ophthalmol Vis Sci.* 2019;60:2551–62.
  47. Hess K, Gliem M, Charbel Issa P, Birtel J, Müller PL, von der Emde L, et al. Mesopic and scotopic light sensitivity and its microstructural correlates in pseudoxanthoma elasticum. *JAMA Ophthalmol.* 2020;138:1272–9.