

1 **OCT-based Visual Field Estimation Using Segmentation-free 3D CNN**
2 **Shows Lower Variability than Subjective Standard Automated Perimetry**

3

4 *Makoto Koyama**, MD¹; *Satoru Inoda*, MD, PhD²; *Yuta Ueno*, MD, PhD³; *Yoshikazu Ito*, MD³;
5 *Tetsuro Oshika*, MD, PhD³; *Masaki Tanito*, MD, PhD⁴

6 ¹ Minamikoyasu Eye Clinic, 2-8-30 Minamikoyasu, Kimitsu-shi, Chiba, Japan

7 ² Department of Ophthalmology, Jichi Medical University, Shimotsuke, Tochigi, Japan

8 ³ Department of Ophthalmology, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

9 ⁴ Department of Ophthalmology, Shimane University Faculty of Medicine, Shimane, Japan

10

11 ***Correspondence:**

12 Makoto Koyama, MD

13 Minamikoyasu Eye Clinic, 2-8-30, Minamikoyasu, Kimitsu-shi, Chiba, 299-1162, Japan

14 E-mail: minamikoyasuganka@gmail.com

15

16 **Key Words:** Visual Field; Optical coherence tomography; Artificial intelligence; Glaucoma;
17 Perimetry

18

19 **Abstract**

20 **Purpose:** To train and evaluate a segmentation-free 3D convolutional neural network (3DCNN)
21 model for estimating visual field (VF) from optical coherence tomography (OCT) images and to
22 compare the residual variability of OCT-based estimated VF (OCT-VF) with that of Humphrey Field
23 Analyzer (HFA) measurements in a diverse clinical population.

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

25 **Participants:** 5,351 patients (9,564 eyes) who underwent macular OCT imaging and Humphrey
26 Field Analyzer (HFA) tests (24-2 or 10-2 test patterns) at a university hospital from 2006 to 2023.
27 The dataset included 47,653 paired OCT-VF data points, including various ocular conditions.

28 **Methods:** We trained a segmentation-free 3DCNN model based on the EfficientNet3D-b0
29 architecture on a comprehensive OCT dataset to estimate VF. We evaluated the model's performance
30 using Pearson's correlation coefficient and Bland–Altman analysis. We assessed residual variability
31 using a jackknife resampling approach and compared OCT-VF and HFA datasets using generalized
32 estimating equations (GEE), adjusting the number of VF tests, follow-up duration, age, and clustering
33 by eye and patient.

34 **Main Outcome Measures:** Correlations between estimated and measured VF thresholds and mean
35 deviations (MDs), and residual variability of OCT-VF and HFA.

36 **Results:** We observed strong correlations between the estimated and measured VF parameters
37 (Pearson's r: 24-2 thresholds 0.893, MD 0.932; 10-2 thresholds 0.902, MD 0.945; all $p < 0.001$).
38 Bland–Altman analysis showed good agreement between the estimated and measured MD, with a
39 slight proportional bias. GEE analysis demonstrated significantly lower residual variability for OCT-
40 VF than for HFA (24-2 thresholds: 1.10 vs. 2.48 dB; 10-2 thresholds: 1.20 vs. 2.48 dB; all $p < 0.001$,
41 Bonferroni-corrected), with lower variability across all test points, severities, and ages, thus
42 highlighting the robustness of the segmentation-free 3DCNN approach in a heterogeneous clinical

43 sample.

44 **Conclusions:** A segmentation-free 3DCNN model objectively estimated VF from OCT images with
45 high accuracy and significantly lower residual variability than subjective HFA measurements in a
46 heterogeneous clinical sample, including patients with glaucoma and individuals with other ocular
47 diseases. The improved reliability, lower variability, and objective nature of OCT-VF highlight its
48 value for enhancing VF assessment and monitoring of various ocular conditions, potentially
49 facilitating earlier detection of progression and more efficient disease management.

50

51 **Introduction**

52 Visual field (VF) testing is crucial for diagnosing and monitoring various ocular conditions,
53 particularly glaucoma, a leading cause of irreversible blindness worldwide.¹⁻⁴ Although the
54 Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, Jena, Germany) remains the gold standard for
55 VF assessment, HFA testing is limited by its subjective nature, high test-retest variability, and time-
56 consuming process.^{1,5} These limitations can lead to delayed detection of disease progression and
57 complicate clinical decision-making.

58 Optical coherence tomography (OCT) has revolutionized ophthalmic imaging, providing
59 high-resolution, objective assessments of ocular structures.⁶ The relationship between structural
60 changes observed in OCT and functional deficits measured by VF testing has led to the integration of
61 OCT with artificial intelligence (AI) to estimate VF directly from OCT images. Early approaches
62 often utilized segmentation-based 2D models.⁷⁻⁹ While these models demonstrated valuable insights,
63 the requirement for manual segmentation or quality checks could be time-consuming and potentially
64 limit the scalability of the approach. This constraint may have posed challenges in preparing large-
65 scale datasets for model training, which could influence the generalizability of VF estimation
66 performance.

67 Recent advancements have led to the emergence of 3D models for VF estimation,
68 potentially capturing more comprehensive structural information.^{10,11} Additionally, some 2D
69 approaches now utilize cross-sectional OCT images without segmentation, offering a simplified
70 method.¹²⁻¹⁴ These developments suggest a trend toward more efficient VF estimation techniques,
71 potentially offering a complementary approach to traditional subjective perimetry through objective,
72 OCT-based assessments.

73 To address the limitations of previous methods and capitalize on recent advancements,
74 building upon our preliminary research published as a Japanese preprint using data from a private

75 ophthalmology clinic,¹⁵ we adopted a segmentation-free, 3D convolutional neural network (3DCNN)
76 model to estimate VF from macular OCT images. Our approach eliminates manual segmentation or
77 labeling, enabling the model to learn from a comprehensive dataset based on disease status without
78 exclusions. This could improve the model's generalizability and reduce bias from selective data
79 inclusion. The primary aims of this study are to (1) train and evaluate a segmentation-free 3DCNN
80 model for objectively estimating VF from OCT images in a diverse clinical population, with a
81 particular focus on glaucoma, and (2) compare the residual variability of OCT-based estimated VF
82 (OCT-VF) with that of subjective HFA measurements, potentially offering a more reliable method
83 for VF assessment.

84

85 **Methods**

86 **Study Design and Participants**

87 The Institutional Review Board of Shimane University Hospital approved this retrospective
88 study (IRB No. KS20230719-3, approved on August 10, 2023), which adhered to the tenets of the
89 Declaration of Helsinki. The study included all patients who underwent macular OCT imaging and
90 VF testing at Shimane University Hospital, a tertiary referral center specializing in glaucoma,
91 between October 1, 2006, and October 19, 2023. Due to the retrospective nature of the study, the IRB
92 waived the requirement for informed consent. We employed an opt-out approach, posting study
93 information on the hospital's website and premises to allow patients to decline participation.

94

95 **Inclusion Criteria**

96 We included eyes that met the following criteria: (1) availability of at least one macular
97 OCT scan with a signal strength index (SSI) ≥ 7 and (2) completion of at least one VF test using the
98 HFA with 30-2, 24-2, or 10-2 test patterns using the Swedish Interactive Threshold Algorithm

99 standard protocol. We trimmed the peripheral points for 30-2 VF tests to match the 24-2 pattern.
100 Consistent with a previous report,⁸ we excluded VF tests with false-positive, false-negative, or
101 fixation loss rates $\geq 33\%$ to ensure the data quality for training the 3DCNN model. We included all
102 eligible eyes regardless of their underlying ocular condition or disease status, ensuring a diverse and
103 representative dataset.

104 To address potential artifacts, we implemented automatic exclusion methods for the dataset.
105 For the upper eyelid artifacts in the 24-2 test pattern, we excluded the VF if the difference in the mean
106 value of the three nasal points in any two adjacent rows of the top three rows exceeded 8 dB. For lens
107 rim artifacts, we divided the fields into four quadrants and excluded those where the difference
108 between the mean outermost and adjacent inner points exceeded 5 dB in three or more quadrants. We
109 applied these criteria to paired data for model training, validation, and testing via 10-fold cross-
110 validation to reduce the impact of potential artifacts on the model's learning and evaluation process.
111 While this approach may deviate from real-world clinical scenarios, it allows for consistent
112 assessment of the model's core performance.

113 We acquired OCT images using the RS-3000, RS-3000 Advance, or RS-3000 Advance2
114 OCT device (Nidek, Gamagori, Japan) with a 9 mm \times 9 mm macular scan protocol. On the basis of
115 the manufacturer's recommendation, we chose the SSI threshold of 7 to ensure the reliability of OCT
116 scans and to exclude those with significant media opacities that could affect image quality.

117

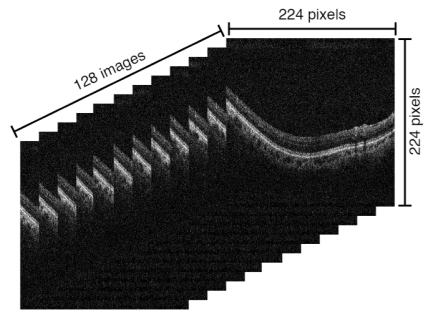
118 **Data Acquisition and Preprocessing**

119 To reduce variability, we constructed time-based regression lines for each VF test point and
120 used them as target values for training the 3DCNN model. For eyes with five or more VF tests, we
121 calculated VF thresholds and mean deviations (MDs) corresponding to the OCT acquisition date
122 using these regression lines. We set the "validity period" for these eyes as 30 days \times the number of

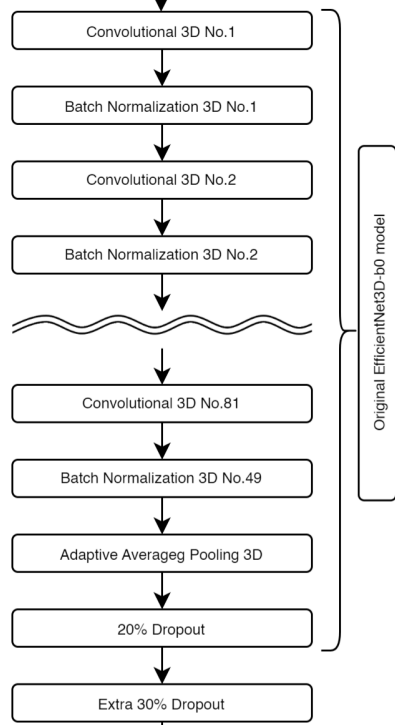
123 VF tests (n), with an upper limit of 240 days. For eyes with two to four VFs, we also constructed
124 regression lines but limited the "validity period" to 90 days, and we did not extend the regression lines
125 beyond the first and last VF tests; instead, we fixed the values at these endpoints. This approach aimed
126 to reduce variability while minimizing potential errors from extrapolation. We excluded data pairs if
127 the interval between the OCT scan and the most recent VF test exceeded the calculated validity period.
128 We set positive regression slopes to zero to account for the progressive nature of glaucomatous VF
129 loss. We included the data for eyes with a single VF test if an OCT scan was performed within 90
130 days of the VF test date.

131 We assigned missing data points a mask value of 1. During training, we limited VF
132 thresholds to 0 to 33 dB, and MDs to 0 to -33 dB, setting values outside these ranges to the respective
133 upper or lower limits. We included all eligible paired data from eyes with multiple OCT scans and
134 VF tests in the analysis.

135



Input: 224 x 224 x 128 3D macular OCT image



Output: 52 points (24-2) + 68 points (10-2) + MDs

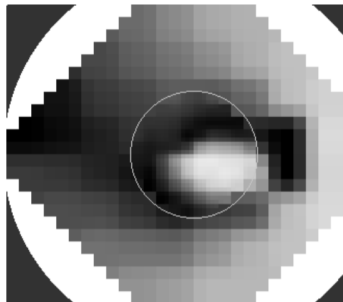


Figure 1. Schematic representation of the segmentation-free 3D convolutional neural network model architecture. We based the model on the EfficientNet3D-b0 architecture and added a 30% dropout layer to mitigate overfitting. The input consists of standardized 224 × 224 × 128 OCT images, and the output includes estimated VF thresholds and MDs for both 24-2 and 10-2 test patterns. OCT = optical coherence tomography; VF = visual field; MD = mean deviation.

148 **Model Architecture and Configuration**

149 We based the segmentation-free 3DCNN model used in this study on the EfficientNet3D-
150 b0 architecture,¹⁶ and added a 30% dropout layer to mitigate overfitting. Figure 1 presents a schematic
151 representation of the model architecture. We trained the model from scratch using the comprehensive
152 OCT dataset, which included scans from patients with glaucoma and other ocular conditions, without
153 any exclusions based on disease status. Table S1 provides an overview of the population
154 characteristics for the dataset used to train the 3DCNN model.

155 We standardized the OCT images to $224 \times 224 \times 128$ resolution and normalized them via
156 min-max normalization. For the HFA24-2 and HFA10-2 datasets, we applied z-score normalization
157 using the mean and standard deviation of the training dataset to ensure consistent data scaling. The
158 model's output consisted of estimated VF thresholds (52 points for the 24-2 test pattern and 68 points
159 for the 10-2 test pattern) and their respective MDs.

160 We horizontally flipped the left eye data and combined them with the right eye data.
161 Following a previous report,⁹ we consistently applied vertical flipping as a data augmentation
162 technique across all phases (training, validation, and testing). During the testing phase, the estimation
163 accuracy improved by averaging the results of both the original and vertically flipped inputs for each
164 OCT image.

165 We trained the model via the Adam optimizer with a mini-batch size of 4, incrementally
166 adjusted the learning rate from $6e-4$ to $1e-3$ over three epochs and then decreased it to $6e-4$ over five
167 epochs. We trained to minimize the mean squared error between the estimated and measured VF data.
168 To account for missing data points, we multiplied the backpropagation calculation by $(1 - \text{mask})$
169 during training, ensuring that the model's learning was unaffected by gaps in the VF data.

170

171 **Model Training and Evaluation**

172 We employed a 10-fold cross-validation approach, randomly dividing patients into training,
173 validation, and test sets at an 8:1:1 ratio. This patientwise split ensured that data from the same patient
174 did not appear in more than one set, preventing data leakage and allowing unbiased model evaluation.
175 We selected the epoch that showed the best performance on the validation set to evaluate the model's
176 performance on the test set, ensuring the use of the model's optimal weights for the final assessment.

177 We calculated the mean absolute error (MAE) and Pearson's and Spearman's correlation
178 coefficients to assess the relationships between the estimated and measured VF parameters. We used
179 Bland–Altman plots to evaluate the agreement between the estimated and measured MDs.
180 Additionally, we analyzed the relationships between the MAE and VF severity, refractive errors, and
181 individual test points to assess the model's performance across different clinical scenarios and VF
182 regions.

183

184 **Residual Variability Analysis Preparation**

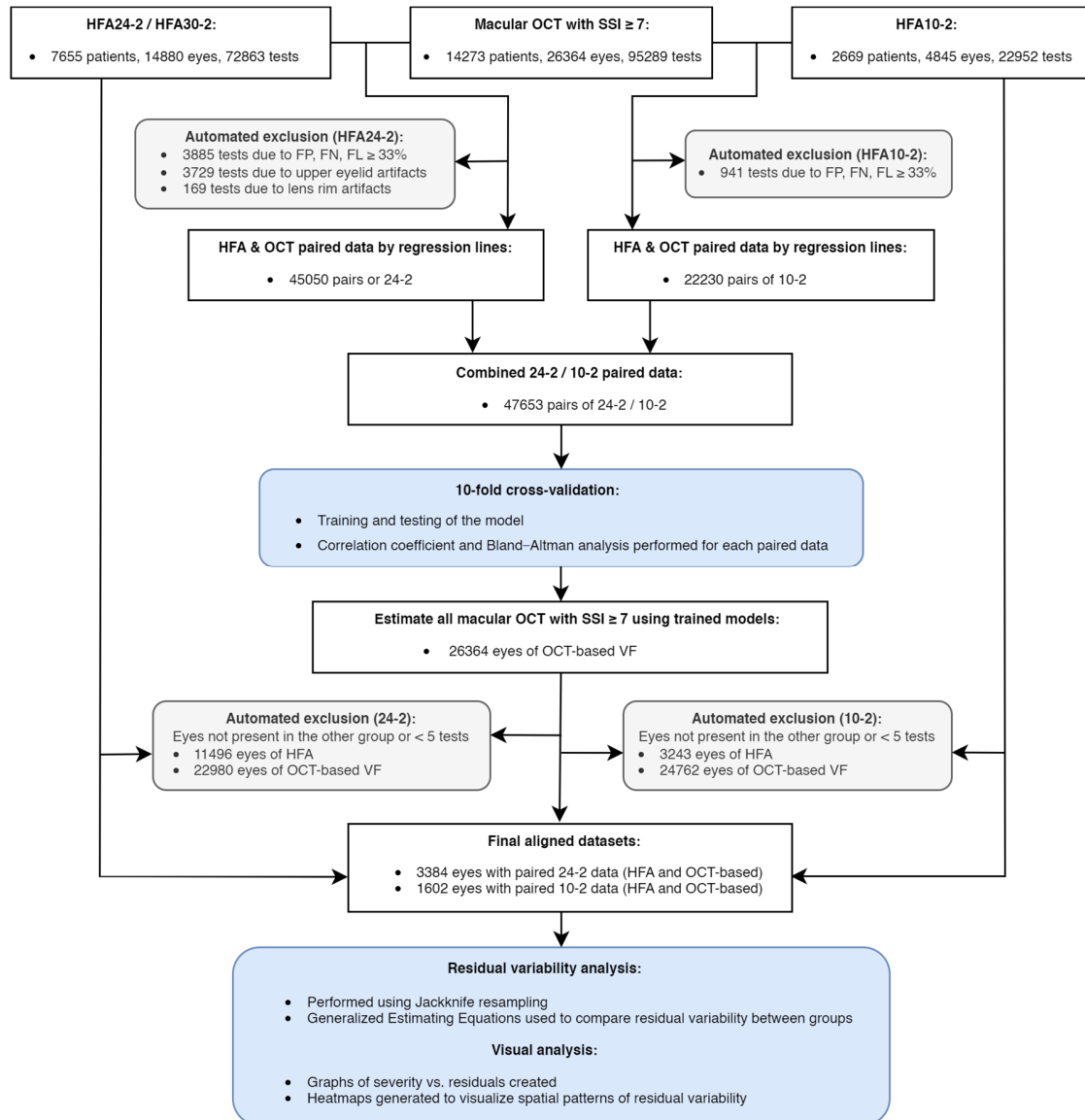
185 To assess the clinical applicability and reliability of our OCT- VF model, we prepared
186 datasets to compare the variability between OCT-VF and measured HFA. We used the trained
187 3DCNN model to estimate VF values for all available macular OCT images with an SSI ≥ 7 . We
188 selected the model for estimation on the basis patient ID, assigning patients in the test set of the 10-
189 fold cross-validation to their corresponding model and patients not included in any fold to a randomly
190 selected model. We included all HFA tests in the analysis, regardless of their false-positive rates,
191 false-negative rates, fixation loss rates, upper eyelid artifacts, or lens rim artifacts. This approach
192 ensured that the comparison reflected real clinical scenarios. To further refine the comparison
193 between the OCT-VF and HFA measurements, we excluded eyes with fewer than 5 VF tests from
194 both groups. Additionally, to ensure a fair comparison, we included only eyes common to the OCT-
195 VF and HFA groups in the analysis, excluding eyes unique to either group.

196

197 **Statistical Analysis**

198 We employed a jackknife resampling approach to compare the residual variability between
199 the OCT-VF and HFA measurements. We calculated residuals for each eye by iteratively excluding
200 one data point, fitting a regression line (at each test point for 24-2 and 10-2 test patterns), computing
201 the residual for the excluded point, and then calculating mean residuals. We used generalized
202 estimating equations (GEE) models to compare the mean residuals between the OCT-VF and HFA
203 datasets, adjusting for the number of tests, follow-up duration, age, and clustering by eye and patient.
204 We visualized relationships between VF severity, age, and residual variability via boxplots with
205 regression lines. We created heatmaps to assess the spatial distribution of mean residuals across VF
206 test points for the OCT-VF and HFA datasets.

207 Figure 2 shows the flowchart of the study design, outlining the data acquisition,
208 preprocessing, model training, residual variability analysis preparation, and statistical analysis steps.
209 We applied only the automated exclusion criteria described above, with no additional manual
210 exclusions. We performed statistical analyses using Python (version 3.11.2) with scikit-learn (version
211 1.41.post1) and statsmodels (version 0.14.2) packages. We implemented the deep learning model
212 using PyTorch (version 2.01).



213
214
215
216
217
218

Figure 2. Flowchart of the study design and data processing. This diagram illustrates the steps involved in data acquisition, preprocessing, model training, residual variability analysis preparation, and statistical analysis. HFA = Humphrey Field Analyzer; OCT = optical coherence tomography; SSI = signal strength index; FP = false positive; FN = false negative; FL = fixation loss.

219

220 **Results**

221 **Model Performance in Estimating Visual Field**

222 The 3DCNN model demonstrated a high correlation between the measured and estimated
223 VF parameters for 24-2 and 10-2 test patterns (Table S2 and Fig. S3). Pearson's correlation
224 coefficients (r) were high for both VF thresholds and MDs (24-2 thresholds $r=0.893$, MD $r=0.932$;
225 10-2 thresholds $r=0.902$, MD $r=0.945$; all $p < 0.001$), indicating high accuracy in estimating both
226 pointwise VF thresholds and global VF parameters. The Bland–Altman analysis demonstrated
227 satisfactory overall concordance between the estimated and measured MDs, with minimal mean
228 differences. Nevertheless, all analyses showed slight proportional biases (Fig. S4).

229 The model demonstrated consistent performance across various VF damage levels,
230 including advanced cases (Fig. S5). The model's estimation accuracy remained relatively stable across
231 various OCT focus values, indicating the minimal impact of refractive errors on performance (Fig.
232 S6). The spatial distribution of MAE for each test point showed generally consistent estimation
233 accuracy across most test locations in 24-2 and 10-2 test patterns, with some regional variations
234 observed (Fig. S7).

235

236 **Residual Variability Analysis using Generalized Estimating Equations (GEE)**

237 Table 3 presents the mean residuals and standard deviations for HFA and OCT-VF datasets.
238 For both 24-2 ($n = 3384$ eyes) and 10-2 ($n = 1602$ eyes) test patterns, the OCT-VF group showed
239 lower mean residuals than the HFA group. For thresholds, OCT-VF vs. HFA mean residuals were
240 1.10 vs. 2.48 dB for 24-2, and 1.20 vs. 2.48 dB for 10-2. For MDs, OCT-VF vs. HFA mean residuals
241 were 0.82 vs. 1.34 dB for 24-2, and 0.87 vs. 1.22 dB for 10-2. We used GEE models to compare the
242 mean residuals between the OCT-VF and HFA datasets, adjusting for the number of tests, follow-up
243 duration, and clustering by eye and patient (Table S4). In all four analyses, the OCT-VF group

244 exhibited significantly lower mean residuals than the HFA group (all $p < 0.001$). We applied a
 245 Bonferroni correction to account for multiple comparisons (adjusted significance level: $0.05/4 =$
 246 0.0125); all p values remained significant after correction.

247

248 **Table 3.** Population characteristics and residual variability analysis results for the HFA and OCT-
 249 VF datasets

Characteristics	HFA24-2	OCT-VF 24-2	HFA10-2	OCT-VF 10-2
Number of patients	1852	1852	889	889
Number of eyes	3384	3384	1602	1602
Age (years)	67.1 ± 13.8	68.5 ± 13.9	67.5 ± 13.8	67.8 ± 13.8
Mean deviation (dB)	-8.58 ± 8.47	$-9.33 \pm 8.36^*$	-11.4 ± 9.7	$-11.9 \pm 9.4^*$
Threshold residuals (dB)	2.48 ± 2.14	1.10 ± 0.89	2.48 ± 2.46	1.20 ± 1.02
MD residuals (dB)	1.34 ± 0.95	0.82 ± 0.66	1.22 ± 0.92	0.87 ± 0.69
Number of tests	12.4 ± 7.6	11.2 ± 4.7	9.7 ± 4.5	12.8 ± 4.5
Follow-up duration (days)	2776 ± 1593	1950 ± 1091	2024 ± 973	1979 ± 1049

250 HFA = Humphrey Field Analyzer; OCT = optical coherence tomography; OCT-VF = OCT-based
 251 estimated visual field; MD = mean deviation. The values are presented as the means \pm standard
 252 deviations.

253 *Estimated values from OCT-VF models.

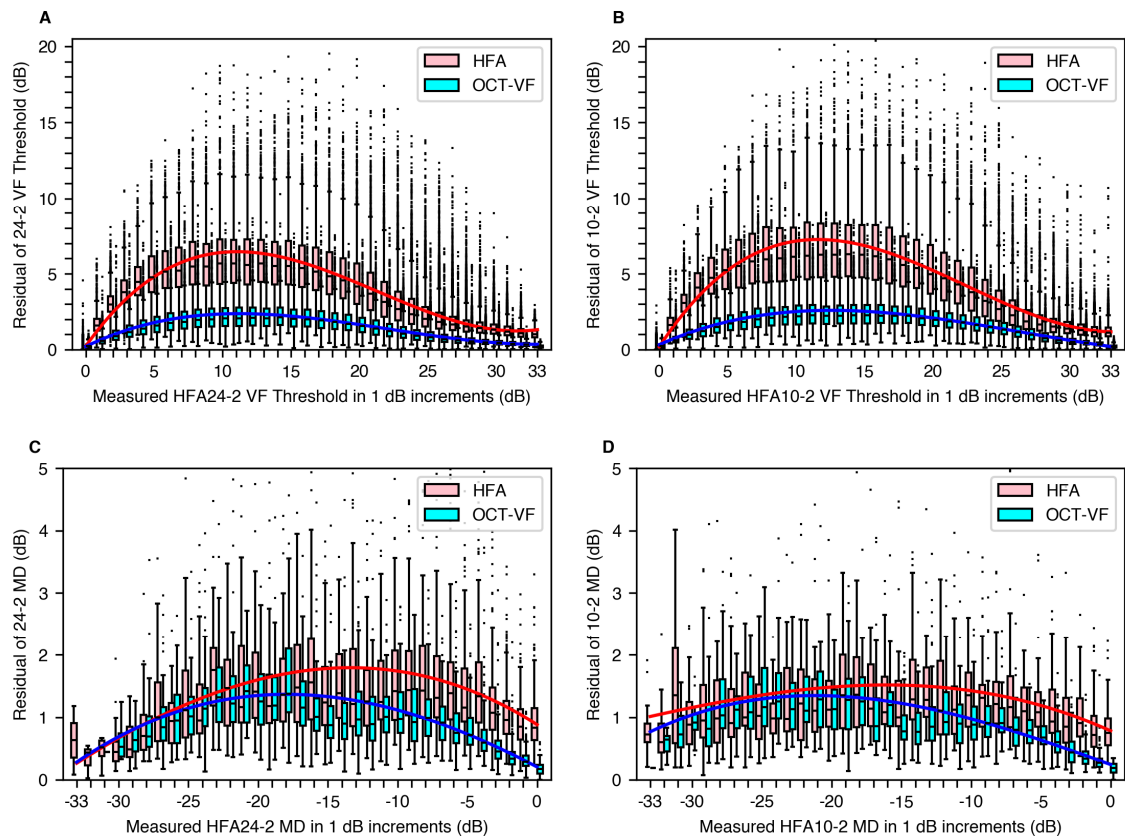
254

255

256 Relationship between VF Severity and Residual Variability

257 Figure 8 presents boxplots with cubic regression lines illustrating the relationship between
 258 VF severity (horizontal axis, dB) and residual variability (vertical axis, dB) for HFA and OCT-VF
 259 datasets. The figures compare the performances of HFA and OCT-VF methods for 24-2 (Fig. 8a and
 260 8c) and 10-2 (Fig. 8b and 8d) test patterns. In all figures, the cubic regression lines for the OCT-VF

261 dataset generally lie below those of the HFA dataset, indicating lower residual variability across most
262 levels of VF severity.
263



264

265 **Figure 8.** Comparison of residual variability between HFA measurements and OCT-VFs across VF
266 severity levels. (A) 24-2 thresholds (pointwise values), (B) 10-2 thresholds (pointwise values), (C)
267 24-2 MD, and (D) 10-2 MD. The horizontal axis represents VF severity, and the vertical axis
268 represents residual variability. Each panel shows boxplots of residuals for each severity increment,
269 along with cubic regression lines. In all panels, the cubic regression lines for the OCT-VF dataset
270 generally lie below those of the HFA dataset, indicating lower residual variability across most levels
271 of VF severity for OCT-VF.

272 HFA = Humphrey Field Analyzer; OCT = optical coherence tomography; OCT-VF = OCT-based
273 estimated visual field; VF = visual field; MD = mean deviation.

274

275

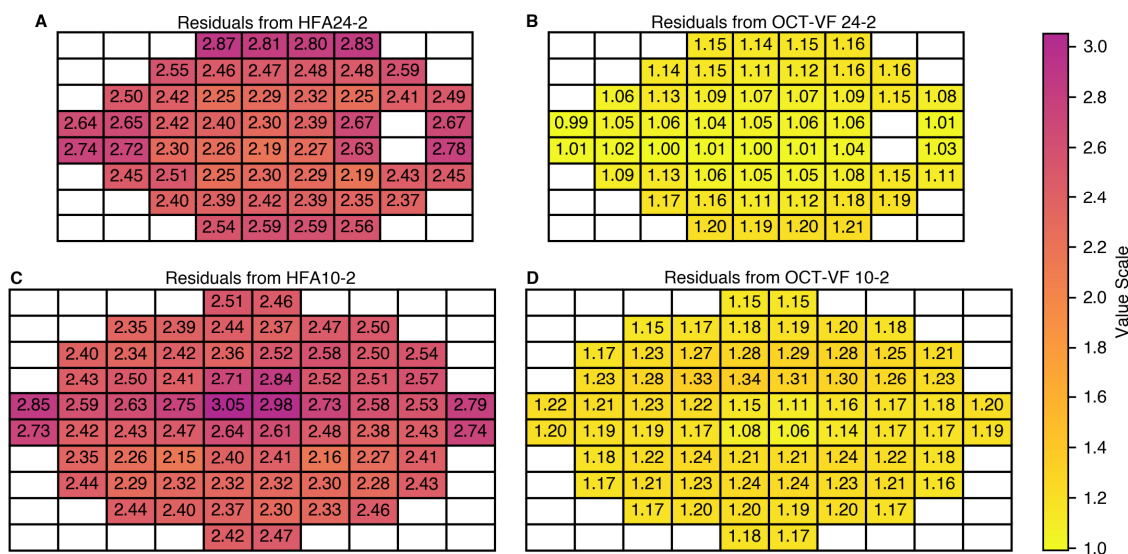
276

277 **Heatmaps of Residual Variability for Each Test Point**

278 Figure 9 presents heatmaps of residual variability for each test point in the HFA (Fig. 9a,
 279 9c) and OCT-VF (Fig. 9b, 9d) datasets. We displayed the heatmaps in a two-dimensional arrangement
 280 mimicking the spatial layout of the VF tests. Figures 9a and 9b show heatmaps for 24-2 test pattern
 281 thresholds, whereas Figures 9c and 9d show heatmaps for 10-2 test pattern thresholds. We
 282 horizontally flipped the left eye data and integrated them with the right eye data.

283 The values in the HFA heatmaps range from 2.15 to 3.05. In contrast, the values in the OCT-
 284 VF heatmaps range from 0.99 to 1.34, demonstrating substantially lower residual variability for the
 285 OCT-VF dataset across all test points.

286



287

288 **Figure 9.** Heatmaps of residual variability for each test point. (A) HFA 24-2, (B) OCT-VF 24-2, (C)
 289 HFA 10-2, and (D) OCT-VF 10-2. The color scale represents the magnitude of residual variability at
 290 each test point. OCT-VF shows consistently lower residual variability than HFA measurements across
 291 all test points, with values ranging from 0.99 to 1.34 dB for OCT-VF compared with 2.15 to 3.05 dB
 292 for HFA measurements. We horizontally flipped the left eye data and integrated them with the right
 293 eye data.
 294 HFA = Humphrey Field Analyzer; OCT = optical coherence tomography; OCT-VF = OCT-based
 295 estimated visual field.
 296

297 **Relationship between Age and Residual Variability**

298 Analysis of the relationship between age and residual variability for HFA and OCT-VF (Fig.
299 S10) revealed that residual variability increased with age for both methods across all measures.
300 However, OCT-VF demonstrated consistently lower residual variability than HFA across all age
301 groups, with a less pronounced increase in variability with age. This was particularly evident for
302 threshold measurements. The difference in residual variability between HFA and OCT-VF widened
303 with increasing age.

304

305 **Discussion**

306 This study demonstrates that a segmentation-free 3DCNN model trained on a
307 comprehensive OCT dataset can estimate VF with significantly lower residual variability than HFA.
308 Using regression-based target values for OCT-VF model training inherently reduces variability
309 compared with raw HFA data. However, this approach's effectiveness depends on the model's
310 estimation accuracy, which our results show to be high. Indeed, our model exhibited strong
311 correlations between estimated and measured VF parameters, with consistent performance across
312 various levels of disease severity, VF test patterns (24-2 and 10-2), and refractive errors. This robust
313 performance addresses challenges reported in previous studies regarding severe VF loss estimation,¹⁰⁻
314 ¹⁴ highlighting the model's potential as a reliable tool for assessing and monitoring VF defects in a
315 diverse clinical population.

316 Compared with HFA, the markedly reduced residual variability of OCT-VF has significant
317 implications for managing various ocular conditions affecting the VF. Our analysis via generalized
318 estimating equations revealed that, compared with HFA, OCT-VF results in significantly lower
319 residual variability for 24-2 and 10-2 test patterns, even after adjusting for potential confounding
320 factors and applying a Bonferroni correction for multiple comparisons. This enhanced consistency

321 could enable earlier detection of disease progression and more timely intervention, potentially
322 slowing vision loss in patients with glaucoma and other conditions.

323 The lower variability observed in OCT-VF across various disease severities underscores the
324 objectivity and reliability of this approach. Our findings demonstrate that OCT-VF provides more
325 consistent results than HFA does, particularly when estimating pointwise sensitivities. This improved
326 reliability is a crucial strength of our study, as it suggests that OCT-based methods could offer more
327 accurate assessments of VF damage, enhancing clinical decision-making and patient management.
328 The spatial analysis of residual variability, as illustrated in our heatmaps, further supports the superior
329 performance of OCT-VF. These visualizations demonstrate reduced variability at each test point for
330 both the 24-2 and 10-2 test patterns, indicating the potential of OCT-VF to provide more reliable VF
331 sensitivity measurements across the entire tested area. Furthermore, analysis of age-related effects
332 revealed that while residual variability increased with age for both methods, OCT-VF maintained
333 relatively lower variability across all age groups, with the difference widening in older populations.
334 This underscores the robustness of OCT-VF in maintaining reliability across diverse patient
335 demographics.

336 Our study has the potential to significantly impact ophthalmic practice, particularly
337 glaucoma management. Various patient-related factors, such as proficiency, age, and cognitive
338 function, often influence the reliability of traditional subjective VF testing.^{1,5,17} By utilizing objective
339 OCT-based methods, we aim to mitigate the impact of these factors while providing a reliable
340 assessment of functional damage. This approach could complement or potentially replace HFA
341 testing in certain clinical scenarios, reducing the burden on patients and healthcare systems.

342 Our segmentation-free approach eliminates manual segmentation, enabling efficient
343 utilization of large-scale, real-world OCT datasets for model training and validation. This approach
344 allows the model to learn effectively without requiring specific clinical information such as disease

345 diagnoses or visual acuity data. This versatility could accelerate the development of AI-based tools
346 for assessing VF defects in various ocular conditions, potentially broadening their applicability in
347 diverse clinical settings.

348 Our study has several limitations. First, despite the use of a large and diverse OCT dataset,
349 we did not categorize or analyze data on the basis of specific clinical factors (e.g., disease type, visual
350 acuity, VF defect pattern), limiting our insights into the model's performance for specific conditions.
351 Second, the single-center nature of our study may restrict the generalizability of our findings. Future
352 research should validate our results via external datasets and explore the model's performance across
353 diverse patient populations and healthcare settings. Finally, we did not directly assess the impact of
354 reduced variability on clinical decision-making or patient outcomes. Further studies are needed to
355 evaluate how the improved reliability of OCT-VF translates into more effective management
356 strategies and better visual function preservation in glaucoma and other VF-affecting diseases.

357 In conclusion, our segmentation-free 3DCNN model has the potential to estimate visual
358 fields with significantly lower residual variability than HFA in a diverse clinical population. The
359 improved reliability and consistency of OCT-based estimated visual fields highlight their potential as
360 a valuable tool for assessing and monitoring visual field defects in various ocular conditions,
361 particularly glaucoma. As we refine and validate this approach, AI-based tools may become integral
362 for managing glaucoma and other ocular conditions affecting the visual field, enabling earlier
363 detection of progression, more efficient monitoring, and ultimately, better preservation of visual
364 function.

365

366 **References**

- 367 1. Delgado MF, Nguyen NTA, Cox TA, et al. Automated perimetry: a report by the American
368 Academy of Ophthalmology. *Ophthalmology*. 2002;109(12):2362-2374.
- 369 2. Weinreb RN, Leung CKS, Crowston JG, et al. Primary open-angle glaucoma. *Nat Rev Dis*
370 *Primers*. 2016;2(1):1-19.
- 371 3. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma
372 and projections of glaucoma burden through 2040: A systematic review and meta-analysis.
373 *Ophthalmology*. 2014;121(11):2081-2090.
- 374 4. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: A
375 review. *JAMA*. 2014;311(18):1901-1911.
- 376 5. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J*
377 *Ophthalmol*. 1989;108(2):130-135.
- 378 6. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science*.
379 1991;254(5035):1178-1181.
- 380 7. Xu L, Asaoka R, Kiwaki T, Murata H, Fujino Y, Yamanishi K. PAMI: A computational
381 module for joint estimation and progression prediction of glaucoma. In: *KDD '21:*
382 *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and*
383 *Data Mining*. Assoc. for Computing Machinery; 2021:3826-3834.
- 384 8. Asaoka R, Xu L, Murata H, et al. A joint multitask learning model for cross-sectional and
385 longitudinal predictions of visual field using OCT. *Ophthalmol Sci*. 2021;1(4):100055.

- 386 9. Hashimoto Y, Kiwaki T, Sugiura H, et al. Predicting 10-2 visual field from optical coherence
387 tomography in glaucoma using deep learning corrected with 24-2/30-2 visual field. *Transl Vis*
388 *Sci Technol.* 2021;10(13):28.
- 389 10. Mohammadzadeh V, Vepa A, Li C, et al. Prediction of central visual field measures from
390 macular OCT volume scans with deep learning. *Transl Vis Sci Technol.* 2023;12(11):5.
- 391 11. Chen Z, Shemuelian E, Wollstein G, Wang Y, Ishikawa H, Schuman JS. Segmentation-free
392 OCT-volume-based deep learning model improves pointwise visual field sensitivity
393 estimation. *Transl Vis Sci Technol.* 2023;12(6):28.
- 394 12. Kihara Y, Montesano G, Chen A, et al. Policy-driven, multimodal deep learning for predicting
395 visual fields from the optic disc and OCT imaging. *Ophthalmology.* 2022;129(7):781-791.
- 396 13. Hemelings R, Elen B, Barbosa-Breda J, et al. Pointwise visual field estimation from optical
397 coherence tomography in glaucoma using deep learning. *Transl Vis Sci Technol.*
398 2022;11(8):22.
- 399 14. Lazaridis G, Montesano G, Afgeh SS, et al. Predicting visual fields from optical coherence
400 tomography via an ensemble of deep representation learners. *Am J Ophthalmol.* 2022;238:52-
401 65.
- 402 15. Koyama M. Three-dimensional convolutional neural network model for estimating
403 glaucomatous visual fields based on optical coherence tomography. *Jxiv.* Preprint posted
404 online December 12, 2022. doi:10.51094/jxiv.170
- 405 16. Shi J. EfficientNet-PyTorch-3D. <https://github.com/shijianjian/efficientnet-pytorch-3d>; 2023
406 Accessed June 24, 2024.

407 17. Ichitani A, Takao E, Tanito M. Roles of cognitive function on visual field reliability indices
408 among glaucoma patients. *J Clin Med.* 2023;12:7119.

409

410 **Data availability statement:**

411 We are unable to make the datasets publicly available due to privacy and ethical considerations
412 related to patient data. We conducted the study on an opt-out basis, without obtaining explicit
413 consent from all participants to release their raw data.

414

415 **Acknowledgements:**

416 The authors would like to thank the Institutional Review Board of Shimane University Hospital for
417 their approval and guidance (IRB No. KS20230719-3, approved on August 10, 2023). We are
418 grateful to the staff at Shimane University Hospital for their support throughout this study. Our
419 sincere appreciation goes to all the patients who participated in this research, making this study
420 possible. This work was partially supported by a Grant-in-Aid for Scientific Research (21K16903)
421 from the Ministry of Education, Culture, Sports, Science and Technology of Japan to Satoru Inoda.
422 The authors declare that the funding body had no role in the design of the study, the collection,
423 analysis, and interpretation of data, or in writing the manuscript.

424

425 **Competing Interests:**

426 Makoto Koyama has potential future royalties from DeepEyeVision Inc. if the product is
427 commercialized, and has purchased stock in the company. He is also engaged in ongoing discussions
428 about potential future product development and commercialization with DeepEyeVision Inc.

429

430 **Funding Statement:**

431 Satoru Inoda received a Grant-in-Aid for Scientific Research (21K16903) from the Ministry of
432 Education, Culture, Sports, Science and Technology of Japan. The sponsor or funding organization had
433 no role in the design or conduct of this research.

434

435 **Author's Contributions:**

436 The first author, MK, was involved in all aspects of the study from conception to manuscript
437 writing. TO and MT supervised the overall research. MT was responsible for data collection. MK
438 developed and implemented the machine learning model-related code. MK, SI, YU, and YI
439 conducted data analysis and interpretation. MK drafted the initial manuscript, and all authors
440 actively participated in discussions to revise and improve the paper. Finally, all authors reviewed
441 and approved the final version of the manuscript.