Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI-derived OCT features track inter-eye contrast-sensitivity differences

Ren Zhang

rzhang@med.wayne.edu

Wayne State University School of Medicine

Research Article

Keywords: Optical coherence tomography, OCT, artificial intelligence, contrast sensitivity

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11575483/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

AI-derived OCT features track inter-eye contrast-sensitivity differences

Ren Zhang
Center for Molecular Medicine and Genetics
Wayne State University School of Medicine
Detroit, MI 48202, USA

rzhang@med.wayne.edu

Abstract

Purpose: AI-based OCT analysis can predict visual function across individuals, but these associations may reflect shared participant characteristics rather than eye-specific information. Comparing fellow eyes tests whether OCT features capture functional differences within the same person. The purpose is to test whether inter-eye OCT feature differences track contrast-sensitivity (CS) differences after visual-acuity (VA) adjustment.

Methods: We analyzed 2,015 AI-READI participants with bilateral OCT and visual-function measurements. A fixed RETFound encoder extracted OCT features. Photopic and mesopic ridge models were developed in training (n = 1,412) and applied unchanged to validation (n = 307) and test (n = 296). Analyses included eye-swap and outcome-permutation controls, VA adjustment, and exploratory retinal-thickness comparisons.

Results: Photopic correlations between predicted and observed inter-eye CS differences were 0.334 in validation and 0.318 in test; predictive R^2 values were 0.095 and 0.101. Mesopic correlations were 0.373 and 0.194; predictive R^2 values were 0.133 and 0.027. All associations exceeded control null expectations. VA-adjusted partial correlations were 0.272 (95% CI, 0.174–0.364) for photopic and 0.171 (95% CI, 0.069–0.264) for mesopic CS. AI features yielded numerically higher photopic correlations than thickness features; combining features did not consistently improve performance.

Conclusion: AI-derived OCT features tracked inter-eye CS differences after VA adjustment, with more consistent photopic performance.

Translational Relevance: AI-derived OCT features may complement visual acuity in assessing contrast sensitivity.

Keywords: Optical coherence tomography; OCT; artificial intelligence; contrast sensitivity

Introduction

Understanding how ocular structure relates to visual function is central to interpreting retinal imaging and assessing the consequences of eye disease. Contrast sensitivity provides a measure of visual function relevant to everyday tasks, complementing standard assessments of visual acuity[1,2]. Contrast sensitivity measures the ability to detect luminance differences and can vary substantially among eyes with similar visual acuity. Optical coherence tomography (OCT) provides noninvasive images of retinal microstructure, creating an opportunity to investigate the structural correlates of these functional differences[3]. Associations between OCT-derived retinal thickness and contrast sensitivity have been reported in healthy adults and in ocular disease, including glaucoma and early diabetic retinopathy[4-6]. Studies in diabetes have also linked contrast sensitivity to retinal structural and microvascular measures assessed using OCT and OCT angiography[7-9]. Structural and functional abnormalities have also been documented in diabetes without clinically evident retinopathy[10,11].

Artificial intelligence (AI), particularly deep learning, offers an opportunity to characterize structural–functional relationships by learning features directly from retinal images. These approaches extend OCT analysis beyond predefined thickness measurements. Previous studies have used deep learning to predict contrast sensitivity and other measures of retinal sensitivity from OCT images[12,13]. Pretrained retinal foundation models now provide transferable image representations, enabling investigation of their functional relevance without training a new image encoder for each task[14].

However, predicting contrast sensitivity across individuals leaves an important question unresolved: do the image representations capture functional differences specific to each eye? A between-person association can reflect both eye-specific variation and characteristics shared by the participant, which retinal images can also encode[15]. Consequently, successful prediction across people does not establish that differences between their eyes are represented. Comparing fellow eyes provides a direct test of this question because age, genetic background and systemic exposures are shared within each pair. If inter-eye OCT differences predict inter-eye contrast sensitivity differences in held-out participants, the association is less readily explained by shared participant characteristics alone. Although within-person structural–functional associations have been investigated using conventional retinal thickness[16], it remains

uncertain whether pretrained OCT representations capture such differences and whether their associations persist after accounting for inter-eye visual acuity differences.

Here, we used bilateral OCT and visual-function measurements from AI-READI to test whether differences in pretrained OCT representations predict photopic and mesopic contrast sensitivity differences between fellow eyes[17,18]. Models were developed in the training partition and applied unchanged to two nonoverlapping held-out participant partitions. We evaluated structural–functional correspondence using eye-swap and outcome-permutation controls, assessed associations after adjustment for inter-eye visual acuity differences, and compared the representations with conventional retinal thickness features. The study thus tests whether pretrained OCT representations capture eye-specific functional information and examines how that information relates to established structural measurements.

Materials and Methods

Data source and study design

We conducted a cross-sectional secondary analysis of the AI-READI dataset, version 3.0.0. The parent study collected standardized retinal imaging and visual-function measurements across four metabolic categories: healthy, prediabetes/lifestyle-controlled diabetes, diabetes treated without insulin, and diabetes treated with insulin[17,18]. Participants were eligible for the primary analysis if both eyes had available RETFound OCT features and complete photopic and mesopic CS and VA measurements. The resulting cohort included 2,015 participants. The dataset's predefined participant-level partitions were retained: training (n = 1,412), validation (n = 307), and test (n = 296). Both eyes and all repeated acquisitions from a participant remained in the same partition. Validation and test provided separate internal evaluations; neither partition was used to fit the primary models or select their hyperparameters. Eligibility was based on data availability rather than restriction to participants with normal VA or no reported ocular disease. CS was measured monocularly using the Mars Letter Contrast Sensitivity Test and expressed in log contrast-sensitivity units[18]. The recorded, condition-specific CS and logMAR values were used in this analysis.

AI feature extraction from OCT images

OCT features were extracted from Topcon Maestro2 3D macula structural OCT volumes using the pretrained OCT RETFound masked-autoencoder model, based on a Vision Transformer Large architecture with 16 × 16-pixel patches and a 1,024-dimensional embedding[14]. The encoder was held fixed throughout the analysis and was not fine-tuned using CS, VA, or other participant labels. Five contiguous central B-scans were selected from each acquisition, corresponding to zero-based frame indices 62–66. Each grayscale B-scan was converted to three identical RGB channels and resized to 224 × 224 pixels. Pixel values were converted to floating point and divided by 255, after which each channel was standardized using its own image-level mean and standard deviation, with 10^{-6} added to the denominator. The five B-scan embeddings were averaged component-wise to obtain one acquisition-level vector; vectors from repeated acquisitions were then averaged component-wise to obtain one 1,024-dimensional vector per eye. For participant p, the right- and left-eye vectors were denoted $Z_{p,R}$ and $Z_{p,L}$, respectively. Each vector contained 1,024 numerical features and was retained in full for downstream modeling.

Inter-eye prediction and comparative analyses

Inter-eye differences were calculated as right minus left. For participant p, the OCT feature-difference vector was defined as:

$$\mathbf{\Delta Z}_p = \mathbf{Z}_{p,R} - \mathbf{Z}_{p,L}$$

CS and VA differences were calculated in the same direction. Separate ridge-regression models were fitted directly to predict photopic and mesopic ΔCS from the 1,024 OCT feature differences. Each feature difference was standardized before regression:

$$x_{p,j} = \frac{\Delta Z_{p,j} - \mu_j}{s_j}$$

Here, μ_j and s_j denote the training-data mean and standard deviation of feature difference j. The predicted inter-eye CS difference was:

$$\Delta CS_p = \beta_0 + \sum_{j=1}^{1024} \beta_j x_{p,j}$$

Thus, the regression combined the 1,024 standardized feature differences into one predicted ΔCS per participant and lighting condition. The coefficients β_j and intercept β_0 were estimated from training participants, with an L2 penalty on the coefficients.

Two controls were evaluated without model refitting. Random within-pair eye swaps independently reversed each participant's OCT difference vector with probability 0.5, leaving observed ΔCS unchanged. Outcome permutations reassigned observed ΔCS across participants within each evaluation partition while retaining the original predictions. VA-adjusted associations were calculated in pooled validation and test participants by correlating residuals of predicted and observed ΔCS after separate linear adjustment for condition-specific ΔVA and evaluation partition.

Retinal thickness was derived from differences between seven device-provided segmentation surfaces in Topcon Maestro2 macular OCTA scans, converted to micrometers using DICOM axial spacing. Acquisitions required finite, correctly ordered boundaries at ≥90% of sampled locations; repeats were summarized by the per-eye median. Six elementary thickness intervals and six 0.5-mm annuli (0–3 mm) yielded 441 fovea-centered features, with regional means weighted by valid sampled locations. The foveal center was defined by the minimum smoothed total retinal thickness within 1 mm of scan center. Single measures were selected by the strongest absolute VA-adjusted training-set partial correlation with ΔCS. Layer subsets were selected from all 63 nonempty interval combinations within 0.5–2.0 mm using five-fold training cross-validation including ΔVA. Selected measures and subsets were fitted as thickness-only linear models and evaluated using held-out VA-adjusted associations. Full-feature ridge models using thickness, RETFound, or combined features used identical complete-case training (n = 1,380), validation (n = 303), and test (n = 291) sets.

Statistical analysis

Characteristics were summarized as mean (SD) or n (%). Analyses used complete cases without imputation. Prediction performance was assessed by Pearson r and predictive R^2, calculated as $1 - \Sigma(y - \hat{y})^2/\Sigma(y - \bar{y})^2$ using the evaluation-partition mean as the reference. Predictive R^2 was distinct from squared Pearson r. Percentile 95% confidence intervals used 2,500 participant-bootstrap resamples for primary correlations, 5,000 for adjusted AI associations, and 4,000 for adjusted thickness associations. Pooled bootstraps were stratified by evaluation partition, with adjustment regressions refitted in each resample. Pearson tests were two-sided. Control analyses used 5,000 randomizations and one-sided empirical P values calculated as (1 + the number of randomized correlations ≥ the observed correlation)/5,001. The reported adjusted permutation procedure shuffled raw ΔCS within partition and recalculated the

partial correlation. Thickness partial-correlation tests were two-sided, with n − 4 degrees of freedom. P values were unadjusted for multiple comparisons. Analyses used Python 3.12.3, NumPy 2.5.2, pandas 3.0.5, SciPy 1.18.1, and scikit-learn 1.9.1. OCT feature extraction used PyTorch 2.7.1, timm 0.9.12, pydicom 3.0.2, and Pillow 12.3.0. Numerical plots were generated with Matplotlib 3.11.2; scatterplots included all observations without jitter or outlier removal.

Results

Paired-eye analysis and participant characteristics

To examine whether AI-derived OCT features capture eye-specific information related to contrast sensitivity, we compared structural and functional differences between fellow eyes. Each participant served as their own control, reducing confounding by shared characteristics such as age, genetic background, and systemic metabolic status. Two complementary controls assessed the specificity of the associations: randomly swapping OCT features between fellow eyes disrupted structural–functional eye matching, whereas permuting inter-eye contrast-sensitivity differences across participants disrupted the correspondence between imaging features and functional outcomes. Models were developed exclusively in the training partition and applied unchanged to the dataset's predefined validation and test partitions, providing two separate internal evaluations in nonoverlapping participant groups (Figure 1).

The primary analysis included 2,015 participants with paired-eye OCT, contrast sensitivity (CS), and visual acuity (VA) data: 1,412 in training, 307 in validation, and 296 in test (Table 1). Mean age was 60.6 years (SD, 11.2), and mean HbA1c was 6.1% (SD, 1.1) among the 1,960 participants with available measurements. The cohort included participants classified as healthy (34.8%), prediabetes/lifestyle controlled (24.7%), diabetes treated without insulin (30.1%), and diabetes treated with insulin (10.5%). Mean age was similar across partitions, although insulin-treated diabetes was less frequent in training (6.6%) than in validation (17.6%) and test (21.6%).

OCT-derived AI features predict inter-eye contrast-sensitivity differences in held-out participants

To test whether AI-derived OCT features capture eye-specific information related to contrast sensitivity, we used inter-eye feature differences to predict inter-eye CS differences in

participants excluded from model development. Separate photopic and mesopic models were trained in the training partition and applied unchanged to both evaluation partitions. Predicted photopic ΔCS correlated positively with observed ΔCS in validation (r = 0.334; 95% CI, 0.138–0.497; P = 1.85×10^{-9}) and test (r = 0.318; 95% CI, 0.166–0.452; P = 2.32×10^{-8}; Figure 2A). Positive associations were also observed for mesopic CS in validation (r = 0.373; 95% CI, 0.181–0.553; P = 1.35×10^{-11}) and test (r = 0.194; 95% CI, 0.036–0.358; P = 0.0008; Figure 2B). Photopic correlations were similar across the two held-out groups, whereas the mesopic correlation was lower in test.

To assess whether these associations depended on correct structural–functional eye matching, we compared the observed correlations with those obtained after randomly swapping OCT features between fellow eyes. Observed correlations exceeded the central 95% range of the eye-swap null distributions in all four evaluations (Figure 2). One-sided empirical P values were 0.0008 and 0.0002 for photopic validation and test, respectively, and 0.0002 and 0.0038 for mesopic validation and test. Permuting observed ΔCS across participants provided a complementary control by disrupting the correspondence between imaging features and functional outcomes. Observed correlations also exceeded those expected under outcome permutation, with empirical P = 0.0002 for both photopic evaluations and mesopic validation, and P = 0.0008 for mesopic test.

Photopic predictions achieved predictive R^2 values of 0.095 in validation and 0.101 in test, indicating similar prediction performance across partitions (Figure 3A,B; Table 2). Mesopic predictive R^2 was 0.133 in validation and 0.027 in test, consistent with the lower test correlation (Figure 3C,D; Table 2). Together, these findings support the presence of eye-specific information in AI-derived OCT features related to inter-eye CS differences, although the proportion of variation captured was modest and mesopic performance varied across evaluation groups.

Associations persist after adjustment for inter-eye visual-acuity differences

The paired-eye design reduces confounding by shared participant characteristics, but does not account for differences in visual acuity between fellow eyes. Because structural changes may relate to both acuity and contrast sensitivity, the OCT-based predictions could reflect functional variation shared with VA. To examine whether the associations extended beyond measured

inter-eye acuity differences, we assessed the relationship between predicted and observed ΔCS after adjustment for condition-specific ΔVA and evaluation partition.

Among the 603 pooled validation and test participants, predicted ΔCS remained positively associated with observed ΔCS under photopic conditions (partial r = 0.272; 95% CI, 0.174–0.364) and mesopic conditions (partial r = 0.171; 95% CI, 0.069–0.264; Figure 4A). Both bootstrap confidence intervals excluded zero, and one-sided permutation tests yielded P = 0.0002 for each condition. The adjusted association was numerically larger for photopic than mesopic CS. These findings indicate that measured inter-eye VA differences did not fully account for the relationship between AI-derived OCT features and inter-eye contrast sensitivity.

Comparison of AI-derived OCT features and conventional retinal-thickness measures for predicting inter-eye contrast-sensitivity differences

To assess how the AI-based findings compared with conventional measures of retinal structure, we examined associations using training-selected thickness measures and compared held-out prediction performance using thickness features, RETFound features, and their combination. Anatomical feature selection and model development were confined to the training partition.

Among 594 held-out participants with the required thickness measurements, predictions from the selected single thickness measure were positively associated with photopic ΔCS after adjustment for ΔVA and evaluation partition (partial r = 0.097; 95% CI, 0.009–0.193; P = 0.019). The corresponding layer-subset model yielded a partial correlation of 0.106 (95% CI, 0.014–0.228; P = 0.010). Mesopic associations were smaller, with partial correlations of 0.084 for the single thickness measure (95% CI, 0.002–0.165; P = 0.046) and 0.080 for the layer-subset model (95% CI, 0.006–0.168; P = 0.052; Figure 4B).

For a comparison on identical participants, models using the full thickness feature set, RETFound features, or their combination were trained in the complete-case training cohort (n = 1,380) and evaluated in validation (n = 303) and test (n = 291). Photopic validation correlations were 0.255 for thickness, 0.336 for AI, and 0.347 for combined features; corresponding test correlations were 0.023, 0.298, and 0.212 (Figure 4C). AI features therefore yielded numerically higher photopic correlations than thickness features in both partitions. Combining the feature sets produced a slightly higher validation correlation than AI alone, but a lower test correlation.

For mesopic CS, validation correlations were 0.241 for thickness, 0.322 for AI, and 0.342 for combined features. Corresponding test correlations were 0.126, 0.126, and 0.146 (Figure 4D). Although AI and combined models had numerically higher correlations than thickness alone in validation, all three models showed lower and similar correlations in test. These exploratory comparisons identified modest associations using conventional thickness measures and numerically stronger photopic prediction using AI features, but did not show a consistent improvement from combining thickness and AI features across evaluation groups.

Discussion

Understanding the relationship between retinal structure and visual function is essential for interpreting information extracted from OCT images. For AI-derived representations, an important question is whether their predictive information corresponds to functional differences between eyes within the same individual. In this study, inter-eye differences in pretrained RETFound OCT features predicted inter-eye contrast-sensitivity differences in held-out participants. Photopic performance was similar across validation and test, whereas mesopic performance was lower in test. The associations exceeded those expected under eye-swap and outcome-permutation controls and remained positive after adjustment for inter-eye visual-acuity differences. These findings support the functional relevance of AI-derived OCT representations at the level of individual eyes.

A strength of this study is the within-person comparison of structural and functional differences. Between-person prediction can reflect characteristics that influence both retinal appearance and visual performance, including age and systemic metabolic status. Comparing fellow eyes reduces the contribution of characteristics shared by the participant and focuses the analysis on variation between eyes. The eye-swap control further tested whether prediction depended on the correspondence between structural and functional eye differences, while outcome permutation tested the association after disrupting participant-level feature–outcome matching. These complementary controls strengthen the evidence that the observed associations depended on the original pairing of imaging features and functional measurements. Training-only model development and unchanged application to two nonoverlapping evaluation groups provided additional evidence of consistency within the cohort, particularly for photopic CS.

The persistence of the associations after VA adjustment suggests that measured acuity differences did not fully account for the relationship between OCT features and CS. This distinction is relevant because acuity and contrast sensitivity characterize related but different aspects of visual performance. Conventional thickness measures also showed modest adjusted associations, while AI features yielded numerically higher photopic prediction correlations in both evaluation groups. Combining thickness and AI features did not consistently improve performance, leaving their complementarity uncertain. Previous studies have reported associations between retinal thickness and CS and have used deep learning to predict CS from OCT images[6,12]. Retinal foundation models provide pretrained representations that can be transferred to downstream tasks[14]. The present study extends this work by demonstrating that differences in these representations are associated with functional differences between fellow eyes in held-out participants. The scientific contribution is therefore evidence that a representation learned without CS labels retains information relevant to within-person variation in visual function. Anatomical characterization of that information could help connect image-based prediction with interpretable structural–functional relationships.

Several limitations should be considered. The cross-sectional design cannot establish temporal or causal relationships, and both evaluation partitions came from the same cohort; external generalizability remains untested. Predictive R^2 was modest, particularly for mesopic CS in test, and the findings do not establish sufficient accuracy to substitute for direct functional testing. Inter-eye differences also incorporate measurement error, and the paired-eye design does not eliminate confounding by asymmetric lens opacity, refraction, image quality, or other ocular factors. VA adjustment cannot resolve all such influences. The encoder sampled five central B-scans rather than the full OCT volume, and the conventional thickness measures came from a different acquisition protocol, limiting interpretation of the feature comparisons. Thickness analyses were exploratory, and numerical performance differences do not establish statistical superiority. Future studies should evaluate external cohorts, incorporate optical and image-quality measures, assess repeatability, and determine which retinal characteristics contribute to the predictive signal. Longitudinal evaluation would further establish whether these features track changes in contrast sensitivity over time.

Conclusion

Inter-eye differences in pretrained AI-derived OCT features predicted contrast-sensitivity differences between fellow eyes in held-out participants, with more consistent performance under photopic conditions. Associations remained after adjustment for measured inter-eye visual-acuity differences, supporting the functional relevance of these representations. External validation and anatomical characterization are needed to establish their broader applicability and potential clinical value.

Statements and Declarations

Funding
This research received no external funding.

Competing Interests
The author declares no competing interests.

Ethical approval
The present study was approved by the Institutional Review Board of Wayne State University.

Consent to Participate
Not applicable.

Author Contributions
R.Z. conceptualized, designed the study, analyzed the data, and wrote the paper.

Ethical Responsibilities of Authors
The author has read, understood, and complied, as applicable, with the statement on Ethical Responsibilities of Authors in the Instructions for Authors.

Data Availability Statement
The data analyzed in this study were obtained from the AI-READI dataset (version 3.0.0), available through FAIRhub at https://doi.org/10.60775/fairhub.3. Access and reuse are subject to the dataset license and applicable access requirements. The pretrained RETFound model and associated code used for OCT feature extraction are available at https://github.com/rmaphoh/RETFound_MAE.

References

1. Owsley C, Sloane ME. Contrast sensitivity, acuity, and the perception of'real-world'targets. *Br J Ophthalmol*. 1987;71(10):791-796.
2. Xiong Y-Z, Kwon M, Bittner AK, Virgili G, Giacomelli G, Legge GE. Relationship between acuity and contrast sensitivity: differences due to eye disease. *Invest Ophthalmol Vis Sci*. 2020;61(6):40-40.
3. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science*. 1991;254(5035):1178-1181.
4. Puell MC, Palomo-Álvarez C, Pérez-Carrasco MJ. Macular inner retinal layer thickness in relation to photopic and mesopic contrast sensitivity in healthy young and older subjects. *Invest Ophthalmol Vis Sci*. 2018;59(13):5487-5493.
5. Fatehi N, Nowroozizadeh S, Henry S, Coleman AL, Caprioli J, Nouri-Mahdavi K. Association of structural and functional measures with contrast sensitivity in glaucoma. *Am J Ophthalmol*. 2017;178:129-139.
6. McAnany JJ, Park JC, Liu K, et al. Contrast sensitivity is associated with outer-retina thickness in early-stage diabetic retinopathy. *Acta Ophthalmol (Copenh)*. 2020;98(2):e224-e231.
7. Joltikov KA, de Castro VM, Davila JR, et al. Multidimensional functional and structural evaluation reveals neuroretinal impairment in early diabetic retinopathy. *Invest Ophthalmol Vis Sci*. 2017;58(6):BIO277-BIO290.
8. Srinivasan S, Sivaprasad S, Rajalakshmi R, et al. Retinal structure–function correlation in type 2 diabetes. *Eye*. 2022;36(10):1865-1871.
9. Ha SK, Ding X, Romano F, et al. Structure-function associations between quantitative contrast sensitivity function and peripapillary optical coherence tomography angiography in diabetic retinopathy. *Invest Ophthalmol Vis Sci*. 2025;66(4):69-69.
10. Sohn EH, van Dijk HW, Jiao C, et al. Retinal neurodegeneration may precede microvascular changes characteristic of diabetic retinopathy in diabetes mellitus. *Proceedings of the National Academy of Sciences*. 2016;113(19):E2655-E2664.
11. Montesano G, Ometto G, Higgins BE, et al. Evidence for structural and functional damage of the inner retina in diabetes with no diabetic retinopathy. *Invest Ophthalmol Vis Sci*. 2021;62(3):35-35.
12. Shamsi F, Liu R, Owsley C, Kwon M. Identifying the retinal layers linked to human contrast sensitivity via deep learning. *Invest Ophthalmol Vis Sci*. 2022;63(2):27-27.
13. Kihara Y, Heeren TF, 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(2):e188029.
14. Zhou Y, Chia MA, Wagner SK, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622(7981):156-163.
15. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature biomedical engineering*. 2018;2(3):158-164.

16. Chien L, Liu R, Girkin C, Kwon M. Higher contrast requirement for letter recognition and macular RGC+ layer thinning in glaucoma patients and older adults. *Invest Ophthalmol Vis Sci*. 2017;58(14):6221-6231.
17. AI-READI-Consortium. AI-READI: rethinking AI data collection, preparation and sharing in diabetes research and beyond. *Nature metabolism*. 2024;6(12):2210-2212.
18. Owsley C, Matthies DS, McGwin G, et al. Cross-sectional design and protocol for artificial intelligence ready and equitable atlas for diabetes insights (AI-READI). *BMJ open*. 2025;15(2):e097449.

Figure legends

Figure 1. Paired-eye study design and AI-based OCT analysis. (1) Optical coherence tomography (OCT), contrast sensitivity (CS), and visual acuity (VA) were assessed for each eye. Within-person comparisons reduce confounding by shared participant-level characteristics. (2) The fixed pretrained RETFound encoder processed five central OCT B-scans per acquisition. Embeddings were averaged across B-scans and repeated acquisitions to obtain a 1,024-dimensional feature vector, Z, for each eye. (3) Inter-eye differences were calculated consistently as right minus left: ΔOCT = Z(right) − Z(left), ΔCS = CS(right) − CS(left), and ΔVA = VA(right) − VA(left). (4) AI-READI's predefined participant-level partitions were retained: training (n = 1,412), validation (n = 307), and test (n = 296). Separate photopic and mesopic ridge models were tuned by five-fold cross-validation within training and applied unchanged to both evaluation partitions. Validation and test served as separate internal evaluations to assess consistency across nonoverlapping participant groups. (5) Analyses compared predicted and observed ΔCS, included structural eye-swap and outcome-permutation controls, and assessed associations after ΔVA adjustment. The scatterplot is schematic; r = ? denotes the association to be estimated.

Figure 2. Eye-specific OCT features track inter-eye contrast-sensitivity differences in held-out participants. Associations between AI-predicted and observed inter-eye contrast-sensitivity differences are shown for (A) photopic and (B) mesopic conditions. Models were trained only in the training set and applied unchanged to the validation (n = 307) and test (n = 296) sets. Points indicate Pearson correlation coefficients with 95% participant-bootstrap confidence intervals. Gray boxes show the central 95% range of the random within-pair eye-swap null distribution, and the dashed vertical line indicates r = 0. P values are from two-sided Pearson correlation tests.

Figure 3. Observed and AI-predicted inter-eye contrast-sensitivity differences. Scatterplots show observed versus predicted ΔCS for photopic validation (A), photopic test (B), mesopic validation (C), and mesopic test (D) participants. Each point represents one participant. Models were trained only in the training set and applied unchanged to validation (n = 307) and test (n = 296) participants. Solid lines show the fitted linear relationship and gray bands indicate 95% confidence intervals. Pearson r, predictive R^2, and two-sided P values are shown in each panel. All observations are displayed without jitter or outlier removal.

Figure 4. AI-derived OCT features and conventional retinal thickness capture inter-eye contrast-sensitivity differences. (A) Partial correlations between AI-predicted and observed inter-eye CS differences after adjustment for inter-eye visual-acuity difference and evaluation split (n = 603). Points indicate partial correlations with 95% bootstrap confidence intervals; P values are from permutation tests. (B) Corresponding associations for training-selected conventional retinal-thickness measures and layer subsets (n = 594). P values are from two-sided partial-correlation tests. (C, D) Held-out prediction performance for conventional thickness features, RETFound AI features, and their combination for photopic (C) and mesopic (D) CS. Models were evaluated on identical validation (n = 303) and test (n = 291) participants. Bars show Pearson correlations; solid bars indicate validation and hatched bars indicate test results. Anatomical feature selection was confined to the training set and these analyses were exploratory.

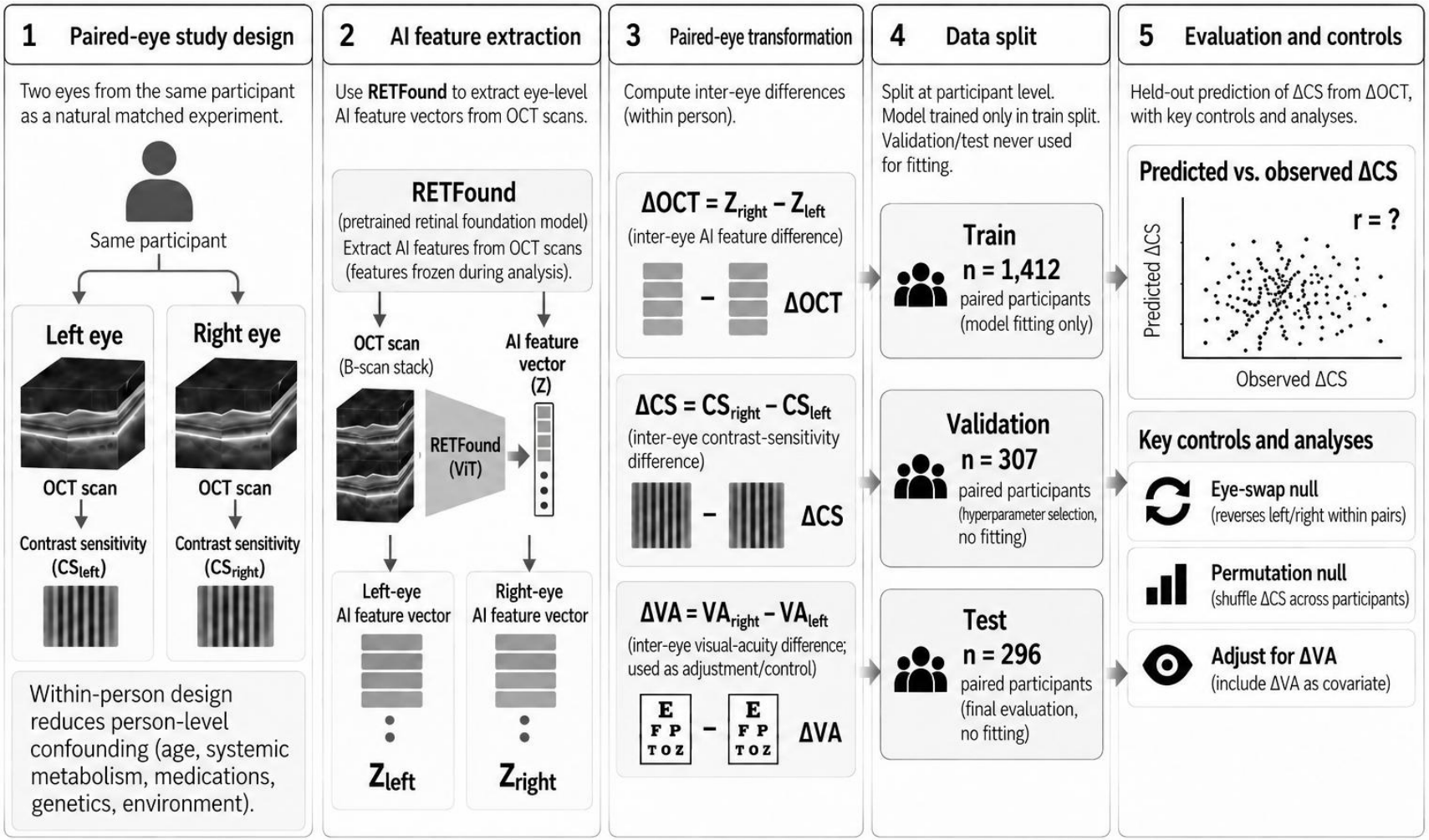

1 Paired-eye study design
Two eyes from the same participant as a natural matched experiment.
Same participant
Left eye
Right eye
OCT scan
OCT scan
Contrast sensitivity (CS_{left})
Contrast sensitivity (CS_{right})
Within-person design reduces person-level confounding (age, systemic metabolism, medications, genetics, environment).
2 AI feature extraction
Use RETFound to extract eye-level AI feature vectors from OCT scans.
RETFound
(pretrained retinal foundation model)
Extract AI features from OCT scans (features frozen during analysis).
OCT scan (B-scan stack)
AI feature vector (Z)
RETFound (ViT)
Left-eye AI feature vector
Right-eye AI feature vector
Z_{left}
Z_{right}
3 Paired-eye transformation
Compute inter-eye differences (within person).
$\Delta OCT = Z_{right} - Z_{left}$
(inter-eye AI feature difference)
ΔOCT
$\Delta CS = CS_{right} - CS_{left}$
(inter-eye contrast-sensitivity difference)
ΔCS
$\Delta VA = VA_{right} - VA_{left}$
(inter-eye visual-acuity difference; used as adjustment/control)
ΔVA
4 Data split
Split at participant level.
Model trained only in train split.
Validation/test never used for fitting.
Train
n = 1,412
paired participants
(model fitting only)
Validation
n = 307
paired participants
(hyperparameter selection, no fitting)
Test
n = 296
paired participants
(final evaluation, no fitting)
5 Evaluation and controls
Held-out prediction of ΔCS from ΔOCT, with key controls and analyses.
Predicted vs. observed ΔCS
r = ?
Predicted ΔCS
Observed ΔCS
Key controls and analyses
Eye-swap null
(reverses left/right within pairs)
Permutation null
(shuffle ΔCS across participants)
Adjust for ΔVA
(include ΔVA as covariate)

Figure 1

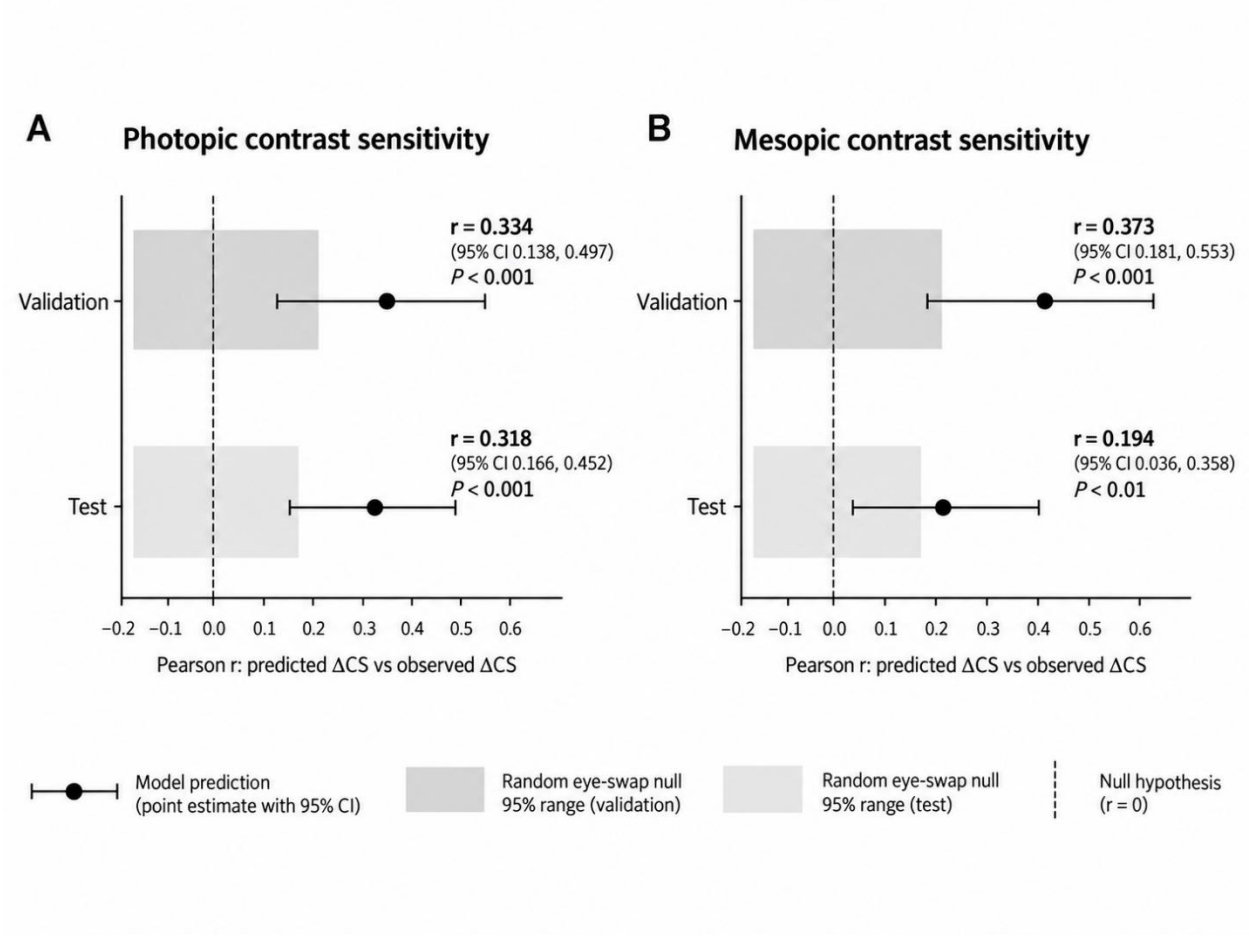

A
Photopic contrast sensitivity
r = 0.334
(95% CI 0.138, 0.497)
P < 0.001
Validation
r = 0.318
(95% CI 0.166, 0.452)
P < 0.001
Test
−0.2 −0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6
Pearson r: predicted ΔCS vs observed ΔCS
B
Mesopic contrast sensitivity
r = 0.373
(95% CI 0.181, 0.553)
P < 0.001
Validation
r = 0.194
(95% CI 0.036, 0.358)
P < 0.01
Test
−0.2 −0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6
Pearson r: predicted ΔCS vs observed ΔCS
Model prediction
(point estimate with 95% CI)
Random eye-swap null
95% range (validation)
Random eye-swap null
95% range (test)
Null hypothesis
(r = 0)

Figure 2

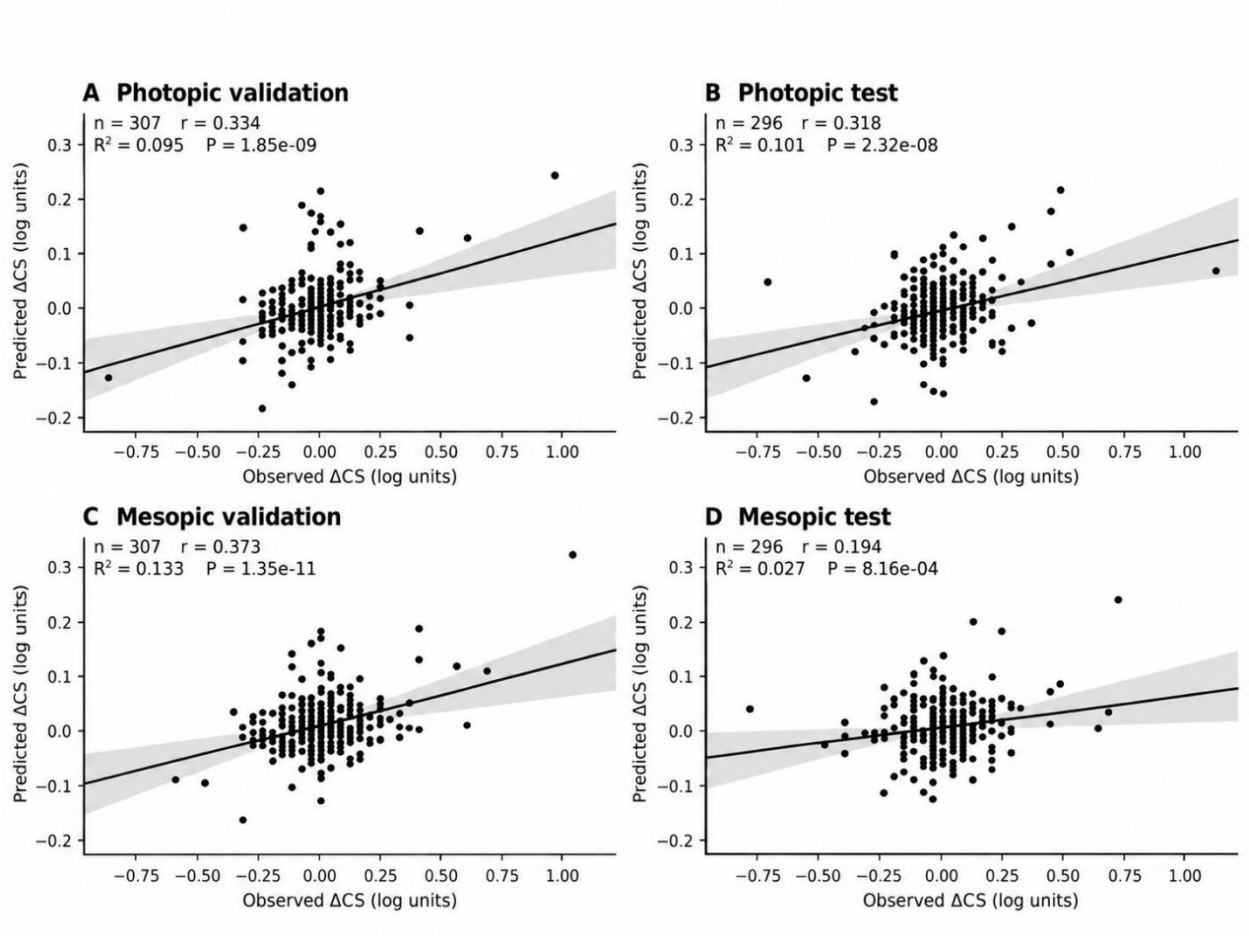

A Photopic validation
n = 307 r = 0.334
R² = 0.095 P = 1.85e-09
B Photopic test
n = 296 r = 0.318
R² = 0.101 P = 2.32e-08
C Mesopic validation
n = 307 r = 0.373
R² = 0.133 P = 1.35e-11
D Mesopic test
n = 296 r = 0.194
R² = 0.027 P = 8.16e-04
Predicted ΔCS (log units)
Observed ΔCS (log units)
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

Figure 3

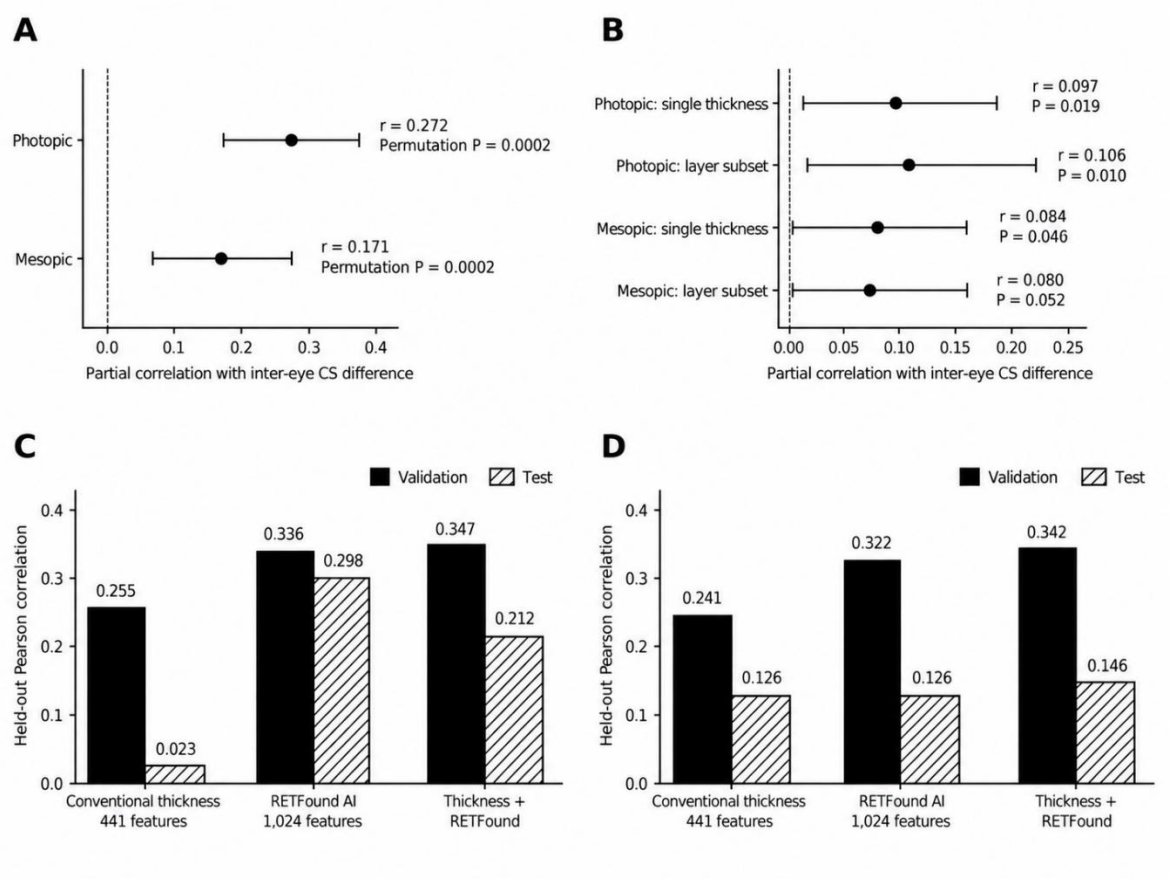

A
Photopic
r = 0.272
Permutation P = 0.0002
Mesopic
r = 0.171
Permutation P = 0.0002
0.0 0.1 0.2 0.3 0.4
Partial correlation with inter-eye CS difference
B
Photopic: single thickness
r = 0.097
P = 0.019
Photopic: layer subset
r = 0.106
P = 0.010
Mesopic: single thickness
r = 0.084
P = 0.046
Mesopic: layer subset
r = 0.080
P = 0.052
0.00 0.05 0.10 0.15 0.20 0.25
Partial correlation with inter-eye CS difference
C
Validation
Test
Held-out Pearson correlation
0.255
0.023
0.336
0.298
0.347
0.212
Conventional thickness 441 features
RETFound AI 1,024 features
Thickness + RETFound
D
Validation
Test
Held-out Pearson correlation
0.241
0.126
0.322
0.126
0.342
0.146
Conventional thickness 441 features
RETFound AI 1,024 features
Thickness + RETFound

Figure 4

Table 1. Participant characteristics and inter-eye visual-function differences

Characteristic	Overall	Training	Validation	Test
Participants, n	2,015	1,412	307	296
Age, years	60.6 (11.2)	60.8 (11.3)	59.9 (10.7)	60.0 (11.3)
HbA1c, %	6.1 (1.1)	6.0 (1.0)	6.3 (1.2)	6.3 (1.2)
Healthy	701 (34.8)	545 (38.6)	77 (25.1)	79 (26.7)
Prediabetes/lifestyle controlled	497 (24.7)	342 (24.2)	79 (25.7)	76 (25.7)
Diabetes treated without insulin	606 (30.1)	432 (30.6)	97 (31.6)	77 (26.0)
Diabetes treated with insulin	211 (10.5)	93 (6.6)	54 (17.6)	64 (21.6)
Photopic ΔCS, log units	-0.010 (0.143)	-0.008 (0.143)	-0.020 (0.138)	-0.009 (0.148)
Photopic ΔVA, logMAR	0.002 (0.173)	0.003 (0.174)	-0.007 (0.192)	0.005 (0.146)
Mesopic ΔCS, log units	0.002 (0.152)	0.001 (0.149)	0.004 (0.162)	0.008 (0.154)
Mesopic ΔVA, logMAR	-0.010 (0.174)	-0.010 (0.172)	-0.016 (0.188)	-0.001 (0.168)

Values are n, n (%), or mean (SD). Δ denotes right minus left. Study-group labels follow the dataset categories. HbA1c was available for 1,960 participants (training, 1,375; validation, 297; test, 288); other listed measures were complete. Sex, race and ethnicity were unavailable in the source metadata. CS, contrast sensitivity; VA, visual acuity; logMAR, logarithm of the minimum angle of resolution.

Table 2. Primary held-out prediction of inter-eye contrast-sensitivity differences

CS condition	Partition	n	Pearson r (95% CI)	Predictive R^2	P value
Photopic	Validation	307	0.334 (0.138 to 0.497)	0.095	1.85e-09
Photopic	Test	296	0.318 (0.166 to 0.452)	0.101	2.32e-08
Mesopic	Validation	307	0.373 (0.181 to 0.553)	0.133	1.35e-11
Mesopic	Test	296	0.194 (0.036 to 0.358)	0.027	0.0008

Models were trained only in the training partition and applied unchanged to validation and test participants. Confidence intervals are participant-bootstrap intervals; P values are from two-sided Pearson tests. Predictive R^2 is calculated from prediction errors and is distinct from squared Pearson r. Values correspond to Figures 2 and 3. Acuity-adjusted associations and exploratory thickness comparisons are presented in Figure 4.