ARTIFICIAL INTELLIGENCE

Suspect glaucoma detection from corneal densitometry supported by machine learning

Andrés García-Jiménez, Alejandra Consejo*

Department of Applied Physics, University of Zaragoza, Zaragoza, Spain

Abstract

Purpose: To discriminate suspect glaucomatous from control eyes using corneal densitometry based on the statistical modeling of the pixel intensity distribution of Scheimpflug images.

Methods: Twenty-four participants (10 suspect glaucomatous and 14 control eyes) were included in this retrospective study. Corneal biomechanics was assessed with the commercial Scheimpflug camera Corvis ST (Oculus). Sets of 140 images acquired per measurement were exported for further analysis. After corneal segmentation, pixel intensities corresponding to different corneal depths were statistically modeled for each image, from which corneal densitometry in the form of parameters α (brightness) and β (homogeneity) was derived. After data pre-processing, parameters α and β were input to six supervised machine learning algorithms that were trained, tested, and compared.

Results: There exists a statistically significant difference in α and β parameters between suspect glaucomatous and control eyes (both, P < 0.05/N, Bonferroni). From the implemented supervised machine learning algorithms, the K-nearest neighbors (K-NN) algorithm reached 83.93% accuracy to discriminate between groups only using corneal densitometry parameters (α and β).

Conclusion: Densitometry of the anterior cornea including epithelium on its own has the potential to serve as a clinical tool for early glaucoma risk assessment.

Introduction

Glaucoma is a multifactorial eye condition that is the leading cause of irreversible vision loss.[1,2] Until it reaches an advanced stage, glaucoma is an asymptomatic disease, so methods of early diagnosis are needed.[3] As glaucoma is a condition that damages the optic nerve, research works aiming toward early glaucoma detection were traditionally focused on the back of the eye, i.e., optic nerve, lamina cribrosa, and retina. However, there exists an increasing interest in investigating the cornea as a target tissue for glaucoma detection and management. The potential role of corneal thickness[4,5] and corneal biomechanics[6,7] in the assessment of glaucoma risk have been investigated. More recently, different studies have found statistically significant differences in corneal densitometry between glaucomatous and control eyes,[7-9] suggesting that corneal densitometry analysis could have clinical applications in the diagnosis and management of glaucoma.

* Corresponding author at: Department of Applied Physics, University of Zaragoza, C/ María de Luna 3, 50018, Zaragoza, Spain.
E-mail address: alejandra.consejo@unizar.es (A. Consejo).

Corneal densitometry is an objective measure of corneal clarity, and it is an indicator of corneal health.[10] Scheimpflug-based corneal tomography is considered the gold standard for corneal densitometry assessment, but, to date, it is exclusively available on one single commercial device. To overcome this limitation, our research group introduced a platform-independent methodology based on the statistical modeling of the pixel intensity distribution of Scheimpflug images that correlates very well with traditional densitometry (overall cornea, $r = 0.89$; $P < 0.001$).[11] The method characterizes corneal densitometry with two different parameters. The first parameter accounts for corneal tissue clarity, as traditional densitometry does, whilst the second parameter accounts for corneal tissue homogeneity. This complete approach toward corneal tissue characterization has proven high accuracy and precision in early keratoconus detection,[12,13] overpassing the performance of clinically available systems.[12,13] This method also investigated corneal densitometry in minor corneal hypoxia[14] and rare genetic conditions.[15] Moreover, a previous work, based on Optical Coherence Tomography (OCT), investigated corneal backscatter based on the statistical modeling of the pixel intensity distribution of glaucomatous corneas.[16] Their results indicated that glaucoma suspects have similar microstructural corneal tissue properties to those exhibited in glaucoma patients but markedly different from that of healthy controls.[16] Consequently, their results suggest that micro parameters related to corneal tissue integrity might be a valuable early glaucoma biomarker.

Additionally, machine learning techniques have repeatedly shown their usefulness in ophthalmology as an objective tool to assist practitioners in diagnosing certain conditions.[17-19] Glaucoma detection is one of the ocular pathologies where the most efforts regarding machine learning implementation have been made.[17] However, to the best of the authors' knowledge, corneal densitometry was not previously investigated as a valuable feature for suspect glaucoma detection.

Consequently, the current work aims to investigate whether corneal densitometry parameters estimated from statistical modeling of the pixel intensity distribution of Scheimpflug images, with the support of supervised machine learning algorithms, can successfully discriminate between suspect glaucomatous and control eyes.

Methodology

Subjects and data collection

Twenty-four participants (24 eyes) were included in this study. Data was retrospectively collected from a previous study.[16] The study adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants. They were adult participants (19 females, 5 males) between 51 and 73 years old. The suspect glaucoma group (G) consisted of 10 eyes. The remaining 14 eyes served as the healthy control eyes (C). As age, IOP was similar in both groups. Participants had an IOP of less than or equal to 18.5 mmHg. As this was a retrospective study, participants were classified as suspect glaucoma or control according to the classification done by an experienced ophthalmologist in the corresponding prospective study.[16] In that prospective study,[16] IOP was assessed with Goldmann applanation tonometry; the optic nerve head, including disc area, rim area, rim volume, cup shape measure, height variation contour, and the mean thickness of the retinal nerve fiber layer was examined using Heidelberg Retinal Tomography (HRT 3; Heidelberg Engineering GmbH, Heidelberg, Germany); visual field parameters including mean deviation and pattern standard deviation with Humphrey Field Analyzer II 750, 24-2 Swedish interactive threshold algorithm (Carl Zeiss Meditec, Inc., Dublin, CA); biometry including the central corneal thickness, anterior chamber depth, axial length, and keratometry was measured with IOLMaster 700 (Carl Zeiss Meditec, Inc., Jena, Germany). Additionally, all participants underwent a comprehensive ophthalmological examination, including corneal biomechanics assessment with the Corvis ST (Oculus, Wetzlar, Germany), a high-speed dynamic Scheimpflug imaging system based on mechanical stimulation of the cornea. Exclusively Corvis ST data, including Scheimpflug images and biomechanical data, were utilized in the current study. Eyes with suspect glaucoma were defined as those with glaucomatous optic disc appearance (i.e., exhibiting disc rim loss or excavation, subjectively assessed optic disc cupping of >0.6, or localized abnormalities of the retinal nerve fiber layer) but normal visual fields. None of the suspects were pharmacologically treated for glaucoma. Healthy controls had healthy optic disc appearance and normal visual field, defined as the absence of glaucomatous and neurologic field defects. Subjects were excluded if they had a history of ocular surgery within 12 months before the study or if they were younger than 40 years with corneal or intraocular disease or neurologic disorders affecting visual fields.

Data analysis

Image processing. Parameters α and β

Corvis ST corneal deformation grayscale images were obtained every 230 μs and had a fixed size of 576×200 pixels, corresponding with an estimated resolution of 14.8 μm/pixel horizontally and 24 μm/pixel vertically. Examples of Corvis ST images along the corresponding corneal histograms are shown in Fig. 1.

After data acquisition, the sets of 140 images acquired per measurement were exported for further analysis, as explained in detail in previous work.[12] In short, the method of data analysis consisted of two steps: corneal segmentation and statistical modeling of the pixel intensity distribution. The axial and lateral dimensions of the ROI were optimized to maximize the amount of signal-carrying data or, in other words, to achieve the highest discriminative power between the investigated groups (suspect glaucomatous vs. control eyes). As a preliminary analysis, the horizontal (lateral) dimension of the ROI was fixed to 160 pixels, corresponding approximately to the central 2.5 mm corneal chord. Regarding the vertical (axial) dimension, three different depths were investigated: (1) full corneal thickness, (2) Stroma, and (3) Anterior cornea including epithelium. Consequently, the vertical dimension of the ROI depended on the corneal depth analyzed. Firstly, in the case of considering the full corneal thickness, the number of vertical pixels in
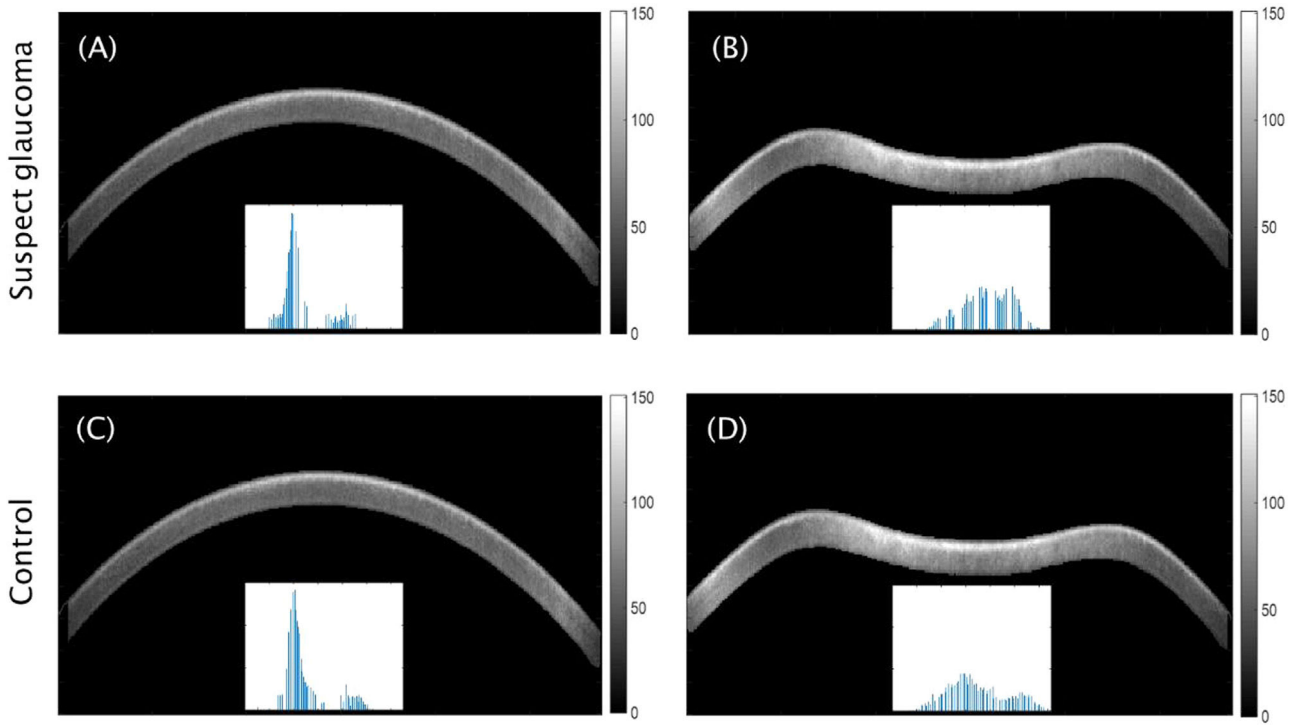
Fig. 1 Corneal Scheimpflug images after segmentation (i.e. without background) of a randomly chosen suspect glaucoma participant (A, B) and a randomly chosen control participant (C, D); before air-puff (A, C) and during air-puff (B, D). Colour bars indicate pixel intensity. Scaled histograms corresponding to corneal pixels are shown for each image.

the ROI was given by the corneal thickness of each subject, as it was investigated in previous work that corneal thickness is not a co-founding factor affecting the statistical analysis of corneal Scheimpflug images.[12] Secondly, when considering the stroma, pixels corresponding to corneal epithelium were omitted, as they carry different statistical information compared to those from the stroma.[11] In the third scenario, anterior cornea including epithelium, a fixed vertical ROI of 8 pixels from the detected anterior corneal border, approximately corresponding to the anterior 190 μm, was considered.

The intensity of the pixels corresponding to a given ROI was statistically modeled with the two-parameter distribution functions that were previously used to model corneal Scheimpflug images (i.e., Weibull, Gamma, and Lognormal distributions).[12] Output parameters were estimated using the method of maximum likelihood from the pixel intensities of the selected ROI in each image. The goodness of fit was assessed by means of the root mean squared error (RMSE) for the three candidate distributions, as in previous literature.[12]

As described in previous work in detail,[12] the output of the candidate two-parameter distribution functions is two parameters that will be further used to feed the supervised machine learning algorithms. These two parameters are the scale parameter (referred to as α in this work) and the shape parameter (referred to as β in this work). Parameter α accounts for corneal clarity (the larger α, the lesser corneal clarity), while parameter β accounts for corneal tissue homogeneity (the larger β, the greater corneal tissue homogeneity).[14] Parameters α and β can have any non-zero positive value.

Cases of study and statistical analysis

As mentioned before, different corneal depths (full cornea, stroma, and anterior cornea including the epithelium) and different statistical models (Weibull, Gamma, and Lognormal distributions) were considered to discriminate between groups (suspect glaucomatous vs. control eyes). Mean group values of $\alpha(t)$ and $\beta(t)$ parameters under every condition were analyzed. Further, individual values of α and β corresponding to the statistical model and corneal depth that would better discriminate between suspect glaucomatous and control eyes were used to feed the supervised machine learning classification algorithms. In addition, two cases were investigated. Case 1 included α and β estimated from the 140 frames available (i.e., 6720 data points = 2 parameters/frame x 140 frames/participant x 24 participants), while case 2 included α and β estimated from the 20 frames range that showed the best group mean discrimination between groups (i.e., 960 data points = 2 parameters/frame x 20 frames/participant x 24 participants).

Statistical analysis was performed using Microsoft Office Excel (Microsoft Office Professional Plus 2016; Microsoft; Redmond, WA, USA). The normality of each parameter, including α and β per frame and biomechanical parameters, was not rejected (Shapiro–Wilk test, $P > 0.05$). The independent two-sample t-test was used to assess differences in parameters under evaluation between suspect glaucoma and control groups. The paired two-sample t-test was used to assess differences in α and β parameters between groups. The level of significance was set to 0.05, and a Bonferroni correction was applied to address the problem of multiple comparisons.

Machine learning classification algorithms

Machine learning classification algorithms were implemented in Python (version 3.8, Python Software Foundation) and were based on the scikit-learn data science library (https://scikit-learn.org). Six supervised algorithms were implemented and tested: (1) Logistic regression, (2) K-nearest neighbors (K-NN) with Euclidean distance and K optimized for each case under investigation (Fig. 2), (3) Kernel support vector machine (SVM) with Gaussian RBF kernel, (4) Naïve Bayes, (5) Decision tree classification, and (6) Random forest classification.

The dataset, imported in .csv format, consisted of two features (α and β) and a labeled output (1 if suspect glaucoma or 0 if control). In case 1 (140 frames) a total of 3360 instances were used as the dataset, while in case 2 (20 frames) a total of 480 instances were used as the dataset.

The architecture of each supervised algorithm follows the same pattern. For data pre-processing, after importing the dataset, missing values were removed. In case 1, 3275 instances remained after missing values depletion (i.e., 2.5% of the whole dataset which was considered acceptable). There was no missing data in the dataset corresponding to case 2.

Each dataset was split into the training set (80 % of the dataset, randomly selected but fixed to train the different algorithms) and the test set (the remaining 20 % of the dataset). This is, in case 1 (140 frames), from the 3275 available instances, a total of 2620 instances were used to train the algorithms, while the remaining 655 instances were used to test the performance of the trained algorithms. Similarly, in case 2 (20 frames), from the 480 available instances, a total of 384 instances were used to train the algorithms, while the remaining 96 instances were used to test the performance of the trained algorithms. Further, feature scaling of α and β based on the standardization technique was applied to the test set and the test set separately in both cases under investigation.

After data pre-processing, each algorithm was trained on the training set and evaluated using the test set. To investigate for potential overfitting each algorithm was also evaluated using the training set. The confusion matrix and the corresponding accuracy (% of correct predictions) were calculated to assess the performance of each architecture. In addition to the statistical analysis, the performance of each test set for each case and architecture was plotted for illustrative purposes.

Results

Biomechanical parameters

Participants from this study had equivalent mean values of age, IOP, and CCT, as shown in Table 1. No statistically significant difference was found in biomechanical parameters between suspect glaucomatous eyes and control eyes, as indicated in Table 1.

Statistical analysis of Scheimpflug images

No statistically significant differences were found between the candidate functions to fit pixel intensity distribution within the ROI (all pair-wise comparisons, P > 0.05), as indicated in Table 2. The Weibull distribution was chosen to fit the pixel intensity distribution in corneal Scheimpflug images. This choice was motivated for two reasons. One reason is that the Weibull function reported the smallest RMSE (Table 2), even though this difference was not statistically significant. The second reason was to maintain consistency with previous literature where the Weibull distribution function was chosen to fit pixel intensity distribution in corneal Scheimpflug images representing various corneal pathologies[12,13,15] and healthy volunteers.[11,14]

The mean parameters $\alpha(t)$ and $\beta(t)$ were good discriminators between suspect glaucoma and control eyes (Fig. 3), with statistically significant differences for both $\alpha(t)$ (paired t-test, P < 0.05/N [N = 140, Bonferroni]) and $\beta(t)$ (paired t-test, P < 0.05/N [N=140, Bonferroni]). This was a good indicator that $\alpha(t)$ and $\beta(t)$ could serve as discriminators between groups. Hence, case 1 of the study used $\alpha(t)$ and $\beta(t)$ parameters estimated from 140 frames. However, from the analysis of the three investigated corneal depths and the dynamic evolution of $\alpha(t)$ and $\beta(t)$ it was concluded that the best discrimination between groups was achieved when analyzing the first frames of the cornea; this is, before mechanical stimulation, as illustrated in Fig. 3. Consequently, individual values of $\alpha(t)$ and $\beta(t)$ corresponding to the first 20 frames of each participant were used as case 2 of the study.

Machine learning classification algorithms

Besides the statistically significant difference achieved in mean $\alpha(t)$ and $\beta(t)$ between groups, the individual $\alpha(t)$ and $\beta(t)$ parameters represented messy data, as shown in Fig. 4. Consequently, standard statistics could not help predict whether a new data point would belong to the suspect glaucomatous group or the control group. Therefore, supervised machine learning algorithms needed to be implemented and tested for both study cases separately.

Figs. 5 and 6 illustrate the performance of the implemented supervised machine learning algorithms. The
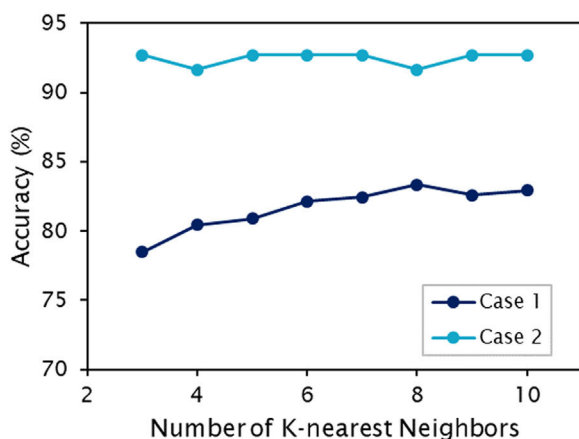

Fig. 2 Performance of the KNN algorithm as a function of the number of K-nearest neighbors chosen. The highest accuracy was reached with K=8 (83.35% accuracy) for case 1 (140 frames), and with K=5 (92.71% accuracy) for case 2 (20 frames).

Table 1 The group statistics for age and biomechanical parameters acquired with Corvis ST: mean values ± SD (range).

Parameter	Glaucoma suspects (n=10)	Control (n=14)	P-value
Age	66 ± 6 (51, 71)	60 ± 7 (48, 73)	0.07
IOP corrected (mmHg)	15.4 ± 1.5 (12.6, 17.3)	16.0 ± 1.4 (13.5, 19.3)	0.71
IOP non-corrected (mmHg)	15.2 ± 1.1 (13.2, 17.0)	16.7 ± 1.0 (13.5, 18.2)	0.07
CCT (μm)	543 ± 22 (510, 577)	565 ± 33 (506, 605)	0.16
A1 − length (mm)	2.2 ± 0.3 (1.8, 2.8)	2.2 ± 0.3 (1.8, 2.9)	0.90
v1 − velocity (m/s)	0.16 ± 0.02 (0.10, 0.19)	0.17 ± 0.01 (0.13, 0.17)	0.75
A2 − length (mm)	1.63 ± 0.50 (0.86, 2.21)	1.85 ± 0.40 (0.98, 2.21)	0.23
v2 − velocity (m/s)	−0.33 ± 0.07 (−0.43, −0.19)	−0.31 ± 0.03 (−0.35, −0.25)	0.40
Peak distance (mm)	5.13 ± 0.39 (4.3, 5.8)	4.83 ± 0.29 (4.17, 5.14)	0.05
Radius of curvature (mm)	7.17 ± 0.50 (6.58, 8.07)	7.54 ± 0.75 (6.32, 8.81)	0.24
Maximum deformation amplitude (mm)	1.18 ± 0.15 (0.95, 1.42)	1.08 ± 0.10 (0.86, 1.2)	0.07

background of the subplots corresponds to the training phase, while the data points correspond to the test set. In other words, suspect glaucomatous eyes (in orange in Figs. 5 and 6) should fall in the orange background while control eyes (in green in Figs. 5 and 6) should fall in the green background, otherwise, there is a misclassification. Corresponding statistics are shown in Table 3. The performance of the algorithms was measured by counting the number of true negatives (number of control eyes correctly classified as control eyes), true positives (number of suspect glaucomatous eyes correctly classified as suspect glaucomatous), false negatives or type II error (number of suspect glaucomatous eyes wrongly classified as control), false positives or type I error (number of control eyes wrongly classified as suspect glaucomatous). Table 4 shows the corresponding statistics for the training set.

Discussion

Diagnosing early-stage glaucoma is still a clinical challenge. This study shows that it is possible to discriminate, at over 80 % accuracy, suspect glaucomatous from control eyes using corneal densitometry parameters (α − corneal tissue clarity, β − corneal tissue homogeneity) with the support of supervised machine learning algorithms (Figs. 5, 6 and Table 3). The results of the current study are in agreement with those of Iskander and colleagues.[16] Using a similar image processing technique as in the present work, they applied statistical modeling of the pixel intensity distribution (i.e., speckle

distribution) of commercial OCT images to investigate corneal tissue differences between glaucoma, suspect glaucoma, and control participants. The parameters of OCT speckle can be indirectly linked to the microstructural properties of the imaged tissue.[16] Consequently, they showed that the corneal microstructure of suspect glaucoma participants is more similar to that of glaucoma than that of control participants.[16]

Even though glaucoma primarily affects the optic nerve, as it is a multifactorial disease challenging to detect early, there is an increasing interest in investigating the cornea as a target tissue for glaucoma risk assessment and glaucoma management. In the current work, no statistically significant differences were found in any of the biomechanical parameters analyzed between suspect glaucomatous and control eyes (Table 1). Previous works on corneal biomechanics and glaucoma found a weak relationship between corneal biomechanical parameters and measures of structural and functional damage in glaucoma.[6,7] As these correlations were weak and only found in developed cases, corneal biomechanics does not seem to be an adequate biomarker for early glaucoma detection. Consequently, biomechanical parameters were not used to feed the supervised machine learning algorithms in the current work.

Furthermore, recent works suggested that corneal densitometry analysis could have a clinical impact on the diagnosis and management of glaucoma.[7-9] Inspired by those previous works, in the current study we investigated corneal densitometry as an early glaucoma biomarker. However, we did not use traditional densitometry values estimated from

Table 2 The goodness of fit in terms of RMSE of the different candidate models to fit corneal pixel intensity distribution, expressed in arbitrary units, for 10 suspect glaucomatous eyes and 14 control eyes.

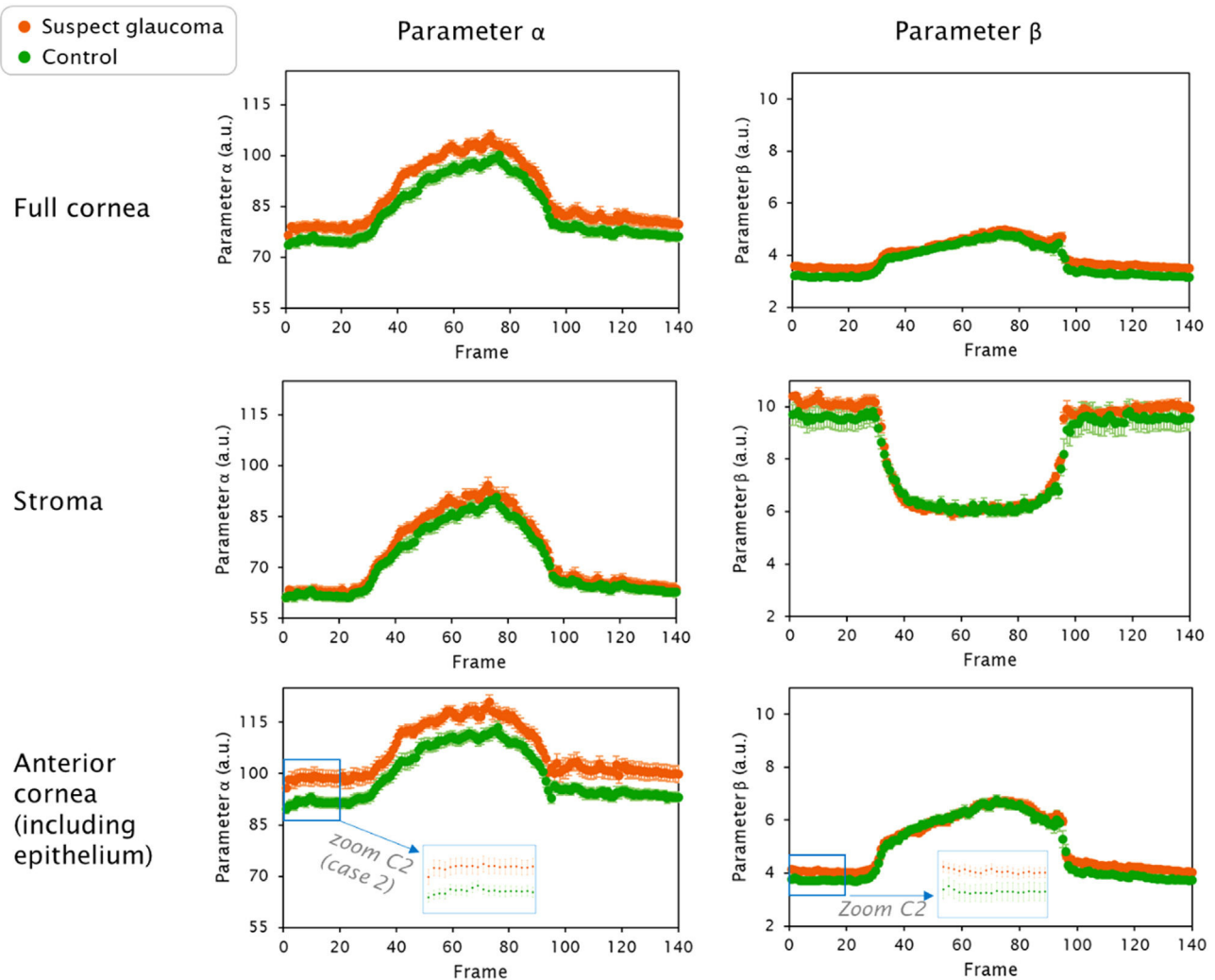|		Model		
		Weibull	Gamma	Lognormal
Full cornea	Control	0.032 ± 0.001	0.032 ± 0.001	0.033 ± 0.001
	Suspect glaucoma	0.027 ± 0.001	0.027 ± 0.001	0.027 ± 0.001
Stroma	Control	0.080 ± 0.002	0.081 ± 0.002	0.081 ± 0.002
	Suspect glaucoma	0.075 ± 0.003	0.075 ± 0.004	0.075 ± 0.004
Anterior cornea (including epithelium)	Control	0.027 ± 0.001	0.028 ± 0.001	0.031 ± 0.004
	Suspect glaucoma	0.021 ± 0.001	0.022 ± 0.001	0.022 ± 0.002

Fig. 3 Mean values of the dynamic evolution of $\alpha(t)$ and $\beta(t)$ parameters, expressed in arbitrary units, extracted from the fitting of Weibull distribution to the pixel intensities of corneal Scheimpflug images for 10 suspect glaucoma participants (orange) and 14 control participants (green). The analysis was done for different vertical (axial) dimensions of the ROI, considering: full corneal thickness (above), stroma excluding the epithelium (middle), and anterior cornea including the epithelium (bottom). Zoom-in is shown in those frames where group differences in $\alpha(t)$ are the largest. Corresponding frames in β (t) are also shown. Corresponding individual data points to the zoomed area were used in case 2 (20 frames). Error bars: SE at 95% confidence level.

Pentacam HR software (Oculus, Wetzlar, Germany). Instead, we used Scheimpflug images acquired with Corvis ST. This choice was motivated by the fact that Corvis ST, in opposition to Pentacam HR, allows exporting raw images. In addition, Corvis ST, again contrarily to Pentacam HR, captures Scheimpflug images during the dynamic evolution of the cornea under mechanical stimulation, which allows a more complete analysis of corneal tissue. However, Corvis ST software does not include a corneal densitometry module. Consequently, this study applied an alternative method to traditional densitometry, already validated on corneal Scheimpflug images under mechanical stimulation.[20]

The method used in the current work to estimate corneal densitometry is based on statistical Weibull modeling of the pixel intensity distribution. Besides giving information on corneal tissue clarity (α parameter), as traditional densitometry does, it offers additional information on corneal tissue homogeneity (β parameter). In previous works

investigating corneal densitometry by means of statistical modeling of the pixel intensity distribution, the epithelium was removed,[12-14,20] as it is known to carry different statistical information compared to the stroma[11] and was therefore considered a potential source of noise. Corneal densitometry is based on light backscatter. As the corneal epithelium and corneal stroma have a different composition, it is expected that light will travel, and consequently scatter, differently in one tissue or the other. Actually, from any standard corneal Scheimpflug image, it is possible to observe bare eye a thin bright layer on the anterior cornea corresponding to the epithelium, while the stroma shows a darker and more uniform appearance. These inner tissue differences justify that when only the stroma was considered, smaller values of $\alpha(t)$ and larger values of $\beta(t)$ were obtained in the current study, in comparison to the case when the anterior cornea, including the epithelium, was analyzed, as shown in Fig. 3.
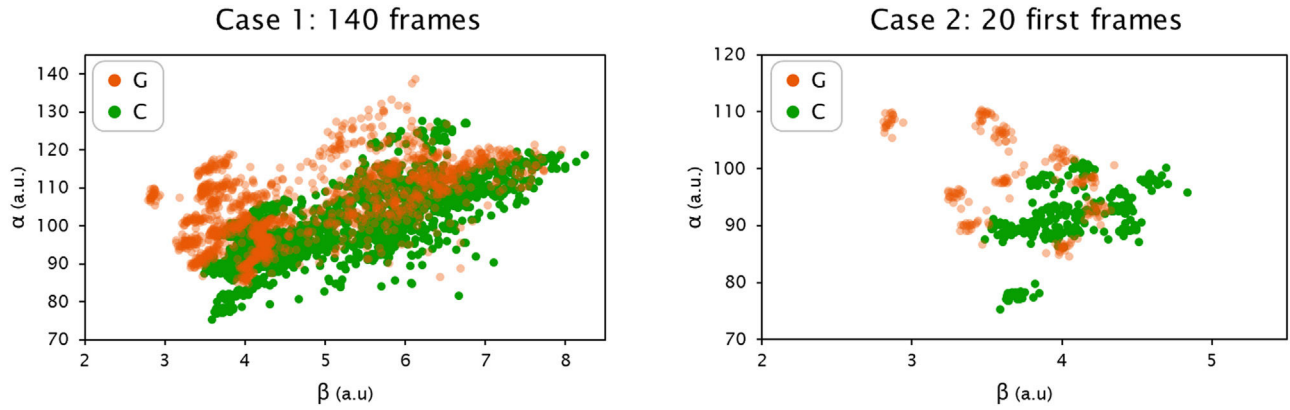
Fig. 4 The complete dataset that was used to feed the different supervised machine learning algorithms in each case under investigation. Data points in case 1 correspond to the $\alpha(t)$ and $\beta(t)$ parameters, expressed in arbitrary units, estimated from the 140 frames available for each subject (3275 data points in total, after removing missing values). Data points in case 2 correspond to the $\alpha(t)$ and $\beta(t)$ parameters, expressed in arbitrary units, estimated from the first 20 frames collected for each subject (480 data points). For both cases: 10 suspect glaucomatous eyes (G, in orange at 60% transparency) and 14 control eyes (C, in green).

In the current study, three different corneal depths were investigated. Various reports have shown that certain glaucoma medications cause alterations in the corneal eye surface.[21,22] However, glaucoma is a pathology that does not affect the cornea directly. Consequently, it is not straightforward to predict which corneal depth would be optimal to discriminate the disease. In the current work, the largest separation between suspect glaucomatous and control eyes in the $\alpha(t)$ parameter was found when considering the anterior cornea, including the epithelium. This result shows that even though pixels corresponding to the epithelium carry different statistical information compared to those from the stroma,[11] this information might be of value in cases where the discrimination between groups is especially challenging.

In the current research, the group mean parameters $\alpha(t)$ and $\beta(t)$ were good discriminators between suspect glaucoma and control eyes (Fig. 3), with statistically significant differences for both $\alpha(t)$ (paired t-test, $P < 0.05/N$ [N = 140, Bonferroni]) and $\beta(t)$ (paired t-test, $P < 0.05/N$ [N=140, Bonferroni]). However, when looking at individual values of $\alpha(t)$ and $\beta(t)$ parameters, discriminating suspect glaucomatous eyes from control eyes became challenging, as illustrated in Fig. 4. However, the combination of these two parameters, $\alpha(t)$ and $\beta(t)$, resulted in a successful way
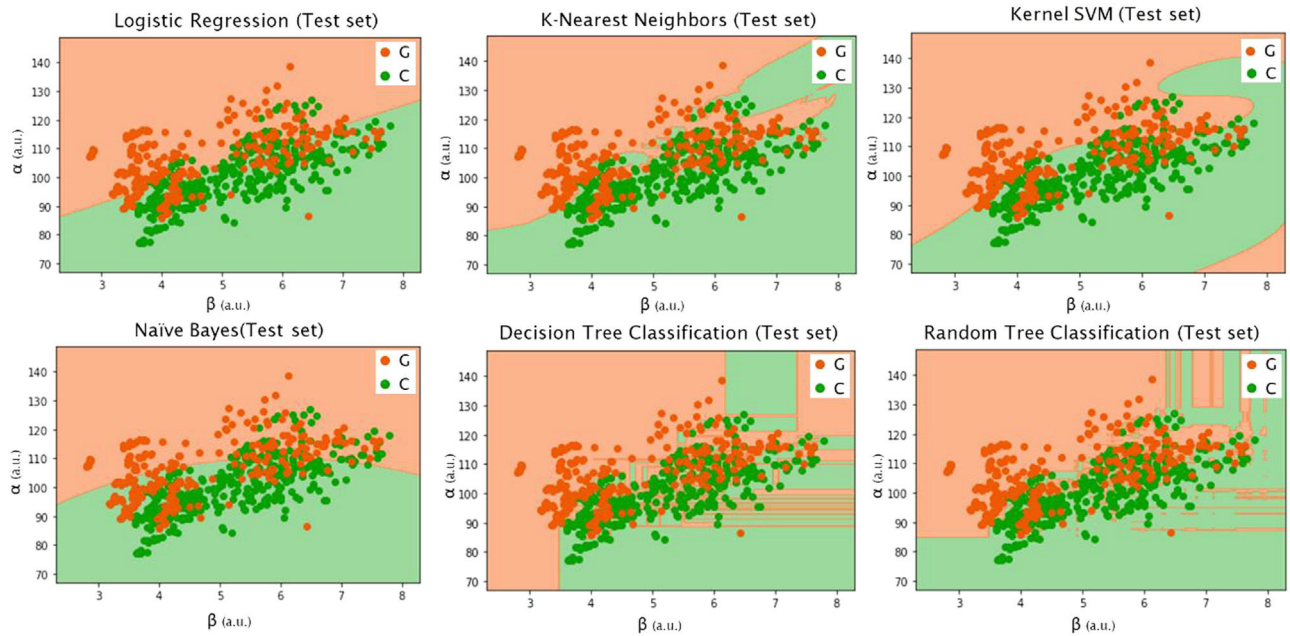

Fig. 5 Graphical representation of the performance of the different supervised machine learning algorithms in case 1 (140 frames). Background colors represent the output of the training step, while the 655 data points correspond to the test set (i.e., 20% of the total dataset that was not used for training). To clarify, suspect glaucomatous eyes (G, in orange) should fall in the orange background while control eyes (C, in green) should fall in the green background, otherwise, there is a misclassification. Corresponding statistics are shown in Table 3.

Fig. 6 Graphical representation of the performance of the different supervised machine learning algorithms in case 2 (20 first frames). Background colors represent the output of the training step, while the 96 data points correspond to the test set (i.e., 20% of the total dataset that was not used for training). To clarify, suspect glaucomatous eyes (G, in orange) should fall in the orange background while control eyes (C, in green) should fall in the green background, otherwise, there is a misclassification. Corresponding statistics are shown in Table 3.
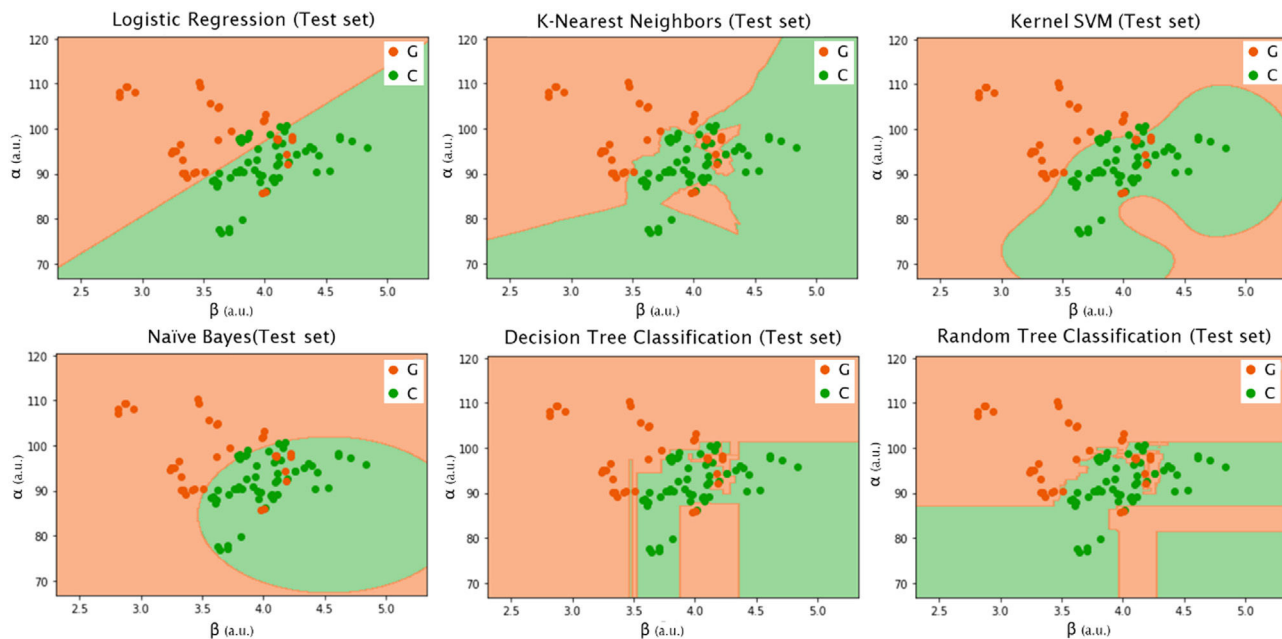
to discriminate between suspect glaucomatous and control eyes (Figs. 5, 6, and Table 3) using supervised machine learning algorithms.

In the machine learning part of the study, two different situations were investigated, case 1 (140 frames, 3360 instances) and case 2 (20 frames, 480 instances). The algorithms that showed the best performance were K-nearest Neighbors (case 1) and Decision Tree Classification (case 2). However, as indicated by Table 4, overfitting occurs in the Decision Tree Classification algorithm and consequently should not be considered a good classifier. Overfitting happens when there is a significant difference in accuracy between training and test sets. Overfitting is undesirable as it implies a lack of generalization beyond the training set. Overfitting is one of the main limitations of Decision Tree Classification and Random Forest Classification algorithms.[17] Regularization techniques can be applied to solve or minimize overfitting as a part of pre-processing. However, this goes beyond the scope of the current study. Consequently, considering the performance on the training set for cases 1 & 2, and the non-overfitting algorithms, K-NN is the classifier that reached the best performance in the current work (83.93 % accuracy in case 1 and 92.7 % in case 2). K-NN is a supervised algorithm based on the principle that instances of a given dataset generally reside near other instances with similar characteristics.[17] The advantages of K-NN include simplicity, speed, and easiness of implementation. The disadvantages of K-NN include not working well with too large datasets or with high dimensions.[18] However, these general disadvantages do not apply to the current work, since the dataset used in the training phase was not too large and we were working in two dimensions (only two parameters, α and β, were used to feed the algorithm). Consequently, as

the inherent disadvantages of the algorithm were minimized, especially regarding dimensionality, we speculate this might be the reason the K-NN algorithm scored the highest (83.93 % accuracy in case 1 and 92.7 % in case 2).

The main limitation of the current work is the limited sample size. Twenty-four eyes are enough to investigate group means of α(t) and β(t) parameters and traditional statistics can be applied to successfully discriminate between groups, as it was shown in the current work and previous literature.[20] However, by definition, machine learning techniques need to learn from a large dataset to be later efficient and achieve high performance. Consequently, 24 eyes is not a sample size large enough to directly apply supervised machine learning methods. To overcome this limitation, instead of using a single α and β per eye (24 instances), we used the complete data available, this is, calculating α and β per eye and frame. Two cases were investigated, case 1 (140 frames, 3360 instances) and case 2 (20 frames, 480 instances). When only considering the 20 first frames, case 2, the highest performance was obtained (92.7 % accuracy), as indicated by Table 3. However, as this data is clustered, as shown in the right part of Fig. 4, this result might be subjected to overfitting. However, in case 1, the performance is still high (83.93 %) and would not be so strongly affected by overfitting. It is also worth mentioning that data quality is often more critical than data quantity to implement a supervised machine learning algorithm successfully.[23] Dirty data, i.e., inaccurate, incomplete, or inconsistent data, often leads to misclassifications.[23] In the current work, we fed the machine learning algorithms with only two parameters (α and β) that proved to be independent and statistically significant different between control and suspect glaucoma eyes according to traditional statistics. Using clean and

Table 3 Performance of the trained supervised algorithms on the test set (20% of the whole dataset, i.e., 655 data points in case 1 (140 frames) and 96 data points in case 2 (20 first frames)). Performance was assessed by the number of true negatives (TN), true positives (TP), false negatives (FN), false positives (FP), accuracy, area under the curve (AUC), and 95% confidence interval (CI). The last column (#) orders the algorithms from best performance (1) to worse performance (6).

Case	Algorithm	TN	TP	FN	FP	TN (%)	TP (%)	FN (%)	FP (%)	Accuracy (%)	AUC	CI	#
140 frames	Logistic Regression	316	164	76	99	48	25	12	15	73.28	0.71	[0.70,0.77]	5
	K-nearest Neighbors (K=8)	348	198	44	65	53	30	7	10	83.34	0.82	[0.81,0.86]	1
	Kernel Support Vector Machine	340	166	52	97	52	25	8	15	77.25	0.75	[0.74,0.81]	4
	Naïve Bayes	326	126	66	137	50	19	10	21	69.00	0.67	[0.67,0.73]	6
	Decision Tree Classification	327	188	65	75	50	29	10	11	78.63	0.77	[0.76,0.82]	3
	Random Forest Classification	340	179	52	84	52	27	8	13	79.24	0.77	[0.76,0.82]	2
20 first frames	Logistic Regression	49	28	10	9	51	29	10	9	80.2	0.79	[0.72,0.88]	5
	K-nearest Neighbors (K=5)	53	36	6	1	55	38	6	1	92.7	0.94	[0.88,0.98]	2
	Kernel Support Vector Machine	49	27	10	10	51	28	10	10	79.2	0.78	[0.71,0.87]	6
	Naïve Bayes	54	28	5	9	56	29	5	9	85.4	0.84	[0.78,0.93]	4
	Decision Tree Classification	54	36	5	1	56	38	5	1	93.8	0.94	[0.89,0.99]	1
	Random Forest Classification	53	36	6	1	55	38	6	1	92.7	0.94	[0.88,0.98]	2

Table 4 Performance of the trained supervised algorithms on the training set (80% of the whole dataset, i.e., 2620 data points in case 1 (140 frames) and 383 data points in case 2 (20 first frames)). Performance was assessed by the number of true negatives (TN), true positives (TP), false negatives (FN), false positives (FP), accuracy, and area under the curve (AUC). Corresponding statistics on the test set are shown in Table 3.

Case	Algorithm	TN	TP	FN	FP	TN (%)	TP (%)	FN (%)	FP (%)	Accuracy (%)	AUC
140 frames	Logistic Regression	1521	1099	0	0	58	42	0	0	100	100
	K-nearest Neighbors (K=8)	1416	783	105	316	54	30	4	12	83.93	0.82
	Kernel Support Vector Machine	1342	625	179	474	51	24	7	18	75.08	0.73
	Naïve Bayes	1288	505	233	594	49	19	9	23	68.43	0.65
	Decision Tree Classification	1521	1099	0	0	58	42	0	0	100	100
	Random Forest Classification	1521	1099	0	0	58	42	0	0	100	100
20 first frames	Logistic Regression	221	162	0	0	58	42	0	0	100	100
	K-nearest Neighbors (K=5)	212	150	6	2	57	41	2	1	94.52	0.94
	Kernel Support Vector Machine	205	108	16	54	54	28	4	14	81.72	0.78
	Naïve Bayes	214	110	7	52	56	29	2	14	84.60	0.82
	Decision Tree Classification	221	162	0	0	58	42	0	0	100	100
	Random Forest Classification	221	162	0	0	58	42	0	0	100	100

meaningful data, as it was done in the current work, enhances the performance and robustness of the training algorithms.[23] Nevertheless, future works should repeat the investigation here presented with a larger sample size.

The current study showed that it is possible to extract valuable information from the anterior cornea, including the epithelium, to discriminate suspect glaucomatous from control eyes with 83.93% accuracy. Even though more work needs to be done, with a larger sample size, the presented results shed some light on the investigation of corneal tissue as a clinical tool for early glaucoma detection and glaucoma management. In conclusion, corneal densitometry on its own has the potential to serve as a biomarker for early glaucoma risk assessment.

Conflicts of interest

The authors report no conflicts of interest and have no proprietary interest in any of the materials mentioned in this article.

Acknowledgements

References

1. Giangiacomo A, Coleman AL. *The Epidemiology of Glaucoma. Glaucoma.* Berlin, Heidelberg: Springer; 2009:13−21.
2. McMonnies CW. Glaucoma history and risk factors. *J Optom.* 2017;10(2):71−78.
3. Cánovas-Serrano Y, Vallés-San-Leandro L, Rodríguez-Izquierdo MÁ, López-Serrano R, Lajara-Blesa J. On the protective role of the blood vessels in glaucomatous damage: a transversal study. *J Optom.* 2021.
4. Chauhan BC, Hutchison DM, LeBlanc RP, Artes PH, Nicolela MT. Central corneal thickness and progression of the visual field and optic disc in glaucoma. *Br J Ophthalmol.* 2005;89 (8):1008−1012.
5. Herndon LW, Weizer JS, Stinnett SS. Central corneal thickness as a risk factor for advanced glaucoma damage. *Arch Ophthal.* 2004;122(1):17−21.
6. Mansouri K, Leite MT, Weinreb RN, Tafreshi A, Zangwill LM, Medeiros FA. Association between corneal biomechanical properties and glaucoma severity. *Am J Ophthalmol.* 2012;153 (3):419−427.
7. Morales-Fernández L, Benito-Pascual B, Pérez-García P, Perucho-González L, Sáenz-Francés F, Santos-Bueso E, García-Bella J, Sánchez-Jean R, García-Feijoo J, Martínez-de-la-Casa JM.

8. Corneal densitometry and biomechanical properties in patients with primary congenital glaucoma. *Can J Ophthalmol.* 2021;56 (6):364−370.
8. Morales-Fernández L, Perucho-González L, Martinez-de-la-Casa JM, Perez-Garcia P, Sáenz-Francés F, Benito-Pascual B, Nieves-Moreno M, García-Bella J, Arriola-Villalobos P, García-Feijoo J. Corneal densitometry and topography in patients with primary congenital glaucoma. *J Fr Ophtalmol.* 2020;43(8):697−703.
9. Molero-Senosiain M, Morales-Fernandez L, Saenz-Frances F, Perucho-Gonzalez L, García-Bella J, Garcia Feijoo J, Martinez-de-la-Casa JM. Corneal properties in primary open-angle glaucoma assessed through scheimpflug corneal topography and densitometry. *J Glaucoma.* 2021;30(5):444−450.
10. Otri AM, Fares U, Al-Aqaba MA, Dua HS. Corneal densitometry as an indicator of corneal health. *Ophthalmology.* 2012 Mar;119 (3):501−508.
11. Consejo A, Jiménez-García M, Rozema JJ. Age-related corneal transparency changes evaluated with an alternative method to corneal densitometry. *Cornea.* 2021;40(2):215−222.
12. Consejo A, Solarski J, Karnowski K, Rozema JJ, Wojtkowski M, Iskander DR. Keratoconus detection based on a single Scheimpflug image. *Transl Vis Sci Technol.* 2020;9(7):36.
13. Consejo A, Jiménez-García M, Issarti I, Rozema JJ. Detection of subclinical keratoconus with a validated alternative method to corneal densitometry. *Transl Vis Sci Technol.* 2021;10(9):32.
14. Consejo A, Alonso-Caneiro D, Wojtkowski M, Vincent SJ. Corneal tissue properties following scleral lens wear using Scheimpflug imaging. *Ophthalmic* Physiol Opt. 2020;40(5):595-606.
15. Tack M, Kreps EO, De Zaeytijd J, Consejo A. Scheimpflug-based analysis of the reflectivity of the cornea in Marfan syndrome. *Transl Vis Sci Technol.* 2021;10(9):32.
16. Iskander DR, Kostyszak MA, DA Jesus, Majewska M, Danielewska ME, Krzyzanowska-Berkowska P. Assessing corneal speckle in optical coherence tomography: a new look at glaucomatous eyes. *Optom Vis Sci.* 2020;97(2):62−67.
17. Consejo A, Melcer T, Rozema JJ. Introduction to machine learning for ophthalmologists. *Semin Ophthalmol.* 2019;34(1):19−41.
18. Vercio LL, Amador K, Bannister JJ, Crites S, Gutierrez A, MacDonald ME, Moore J, Mouches P, Rajashekar D, Schimert S, Subbanna N. Supervised machine learning tools: a tutorial for clinicians. *J Neural Eng.* 2020;17(6): 062001. 19.
19. Issarti I, Rozema JJ. Basics of artificial intelligence for ophthalmologists. *Artificial Intell Ophthalmol.* 2021:17−30. Springer, Cham.
20. Consejo A, Gławdecka K, Karnowski K, Solarski J, Rozema JJ, Wojtkowski M, Iskander DR. Corneal properties of keratoconus based on Scheimpflug light intensity distribution. *Invest Ophthalmol Vis Sci.* 2019;60(8):3197−3203.
21. Mastropasqua R, Agnifili L, Fasanella V, Lappa A, Brescia L, Lanzini M, Oddone F, Perri P, Mastropasqua L. In vivo distribution of corneal epithelial dendritic cells in patients with glaucoma. *Invest Ophthalmol Vis Sci.* 2016;57(14):5996−6002.
22. Noecker RJ, Herrygers LA, Anwaruddin R. Corneal and conjunctival changes caused by commonly used glaucoma medications. *Cornea.* 2004;23(5):490−496.
23. Qi Z, Wang H, Li J, Gao H. Impacts of dirty data: and experimental evaluation. arXiv preprint arXiv:1803.06071. 2018 Mar 16.