*Retrospective Clinical Research Report*

*Journal of* INTERNATIONAL MEDICAL RESEARCH

Journal of International Medical Research
50(6) 1–15

DOI: 10.1177/03000605221108100
journals.sagepub.com/home/imr
SAGE

# Correlation between corneal dynamic responses and keratoconus topographic parameters

**Hsi-Yun Tai[1], Jun-Ji Lin[1], Yi-Hung Huang[1], Po-Jen Shih[2], I-Jong Wang[3] and Jia-Yush Yen[1]**

## Abstract

**Objective:** To investigate the correlation between corneal biomechanical properties and topographic parameters using machine learning networks for automatic severity diagnosis and reference benchmark construction.

**Methods:** This was a retrospective study involving 31 eyes from 31 patients with keratonus. Two clustering approaches were used (i.e., shape-based and feature-based). The shape-based method used a keratoconus benchmark validated for indicating the severity of keratoconus. The feature-based method extracted imperative features for clustering analysis.

**Results:** There were strong correlations between the symmetric modes and the keratoconus severity and between the asymmetric modes and the location of the weak centroid. The Pearson product-moment correlation coefficient (PPMC) between the symmetric mode and normality was 0.92 and between the asymmetric mode and the weak centroid value was 0.75.

**Conclusion:** This study confirmed that there is a relationship between the keratoconus signs obtained from topography and the corneal dynamic behaviour captured by the Corvis ST device. Further studies are required to gather more patient data to establish a more extensive database for validation.

## Keywords

Corneal dynamic characteristics, Corvis ST, keratoconus, machine learning

Date received: 1 September 2021; accepted: 31 May 2022

[1]Department of Mechanical Engineering, National Taiwan University, Taipei
[2]Department of Biomedical Engineering, National Taiwan University, Taipei
[3]Department of Ophthalmology, National Taiwan University Hospital, Taipei

**Corresponding author:**
Jia-Yush Yen, Department of Mechanical Engineering, National Taiwan University of Science and Technology, No. 43, Keelung Rd., Sec. 4, Da'an Dist., Taipei, 106335.
Email: jyen@mail.ntust.edu.tw

## Introduction

Keratoconus is a corneal ectasia in which the cornea becomes conical due to progressive thinning and there is a gradual corneal protrusion.[1] Analysis of the corneal topography is the most widely used method to identify keratoconus.[2] However, this method relies on the subjective analysis of topographic maps, which makes detecting early-stage keratoconus, or keratoconus without obvious symptoms, clinically difficult and different observers may introduce bias into the diagnoses.[2–6] Detecting biomechanical instability, which is an early indicator of the disease, is also problematic.[7] In the early stages of keratoconus, subtle changes in the corneal microscopic structure may already be evident, resulting in abnormal mechanical stability that may occur even before notable changes in corneal morphology are detected.[7] Therefore, a combination of clinical devices to measure biomechanical properties and an objective diagnostic approach is required.

The Corvis ST device is widely used in the clinical evaluation of corneal biomechanics.[8–11] It is a non-contact tonometer that uses an ultrahigh-speed Scheimpflug camera to monitor corneal behaviour during an air-puff test. This allows visualization of a large set of biomechanical deformation response parameters. However, the Corvis ST does not currently provide an automatic analysis of the corneal biomechanics. Two indices have been proposed for this purpose: the Corvis Biomechanical Index (CBI)[12] and the Tomographic Biomechanical Index (TBI).[4] The CBI is based on a logistic regression analysis that uses Corvis ST response parameters to distinguish keratoconic from normal eyes. The TBI combines corneal tomography (using a Pentacam system) and biomechanical parameters from the Corvis ST device to assist in keratoconus detection.

Other methods based on Corvis-ST outputs for discriminating keratoconic from normal eyes include an assessment of two stiffness parameters based on the deformation profiles generated by Corvis ST;[13] a keratoconic cornea is softer than a normal cornea, so corneal softness may be useful as an identifier of keratoconus. Indeed, the Corvis ST parameters of Max Inverse Radius, deformation amplitude (DA) Ratio, Pachy Slope, biomechanical corrected IOP (bIOP) and stiffness have achieved high accuracy in detecting keratoconus.[14,15] Studies have shown that these parameters exhibit excellent repeatability (interclass correlation coefficient $\geq 0.90$) and discriminative ability in diagnosing keratoconic eyes.[16,17]

To more fully utilize objective measurements from Corvis ST, a previous study examined the ability of modal analysis to reflect keratoconus severity.[18] However, the attempt to extract a correlation between the Legendre modal decomposition and the keratoconus severity did not yield a satisfactory result because the association was too complex. To resolve this difficulty, we propose utilizing machine learning to extract the correlation using a shape-based method and a feature-based method. Both of these methods use the dynamic corneal deformation response for their input; the shape-based method establishes benchmarks relating the corneal dynamic responses to the keratoconus topographic patterns; the feature-based clustering system divides the patient data into various keratoconus severities for diagnostic reference. We suggest that these approaches enable a superior automatic unbiased interpretation of the Corvis ST data.

To the best of our knowledge, no previous keratoconus detection method has considered dynamic response as a time sequence and provided means to extract

the corneal biomechanical properties. Therefore, the purpose of this present study was to combine Corvis ST data with machine learning networks for automatic severity diagnosis and reference benchmark construction.

## Methods

### Patients

This retrospective study was performed from February 2013 to February 2014 and included data from 31 eyes of 31 patients with keratonus eyes who attended National Taiwan University Hospital during the study period. For all subjects, topographic and biomechanical data were measured using the Tomey TMS-4 Corneal Topographer System and Corvis ST system, respectively. The severity of keratoconus was quantified using the Keratoconus Index (KCI) and Keratoconus Severity Index (KSI) from the topography report (data not shown).[19,20]

The reporting of this study conforms to STROBE guidelines.[21] The study was approved by the Research Ethics Committee C National Taiwan University Hospital, 7, Chung-Shan South Road, Taipei, Taiwan 100, R.O.C (NTUH-REC No.: 201607049RINA), and all participants provided written informed consent.

### Data pre-processing: Legendre decomposition

Image processing was used to determine the anterior corneal contour (white lines in Figure 1) from the Corvis ST images, which was represented as a function $F(x)$, where $x$ was measured from the centre of the cornea across the anatomic sagittal plane. Each eye had 140 contours recorded. The various $F_i(x)$ functions for the $i$ acquired images were then used for corneal dynamic analysis.

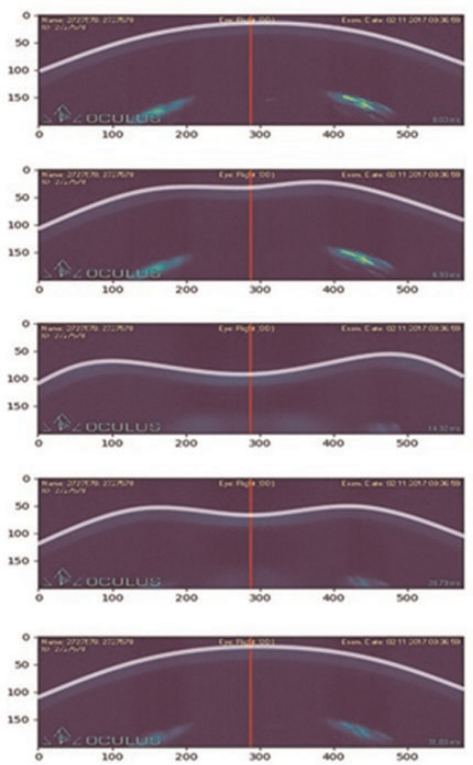

**Figure 1.** Sequential images of corneal deformation induced by an air puff were obtained from the Corvis ST device. The images show the deformation of the sagittal corneal plane subject to an air puff. The white lines show the contours extracted by image processing. (Unit: $10^{-2}$ mm, the direction of the air puff: +Y-axis; X-axis: the direction perpendicular to the air puff and the normal of the eye, Y-axis: the direction parallel to the air puff and the normal of the eye).

The contours described by the $F_n(x)$ functions were expanded using Legendre polynomials into Legendre modal shapes with corresponding modal parameters $A_{0\ell}$:

$$F_n(x) = \sum_{\ell=0}^{\infty} A_{0\ell} P_\ell^0 (\cos\theta) \qquad (1)$$

Integrating both sides with some suitable arrangement yielded

$$\int_0^{\pi} F_n(x)P_l^0(\cos\theta)\sin\theta \mathrm{d}\theta = \int_0^{\pi} \sum_{\ell=0}^{\infty} A_{0\ell}P_{\ell}^0(\cos\theta)P_{l'}^0(\cos\theta)\sin\theta \mathrm{d}\theta \quad (2)$$

$$= A_{0l}\frac{2}{2l+1} \quad (3)$$

The modal parameter, $A_{0\ell}$, in (2) varied with time, with each value for the $l$th mode of the Legendre polynomials described by

$$A_{0\ell} = \frac{2\ell+1}{2}\int_0^{\pi} F_n(x)P_{\ell}^0(\cos\theta)\sin\theta \mathrm{d}\theta, \quad l \in N \quad (4)$$

The modal shapes of the six primary modes were characterized by $P_{\ell}^0(\cos\theta)$, with $l = 0$–5.

Mode 0 [i.e., $P_0(\cos\theta)$] represented corneal vibration in the radial direction, with a positive value indicating contracting inward and a negative value indicating expanding outward. Mode 1 [i.e., $P_1(\cos\theta)$] represented corneal lateral movement, with positive and negative values representing movement toward the left and right, respectively. Modes 2, 3, 4, and 5 formed two, three, four, and five nodal points on the cornea, respectively.

Because the variation in corneal vibrations resulted in differences between modal parameters $A_{0l}$, the corneal vibrations could be evaluated by comparing these variations.

## Machine Learning

### *Shape-based clustering method*

The corneal response was too complicated for direct shape comparisons, and so this study compared the 'decomposed' modes to determine if a meaningful benchmark for diagnosing keratoconus could be extracted. The variation of the $A_{0l}$ modal parameters extracted from waveform function $F(x)$ formed the image sequence obtained using the Corvis ST device. The shape-based method then decomposed each waveform into Legendre modes, recorded changes in the model parameters, and then composed time sequence data for analysis. The modal parameter sequences can be classified using $k$-means clustering, as shown in Figure. 2.

The objective function for clustering minimizes the sum of squared errors between a cluster centre and its neighbours. The amplitude and offset of the variation waveform are meaningful in diagnosing keratoconus, and so, unlike most other time series, no pre-scaling or *a priori* transition invariance analysis was performed.[22] As represented in (5), the distance between each member of the cluster and the centroid was measured as the Euclidean distance. The $k$-means clustering algorithm converged, and stopped when the sum of squared errors, as represented in (6), was less than the threshold value of $10^{-6}$:

$$E = \sum_{k=1}^{K}\sum_{x\in C_k} d(x, m_k)^2 \quad (5)$$

$$d(x, m_k) = ||x, m_k|| = \sqrt{(x_1 - m_{k1})^2 + \cdots + (x_n - m_{kn})^2} \quad (6)$$

where $C_k$ was the $k$th cluster, $m_k$ was the cluster centroid $C_k$, and $d(x, m_k)$ was the distance between an instance $x$ and centroid $m_k$.

The number of clusters was a crucial parameter to be determined. We used the silhouette score to determine the number of clusters separately for each Legendre mode.

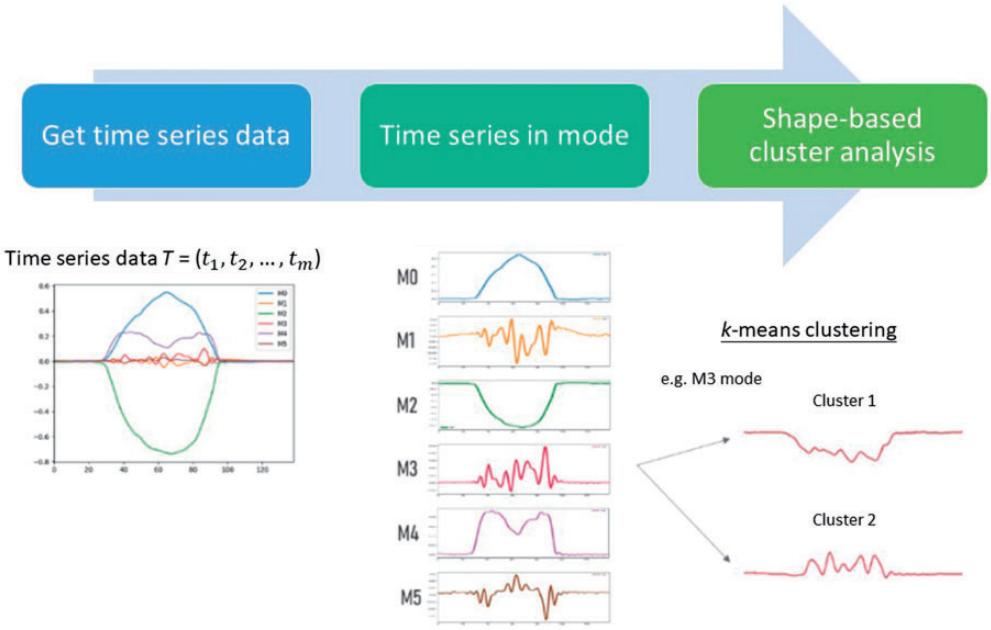


**Figure 2.** The Legendre modes and the workflow of the shape-based clustering method. (X-axis on time series data: time (Unit: s), Y-axis on time series data: amplitude of each mode).

## *Feature-based clustering method*

In contrast to the shape-based method, which compares waveforms in the time domain, the feature-based method clusters the data based on a set of extracted statistical features.[23,24] This approach reduces the dimensionality of the original time series without losing any essential information from the data. Instead of examining variations of the modal parameters, the feature-engineering process uses statistical analysis to extract a set of features with clustering analysis then applied to cluster the data into groups that share common characteristics.

*Feature engineering.* We used the tsfresh Python package[25] to extract meaningful features from the test data. The tsfresh package provides an automated process for feature extraction and feature selection, and results in a data frame comprising approximately 4000 extracted features. The resultant feature parameters include maximum (max), minimum (min), and median values, and also the number of peaks extracted from the time series.

A feature scaling process such as standardization and min-max scaling then transforms the features into standard ranges. This standardization process ensures that the data have zero mean and unit variance, while min-max scaling sets all values into the range [0,1]. The standardization and normalization processes transform the data, which initially have different scales, into comparable units:

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x} \tag{7}$$

where $\mu_x$ and $\sigma_x$ are the mean and standard deviation of a specific characteristic, respectively. Also,

$$x_{norm}^{(i)} = \frac{x^{(i)} - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

where $x_{min}$ and $x_{max}$ are the minimum and maximum values of the specific characteristic, respectively.

*Dimensionality reduction.* There are two main categories of dimensionality reduction: feature selection and feature extraction. The feature selection process directly selects a subset of relevant features from the original data set, while the feature extraction process projects the original data set onto a smaller feature space. Both methods remove redundant and irrelevant features without losing much information. Feature extraction was found to be necessary in this study for extracting better representations of complex data.

Feature extraction can be performed either linearly or nonlinearly. This study used both techniques to compare their effectiveness and determine the best process for later diagnosis. Linear techniques use a linear combination of the original variables to reduce the feature dimension.

A variable matrix in $p$-dimensional subspace can be extracted,[26] whose axes effectively represent the original data. For an observational matrix $X$ with $n$ samples and $p$ features, X can be approximated by calculating the product of matrices $H$ and $W$:

$$X_{n\times p} \approx H_{n\times k} W_{k\times p}, k < p \qquad (9)$$

$X$ is the observation matrix with rows representing the samples and columns representing various features; $H$ is the projected feature matrix (the new representation of the observation matrix) of the $k$ transformed principal components; $W$ is a linear transformation matrix containing the weightings of the $k$ principal components. The linear subspaces are inadequate for data sets containing nonlinear structures, and so a nonlinear dimensionality reduction method also needs to be considered. Famous nonlinear methods include locally linear embedding, Laplacian eigenmaps, $t$-distributed stochastic neighbour embedding, and isometric mapping (Isomap).[27,28]

This study compared the effectiveness of two linear methods (i.e., principal-components analysis [PCA] and non-negative matrix factorization [NMF])[29,30] and the Isomap nonlinear method.

(I) Linear dimension reduction

The PCA and NMF linear dimension reduction methods have different constraints imposed on the weightings ($W$) and the transformed features ($H$). PCA reduces the data dimension by retaining only principal components with the largest variances.[31] $W$ and $H$ matrices in PCA must be orthogonal but can have arbitrary signs. PCA requires the vector components to be orthogonal, while NMF factorizes non-negative data set $X$ into two non-negative matrices $W'$ and $H'$, which are then easier to interpret than are the matrices with arbitrary signs.[32]

(II) Nonlinear dimension reduction

Isomap is the nonlinear generalization of classical multidimensional scaling used to find a lower-dimensional embedding by preserving a pairwise distance matrix in the original space.[33] This approach captures the geodesic manifold distances between all pairs of data points: neighbouring points are approximated by finding their shortest paths. A sequence of short distances between neighbouring data points are then summed for faraway points.[31] This means that intrinsic nonlinear geometry is better represented than when using the Euclidean distance in linear dimensionality reduction.

*Cluster analysis.* Cluster analysis divides the observed data set into natural groupings of data sharing common characteristics. In the feature-based method, cluster analysis

with hierarchical clustering is performed using an agglomerative (bottom-up) algorithm that merges the most-similar objects by considering the cluster distance until all objects are within the cluster. The cluster distance is calculated using the Ward linkage method, and the variance within each cluster merged with the Euclidean distance metric is minimized.[34,35]

One advantage of hierarchical clustering is that it represents the similarity between data graphically in a dendrogram, which allows users to then determine the number of clusters ($k$) by cutting the dendrogram at a suitable level.

The Ward linkage for NMF is defined as

$$d(C_i, C_j) = \sum_{a \in C_i \cup C_j} \|a - \mu\| \tag{10}$$

This present study applied both the average silhouette score and the Calinski–Harabasz (CH) index to determine the optimal number of clusters.[36,37] The silhouette score measures how far the clusters are from each other. The silhouette coefficient of each instance is determined as

$$s(i) = \frac{b^{(i)} - a^{(i)}}{\max\{a^{(i)}, b^{(i)}\}}, \text{ if } |C_i| > 1$$

$$s(i) \in (-1,1) \tag{11}$$

where $a$ is the mean distance between instances within the same cluster and $b$ is the distance between an instance and the nearest foreign cluster.

The CH index is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion:

$$c(i) = \frac{tr(B_k)}{tr(W_k)} \frac{n-k}{k-1} \tag{12}$$

where $B_k$ and $W_k$ are the between- and within-cluster sums of squares for the $k$ clusters. The estimated number of clusters is the $k$ value that maximizes the CH index.

Three cluster analyses using different dimensionality reduction methods were performed using the clinical data points provided from the 31 eyes measured using the Corvis ST device.

## Results

This study collected data from the Corvis ST outputs and extracted anterior cornea contours for analysis from the 31 keratoconus eyes. Applying Legendre polynomial decomposition to the sequential contour data yielded the first six modes for the subsequent machine-learning analysis. These time-dependent modal parameters were used as input for clustering to investigate the relationship between the characteristics of keratoconus and their corresponding modal parameter profiles in the air-puff test.

### *Shape-based clustering*

The shape-based method directly clusters the data using the Legendre mode parameters. Modal parameter waveforms of the symmetric modes, denoted as M0, M2, and M4, were directly used as input for clustering without any pre-treatment because the amplitude was a crucial parameter for the symmetric modes. By contrast, the variation in the modal parameters was more meaningful than the amplitudes for asymmetric modes M1, M3, and M5. Thus, a *Time Series Scaler Mean Variance* was used to reduce the waveforms to signals with zero mean and unit variance.[38,39] The silhouette score for 2–14 clusters was calculated to determine the optimal number of clusters for each mode separately.

*Correlation between symmetric modes and severity.* The biomechanical interpretation of keratoconus is that it reflects reduced

corneal stiffness, a property that represents the ability of the cornea to withstand external forces. This mainly affects the symmetrical modes in the response, and so only these modes were considered in this section. As shown in Figure 3, 80% of the data were randomly selected to construct the benchmark. The M0 and M2 modes resulted in two groups: the severe group with higher amplitudes and the mild group with lower amplitudes. The M4 mode resulted in four groups: the severe groups included cases with high amplitudes as well as those with apparent depressions in the middle, while the other cases belonged to the mild groups. The above comparisons were made based on the mean of the KCI and KSI severity indices [i.e., $(KCI + KSI)/2$] from the topography report.

The validation process used the remaining 20% of the data with the corresponding silhouette score to determine how close the test data were to the benchmark groups. A similarity index was defined as

$$S_{mild} = \frac{b - a}{\max\{a, b\}} \tag{13}$$

where $a$ is the Euclidean distance between data of interest and the mild group, $b$ is the Euclidean distance between the data of interest and the severe group, and $S_{\text{mild}} > 0$ indicates that the data of interest exhibit characteristics of the mild group. Ten iterations in total were performed.

To examine whether the dynamic response could effectively reflect the severity of corneal ectasia, we assessed the Pearson product-moment correlation (PPMC) between the similarity index (with the mild group) and the reciprocal of the mean of the KCI and KSI severity indices [i.e., $2/(KCI + KPI)$], to reflect normality. Symmetric mode index, $I_{symmetric}$, was defined as below. Figure 4 shows that the symmetric modes were correlated strongly with the KCI and KSI severity indices from the topography report. The coefficients for the modal parameters in (14) were derived from many experimental trials to best represent the contributions of various modes.

$$I_{Symmetric} = 0.36 \times M_0 + 0.56 \times M_2 + 0.43 \times M_4 \tag{14}$$

*Correlation between asymmetric modes and weak region.* This study further examined if dynamic analysis can also provide geometric information on corneal topography. It is reasonable to assume that the weakest

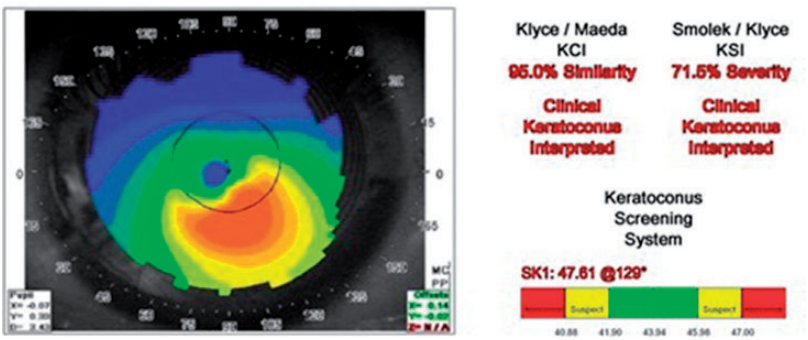


**Figure 3.** Topography report for a selected keratoconus case. KCI: Keratoconus Index, KSI: Keratoconus Severity Index.

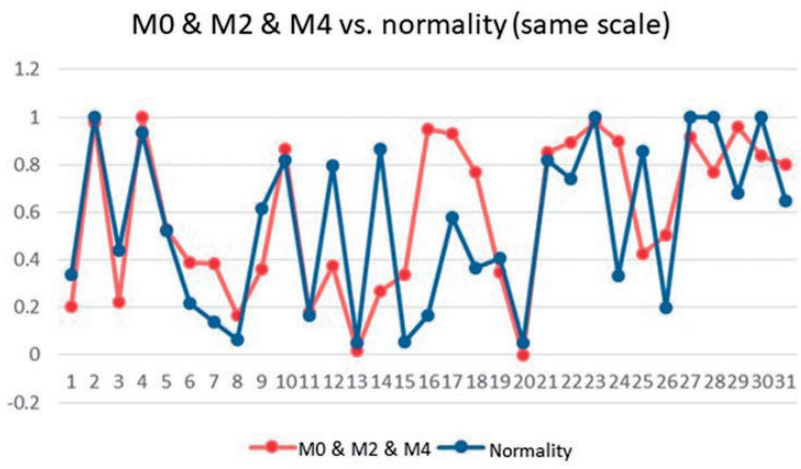


**Figure 4.** Correlation between the similarity index for the symmetric modes $I_{symmetric}$ and the reciprocal severity (both with min-max scaling; product-moment correlation coefficient [PPMC] coefficient = 0.62 [(high degree]). X-axis: Corvis ST data points, Y-axis: correlation coefficient.

region of the cornea will contribute the most to the topographic changes and will also affect the dynamics of the asymmetric modes. Establishing the correlation requires a geometric description of the weak region of the cornea. This study extracted the centroid of the location of the weak region from the topography map (i.e., red and orange parts of keratoconus from the topographic data in Figure 5). Because the cornea is circular, it was easy to describe a geometric location by the distance from the centroid to the centre of the cornea ($d$) and the angle of the centroid ($\phi$). The scalar centroid function was defined as

$$I_{centroid} = \mathrm{f}\ (d, \phi) = d \times \frac{1}{|\phi - 60^{\circ}|} \qquad (15)$$

where $\phi = \tan^{-1}\frac{y}{x}$. A weak region with a centroid closer to $\phi = 60^{\circ}$ and farther from the corneal center resulted in a larger value. Figure 5 shows the extraction of the weak region based on a topography report.

Again, 80% of Legendre data were randomly selected to construct benchmarks for comparison by using $k$-means clustering. The M1 and M3 modes were both clustered into two groups: one with fluctuations mostly on one side was defined as the 'deviation group' and the other with two-sided fluctuations was defined as the 'oscillation group'. For the M5 mode, the one group exhibiting only one-sided fluctuations was defined as the 'deviation group', and the other two were defined as the 'oscillation groups'.

The other 20% of data were used as the validation set to discern the similarity within constructed groups by calculating the total Euclidean distance between the data of interest and the benchmark constructed in the previous step. The silhouette score was derived to determine how similar the data of interest were to the deviation group:

$$S_{deviation} = \frac{d - c}{\max\{c, d\}} \qquad (16)$$

where $c$ is the Euclidean distance between data of interest and the deviation group, $d$ is the Euclidean distance between data of interest and the oscillation group,

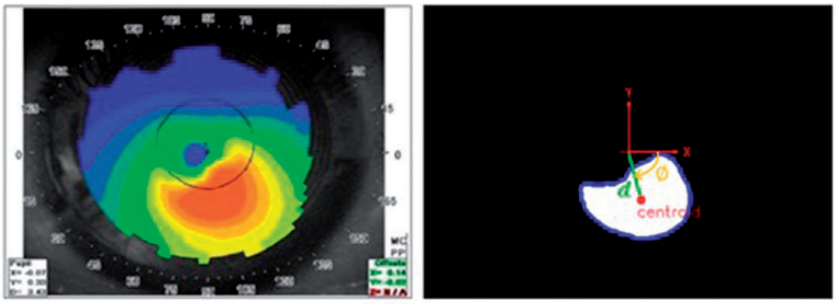

**Figure 5.** Geographical parameters extracted from a topography report to describe the weakest region.

and $S_{deviation} > 0$ indicates that the data of interest were highly similar to the deviation group. In total, 10 iterations were performed.

We first checked the PPMC between $S_{deviation}$ and $I_{centroid}$ for the M1, M3, and M5 modes, separately, and confirmed that they were strongly correlated. To enable a single index to be used, we proposed an asymmetric index $I_{asym}$. Again, the coefficients for the modal parameters in (17) resulted from many trial tests.

$$I_{asym} = 0.49 \times M_1 + 0.48 \times M_3 + 0.33 \times M_5 \quad (17)$$

The PPMC of $I_{asym}$ was strongly correlated with $I_{centroid}$ (Figure 6).

## Feature-based clustering

The feature-based method uses all of the Legendre mode shape data for feature extraction and selection. The feature-based clustering uses three dimensionality reduction methods for feature extraction and clustering: PCA, NMF, and Isomap. Using the Tsfresh Python package resulted in eight groups for both the symmetric and asymmetric mode data. The average modal parameter of each group was first derived and then used to calculate the Euclidean distance with the benchmark constructed through the shape-based clustering with 31 Corvis ST data points. A benchmark with a shorter Euclidean distance indicated higher similarity.

*Principal-components analysis (PCA) and non-negative matrix factorization (NMF).* PCA and NMF clustered the data into eight groups. The similarity with the mild group ($S_{mild}$) was derived using the symmetric modes, and that with the deviation group ($S_{deviation}$) was derived using asymmetric modes; zero was the reference point, where the mild and deviation groups were defined in (13) and (15), respectively.

The mean of the KCI and KSI severity indices and the weak centroid value were also obtained for each group based on the topographic data, with 0.5 as a reference point. There was a strong PPMC between the similarity to the mild groups and normality, with a coefficient of 0.94 (Figure 7). However, the similarity index for the deviation groups and the weak centroid value of keratoconus exhibited a weak correlation (PPMC coefficient = –0.26).

*Isomap.* The Isomap analysis also clustered the data into eight groups. Again, the similarity between the mild and deviation groups was examined for both the symmetric and asymmetric modes, with zero as the reference point. Mean KCI and KSI severity indices and the weak centroid value were

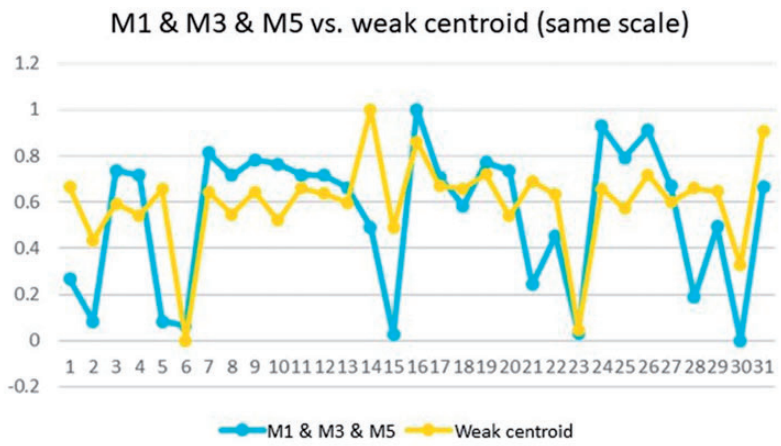


**Figure 6.** Correlation between the asymmetric modes and the weak centroid value [both with min–max scaling; product-moment correlation coefficient (PPMC) coefficient = 0.53 (high degree). X-axis: Corvis ST data points, Y-axis: correlation coefficient.

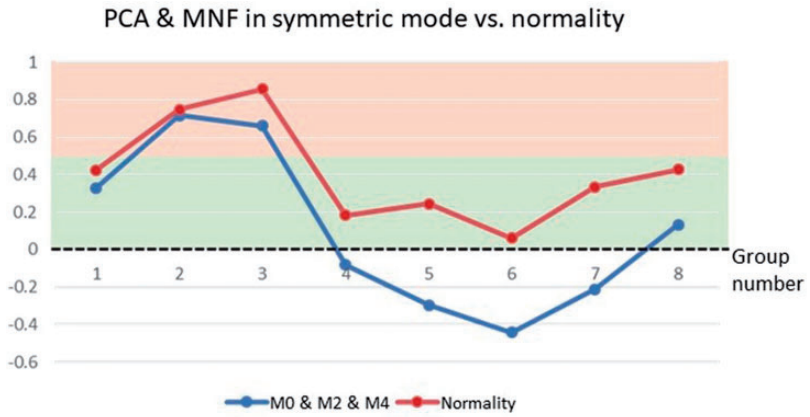


**Figure 7.** Correlation between symmetric modes and normality (product-moment correlation coefficient [PPMC] coefficient = 0.94 [high degree]). PCA, principal-components analysis; NMF, non-negative matrix factorization. Y-axis: correlation coefficient.

also obtained for each group based on the topographic data, with 0.5 as a reference point. Again, there was a strong PPMC between the mild group in the symmetric modes and normality (PPMC coefficient = 0.92). In addition, there was a strong correlation between the deviation groups in the asymmetric modes and the weak centroid value of keratoconus data (PPMC coefficient = 0.76) (Figure 8).

## Discussion

This study evaluated the correlation between the Corvis ST corneal dynamic response and the topographic patterns of keratoconus. The Legendre polynomial expansion was used to expand the response waveforms into different modes with corresponding modal parameters, which resulted in six primary modes exhibiting significant

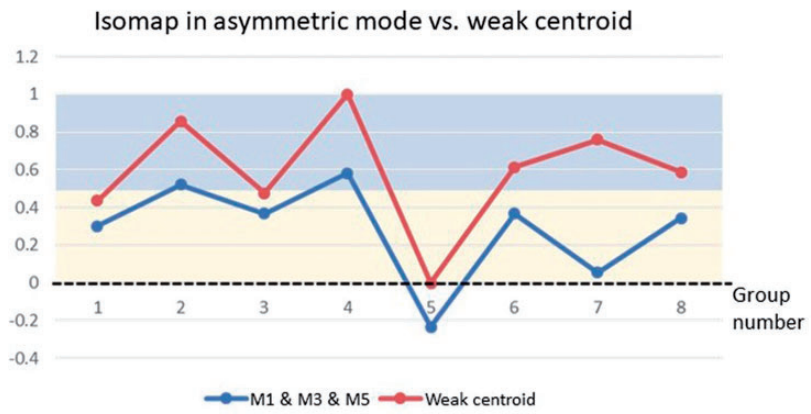


**Figure 8.** Correlation between asymmetric modes and weak centroid (product-moment correlation coefficient (PPMC) coefficient = 0.76 (high degree). Y-axis: correlation coefficient.

time-varying characteristics. Shape-based and feature-based machine-learning analyses were then applied to these modes to identify their relationships with the keratoconus features.

In the shape-based method, *k*-means clustering was used to construct a benchmark for relating the corneal dynamic responses to the keratoconus topographic patterns. The analysis showed a strong positive correlation between severity and the symmetric Legendre mode; a soft weak region, which is an indication of severe keratoconus, induced a large amplitude in symmetric modes due to low deformation resistance. The keratoconus features were also correlated with the asymmetric modes. When the weak region was close to the centre and far from 60°, the asymmetric modes of the cornea became more significant.

Based on the results of the feature-based method, it can be concluded that clustering with Isomap produced results that corresponded closely to the benchmarks constructed with *k*-means clustering. The results for groups 4, 5, and 6 in PCA and groups 3 and 6 in Isomap indicated severe groups that tended to exhibit large deformation amplitudes in symmetric modes. Both PCA and Isomap demonstrated a strong correlation between symmetric modes and severity of keratoconus.

Discriminating the keratoconus features from the response is difficult. The coefficient for the PPMC between the asymmetric mode and the weak centroid value reached only 0.76 in the Isomap analysis. It might be possible to improve this by modifying the definition of the weak region and the centroid function. Most current techniques rely on the subjective analysis of topographic maps, making the detection of early-stage keratoconus without obvious symptoms clinically difficult.

Our study had some limitations. Firstly, the dynamic simulation used in this study does not account for fluid-structure interaction and the internal structure of the eyeball because the objective was to investigate trends. Therefore, to obtain a more realistic simulation of non-contact tonometry, these factors need to be taken into consideration. Secondly, although the keratoconic benchmark achieves a high correlation with specific keratoconic characteristics the amount and the quality of the data may affect the clustering result. Therefore, more data are

required to establish more accurate model. Finally, after establishing a more accurate model, the system could be embedded into the Corvis ST program to aid in the diagnosis of keratoconus.

In conclusion, this study confirmed that there is a relationship between the keratoconus signs obtained from topography and the corneal dynamic behaviour captured by the Corvis ST device. Strong correlations were evident between the keratoconus severity and the symmetric modes and between the keratoconus features (weak region) and the asymmetric modes. In addition, the machine-learning clustering system classified the keratoconus responses into different features, and a strong relationship was found between the system and the benchmarks that were constructed. Further studies are required to gather more patient data to establish a more extensive database for validation. It is also essential to develop the algorithm into an easy-to-use automatic process for physicians.

## Declaration of conflicting interests

The authors declare that there are no conflicts of interest.

## Funding

This research receives no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## ORCID iD

Jia-Yush Yen https://orcid.org/0000-0001-8795-9211

## References

1. Kymionis GD, Plaka AD and Kontadakis GA: Clinical signs and differential diagnosis of keratoconus. Chapter 6 in: *Textbook on Keratoconus. New Insights* (by Barbara A): Jaypee Brothers Medical Publishers (P) Ltd. 2011: pp 33–39.
2. Heidari Z, Hashemi H, Mohammadpour M, et al. Evaluation of corneal topographic, tomographic and biomechanical indices for detecting clinical and subclinical keratoconus: a comprehensive three-device study. *Int J Ophthalmol* 2021; 14: 228–239.
3. Nguyen BA, Reilly MA and Roberts CJ. Biomechanical contribution of the sclera to dynamic corneal response in air-puff induced deformation in human donor eyes. *Exp Eye Res* 2020; 191: 107904.
4. Maklad O, Eliasy A, Chen KJ, et al. Fluid-Structure Interaction Based Algorithms for IOP and Corneal Material Behavior. *Front Bioeng Biotechnol* 2020; 8: 970.
5. Liu J, Wang Y, Zou H, et al. Influence of corneal shape parameters on corneal deformation responses measured with a Scheimpflug camera. *Int Ophthalmol* 2021; 41: 2853–2859.
6. Herber R, Vinciguerra R, Lopes B, et al. Repeatability and reproducibility of corneal deformation response parameters of dynamic ultra-high-speed Scheimpflug imaging in keratoconus. *J Cataract Refract Surg* 2020; 46: 86–94.
7. Andreassen TT, Simonsen AH and Oxlund H. Biomechanical properties of keratoconus and normal corneas. *Exp Eye Res* 1980; 31: 435–441.
8. Wang LK, Tian L and Zheng YP. Determining in vivo elasticity and viscosity with dynamic Scheimpflug imaging analysis in keratoconic and healthy eyes. *J Biophotonics* 2016; 9: 454–463.
9. Lanza M, Iaccarino S and Bifani M. In vivo human corneal deformation analysis with a Scheimpflug camera, a critical review. *J Biophotonics* 2016; 9: 464–477.
10. Joda AA, Shervin MM, Kook D, et al. Development and validation of a correction equation for Corvis tonometry. *Comput Methods Biomech Biomed Engin* 2015; 19: 943–953.
11. Hon Y and Lam AK. Corneal deformation measurement using scheimpflug noncontact tonometry. *Optom Vis Sci* 2013; 90: e1–e8.
12. Vinciguerra R, Ambrósio R Jr, Elsheikh A, et al. Detection of Keratoconus With a New Biomechanical Index. *J Refract Surg* 2016; 32: 803–810.
13. Roberts CJ, Mahmoud AM, Bons JP, et al. Introduction of two novel stiffness

parameters and interpretation of air Puff–Induced biomechanical deformation parameters with a dynamic scheimpflug analyzer. *J Refract Surg* 2017; 33: 266–273.
14. Elham R, Jafarzadehpur E, Hashemi H, et al. Keratoconus diagnosis using Corvis ST measured biomechanical parameters. *J Curr Ophthalmol* 2017; 29: 175–181.
15. Mercer RN, Waring GO 4th, Roberts CJ, et al. Comparison of corneal deformation parameters in keratoconic and normal eyes using a non-contact tonometer with a dynamic ultra-high-speed Scheimpflug camera. *J Refract Surg* 2017; 33: 625–631.
16. Yang K, Xu L, Fan Q, et al. Repeatability and comparison of new Corvis ST parameters in normal and keratoconus eyes. *Sci Rep* 2019; 9: 1–10.
17. Chan TC, Wang YM, Yu M, et al. Comparison of corneal dynamic parameters and tomographic measurements using Scheimpflug imaging in keratoconus. *Br J Ophthalmol* 2018; 102: 42–47.
18. Shih PJ, Cao HJ, Huang CJ, et al. A corneal elastic dynamic model derived from Scheimpflug imaging technology. *Ophthalmic Physiol Opt* 2015; 35: 663–672.
19. Kolozsvári BL, Petrovski G, Gogolák P, et al. Association between mediators in the tear fluid and the severity of keratoconus. *Ophthalmic Research* 2014; 51: 46–51.
20. Millodot M, Shneor E, Albou S, et al. Prevalence and associated factors of keratoconus in Jerusalem: a cross-sectional study. *Ophthalmic epidemiology* 2011; 18: 91–97.
21. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577.
22. Batista GE, Keogh EJ, Tataw OM, et al. CID: an efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery* 2014; 28: 634–669.
23. Räsänen T and Kolehmainen M: Feature-based clustering for electricity use time series data. In: Adaptive and Natural Computing Algorithms. (Kolehmainen, M., et al. (Eds). Berlin: Springer, ICANNGA 2009, Lecture Notes in Computer Science, vol 5495 pp 401-412. Available from: https://doi.org/10.1007/978-3-642-04921-7_41
24. Soni B, Das PK, Thounaojam DM. CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection. *IeT Image Processing* 2018 February;12: 167–178.
25. Christ M, Braun N, Neuffer J, et al. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing* 2018; 307: 72–77.
26. Fodor I. A Survey of Dimension Reduction Techniques. United States. [Internet] 2002 Available from https://doi.org/10.2172/15002155
27. Taşkin G and Crawford MM. An out-of-sample extension to manifold learning via meta-modeling. IEEE Trans Image Process 2019 May 15. doi: 10.1109/TIP.2019.2915162. Epub ahead of print.
28. Leon-Medina JX, Anaya M, Pozo F, et al. Nonlinear feature extraction through manifold learning in an electronic tongue classification task. *Sensors* 2020; 20: 4834.
29. Lee DD and Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 1999; 401: 788–791.
30. Lee D and Seung HS. Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems 13 – Proceedings of the 2000 Conference, NIPS 2000. Advances in Neural Information Processing Systems, Neural information processing systems foundation, 14th Annual Neural Information Processing Systems Conference, NIPS 2000, Denver, CO, United States, 11/27/00.
31. Tenenbaum JB, De Silva V and Langford JC. A global geometric framework for non-linear dimensionality reduction. *Science* 2000; 290: 2319–2323.
32. Li SZ, Hou XW, Zhang HJ, et al. Learning spatially localized, parts-based representation. In: IEEE Conference on Computer Vision and Pattern Recognition CVPR: 2001: pp 207–212.
33. Bengio Y, Paiement J-f, Vincent P, et al. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. *Advances in neural information processing systems* 2003; 16.

34. Lakhina A, Crovella M, Diot C. Mining anomalies using traffic feature distributions. *ACM SIGCOMM computer communication review* 2005; 35: 217–228.
35. Aghabozorgi S, Shirkhorshidi AS, Wah TY. Time-series clustering–a decade review. *Information systems*. 2015; 53: 16–38.
36. Lord E, Willems M, Lapointe F-J, et al. Using the stability of objects to determine the number of clusters in datasets. *Information Sciences* 2017; 393: 29–46.
37. De Amorim RC, Hennig C. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information sciences* 2015; 324: 126–145.
38. Ou G, Wakabayashi K, Satoh T. Searching Behavior Analysis of Online Shopping Based on Information Content of Query Words. In: *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*; 2019: Jul 7 (pp. 43–48). IEEE.
39. von Csefalvay C. Vector quantisation and partitioning of COVID-19 temporal dynamics in the United States. arXiv preprint arXiv: :2008.00559:2020 Aug 2.