

 Open access • Posted Content • DOI:10.1101/2021.02.07.21251275

Circular functional analysis of OCT data for precise identification of structural phenotypes in the eye. — Source link

Hasnat Ali, Brian Wainwright, Alexander M. Petersen, Ganesh B. Jonnadula ...+6 more authors

Institutions: L V Prasad Eye Institute, University of California, Santa Barbara, Birla Institute of Technology and Science

Published on: 09 Feb 2021 - Scientific Reports (Cold Spring Harbor Laboratory Press)

Topics: Population

Related papers:

- [Unifying structure analysis and surrogate-driven function regression for glaucoma OCT image screening](#)
- [Finding New Diagnostic Information for Detecting Glaucoma using Neural Networks](#)
- [Did the OCT Show Progression Since the Last Visit](#)
- [Glaucoma progression detection using nonlocal Markov random field prior.](#)
- [Structural classification of glaucomatous optic neuropathy](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/circular-functional-analysis-of-oct-data-for-precise-28qmnmsi99>

Title: Circular functional analysis of OCT data for precise identification of structural phenotypes in the eye

Authors:

Md. Hasnat Ali^{1,2}, Brian Wainwright³, Alexander Petersen^{3,4}, Ganesh B. Jonnadula¹, Meghana Aruru⁵, Harsha L. Rao^{6,7}, M. B. Srinivas², S. Rao Jammalamadaka³, Sirisha Senthil¹, and Saumyadipta Pyne^{5,8,*}

Author Affiliations:

¹ L. V. Prasad Eye Institute, Hyderabad, Telangana, India.

² BITS Pilani, Hyderabad Campus, Hyderabad, Telangana, India.

³ Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, USA.

⁴ Department of Statistics, Brigham Young University, Provo, UT, USA.

⁵ Health Analytics Network, Pittsburgh, PA, USA.

⁶ Narayana Nethralaya Eye Hospital, Bengaluru, Karnataka, India.

⁷ University Eye Clinic Maastricht, Maastricht University Medical Center, Maastricht, the Netherlands.

⁸ Public Health Dynamics Laboratory, and Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA.

***Corresponding Author:**

Saumyadipta Pyne, email: spyne@healthanalytics.net

Abstract:

Progressive optic neuropathies such as glaucoma are major causes of blindness globally. Multiple sources of subjectivity and analytical challenges are often encountered by the clinicians in the process of early diagnosis and clinical management of these diseases. In glaucoma, the structural damage is often characterized by neuroretinal rim (NRR) thinning of the optic nerve head, and other clinical parameters. Optical coherence tomography (OCT) is a popular and quantitative eye imaging platform for precise and reproducible measurement of such parameters in the clinic.

Baseline structural heterogeneity in the eyes can play a key role in the progression of optic neuropathies, and thus present challenges to clinical decision-making. To address this, large and diverse normative OCT databases with mathematically precise description of phenotypes can help with early detection and characterization of the different phenotypes that are encountered in the clinic. In this study, we generated a new large dataset of OCT generated high-resolution circular data on NRR phenotypes, along with other clinical covariates, of nearly 4,000 healthy eyes as part of a well-established clinical cohort (LVPEI-GLEAMS) of Asian Indian participants.

In this study, we (1) generated high-resolution circular OCT measurements of NRR thickness in a given eye, (2) introduced CIFU, a new computational pipeline for Circular Functional data modeling and analysis that is demonstrated using the OCT dataset, and (3) addressed the disparity of representation of the Asian Indian population in normative OCT databases. We demonstrated CIFU by unsupervised circular functional clustering of the OCT NRR data, meta-clustering to characterize the clustering output using clinical covariates, and presenting a circular visualization of the results. Upon stratification by age, we identified a healthy NRR phenotype cluster in the age group 40-49 years with predictive potential for glaucoma.

1. Introduction

Progressive optic neuropathies such as glaucoma can cause irreversible blindness, especially when left untreated or diagnosed late. Indeed, early detection and management hold the key to slowing the progressive loss of vision and preventing blindness due to many chronic and age-related degenerative eye diseases. Glaucoma, for instance, is the second-leading cause of blindness worldwide¹. In 2020, an estimated 80 million individuals worldwide had glaucoma and this number is expected to increase to over 111 million by 2040².

There are multiple sources of subjectivity and analytical challenges that are often encountered by the clinicians in the process of early diagnosis and clinical management of these diseases. In glaucoma, the functional damage is established most commonly by the occurrence of visual field (VF) loss whereas the structural damage is often characterized by neuroretinal rim (NRR) thinning of the optic nerve head (ONH), and loss of retinal nerve fibers, which are the axons of retinal ganglion cells (RGC). On the functional side, while standard automated perimetry (SAP) has been the gold standard for detection of VF loss, often 30% of RGC loss may have already occurred before VF defects could be detected by SAP³.

On the structural side, biological heterogeneity of ONH phenotypes, with or without any neuropathy, can present challenges to clinical decision-making. For instance, NRR area has been found to normally decline at the rate of 0.28%-0.39% per year⁴. There is no single, specific management guidance for patients with diverse morphology of ONH⁵. For instance, to assess the progression of glaucoma, one of the parameters assessed is the optic cup-to-disc ratio (CDR) which is calculated by comparing the diameter of the “cup” portion of the optic disc with the total diameter of the latter. Yet, while a large CDR may indicate glaucoma or other pathology, deep yet stable (over age) cupping, i.e., a normal physiologically large optic disc cup, can occur due to genetic factors in the absence of any disease or associated clinical covariates (e.g., high intraocular pressure)⁶. It is of great importance that sources of natural variation are rigorously understood thereby controlling for subjectivity in diagnosis.

In the clinic, non-invasive, high-resolution eye imaging platforms such as spectral-domain optical coherence tomography (OCT) provide excellent glaucoma diagnostic performance, especially in early stages of disease^{7,8}. The quantitative and reproducible OCT data provide objective measurement of ONH parameters, NRR area, retinal nerve fiber layer (RNFL), macular thickness, etc., which are used by clinicians to identify structural damage. For instance, Zeiss Cirrus HD-OCT platform uses the clinically invisible but OCT detectable Bruch's membrane opening (BMO) as the landmark to measure the amount of NRR tissue in the optic nerve. It has been reported that NRR thickness calculation by Cirrus HD-OCT has high reproducibility and glaucoma diagnostic ability, and a low rate of incorrect optic disc margin detection⁹⁻¹¹. The platform generates a comparative report of a patient's data based on its normative database.

The OCT platform performs circular scans of the eye which, as in many biomedical technologies, are examples of measurements that are recorded or indexed at different directions, say, at given angular positions around a central point. Unlike the analyses of “linear” data points that reside on the real line or Euclidean spaces, directional data requires special and altogether different treatment. For instance, a direction in two-dimensional plane can be represented as a point on the circumference of a unit circle, or simply as an angle, but neither representation is unique, as both depend on the selection of some appropriate “zero-direction” from which to start measuring, as well as the sense of rotation, viz., clockwise or anti-clockwise. The unique properties of circular data – for instance, if one wishes to compare two such scans with a distance measure – are appropriately addressed by the field of circular statistics¹².

Traditional OCT analysis may involve division of the circle around ONH into 4 fixed quadrants, or 12 clock-hours, to record measurements at these sectors. In this study, we extended such data collection to divide the same circle (of total 360 degrees) into much finer segments of 2 degrees each, and thus, generate 180 circular data points for each clinical sample (human eye). These rich and evenly spaced high-resolution circular (HRC) data allows for natural application of functional data analysis (FDA) where the data are not viewed as points but as curves or mathematical functions. Not to be confused with alternate usage of the term “function” (such as in vision), FDA is increasingly popular in biomedical informatics due to the emergence of new monitoring technologies that can record data as curves¹³⁻¹⁵. The HRC OCT data for each sample can be

modeled as a circular function or curve, with angle and magnitude being the independent and dependent variables, respectively. In this study, we describe methods for novel circular functional analysis of OCT data, and apply it for unsupervised identification of NRR phenotypes in healthy eyes.

Progressive and degenerative eye diseases benefit from pre-existing knowledge and cumulative collection and description of normal phenotypes as these may help to identify early and characterize precisely the new phenotypes that emerge over time. While some normative OCT databases do exist, they are generally limited in their size and diversity. Thus, the breakup of their ethnic representation may not reflect the actual epidemiologic distribution of the disease. For instance, only 1% of the popular Cirrus HD-OCT platform's normative database is of Asian Indian origin¹⁶, although India contributes to more than 12% of the global cases of both primary open-angle and primary angle-closure glaucomas¹⁷. Towards this, we generated a new large HRC OCT dataset on NRR phenotypes, along with other clinical covariates, of 3973 healthy eyes as part of a well-established clinical cohort (LVPEI-GLEAMS) at the L.V. Prasad Eye Institute, Hyderabad, India.

The main objectives of the present study are to: (1) generate HRC OCT data in the form of 180 circular measurements of NRR thickness in a given eye, (2) introduce CIFU, a computational pipeline for Circular Functional data modeling and analysis that is demonstrated using the OCT dataset, and (3) address the disparity of representation of the Asian Indian population in normative OCT databases. In the next section, we describe the clinical cohort and the protocol for data generation as well as the algorithm of CIFU for unsupervised circular functional clustering of the OCT NRR data, followed by meta-clustering to characterize the clustering output using clinical covariates of glaucoma. In the following section, the results of CIFU analysis are described with help of circular visualization. In particular, upon stratification of the samples by age, we identified a healthy NRR phenotype cluster in the age group 40-49 years, and having the highest mean values of cup volume and average CDR among all clusters, with predictive potential for glaucoma. We end with discussion of the CIFU approach and its potential applications to future work.

2. Data and Methods

2.1 Data

All participants were selected from a population-based study LVPEI-GLEAMS (LVPEI Glaucoma Epidemiology And Molecular Genetic Study¹⁷) conducted by the L.V. Prasad Eye Institute (LVPEI), Hyderabad, India. Data were collected on a total of 3973 (OD:1981; OS:1992) healthy eyes of 2222 participants from the southern Indian state of Andhra Pradesh, India. Written informed consent was obtained from all participants to participate in the study, and the ethics and review committee of the LVPEI reviewed and approved the methodology and was conducted in strict adherence to the tenets of the Declaration of Helsinki. The inclusion criteria used were age ≥ 40 years, best-corrected visual acuity of 20/40 or better, spherical equivalent of ± 6 Diopters, good quality stereo optic disc photographs, and no media opacities (signal strength ≥ 6). The exclusion criteria used were intraocular surgery within the previous 6 months, and any retinal (including macular) or neurologic diseases other that could confound the structural measurements with SD-OCT.

Healthy eyes were defined by the absence of anterior and posterior pathology. Each digital optic disc photograph was evaluated by three glaucoma specialists independently. The specialists were masked to the other clinical findings and the other imaging outcomes of the subjects. Eyes were excluded from the study in case of any disagreements among the specialists. All participants underwent a comprehensive ophthalmic examination which included detailed medical and systemic history. The means of clinical determination included best-corrected visual acuity measurement, slit-lamp photographs (Topcon, Bauer Drive, Oakland, NJ), Goldmann applanation tonometry (Hagg-Streit AT 900, Hagg-Streit AG, Switzerland), gonioscopy with a Sussman four mirror gonioscope (Volk Optical Inc, Mentor, Ohio, USA), dilated fundus examination, central corneal thickness (CCT) assessment, Humphrey visual fields (HVF) with 24-2 Swedish Interactive threshold algorithm (SITA; Carl Zeiss Meditec Inc. Dublin, CA). Visual fields (VF) were considered if false positive, false negative, and fixation losses were less than 20%, and all the stereophotographs of the optic disc had good quality.

In addition, digital optic disc photography and spectral-domain optical coherence tomography imaging with Cirrus HD-OCT (software version 9.0.0.281; Carl Zeiss Meditec, Dublin, CA, USA) were used. This is a computerized instrument that acquires and analyzes cross-sectional and three-dimensional tomograms of the eye using spectral-domain optical coherence tomography (SD-OCT) technology. The algorithm of the instrument automatically identifies the optic disc margin as the termination of Bruch's membrane (BM). BM opening (BMO) is used as the landmark to measure the amount of neuro-retinal rim tissue in the optic nerve. Optic Disc Cube 200x200 protocol was used to scan the ONH and peripapillary area through a 6 mm square grid, which consists of 200 horizontal linear B-scans and each composed of 200 A-scans. First, the Cirrus HD-OCT algorithm identifies the center of ONH and then automatically places a calculation circle of 3.46 mm diameter evenly around it. The circular scan starts at an extreme temporal point and moves around the ring in the superior direction, then nasal, then inferior, then back to temporal (TSNIT). The circular measurements are made clockwise for the right eye and counter-clockwise for the left eye. NRR thickness is measured by the amount of neuro-retinal tissue in the optic nerve around the entire edge of the optic disc. Zeiss Cirrus HD-OCT used Bruch's membrane opening–minimum rim width (BMO-MRW) to measure the rim area. The BMO-MRW is the shortest distance from BMO to the retinal internal limiting membrane. The advanced export functionality was used to record the NRR thickness values at 180 points in TSNIT order spaced evenly by 2 degrees (from 2°–360°) around the circle. We refer to this as our NRR OCT high-resolution circular (HRC) data. The data were stratified into 3 age groups: (1) 40-49 years, (2) 50-59 years, and (3) 60 years and older.

2.2 Methods

We describe CIFU pipeline for circular functional modeling and clustering of HRC OCT data, followed by metaclustering based clinical characterization of the clusters identified by CIFU.

Circular Functional Modeling and Clustering

First, we introduce a method for clustering OCT HRC data into K homogeneous groups of samples (i.e., eyes). As an unsupervised approach, the method is based only on HRC data as input, and not

any clinical variables of the samples. While the actual OCT measurements are taken on a discrete grid of angles around the center of ONH, in principle, these finely indexed measurements are assumed to vary continuously around a circular scale ranging from 0 to 360 degrees, and wrapped around. Thus, for statistical modeling, we will refer to the OCT data as a collection of n circular curves $X_1(t), X_2(t) \dots, X_n(t)$, where $X_i(t)$ represents the OCT value (of NRR thickness) in the i^{th} sample ($i = 1, 2, \dots, n$) measured radially at each point t (here, $t = 1, 2, \dots, 180$) and around a common center. The angular indices are aligned for all samples and spaced 2 degrees apart starting from a common 0 degrees.

To allow for comparison of the curves based on their shapes rather than the magnitude, we normalize each curve $X_i(t)$ by dividing it by $\int_0^{2\pi} X_i(t) dt$.

Our modeling begins with the use of p basis functions for capturing the functional nature of data. If $\psi_1, \psi_2, \dots, \psi_p$ are the basis functions with the associated basis expansion coefficients γ_{ij} , where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p$, then the functional approximation for the i^{th} curve at point t , $X_i(t)$, is given by

$$X_i(t) \approx \sum_{j=1}^p \gamma_{ij} \psi_j(t) = X_i^p(t). \quad (1)$$

Our objective is to cluster each of the circular curves described as above into a pre-specified number (K) of clusters. Towards this, we used a discriminative functional mixture (DFM) model given by Bouveyron et al. (2015)¹⁸ in which $\gamma_i = (\gamma_{i1}, \dots, \gamma_{ip})^t$ of curve X_i follows a finite mixture model of K Gaussian components with the density function

$$f(\gamma) = \sum_{k=1}^K \pi_k \phi(\gamma; U\mu_k, U^t \Sigma_k U + \Xi), \quad (2)$$

where $\pi_k \geq 0$ is the mixing proportion of the k^{th} component (i.e., the cluster k) such that $\sum_{k=1}^K \pi_k = 1$, and ϕ is the standard Gaussian density function. Here, U is a $p \times d$ orthogonal matrix mapping the basis coefficients γ into the discriminative subspace (of dimension $d < p$) through a linear transformation. Similarly, μ_k and Σ_k are the mean vector and covariance matrix (for cluster k) of γ mapped into the discriminative subspace, and the noise of the above transformation is normally distributed with mean zero and covariance Ξ .

The DFM model was fit with an EM algorithm implemented in the R package funFEM¹⁹. Following the idea of constraining variance parameters in Fraley and Raftery (1999)²⁰, the funFEM package allows 12 different choices of DFM models. As an initial step, using different restrictions on the noise covariance matrix \mathcal{E} , a preliminary model search was run with NRR data, with outlier curves included, over all 12 models and a flexible set of (61) basis functions. Based on the results of the initial run, and after removing the outlier curves by equation (3), the clustering algorithm was run using a smaller selective set of (21) basis functions.

A given curve $X_i(t)$ is considered an outlier if it exceeds the following threshold

$$X_{out} = \max_t \bar{X}_i(t) + 3.5 \times \max_t sd(X_i(t)), \quad t = 1, \dots, 180; i = 1, \dots, n \quad (3)$$

To avoid model overfitting, we determined the smallest number of basis functions (p) that recover the input curves sufficiently well, as determined by the fraction of variation explained (FVE) as described below. Let the sample mean of n given curves be

$$\bar{X}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t), \quad t = 1, 2, \dots, 180. \quad (4)$$

Then the total variation (TV) is given by

$$TV = \frac{1}{n-1} \sum_{i=1}^n \int (X_i(t) - \bar{X}(t))^2 dt \quad (5)$$

and the fraction of variation explained (FVE) by

$$FVE = \frac{TV - \frac{1}{n-1} \sum_{i=1}^n \int (X_i^p(t) - X_i(t))^2 dt}{TV}. \quad (6)$$

We used FVE as the criterion for selecting an optimal number of basis functions (p).

Finally, the number of clusters identified by the DFM models for each age group was determined by 3 known model selection criteria AIC, BIC and ICL.

For intuitive visualization of the clustering results, we plotted the curves of every cluster in a distinct color using a circular scale. The mean curve of each cluster, as computed by (4), is included as a bold black curve, which serves as a cluster-specific template.

For comparison of our circular functional clustering method with a popular non-circular data clustering approach, we used the Partitioning Around Medoids (PAM) as implemented in R. The optimal number of clusters identified by PAM was selected using the Average Silhouette Width (ASW), while Dunn Index was calculated as a measure of inter-cluster variation. We used the R packages ‘factoextra’ and ‘fpc’ for PAM clustering, validation and visualization.

Metaclustering and Clinical Characterization of Clusters

In the metaclustering step, the clusters identified by circular functional data were grouped based on their samples’ similarity in terms of a selected set of clinical variables that are known covariates of glaucoma. A feature selection step was performed simultaneously to detect the covariates that were the most distinctive across the metaclusters. The metaclustering workflow consists of the following steps:

1. In each age group, each circular functional cluster C was represented by the mean values of a set of clinical covariates of the samples in C .
2. We performed agglomerative hierarchical clustering of the clusters given by their mean covariate data with complete linkage, and a tuning parameter ($wbound$) to select the covariates that are the most distinctive across the metaclusters.
3. We plotted the metaclusters (identified by Step 2) with age group-specific dendrograms. A flat cut of the dendrograms at a common height threshold was used to distinguish the metaclusters in each age group, which were shown as subtrees of corresponding colors.
4. We visualized using contour plots the corresponding metaclusters of each age group to compare the distributions of the selected covariates across the metaclusters as well as the age groups.

The set of 9 covariates used in step 1, selected for their clinical relevance, are shown in Table 2. The R package ‘sparcl’²¹ was used for agglomerative and sparse hierarchical metaclustering in step 2; the arguments of the $wbound$ parameter were varied from 2 to 5.

3 Results

The CIFU pipeline was run with HRC NRR phenotype data collected with OCT and clinical assessment of a normal cohort consisting of 3973 healthy eyes. The steps of the pipeline began with stratification of the OCT and clinical data by age into 3 age groups with (1) 1841, (2) 1351, and (3) 781 samples respectively. The list of clinical variables are summarized in Table 1. An identical sequence of steps of analysis was followed by CIFU within each age group.

The 180-point HRC data for each sample (eye) were modeled using $p = 11$ Fourier basis functions. We chose $p = 11$ since it was the smallest value of p for which FVE , as given in equation (6), exceeded 99% (Supplementary Figure S1). The curves were normalized and aligned to a common starting angle of 0 degrees to allow for comparison of their shapes around the center of ONH. Using equation (3), the outlier OCT samples were removed: 6 samples from age group 1, 1 from age group 2, and 5 from age group 3. Then, within each age group, the curves were clustered by a discriminative functional mixture (DFM) model as described in equation (2). The number of clusters (K) for each age group was determined by 3 different well-known criteria: AIC, BIC, and ICL. These criteria showed overall strong agreement attesting to optimal model selection as seen in Figure 1. Thus, we determined K , the number of clusters, for age group 1, 2, and 3 as 7, 8, and 6 respectively. The size of each cluster is shown in Table 2.

The results of our circular functional clustering are shown in Figure 2(a)-(c) as a panel of K clusters for each age group. Each cluster C within a panel consists of the circular curves for the samples that belong to C (all shown in a common color specific to C) based on the similarity of their functional representation. To gain an intuitive understanding of the 180-point HRC OCT data on NRR phenotypes, we used a visualization of curves as represented on a common circular scale. Unsupervised clustering of the circular functions revealed various NRR patterns in the identified clusters, some of which were distinctive whereas others have subtle differences. Notably, the visualization reveals the unique mean shape (or NRR “template”) of each cluster as shown by a bold black circular curve in each plot of Figure 3.

The circular curve visualization allows several interesting observations. We note the consistent dip at the temporal (T) region near 0 degrees, which is a characteristic feature shared by the templates of all clusters. This is supported by the well-known ISNT rule²² according to which, from the center of ONH, the rim is the thinnest at the temporal (T) region. Interestingly, we also observed various shapes and features in the cluster templates (such as distinctive protrusions, notches, tilts, etc.) that appear as well as vary continuously in different (non-T) regions around the circle. The clustering solution allows us to record the intra-cluster variation which could be used to quantitatively compare the dynamics (say, the rates of focal change) of corresponding clusters across age groups. In this regard, we note that non-circular (PAM) clustering of the same NRR HRC data yielded only 2 clusters of low phenotypic variation in every age group (Supplementary Figures S2 and S3) as noted by their small values of Dunn Index for every age group (1: 0.0548; 2: 0.044; 3: 0.0557).

In order to establish a correspondence among the clusters in different age groups as well as to characterize the samples that belong to each cluster, we conducted a metaclustering analysis. In this step, we clustered the clusters based on a set of 9 clinical variables of the samples therein. These are known covariates of glaucoma, and no NRR data from the previous clustering step was used. The results of sparse hierarchical metaclustering are shown in Figure 5. The dendrograms reveal the similarities among the clusters in terms of their mean sample covariates as well as the counts of metaclusters identified at different levels of each dendrogram. Based on flat cuts of all the dendrograms (at a common height threshold of 0.1), we identified 3 metaclusters $\{M_1^1, M_2^1, M_3^1\}$ for age group 1; 2 metaclusters $\{M_1^2, M_2^2\}$ for age group 2; and 2 metaclusters $\{M_1^3, M_2^3\}$ for age group 3. Notably, all the dendrograms show the metacluster M_2^\bullet (pink subtree) to be more heterogeneous in every age group than the metacluster M_1^\bullet (blue subtree). Among the youngest participants, i.e., in age group 1, the metacluster M_3^1 (consisting of the original cluster 4) is distinct from the metacluster M_2^1 .

A feature selection step, performed along with metaclustering, identified the optic disc cup volume and the average cup-to-disk ratio (CDR) of an eye as the most significant features in terms of the contributions of the different covariates to the metaclustering. These distinctive covariates allow us to register the correspondence of the metaclusters across the different age groups in Figure 6,

which shows the contour plots of the metaclusters in their matched colors. The 3 metaclusters M_1^\bullet shown in blue have the smallest mean values of cup volume and CDR, the 3 metaclusters M_2^\bullet shown in pink have higher mean values of these covariates, and the single unmatched metacluster M_3^1 shown in red has the highest mean values (Table 2).

It is interesting to consider the unmatched metacluster M_3^1 which not only has the highest mean values of the covariates (cup volume and average CDR) but is, in fact, comprised of a single, distinct cluster based on the OCT NRR phenotype data (Figure 5a). Here we note that notwithstanding a large value of CDR (especially > 0.5), cupping by itself is not indicative of glaucoma. In fact, it is known that deep but stable cupping can occur due to hereditary reasons without glaucoma (see Discussion). Rather, it is a change in these ONH parameters with age of the participants that is a clinical indicator of glaucoma. Since the samples included in the present study contain only healthy eyes as determined clinically by agreement of multiple glaucoma specialists, the presence of this unmatched cluster only in the youngest age group serves as a signature of healthy NRR phenotype with predictive potential for glaucoma. That is, the corresponding metacluster with such high values of these covariates among the older age groups would have the likelihood of progressing to glaucoma, and thus, is unlikely to be represented in a cohort of only healthy eyes, as we have in the present study.

4 Discussion

Unsupervised learning of the heterogeneity of normative ONH phenotypes in a given population can provide a more comprehensive understanding of the diversity of baselines that exist for degenerative neuropathies. Such knowledge is particularly useful in the case of glaucomas for which different ONH parameters play a combined role in early detection. For example, in a non-glaucoma multiethnic cohort of Asian individuals, the inter-eye RNFL profile was found by OCT to be less symmetric in Malays and Indians than that in Chinese²³. Not only are the structural characteristics of individual eyes known to vary racially, even their rates of change over time could be different across population groups. For instance, the rate of change of BMO-MRW was recorded as $-1.82 \mu\text{m}/\text{year}$ and $-2.20 \mu\text{m}/\text{year}$ in glaucoma suspect eyes of European and African descents respectively²⁴. In another multi-centred normal population study, both age-related decline and between-subject variability in BMO-MRW were observed²⁵. Indeed, even the manufacturers

of OCT technology noted racial differences in optic disc area, CDR, cup volume, and RNFL thickness when measured using their platform²⁶.

The presence of phenotypic heterogeneity makes it less justified to apply common, universal thresholds for clinical determination of glaucomatous damage in different population groups using OCT measurements, particularly in the early stages of the disease when the baselines could have stronger initial effects. To account for the effects of normal variation in ONH parameters, large and racially representative normative databases of healthy eye OCT phenotypes should be created. However, often such collections of healthy samples tend to be small or moderately sized, e.g., the normative database of Cirrus HD-OCT platform included just 284 subjects²⁶. In that cohort, Caucasians represented 43%, Chinese 24%, African Americans 18%, Hispanics 12%, and others 6%. The representation of Asian Indians, in contrast, was about 1% of the Cirrus HD-OCT cohort, which does not adequately reflect the 2020 projections about India to become the second in global glaucoma numbers, surpassing Europe²⁷. Thus, more than 16 million Indians could be affected by glaucomas, and nearly 1.2 million get blind from the disease. Some resources such as the HRT3 Normative Database, while including 104 Indian individuals, did not improve the diagnostic sensitivity or specificity for glaucoma in that group for the potential reasons of limited sample size and intra-racial variation of ocular topography²⁸.

In this study, we leverage on the large population-based LVPEI-GLEAMS study to generate new and relatively large OCT dataset based on nearly 4000 samples from normal Asian Indian participants. In fact, given that the recruitment of all the study participants was from a single geographic region (namely, the state of Andhra Pradesh), the scope of intra-racial variation to affect this dataset is limited. Moreover, the relatively large sample size of the data allowed us to identify a variety of clusters of NRR phenotypes, including the signature (cluster 4) with predictive potential for glaucoma consisting of 6.6% of all samples in the youngest age group (40-49 years). The absence of its corresponding cluster in the older healthy age groups despite their considerable sample sizes (total of 2132 samples of age 50 years and above) leads to a reasonable supposition of its potential pathological progression with increase in age, thereby resulting in lack of subsequent representation in a healthy cohort such as in the present study. Such phenotypic decline

is consistent with the findings from a prospective longitudinal study that found the rate of age-related, glaucomatous global (or global percentage) rim area loss to be 3.7 (5.4) times faster as compared to healthy eyes²⁹.

Importantly, independent support for the identified signature relies on its clinical characterization in terms of covariates such as cup volume and CDR, which are useful parameters for diagnosis of glaucoma suspect³⁰. Despite its normal mean value of intraocular pressure (IOP) as is expected of healthy eyes, the signature cluster has the highest mean values of average CDR and cup volume of all metaclusters across all 3 age groups (Table 2). In a recently published longitudinal study that started from baseline values and was run over a 5-year period, the covariates which had statistically significant increase in glaucomatous progression included CDR and cup volume³¹. We understand that it would perhaps be ideal to follow-up healthy individuals and measure the changes in their clinical covariates as they age in order to classify the ONH phenotypes via supervised learning. The approach of CIFU, in comparison, involves unsupervised learning of different high-resolution phenotypes in age-stratified data and their characterization using key covariates, which is far less time-consuming and yet has the potential to produce a clinically insightful database for diverse populations with high phenotypic heterogeneity.

In addition to its sample size and racial representation, perhaps the most noteworthy feature of the present OCT dataset is its unique HRC measurements of NRR thickness. These samples, collected at every 2 degrees, extend the typical use of such measurements recorded at either 4 quadrants or 12 clock-hours, and even 48 angular positions³², to higher-dimensional analysis. As clustering with curves show, the focal variations could be more nuanced than that suggested by a general rule, e.g., ISNT, and a capability to “zoom” into finer angular divisions can reveal further patterns³³. In low-resolution data, it may be difficult to detect focal changes within the confines of pre-determined inflexible sectors. Moreover, the templates of the clusters could also be compared using known tests in shape analysis³⁴. Be it circular data from OCT, or optic phenotypes in general, they seem suitable as candidate applications of circular statistics, and yet, we are unaware of any major previous studies in this regard. Further, the high-resolution also allowed the HRC data to be closely approximated by continuous curves, and thus, specified by the corresponding functional representation. While clustering of circular point data^{12,35,36} as well as clustering of curves³⁷⁻³⁹ and

(non-circular) functional data^{13,14,19,40-42} have been addressed by past studies, the present clustering of curves in the form of circular functions is possibly a novel application.

We understand that the present study has certain limitations. As we noted above, a prospective cohort study would be better suited to validate the predictive glaucomatous potential of the identified NRR phenotypic signature. We plan to address this in our future work. While HRC data could be accessed from the OCT scans, it is not commonly done by the clinical protocols. We hope that by adding user-friendly interfaces to CIFU, we may promote such data acquisition and analysis. Indeed, there are several distinct advantages of our approach which could be built upon further in future studies. The numeric representation of the functional mixture model parameters could be used to compare the normal against disease ONH phenotypes allowing us to characterize any changes therein with precision and rigor. Once a database of phenotypic parameters is developed, known measures of shapes and distances between curves could be used for objective clinical classification of new samples. Applied to longitudinal analyses, our high-resolution modeling could identify intermediate, or previously uncharacterized, stages of disease progression, especially by focusing on variations within fine angular sections. Straightforward extensions are feasible for similar circular data such as RNFL phenotypes and other optic neuropathies as well as related eye imaging platforms, e.g., OCT-Angiography (OCTA). As we have demonstrated for other biomedical platforms^{39,43-46}, the new pipeline CIFU could be enhanced incrementally with different functionalities, say, to increase computational efficiency or capture the perspective of the clinical experts. The circular curve visualization introduced in the present study may lead to a more user-friendly tool for clinical purposes as we plan to make it interactive, with advanced capabilities to jointly handle data and metadata, in our future work.

Acknowledgements

AP is supported by NSF grant DMS-1811888. The population-based cohort was funded by Hyderabad Eye Research institute.

Figures

Figure 1: The values of model selection criteria AIC, BIC, and ICL corresponding to fitting of a DFM model of K clusters to OCT NRR samples of age group 1 shown in (a), 2 in (b) and 3 in (c). The optimal DFM models for the age groups 1, 2 and 3 were selected for $K=7, 8$ and 6 respectively, beyond which no significant gain was noted.

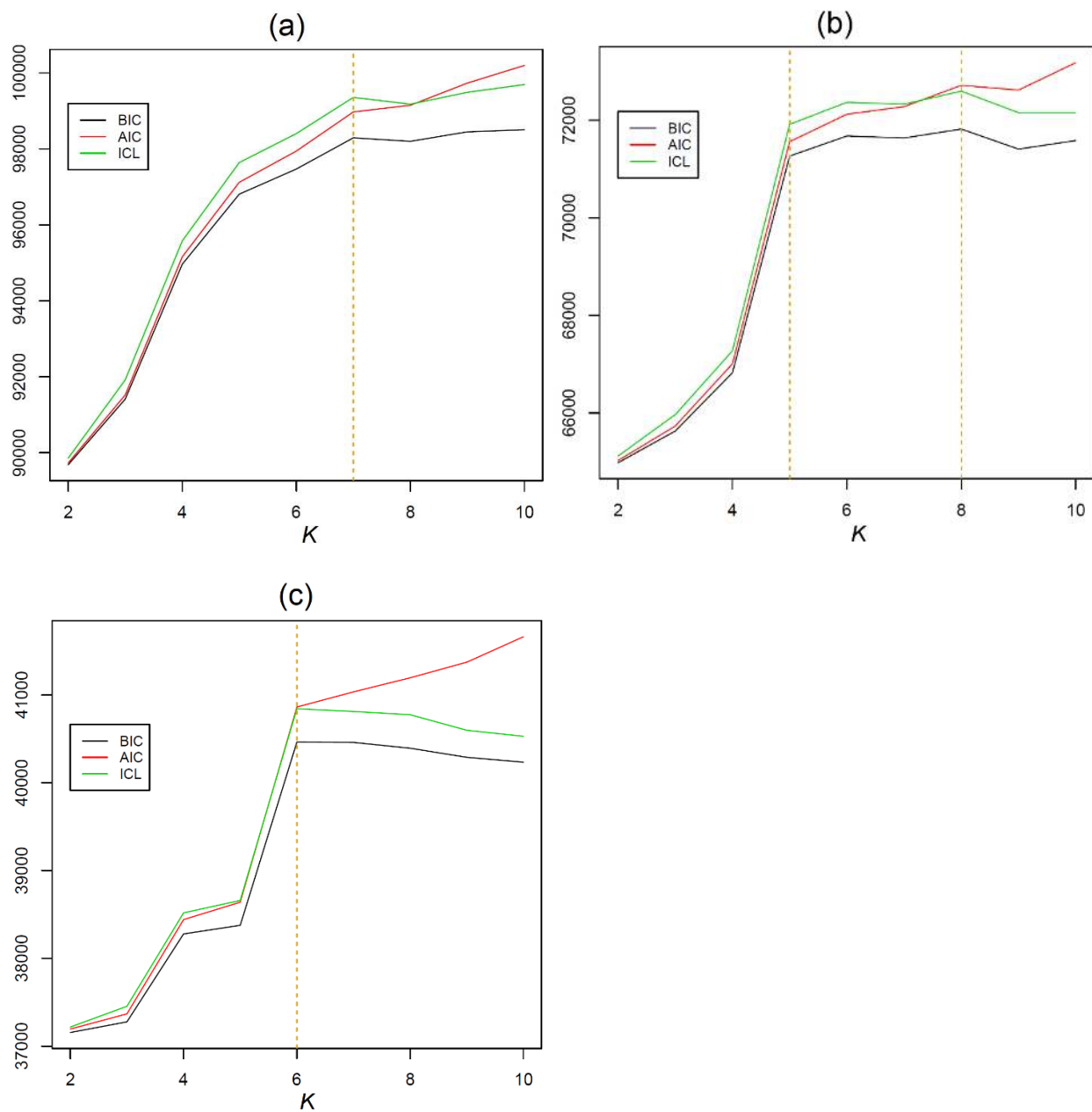
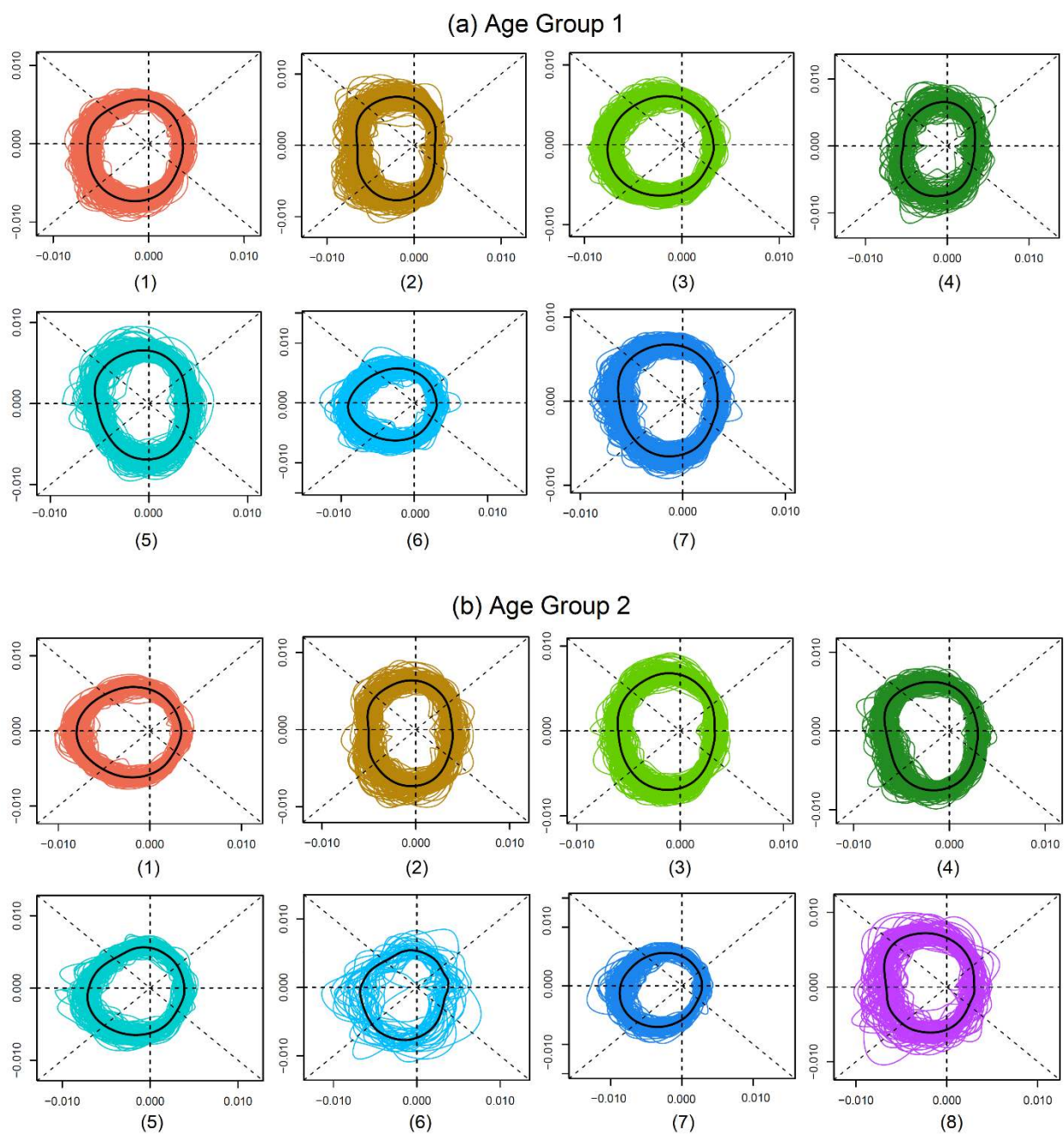


Figure 2: The results of clustering the OCT NRR data of age groups 1, 2, and 3 are shown respectively in panels (a), (b), and (c) using a circular functional representation. The curves belonging to each cluster are shown in a distinct color within a panel. The sample mean of each cluster is shown as a bold black circular curve.



(c) Age Group 3

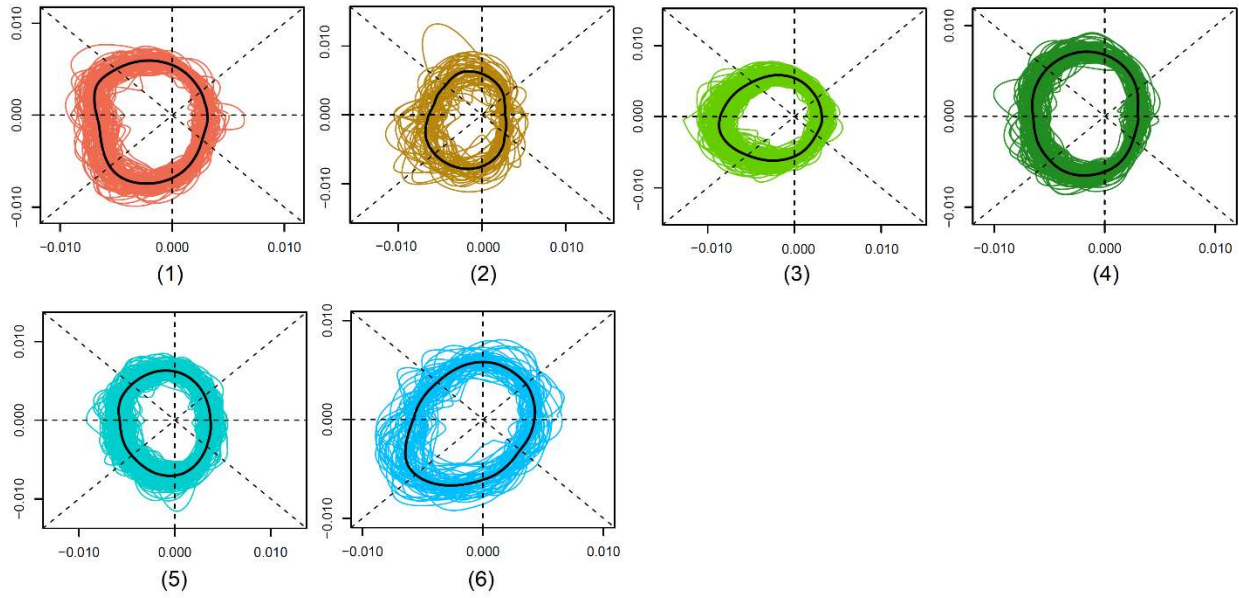


Figure 3: The results of metaclustering are shown using dendrograms for age groups (a) 1, (b) 2, and (c) 3. The leaves denote the Id-s of the circular functional clusters of OCT NRR data. The metaclusters are obtained by a flat cut of each dendrogram at the common height of 0.1, and the labels and subtrees representing them are shown in different colors.

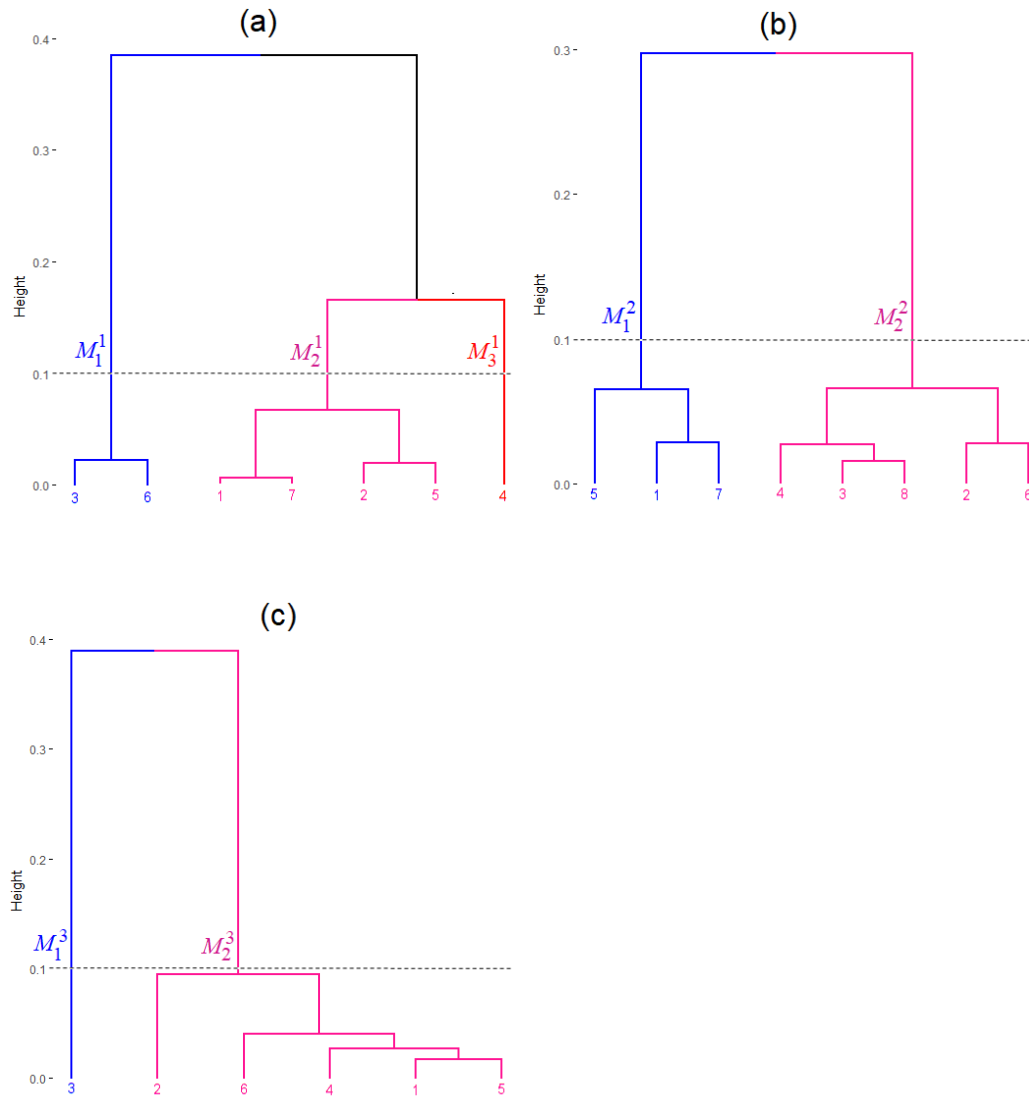
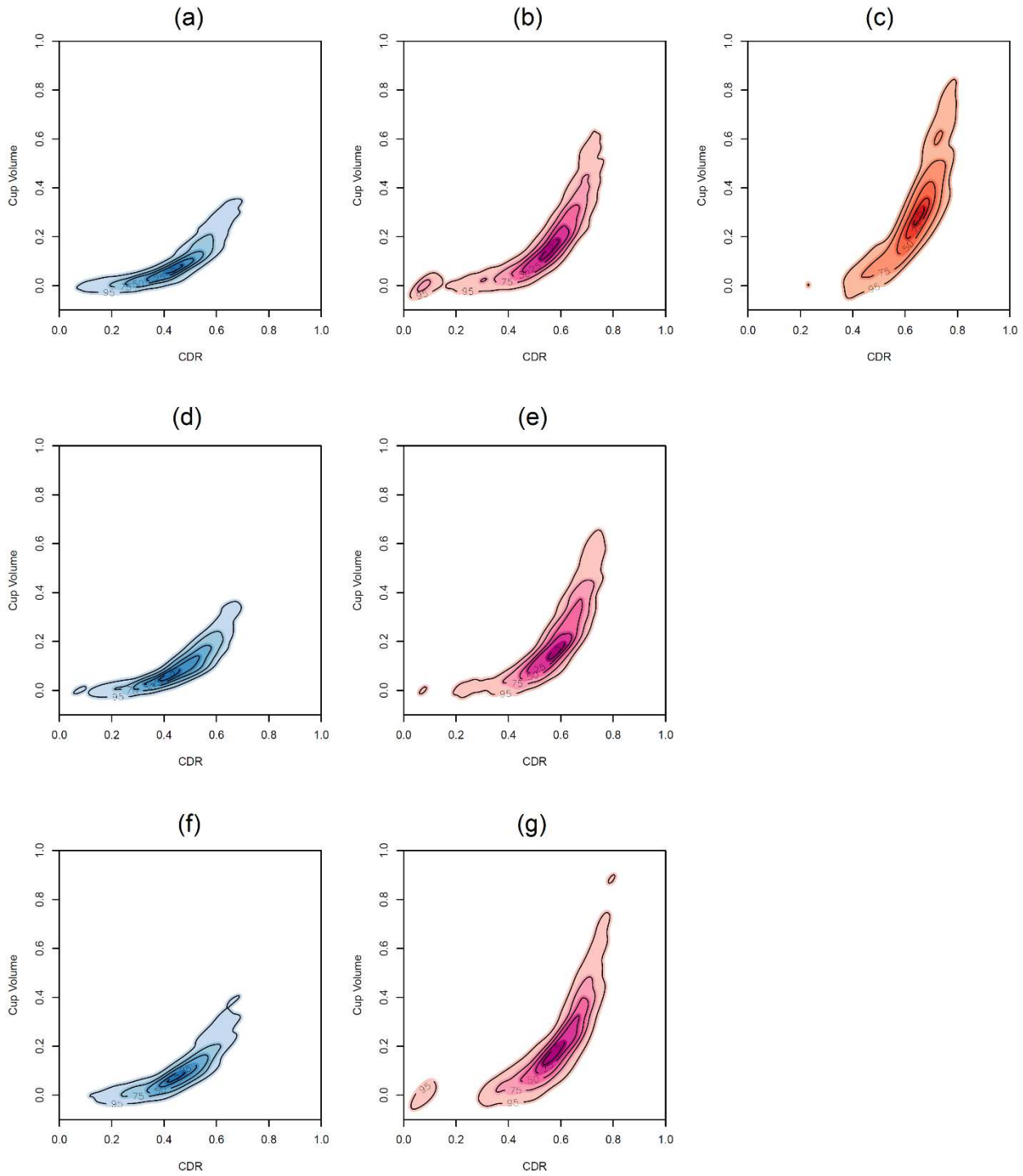


Figure 4: Contour plots of the distributions of clinical covariates volume and average cup CDR of the samples belonging to metaclusters of age groups 1: (a) M_1^1 , (b) M_2^1 , (c) M_3^1 ; 2: (d) M_1^2 , (e) M_2^2 ; and 3: (f) M_1^3 , (g) M_2^3 . The corresponding metaclusters are shown in matched colors.



Tables

Table 1: The clinical variables of the study participants in the three age groups. The units are given in parentheses. The asterisk(*) denotes that a variable is described as mean±sd. (N: Number of samples; OD: Oculus Dexter; OS: Oculus Sinister; BCVA LogMAR: Best Corrected Visual Acuity Logarithm of the Minimum Angle of Resolution; IOP: IntraOcular Pressure; CCT: Central Corneal Thickness; CDR: Cup-to-Disc Ratio.)

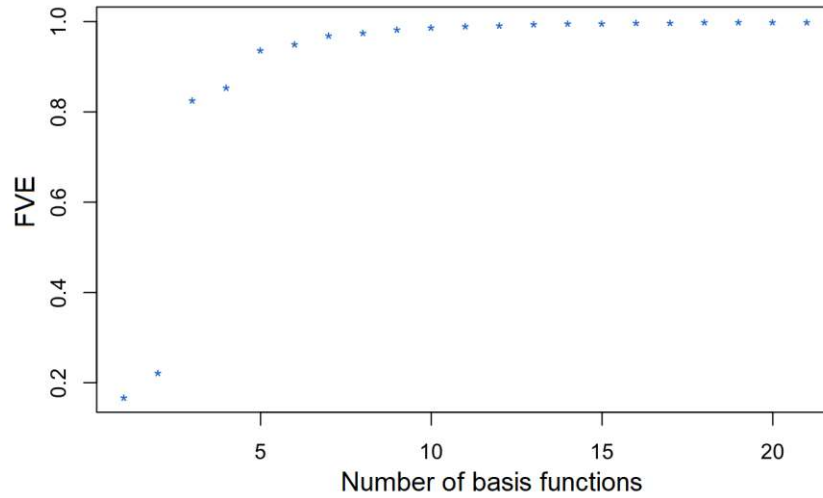
Clinical variables	Age Group1	Age Group2	Age Group3
Number of eyes, N (OD/OS)	1841(917/924)	1351(677/674)	781(387/394)
Age (years)*	43.95±2.83	53.18±2.76	64.63±5.39
Gender, N (Female/Male)	1256/585	755/596	372/409
BCVA LogMAR*	0.01±0.04	0.03±0.08	0.09±0.11
Spherical equivalent (diopter)*	0.02±0.75	0.16±1.05	-0.17±1.3
IOP (mmHg)*	12.58±2.34	12.49±2.41	12.13±2.38
CCT (µm)*	525.7±32.22	524.28±31.96	517.14±31.62
Axial length (mm)*	22.59±0.74	22.63±0.71	22.6±0.8
Family history of Glaucoma, n, (no/yes)	1835/6	1345/6	781/0
Diabetes mellitus, n, (no/yes)	1720/121	1163/188	666/115
Hypertension, n, (no/yes)	1696/145	1083/268	575/206
RIM AREA (mm ²)*	1.36±0.22	1.34±0.23	1.33±0.25
DISC AREA (mm ²)*	1.96±0.35	1.97±0.35	2.01±0.37
AVERAGE CDR*	0.52±0.15	0.53±0.14	0.55±0.14
AVERAGE THICKNESS (µm)*	94.36±9.25	92.3±9.57	90.73±9.9
VERTICAL CDR*	0.49±0.15	0.5±0.13	0.52±0.13
CUP VOLUME (mm ³)*	0.18±0.16	0.18±0.15	0.2±0.18
DISC DIAMETER (mm)*	1.49±0.15	1.51±0.15	1.53±0.16

Table 2: The clinical covariates used for metaclustering in the three age groups. The two most significant covariates due to feature selection are shown in bold. The values of each variable in a metacluster are described as mean±sd. The units are given in parentheses. (*N*: the number of samples; IOP: IntraOcular Pressure; CCT: Central Corneal Thickness; CDR: Cup-to-Disc Ratio.)

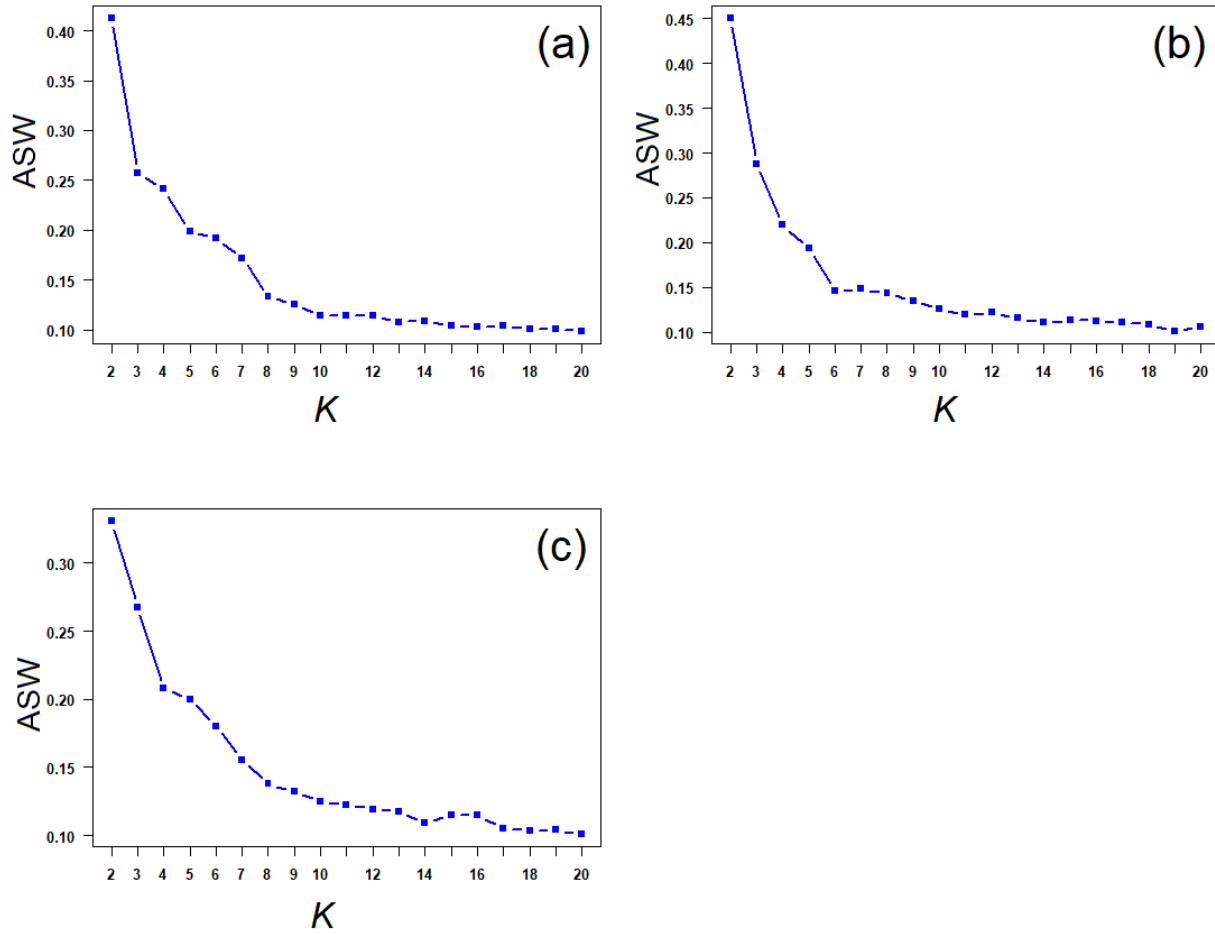
Metacluster {Clusters}	Age Group 1			Age Group 2		Age Group 3	
	<i>M</i> ₁ {3,6}	<i>M</i> ₂ {1,2,5,7}	<i>M</i> ₃ {4}	<i>M</i> ₁ {1,5,7}	<i>M</i> ₂ {2,3,4,6,8}	<i>M</i> ₁ {3}	<i>M</i> ₂ {1,2,4,5,6}
Metacluster <i>N</i> (percent)	549 (29.8)	1170 (63.6)	122 (6.6)	547 (40.5)	804 (59.5)	212 (27.1)	569 (72.9)
IOP (mmHg)	12.68±2.31	12.54±2.33	12.54±2.61	12.34±2.31	12.6±2.48	12.12±2.48	12.13±2.34
CCT (µm)	525.93±31.08	526.24±32.86	519.53±30.68	523.06±31.53	525.09±32.23	518.71±33.58	516.56±30.87
Axial Length (mm)	22.55±0.75	22.59±0.74	22.7±0.67	22.63±0.69	22.63±0.72	22.8±0.72	22.53±0.82
RIM AREA (mm ²)	1.44±0.23	1.32±0.21	1.25±0.19	1.42±0.24	1.29±0.21	1.42±0.25	1.3±0.24
DISC AREA (mm ²)	1.86±0.29	1.99±0.35	2.17±0.4	1.89±0.32	2.02±0.36	1.88±0.32	2.06±0.38
DISC DIAMETER (mm)	1.47±0.13	1.5±0.15	1.59±0.17	1.48±0.14	1.52±0.15	1.48±0.14	1.55±0.16
VERTICAL CDR	0.45±0.13	0.49±0.15	0.56±0.14	0.47±0.14	0.52±0.13	0.48±0.12	0.53±0.14
AVERAGE CDR	0.44±0.13	0.54±0.15	0.62±0.12	0.46±0.14	0.57±0.13	0.46±0.12	0.58±0.14
CUP VOLUME (mm³)	0.1±0.1	0.2±0.16	0.3±0.19	0.12±0.12	0.22±0.16	0.11±0.11	0.23±0.19

Supplementary Figures

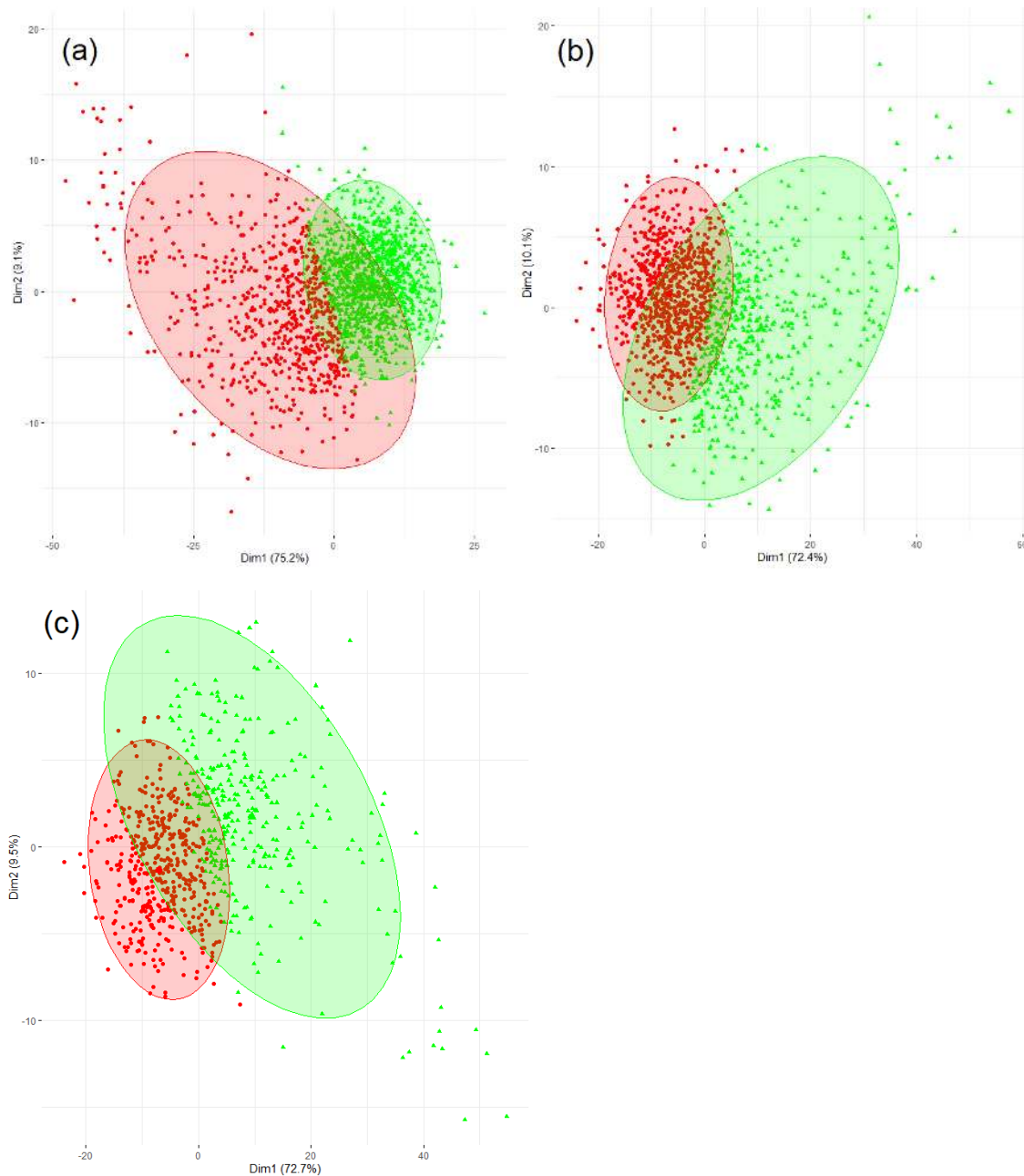
Supplementary Figure S1: The fraction of variation explained (*FVE*) by models with different choices of the number (p) of basis functions used for the functional representation of OCT data.



Supplementary Figure S2: Average Silhouette Width (ASW) for different choices of the number of clusters (K) due to PAM clustering of OCT HRC data for age groups (a) 1, (b) 2, and (c) 3. The value of ASW is maximum for $K=2$ for all three age groups.



Supplementary Figure S3: Results of PAM clustering OCT HRC data for age groups (a) 1, (b) 2, and (c) 3. Each scatterplot shows $K=2$ clusters (in red and green) as identified by PAM based on maximum ASW. Each 2D point represents a sample's first two principal components of HRC data.



References

1. Global Initiative for the Elimination of Avoidable Blindness: Action Plan 2006-2011. 2007. <https://apps.who.int/iris/handle/10665/43754>. Accessed January 24, 2021.
2. Tham Y-C, Li X, Wong T, Quigley H, Aung T, Cheng C-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081-2090.
3. Kerrigan–Baumrind LA, Quigley HA, Pease ME, Kerrigan DF, Mitchell RS. Number of Ganglion Cells in Glaucoma Eyes Compared with Threshold Visual Field Tests in the Same Persons. *Investig Ophthalmol Vis Sci*. 2000;41:741-748.
4. Garway-Heath D, Wollstein G, Hitchings R. Aging changes of the optic nerve head in relation to open angle glaucoma. *Br J Ophthalmol*. 1997;81(10):840-845. doi:10.1136/bjo.81.10.840
5. Prata TS, Dorairaj S, Trancoso L, et al. Eyes with large disc cupping and normal intraocular pressure: using optical coherence tomography to discriminate those with and without glaucoma. *Med hypothesis, Discov Innov Ophthalmol J*. 2014;3(3):91-98.
6. Kwon Y, Fingert J, Kuehm M, Alward W. Primary open-angle glaucoma. *New Engl J Med*. 2009;360(11):1113-1124.
7. Leung CKS, Lam S, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: analysis of the retinal nerve fiber layer map for glaucoma detection. *Ophthalmology*. 2010;117(9):1684-1691. doi:10.1016/j.ophtha.2010.01.026. Epub 2010 Jul 21
8. Leung CKS, Choi N, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: Pattern of RNFL defects in glaucoma. *Ophthalmology*. 2010;117(12):2337-2344. doi:10.1016/j.ophtha.2010.04.002
9. Mwanza JC, Chang RT, Budenz DL, et al. Reproducibility of peripapillary retinal nerve fiber layer thickness and optic nerve head parameters measured with cirrus HD-OCT in glaucomatous eyes. *Investig Ophthalmol Vis Sci*. 2010;51(11):5724-5730. doi:10.1167/iovs.10-5222
10. Mwanza JC, Oakley JD, Budenz DL, Anderson DR. Ability of cirrus HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes. *Ophthalmology*. 2011;118(2):241-248.e1. doi:10.1016/j.ophtha.2010.06.036
11. Sung KR, Na JH, Lee Y. Glaucoma diagnostic capabilities of optic nerve head parameters as determined by cirrus HD optical coherence tomography. *J Glaucoma*. 2012;21(7):498-504. doi:10.1097/IJG.0b013e318220dbb7
12. Jammalamadaka SR, SenGupta A. *Topics in Circular Statistics*. World Scientific, 2001. doi:10.1142/4031
13. Ieva F, Paganoni AM, Davide P, Valeria V. Multivariate functional clustering for the morphological analysis of electrocardiograph curves. *Appl Stat*. 2012;62(3):401-418. doi:<https://doi.org/10.1111/j.1467-9876.2012.01062.x>
14. Jacques J, Preda C. Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing*. 2013;112:164-171. doi:10.1016/j.neucom.2012.11.042
15. Wang J-L, Chiou J-M, Müller H-G. Functional Data Analysis. *Annu Rev Stat Its Appl*. 2016;3:257-295. doi:10.1146/ANNUREV-STATISTICS-041715-033624
16. Carl Zeiss Meditec. Inc D, CA, USA. Cirrus HD-OCT User Manual, 2015.
17. Thylefors B, Négrel A. The global impact of glaucoma. *Bull World Heal Organ*. 1994;72(3):323-326.
18. Addepalli UK, Jonnadula GB, Garudadri CS, et al. LV Prasad Eye Institute Glaucoma Epidemiology and Molecular Genetic Study (LVPEI- GLEAMS). Report 1: study design and research methodology. *Ophthalmic Epidemiol*. 2013;20(3):188-195. doi:10.3109/09286586.2013.792938
19. Bouveyron C, Côme E, Jacques J. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *Ann Appl Stat*. 2015;4:1726-1760.
20. Fraley C, Raftery AE. MCLUST: Software for Model-Based Cluster Analysis. *J Classif*. 1999;16(2):297-306. doi:10.1007/s003579900058

21. Witten D, Tibshirani R. A framework for feature selection in clustering. *J Am Stat Assoc.* 2010;105(490):713-726.
22. Jonas JB, Gusek GC, Naumann GO. Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes. *Invest Ophthalmol Vis Sci.* 1988;29(7):1151-1158.
23. Tao Y, Tham YC, Chee ML, et al. Profile of retinal nerve fibre layer symmetry in a multiethnic Asian population: The Singapore Epidemiology of Eye Diseases study. *Br J Ophthalmol.* 2020;104:836-841. doi:10.1136/bjophthalmol-2019-314894
24. Bowd C, Zangwill LM, Weinreb RN, et al. Racial differences in rate of change of spectral domain OCT-measured minimum rim width and retinal nerve fiber layer thickness. *Am J Ophthalmol.* 2018;196:154-164. doi:10.1016/j.ajo.2018.08.050
25. Chauhan, Balwantray C. Danthurebandara, Vishva M, Sharpe GP, Demirel S, et al. Bruch's Membrane Opening-Minimum Rim Width and Retinal Nerve Fibre Layer Thickness in a Normal White Population. A Multi-centre Study. *Ophthalmology.* 2015;122(9):1786-1794. doi:10.1016/j.ophtha.2015.06.001.Bruch
26. Knight OJ, Girkin CA, Budenz DL, Durbin MK, Feuer WJ. Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by cirrus HD-OCT. *Arch Ophthalmol.* 2012;130(3):312-318. doi:10.1001/archophthalmol.2011.1576
27. Quigley H, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol.* 2006;90(3):262-267. doi:10.1136/bjo.2005.081224
28. Rao HL, Babu GJ, Sekhar GC. Comparison of the Diagnostic Capability of the Heidelberg Retina Tomographs 2 and 3 for Glaucoma in the Indian Population. *Ophthalmology.* 2010;117(2):275-281. doi:10.1016/j.ophtha.2009.06.062
29. Hammel N, Belghith A, Bowd C, et al. Rate and pattern of rim area loss in healthy and progressing glaucoma eyes. *Ophthalmology.* 2016;123(4):760-770. doi:10.1016/j.ophtha.2015.11.018
30. Garway-Heath DF, Ruben ST, Viswanathan A, Hitchings RA. Vertical cup/disc ratio in relation to optic disc size: Its value in the assessment of the glaucoma suspect. *Br J Ophthalmol.* 1998;82(10):1118-1124. doi:10.1136/bjo.82.10.1118
31. Siesky B, Wentz SM, Januleviciene I, et al. Baseline structural characteristics of the optic nerve head and retinal nerve fiber layer are associated with progressive visual field loss in patients with open-angle glaucoma. *PLoS One.* 2020;15(8):1-11. doi:10.1371/journal.pone.0236819
32. Reis ASC, O'Leary N, Yang H, et al. Influence of clinically invisible, but optical coherence tomography detected, optic disc margin anatomy on neuroretinal rim evaluation. *Invest Ophthalmol Vis Sci.* 2012;53:1852-1860. doi:10.1167/iovs.11-9309
33. Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in early glaucoma. *Br J Ophthalmol.* 1998;82(4):352-361. doi:10.1136/bjo.82.4.352
34. Dryden IL, Mardia K V. Statistical Shape Analysis: With Applications in R, 2nd Edition. Wiley Series in Probability and Statistics. Published 2016. ISBN: 978-0-470-69962-1.
35. Abraham C, Molinari N, Servien R. Unsupervised clustering of multivariate circular data. *Stat Med.* 2013;32(8):1376-1382. doi:10.1002/sim.5589
36. Rodríguez CE, Núñez-Antonio G, Escarela G. A Bayesian mixture model for clustering circular data. *Comput Stat Data Anal.* 2020;143:106842. doi:10.1016/j.csda.2019.106842
37. Gaffney S, Smyth P. Joint probabilistic curve clustering and alignment. In: *Proceedings of the 17th International Conference on Neural Information Processing Systems.* 2004:473-480. <https://dl.acm.org/doi/10.5555/2976040.2976100>.
38. Heard NA, Holmes CC, Stephens DA. A quantitative study of gene regulation involved in the immune response of Anopheline mosquitoes: An application of Bayesian hierarchical clustering of curves. *J Am Stat Assoc.* 2006;101(473):18-29. doi:10.1198/016214505000000187
39. Ray S, Pyne S. A Computational Framework to Emulate the Human Perspective in Flow Cytometric Data Analysis. *PLoS One.* 2012;7(5):e35693. doi:<https://doi.org/10.1371/journal.pone.0035693>
40. James GM, Sugar CA. Clustering for sparsely sampled functional data. *J Am Stat Assoc.* 2003;98(462):397-408. doi:10.1198/016214503000189

41. Ray S, Mallick B. Functional clustering by Bayesian wavelet methods. *J R Stat Soc Ser B (Statistical Methodol)*. 2006;68(2):305-332. doi:10.1111/j.1467-9868.2006.00545.x
42. Jones PN, McLachlan GJ. Fitting finite mixture models in a regression context. *Aust J Stat*. 1992;34(2):233-240. doi:10.1111/j.1467-842X.1992.tb01356.x
43. Ho H, Pyne S, Lin T. Maximum likelihood inference for mixtures of skew Student-t-normal distributions through practical EM-type algorithms. *Stat Comput*. 2012;22:287-299.
44. Pyne S, Hu X, Wang K, et al. Automated high-dimensional flow cytometric data analysis. *PNAS*. 2009;106(21):8519-8524.
45. Pyne S, Lee S, Wang K, et al. Joint modeling and registration of cell populations in cohorts of high-dimensional flow cytometric data. *PLoS One*. 2014;9(7):e100334.
46. Qi Y, Fang Y, Sinclair D, et al. High-Speed Automatic Characterization of Rare Events in Flow Cytometric Data. *PLoS One*. 2020;15(2):e0228651.