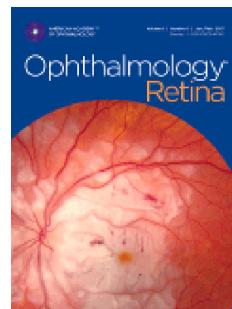


Journal Pre-proof

Linking Function and Structure with ReSenseNet: Predicting Retinal Sensitivity from Optical Coherence Tomography using Deep Learning

Philipp Seeböck, PhD, Wolf-Dieter Vogl, PhD, Sebastian M. Waldstein, MD, PhD, Jose Ignacio Orlando, PhD, Magdalena Baratsits, Thomas Alten, Mustafa Arikan, MSc, Georgios Mylonas, MD, PhD, Hrvoje Bogunović, PhD, Ursula Schmidt-Erfurth, MD



PII: S2468-6530(22)00043-4

DOI: <https://doi.org/10.1016/j.oret.2022.01.021>

Reference: ORET 1164

To appear in: *Ophthalmology Retina*

Received Date: 21 October 2021

Revised Date: 26 January 2022

Accepted Date: 31 January 2022

Please cite this article as: Seeböck P., Vogl W.-D., Waldstein S.M., Orlando J.I., Baratsits M., Alten T., Arikan M., Mylonas G., Bogunović H. & Schmidt-Erfurth U., Linking Function and Structure with ReSenseNet: Predicting Retinal Sensitivity from Optical Coherence Tomography using Deep Learning, *Ophthalmology Retina* (2022), doi: <https://doi.org/10.1016/j.oret.2022.01.021>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Inc. on behalf of American Academy of Ophthalmology

1 **Linking Function and Structure with ReSenseNet:**
 2 **Predicting Retinal Sensitivity from Optical Coherence**
 3 **Tomography using Deep Learning**

4 Philipp Seeböck, PhD^{1,2}; Wolf-Dieter Vogl, PhD¹; Sebastian M. Waldstein,
 5 MD, PhD¹; Jose Ignacio Orlando, PhD¹; Magdalena Baratsits³; Thomas Alten¹;
 6 Mustafa Arikan, MSc¹; Georgios Mylonas, MD, PhD⁴; Hrvoje Bogunović,
 7 PhD¹; Ursula Schmidt-Erfurth, MD^{1,2}

8 **Affiliations:**

9 ¹ OPTIMA - Laboratory for Ophthalmic Image Analysis, Department of
 10 Ophthalmology and Optometry, Medical University of Vienna, Vienna, Austria

11 ² Vienna Reading Center, Department of Ophthalmology and Optometry,
 12 Medical University of Vienna, Vienna, Austria.

13 ³ Vienna Clinical Trial Center, Department of Ophthalmology and Optometry,
 14 Medical University of Vienna, Vienna, Austria.

15 ⁴ Department of Ophthalmology and Optometry, Medical University of
 16 Vienna, Vienna, Austria.

17 **Short Title:** Predict Function from OCT using AI

18 **Address for Correspondence:**

19 Ursula Schmidt-Erfurth, MD
 20 Professor and Chair, Department of Ophthalmology, Medical University of
 21 Vienna
 22 Spitalgasse 23, 1090 Vienna, Austria
 23 Tel.: +43 1 40400 79310; Fax: +43 1 40400 79320
 24 Email: ursula.schmidt-erfurth@meduniwien.ac.at

Funding/Support: The financial support by the Christian Doppler Research Association, the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the FWF (Austrian Science Fund; Grant number FG 9-N) is gratefully acknowledged. We thank the NVIDIA Corporation for GPU donation.

Competing Interests: PS: none; WDV: none; SMW: Consultancy for Bayer, Novartis, Böhringer Ingelheim, Santen, Speaker Fees for Roche, Support for attending meetings and/or travel from Bayer, Novartis, Roche; JIO: none; MB: none; TA: none; MA: none; GM: none; HB: Speaker Fees for Bayer, Roche; SE: Scientific consultancy for Genentech, Novartis, Roche, RetInSight, Research support from Kodiak, Apellis

26 This work had been accepted for Poster Presentation at the Annual Meeting of
27 the Association of Research in Vision and Ophthalmology 2019 and
28 accordingly published as an abstract in IOVS.

29 **Keywords:** retina, artificial intelligence, deep learning, optical coherence
30 tomography, microperimetry, age-related macular degeneration, clinical trial
31 endpoints

32

33 **Précis:** To address limitations of conventional retinal function measures, we
34 propose a deep learning method that provides functional information directly
35 from standard imaging (OCT), showing the way from subjective
36 psychophysical tests towards objective functional measures.

37

38 **Abstract**

39 **Purpose:** Currently used measures of retinal function are limited by being
40 subjective, non-localized and/or taxing for patients. To address these
41 limitations, we sought to develop and evaluate a deep learning (DL) method to
42 automatically predict a functional endpoint (retinal sensitivity) from
43 structural optical coherence tomography (OCT) images.

44 **Design:** Retrospective cross-sectional study.

45 **Subjects:** In total, 714 volumes of 289 patients were used in this study.

46 **Methods:** A novel deep learning algorithm was developed to automatically
47 predict a comprehensive retinal sensitivity map from OCTs. 463 SD-OCT
48 volumes from 174 patients and their corresponding microperimetry
49 examinations (Nidek MP-1) were used for development and internal
50 validation, with a total of 15,563 retinal sensitivity measurements. Patients
51 presented either a healthy macula, early or intermediate AMD, choroidal
52 neovascularization (CNV) or geographic atrophy (GA). In addition, an
53 external validation with 251 volumes of 115 patients was performed,
54 comprising three different patient populations suffering from diabetic
55 macular edema (DME), retinal vein occlusion (RVO) or epiretinal membrane
56 (ERM).

57 **Main Outcome Measures:** We evaluated the performance of the algorithm
58 using Mean Absolute Error (MAE), limits of agreement (LoA), and correlation
59 coefficients of point-wise sensitivity (PWS) and mean sensitivity (MS).

60 **Results:** The algorithm achieved a MAE of 2.34dB/1.30dB for point-wise
61 sensitivity (PWS)/mean sensitivity (MS), LoA of 5.70/3.07 for PWS/MS and
62 Pearson and Spearman correlation coefficients of 0.66/0.84 and 0.68/0.83
63 for PWS/MS. On the external test set, the method achieved a MAE of
64 2.73dB/1.66dB for PWS/MS.

65 **Conclusions:** The proposed approach allows predicting retinal function at
66 each measured location directly from the OCT scan, demonstrating how
67 structural imaging can serve as a surrogate of visual function. Prospectively,
68 the approach may help to complement retinal function measures, explore the
69 linkage between image-based information and retinal functionality, improve
70 disease progression monitoring or provide objective surrogate measures for
71 future clinical trials.

72

73 **Introduction**

74 Primary endpoints of therapeutic interventions are patient-centered variables
75 that directly represent the clinical outcome of interest and are used to prove
76 the efficacy of therapies¹. Alternatively, surrogate endpoints describe physical
77 signs, laboratory measures and/or anatomical characteristics that are
78 expected to reflect changes in disease progression and therefore in primary
79 endpoints.² Its usage can be preferable over primary endpoints for their cost-
80 effectiveness, reproducibility or because they are easier and faster to obtain.³
81 Moreover, surrogates may be used as supportive endpoints in clinical studies
82 to assess characteristics of the therapeutic effect in more detail.⁴ In
83 ophthalmology, the most relevant standard primary endpoint is central visual
84 acuity (VA), even though it can be unreliable and does not provide a
85 comprehensive measurement of the entire visual function, remaining a poor
86 functional endpoint for many retinal diseases due to its origin exclusively
87 based on foveal performance.^{5,6}

88 Microperimetry offers the opportunity to overcome this limitation by
89 projecting a localized visual stimulus onto multiple positions of the retina.⁷ It
90 is the most comprehensive test of macular function and provides a non-
91 invasive measurement of focal and global retinal sensitivity. It has been shown
92 that microperimetry is more sensitive to macular functional deficit than best-
93 corrected VA (BCVA) and low luminance VA (LLVA).⁸ However, several
94 issues make microperimetry difficult to apply in practice: the examination
95 takes time and can be exhausting for patients⁹; the measurements quality
96 depends on patient motivation and disease¹⁰; and only a limited number of

97 points can be tested, making small areas of sensitivity loss difficult to capture.

98 **11**

99 These drawbacks are undermining its usage as surrogate endpoint in clinical
100 trials and research, when investigating the association between structural
101 information and visual function. Having reliable, localized and easy to obtain
102 measurements of retinal sensitivity would provide a powerful tool to facilitate
103 and advance processes both in research and clinical practice.

104 This is particularly relevant in age-related macular degeneration (AMD), one of
105 the worldwide leading causes of blindness and visual impairment.^{12, 13} AMD
106 leads to progressive destruction of foveal integrity, first anatomically and
107 consecutively in terms of visual function. This destruction ranges from little
108 functional deficiency in early stages to severe and irreversible visual loss in
109 advanced stages.^{13, 14, 15} While BCVA testing with standardized test charts is
110 routinely used, a reliable, objective and clinically applicable endpoint is still
111 missing in this field. Its limitations have been recently demonstrated by
112 comparing BCVA and LLVA with contrast sensitivity (CS), with CS showing a
113 better correlation with both morphological features and vision-related quality
114 of life in AMD patients.^{16, 17}

115 In this work, we propose to combine the advantages of high-resolution
116 spectral domain optical coherence tomography (SD-OCT) imaging,
117 topographic microperimetry examinations and artificial intelligence (AI) to
118 learn relationships between structure and function of the central retina in a
119 fully automated way. In general, predicting sensitivity measurements directly
120 from OCT allows using non-invasive standard imaging as a surrogate

121 endpoint for retinal sensitivity, complementing VA and microperimetry in its
122 application. This may help to explore the linkage between image-based
123 information and retinal functionality, assess retinal function in patient groups
124 where microperimetry cannot be applied, retrospectively analyze pre-existing
125 databases from a functional perspective (i.e. where no microperimetry had
126 been conducted), improve disease progression monitoring or provide
127 objective surrogate measures for future clinical trials. We introduce a novel
128 deep learning algorithm to regress an en-face 2D retinal sensitivity map from
129 a 3D SD-OCT volume, where a retinal sensitivity value is assigned to each
130 location in the retina.

131

132

133 **Methods**

134 Image Dataset

135 This study was conducted in adherence to the tenets of the Declaration of
136 Helsinki, and ethics approvals were obtained by the Ethics Committee of the
137 Medical University of Vienna Submission Nr 1246/2016, 2094/2018 and
138 2095/2018. The study is a retrospective data analysis and the ethics
139 committee decided that informed patient consent is not necessary.

140 We extracted and merged data from two large-scale in-house datasets,
141 consisting of 10,630 SD-OCT scans and 9,765 microperimetry acquisitions,
142 respectively. 2,123 visits of 587 eyes of 473 patients had corresponding SD-
143 OCT and MP acquisitions on the same day. After applying the exclusion

144 criteria, the final primary dataset included 463 OCT volumes and
145 corresponding microperimetry measures of 195 eyes of 174 patients (see
146 **Figure 1**). The dataset consists of healthy subjects ($n_{visits}=50$, $n_{patients}=29$),
147 patients with early/intermediate AMD ($n_{visits}=151$, $n_{patients}=66$), CNV
148 ($n_{visits}=186$, $n_{patients}=49$) or GA ($n_{visits}=76$, $n_{patients}=30$). In total 15,563
149 micropertimetric retinal sensitivities were measured. The data has been
150 acquired between April 12, 2007 and October 2, 2014.

151 For external validation , we used another test set with 9263 retinal sensitivity
152 measurements from 251 visits of 115 patients, suffering from diabetic macular
153 edema (DME, $n=51$), retinal vein occlusion (RVO, $n=44$) or epiretinal
154 membrane (ERM, $n=20$).

155

156 All microperimetry samples were mapped on the corresponding OCT volumes
157 by using a semi-automated registration procedure, aligning the fundus image
158 of the microperimetry device with an en-face projection of the OCT volume
159 (see **Supplement**).

160 All OCT volumes were acquired using Cirrus HD-OCT (Carl Zeiss Meditec,
161 Dublin CA) with a $2\text{mm} \times 6\text{mm} \times 6\text{mm}$ field of view centered at fovea and a
162 $1024 \times 512 \times 128$ voxel image resolution. Microperimetry data was acquired
163 using Nidek Microperimetry-1 (MP-1, Nidek Technologies Srl, Padova, Italy),
164 using four different test patterns with 25, 33, 41 and 61 points, respectively.
165 Goldmann III stimuli were used, with intensity values ranging from 0 to 20
166 decibels (dB) on a 1 (0.1 log) step scale, with 0 dB representing the brightest
167 luminance of 127 cd/m^2 .

168 Deep Learning Method

169 We developed an innovative AI approach that regresses retinal sensitivity
170 values for each A-scan in a given OCT volume (**Figure 2**). The proposed
171 method is based on deep learning, a state-of-the-art machine learning
172 technique. By means of a large training dataset of SD-OCT volumes and their
173 associated microperimetry points, our method automatically learns the
174 relationship between structure (OCT) and function (retinal sensitivity).

175 Our novel method, namely *ReSensNet*, is a convolutional neural network that
176 uses a 3D patch of the OCT volume as input, and regresses a 2D en-face
177 retinal sensitivity map as output, with a continuous output value for each A-
178 Scan. This yields faster prediction during application, compared to a model
179 that would predict only a single value from the patch. The final model can be
180 iteratively applied to retrieve a high-resolution en-face sensitivity map for the
181 whole volume.

182 The proposed network has an encoder-decoder structure. The encoder learns
183 abstract features, captures the context and reduces the spatial dimensionality
184 in the 3D domain. The decoder operates in the 2D domain, using the 3D
185 context provided by the encoder to generate the 2D 'en-face' output. In
186 standard architectures, this context is provided at multiple levels of depth by
187 simple skip connections that directly connect the encoder and the decoder.¹⁶
188 Due to the mismatch of dimensions (3D vs. 2D) between encoder and decoder,
189 we introduce novel "*3D-to-2D blocks*" that replace direct skip connections (see
190 **Supplement**).

191

192 Experimental Setup

193 The primary dataset was randomly divided on patient level into 70% training,
194 10% validation and 20% test set, so data of one patient only appears in exactly
195 one set. This means that in case of multiple volumes of the same patient, all
196 volumes are added to the same subset. While the model learns from the
197 training set, the validation set is used to evaluate performance during
198 development and to select the final model. The test set serves as an
199 independent data set and is used to provide a final unbiased evaluation of
200 model's generalization performance. The split resulted in 13,415/772/1,376
201 retinal sensitivity measurements, 400/22/41 volumes or 126/16/32 subjects
202 for the training/validation/test set. The external test set was used for
203 evaluation of the algorithm's generalization performance on additional
204 diseases. Centered at each retinal sensitivity measurement, we extracted 3D-
205 patches of size $1024 \times 64 \times 16$ voxels as input and corresponding retinal
206 sensitivity values as target output.

207

208 Statistical Evaluation / Validation of Regression-Performance

209 We evaluated the performance of our proposed model to regress the retinal
210 sensitivity values from OCT volumes on the test set, using the sensitivity
211 measurements ($n=1,376$) as ground-truth. Mean absolute error (MAE) of
212 point wise sensitivity (PWS) and mean sensitivity (MS) between regressed and
213 ground-truth values were used as metrics.^{11, 17, 18} To account for potential bias
214 caused by multiple scans from the same patient, we computed the average
215 error per patient, and subsequently the mean over these values, for both

216 metrics. We also calculated the limits of agreement (LoA) ¹⁹ for PWS and MS,
 217 defined as:

218
$$\text{limits of agreement} = \text{MSD} \pm 1.96 \times \sqrt{\frac{\sum(d_1 - d_2)^2}{n}}$$

219 where $\text{MSD} = \sum \frac{(d_1 - d_2)}{n}$ is the mean signed difference between the two
 220 compared measures d_1 and d_2 . As defined by Bland and Altman, LoA can be
 221 compared with the coefficient of repeatability (CR) ^{19, 20}. This allows a
 222 comparison of LoA of our method with previously published test-retest results
 223 based on CR, meaning that if the intervals of both are similar, then the lack of
 224 agreement between d_1 and d_2 can be explained by lack of repeatability. ^{17, 21, 19},
 225 ²⁰ This comparison is of particular relevance as microperimetry is known for
 226 its relatively high test-retest variability ¹⁷ and allows putting the evaluation
 227 results into better context. Both the linear and monotonic relationship
 228 between regressed and ground-truth values was assessed with Pearson and
 229 Spearman correlation coefficients. Statistically significant differences in
 230 PWS/MS between disease groups in the primary test set were evaluated with
 231 two-sided Kruskal-Wallis tests with $\alpha = 0.05$ and Bonferroni–Holm correction
 232 for multiple comparisons. In addition, qualitative results are presented.

233

234

235

236

237 **Results**238 Quantitative Results

239 **Table 1** summarizes the quantitative results in the test set. The total MAE
 240 over all sensitivity measurements was 2.34 dB for PWS and 1.30 dB for MS.
 241 The lowest MAE for PWS was found for healthy patients (2.03 dB), while
 242 patients with CNV showed the highest MAE (2.53 dB). With respect to LoA,
 243 the mean signed difference and the corresponding limits were -0.57 ± 5.70 dB
 244 for PWS and -0.47 ± 3.07 dB for MS, calculated over the whole test set. The
 245 ranges of LoA were [-6.27,5.13] for PWS and [-3.54,2.6] for MS, as depicted in
 246 the Bland-Altman plots of **Figure 3(a)(b)**. Boxplots illustrating the MAE
 247 distribution of PWS and MS are shown in **Figure 3(c)(d)**. Statistically
 248 significant differences were only observed in PWS between the healthy and CNV
 249 groups ($p=0.024$). The Pearson/Spearman correlation coefficient was
 250 0.66/0.68 for PWS and 0.84/0.83 for MS over the entire test set.

251 **Table 2** presents quantitative results on the external test. The total MAE was
 252 2.73 dB for PWS and 1.66 dB for MS. More detailed results are provided in the
 253 **Supplement**, demonstrating that the proposed *ReSensNet* outperformed 1D
 254 and 2D methods as well as a baseline following the image projection network
 255 (IPN) architecture proposed by Li et al.²²

256

257 Qualitative Results correspond to established knowledge of the
 258 pathophysiology of AMD

259 **Figure 4** illustrates qualitative results on the test set on four different
260 exemplary conditions (Healthy, early/intermediate AMD, CNV, GA). On the
261 right hand side, the en-face maps for the original microperimetry
262 measurements, the prediction map, the regression map visualized only in the
263 same regions as the original microperimetry measurements and the model
264 activation maps are shown for each subject. The activations are generated
265 using integrated gradients²², highlighting areas considered as relevant by the
266 method to produce each prediction. Moreover, five B-scans of the input OCT
267 volume (slices 32/48/64/80/96), their activation maps, an overlay of both
268 and their retinal sensitivity predictions are shown on the left side of **Figure 4**
269 for each subject.

270 In the first rows, the healthy macula shows evenly distributed activations over
271 the scan volume (see en-face activation map). At the same time, the predicted
272 sensitivity response is high and homogeneous, which is reflected in the
273 microperimetry measurements and plausible for a healthy macula.
274 Additionally, a lower retinal sensitivity is predicted at the border of the scan.
275 This is likely due to the black borders introduced by the motion correction
276 algorithm on the left and right sides of each B-scan, which might be
277 interpreted by the method as photoreceptor dysfunction (no retina - no
278 function). The second example corresponds to an early AMD case with
279 drusen-covered areas. Compared to the healthy sample, a lower sensitivity is
280 predicted in this eye in general. The model activations are observed to be
281 much higher in regions with drusen than in other areas. A reduced retinal
282 sensitivity is predicted in the area of the largest drusen in volume (center of
283 slice 64). This is in line with recent observations, associating drusen size with

284 sensitivity loss.²⁴ The third row presents the retina of a CNV patient. Low
285 sensitivity values are predicted in abnormal regions such as areas presenting
286 with pathological fluid. These zones also exhibit high activation values, as
287 observed in the en-face activation maps. This aligns with the currently known
288 association between retinal fluid and sensitivity loss.²⁵ Finally, the GA patient
289 in the last row demonstrates a distinct low retinal sensitivity prediction in the
290 area of photoreceptor loss, consistent with clinical observations.²⁵
291 Particularly high activations are observed in these regions. In addition,
292 outliers with exceptional high error on the test set are visualized in **Figure 5**.

293

294

295 **Discussion**

296 OCT enables non-invasive investigation of retinal morphology and
297 pathological alterations of various diseases at a micrometer resolution,
298 becoming a cornerstone for retinal analysis in clinical routine and scientific
299 studies. OCT scans are routinely analyzed by ophthalmologists for diagnosis,
300 determining treatments, or to inform relevant clinical decisions. However,
301 regulatory authorities have not yet accepted OCT as an endpoint in clinical
302 trials, as a direct link between visual function and morphological imaging has
303 yet to be established. As a consequence, visual acuity remains by far the most
304 common endpoint in retinal therapeutic trials, despite its well-known deficits
305 due to the psycho-physical nature of the test method and its limited
306 representative value for overall macular function.

307 Establishing associations between retinal micro-structures and function using
308 OCTs and VA measurements is challenging for multiple reasons. First,
309 researchers are overwhelmed by the large amount of data (i.e. numerous OCT
310 volumes with millions of pixels each), meaning that a manual inspection and
311 analysis of OCT image data on a large-scale basis is practically unfeasible.
312 Here, automated analysis can facilitate the study of large OCT datasets. AI
313 approaches based on deep learning have recently shown impressive results in
314 ophthalmic OCT imaging.¹² Second, the non-localized nature of the VA test
315 hinders a direct comparison of structure-function relationship. Conceptually,
316 this limitation can be overcome by microperimetry examinations, which offer
317 corresponding topographic measurements of retinal sensitivity values.
318 Practically, however, microperimetry is limited due to effort and expense of
319 the examination, the number of points that can be tested and its
320 reproducibility.

321 In this context, the proposed deep learning method could complement and
322 extend currently known functional measures and endpoints. The model
323 provides a retinal sensitivity map from the OCT volume in a fully automated
324 way (see **Figure 2**), without requiring any additional costly and time-
325 consuming examination. Assuming that the deep learning model has correctly
326 learned the structure-function relation during training, using it for regressing
327 the sensitivity map may allow to exploit standard imaging (OCT) as a
328 surrogate endpoint for retinal function. Besides a potential reduction of the
329 burden for patients and examiners, this would also allow assessing retinal
330 function in patient groups where microperimetry cannot be applied, e.g.
331 children or adults with loss in central fixation or other disabilities.

332 Furthermore, the proposed method produces a detailed topographic map,
333 providing a sensitivity value for each individual A-scan (see **Figure 4**). This
334 helps to detect small areas of sensitivity loss and track small deviations, which
335 can in turn be important for early diagnosis, monitoring or patient
336 management.

337 The obtained sensitivity map is also reproducible, as the proposed *ReSensNet*
338 is will always produce the same output given the same input. In contrast, the
339 reproducibility of microperimetry measurements is known to be limited, e.g. it
340 depends on the disease type and stage and is influenced by the learning effect
341 or the patients motivation. ^{10, 21, 26, 27}

342 The method can be used to analyze the retina and its properties in a more
343 comprehensive way, collecting many more focal areas than BCVA which
344 exclusively derives from the fovea. The detailed sensitivity map can be used in
345 large-scale analyses to extract knowledge about diseases and their
346 progression. This is not limited to prospective studies, but also allows
347 retrospective analysis of pre-existing databases from a functional perspective,
348 e.g. data where no microperimetry had been conducted.

349 Moreover, the output of the model could be used to guide and/or complement
350 a subsequent microperimetry examination. Finally, the model itself can be
351 analyzed to evaluate the automatically learned relation between structure and
352 function, e.g. by means of activation maps (**Figure 4**).

353 We qualitatively observed that these activations were distributed more
354 uniformly in healthy scans compared to diseased ones. We hypothesize that
355 the network only regresses a high sensitivity value if the overall appearance is

356 healthy, leading to the necessity to check the whole input image and spreading
357 the activations more evenly. In contrast, a low sensitivity value can be caused
358 by a lesion that is restricted locally, causing the network to focus on these
359 anomalous regions and to assign them a higher importance. This theory is
360 supported by the qualitative finding that the importance is higher in diseased-
361 areas when comparing en-face sensitivity and activation maps, linking the
362 importance with regression of a lower retinal sensitivity.

363 Another observation was that a certain amount of activation (i.e. importance
364 for the prediction) is located in the area of the choroidal layers below the
365 retina. We hypothesize that on the one hand this can be explained by detected
366 hyper-transmission that occurs with geographic atrophy. On the other hand,
367 functional changes may be reflected in or correlated with structural changes of
368 choroidal layers visible in OCT, which have been learned by the model. Even
369 though not verified, this shows the explorative potential of the method,
370 possibly leading to new hypotheses. Further discussion about clinical
371 observations in our results is provided in the **Supplement**.

372 A model that has properly learned the structure-function relationship between
373 OCT volume and retinal sensitivity is expected to lie within the test-retest
374 variability. While this information is not available for our patient population,
375 previous work has reported a test-retest variability of the Nidek MP-1 device
376 in patients with a range of macular diseases.¹⁷ Chen et al.¹⁷ reported a CR of
377 5.56 for PWS and 1.81 for MS, meaning that 5% of repeated microperimetry
378 measurements would have an error which is larger than 5.56 dB for PWS and
379 likewise for MS. In comparison, our results with respect to LoA lie in a similar
380 range as the reported CR values, indicating that our model is able to achieve a

381 performance close to the test-retest variability of the microperimetry device.
382 This in turn means that the lack of agreement between the regressed and
383 measured sensitivity values to a certain extent can be explained by the lack of
384 repeatability of the microperimetry test itself. Moreover, the correlation
385 coefficients show a strong correlation between the regressed and ground-truth
386 values. These results indicate that the *ReSensNet* has successfully learned
387 basic structure-function relationships between OCT scans and retinal
388 sensitivity values in our dataset. This statement is also supported by the
389 results on the external test set, demonstrating that the learned model
390 generalizes to new diseases not used for training to a certain extent. We
391 hypothesize that the model learned fundamental structure-function
392 relationships, resulting in only a slightly lower performance on new diseases.

393 However, we also noticed a number of mismatches of the predicted and
394 measured values (see outliers in Figure 3c). In Figure 5a, both the original
395 sensitivity measurement and the model prediction show a reduced sensitivity
396 value in the area of the druse. At the same time, this pathology seems to
397 impact the prediction less than the sensitivity measurement. In Figure 5b and
398 5c, the presence of certain structural changes (subretinal fluid in 5b, hyper-
399 reflective foci in 5c) impact the prediction to a greater degree than the actual
400 sensitivity. We believe that these results are related to the known limited
401 association between structural changes and their impact on function in AMD
402 30. The prediction accuracy of the model might further be restricted by the
403 image quality. In Figure 5d, the differentiation of neurosensory retinal layers
404 is limited due to low signal-to-noise ratio, causing an underestimation of the
405 true retinal sensitivity. In Figure 5e and 5f we can see original microperimetry

406 measurements that are close to an adjacent region with a similar predicted
407 sensitivity value. We hypothesize that the location of these measurement
408 points are slightly off due to an inaccurate registration of the microperimetry
409 data and the OCT volumes, causing a high error in the quantitative evaluation.
410 In general, a certain extent of outliers must be expected when evaluating
411 microperimetry data due to the relatively high test-retest variability of the
412 examination itself, even in healthy subjects. Moreover, this is the first work
413 proposing a deep learning model to predict retinal sensitivity for AMD, DME,
414 RVO and ERM patients. While we are convinced that the proposed approach
415 has a great potential, more work beyond this proof-of-concept study is
416 required.

417 Investigating the structure-function relationship using AI ^{11, 28, 29, 30, 31} is an
418 active field of research in general. In contrast to previous work, our model
419 takes 3D context into account, which is important for the performance of the
420 model (see **Supplement**). This is especially relevant, as the correlation
421 between OCT and microperimetry in AMD is more complex than in other
422 diseases (e.g. Macular Telangiectasia Type 2), ^{25, 32} making our task
423 particularly challenging. Moreover, previously published work on
424 simultaneous evaluation of microperimetry and OCT scans report datasets
425 with patient numbers below one hundred, ^{33, 34, 35, 36, 37, 38, 25, 39, 40} which may
426 be related to the high cost of microperimetry examinations. In comparison, we
427 developed and evaluated our model on a larger number of samples, namely
428 714 volumes of 289 patients.

429 As it is known that the dataset size affects the performance of deep learning
430 models and the number of samples is limited in general due to the costly

431 nature of microperimetry acquisitions, ⁴⁴ we included microperimetry data
432 acquired with different test patterns, included both eyes of some patients and
433 multiple visits of one eye if available. On the one hand, considering the case of
434 similar samples in the training data, we hypothesize that this can help the
435 model to learn to differentiate between noise in the original microperimetry
436 measurements (e.g. due to test-retest variability) and the actual true
437 underlying sensitivity differences caused by morphological changes. On the
438 other hand, multiple scans of one patient can also differ significantly (e.g.
439 unilateral late AMD with no late AMD in the fellow eye ^{45, 46}), offering the
440 opportunity for the AI model to achieve better generalizability by learning
441 from a dataset with more variations. Furthermore, the varying number of test
442 points per patient means that patients with a larger number of samples have
443 potentially more influence on model training compared to other patients with
444 fewer test points in the training set. In principle this is meaningful, as a
445 sample with more test points is likely to contain more information. On the
446 contrary, using this heterogeneous dataset also constitutes a limitation of this
447 study, in particular since a negative impact on training is possible. At the same
448 time, the quantitative evaluation remains unaffected as we accounted for
449 potential bias across patients in the statistical methods.

450 A limitation of the proposed model is that it cannot replace the “real”
451 microperimetry examination, at least at this early stage of development.
452 However, we believe the application of the presented method lies in the
453 complementation of currently used measures of function (rather than
454 replacement), as it provides new opportunities to examine and assess retinal
455 function. Another limitation is that activation maps do not explain feature

456 interactions or combinations and do not provide global importances across
457 the entire dataset. However, activation maps for individual examples can still
458 be useful to gain new insights and explore the characteristics of the trained
459 deep learning model. The fact that the number of patients in the test set per
460 disease was limited and that no test-retest reliability evaluation was available
461 for the used datasets are limitations as well. Nevertheless, previously
462 published work allowed putting the obtained results in context. More research
463 with respect to this topic is left for future work.

464 It would be also interesting to focus future work on how the algorithm
465 performs in areas with specific common OCT artefacts, although we
466 qualitatively did not observe any influence by e.g. posterior shadowing from
467 retinal vessels. Moreover, as lower retinal sensitivity is predicted at the border
468 of the scan (due to black borders), future use should take this into account
469 when calculating mean sensitivity. Another limitation of our study is that the
470 information about age was not available in the used dataset. Future work
471 should take the potential influence of age into account when validating, in
472 particular when aiming to transfer such a model to real world clinical practice.

473 In conclusion, we propose a novel AI method to predict detailed retinal
474 sensitivity maps (function) from OCT volumes (structure). It has potential to
475 overcome limitations of conventional function measures and to use OCT as a
476 surrogate endpoint for visual function of patients. We are convinced that this
477 paper contributes to a paradigm shift from subjective psychophysical tests
478 towards objective functional measures based on fast, non-invasive imaging
479 and artificial intelligence. While the proposed approach has potential for
480 multiple application scenarios such as complementing VA and microperimetry

481 or expanding our abilities to investigate retinal diseases from a functional
482 perspective, more work beyond this proof-of-concept study is required to
483 transfer such a model to real world clinical practice. Nevertheless, this work
484 constitutes a promising step towards exploring the linkage between image-
485 based information and function, its underlying concept being transferable to
486 other imaging modalities, diseases or functional targets.

487 **References**

1. Medeiros FA. Biomarkers and surrogate endpoints in glaucoma clinical trials. *British Journal of Ophthalmology*. 2015;99:599–603.
2. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*. 1989;8:431–440.
3. Cole ED, Ferrara D, Novais EA, Louzada RN, Waheed NK. Clinical trial endpoints for optical coherence tomography angiography in neovascular age-related macular degeneration. *Retina*. 2016;36:S83–S92.
4. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Statistics in medicine*. 2012;31:2973–2984.
5. Elliott DB. The good (logMAR), the bad (Snellen) and the ugly (BCVA, number of letters read) of visual acuity measurement. *Ophthalmic and Physiological Optics*. 2016;36:355–358.
6. Cocce KJ, Stinnett SS, Luhmann UFO, et al. Visual function metrics in early and intermediate dry age-related macular degeneration for use as clinical trial endpoints. *American journal of ophthalmology*. 2018;189:127–138.
7. Midena E, Pilotto E. Microperimetry in age-related macular degeneration. *Eye*. 2017;31:985.
8. Wu Z, Ayton LN, Guymer RH, Luu CD. Low-luminance visual acuity and microperimetry in age-related macular degeneration. *Ophthalmology*. 2014;121:1612–1619.
9. Csaky KG, Patel PJ, Sepah YJ, et al. Microperimetry for geographic atrophy secondary to age-related macular degeneration. *Survey of ophthalmology*. 2019;64:353–364.
10. Cideciyan AV, Swider M, Aleman TS, et al. Macular function in macular degenerations: repeatability of microperimetry as a potential outcome measure for ABCA4-associated retinopathy trials. *Investigative ophthalmology & visual science*. 2012;53:841–852.
11. Kihara Y, Heeren TFC, Lee CS, et al. Estimating Retinal Sensitivity Using Optical Coherence Tomography With Deep-Learning Algorithms in Macular Telangiectasia Type 2. *JAMA network open*. 2019;2:e188029–e188029.
12. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. *Progress in retinal and eye research*. 2018.
13. Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular

- degeneration. *The Lancet*. 2012;379:1728–1738.
14. Lamoureux EL, Mitchell P, Rees G, et al. Impact of early and late age-related macular degeneration on vision-specific functioning. *British Journal of Ophthalmology*. 2011;95:666–670.
 15. Sevilla MB, McGwin Jr G, Lad EM, et al. Relating retinal morphology and function in aging and early to intermediate age-related macular degeneration subjects. *American journal of ophthalmology*. 2016;165:65–77.
 16. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention, 2015.
 17. Chen FK, Patel PJ, Xing W, et al. Test-retest variability of microperimetry using the Nidek MP1 in patients with macular disease. *Investigative ophthalmology & visual science*. 2009;50:3464–3472.
 18. Liefers B, González-Gonzalo C, Klaver C, van Ginneken B, Sánchez CI. Dense Segmentation in Selected Dimensions: Application to Retinal Optical Coherence Tomography. Paper presented at: Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning, 2019.
 19. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical methods in medical research*. 1999;8:135–160.
 20. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*. 1986;327:307–310.
 21. Wong EN, Mackey DA, Morgan WH, Chen FK. Intersession test-retest variability of conventional and novel parameters using the MP-1 microperimeter. *Clinical ophthalmology (Auckland, NZ)*. 2016;10:29.
 22. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. Paper presented at: Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017.
 23. Schlanitz FG, Bogunovic H, Vogl WD, et al. The Impact of Drusen on Retinal Sensitivity in non-exudative Age-Related Macular Degeneration: A point-to-point Analysis. *Submitted to Investigative Ophthalmology & Visual Science: ARVO abstract*. 2020.
 24. Sulzbacher F, Kiss C, Kaider A, et al. Correlation of SD-OCT features and retinal sensitivity in neovascular age-related macular degeneration. *Investigative ophthalmology & visual science*. 2012;53:6448–6455.

25. Palkovits S, Hirnschall N, Georgiev S, Leisser C, Findl O. Test–Retest Reproducibility of the Microperimeter MP3 With Fundus Image Tracking in Healthy Subjects and Patients With Macular Disease. *Translational Vision Science & Technology*. 2018;7:17–17.
26. Wu Z, Ayton LN, Guymer RH, Luu CD. Intrasession test–retest variability of microperimetry in age-related macular degeneration. *Investigative ophthalmology & visual science*. 2013;54:7378–7385.
27. Aslam TM, Zaki HR, Mahmood S, et al. Use of a neural net to model the impact of optical coherence tomography abnormalities on vision in age-related macular degeneration. *American journal of ophthalmology*. 2018;185:94–100.
28. von der Emde L, Pfau M, Dysli C, et al. Artificial intelligence for morphology-based function prediction in neovascular age-related macular degeneration. *Scientific reports*. 2019;9:1–12.
29. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnav R. Inference of visual field test performance from OCT volumes using deep learning. *arXiv preprint arXiv:1908.01428*. 2019.
30. Pfau M, von der Emde L, Dysli C, et al. Determinants of cone and rod functions in geographic atrophy: AI-based structure-function correlation. *American journal of ophthalmology*. 2020;217:162–173.
31. Issa PC, Troeger E, Finger R, Holz FG, Wilke R, Scholl HPN. Structure-function correlation of the human central retina. *PLoS One*. 2010;5:e12864.
32. Acton JH, Smith RT, Hood DC, Greenstein VC. Relationship between retinal layer thickness and the visual field in early age-related macular degeneration. *Investigative ophthalmology & visual science*. 2012;53:7618–7624.
33. Forte R, Querques G, Querques L, Leveziel N, Benhamou N, Souied EH. Multimodal evaluation of foveal sparing in patients with geographicatrophy due to age-related macular degeneration. *Retina*. 2013;33:482–489.
34. Hariri AH, Tepelus TC, Akil H, Nittala MG, Sadda SR. Retinal sensitivity at the junctional zone of eyes with geographic atrophy due to age-related macular degeneration. *American journal of ophthalmology*. 2016;168:122–128.
35. Echols BS, Clark ME, Swain TA, et al. Hyperreflective foci and specks are associated with delayed rod-mediated dark adaptation in nonneovascular age-related macular degeneration. *Ophthalmology Retina*. 2020;4:1059–1068.
36. Iaculli C, Barone A, Scudieri M, Palumbo MG, Delle Noci N. Outer retinal tubulation: characteristics in patients with neovascular age-related macular

- degeneration. *Retina*. 2015;35:1979–1984.
37. Pilotto E, Convento E, Guidolin F, et al. Microperimetry features of geographic atrophy identified with en face optical coherence tomography. *JAMA ophthalmology*. 2016;134:873–879.
38. Takahashi A, Ooto S, Yamashiro K, et al. Photoreceptor damage and reduction of retinal sensitivity surrounding geographic atrophy in age-related macular degeneration. *American journal of ophthalmology*. 2016;168:260–268.
39. Wu Z, Ayton LN, Luu CD, Guymer RH. Relationship between retinal microstructures on optical coherence tomography and microperimetry in age-related macular degeneration. *Ophthalmology*. 2014;121:1445–1452.

489

Figure Legends:

Figure 1: Dataset preparation as a flow diagram. First, we merged two large in-house databases (Microperimetry MP-1, Cirrus SD-OCT) from our department. Second, we excluded unreliable microperimetry acquisitions. Third, cases were excluded if registration could not be done (e.g. due to bad image quality). In the last step, we excluded samples with a diagnosis different than healthy, early- or intermediate AMD, CNV or GA. *AMD: age-related macular degeneration. CNV: choroidal neovascularization. GA: geographic atrophy. RVO: retinal vein occlusion. DME: diabetic macular edema.*

Figure 2: Overview of the proposed deep learning method. Our convolutional neural network generates 2D output patches from 3D input patches (multiple neighboring A-scans within and across B-Scans). It regresses a 2D sensitivity map covering all A-scans by being iteratively applied across the whole 3D-OCT volume. The proposed '3D-to-2D blocks' between the 3D-encoder and the 2D-decoder are depicted as black arrows. Details of the method are provided in the **Supplement**.

Figure 3: Plots showing quantitative results on the test set. Bland-Altman plots of the regressed and measured PWS (a) and MS values (b) of the test set in dB. The mean difference and the limits of agreement are denoted as gray dashed lines. Boxplots of the mean absolute error (MAE) on the test set

in dB, between the measured and regressed values of the proposed *ReSensNet* model, are shown in (c) for PWS and (d) for MS values.

Figure 4: Qualitative results on the test set for four exemplary AMD stages. The integrated gradients method²² was used to calculate activation maps of the input, highlighting the areas that were most relevant to a given prediction. For each subject the activation map (first row), an overlay of the map on the B-scan (second row), the original B-Scan (third row), and the corresponding regression result (fourth row) are shown. On the right side, the en-face activation map (median projection, first row and first column), the en-face sensitivity regression map of the proposed *ReSensNet* (first row, second column), the original microperimetry measurements (second row, first column) and the regression map visualized only in the same regions as the original microperimetry measurements (second row, second column) are illustrated. The black arrows next to the en-face maps denote the positions of the corresponding visualized B-scans (slice 32/48/64/80/96).

Figure 5: Additional qualitative results on the test set showing outliers with exceptionally high errors. For each sample the B-Scan, the corresponding predicted sensitivity regression values of the proposed ReSensNet ('pred.'), and the original microperimetry measurements ('orig. meas.') are illustrated one below the other. Yellow arrows in the B-scans denote areas of interest.

List of Supplemental Digital Content:

OphRet_SDC 1_2021-10-21_Function_Structure.pdf

Linking Function and Structure with ReSenseNet: Predicting Retinal Sensitivity from Optical Coherence Tomography using Deep Learning

Philipp Seeböck, PhD^{1,2}; Wolf-Dieter Vogl, PhD¹; Sebastian M. Waldstein, MD, PhD¹; Jose Ignacio Orlando, PhD¹; Magdalena Baratsits³; Thomas Alten¹; Mustafa Arikán, MSc¹; Georgios Mylonas, MD, PhD⁴; Hrvoje Bogunović, PhD¹; Ursula Schmidt-Erfurth, MD^{1,2}

Précis: To address limitations of conventional retinal function measures, we propose a deep learning method that provides functional information directly from standard imaging (OCT), showing the way from subjective psychophysical tests towards objective functional measures.

Tables:

Table 1: Quantitative results of the proposed deep learning model ReSensNet on the test set.

	Healthy	Early/Intermediate	CNV	GA	Total
PWS	2.03 (\pm 0.28)	2.34 (\pm 0.56)	2.53 (\pm 0.62)	2.46 (\pm 0.46)	2.34 (\pm 0.53)
MS	1.53 (\pm 0.56)	1.53 (\pm 0.79)	1.25 (\pm 0.93)	0.89 (\pm 0.26)	1.30 (\pm 0.73)

The patient-wise calculated mean absolute error (MAE) is depicted in decibel (dB) for point-wise sensitivity (PWS) and mean sensitivity per volume (MS). The standard deviation is provided in parentheses.

Tables:

Table 2: Quantitative results of the proposed deep learning model ReSensNet on the external test set, including patients with DME, RVO or ERM.

	DME	RVO	ERM	Total
PWS	2.86 (± 0.98)	2.60 (± 0.87)	2.70 (± 0.70)	2.73 (± 0.90)
MS	1.76 (± 1.49)	1.43 (± 1.17)	1.93 (± 1.15)	1.66 (± 1.33)

The mean absolute error (MAE) is depicted in decibel (dB) for point-wise sensitivity (PWS) and mean sensitivity per volume (MS). The standard deviation is provided in parentheses.

10,630 Potential Cirrus OCTs

9,765 Potential 'Nidek MP-1' microperimetry examinations with color fundus image

2,123 Linked visits of 587 eyes of 473 patients

975 Excluded (one or more invalid sensitivity values; Reliability test false positive rate > 30%; Instable fixation-point distribution)

1,148 Visits of 442 eyes of 383 patients

411 Excluded (samples could not be registered)

737 Visits of 328 eyes of 301 patients

23 Excluded (samples contained other diseases)

Final Dataset

463 Visits of 195 eyes of 174 patients (healthy, early- or intermediate AMD, CNV, GA)

External Test Set

251 Visits of 115 patients (DME, RVO, ERM)

