

Check for updates

Classifying glaucoma exclusively with OCT: comparison of three clustering algorithms derived from machine learning

Marc Biarnés^{1,2,7}, Néstor Ventura-Abreu ^{3,4,5,7}, Ignacio Rodríguez-Una⁶, Francesc Franquesa-Garcia ³, Sofia Batlle-Ferrando³, María Teresa Carrión-Donderis ³, Rafael Castro-Domínguez³, Elena Millá^{3,4}, María Jesús Muniesa^{3,4} and Marta Pazos ^{3,4 ×}

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2023

BACKGROUND/AIMS: To objectively classify eyes as either healthy or glaucoma based exclusively on data provided by peripapillary retinal nerve fiber layer (pRNFL) and ganglion cell-inner plexiform (GCIPL) measurements derived from spectral-domain optical coherence tomography (SD-OCT) using machine learning algorithms.

METHODS: Three clustering methods (k-means, hierarchical cluster analysis -HCA- and model-based clustering-MBC-) were used separately to classify a training sample of 109 eyes as either healthy or glaucomatous using solely 13 SD-OCT parameters: pRNFL average and sector thicknesses and GCIPL average and minimum values together with the six macular wedge-shaped regions. Then, the best-performing algorithm was applied to an independent test sample of 102 eyes to derive close estimates of its actual performance (external validation).

RESULTS: In the training sample, accuracy was 91.7% for MBC, 81.7% for k-means and 78.9% for HCA (*p* value = 0.02). The best MBC model was that in which subgroups were allowed to have variable volume and shape and equal orientation. The MBC algorithm in the independent test sample correctly classified 98 out of 102 cases for an overall accuracy of 96.1% (95% CI, 92.3–99.8%), with a sensitivity of 94.3 and 100% specificity. The accuracy for pRNFL was 92.2% (95% CI, 86.9–97.4%) and for GCIPL 98.0% (95% CI, 95.3–100%).

CONCLUSIONS: Clustering algorithms in general (and MBC in particular) seem promising methods to help discriminate between healthy and glaucomatous eyes using exclusively SD-OCT-derived parameters. Understanding the relative merits of one method over others may also provide insights into the nature of the disease.

Eye; https://doi.org/10.1038/s41433-023-02785-5

INTRODUCTION

Glaucoma is one of the main causes of global irreversible blindness [1]. Still today, the gold standard for diagnosis is the glaucoma expert's assessment derived from integrating the visual field-testing results with the optic disc examination. However, relying only on the ophthalmologists' criteria could be suboptimal [2], since the agreement in the evaluation of optic disc glaucomatous damage [3] or progression [4] on fundus exam even among glaucoma specialists is not perfect. Moreover, in realworld conditions, the first diagnosis is often made by practice trainees and comprehensive ophthalmologists that have been shown to underestimate clinical findings [5], potentially leading to late diagnosis and referrals.

On the other hand, spectral-domain optical coherence tomography (SD-OCT) has become a game-changer in glaucoma diagnosis. It provides objective peripapillary retinal nerve fiber layer (pRNFL) and macular measurements (like Ganglion cell-inner plexiform layer -GCIPL-) comparing them with a normative database, achieving an excellent ability to discriminate glaucoma from healthy eyes [6]. In fact, adding SD-OCT data to the glaucoma specialist assessment has been shown to increase agreement between examiners [7], decreasing the number of unnecessary referrals from screening programs [8]. Although current knowledge suggests that combining different peripapillary and macular measurements further improves SD-OCT diagnostic accuracy [9], the best set of parameters is yet to be determined.

Could we find, then, a way to detect glaucoma patients only using SD-OCT parameters? One potential way could be clustering algorithms (CA). CA, derived from unsupervised machine learning, classifies patients according to a set of features using the clinical exam as the ground truth. They can group a set of eyes without supervision in a way that eyes in the same group (a *cluster*) are more similar to each other than to those in other groups/clusters. This is a common technique for statistical data analysis used in many fields including pattern recognition, image analysis, information retrieval, bioinformatics, and data compression.

Two of the most used clustering algorithms are heuristic and include k-means and hierarchical clustering analysis (HCA) [10]. In k-means the whole sample is separated into "k" groups pre-

¹Oftalmologia Mèdica i Quirúrgica (OMIQ) Research, Sant Cugat del Vallès, Spain. ²Institut de la Màcula (Hospital Quirón-Teknon), Barcelona, Spain. ³Institute of Ophthalmology. Hospital Clínic Barcelona, Barcelona, Spain. ⁴Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. ⁵Hospital Sagrat Cor, Barcelona, Spain. ⁶Instituto Oftalmológico Fernández-Vega. Fundación de Investigación Oftalmológica, University of Oviedo, Oviedo, Spain. ⁷These authors contributed equally: Marc Biarnés, Néstor Ventura-Abreu [©]email: pazos@clinic.cat

specified by the researcher that minimize within-cluster and maximize between-cluster variances. In (agglomerative) HCA, all patients begin as their own cluster and are gathered one at a time given their similarities, until they all form a single cluster. On the contrary, model-based clustering (MBC) is based on a statistical approach; with assumptions that can be tested considering that data must fit a certain model [11]. Each patient has a given probability of belonging to a specific cluster allowing a certain variation of means, variances, and shapes of clusters, which is probably more realistic.

The purpose of this study was to evaluate the performance of these three algorithms to objectively classify eyes as either healthy or glaucoma based exclusively on data provided by pRNFL and GCIPL measurements derived from SD-OCT.

METHODS

Study design

This retrospective cohort study included a training sample of consecutive patients attending the Glaucoma and Ophthalmology Primary Care Departments at Hospital Clínic Barcelona from June 2019 to June 2021. Similarly, a secondary test sample for external validation was obtained from patients visited consecutively from July 2021 to July 2022 in the Glaucoma and Ophthalmology Primary Care Departments of three different Spanish centres: Hospital Clínic Barcelona (Barcelona), Hospital Sagrat Cor (Barcelona), and Instituto Fernández-Vega (Oviedo).

The study protocol was approved by Hospital Clinic of Barcelona Ethics committee and adhered to the tenets of the Declaration of Helsinki. A waiver of the written informed consent was granted due to the retrospective nature of the study and because the data were collected from regular clinical practice.

Eligibility criteria and clinical exam

Inclusion criteria were as follows: age ≥ 18 years old, spherical equivalent within ± 5.0 diopters (D), astigmatism ≤ -3.0 D, best-corrected visual acuity (BCVA) $\geq 20/40$, and open-angle on gonioscopy. Exclusion criteria included corneal or retinal diseases, amblyopia, systemic or neurologic disorders that could affect test results, intraocular surgery other than uncomplicated phacoemulsification less than 6 months before the examination, and poor image quality (see OCT section). Only one eye from each patient was included in the study; if both eyes met the eligibility criteria, the study eye was selected with simple randomization (1:1 ratio).

All participants underwent a complete ophthalmic examination with BCVA, pachymetry, slit-lamp examination, Goldmann applanation tonometry, gonioscopy; and 24-2 SITA-Standard Automated Perimetry (SAP) with the Humphrey Field Analyzer, and Cirrus[®] high-definition OCT imaging (Carl Zeiss Meditec Inc., Dublin, CA, USA). Clinical examination and testing were performed within 2 months of each other. All patients had reliable SAP with fixation losses, false positives, and negatives <25%. Both sets of healthy subjects and glaucoma patients were examined by the three glaucoma specialists (MP, NVA, and IRU). Healthy subjects were recruited in the Ophthalmology Primary Care unit who came for an ordinary ophthalmic examination, patients who came for refraction, or requiring a medical certificate assuring that they did not present any ophthalmic disease. Healthy controls had a normal optic nerve head appearance, intraocular pressure (IOP) ≤ 21 mmHg, and a normal SAP. Glaucoma was defined as untreated IOP > 21 mmHg, glaucomatous optic disc appearance (neuroretinal rim thinning and/or notching and/or peripapillary hemorrhages) with a correspondent, reproducible visual field defect as described by Anderson et al. [12]. SD-OCT was not used to make the diagnosis.

Optical coherence tomography acquisition and analysis

SD-OCT scans were performed with Cirrus^{*} high-definition device using the standard Optic Disc Cube 200 × 200 protocol centered on the optic nerve head and the Macular Cube 200 × 200 centered on the fovea. For this study, only 13 parameters: 5 pRNFL thickness (average, superior, inferior, temporal, and nasal sectors) and 8 GCIPL thicknesses (average and minimum values, and the six wedge-shaped sectors: superior temporal, superior, superior nasal, inferior nasal, inferior and inferior temporal) were considered. Only images with good quality (signal strength > 6/10), no artifacts, and correct segmentation, as reviewed by the three glaucoma specialists, were included.

Statistical analysis

The normality—and homoscedasticity in normally distributed variables were checked. The sample was described using the mean and standard deviation (SD) for normally distributed quantitative variables, median (interquartile range) for non-normally distributed quantitative variables, and n (percentage) for categorical variables. A comparison of baseline characteristics between healthy and glaucomatous eyes was made using unpaired Student's t test, Mann–Whitney, or Fisher exact test, as appropriate.

The variables inspected were derived exclusively from the 13 SD-OCT parameters mentioned previously. To gain early insights into the ability of these parameters to classify eyes as either glaucomatous or healthy, the first two components derived from principal component analysis (PCA) were plotted (SM Fig. 1).

The main analyses involved two phases. In the first phase, three clustering methods were used separately to classify each eye in the training sample as healthy or glaucomatous using solely those 13 SD-OCT variables:

- In "k-means" the number of groups is pre-specified by the researcher and the corresponding centroids or points in space representing the reference for each cluster are randomly chosen. Each eye is assigned to the closer centroid based on its similarities to the 13 variables. The position of each centroid is updated based on the mean values for each variable of eyes belonging to it, and each eye is then reallocated to its nearest new centroid. This process is repeated until no eye changes its centroid. The whole process was repeated after n = 100 random starts.
- In agglomerative "HCA" each eye begins as its own cluster and they are merged one at a time with the closest observation, based on the Euclidean distance until all eyes form a single cluster. When ≥2 eyes form a cluster, the measure of similarity (linkage) was taken as the average distance between each observation in one cluster and each observation in other clusters.
- "MBC" models the probability of each eye belonging to each group. This allows the clusters to have different sizes, shapes, and orientations; the combination of these three characteristics generates 14 models. The optimal model was that with the best Bayesian Information Criterion. The expectation-maximization algorithm was used for maximum-likelihood estimation and a hierarchical clustering algorithm was used to define its optimal starting value. For a more detailed explanation of these methods, see Giordani et al. [10].

In the second phase, the best-performing algorithm in the training sample was applied to an independent (test) sample of healthy and glaucomatous eyes from hospitals in different geographical locations. This was used for external validation of results.

The primary outcome was the accuracy of each clustering algorithm, defined as the number of correctly classified eyes divided by the total number of eyes. Secondary outcomes included the diagnostic parameters sensitivity, specificity, positive/negative predictive values, false positives (FP)/negatives (FN), and the adjusted Rand index (ARI), a measure of agreement between two classifications that ranges from 0 (random partition) to 1 (perfect agreement) [13]. These parameters were also compared between the best-performing algorithm and the color-coded analysis of the pRNFL and GCIPL sector parameters on SD-OCT (classified as glaucoma when one or more sectors were yellow [borderline] or red [abnormal], and as healthy otherwise). Venn diagrams were plotted to determine concordant/discordant cases of each approach with the ground truth.

The analyses were conducted using software R (version 4.0.1, available on www.r-project.org, and its library mclust, version 5.4.9) and Stata IC (version 15; StataCorp, College Station, Texas, US). A p value < 0.05 was considered statistically significant.

RESULTS

We consecutively reviewed the medical records of 10,457 (training sample) and 3823 (test sample) subjects from the Glaucoma and Ophthalmology Primary Care Departments. From each set, 160 and 135 patients were eligible and further evaluated; after a second review, 109 and 102 patients meeting all the inclusion/exclusion criteria were included in the study; the reasons for exclusion are detailed in SM Fig. 2. The distribution of

the visual field mean deviation values for all 80 healthy and 131 glaucoma eyes is represented in SM Fig. 3.

Training sample

This sample included 109 eyes of 109 patients. There were 62 females (56.9%), the mean age was 68.1 years (SD 12.0), and there were 48 healthy and 61 glaucomatous eyes (Table 1). Of those with glaucoma, 43 and 18 had early and moderate disease, respectively (SM Fig. 3).

In the scatterplot matrix of the SD-OCT parameters, there was a fair to good correlation among all variables with a trend for the glaucomatous eyes to occupy the lower left location of the graph (thinner thicknesses) (SM Fig. 4). When representing the PCA of these 13 SD-OCT parameters, we found that the first two components gathered 82.0% of all the variance and that despite a considerable overlap of some healthy and glaucomatous eyes, there was a substantial clustering of eyes by diagnosis (SM Fig. 5).

Algorithm accuracy (Table 2 and SM Table 1) was highest for MBC (91.7%), followed by k-means (81.7%) and HCA (78.9%; *p* value = 0.02). The best MBC model was that in which subgroups were allowed to have variable volume and shape and equal orientation ("VVE"). K-means and HCA showed a similar and reasonable ability to separate observations into the healthy and glaucoma subgroups, but the separation was sharper for k-means than for HCA, explaining its slightly better accuracy results (SM Fig. 6). The classical graph for HCA, the dendrogram or tree plot, is shown in SM Fig. 7.

Regarding the classification metrics for each algorithm (SM Table 2), MBC obtained the best outcomes in every parameter measured, while k-means and HCA showed similar results between them, with modest sensitivities (~65%) and a rate of false negatives around 30%. The ARI provided a clear separation of values between algorithms, with the best results for MBC (0.69), followed by k-means (0.40) and HCA (0.33).

Test sample

Given the previous results, MBC was chosen as the best algorithm and then applied to the test sample for external validation. This consisted of 102 eyes of 102 patients from Hospital Sagrat Cor (n = 43), Hospital Clínic (n = 34), and Instituto Oftalmológico Fernández-Vega (n = 25). There were 32 healthy and 70 glaucomatous eyes (Table 1). Of those with glaucoma, 37 had early, 25 had moderate, and 8 advanced disease (SM Fig. 8). A description by site is provided in SM Table 3. Again, there was a good correlation between all SD-OCT parameters with glaucomatous eyes showing thinner layer thicknesses, as expected (SM Fig. 9).

The best model in the test set using MBC was again the VVE, which correctly classified 98 out of 102 cases for an overall accuracy of 96.1% (95% Cl, 92.3-99.8%). There were 4 FNs (glaucomatous eyes classified as healthy by the MBC algorithm): three of them were early stage, and in one the damage was earlymoderate (MD -6.12 dB). If we exclude the 8 advanced glaucoma cases from the test sample, the accuracy would be 95.7% (95% Cl, 89.5-98.8%). If we now further restrict the analysis to a comparison of healthy and early glaucoma cases, the accuracy of the MBC model would then be 66/69 or 95.7% (95% Cl, 87.8-99.1%). The diagnostic accuracy of OCT conventional parameters was as follows: 92.2% (95% Cl, 86.9-97.4%) for sector pRNFL and 98.0% (95% CI, 95.3–100%) for GC-IPL. There were no statistically significant differences between the accuracy of MBC and GCIPL (p value = 0.41), MBC and pRNFL (p value = 0.23), and they were borderline for the GCIPL-pRNFL comparison (p = 0.052). The corresponding 2×2 tables are shown in Table 3, additional metrics in Table 4, and Venn diagrams in Fig. 1.

DISCUSSION

Clustering algorithms are widely used data-analyses techniques. They provide an unbiased way to classify observations, although

Eye

Table 1. Baseline features of study participants as classified in twogroups, normal and glaucoma in the training and test samples.

Training sample (n = 109)

	Healthy, n = 48	Glaucoma, <i>n</i> = 61	p value ^a	
Age, years	61.1 (11.5)	73.5 (9.4)	<0.0001 ^a	
Sex, females	33 (68.8)	29 (47.5)	0.03 ^a	
IOP, mmHg*	16 (4)	24 (3)	<0.0001 ^b	
Mean deviation, dB	—0.30 (1.05)	-5.14 (2.44)	<0.0001 ^a	
Pachymetry, μm	548.9 (40.3)	538.3 (37.0)	0.21ª	
Quality, papilar	7.8 (1.1)	7.4 (0.9)	0.04 ^a	
Quality, macular	8.6 (0.9)	7.9 (1.2)	0.0006 ^a	
Test sample ($n = 102$)				

Healthy, Glaucoma, n = 70 p value n = 32Age, years 57.7 (10.8) 71.4 (12.5) < 0.0001^a 0.52^a Sex, females 20 (62.5) 38 (54.3) IOP, mmHg < 0.0001^a 16.6 (3.0) 24.2 (3.0) -0.03(1.29)< 0.0001^b Mean deviation, -5.90(4.29)dB* Pachymetry, µm 0.55^a 538.3 (43.0) 533.3 (34.4) Quality, papilar 8.1 (1.3) 7.8 (1.1) 0.26^a Quality, macular 8.9 (1.2) 8.6 (1.2) 0.35^a

Values represent mean (standard deviation) or *median (interquartile range) for quantitative and *n* (%) for categorical variables. *dB* decibels, *IOP* intraocular pressure.

ab decidels, *IOP* int *at*-test.

^bMann–Whitney U test.

[ahle	2	Comparison	of accuracies	of the	different	algorithms
iavie	2.	Companson		or the	unerent	algorithins.

Algorithm	Accuracy (95% CI), %	p value
K-means	81.7 (73.1-88.4)	0.02
Hierarchical clustering analysis	78.9 (70.0–86.1)	
Model-based clustering	91.7 (84.9–96.2)	

Model-based clustering showed the highest accuracy.

input from the researcher is required for variable selection, distance, and initialization method, among others [14]. Surprisingly, these algorithms have hardly been employed in glaucoma, and the very few works available, have applied these techniques to find subgroups in normal-tension [15] and juvenile-onset glaucoma [16], molecular [17] associations with the disease, and rates of progression depending on patterns observed in optic nerve imaging [18], but not in classification. Most of these studies have used HCA and none of them has applied more recent techniques based on MBC.

The purpose of our study was to evaluate three clustering algorithms derived exclusively from data provided by pRNFL and GCIPL parameters from the SD-OCT to obtain an unsupervised glaucoma classification model. We found that MBC was superior to k-means and HCA, with accuracies of 91.7% for MBC, 81.7% for k-means, and 78.9% for HCA (p value = 0.02; Table 2). Likewise, other measures of discrimination (SM Table 2) and the ARI were

also higher for MBC (0.69 vs. 0.40 for k-means and 0.33 for HCA). The MBC algorithm was then externally validated in a different sample (test sample) achieving even better accuracy (96.1%) with a sensitivity of 94.3 and 100% specificity. When compared to SD-OCT standard color-coded classification, MBC accuracy was not inferior to the best-performing SD-OCT parameter, GCIPL (96.1% vs. 98.0%, *p* value = 0.41). Classification based on pRNFL numerically showed the lowest accuracy (92.2%) although, again, differences were not statistically significant to either MBC or GCIPL (*p* values > 0.05).

There are several possible explanations for these results. First, even though pRNFL thickness usually outperforms the rest of individual SD-OCT parameters [19, 20] there are previous publications in which GCIPL diagnostic performance has been shown to be comparable to that of pRNFL [20] or even higher [21], although these differences were found to be not clinically relevant. Second, the color-coded classification values of pRNFL and GCIPL are obtained from a comparison with a built-in age-adjusted normative database [22], whereas the MBC algorithm is based only on SD-OCT parameters without taking age or any other input into account. Third, our definition used to classify eyes as glaucomatous on pRNFL and GCIPL was biased toward increased sensitivity (a single borderline or yellow sector was considered as indicative of the disease), which may explain the slightly improved sensitivity of GCIPL relative to MBC, at the expense of some loss on specificity. Overall, the results with the MBC algorithm are promising, especially considering that the sample included predominantly cases with early disease and that the parameters were only derived from SD-OCT without using any other clinical or demographical variable, which allows room for improvement.

Why does MBC achieve better outcomes than the other models? A potential explanation is that in classical heuristic methods, clusters are required to be spherical; that is, the distances between the center of the cluster and the more peripheral individuals are forced to be the same. In MBC they

are allowed to be elliptical, with different shapes and orientations in the multivariable space formed by all pRNFL and GCIPL measurements [23]. In fact, the optimal MBC model was the VVE, which allows for variable size, variable shape, and equal orientation of the major axes of the ellipse in each subgroup. This suggests that the increased flexibility of this model fitted better the data, allowing improvements in the classification.

What are the potential applications of these algorithms? SD-OCT can detect, quantify, and monitor structural glaucomatous damage objectively. However, SD-OCT testing is not yet included in the definition of diagnosis or progression of the disease [24]. Furthermore, SD-OCT changes typically precede functional abnormalities in early glaucoma in which visual defects are often very scarce or even absent [20, 25]. Considering our results, we believe that SD-OCT-based algorithms can be useful in a screening context or when a glaucoma diagnosis cannot be made with certainty. In this scenario, highly specific tests like the one designed here could be very useful, since only patients with the disease will screen positive, optimizing resources for further necessary diagnostic procedures. Despite the very good sensitivity values (>90%), in an irreversible and potentially treatable disease like glaucoma, FNs are also concerning. We reviewed the four FNs of the testing sample, finding that all of them had very focal damage that could have been missed in the SD-OCT regional analyses. Since these clustering modeling techniques are objective and unsupervised, they may also provide new insights into the nature of glaucomatous neuropathy. Some studies have found ganglion cell layer thinning at early stages of the disease [26], but to date, pRNFL thickness measurements have not been overcome by other SD-OCT parameters [20]. Although it is possible that axonal thinning occurs before soma degeneration [27, 28], it may also be that current technologies are not able to completely identify prompt damage occurring in different areas at the same time. Moreover, the location of the glaucomatous structural loss can be very variable among patients, and even

Table 3. Contingency (2×2) table of the classification into healthy or glaucoma cases in the test set using model-based clustering, sector pRNFL and sector GCIPL.

	Clinical exam (ground truth)			
	Glaucoma	Healthy		
MBC				
Diseased	66	0		
Non-diseased	4	32		
pRNFL				
Diseased	63	1		
Non-diseased	7	31		
GCIPL				
Diseased	69	1		
Non-diseased	1	31		

GCIPL ganglion cell-inner plexiform layer, *MBC* model-based clustering, *pRNFL* peripapilar retinal nerve fiber layer.





Table 4. Classification parameters for the model-based clustering with the VVE covariance matrix, sector pRNFL and sector GCIPL in the test sample.

	-		•				•
Method Ser	nsitivity	Specificity	PPV	NPV	FP	FN	ARI
MBC 94.	.3 (86.0–98.4)	100 (89.1–100)	100 (89.1–100)	88.9 (75.5–95.4)	0 (0.0–10.9)	5.7 (1.6–14.0)	0.85
pRNFL 90.	.0 (80.5–95.9)	96.9 (83.8–99.9)	98.4 (91.6–100.0)	81.6 (65.7–92.3)	3.1 (0.0–16.2)	10 (4.1–19.5)	0.71
GCIPL 98.	.6 (92.3–100)	96.9 (83.8–99.9)	98.6 (92.3–100)	96.9 (83.8–99.9)	3.1 (0.0–16.2)	1.4 (0.0–7.7)	0.92

ARI adjusted Rand index, FN false negatives, FP false positives, GCIPL ganglion cell-inner plexiform layer, HCA hierarchical cluster analysis, MBC model-based clustering, NPV negative predictive value, PPV positive predictive value, pRNFL peripapilar retinal nerve fiber layer.

though the superior and inferior disc thinning ("vulnerability zones") are the most common, widespread damage has also been identified [29]. Actually, in our recent work, a diffuse representation of both macular and pRNFL parameters was found to have the most discriminative ability between healthy and glaucomatous subjects [30]. Currently, different combinations of peripapillary and macular SD-OCT parameters have been able to achieve better diagnostic performances than using isolated SD-OCT parameters. In this regard, Hood and colleagues have extensively analyzed the macular and peripapillary areas, addressing the fact that the RNFL -axons- (macula and optic nerve) and ganglion cell layer -cell bodies- (macula) damage are not separate [29] and that the use of probability maps including both areas probably improves structure-function relationship [31]. Although their research is based on probability maps and ours is derived from thickness measurements, using information coming from both areas may help the clinician to detect glaucoma cases that, relying only on pRNFL thickness, would have been either missed [29], or erroneously classified [32], as observed in the four FNs missed by our algorithm. However, further research is needed to elucidate if both methods are addressing the same clinical problem since combining both probability maps and thickness could potentially worsen the model specificity.

Another area of discussion is the optimization of the model variable selection. In this regard, some SD-OCT parameters may have little relevance for the detection of glaucoma and therefore may only induce noise in the model (i.e., nasal pRNFL); others may be redundant (does average pRNFL or GCIPL measurements increase accuracy if all sectors in these locations have already been considered?). The present research was performed with Cirrus OCT: parameters such as isolated macular retinal nerve fiber layer (Spectralis SD-OCT) [30], or the lack of agreement between the devices [33] may influence the outcomes using the clustering approach in different instruments. Although using the instrument probability color-coded values (green, yellow, red) is more similar to what we do in clinical practice (especially among comprehensive ophthalmologists), it may have a different accuracy behavior than the usually reported quantitative isolated SD-OCT best cut-off values, and that might have led to the previously mentioned increased sensitivity of GCIPL

In this study using only OCT data, despite not using any clinical variable for the model configuration like IOP, central corneal thickness or age—all very known risk factors for glaucoma—the MBC algorithm showed very good accuracy with little overfitting in the test sample. Nevertheless, since color-coded values obtained by the in-built comparative age-adjusted database were used as a classificatory comparator, the effect of age was somehow indirectly incorporated into the model which could also further explain its excellent results. Adding other relevant clinical characteristics (hemorrhages, IOP, MD...) may further improve the model outcomes.

Our study has some limitations. First, its exploratory nature. Second, the sample size: the relatively small sample size limits the range of presentations of both healthy and glaucoma cases. However, we believe these results are still applicable, considering that the characteristic glaucomatous focal and/or diffuse thinning of the pRNFL and GCIPL are aspects well captured by our model and the good results obtained in the validation in an independent sample. Third, the sampling method. In trials with less than 100 subjects per subgroup, block randomization is advisable to reduce the risk of imbalance [34, 35]; however, it is not expected to have a significant impact on the results because of the retrospective, non-interventional design of the study; and also because the randomization was done for sampling between the right or left eye from the same (either a healthy control or a glaucoma case) patient, and not between glaucoma and healthy patients. Fourth, our results may have been positively influenced by a similar age difference between healthy and glaucoma eyes in both the training and the test sets. However, despite not directly including age in the model, the MBC algorithm showed very good accuracy with little overfitting in the test sample. Evaluating model accuracy in samples with a different age distribution from that observed in our samples will shed light on this area. Additionally, we expect that adding other relevant characteristics (peripapillary hemorrhages, IOP, pachymetry, MD) may overcome any potential bias introduced by a shift in age and other non-detected imbalances between groups and improve the model outcomes. Fifth, the diagnostic accuracy overestimation in case-control studies. A perfect reference standard for glaucoma diagnosis is still lacking and therefore, glaucoma suspects are not commonly involved in diagnostic accuracy studies evaluating OCT for glaucoma detection, which often adopt a casecontrol design, like in our study. Such design includes healthy controls separately from glaucoma cases; thus, it can overestimate accuracy and reduce the applicability of the results to daily practice. Therefore, ocular hypertensives and eyes with pre-perimetric glaucoma were not considered and the present algorithm cannot be applied to those eyes. Long-term follow-up in uncertain cases would be useful to determine their conversion rate and the corresponding performance of the algorithm, but this is beyond the scope of the current study. Sixth, the performance of these methods specifically in early glaucoma, where the diagnosis is more challenging, should be evaluated in future studies to minimize spectrum bias; the current sample included almost 30% of cases with moderate disease, which are arguably easier to detect. Finally, the study design must minimize incorporation bias, a common issue in clinical glaucoma research. These potential improvements in classification accuracy deserve further consideration.

In summary, clustering algorithms in general (and MBC in particular) seem promising methods to help discriminate between healthy and glaucomatous eyes using exclusively SD-OCT-derived parameters. Understanding their relative merits may also provide insights into the nature of the disease.

SUMMARY

What was known before

- Adding optic nerve and macular optical coherence tomography (OCT) parameters to the visual field and optic nerve examinations for glaucoma diagnosis has increased the agreement among specialists.
- Combining OCT parameters could further improve diagnostic accuracy, but the best combination is yet to be determined.

What this study adds

- Unsupervised clustering algorithms (especially model-based clustering) using only OCT parameters can discriminate between healthy and glaucoma cases with very good accuracy, sensitivity, and specificity.
- A further comprehension of the merits of clustering models for classifying the cases could improve knowledge (structurewise) of the disease.
- These algorithms could be implemented for glaucoma screening.

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

REFERENCES

- Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121:2081–90.
- Greenfield DS, Weinreb RN. Role of optic nerve imaging in glaucoma clinical practice and clinical trials. Am J Ophthalmol. 2008;145:598–603.
- Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. Ophthalmology. 1992;99:215–21.
- Jampel HD, Friedman D, Quigley H, Vitale S, Miller R, Knezevich F, 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.
- Susanna R, De Moraes CG, Cioffi GA, Ritch R. Why do people (still) go blind from glaucoma? Transl Vis Sci Technol. 2015;4:1.
- Chen TC, Hoguet A, Junk AK, Nouri-Mahdavi K, Radhakrishnan S, Takusagawa HL, et al. Spectral-domain OCT: helping the clinician diagnose glaucoma. Ophthalmology. 2018;125:1817–27.
- Blumberg DM, De Moraes CG, Liebmann JM, Garg R, Chen C, Theventhiran A, et al. Technology and the glaucoma suspect. Invest Ophthalmol Vis Sci. 2016;57:OCT80–5.
- Ramachandran R, Joiner DB, Patel V, Popplewell D, Misra P, Kaplan CM, et al. Comparison between the recommendations of glaucoma specialists and OCT report specialists for further ophthalmic evaluation in a community-based screening study. Ophthalmol Glaucoma. 2022;5:602–13.
- Mwanza JC, Warren JL, Budenz DL. Utility of combining spectral domain optical coherence tomography structural parameters for the diagnosis of early Glaucoma: a mini-review. Eye Vis. 2018;5:9.
- Giordani P, Ferraro MB, Martella F. An introduction to clustering with R. Vol 1. 1st ed. Singapore: Springer; 2020. p. 20.
- Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc. 2002;97:611–31.
- Anderson DR, Patella VM. Automated static perimetry. 2nd ed. St. Louis: Mosby; 1999.
- 13. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193-218.
- Biarnés M, Ferraro LL, Garcia M, Delcourt C, Lengyel I, Monés J. Re: Keenan et al.: Cluster analysis and genotype-phenotype assessment of geographic atrophy in age-related macular degeneration: AREDS2 Report 25 (Ophthalmology. 2021;5(11):1061-1073). Ophthalmol Retin. 2022;6:333–4.
- Bae HW, Ji Y, Lee HS, Lee N, Hong S, Seong GJ, et al. A hierarchical cluster analysis of normal-tension glaucoma using spectral-domain optical coherence tomography parameters. J Glaucoma. 2015;24:328–33.
- Birla S, Gupta D, Somarajan BI, Gupta S, Chaurasia AK, Kishan A, et al. Classifying juvenile onset primary open angle glaucoma using cluster analysis. Br J Ophthalmol. 2020;104:827–35.
- Howell GR, Macalinao DG, Sousa GL, Walden M, Soto I, Kneeland SC, et al. Molecular clustering identifies complement and endothelin induction as early events in a mouse model of glaucoma. J Clin Invest. 2011;121:1429–44.
- Yan S, Abidi SSR, Artes PH. Analyzing sub-classifications of glaucoma via SOM based clustering of optic nerve images. Stud Health Technol Inf. 2005;116:483–8.
- Schulze A, Lamparter J, Pfeiffer N, Berisha F, Schmidtmann I, Hoffmann EM. Diagnostic ability of retinal ganglion cell complex, retinal nerve fiber layer, and optic nerve head measurements by Fourier-domain optical coherence tomography. Graefes Arch Clin Exp Ophthalmol. 2011;249:1039–45.
- Pazos M, Dyrda AA, Biarnés M, Gómez A, Martín C, Mora C, et al. Diagnostic accuracy of spectralis SD OCT automated macular layers segmentation to discriminate normal from early glaucomatous eyes. Ophthalmology. 2017;124:1218–28.
- Mwanza JC, Durbin MK, Budenz DL, Sayyad FE, Chang RT, Neelakantan A, et al. Glaucoma diagnostic accuracy of ganglion cell-inner plexiform layer thickness: comparison with nerve fiber layer and optic nerve head. Ophthalmology. 2012;119:1151–8.
- Mwanza JC, Durbin MK, Budenz DL. Interocular symmetry in peripapillary retinal nerve fiber layer thickness measured with the cirrus HD-OCT in healthy eyes. Am J Ophthalmol. 2011;151:514–21.e1.
- 23. Scrucca L, Fop M, Murphy BT, Raftery AE. Mclust. R J. 2016;8:289-317.
- 24. Medeiros FA, Tatham AJ. Structure versus function in glaucoma: the debate that doesn't need to be. Ophthalmology. 2016;123:1170–2.

- Kuang TM, Zhang C, Zangwill LM, Weinreb RN, Medeiros FA. Estimating lead time gained by optical coherence tomography in detecting glaucoma before development of visual field defects. Ophthalmology. 2015;122:2002–9.
- Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. Invest Opthalmol Vis Sci. 2012;53:6939–46.
- Quigley HA, Addicks EM. Chronic experimental glaucoma in primates. II. Effect of extended intraocular pressure elevation on optic nerve head and axonal transport. Invest Ophthalmol Vis Sci. 1980;19:137–52.
- Martin KRG, Quigley HA, Valenta D, Kielczewski J, Pease ME. Optic nerve dynein motor protein distribution changes with intraocular pressure elevation in a rat model of glaucoma. Exp Eye Res. 2006;83:255–62.
- Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). Prog Retin Eye Res. 2017;57:46–75.
- Pazos M, Biarnés M, Blasco-Alberto A, Dyrda A, Luque-Fernández MÁ, Gómez A, et al. SD-OCT peripapillary nerve fibre layer and ganglion cell complex parameters in glaucoma: principal component analysis. Br J Ophthalmol. 2021;105:496–501.
- Hood DC, Raza AS, de Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. Prog Retin Eye Res. 2013;32:1–21.
- Hood DC, la Bruna S, Tsamis E, Thakoor KA, Rai A, Leshno A, et al. Detecting glaucoma with only OCT: implications for the clinic, research, screening, and Al development. Prog Retin Eye Res. 2022;90:101052.
- Patel NB, Wheat JL, Rodriguez A, Tran V, Harwerth RS. Agreement between retinal nerve fiber layer measures from spectralis and cirrus spectral domain OCT. Optom Vis Sci. 2012;89:E652–66.
- Lim CY, In J. Randomization in clinical studies. Korean J Anesthesiol. 2019;72:221–32.
- Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. Lancet. 2002;359:966–70.

AUTHOR CONTRIBUTIONS

MB contributed to design of the study, data analysis, interpretation, and writing. NVA was responsible for data acquisition and collection, data analysis, interpretation, and writing. RCD, FF, SBF, MTCD, and IRU collected data. MJM and EM data interpretation, review. MP contributed to conception and design, data acquisition, interpretation, writing and preparation of the manuscript.

COMPETING INTERESTS

MB, FF, SBF, MTCD, RCD, EM, MJM: none. NVA and IRU: Lecturer for Carl Zeiss Meditec. MP: Lecturer and Consultant Carl Zeiss Meditec, Editorial board member for *Eye*.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41433-023-02785-5.

Correspondence and requests for materials should be addressed to Marta Pazos.

Reprints and permission information is available at http://www.nature.com/ reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.