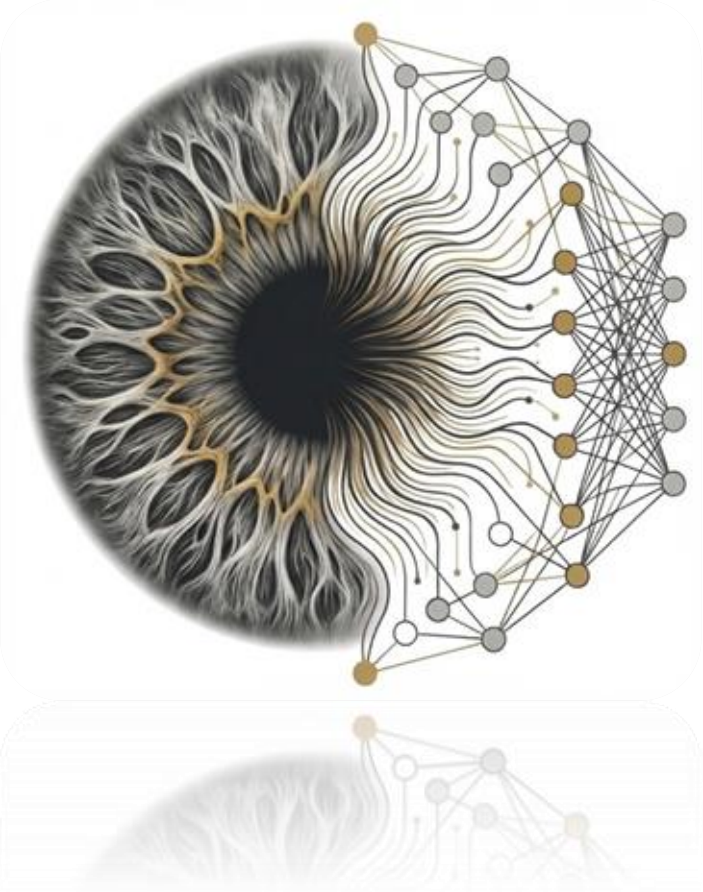
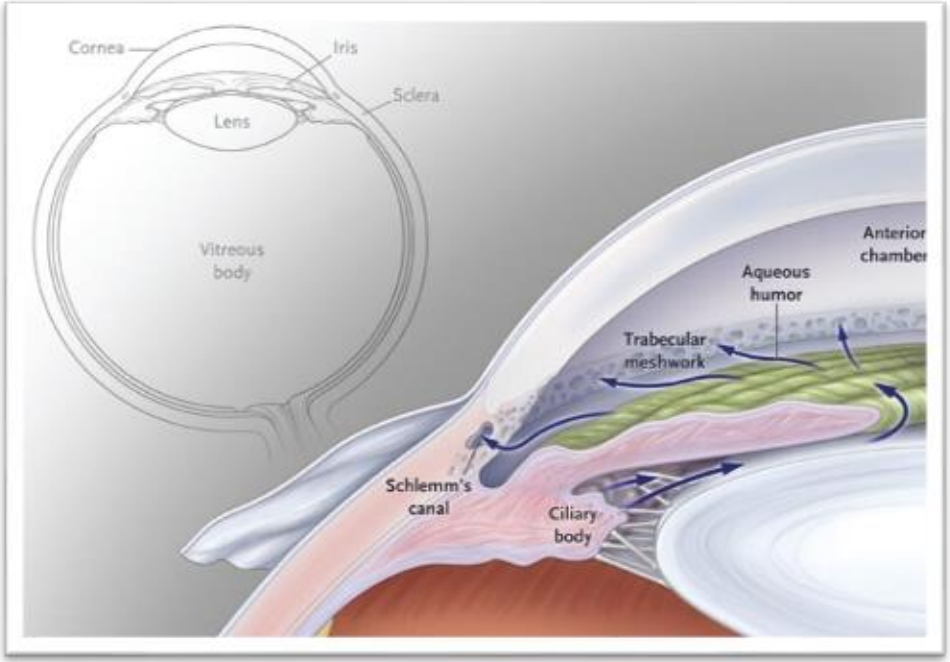
AI in Focus

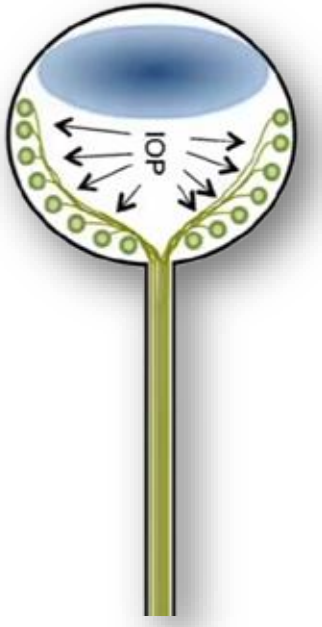
A New Era for Glaucoma Diagnosis

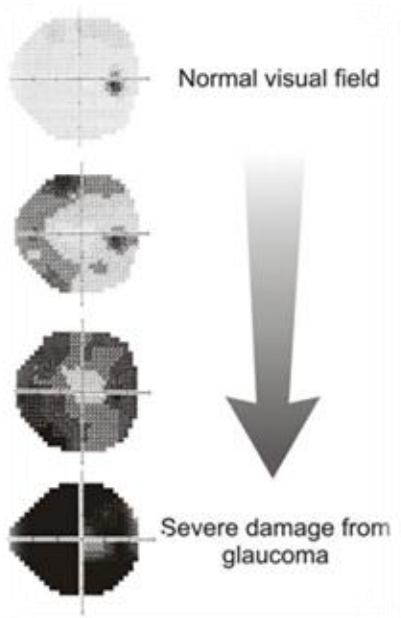


Glaucoma: Pathophysiology

Blocked aqueous humor flow → increase IOP → optic nerve cupping → **RGC death**

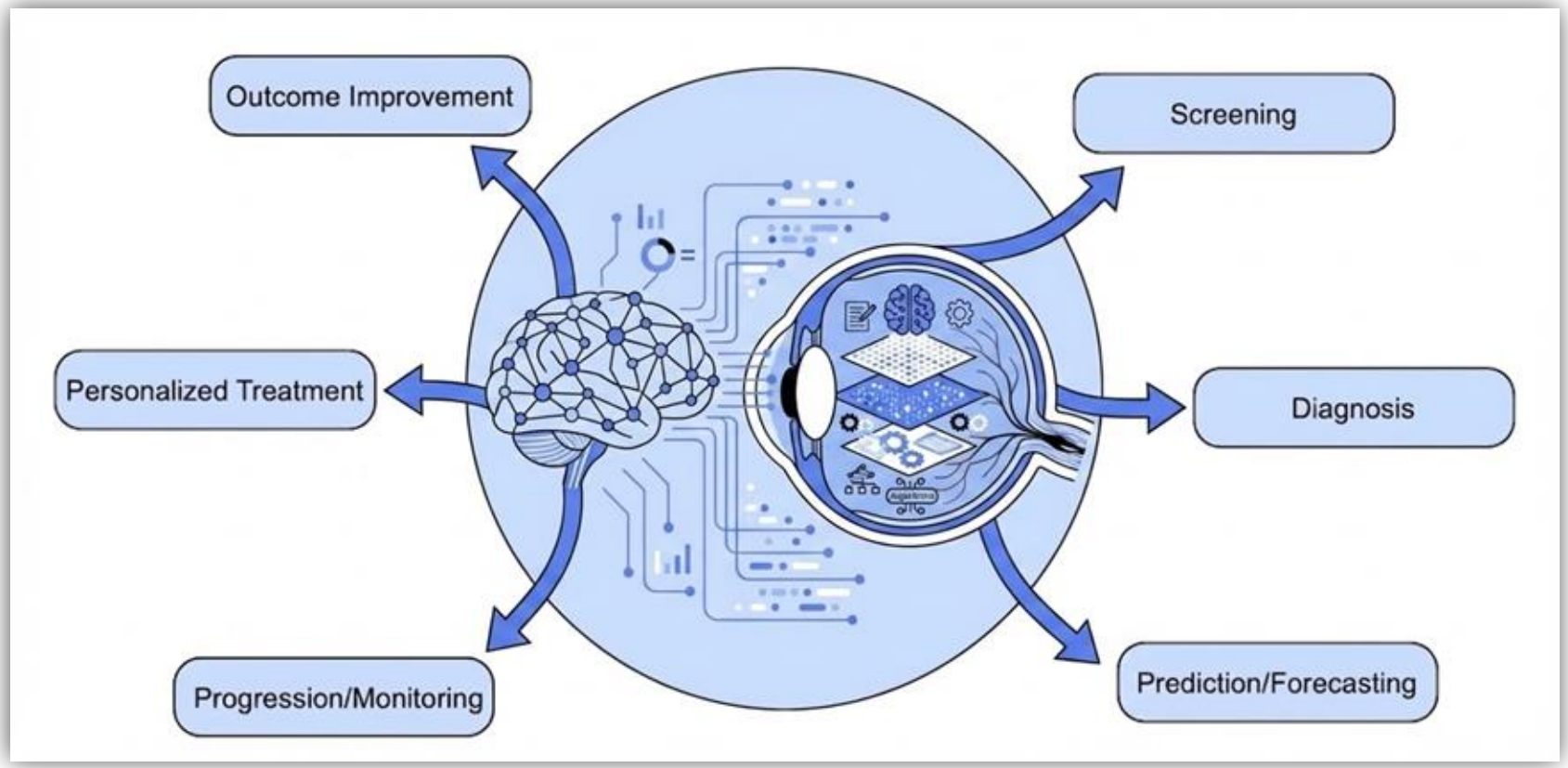






Lower IOP to reduce mechanical stress: eye drops, laser trabeculoplasty, shunts or filtering surgery

Artificial Intelligence in Glaucoma



Roadmap

Functional & Clinical Diagnosis
Glaucoma diagnosis began with functional visual-field tests and clinical exams.

Image-based Deep Learning (CNN)
Convolutional neural networks enabled automated detection/segmentation from retinal images.

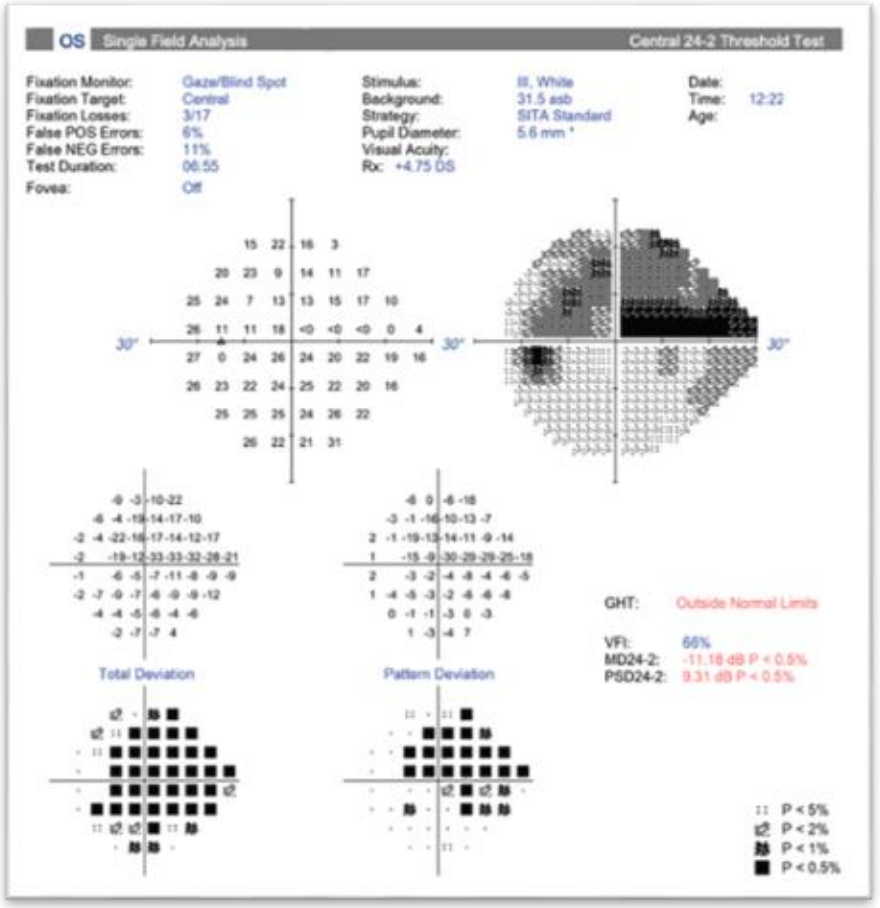
Multimodal Data Integration
Models began integrating imaging, visual field, and clinical EHR data for glaucoma assessment

Genetics & Risk Stratification
Detect progression risk early with genetic phenotyping and risk stratification

Foundation Models & Multimodal AI
Large datasets (millions) and foundation models will drive diagnosis and generalized ophthalmology care.

Visual fields map functional sensitivity across the retina

- Measure the light sensitivity threshold across central and peripheral retina
- Results plotted as maps
  - **Total Deviation map**: how each point differs from age-matched normals
  - **Pattern Deviation map**: adjusts for generalized sensitivity loss and highlights true localized defects

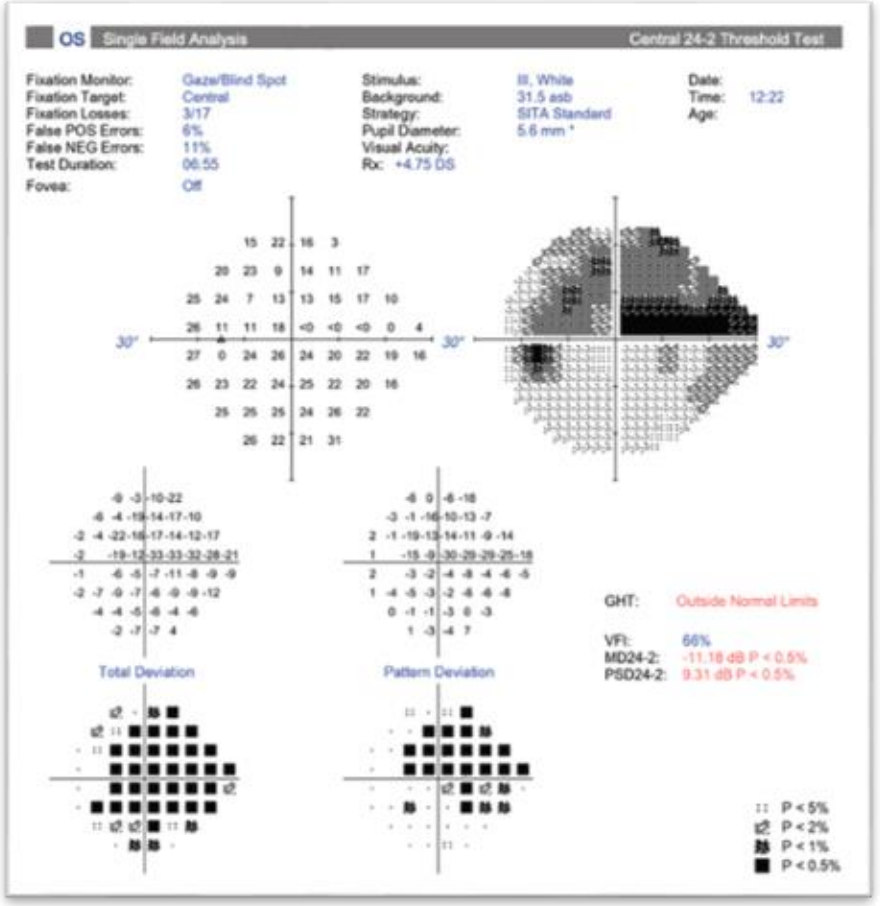


Visual fields map functional sensitivity across the retina

- Measure the light sensitivity threshold across central and peripheral retina
- Results plotted as maps
  - **Total Deviation map**: how each point differs from age-matched normals
  - **Pattern Deviation map**: adjusts for generalized sensitivity loss and highlights true localized defects

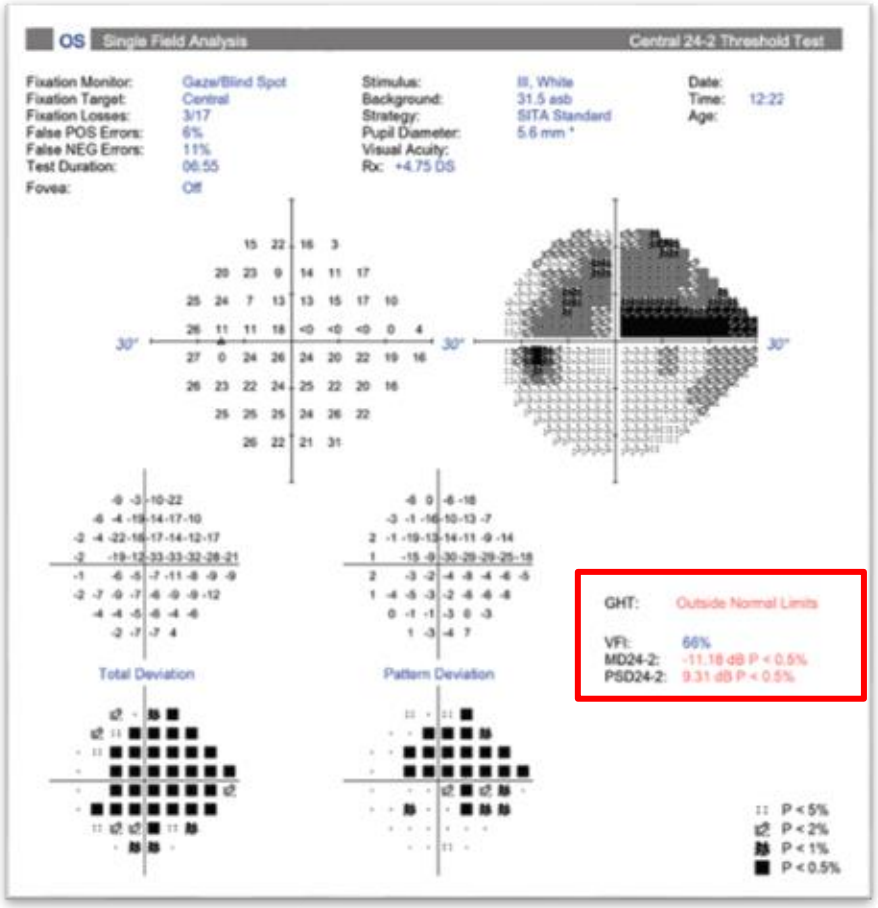
○ VF testing is inherently a noisy, high-variability measurement

How do we track progression?



Progression on visual fields is determined by longitudinal changes in global indices

- Key summary indices include VFI, MD, and PSD, which quantify severity and pattern of loss
- Global Performance
  - Track global indices (MD, VFI) over multiple tests to estimate **rate of progression (dB/year)**
    - A “fast” progression may be ~ −1.5 to −2.0 dB/year
  - Two or more test points within or adjacent to an existing scotoma worsen by ≥ 10 dB (or by ≥ 3× the average short-term variability), confirmed on two subsequent fields

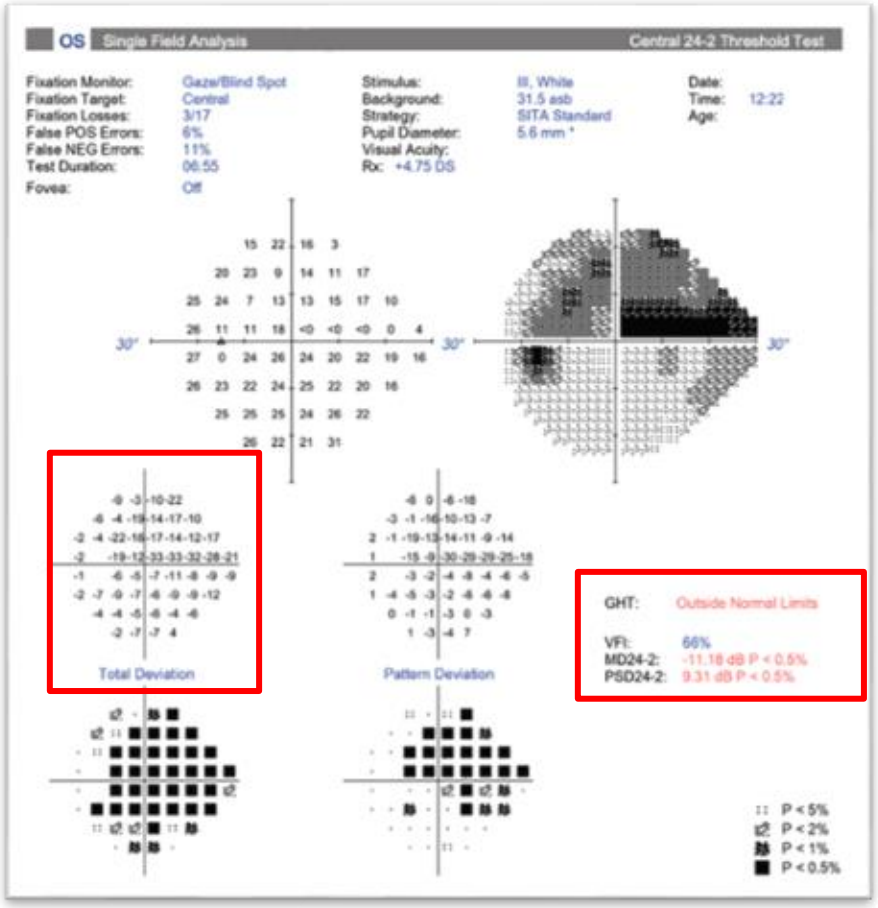


Progression on visual fields is determined by longitudinal changes in global indices

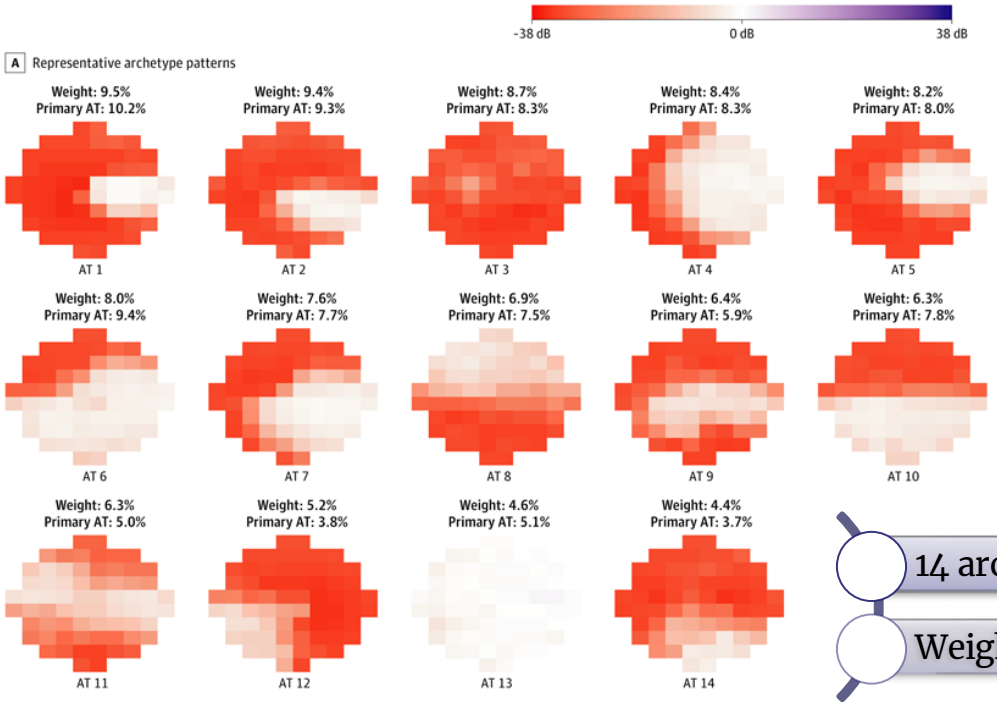
- Key summary indices include VFI, MD, and PSD, which quantify severity and pattern of loss.
- Global Performance
  - Track global indices (MD, VFI) over multiple tests to estimate **rate of progression (dB/year)**
    - A “fast” progression may be ~ −1.5 to −2.0 dB/year
  - Two or more test points within or adjacent to an existing scotoma worsen by ≥ 10 dB (or by ≥ 3× the average short-term variability), confirmed on two subsequent fields

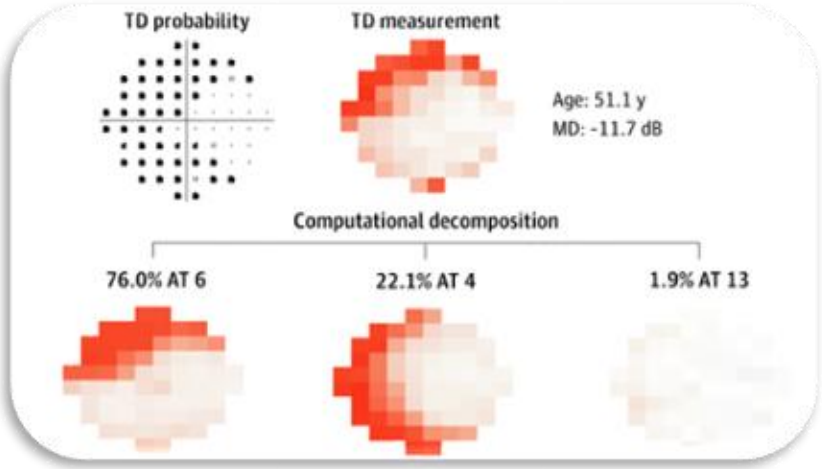
Are we losing information from global metrics?

Are different 2D patterns associated with progression?



Unsupervised learning reveals fundamental visual-field patterns, enabling standardized representation of glaucomatous loss.





- 14 archetypes (AT) that can decompose a visual field
- Weighted sum of AT reconstructs original field

Distinct archetypal patterns stratify progression risk, linking VF phenotypes to meaningful clinical outcomes.

JAMA Ophthalmology | Original Investigation

Characterization of Central Visual Field Loss in End-stage Glaucoma by Unsupervised Artificial Intelligence

Mengyu Wang, PhD; Jorryt Tichelaar, MS; Louis R. Pasquale, MD; Lucy Q. Shen, MD; Michael V. Boland, MD, PhD; Sarah R. Wellik, MD; Carlos Gustavo De Moraes, MD; Jonathan S. Myers, MD; Pradeep Ramulu, MD, PhD; MiYoung Kwon, PhD; Osamah J. Saeedi, MD; Hui Wang, PhD; Neda Baniasadi, MD, PhD; Dian Li, MS; Peter J. Bex, PhD; Tobias Elze, PhD

Invited Commentary page 196
Supplemental content

IMPORTANCE Although the central visual field (VF) in end-stage glaucoma may substantially vary among patients, structure-function studies and quality-of-life assessments are impeded by the lack of appropriate characterization of end-stage VF loss.

OBJECTIVE To provide a quantitative characterization and classification of central VF loss in end-stage glaucoma.

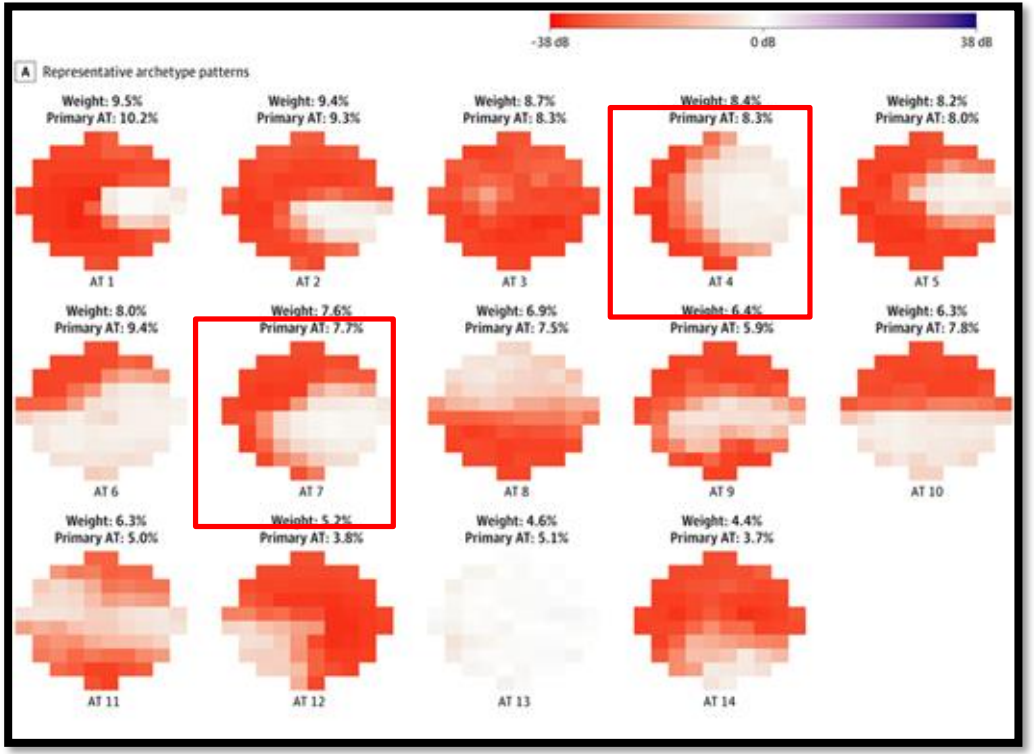
DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study collected data from 5 US glaucoma services from June 1, 1999, through October 1, 2014. A total of 2912 reliable 10-2 VFs of 1103 eyes from 1010 patients measured after end-stage 24-2 VFs with a mean deviation (MD) of −22 dB or less were included in the analysis. Data were analyzed from March 28, 2018, through May 23, 2019.

MAIN OUTCOMES AND MEASURES Central VF patterns were determined by an artificial intelligence algorithm termed *archetypal analysis*. Longitudinal analyses were performed to investigate whether the development of central VF defect mostly affects specific vulnerability zones.

RESULTS Among the 1103 patients with the most recent VFs, mean (SD) age was 70.4 (14.3) years; mean (SD) 10-2 MD, −21.5 (5.6) dB. Fourteen central VF patterns were determined, including the most common temporal sparing patterns (304 [27.5%]), followed by mostly nasal loss (280 [25.4%]), hemifield loss (169 [15.3%]), central island (120 [10.9%]), total loss (91 [8.3%]), nearly intact field (56 [5.1%]), inferonasal quadrant sparing (42 [3.8%]), and nearly total loss (41 [3.7%]). Location-specific median total deviation analyses partitioned the central VF into a more vulnerable superonasal zone and a less vulnerable inferotemporal zone. At 1-year and 2-year follow-up, new defects mostly occurred in the more vulnerable zone. Initial encroachments on an intact central VF at follow-up were more likely to be from nasal loss (91 [18.4%], $P < .001$). One of the nasal loss patterns had a substantial chance at 2-year follow-up (38 [11.0%], $P = .004$) to shift to total loss, whereas others did not.

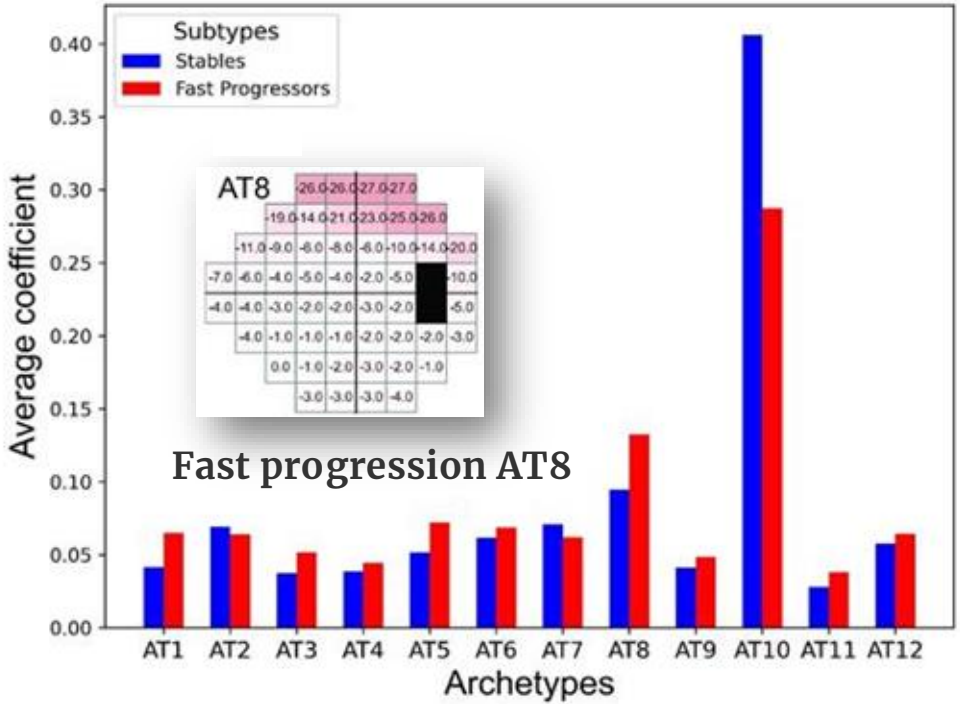
CONCLUSIONS AND RELEVANCE In this study, central VF loss in end-stage glaucoma was found to exhibit characteristic patterns that might be associated with different subtypes. Initial central VF loss is likely to be nasal loss, and 1 specific type of nasal loss is likely to develop into total loss.

Archetype 4 dominant: ↑ **chance** of total loss progression
Archetype 7 dominant: similar presentation, ↓ risk compared to 4



Distinct archetypal patterns stratify progression risk, linking VF phenotypes to meaningful clinical outcomes.

Mean AT coefficient of VF patterns of eyes in the **stable vs rapid** progressor subtypes.



OPEN ACCESS

ARVO Annual Meeting Abstract | June 2025

Identification of Visual Field Progression Subtypes and Factors Influencing Faster Progression Using Artificial Intelligence

Siamak Yousefi; Xiaoqin Huang; Asma Poursorousht; Chris A Johnson; Louis R Pasquale; Michael V Boland

+ Author Affiliations & Notes

Investigative Ophthalmology & Visual Science June 2025, Vol.66, 421. doi:

Key Takeaway:

- Global measures are reductionist and mask different subtypes of progression
- AA will help provide more granular analysis of functional progression data in glaucoma.

Roadmap

Functional & Clinical Diagnosis
Glaucoma diagnosis began with functional visual-field tests and clinical exams.

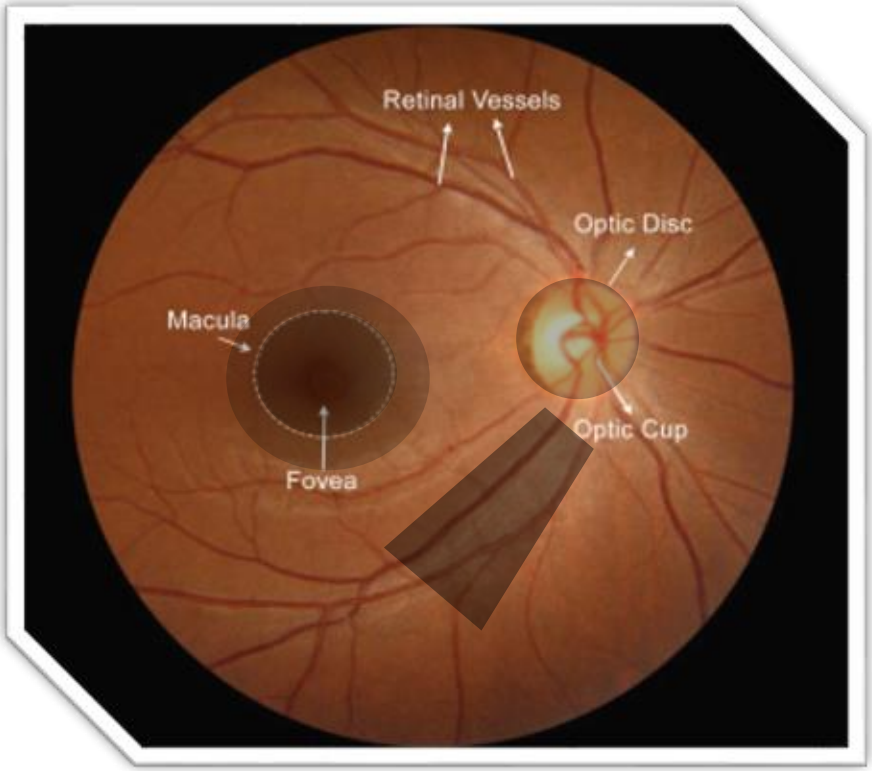
Image-based Deep Learning (CNN)
Convolutional neural networks enabled automated detection/segmentation from retinal images.

Multimodal Data Integration
Models began integrating imaging, visual field, and clinical EHR data for glaucoma assessment

Genetics & Risk Stratification
Detect progression risk early with genetic phenotyping and risk stratification

Foundation Models & Multimodal AI
Large datasets (millions) and foundation models will drive diagnosis and generalized ophthalmology care.

Fundus Glaucoma Markers

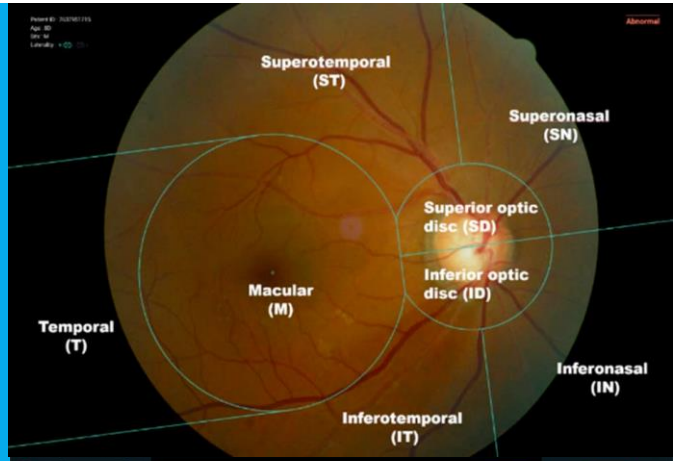


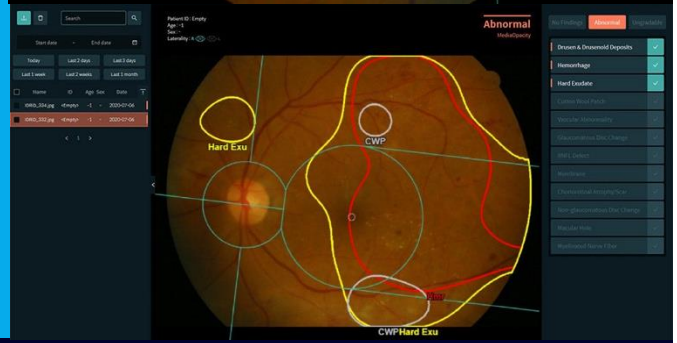
OPTIC NERVE HEAD GEOMETRY	• Vertical / horizontal CDR • Rim area & sectoral rims • Minimum rim width
RNFL & STRUCTURAL DAMAGE	• RNFL defect detection • Reflectance / texture loss • S–I symmetry index
PERIPAPILLARY & VASCULAR	• β- / γ-zone PPA area • Disc hemorrhage detection • Disc–fovea angle
GLOBAL & ASYMMETRY INDICES	• Inter-eye asymmetry (CDR/RNFL) • Glaucoma probability score • Structural severity class

01: Precision Ophthalmology

AI has the potential to transform ophthalmic imaging into a rich **biomarker data layer** providing micrometer-level changes over time and will help overcome the ambiguity surrounding glaucoma progression and management.

Standardized fundus biomarkers unlock earlier detection, consistent grading, and scalable screening across care settings.





Standardize ONH measurements

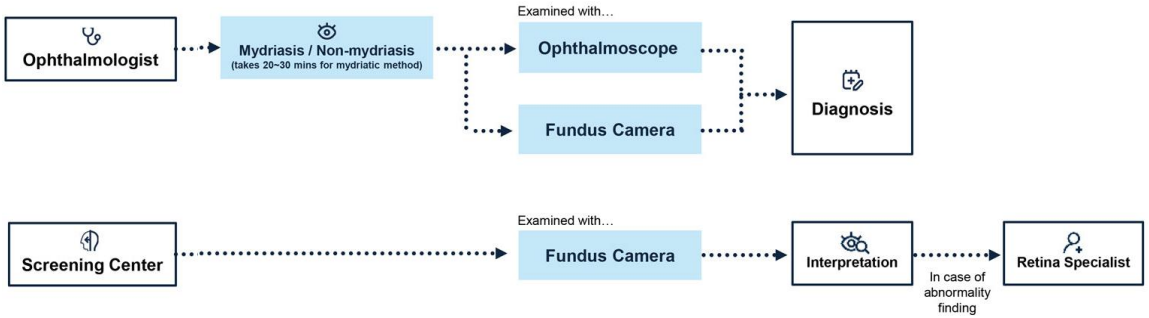
(CDR, rim width, neuroretinal rim area, RNFL defect localization) along with micro-patterns of retinal pathology

Early Detection of Structural Damage

(RNFL thickness, GCC thinning, BMO-MRW) detectable before functional loss

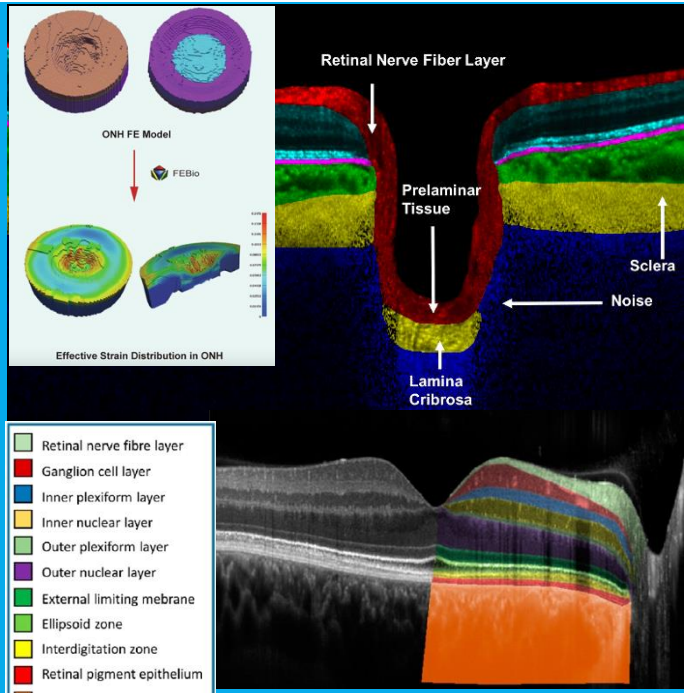
Rapid, Low-Cost Screening at Scale

mass population screening in primary care, retail clinics, and telemedicine networks.



Summary & Impact

OCT will evolve into a unified structural monitoring tool, capturing subtle changes and modeling biomechanics of disease



Layer segmentation with micron precision

Glaucoma diagnosis began with functional visual-field tests and clinical exams.

Highly sensitive structural biomarkers

(RNFL thickness, GCC thinning, BMO-MRW) detectable before functional loss.

Device-agnostic measurements

Enabling consistent metrics across varying OCT manufacturers and clinical site – harmonizing access to structured data

3D structural modeling of the optic nerve

providing insight into cup depth, lamina cribrosa displacement, and biomechanical stress.

Summary & Impact

OCT demonstrates diagnostic edge in glaucoma classification task

AMERICAN ACADEMY OF OPHTHALMOLOGY

Deep Learning with Disc Photos or OCT Scans in Glaucoma Detection

Akhilesh Kotnis, Jo Young Chang, MBBS, MPH, Iyad Majid, Lucy Q. Shen, MD, Mengyu Wang, PhD

Objective: To determine whether a deep learning (DL) model using retinal nerve fiber layer thickness (RNFLT) maps from OCT scans can detect glaucoma, defined by functional visual field (VF) impairment, more accurately than a DL model using disc photos (DPs). A secondary objective was to assess the diagnostic performance of these DL models across demographic groups (race, sex, and ethnicity).

Design: Retrospective cohort study at a tertiary glaucoma center utilizing OCT and DP datasets collected between 2015 and 2022.

Participants: Out of the 16 936 DP and OCT image sets, patients with Cirrus OCT images with a quality score ≥6 of 10 and reliable 24-2 Humphrey VF tests (fixation loss ≤33%, false-negative rate ≤20%, false-positive rate ≤20%), taken within 30 days of OCT, were included. Disc photos were obtained within 6 months of OCT. Data were randomly selected for training and testing of the DL models.

Testing: Development of DL models utilizing either OCT RNFLT maps or DPs to detect glaucoma based on VF-defined functional impairment.

Main Outcome Measures: The primary outcome was the area under the curve (AUC) for glaucoma detection, comparing the OCT-based DL model with the DP-based model. The secondary outcome was the AUC across demographic groups.

Results: The OCT-based DL model achieved an AUC of 0.90, significantly outperforming the DP-based model (AUC = 0.86, $P < 0.005$) with superior performance consistent across demographic groups. The OCT and DP model accuracies varied significantly by demographic groups. For the OCT model, AUCs were 0.93, 0.92, and 0.92 for Asians, Blacks, and Whites ($P < 0.005$); 0.89 for women versus 0.93 for men ($P < 0.005$); and 0.93 for Hispanics versus 0.94 for non-Hispanics ($P < 0.005$). For the DP model, corresponding AUCs for race were 0.87, 0.90, and 0.82 ($P < 0.005$); for sex, 0.856 versus 0.862 ($P < 0.005$); and for Hispanics, 0.85 versus 0.79 ($P < 0.005$).

Conclusions: When glaucoma diagnosis was based on functional deficit, the OCT-based DL model offered greater accuracy in detecting glaucoma than the DP-based model, likely due to its use of objective and quantitative RNFLT measurements. This work supports the use of OCT-based DL models for glaucoma detection, while observed demographic disparities underscore the need for equitable datasets to ensure fair DL-driven glaucoma diagnosis across populations.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science* 2025;5:100577 © 2025 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplemental material available at www.ophthalmologyscience.org.

Glaucoma is a group of progressive optic neuropathies marked by the gradual loss of peripheral and central vision.[1] As the second-leading cause of blindness worldwide, early and timely detection and treatment are important to preserve visual function.[2] However, glaucoma often remains asymptomatic until in later stages, resulting in delayed diagnosis.[3] Additionally, over half of patients are unaware they have the disease.[4,5] The prevalence of glaucoma and its risk factors also vary by race, gender, and ethnicity.[6,7]

Glaucoma screening typically occurs during primary care visits or annual eye checkups. However, individuals lacking access to these services—especially those from lower socioeconomic backgrounds—are more likely to be diagnosed at an advanced stage.[8–10]

To detect and monitor glaucoma, ophthalmologists typically use 2-dimensional color disc photos (DPs) and 3-dimensional (3D) OCT scans. Disc photo offers the clinical advantage of providing true color representation for documentation of optic nerve appearance, such as a disc hemorrhage.[11] In contrast, 3D OCT scans offer detailed and quantitative information about the different retinal layers, such as measurements of retinal nerve fiber layer thickness (RNFLT), which enables better glaucoma diagnosis and monitoring.[12]

While traditional imaging techniques remain invaluable in glaucoma detection and monitoring, recent advancements in artificial intelligence (AI) and deep learning (DL) algorithms have shown promise to significantly improve

https://doi.org/10.1016/j.xops.2025.100577
ISSN: 2666-9145

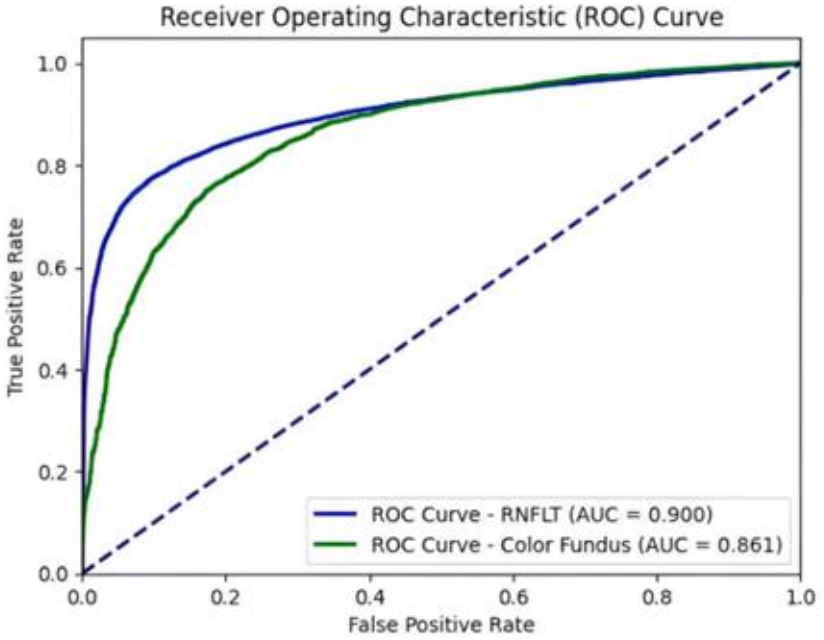


Figure 2. Comparison of DL models on glaucoma detection. Receiver operating characteristic curve of the DL model utilizing disc photos is shown in green and the model using OCT is shown in blue. Area under the curve of the DL model using OCT was 0.90; AUC of the DP model was 0.86. AUC = area under the curve; DL = deep learning; DP = disc photo; RNFLT = retinal nerve fiber layer thickness.

OCT demonstrates diagnostic edge in glaucoma classification task

AMERICAN ACADEMY OF OPHTHALMOLOGY

Deep Learning with Disc Photos or OCT Scans in Glaucoma Detection

Abhilash Kavuru,[1,*] Ji Young Chang, MBBS, MPH,[1,*] Iyad Majid,[1,*] Lucy Q. Shen, MD,[1,**] Mengyu Wang, PhD[1,2,**]

Objective: To determine whether a deep learning (DL) model using retinal nerve fiber layer thickness (RNFLT) maps from OCT scans can detect glaucoma, defined by functional visual field (VF) impairment, more accurately than a DL model using disc photos (DPs). A secondary objective was to assess the diagnostic performance of these DL models across demographic groups (race, sex, and ethnicity).

Design: Retrospective cohort study at a tertiary glaucoma center utilizing OCT and DP datasets collected between 2015 and 2022.

Participants: Out of the 16 936 DP and OCT image sets, patients with Cirrus OCT images with a quality score ≥6 of 10 and reliable 24-2 Humphrey VF tests (fixation loss ≤33%, false-negative rate ≤20%, false-positive rate ≤20%), taken within 30 days of OCT, were included. Disc photos were obtained within 6 months of OCT. Data were randomly selected for training and testing of the DL models.

Testing: Development of DL models utilizing either OCT RNFLT maps or DPs to detect glaucoma based on VF-defined functional impairment.

Main Outcome Measures: The primary outcome was the area under the curve (AUC) for glaucoma detection, comparing the OCT-based DL model with the DP-based model. The secondary outcome was the AUC across demographic groups.

Results: The OCT-based DL model achieved an AUC of 0.90, significantly outperforming the DP-based model (AUC = 0.86, $P < 0.005$) with superior performance consistent across demographic groups. The OCT and DP model accuracies varied significantly by demographic groups. For the OCT model, AUCs were 0.93, 0.92, and 0.92 for Asians, Blacks, and Whites ($P < 0.005$); 0.89 for women versus 0.93 for men ($P < 0.005$); and 0.93 for Hispanics versus 0.94 for non-Hispanics ($P < 0.005$). For the DP model, corresponding AUCs for race were 0.87, 0.90, and 0.82 ($P < 0.005$); for sex, 0.86 versus 0.862 ($P < 0.005$); and for Hispanics, 0.85 versus 0.79 ($P < 0.005$).

Conclusions: When glaucoma diagnosis was based on functional deficit, the OCT-based DL model offered greater accuracy in detecting glaucoma than the DP-based model, likely due to its use of objective and quantitative RNFLT measurements. This work supports the use of OCT-based DL models for glaucoma detection, while observed demographic disparities underscore the need for equitable datasets to ensure fair DL-driven glaucoma diagnosis across populations.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. Ophthalmology Science 2025;5:100577 © 2025 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplemental material available at www.ophthalmologyscience.org.

Glaucoma is a group of progressive optic neuropathies marked by the gradual loss of peripheral and central vision.[1] As the second leading cause of blindness worldwide, early and timely detection and treatment are important to preserve visual function.[2] However, glaucoma often remains asymptomatic until in later stages, resulting in delayed diagnosis.[3] Additionally, over half of patients are unaware they have the disease.[4,5] The prevalence of glaucoma and its risk factors also vary by race, gender, and ethnicity.[6,7]

Glaucoma screening typically occurs during primary care visits or annual eye checkups. However, individuals lacking access to these services—especially those from lower socioeconomic backgrounds—are more likely to be diagnosed at an advanced stage.[8–10]

To detect and monitor glaucoma, ophthalmologists typically use 2-dimensional color disc photos (DPs) and 3-dimensional (3D) OCT scans. Disc photo offers the clinical advantage of providing true color representation for documentation of optic nerve appearance, such as a disc hemorrhage.[11] In contrast, 3D OCT scans offer detailed and quantitative information about the different retinal layers, such as measurements of retinal nerve fiber layer thickness (RNFLT), which enables better glaucoma diagnosis and monitoring.[12]

While traditional imaging techniques remain invaluable in glaucoma detection and monitoring, recent advancements in artificial intelligence (AI) and deep learning (DL) algorithms have shown promise to significantly improve

https://doi.org/10.1016/j.xops.2025.100577
ISSN: 2666-9145/25/

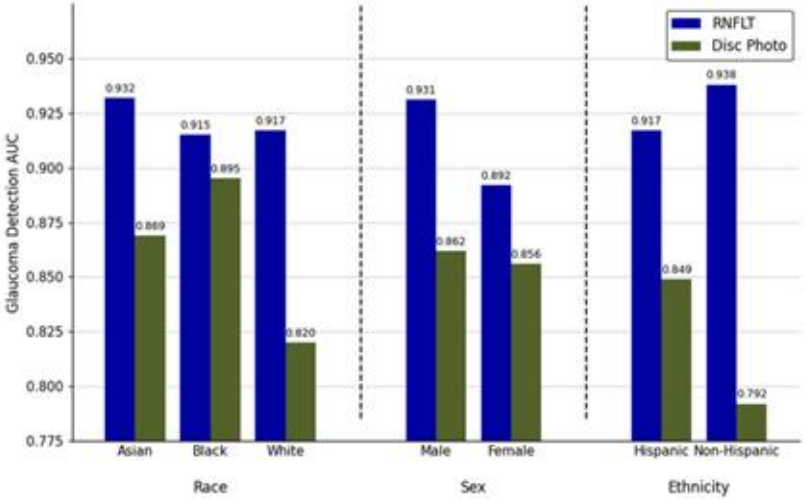
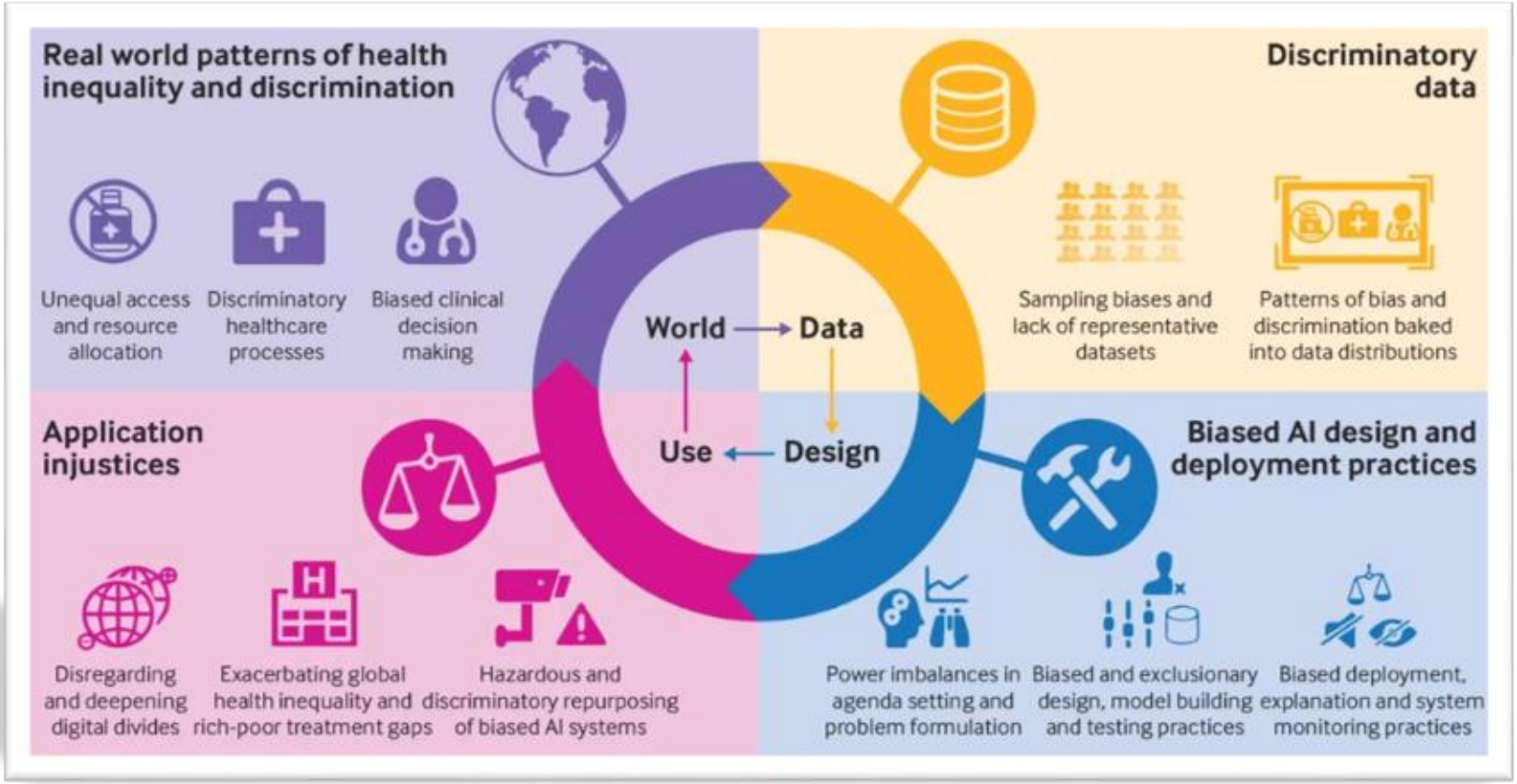


Figure 3. Comparison of artificial intelligence glaucoma detection models performance measured by AUC between OCT (in blue) and DP (in green) imaging modalities across various demographic factors (race, sex, and ethnicity). The blue bar represents the AUC of OCT model, and the green bar represents the AUC of DP model. AUC = area under the curve; DP = disc photo; RNFLT = retinal nerve fiber layer thickness.

Key Takeaway:

- Essential we interrogate equity of AI model performance
- OCT may provide more consistent/quantifiable features that correlate with glaucoma

Ensuring Equitable AI in Ophthalmology: Bias Across the Pipeline Impacts Glaucoma Care



Roadmap

Functional & Clinical Diagnosis
Glaucoma diagnosis began with functional visual-field tests and clinical exams.

Image-based Deep Learning (CNN)
Convolutional neural networks enabled automated detection/segmentation from retinal images.

Multimodal Data Integration
Models began integrating imaging, visual field, and clinical EHR data for glaucoma assessment

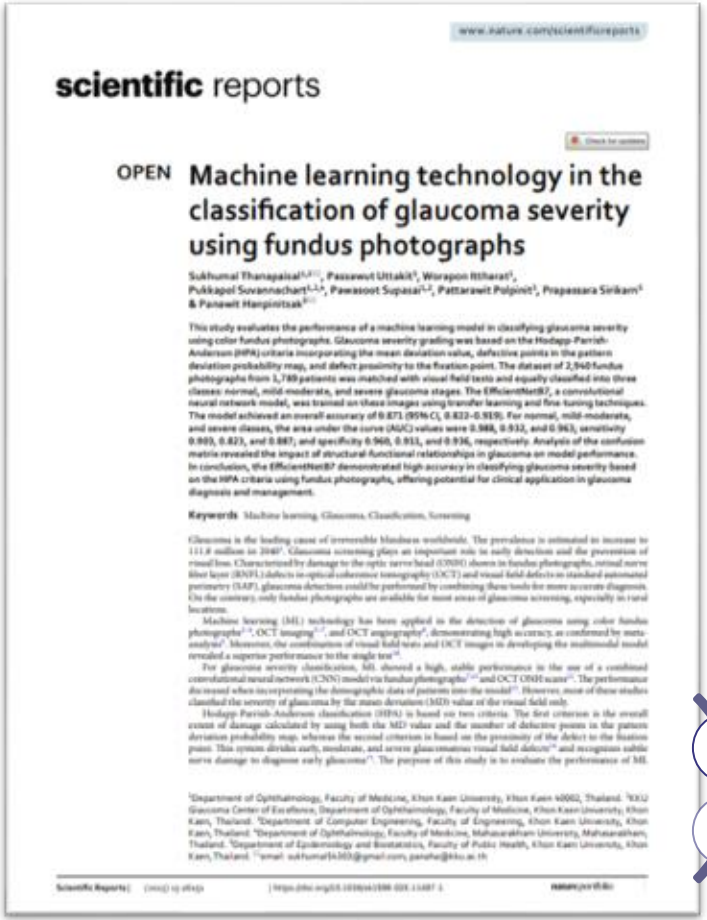
Genetics & Risk Stratification
Detect progression risk early with genetic phenotyping and risk stratification

Foundation Models & Multimodal AI
Large datasets (millions) and foundation models will drive diagnosis and generalized ophthalmology care.

02: Digital Microscope

By training on paired multimodal data, we can **inference gold standard clinical metrics** through widely accessible imaging, unlocking advanced diagnostics at a fraction of the time/cost

Transfer learning of VF progression endpoints to fundus photo



www.nature.com/scientificreports

scientific reports

OPEN

Machine learning technology in the classification of glaucoma severity using fundus photographs

Sukhumal Thanapaisal[1,2*], Passawut Uttakit[1], Worapon Ittharat[1], Pukkapol Suvannachart[1,3,4], Pawasoot Supasai[1,2], Pattarawit Polpinit[1], Prapassara Sirikarn[1] & Panawit Hanpinitsak[1*]

This study evaluates the performance of a machine learning model in classifying glaucoma severity using color fundus photographs. Glaucoma severity grading was based on the Hodapp-Parrish-Anderson (HPA) criteria incorporating the mean deviation value, defective points in the pattern deviation probability map, and defect proximity to the fixation point. The dataset of 2,940 fundus photographs from 1,780 patients was matched with visual field tests and equally classified into three classes: normal, mild-moderate, and severe glaucoma stages. The EfficientNetB7, a convolutional neural network model, was trained on these images using transfer learning and fine tuning techniques. The model achieved an overall accuracy of 0.871 (95% CI, 0.822–0.919). For normal, mild-moderate, and severe classes, the area under the curve (AUC) values were 0.988, 0.932, and 0.963; sensitivity 0.903, 0.823, and 0.887; and specificity 0.960, 0.933, and 0.936, respectively. Analysis of the confusion matrix revealed the impact of structural-functional relationships in glaucoma on model performance. In conclusion, the EfficientNetB7 demonstrated high accuracy in classifying glaucoma severity based on the HPA criteria using fundus photographs, offering potential for clinical application in glaucoma diagnosis and management.

Keywords Machine learning, Glaucoma, Classification, Screening

Glaucoma is the leading cause of irreversible blindness worldwide. The prevalence is estimated to increase to 111.8 million in 2040[1]. Glaucoma screening plays an important role in early detection and the prevention of visual loss. Characterized by damage to the optic nerve head (ONH) shown in fundus photographs, retinal nerve fiber layer (RNFL) defects in optical coherence tomography (OCT) and visual field defects in standard automated perimetry (SAP), glaucoma detection could be performed by combining these tools for more accurate diagnosis. On the contrary, only fundus photographs are available for most areas of glaucoma screening, especially in rural locations.

Machine learning (ML) technology has been applied in the detection of glaucoma using color fundus photographs[2–5], OCT imaging[6,7], and OCT angiography[8], demonstrating high accuracy, as confirmed by meta-analyses[9]. Moreover, the combination of visual field tests and OCT images in developing the multimodal model revealed a superior performance to the single test[10].

For glaucoma severity classification, ML showed a high, stable performance in the use of a combined convolutional neural network (CNN) model via fundus photographs[11,12] and OCT ONH scans[13]. The performance also soared when incorporating the demographic data of patients into the model[13]. However, most of these studies classified the severity of glaucoma by the mean deviation (MD) value of the visual field only.

Hodapp-Parrish-Anderson classification (HPA) is based on two criteria. The first criterion is the overall extent of damage calculated by using both the MD value and the number of defective points in the pattern deviation probability map, whereas the second criterion is based on the proximity of the defect to the fixation point. This system divides early, moderate, and severe glaucomatous visual field defects[14] and recognizes subtle nerve damage to diagnose early glaucoma[15]. The purpose of this study is to evaluate the performance of ML

[1]Department of Ophthalmology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. [2]KKU Glaucoma Center of Excellence, Department of Ophthalmology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. [3]Department of Computer Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand. [4]Department of Ophthalmology, Faculty of Medicine, Mahasarakham University, Mahasarakham, Thailand. [5]Department of Epidemiology and Biostatistics, Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand. *email: sukhumal1603@gmail.com; panahe@kku.ac.th

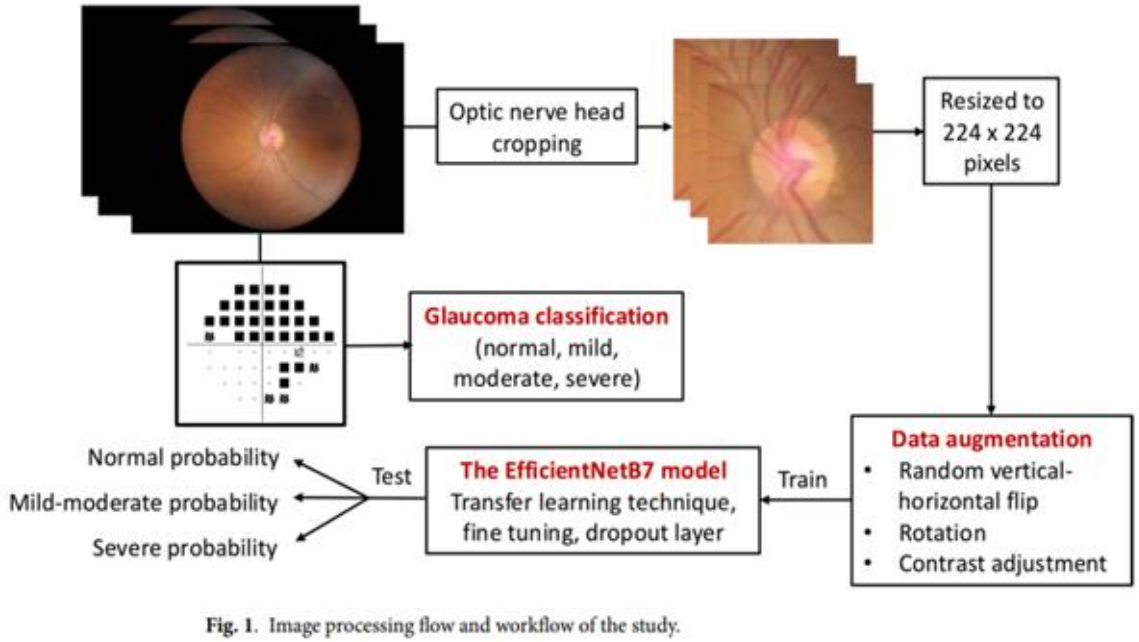
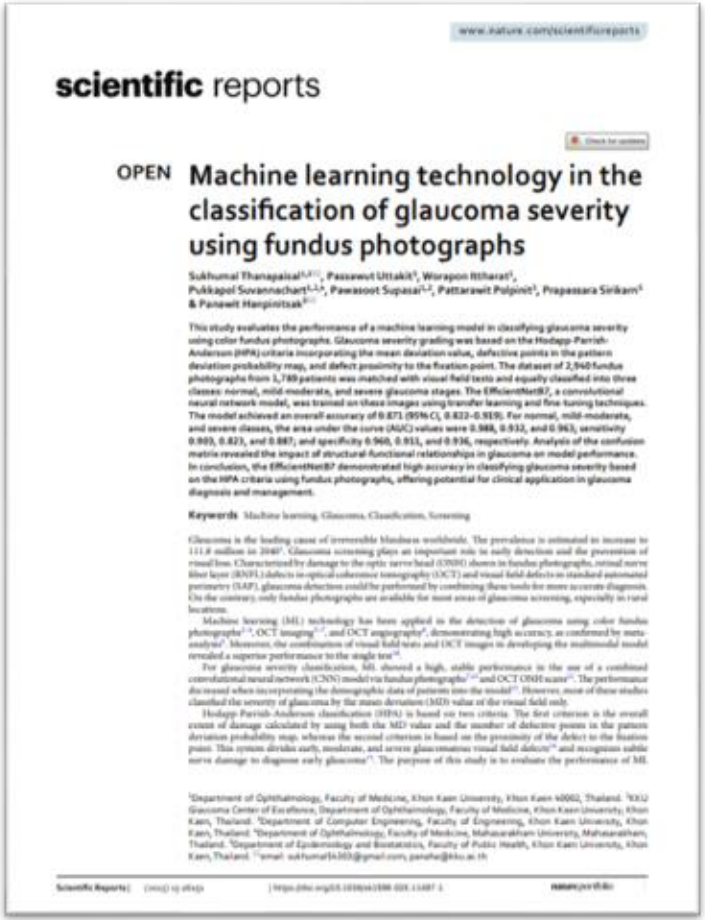


Fig. 1. Image processing flow and workflow of the study.

- ~3K paired images; ~1.8K patients; **normal, mid-moderate, severe**
- Predict functional endpoint from fundus with paired data modality

Transfer learning of VF progression endpoints to fundus photo



www.nature.com/scientificreports

scientific reports

OPEN

Machine learning technology in the classification of glaucoma severity using fundus photographs

Sukhumal Thanapaisal[1,2*], Passawut Uttakit[1], Worapon Ritharat[1], Pukkapol Suvannachart[1,3,4], Pawasoot Supasai[1,2], Pattarawit Polpinit[1], Prapassara Sirikarn[1] & Panawit Hanpinitsak[1*]

This study evaluates the performance of a machine learning model in classifying glaucoma severity using color fundus photographs. Glaucoma severity grading was based on the Hodapp-Parrish-Anderson (HPA) criteria incorporating the mean deviation value, defective points in the pattern deviation probability map, and defect proximity to the fixation point. The dataset of 2,940 fundus photographs from 1,780 patients was matched with visual field tests and equally classified into three classes: normal, mild-moderate, and severe glaucoma stages. The EfficientNetB7, a convolutional neural network model, was trained on these images using transfer learning and fine-tuning techniques. The model achieved an overall accuracy of 0.871 (95% CI, 0.822–0.919). For normal, mild-moderate, and severe classes, the area under the curve (AUC) values were 0.988, 0.932, and 0.963; sensitivity 0.903, 0.823, and 0.887; and specificity 0.960, 0.911, and 0.936, respectively. Analysis of the confusion matrix revealed the impact of structural-functional relationships in glaucoma on model performance. In conclusion, the EfficientNetB7 demonstrated high accuracy in classifying glaucoma severity based on the HPA criteria using fundus photographs, offering potential for clinical application in glaucoma diagnosis and management.

Keywords Machine learning, Glaucoma, Classification, Screening

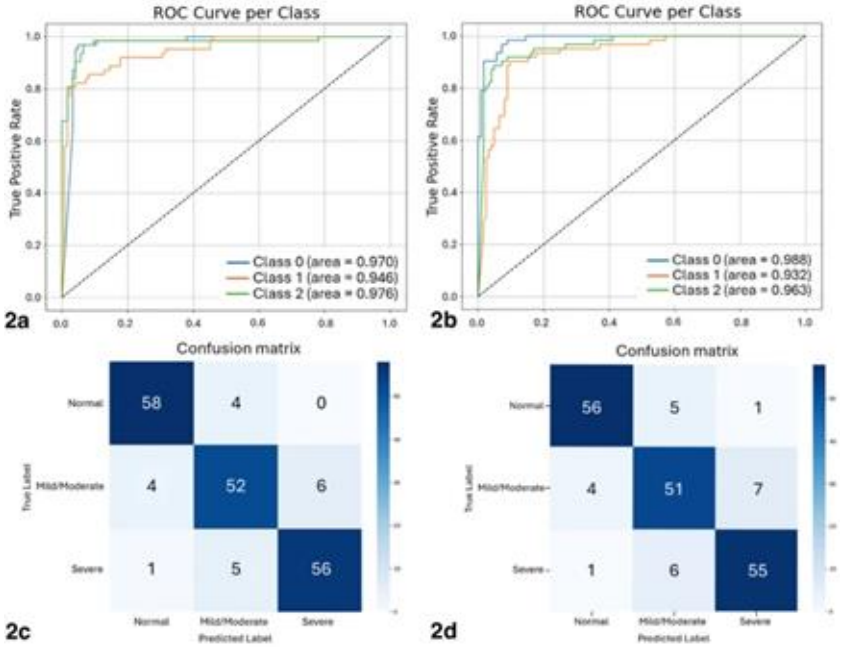
Glaucoma is the leading cause of irreversible blindness worldwide. The prevalence is estimated to increase to 111.8 million in 2040[1]. Glaucoma screening plays an important role in early detection and the prevention of visual loss. Characterized by damage to the optic nerve head (ONH) shown in fundus photographs, retinal nerve fiber layer (RNFL) defects in optical coherence tomography (OCT) and visual field defects in standard automated perimetry (SAP), glaucoma detection could be performed by combining these tools for more accurate diagnosis. On the contrary, only fundus photographs are available for most areas of glaucoma screening, especially in rural locations.

Machine learning (ML) technology has been applied in the detection of glaucoma using color fundus photographs[2–5], OCT imaging[6,7], and OCT angiography[8], demonstrating high accuracy, as confirmed by meta-analyses[9]. Moreover, the combination of visual field tests and OCT images in developing the multimodal model revealed a superior performance to the single test[10].

For glaucoma severity classification, ML showed a high, stable performance in the use of a combined convolutional neural network (CNN) model via fundus photographs[11,12] and OCT ONH scans[13]. The performance also soared when incorporating the demographic data of patients into the model[13]. However, most of these studies classified the severity of glaucoma by the mean deviation (MD) value of the visual field only.

Hodapp-Parrish-Anderson classification (HPA) is based on two criteria. The first criterion is the overall extent of damage calculated by using both the MD value and the number of defective points in the pattern deviation probability map, whereas the second criterion is based on the proximity of the defect to the fixation point. This system divides early, moderate, and severe glaucomatous visual field defects[14] and recognizes subtle nerve damage to diagnose early glaucoma[15]. The purpose of this study is to evaluate the performance of ML

[1]Department of Ophthalmology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. [2]KKU Glaucoma Center of Excellence, Department of Ophthalmology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. [3]Department of Computer Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand. [4]Department of Ophthalmology, Faculty of Medicine, Mahasarakham University, Mahasarakham, Thailand. [5]Department of Epidemiology and Biostatistics, Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand. *email: sukhumal1630@gmail.com; panahe@kku.ac.th



Key Takeaway:

- Strong ROC performance across all severity levels
- Overcome measurement noise from Visual Fields

Validating AI-Derived Biomarkers at Population Scale

RNFL Thickness in a Population-Based Cohort: The Canadian Longitudinal Study on Aging M2M (Machine-to-Machine) Study

ALI AZIZI, DOUGLAS R. DA COSTA, RAFAEL SCHERER, DAVINA A. MALEK, GUSTAVO A. SAMICO, AND FELIPE A. MEDEIROS

• PURPOSE: To evaluate factors associated with retinal nerve fiber layer (RNFL) thickness in the Canadian Longitudinal Study on Aging (CLSA) using the Machine-to-Machine (M2M) deep learning model applied to fundus photographs.
• DESIGN: Cross-sectional study.
• SUBJECTS: Participants from the baseline Comprehensive Cohort of the CLSA.
• METHODS: This study included 28,114 CLSA participants aged 45 to 85 years with gradable baseline fundus photographs. The M2M model, trained on optical coherence tomography (OCT) data, was applied to estimate RNFL thickness from disc-centered fundus images. For participants with images from both eyes, the mean RNFL thickness of the 2 eyes was used. Associations between M2M-predicted RNFL thickness and age, sex, ethnicity/race, and self-reported glaucoma were analyzed using linear regression models adjusted for covariates.
• MAIN OUTCOME MEASURES: M2M-predicted RNFL thickness, age, age groups, sex, ethnicity/race, and self-reported glaucoma.
• RESULTS: The mean age of participants was 62.6 ± 10.1 years, and 51% were women. Self-reported glaucoma was present in 4.8% of the participants. The mean M2M-predicted RNFL thickness was 90.9 ± 9.2 μm. Age was inversely associated with RNFL thickness (Pearson's $r = -0.16$; $p < .001$), with each additional year associated with a 0.15 μm decrease ($p < .001$); after adjustment for covariates, the association remained significant ($\beta = -0.11$; $p < .001$). Participants with self-reported glaucoma exhibited significantly thinner RNFL (82.6 ± 12.9 μm) compared to those without (91.4 ± 8.7 μm; $p < .001$). RNFL thickness was slightly greater in women than in men ($p < .001$), and differences were observed across ethnicity/race groups ($p < .001$).
• CONCLUSIONS: The M2M model provided robust estimates of RNFL thickness from fundus photographs in a large population-based cohort. The observed associations between RNFL thickness, age, and glaucoma status were consistent with previous OCT-based findings, supporting the utility of the model for scalable structural assessments in epidemiological studies. (Am J Ophthalmol 2026;281:395–403. © 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/))

THE CANADIAN LONGITUDINAL STUDY ON AGING (CLSA) is a major national research initiative designed to enhance understanding of aging and its impact on health and well-being in the Canadian population.[1] It involves a diverse cohort of over 51,000 participants aged 45 to 85, who provided in-depth data through interviews, physical assessments, and biological samples. With baseline data collection completed in 2015, the study will follow participants every 3 years until 2035 or their death. As part of the ophthalmic evaluation, participants underwent visual acuity, tonometry, and retinal scans (fundus photography). Analyzing this extensive dataset may provide valuable insights into the effects of aging on ocular health.

A key limitation of the ocular imaging data in the CLSA is its reliance on fundus photography. While fundus images provide documentation of the retina and optic nerve head, their interpretation often requires subjective assessment by experts, which can be poorly reproducible and inconsistent.[2] This is particularly relevant in the context of glaucoma, the leading cause of irreversible blindness worldwide.[3] Glaucoma is characterized by progressive retinal ganglion cell loss and thinning of the retinal nerve fiber layer (RNFL),[4] changes that can be difficult to assess from fundus photographs alone. Subjective grading of these images has been shown to have significant interobserver variability,[5] reducing their reliability for large-scale epidemiological studies.

To address these limitations, the Machine-to-Machine (M2M) deep learning model was developed to provide an objective, quantitative assessment of neural damage from

Supplemental Material available at AJO.com.
Accepted for publication September 27, 2025.
Duncan Fellow Eye Institute, University of Miami, Miami, Florida, USA.
Inquiries to Felipe A. Medeiros, Duncan Fellow Eye Institute, Miami, FL, USA; e-mail: fmedeiros@med.miami.edu

© 2025 The Author(s). Published by Elsevier Inc.
THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY-NC-ND LICENSE

0002-9394/$36.00
https://doi.org/10.1016/j.ajo.2025.09.015

395

TABLE 2. Multivariable Linear Regression Analysis of Machine-to-Machine (M2M)-Predicted Retinal Nerve Fiber Layer (RNFL) Thickness Adjusted for Age, Sex, Ethnicity/Race, and Self-Reported Glaucoma Status.

Variable	Coefficient (β)	95% CI	*P*-Value
Age (Per Year)	−0.11	[−0.12, −0.10]	**<.001**
Sex (Female)	0.93	[0.72, 1.14]	**<.001**
Ethnicity/Race (White)	−1.58	[−2.14, −1.03]	**<.001**
Self-Reported Glaucoma Status (Yes)	−6.78	[−7.65, −5.92]	**<.001**
Age × Glaucoma (Interaction)	−0.14	[−0.21, −0.07]	**<.001**
Constant	92.34	[91.78, 92.89]	**<.001**

M2M = Machine-to-Machine; RNFL = Retinal Nerve Fiber Layer.
Note: Age was centered at the participants' mean age (62.6 years).

Key Takeaway:

- M2M-predicted RNFL thickness demonstrated significant associations with age and self-reported glaucoma status
- Deep learning can unlock quantitative structural assessments in massive epidemiological cohorts where performing OCT on everyone simply isn't feasible

Roadmap

Functional & Clinical Diagnosis
Glaucoma diagnosis began with functional visual-field tests and clinical exams.

Image-based Deep Learning (CNN)
Convolutional neural networks enabled automated detection/segmentation from retinal images.

Multimodal Data Integration
Models began integrating imaging, visual field, and clinical EHR data for glaucoma assessment

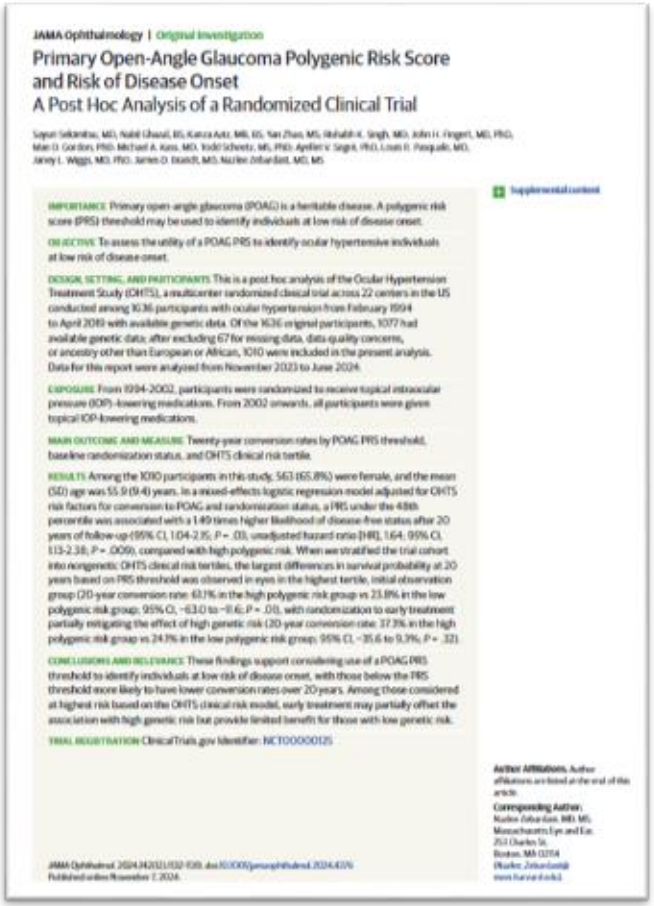
Genetics & Risk Stratification
Detect progression risk early with genetic phenotyping and risk stratification

Foundation Models & Multimodal AI
Large datasets (millions) and foundation models will drive diagnosis and generalized ophthalmology care.

03: Genomic Signature of Disease

From polygenic risk at the population level to cellular programs revealed by single-cell sequencing, genomics uncovers both **how glaucoma develops** and **who is most at risk**...advancing precision intervention.

PRS identifies patients with higher progression risk



JAMA Ophthalmology | Original Investigation

Primary Open-Angle Glaucoma Polygenic Risk Score and Risk of Disease Onset
A Post Hoc Analysis of a Randomized Clinical Trial

Sayan Sekhar Mitra, MD; Nabil Ghaseb, BS; Kanza Aziz, MB, BS; Yan Zhao, MS; Rishabh K. Singh, MD; John H. Fingert, MD, PhD; Mae O. Gordon, PhD; Michael A. Kass, MD; Todd Scheetz, ML, PhD; Ayellet V. Segrè, PhD; Louis R. Pasquale, MD; Janey L. Wiggs, MD, PhD; James D. Brandt, MD; Nazlee Zebardast, MD, MS

IMPORTANCE Primary open-angle glaucoma (POAG) is a heritable disease. A polygenic risk score (PRS) threshold may be used to identify individuals at low risk of disease onset.

OBJECTIVE To assess the utility of a POAG PRS to identify ocular hypertensive individuals at low risk of disease onset.

DESIGN, SETTING, AND PARTICIPANTS This is a post hoc analysis of the Ocular Hypertension Treatment Study (OHTS), a multicenter randomized clinical trial across 22 centers in the US conducted among 1636 participants with ocular hypertension from February 1994 to April 2010 with available genetic data. Of the 1636 original participants, 1077 had available genetic data; after excluding 67 for missing data, data quality concerns, or ancestry other than European or African, 1010 were included in the present analysis. Data for this report were analyzed from November 2023 to June 2024.

EXPOSURE From 1994-2002, participants were randomized to receive topical intraocular pressure (IOP)-lowering medications. From 2002 onwards, all participants were given topical IOP-lowering medications.

MAIN OUTCOME AND MEASURE Twenty-year conversion rates by POAG PRS threshold, baseline randomization status, and OHTS clinical risk tertile.

RESULTS Among the 1010 participants in this study, 563 (65.8%) were female, and the mean (SD) age was 55.9 (9.4) years. In a mixed-effects logistic regression model adjusted for OHTS risk factors for conversion to POAG and randomization status, a PRS under the 40th percentile was associated with a 1.49 times higher likelihood of disease-free status after 20 years of follow-up (95% CI, 1.04-2.15; $P = .03$; unadjusted hazard ratio [HR], 1.64; 95% CI, 1.13-2.38; $P = .009$), compared with high polygenic risk. When we stratified the trial cohort into nongenetic OHTS clinical risk tertiles, the largest differences in survival probability at 20 years based on PRS threshold was observed in eyes in the highest tertile, initial observation group (20-year conversion rate: 61.1% in the high polygenic risk group vs 23.8% in the low polygenic risk group; 95% CI, −63.0 to −11.6; $P = .01$), with randomization to early treatment partially mitigating the effect of high genetic risk (20-year conversion rate: 37.3% in the high polygenic risk group vs 24.1% in the low polygenic risk group; 95% CI, −35.6 to 9.3%; $P = .32$).

CONCLUSIONS AND RELEVANCE These findings support considering use of a POAG PRS threshold to identify individuals at low risk of disease onset, with those below the PRS threshold more likely to have lower conversion rates over 20 years. Among those considered at highest risk based on the OHTS clinical risk model, early treatment may partially offset the association with high genetic risk but provide limited benefit for those with low genetic risk.

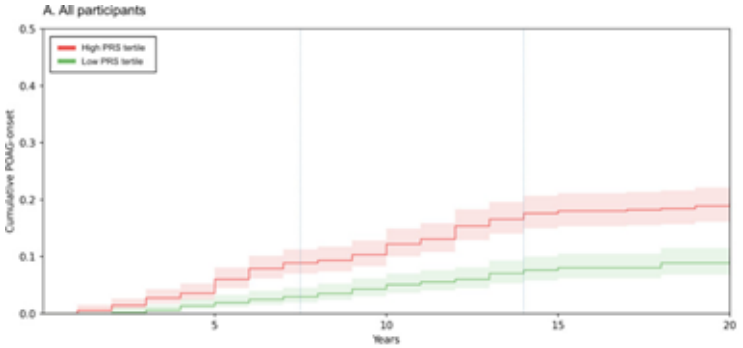
TRIAL REGISTRATION ClinicalTrials.gov Identifier: NCT00000125

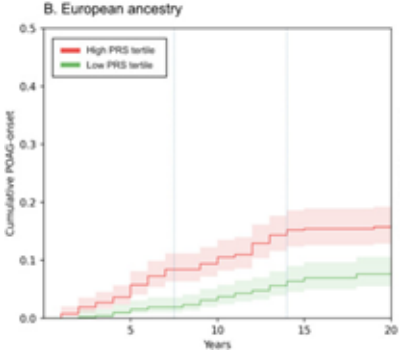
Supplemental content

Author Affiliations. Author affiliations are listed at the end of this article.

Corresponding Author: Nazlee Zebardast, MD, MS, Massachusetts Eye and Ear, 253 Charles St, Boston, MA 02114 (Nazlee_Zebardast@meei.harvard.edu).

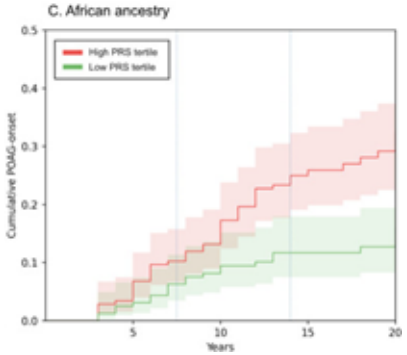
JAMA Ophthalmol. 2024;142(12):1132-1139. doi:10.1001/jamaophthalmol.2024.4376
Published online November 7, 2024.



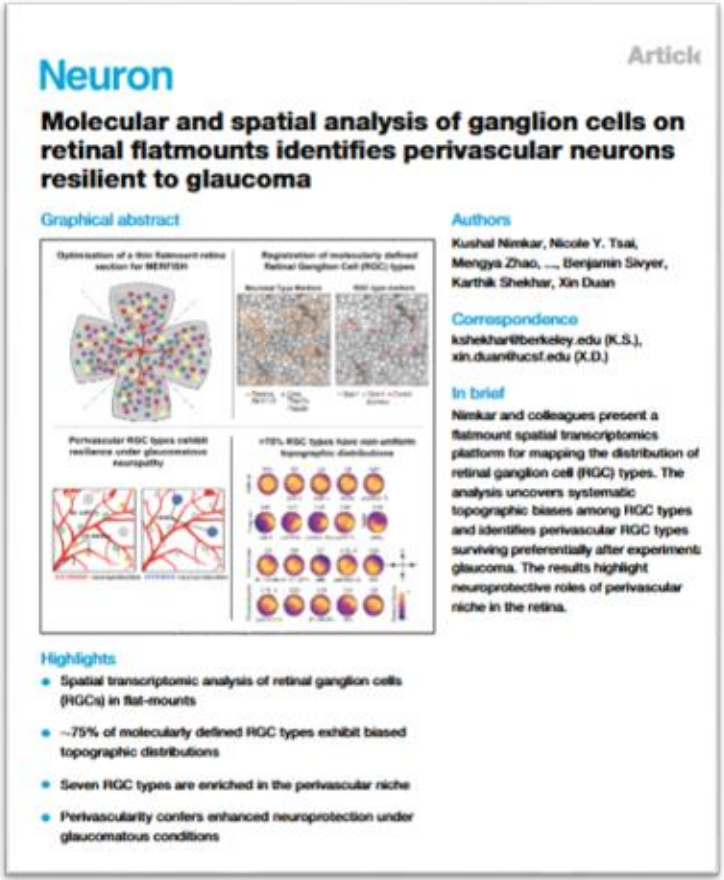


Key Takeaway:

- PRS is calculated by aggregating thousands of glaucoma-associated genetic variants
- Higher PRS values strongly correlate with increased lifetime glaucoma risk, earlier disease onset, and faster structural progression.
- AI models can combine PRS with imaging and clinical data to improve risk prediction, stratify patients, and personalize screening strategies.



Unsupervised clustering of scRNA-seq identifies 45 clusters of phenotypic RGC expression/cell types



Neuron

Artick

Molecular and spatial analysis of ganglion cells on retinal flatmounts identifies perivascular neurons resilient to glaucoma

Graphical abstract

Authors

Kushal Nimkar, Nicole Y. Tsai, Mengya Zhao, ..., Benjamin Sivyer, Karthik Shekhar, Xin Duan

Correspondence

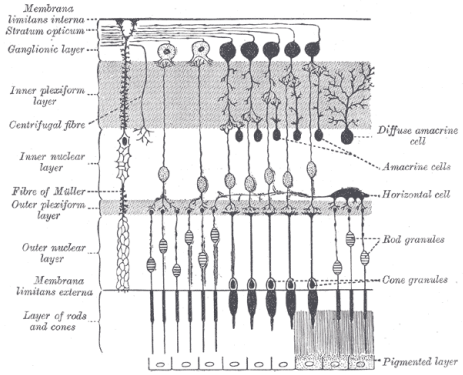
kshekhar@berkeley.edu (K.S.), xin.duan@ucsf.edu (X.D.)

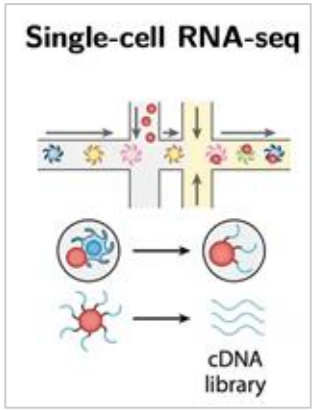
In brief

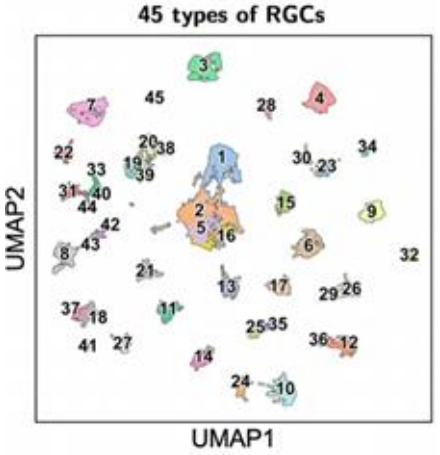
Nimkar and colleagues present a flatmount spatial transcriptomics platform for mapping the distribution of retinal ganglion cell (RGC) types. The analysis uncovers systematic topographic biases among RGC types and identifies perivascular RGC types surviving preferentially after experimental glaucoma. The results highlight neuroprotective roles of perivascular niche in the retina.

Highlights

- Spatial transcriptomic analysis of retinal ganglion cells (RGCs) in flat-mounts
- ~75% of molecularly defined RGC types exhibit biased topographic distributions
- Seven RGC types are enriched in the perivascular niche
- Perivascularity confers enhanced neuroprotection under glaucomatous conditions







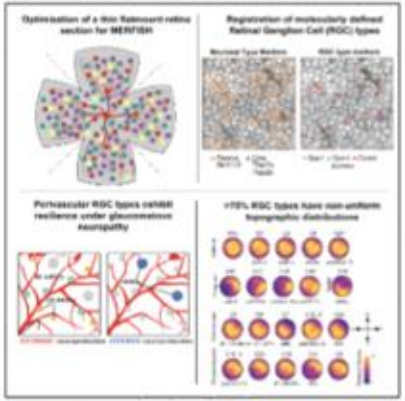
Perivascular retinal ganglion cells show resilience to glaucomatous damage

Neuron

Article

Molecular and spatial analysis of ganglion cells on retinal flatmounts identifies perivascular neurons resilient to glaucoma

Graphical abstract



Authors

Kushal Nimkar, Nicole Y. Tsai, Mengya Zhao, ..., Benjamin Sivyer, Karthik Shekhar, Xin Duan

Correspondence

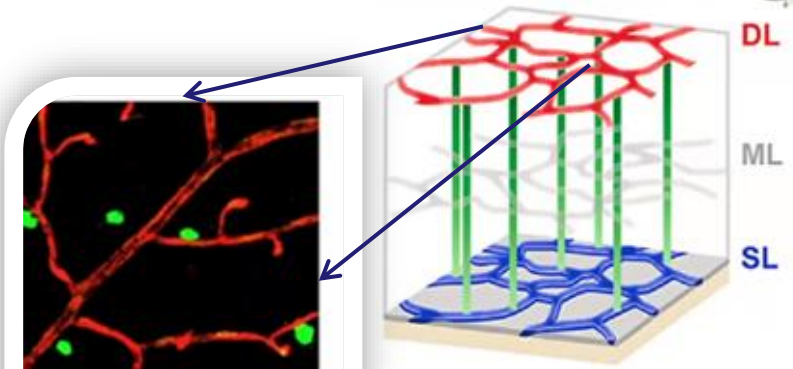
kshekhar@berkeley.edu (K.S.), xin.duan@ucsf.edu (X.D.)

In brief

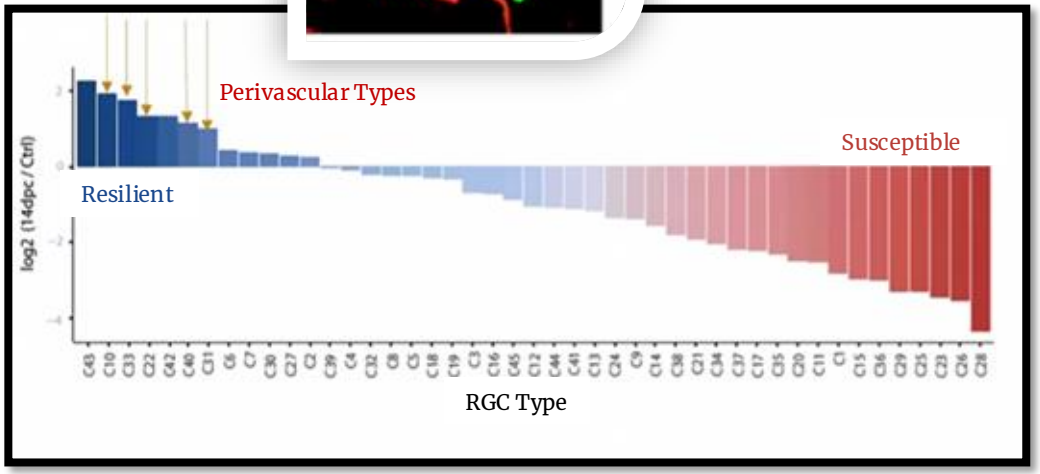
Nimkar and colleagues present a flatmount spatial transcriptomics platform for mapping the distribution of retinal ganglion cell (RGC) types. The analysis uncovers systematic topographic biases among RGC types and identifies perivascular RGC types surviving preferentially after experimental glaucoma. The results highlight neuroprotective roles of perivascular niche in the retina.

Highlights

- Spatial transcriptomic analysis of retinal ganglion cells (RGCs) in flat-mounts
- ~75% of molecularly defined RGC types exhibit biased topographic distributions
- Seven RGC types are enriched in the perivascular niche
- Perivascularity confers enhanced neuroprotection under glaucomatous conditions



Toma et al., Cell, 2024



Roadmap

Functional & Clinical Diagnosis
Glaucoma diagnosis began with functional visual-field tests and clinical exams.

Image-based Deep Learning (CNN)
Convolutional neural networks enabled automated detection/segmentation from retinal images.

Multimodal Data Integration
Models began integrating imaging, visual field, and clinical EHR data for glaucoma assessment

Genetics & Risk Stratification
Detect progression risk early with genetic phenotyping and risk stratification

Foundation Models & Multimodal AI
Large datasets (millions) and foundation models will drive diagnosis and generalized ophthalmology care.

04: Foundation Model Future

Foundation models transform ophthalmology by creating universal, device-agnostic representations that enable earlier diagnosis, reliable progression prediction, and equitable performance across diverse patients and imaging systems...**all from a unified model**

From Narrow, Task-Specific Models to Generalized Intelligence

Traditional AI Models

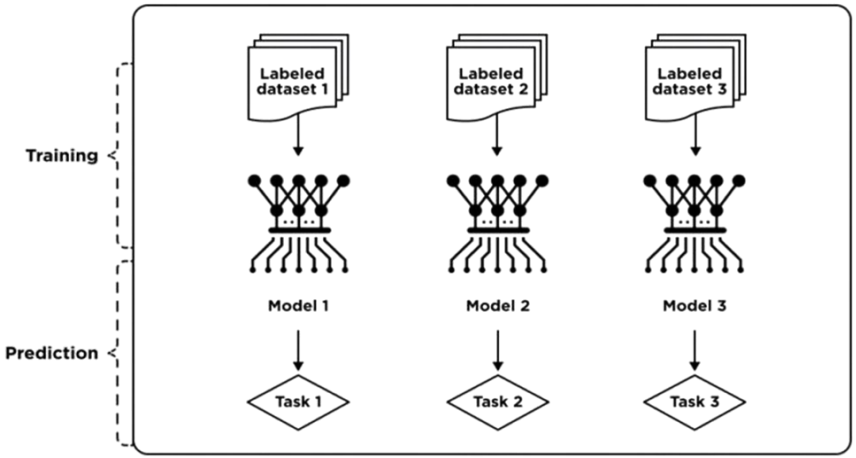
- Task-Specific Training
- Large Labeled Datasets Required
- Limited Generalization

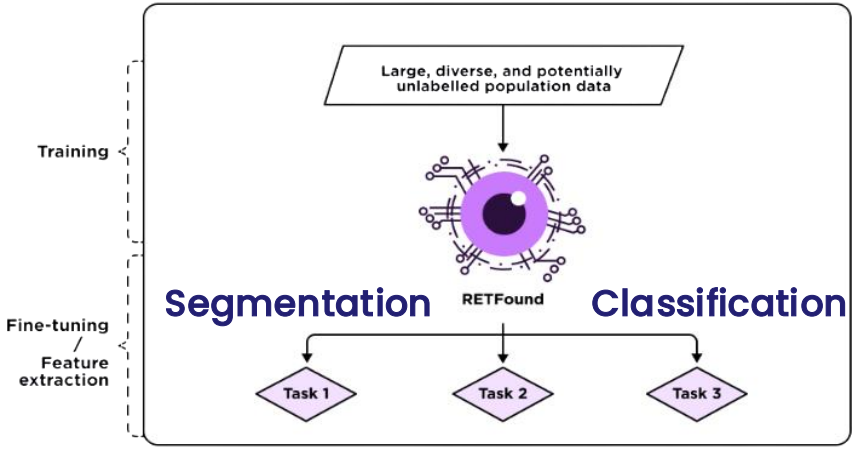
VS

01 02 03

Foundation Models

- Multi-Task, Multi-Modal Learning
- Leverages Massive Unlabeled Data
- Fine-Tuning with Small Datasets

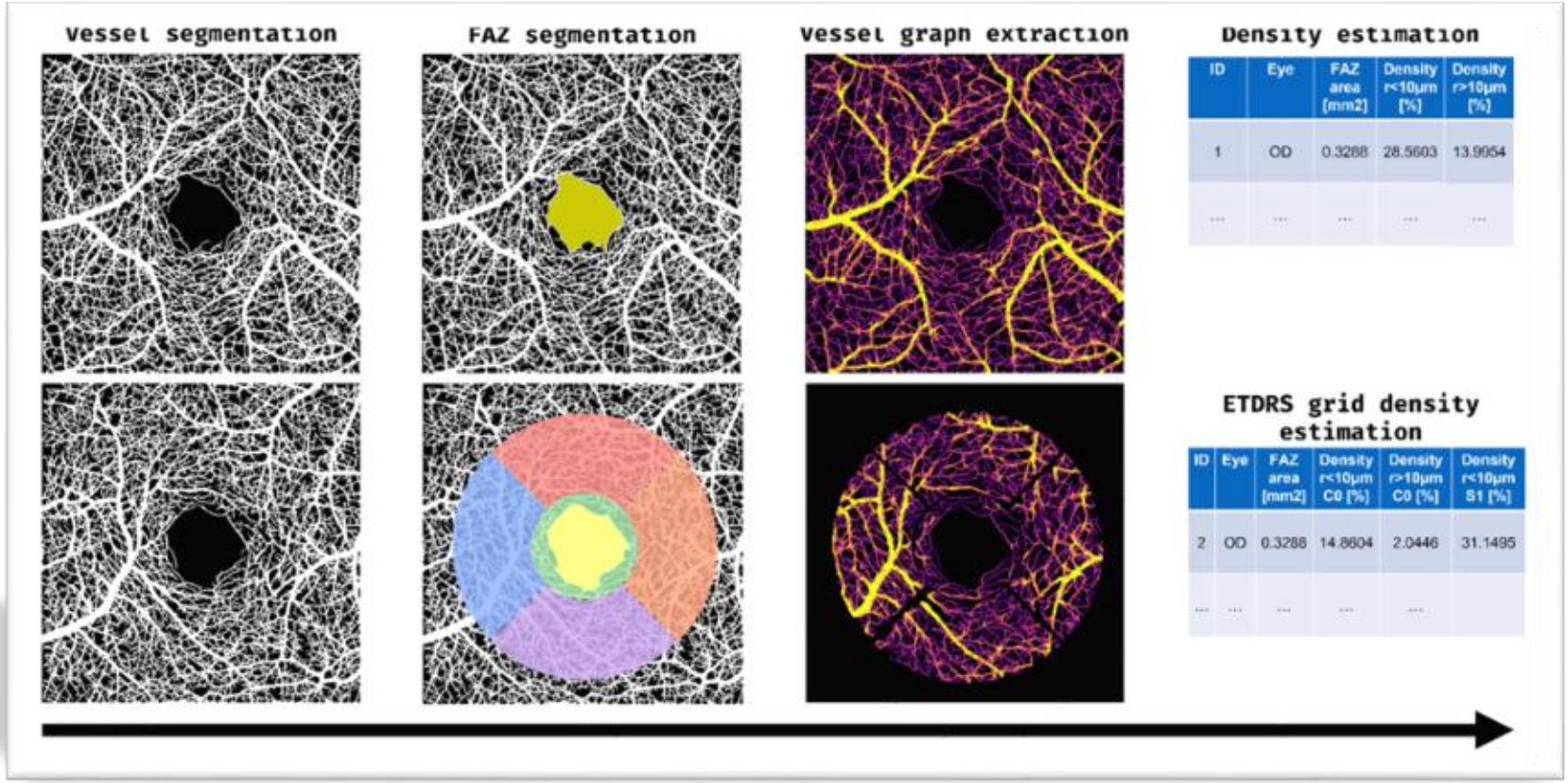




05: Oculomics Future

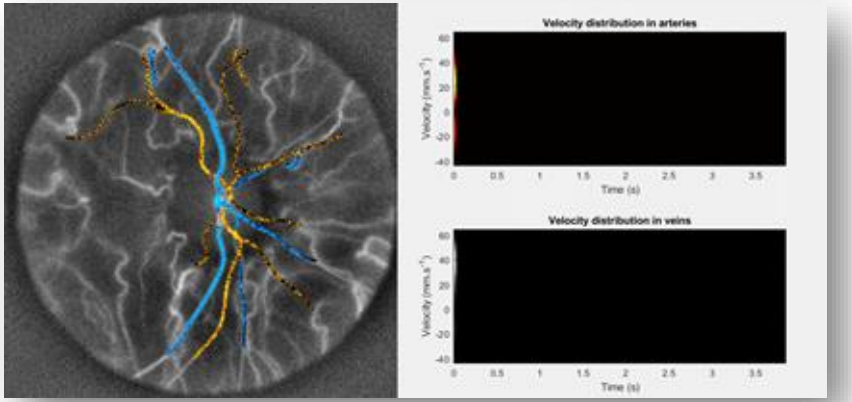
Emerging retinal imaging technologies combined with AI-enabled modeling now offer cellular-level structural and vascular insights, positioning the eye as a **non-invasive window** into systemic disease.

Unified OCTA analysis: automated vessel maps, FAZ segmentation, vascular graphs, and localized perfusion metrics

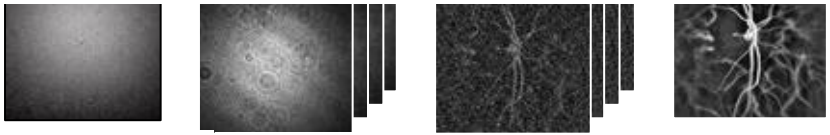


computed imaging for non-invasive angiography

Doppler holography

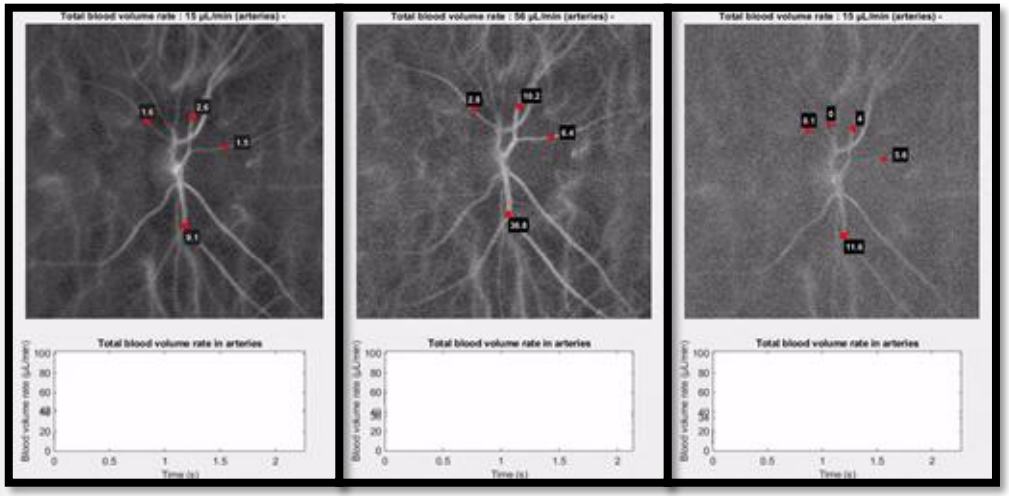


- Open-source image rendering and analysis software
- Estimation of hemodynamical and rheological parameters
- Technology transfer to ophthalmic device companies and academic laboratories, worldwide



raw frames | Space transformation | Time transformation

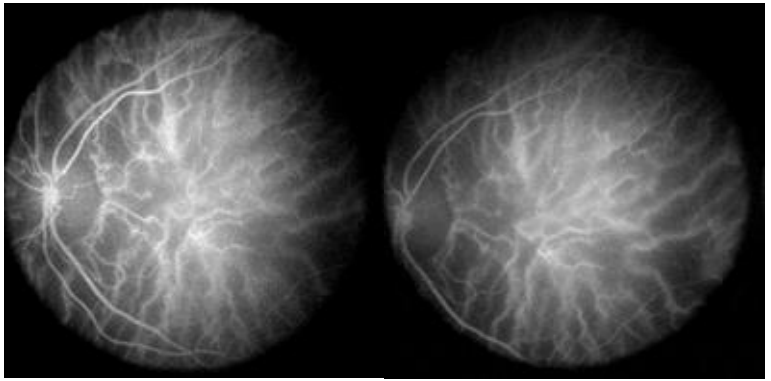
Estimated total retinal blood flow vs. laser exposure



exposure : 3.8 mW | exposure : 1.6 mW | exposure : 0.7 mW

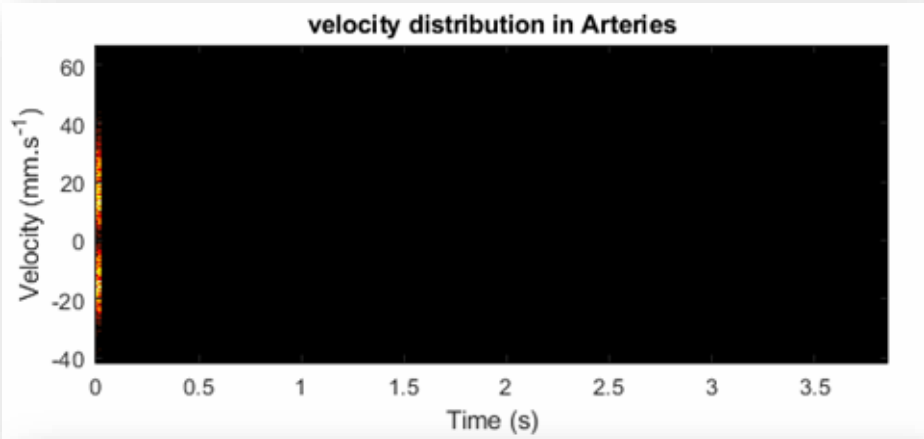
anterior segment

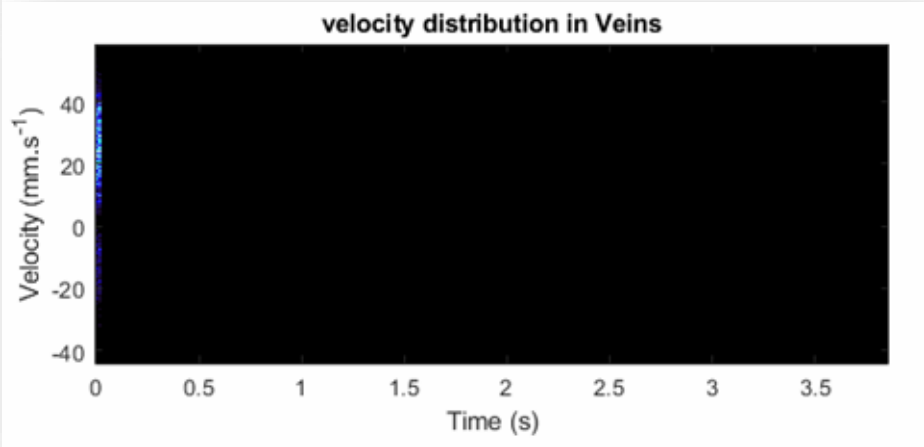
device

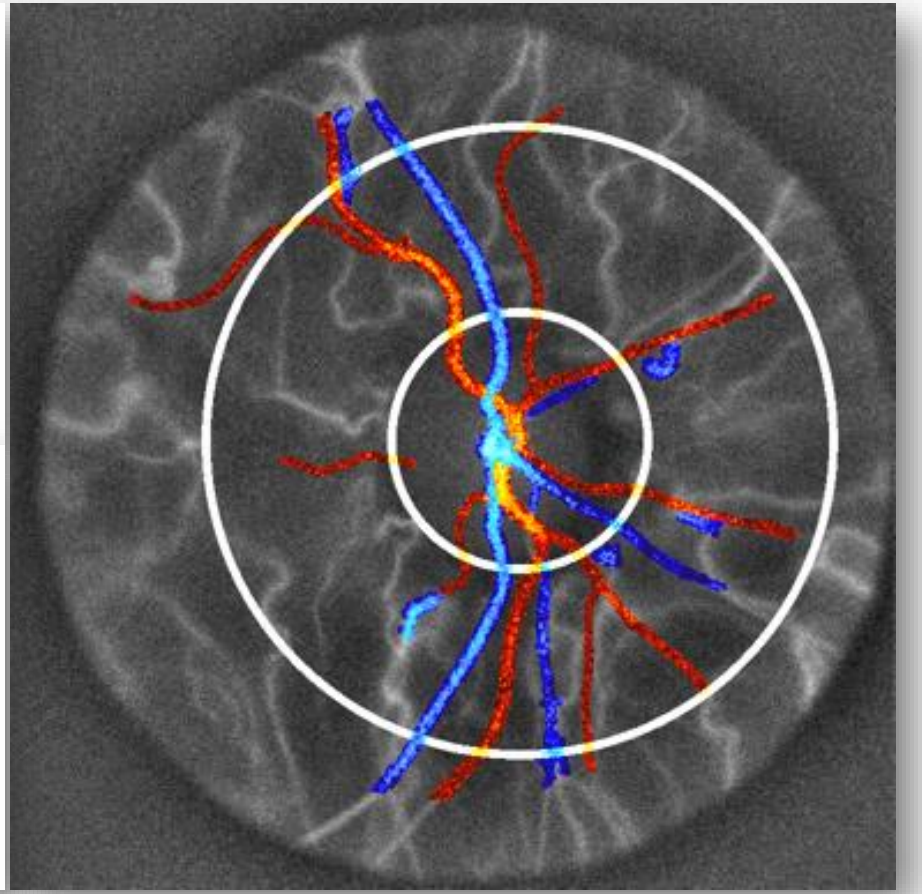


Real-time computational imaging **with 34,000 frames/s. 384x384 pixel frame.**

Average velocity profile estimate from CRA and CRV

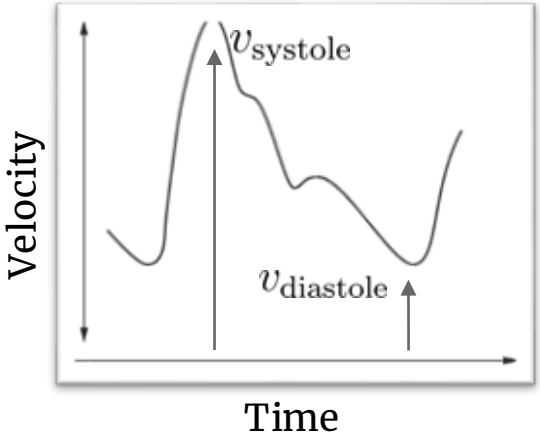




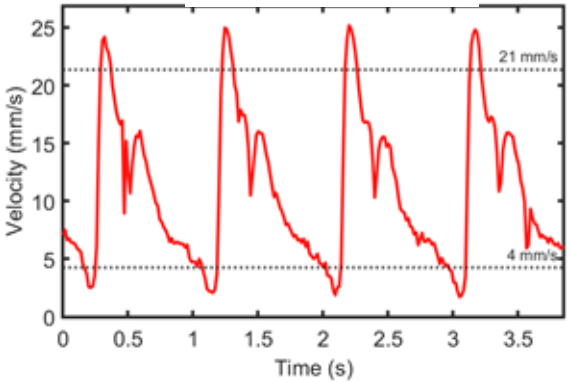


Elevated Arterial Resistivity Reflects Vascular Dysfunction in Glaucoma

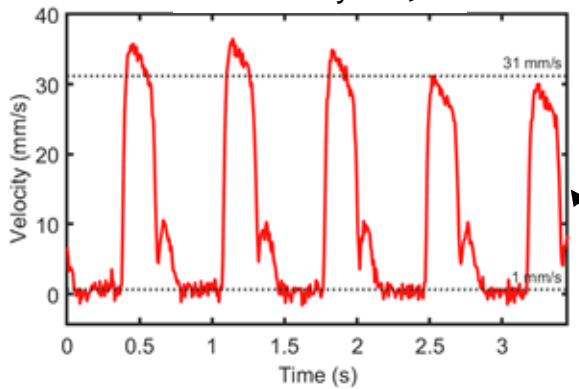
$$RI = \frac{v_{systole} - v_{diastole}}{v_{systole}}$$



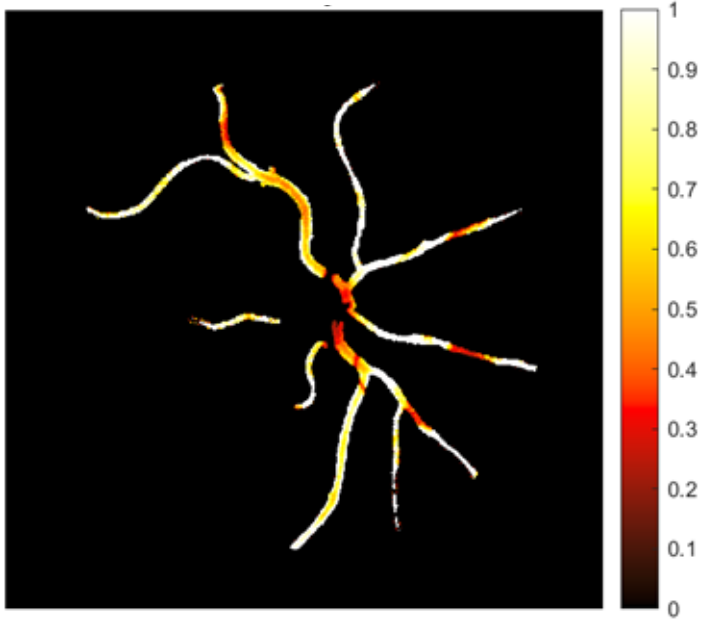
ARI velocity = 0.69



ARI velocity = 0.98

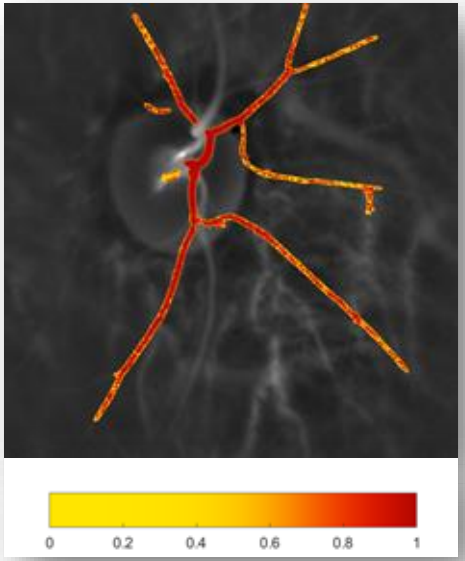


ARI velocity = 0.69

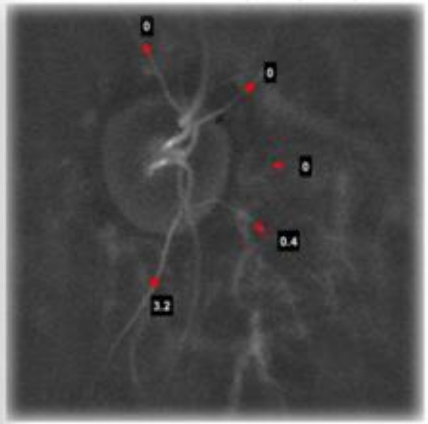


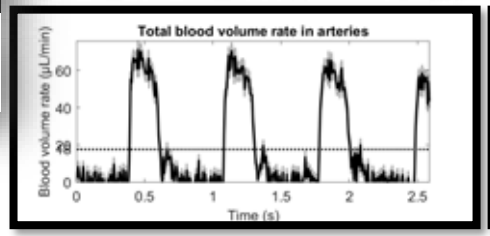
glaucomatous patient

IOP Elevation leading to reduction in CRA blood volume and increased arterial resistivity



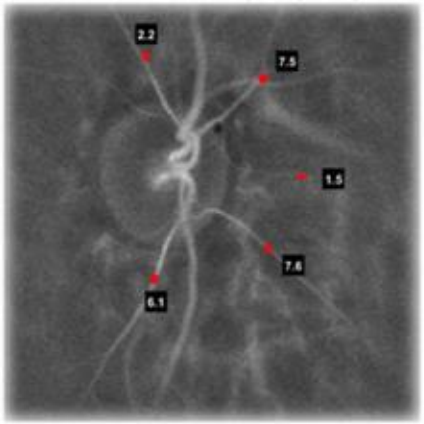
ARI index : 1

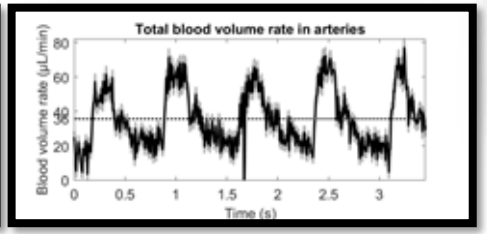




Arterial blood flow: 18 µL/min
Arterial resistivity index: 1

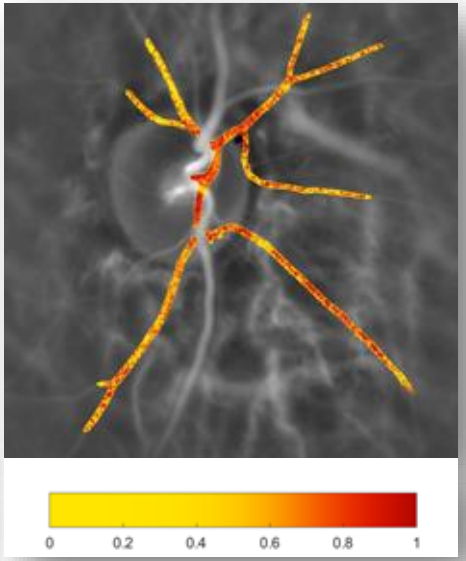
OD, IOP = 38 mmHg





Arterial blood flow: 36 µL/min
Arterial resistivity index: 0.729

OD, IOP = 7 mmHg, post-trabeculectomy



ARI index : 0.729

06: Closing Remarks

The essential ingredients for improving patient outcomes in the age of AI are the same ones that have always defined medicine: scientific rigor, human compassion, daring imagination, critical thinking, thoughtful leadership, multidisciplinary collaboration, and... a healthy dose of luck!