



An Artificial Intelligence Enabled System for Retinal Nerve Fiber Layer Thickness Damage Severity Staging

Siamak Yousefi, PhD,^{1,2} Xiaoqin Huang, PhD,¹ Asma Poursoroush, MSc,¹ Julek Majoor, PhD,³ Hans Lemij, MD,³ Koen Vermeer, PhD,³ Tobias Elze, PhD,⁴ Mengyu Wang, PhD,⁴ Kouros Nouri-Mahdavi, MD, MPH,⁵ Vahid Mohammadzadeh, MD,⁵ Paolo Brusini, MD,⁶ Chris Johnson, PhD⁷

Purpose: To develop an objective glaucoma damage severity classification system based on OCT-derived retinal nerve fiber layer (RNFL) thickness measurements.

Design: Algorithm development for RNFL damage severity classification based on multicenter OCT data.

Subjects and Participants: A total of 6561 circumpapillary RNFL profiles from 2269 eyes of 1171 subjects to develop models, and 2505 RNFL profiles from 1099 eyes of 900 subjects to validate models.

Methods: We developed an unsupervised *k*-means model to identify clusters of eyes with similar RNFL thickness profiles. We annotated the clusters based on their respective global RNFL thickness. We computed the optimal global RNFL thickness thresholds that discriminated different severity levels based on Bayes' minimum error principle. We validated the proposed pipeline based on an independent validation dataset with 2505 RNFL profiles from 1099 eyes of 900 subjects.

Main Outcome Measures: Accuracy, area under the receiver operating characteristic curve, and confusion matrix.

Results: The *k*-means clustering discovered 4 clusters with 1382, 1613, 1727, and 1839 samples with mean (standard deviation) global RNFL thickness of 58.3 (8.9) μm , 78.9 (6.7) μm , 87.7 (8.2) μm , and 101.5 (7.9) μm . The Bayes' minimum error classifier identified optimal global RNFL values of $> 95 \mu\text{m}$, 86 to 95 μm , 70 to 85 μm , and $< 70 \mu\text{m}$ for discriminating normal eyes and eyes at the early, moderate, and advanced stages of RNFL thickness loss, respectively. About 4% of normal eyes and 98% of eyes with advanced RNFL loss had either global, or ≥ 1 quadrant, RNFL thickness outside of normal limits provided by the OCT instrument.

Conclusions: Unsupervised machine learning discovered that the optimal RNFL thresholds for separating normal eyes and eyes with early, moderate, and advanced RNFL loss were 95 μm , 85 μm , and 70 μm , respectively. This RNFL loss classification system is unbiased as there was no preassumption or human expert intervention in the development process. Additionally, it is objective, easy to use, and consistent, which may augment glaucoma research and day-to-day clinical practice.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science* 2024;4:100389 © 2023 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/>).

Glaucoma is a heterogeneous group of disorders that represents the second leading cause of blindness overall, affecting up to 91 million individuals worldwide.^{1,2} As an optic neuropathy, glaucoma is characterized by distinct structural and functional changes that ultimately may impact patients' vision-related quality of life.³ Accumulating evidence suggests that detectable glaucomatous axonal loss as measured by OCT precedes detectable visual field (VF) impairment.^{4,5} As OCT can quickly provide objective, 3-dimensional, depth-encoded information from the retina and optic disc, it has recently become one of the most frequently used tools in clinical practice to assess glaucoma and monitor its progression.

Accurate staging of the severity of glaucoma-induced damage is an important component of guiding glaucoma

management. It can provide prognostic information for therapy adjustment based on the type and extent of damage. It can also enhance both the monitoring of progression and evaluation of the treatment efficacy, thereby improving prognosis plans and maintaining vision-related quality of the life for patients.⁶ It could essentially establish a common ground for glaucoma research and clinical practice.⁶

Currently available OCT instruments provide some quantitative data and visualizations for determining if a given measurement is outside the range of normal retinal nerve fiber layer (RNFL) thickness measurements. However, OCT instruments do not currently provide a severity scale for RNFL loss. Thus, staging the severity of structural damage based on OCT has remained largely subjective with poor agreement among experts.^{7,8} It is our aim to a develop

an improved staging system based on unsupervised and supervised machine learning models applied to OCT-derived RNFL thickness measurements, that will be unbiased, objective, consistent, reproducible, clinician-friendly, and most importantly, easy to use.⁶ Unlike many machine learning models that require expert-annotated datasets, our models require no labeled data and include no human expert intervention, thus generating unbiased and objective outcomes.

Numerous approaches have been proposed for glaucoma damage staging and disease characterization based on VF tests only.^{9–17} In contrast, methods for staging glaucoma damage based on OCT are rare.^{18,19} An OCT-based glaucoma staging system was proposed by Brusini¹⁸ in which superior and inferior quadrant average RNFL thickness measurements were used to stage glaucoma.¹⁸ Brusini also suggested another model for staging glaucoma based on RNFL profiles derived from scanning laser polarimeter (GDx) instrument.¹⁹ However, this model was not extended to OCT-derived RNFL profiles.

Machine learning models of artificial intelligence systems have shown great promise for addressing different challenges in glaucoma, although most of these models have been applied to glaucoma diagnosis or progression detection.^{20–29} In this paper, we propose unsupervised and supervised machine learning models to stage structural RNFL loss. Our system may be used to gain some insights to glaucoma when it is interpreted along with other potential ocular comorbidities that may impact RNFL.

Methods

Subjects and Data

Discovery Dataset. The discovery dataset included 6561 reliable OCT circle scans from 2269 eyes of 1171 patients (who visited our glaucoma clinic) collected from the Spectralis instrument (Heidelberg Engineering). We excluded OCT profiles of visits that were collected less than a year apart from the previous visit. This dataset was used to develop the RNFL damage severity classification system.

Independent Validation Dataset. The validation dataset included 2505 reliable OCT circle scan profiles collected from 3 different institutes. A total of 691 OCT profiles from 691 eyes of 691 subjects who visited Mass Eye and Ear glaucoma service, 154 OCT profiles from 154 eyes of 82 normal subjects who participated in the Advanced Glaucoma Intervention Study at the University of California, Los Angeles, and 1660 OCT profiles from 254 eyes of 127 glaucoma patients who visited the Rotterdam Eye Hospital in the Netherlands, were collected. Details of the subjects' characteristics are presented in Table 1.

We received institutional review board approval to perform this secondary data analysis study and were compliant with the tenets of the Declaration of Helsinki. This was primarily a secondary de-identified data analysis and thus was exempt from obtaining consent. However, for one small subset of data from the University of California, Los Angeles center, consent had been previously obtained from the patients.

Overview of the Pipeline

Figure 1 shows the diagram of the RNFL damage severity classification system. We first developed an unsupervised

clustering approach based on *k*-means to find clusters with similar OCT profiles by inputting 64 averaged sectors, 6 general sectors, and 1 global RNFL parameter. We then investigated the optimal number of clusters objectively. To assure whether the clustering step was stable, and whether the clusters were reproducible, we repeated the clustering several times, each time randomly selecting a subset of OCT profiles (without replacement) and assessing cluster memberships. We then performed post hoc analyses to label statistical clusters and generate clinical clusters based on means of global RNFL thickness values of each cluster. We then computed the optimal global RNFL thresholds that separated the clusters with the highest accuracy (i.e., minimum error classifier based on Bayes theorem which is equivalent to highest area under the receiver operating characteristic curve) and generated the new clusters (we called them clinical clusters). We further verified the reproducibility of the global RNFL thresholds based on an independent validation dataset. We then established an objective RNFL damage severity classification system based on the ascertained clinical clusters and identified global RNFL thresholds. As OCT data for developing this pipeline were primarily collected from glaucoma clinics, this model may be used to stage RNFL loss in patients with glaucoma, provided considering the impact of other comorbidities such as myopia or macular edema on RNFL thickness.

Preprocessing

We excluded all OCT profiles in which the signal strength was < 15 based on the vendor recommendation. OCT segmentation was manually evaluated at each contributing center to exclude scans with segmentation error. To generate 64 RNFL sectors, we averaged the segmented RNFL of every 12 A-scans as the initial profile of circle scans is composed of 768 A-scans (Fig 1; a ring with 64 local sectors around the optic disc). Averaging segmented A-scans (provided by the Spectralis Software) reduces the effect of variation due to possible imaging misalignment or previous anatomic differences in different eyes.²⁸ The input to the unsupervised model included 64 RNFL sectors along with 7 instrument-generated general sectoral and global RNFL thickness measurements.

Unsupervised Machine Learning

We developed a *k*-means unsupervised clustering³⁰ to partition eyes to *k* clusters by dividing data in the 71-dimensional space into disjoint groups by employing an objective function that minimizes the distances within RNFL profiles in each group while maximizing the distance among clusters.³¹ We then used the Silhouette metric to objectively identify the optimal number of clusters.³²

To verify that our datasets were representative, we included eyes across the glaucoma continuum, particularly eyes from both ends of the glaucoma spectrum in terms of RNFL loss, to assure an unbiased and reproducible analysis. We visualized the RNFL loss distribution in both discovery and validation datasets in terms of global RNFL thickness to assure input-data representativeness (Fig 2).

We further assessed the stability of clustering to assure that clusters were reproducible. We therefore selected subsets of RNFL in the discovery dataset randomly and repeated *k*-means clustering several times and computed the membership accuracy. We then computed the overall percentage agreement (membership accuracy) as the percentage of the eyes that were consistently assigned to same cluster.

We performed a post hoc analysis to annotate and assign clinical labels to the statistical clusters. We computed the global RNFL thickness values of the eyes in each cluster to identify levels of RNFL loss. We then labeled the clusters as normal (cluster with highest mean global RNFL thickness), early, moderate, and

Table 1. Characteristics of Global and Sectoral RNFL Thickness Values of Eyes Used in this Study (SD).

Parameters	Discovery Dataset	Validation Dataset
Number of OCT visits	6561	2505
Number of subjects	1171	900
Number of eyes	2268	1099
Number of visits, mean (SD)	1.9 (\pm 1.95)	1.3 (\pm 2.7)
Length of follow-up visits (yrs), mean (SD)	2.2 (\pm 1.9)	1.2 (\pm 2.4)
Global RNFL thickness (μ m), mean (SD)	83.2 (\pm 17.2)	74.9 (\pm 20.1)
Age (yrs), mean (SD)	64.5 (\pm 13.2)	63.0 (\pm 12.3)

RNFL = retinal nerve fiber layer; SD = standard deviation.

advanced (cluster with lowest mean global RNFL thickness) stages of RNFL loss based on their global RNFL thickness values. It is worth mentioning that in the absence of RNFL-impacting comorbidities such as myopia or macular edema, this model can stage glaucoma severity based on the RNFL thickness.

Glaucoma Damage Severity Classification System Based on RNFL Thickness Profiles

Our unsupervised machine learning model can be used as an RNFL (and glaucoma) damage severity staging system; however, a major weakness is its dependency on a machine learning system, which in turn is considered not user-friendly by clinicians, and thus unlikely to receive significant acceptance and widespread clinical utility, especially in a very busy medical setting.

We therefore developed another supervised machine learning model to identify global RNFL thickness thresholds that discriminate statistical clusters (different RNFL/glaucoma severity levels) with highest accuracy. We employed Bayes' minimum error classifier to identify optimal global RNFL thresholds that discriminated statistical clusters with minimum error.³³ We then identified the new "clinical clusters" based on the determined global RNFL thresholds. We also adjusted the identified thresholds based on the mean of global RNFL thinning due to normal aging to make this model applicable for glaucoma staging as well. More specifically, we used the statistics from the normative database of the Spectralis instrument. The mean age of subjects in the Spectralis normative database was 48.2 years and the mean rate of global RNFL thinning due to aging was $-0.075 \mu\text{m}/\text{year}$. The new thresholds were subsequently calculated as the old thresholds minus the difference of the mean cluster age and normative database age multiplied by the rate.

There was no expert intervention in determining clusters, identifying the optimal number of clusters, recognizing global RNFL thresholds with highest accuracy, or establishing clinical clusters. Using global RNFL thresholds provides a simple and easy-to-use glaucoma damage severity classification system that maximizes its clinical utility. We also compared the identified global RNFL thresholds of the proposed staging system against previously developed RNFL-based classification systems.

Validation of the Pipeline

We merged 3 independent datasets from the Rotterdam Eye Hospital, the University of California, Los Angeles, and Mass Eye and Ear to generate a dataset that is as representative as possible of a general clinical population. The independent dataset included 2505 RNFL profiles and was used to validate findings. We first investigated whether we could discover (reproduce) statistical clusters with similar properties we identified based on the discovery dataset. We thus applied the k -means clustering (using same parameters) on the validation dataset and identified statistical clusters. We then evaluated the optimal number of clusters using the Silhouette metric, as was used for the discovery dataset. We then labeled clusters based on the mean of global RNFL of OCT profiles in clusters and compared corresponding clusters based on the discovery and validation datasets. We then identified the RNFL thresholds that discriminate clinical clusters based on the validation dataset and compared them against RNFL thresholds that we had previously identified based on the discovery dataset.

We used generalized estimating equations (GEEs)³⁴ to compare different characteristics of subjects, such as age, in the discovery and validation datasets. We also used GEEs to compare corresponding clusters that were identified based on the discovery and

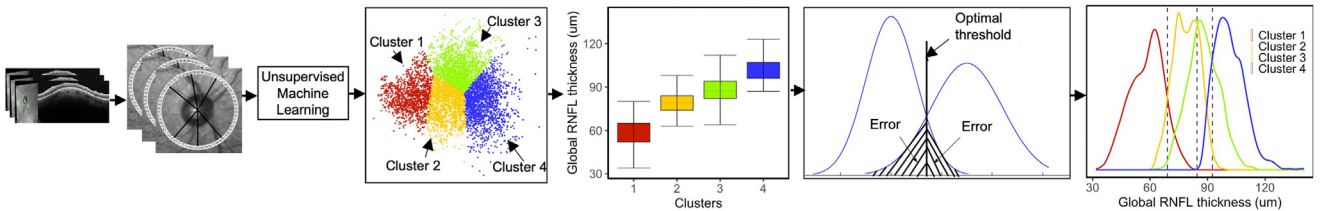


Figure 1. Diagram of the pipeline. Circumpapillary retinal nerve fiber layer (RNFL) thickness in profiles with 768 A-scans were averaged to generate 64 sectors. All 64 RNFL sectors along with general sectoral and global measurements were input to the unsupervised clustering algorithm to identify statistical clusters. Clinical labels were assigned to clusters based on mean global RNFL thickness of eyes in each cluster. Bayes' minimum error classifier was used to identify optimal RNFL thresholds for the RNFL damage severity classification system. The 4 identified clusters are presented in blue, green, orange, and red that correspond to normal eyes and eyes at the early, moderate, and advanced stages of RNFL loss, respectively.

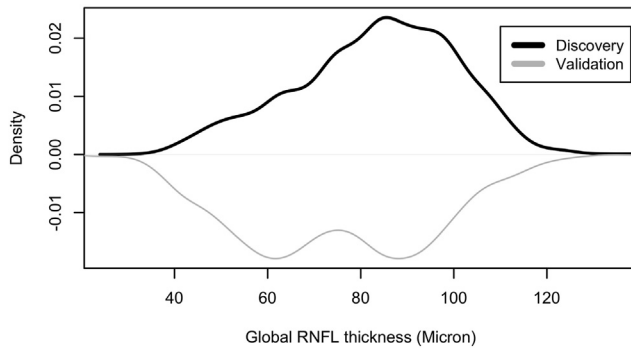


Figure 2. Distribution of global retinal nerve fiber layer (RNFL) thickness measurement in the discovery dataset (black curve; top) and the independent validation dataset (gray curve; bottom).

independent datasets, as we had included both eyes of some of the subjects in datasets. Machine learning and statistical analyses were performed in Python 3.8 and R (version 4.0.3) platforms.

Results

The demographic characteristics of the subjects are presented in Table 1. Briefly, the discovery dataset included 6561 OCT profiles from 2268 eyes of 1171 subjects with a mean (standard deviation [SD]) age of 64.5 (13.2) years and a mean global RNFL thickness of 83.2 (17.2) μm . The independent validation dataset included 2505 OCT profiles from 1099 eyes of 900 subjects with a mean age of 63.0 (12.3) years and a mean global RNFL thickness of 74.9 (20.1) μm . Figure 2 shows the global RNFL thickness distributions of eyes in the discovery and validation datasets. Eyes in the discovery dataset had greater global RNFL thickness compared with eyes in the validation dataset ($P < 0.01$; GEE model). The difference between the mean age of subjects in the discovery and validation datasets was only ~ 1.5 years, but it was statistically significant ($P < 0.01$; GEE model).

Statistical Clusters Based on the Discovery Dataset

Based on the discovery dataset, Silhouette suggested that the optimal number of statistical clusters was 4. However, we performed clustering based on 3 clusters because the knee-plot suggested 3 or 4 clusters (confusing knee-plot curves). Figure 3 shows the scatter plot of the first and second principal components of the RNFL thickness values in 4 statistical clusters. On average, $> 98.0\%$ of the eyes were assigned to the same cluster based on repeated k -means clustering applied 5 times, with each time randomly selecting different subsets of the discovery dataset without replacement (Table 2). Thus, we were able to confirm that the clusters were stable and reproducible.

We annotated statistical clusters by assigning normal, early, moderate, and advanced labels to clusters based on their respective global RNFL thickness values. The mean (SD) global RNFL thickness in normal, early, moderate, and

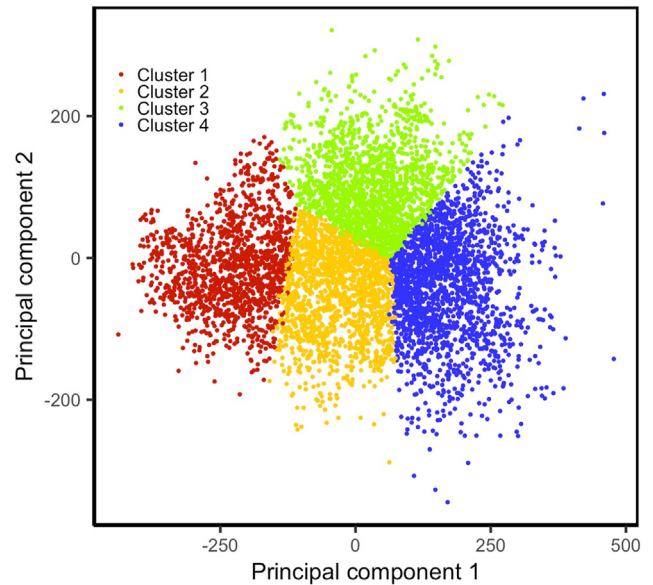


Figure 3. Scatter plot of the statistical clusters. Principal component analysis (PCA) was applied on retinal nerve fiber layer profiles to reduce the dimensionality from 64 dimensions to 2 dimensions for visualization. The eyes in the statistical clusters are grouped together in the PCA space. Colors blue, green, orange, and red correspond to normal eyes and eyes at the early, moderate, and advanced stages of glaucoma, respectively.

advanced clusters was 101.5 (7.9) μm , 87.8 (8.2) μm , 78.9 (6.7) μm , and 58.3 (8.9) μm , respectively. The number of OCT profiles in these 4 statistical clusters was 1839, 1727, 1613, and 1382, respectively.

Based on the statistical clusters, the proportion of eyes with global RNFL thickness, or ≥ 1 out of 6 general sectoral regions, being outside the normal limit (ONL) in the normal, early, moderate, and advanced clusters was 4% (83 OCT profiles), 17% (293 OCT profiles), 58% (940 OCT profiles), and 98% (1357 OCT profiles), respectively.

Clinical Clusters and RNFL Damage Severity Classification System Based on the Discovery Dataset

The mean age (SD) of the subjects in normal, early, moderate, and advanced clusters was 58.3 (15.0), 63.8 (12.0), 67.4 (11.4), and 68.7 (11.3) years, respectively. The Bayes' minimum error classifier identified global RNFL thickness of 95 μm , 85 μm , and 70 μm as optimal thresholds for discriminating normal, early, moderate, and advanced stages of glaucoma (Fig 4). The adjusted thresholds due to normal aging were 96 μm , 86.5 μm , and 71.5 μm ; these were optimal thresholds for discriminating normal, early, moderate, and advanced stages of RNFL loss (or glaucoma). The number of OCT profiles in 4 clinical clusters (based on the identified RNFL thresholds) was 1715, 1453, 1910, and 1483, respectively.

Based on the established global RNFL thresholds, the proportion of eyes with global RNFL thickness, or ≥ 1 out

Table 2. Membership Accuracy of the Clustering Based on 10-times Repeating k -Means on Randomly Selected (without Replacement) Subsets of the Discovery Dataset

Percentage of Discovery Dataset (%)	Mean	Range
90	0.994	[0.976–0.999]
80	0.992	[0.978–0.996]
70	0.990	[0.983–0.995]
60	0.982	[0.959–0.995]
40	0.962	[0.884–0.991]

of 6 general sectoral regions, being ONL was about 3% (48 OCT profiles), 13% (187 OCT profiles), 51% (981 OCT profiles), and 98% (1457 OCT profiles), respectively.

Validating Findings Based on the Validation Dataset

The mean age (SD) of the subjects in normal, early, moderate, and advanced clusters was 55.9 (13.3), 59.2 (12.4), 65.0 (11.6), and 66.0 (10.6) years, respectively. Based on the validation dataset with 2505 OCT profiles, we observed 4 clusters that were optimum according to the Silhouette and knee-plot. We then annotated the statistical clusters by assigning normal, early, moderate, and advanced labels to clusters based on their respective global RNFL thickness values. The mean (SD) global RNFL thickness of eyes in normal, early, moderate, and advanced clinical clusters was 103.8 (9.2) μm , 86.6 (6.6) μm , 80.7 (7.9) μm , and 56.2 (9.4) μm , respectively. The number of OCT profiles in normal, early, moderate, and advanced clusters was 422, 474, 502, and 1107, respectively. The Bayes' minimum error classifier identified global RNFL thickness values of 93 μm , 87 μm , and 70 μm as optimal thresholds for discriminating normal, early, moderate, and advanced stages of glaucoma.

In comparison to initial statistical clusters, the mean accuracy of clinical clusters to discriminate normal eyes and eyes at the early, moderate, and advanced stages of structural RNFL loss, based solely on global RNFL thickness measurement, was approximately 78%. Figure 5 shows the confusion matrix of the classification system (clinical clusters) based on the identified global RNFL thresholds.

The adjusted thresholds due to normal aging were 94 μm , 88.5 μm , and 71.5 μm as optimal thresholds for discriminating normal, early, moderate, and advanced stages of glaucoma. The identified mean global RNFL thickness values in the statistical clusters and the identified global RNFL threshold values based on the validation dataset were supportive of the identified factors from the discovery dataset.

Discussion

We developed unsupervised and supervised machine learning models to identify the level of structural severity based on OCT data without human expert intervention. The unsupervised k -means model discovered 4 clusters with similar OCT profiles that were labeled to normal, early, moderate, and advanced stages based on their respective mean global RNFL thickness values. Rather than proposing a glaucoma damage severity staging system based on the initial k -means clustering (which is complex and thus unlikely to be used by clinicians), we identified clinical clusters corresponding to normal eyes and eyes at early, moderate, and advanced stages of structural loss based on global RNFL thickness values. Specifically, we developed a supervised Bayes' minimum error classifier to identify global RNFL thresholds that could discriminate clusters with highest accuracy (minimum error). We found that global RNFL thickness values of 95 μm , 85 μm , and 70 μm were optimal thresholds for discriminating normal, early,

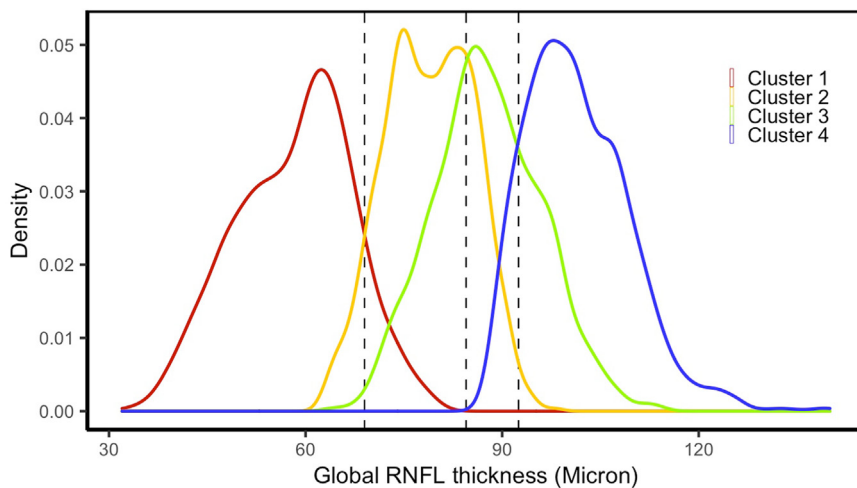


Figure 4. Global retinal nerve fiber layer (RNFL) thickness profiles in 4 statistical clusters based on the discovery dataset. The optimal RNFL thresholds (dashed lines) were identified based on the Bayes' minimum error classifier. Colors blue, green, orange, and red correspond to the RNFL distribution of normal eyes and eyes at the early, moderate, and advanced stages of glaucoma, respectively.

	Class	Normal	Early	Moderate	Advanced
Predicted Class	Normal	401	68	22	0
	Early	21	176	88	0
	Moderate	0	277	340	41
	Advanced	0	3	52	1066
True Class					

Figure 5. Confusion matrix of the proposed glaucoma damage severity staging system based on global retinal nerve fiber layer thresholds compared with the initial statistical clusters that were derived by unsupervised machine learning model only.

moderate, and advanced stages of RNFL loss without considering the impact of age. The age-adjusted thresholds based on insights from the Spectralis normative database deviated only 1 to 1.5 μm from these thresholds. As the mean age of subjects in Spectralis database was 48.8 years, the mean RNFL of new subjects would need to be age-adjusted then compared against our severity staging system. It is worth mentioning that all development steps were unbiased as there was no expert intervention in identifying RNFL thresholds.

The first step of our algorithm included an unsupervised clustering based on k -means. As there is no definite metric to evaluate whether the outcome of the unsupervised clustering model is stable or not, we repeated the clustering algorithm several times, each time randomly selecting a subset of samples and recording the class memberships (Table 2). We observed a high level of membership accuracy ($> 96\%$ in all evaluations) based on 5 subsets of datasets and 10 repeats of the clustering algorithm for each subset. We then assessed whether the number of clusters was optimum based on Silhouette metric and knee-plot visualization.³² The Silhouette metric and knee-plot suggested that the optimal number of clusters was 4. We thus performed clustering based on 4 clusters corresponding to 4 severity levels.

There are several glaucoma staging systems based on VFs that have been derived based on fully or partially subjective criteria with expert knowledge integration.^{9,11,13–16,35,36} For instance, the widely used Hodapp-Parrish-Anderson VF staging system suggests several mean deviation thresholds, along with assessment of several VF test points including those within the central 5 degrees to discriminate early, moderate, and advanced stages of glaucoma.⁹ The basis for these proposed thresholds, however, is mostly subjective. Additionally, it is highly challenging to calculate test point computations in the day-to-day clinical setting. We, however, employed unbiased, unsupervised machine learning along with objective supervised Bayes statistical analysis to identify global RNFL thresholds that determine the severity levels of structural loss. We identified that the global RNFL thickness threshold for discriminating normal eyes is about 95 μm ; the global RNFL thickness threshold for discriminating early and moderate stages of RNFL loss is approximately 85 μm ; and the global RNFL thickness threshold for identifying advanced stages of RNFL loss is about 70 μm . A previous study on the ability of OCT to discriminate different stages

of glaucoma found that global RNFL thickness thresholds of 72.5 μm and 97.5 μm discriminated early, moderate and advanced stages of glaucoma (3 stages).³⁷ As OCT data for developing our model were primarily collected from glaucoma clinics, this model may be used to stage the severity level of RNFL loss in patients with glaucoma, however, physicians would need to be aware of other, potentially RNFL-impacting comorbidities, such as myopia and macular edema, when interpreting findings. Although our model could discriminate 4 stages of glaucoma severity, our proposed threshold for discriminating normal and early stage of glaucoma (95 μm) is close to the threshold suggested in this previous study.

Medeiros et al³⁸ proposed a combined structure-function index to detect glaucoma and to stage glaucoma according to 3 levels of normal, preperimetric, and perimetric glaucoma. This index is derived based on complex equations to calculate VF-derived and OCT-derived estimates of the total number of retinal ganglion cells by analyzing multiple parameters including age, mean deviation, global RNFL thickness, and VF sensitivity expressed in decibels (dB) at different eccentricities. The area under the receiver operating characteristic curve for discriminating early from moderate glaucoma, and moderate from advanced glaucoma, were 0.94 and 0.96, respectively. We, in contrast, propose 3 global RNFL thresholds to distinguish 4 different stages of glaucoma. Moreover, rather than offering a complex model that uses several parameters that make the system complex, we propose only a widely used and available index of global RNFL thickness value to stage glaucoma.

Brusini proposed a glaucoma staging system based on the average RNFL thickness in the superior and inferior quadrants to stage glaucoma into 6 severity levels.¹⁸ The reported sensitivity and specificity for discriminating normal eyes from eyes with glaucoma were 95.2% and 91.9%, respectively. Although the Nidek RS-3000 normative dataset has been used to delineate the curvilinear lines to discriminate normal from borderline and borderline from abnormal, the rest of the curvilinear lines are specified by arbitrary assumptions, based on clinical assessment of the optic disc, using the Optic Disc Damage Staging System³⁹ that was previously developed by the author, and additional data from both Glaucoma Detection with Variable Corneal Compensation (GDx VCC) and Heidelberg Retina Tomograph. This OCT Glaucoma Staging System includes 6 nonlinear equations, each with 3 parameters. Unfortunately, the complexity of this model decreases the

likelihood it would be readily incorporated into clinical use. However, the author has developed a software that requires inputting RNFL thickness values for the superior and inferior quadrant. The issue then becomes downloading and using a third-party software for staging glaucoma.

Although the staging systems proposed by Medeiros, Brusini et al, Mills et al, and others^{14–16,36} used a few hundred eyes to develop and test models, we used large datasets with > 7000 OCT profiles to develop and validate models independently. Unlike previously developed OCT-based staging systems, ours uses global RNFL thickness as the sole parameter. We have several reasons for using a single parameter. First, we intend to provide an easy-to-use approach that is ultimately suited for use in day-to-day clinical practice and glaucoma research. Second, global RNFL is a widely used parameter that is readily available, in contrast to fine sectoral/voxel analysis, local RNFL thickness assessments, and additional criteria based on parameters that are not necessarily readily available. Third, it is not labor intensive, and thus makes staging simple for clinical use.

Our data suggest that sole, unsupervised *k*-means clustering can be used to create a glaucoma damage severity staging system. Why did we bother to develop a second machine learning model based on Bayes theorem to identify global RNFL thickness thresholds alone? The answer is that although a glaucoma staging system based on the unsupervised *k*-means clustering mode may be more accurate, it would utilize a complex mapping model applied to 71 RNFL parameters, an operation unlikely to be routinely utilized in a clinical setting. However, our simplified glaucoma damage severity staging system is based on global RNFL thickness measurements that are readily available to clinicians as well as researchers and easy to work with. Based on initial *k*-means clustering, we observed that the proportion of normal eyes with global RNFL thickness or ≥ 1 general sectoral region being ONL was 4%, although this proportion dropped to 3% for the staging system based on global RNFL thickness thresholds. Moreover, based on both the initial *k*-means clustering and the proposed staging system based on global RNFL thresholds, 98% of the eyes in the advanced stages of glaucoma had global or ≥ 1 general sectoral region ONL. We are simply complying with the Occam's razor principle that suggests the superiority of a simple compared with a complex model, if the accuracy is not significantly compromised.⁴⁰

Several studies, including ours, have proposed unsupervised and supervised machine learning approaches for detecting glaucoma progression based on OCT data or identifying patterns of RNFL loss in patients with glaucoma.^{28,41,42} In this study, however, we developed unsupervised and supervised machine learning models to stage glaucoma based on OCT. Although glaucoma progression and staging are both important clinical measurements, establishing glaucoma damage stage may be particularly critical and valuable in determining treatment options and even detecting progression. We used a large discovery dataset to develop models and employed another independent validation dataset to validate models. We showed that the identified global RNFL thresholds based

on the validation datasets supported the identified global RNFL thresholds based on the discovery dataset. The small discrepancy between global RNFL threshold values derived from the discovery and validation datasets may be explained by the fact that the number of eyes at the moderate to advanced stages of glaucoma were different in the discovery and validation datasets (Fig 2). Moreover, this discrepancy is somewhat universal to many studies and not specific to our study as even the summary statistics of the RNFL data in normative datasets of different instruments could also be different.

Most glaucoma staging methods have limited clinical utility for several major reasons. First, models may be simple and easy to use but may also be somewhat subjective, not standardized, and poorly reproducible. Second, models may be more accurate and quite standardized but may also be too complicated and time-consuming to be utilized day-to-day in a clinical setting.⁶ Third, models may be developed based on a limited number of OCT samples. We, however, proposed a glaucoma damage severity classification system based on large numbers of OCT samples that is simple, easy to use, clinician-friendly, and quick, yet precise enough to be used in glaucoma research and clinical practice. We ultimately suggest using both OCT and VF¹⁷ staging systems to obtain a fuller portrayal of glaucoma severity.

Our study does have several limitations as well. First, both datasets were retrospective, thus providing limited information on true progression of eyes to higher severity levels. Second, the datasets were collected from glaucoma clinics with no information regarding possible comorbidities such as cataract, myopia, macula edema, or surgical history, which may impact findings. Third, OCT profiles were collected from the Spectralis instrument only; thus, data from other vendors would be desirable to verify findings. Fourth, use of the global RNFL thickness parameter only for RNFL damage severity classification in patients with glaucoma may miss those with early stage localized RNFL loss. Fifth, we were unable to collect VFs from corresponding eyes to assess VF severity of eyes in different clusters. Finally, our data may not be fully representative of the cases encountered in some clinical practices, although the use of a validation dataset from multiple centers minimizes this issue. Future studies are desirable to further investigate and address these limitations.

In conclusion, we developed a glaucoma damage severity classification system based on unsupervised and supervised machine learning models that is unbiased and objective. We discovered 4 clusters of RNFL profiles, evaluated the quality of learning based on several objective metrics, and evaluated the reproducibility and optimal number of clusters. We then employed supervised learning, based on Bayes' minimum error criteria, to identify optimal global RNFL thresholds that define 4 severity levels of glaucoma. Our structural damage severity classification system suggested global RNFL thickness values of 95 μm , 85 μm , and 70 μm are optimal thresholds for discriminating normal, early, moderate, and advanced stages of RNFL loss. The proposed RNFL damage severity classification system is simple, easy to use, clinician-friendly, and unbiased. It is

based on a readily available OCT global parameter and uses simple thresholds; thus, it may augment glaucoma research

and clinical practice to gain insight to glaucoma-induced RNFL loss.

Footnotes and Disclosures

Originally received: October 7, 2022.

Final revision: August 8, 2023.

Accepted: August 18, 2023.

Available online: August 23, 2023. Manuscript no. XOPS-D-22-00212R1.

¹ Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, Tennessee.

² Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee.

³ Rotterdam Ophthalmic Institute, The Rotterdam Eye Hospital, Rotterdam, The Netherlands.

⁴ Schepens Eye Research Institute of Massachusetts Eye and Ear, Harvard Medical School, Boston, Massachusetts.

⁵ Department of Ophthalmology, University of California Los Angeles, Los Angeles, California.

⁶ Department of Ophthalmology, “Città di Udine” Health Center, Udine, Italy.

⁷ Department of Ophthalmology & Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, Iowa.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The authors have made the following disclosures: S.Y.: Consulting – GlobeCheck; Received instruments – M&S Technologies, Remidio; Co-founder – AEye.

D.J.: Personal fees – M&S Technologies.

C.J.: Personal fees – M&S Technologies.

The other authors have no proprietary or commercial interest in any materials discussed in this article.

This work was supported by NIH Grants EY033005, EY031725, and a Challenge Grant from Research to Prevent Blindness (RPB), New York.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

HUMAN SUBJECTS: Human subjects were included in this study. We received institutional review board approval to perform this secondary data analysis study. This research adhered to the tenets of the Declaration of Helsinki. This was primarily a secondary de-identified data analysis and thus was exempt from obtaining consent. However, for one small subset of data from the University of California, Los Angeles center, consent had been previously obtained from the patients.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Yousefi, Huang, Poursorouh, Brusini, Johnson

Data collection: Yousefi, Huang, Majoor, Lemij, Vermeer, Elze, Wang, Nouri-Mahdavi, Mohammadzadeh, Brusini, Johnson

Analysis and interpretation: Yousefi

Obtained funding: Yousefi

Overall responsibility: Yousefi, Brusini, Johnson

Abbreviations and Acronyms:

GEE = generalized estimating equation; **ONL** = outside the normal limit; **RNFL** = retinal nerve fiber layer; **SD** = standard deviation; **VF** = visual field.

Keywords:

Artificial intelligence, Glaucoma, Glaucoma severity damage, Optical coherence tomography, Retinal nerve fiber layer (RNFL), Staging, Unsupervised machine learning.

Correspondence:

Siamak Yousefi, PhD, University of Tennessee Health Science Center, 930 Madison Ave., Suite 726, Memphis, TN 38163. E-mail: siamak.yousefi@uthsc.edu

References

1. Quigley HA. Number of people with glaucoma worldwide. *Br J Ophthalmol*. 1996;80:389–393.
2. Goldberg I. How common is glaucoma worldwide? In: Weinreb RN, Kitazawa Y, Krieglstein GK, eds. *Glaucoma in the 21st Century*. London: Mosby International; 2000:3–8.
3. Foster PJ, Buhmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol*. 2002;86:238–242.
4. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol*. 1989;107:453–464.
5. Sample PA. Glaucoma is present prior to its detection with standard automated perimetry: is it time to change our concepts? *Graefes Arch Clin Exp Ophthalmol*. 2003;241:168–169.
6. Brusini P, Johnson CA. Staging functional damage in glaucoma: review of different classification methods. *Surv Ophthalmol*. 2007;52:156–179.
7. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc*. 1976;74:532–572.
8. Jampel HD, Friedman D, Quigley H, et al. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol*. 2009;147:39–44.e1.
9. Hodapp E, Parrish RK, Anderson DR. *Clinical Decisions in Glaucoma*. CV Mosby; 1993:52–61.
10. Advanced glaucoma intervention study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994;101:1445–1455.
11. Brusini P. Clinical use of a new method for visual field damage classification in glaucoma. *Eur J Ophthalmol*. 1996;6:402–407.
12. Brusini P, Filacorda S. Enhanced Glaucoma Staging System (GSS 2) for classifying functional damage in glaucoma. *J Glaucoma*. 2006;15:40–46.
13. Mills RP, Budenz DL, Lee PP, et al. Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease. *Am J Ophthalmol*. 2006;141:24–30.
14. Flammer J. The concept of visual field indices. *Graefes Arch Clin Exp Ophthalmol*. 1986;224:389–392.
15. Heijl A, Lindgren G, Olsson J. A package for the statistical analysis of visual fields. In: Greve EL, Heijl A, eds. *Seventh International Visual Field Symposium, Amsterdam, September 1986*. Springer Netherlands; 1987:153–168.

16. Bebie H, Flammer J, Bebie T. The cumulative defect curve: separation of local and diffuse components of visual field damage. *Graefes Arch Clin Exp Ophthalmol*. 1989;227:9–12.
17. Huang X, Saki F, Wang M, et al. An objective and easy-to-use glaucoma functional severity staging system based on artificial intelligence. *J Glaucoma*. 2022;31:626–633.
18. Brusini P. OCT Glaucoma Staging System: a new method for retinal nerve fiber layer damage classification using spectral-domain OCT. *Eye (Lond)*. 2018;32:113–119.
19. Brusini P. GDx staging system: a new method for retinal nerve fiber layer damage classification. *J Glaucoma*. 2011;20:287–293.
20. Bowd C, Weinreb RN, Balasubramanian M, et al. Glaucomatous patterns in Frequency Doubling Technology (FDT) perimetry data identified by unsupervised machine learning classifiers. *PLOS ONE*. 2014;9:e85941.
21. Yousefi S, Goldbaum MH, Balasubramanian M, et al. Learning from data: recognizing glaucomatous defect patterns and detecting progression from visual field measurements. *IEEE Trans Biomed Eng*. 2014;61:2112–2124.
22. Yousefi S, Goldbaum MH, Zangwill LM, et al. Recognizing patterns of visual field loss using unsupervised machine learning. *Proc SPIE Int Soc Opt Eng*. 2014;2014.
23. Yousefi S, Balasubramanian M, Goldbaum MH, et al. Unsupervised Gaussian mixture-model with expectation maximization for detecting glaucomatous progression in standard automated perimetry visual fields. *Transl Vis Sci Technol*. 2016;5:2. <https://doi.org/10.1167/tvst.5.3.2>.
24. Elze T, Pasquale LR, Shen LQ, et al. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J R Soc Interface*. 2015;12.
25. Wang M, Shen LQ, Pasquale LR, et al. Artificial intelligence classification of central visual field patterns in glaucoma. *Ophthalmology*. 2020;127:731–738.
26. Yousefi S, Elze T, Pasquale LR, et al. Monitoring glaucomatous functional loss using an artificial intelligence-enabled dashboard. *Ophthalmology*. 2020;127:1170–1178.
27. Thakur A, Goldbaum M, Yousefi S. Convex representations using deep archetypal analysis for predicting glaucoma. *IEEE J Transl Eng Health Med*. 2020;8:3800107.
28. Mahotra S, Wang M, Elze T, et al. Patterns of retinal nerve fiber layer loss in patients with glaucoma identified by deep archetypal analysis. *IEEE International Conference on Big Data*. 2020:3775–3782.
29. Nouri-Mahdavi K, Mohammadzadeh V, Rabiolo A, et al. Prediction of visual field progression from OCT structural measures in moderate to advanced glaucoma. *Am J Ophthalmol*. 2021;226:172–181.
30. Hartigan JA. *Clustering Algorithms*. John Wiley & Sons Inc.; 1975.
31. Telgarsky M, Vattani A. Hartigan’s method: k-means clustering without Voronoi. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. *PMLR*. 2020;9:820–827.
32. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comp Appl Math*. 1987;20:53–65.
33. Huang K, Yang H, King I, et al. The minimum error minimax probability machine. *J Mach Learn Res*. 2004;5:1253–1286.
34. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44:1049–1060.
35. Keltner JL, Johnson CA, Cello KE, et al. Classification of visual field abnormalities in the ocular hypertension treatment study. *Arch Ophthalmol*. 2003;121:643–650.
36. Shin Y, Suzumura H, Furuno F. Classification of glaucomatous visual field defects using the Humphrey Field Analyzer box plots. In: Mills RP, Heijl A, eds. *Perimetry Update*. Amsterdam: Kugler Publications; 1991:235–243.
37. Elbendary AM, Mohamed Helal R. Discriminating ability of spectral domain optical coherence tomography in different stages of glaucoma. *Saudi J Ophthalmol*. 2013;27:19–24.
38. Medeiros FA, Lisboa R, Weinreb RN, et al. A combined index of structure and function for staging glaucomatous damage. *Arch Ophthalmol*. 2012;130:1107–1116.
39. Brusini P, Zeppieri M, Tosoni C, et al. Optic disc damage staging system. *J Glaucoma*. 2010;19:442–449.
40. Schaffer J. What not to multiply without necessity. *Australas J Philos*. 2015;93:644–664.
41. Yousefi S, Goldbaum MH, Balasubramanian M, et al. Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points. *IEEE Trans Biomed Eng*. 2014;61:1143–1154.
42. Wang M, Shen LQ, Pasquale LR, et al. An artificial intelligence approach to assess spatial patterns of retinal nerve fiber layer thickness maps in glaucoma. *Transl Vis Sci Technol*. 2020;9:41.