Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A label-free machine learning atlas of corneal aberration clusters from dual-surface Zernike coefficients: a retrospective cross-sectional study of 3,533 eyes

Orhan Karakulak

drorhankarakulak@gmail.com

Kirikhan State Hospital

Emre Ayıntap

Izmir City Hospital, University of Health Sciences

İpek Çıkmazkara

Ministry of Health Izmir City Hospital

Atılım Armağan Demirtaş

Izmir City Hospital, University of Health Sciences

Hasan Aytoğan

Izmir Tepecik Training and Research Hospital, University of Health Sciences

Tuncay Küsbeci

Izmir City Hospital, University of Health Sciences

Research Article

Keywords: Corneal aberrations, Zernike polynomials, Unsupervised machine learning, Keratoconus, Corneal tomography, Cluster analysis, Severity gradient

Posted Date: April 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9420086/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Scalar indices summarise corneal aberrations by magnitude, but they do not describe the overall wavefront pattern. We investigated whether the full set of anterior and posterior corneal Zernike coefficients could identify candidate aberration clusters without using diagnostic labels.

Methods

We analysed Pentacam HR examinations from 3,533 right eyes at a single tertiary centre. Second- to sixth-order Zernike coefficients from both corneal surfaces were combined, block-normalised, reduced by PCA for discovery-set cluster-number selection, and analysed with spherical k-means within a 70/30 internal discovery-replication framework. Standard tomographic indices were reserved for post hoc interpretation, and selected partitions were refit on the full cohort for descriptive summarisation.

Results

Among the examined partitions, the optimum of the prespecified study composite was a two-cluster solution, separating a normal-to-mild cluster (n = 2,472; median BAD-D 1.52) from an ectatic-severe cluster (n = 1,061; median BAD-D 7.25). Because total RMS predicted this partition with 84.6% accuracy, it largely reflected severity. A three-cluster model added an intermediate cluster (n = 915; median BAD-D 2.12, within the suspect range; 36.5% normal, 23.7% suspicious, 39.8% abnormal) and reduced RMS-based predictability to 63.2%, indicating added pattern information beyond scalar severity alone. A four-cluster model produced over-segmentation, with one cluster pair no longer distinguishable on post hoc tomographic indices. Internal split-sample reproducibility was strongest for K = 2 and lower for K = 3 (for K = 3: ARI 0.607; cosine similarity 0.886).

Conclusions

Unsupervised clustering of dual-surface Zernike coefficients organised corneal aberrations in a manner not fully determined by scalar severity alone, while also identifying an intermediate cluster near the BAD-D suspect range. A two-cluster partition was the most stable solution but largely mirrored overall aberration severity; a three-cluster partition was less stable yet added a clinically interpretable intermediate cluster that scalar severity alone did not capture, whereas a four-cluster partition produced clusters that could no longer be reliably distinguished on post hoc tomographic indices. These findings support unsupervised clustering as a descriptive, hypothesis-generating atlas of candidate aberration clusters rather than a definitive phenotype taxonomy.

Background

The cornea accounts for approximately two-thirds of the eye's total refractive power [1] and is the primary source of optical aberrations [2]. Distortions in the corneal wavefront are defined using Zernike polynomials in accordance with the OSA/ANSI standard [3, 4]: second-order terms represent defocus and regular astigmatism, whilst third-order terms represent coma and trefoil. Fourth-order terms include tetrafoil, secondary astigmatism and spherical aberrations [5]. Clinicians routinely reduce these data to a scalar value, such as the root mean square (RMS), which serves as a general indicator of aberration severity, thereby obtaining a simpler representation [6]. Whilst RMS values are useful for screening and severity grading, they do not, on their own, provide information about the appearance of the aberrations or the shape of the cornea [7, 8]. Furthermore, they do not show how the individual aberration components change together across both corneal surfaces or what overall shape they form [9–11]. Two corneas with the same RMS value may have very different compositions in terms of coma, trefoil and spherical aberrations, and this can lead to different optical outcomes following refractive surgery or with different intraocular lens designs [12–14].

Anterior and posterior corneal aberrations have been examined in detail separately, and the interaction between them may be significant in ectasia screening [15]. In cataract or refractive surgery, this may be clinically significant in many cases, as selecting an intraocular lens suited to the patient's corneal aberration profile can improve retinal image quality [16, 17]. Higher-order aberrations (HOA) on the posterior surface are approximately one-quarter of those on the anterior surface. However, in keratoconus, coma and trefoil on the posterior surface progress in the opposite direction to the anterior surface and partially counterbalance the anterior aberrations [18]. In the detection of subclinical keratoconus, anterior surface vertical coma (Z3 − 1) is the strongest single predictor (AUC 0.922–0.980); the addition of posterior surface coefficients did not significantly increase this already high accuracy [19, 20]. These controlled analyses demonstrate the diagnostic value of individual parameters in the detection of ectasia. To the best of our knowledge, it has not yet been fully determined how the complete set of coefficients obtained from both surfaces collectively defines the optical profile, nor whether this profile can be used for classification according to candidate aberration clusters.

Current corneal classification relies largely on scalar grading systems. The severity of keratoconus has traditionally been graded using the Amsler–Krumeich system, which is based on anterior keratometry and corneal thickness, however, this system fails to capture posterior surface changes and subclinical ectasia [21]. Alió and Shabayek [22] proposed grading based on anterior coma-like RMS and demonstrated a strong correlation (r = 0.833) between aberration magnitude and mean keratometry. Belin and colleagues developed the ABCD grading system, which integrates anterior and posterior curvature, pachymetry and visual acuity [23]. In contrast, BAD is a screening tool based on regression analysis of nine tomographic parameters; with a BAD-D value > 1.88, it exhibits a false-positive rate of 2.5% and a false-negative rate of less than 1%, emerging as a highly sensitive marker for early ectasia screening [24]. However, this high level of technological success in disease detection has not yet been fully reflected in the classification of the full spectrum of disease severity. Indeed, the 2015 Global

Keratoconus Consensus acknowledged that existing grading systems (e.g., Amsler–Krumeich) are inadequate in reflecting tomographic and morphological progression; it emphasised the need for new classification approaches that integrate clinical and functional findings with tomographic parameters [25]. These limitations are not unique to ectasia: in both refractive surgery and cataract surgery, scalar indices do not provide information on which aberration modes are dominant in a given cornea, yet it is precisely this composition that determines visual outcomes [8, 26–28]. This gap between the devices' ability to detect the condition and the literature's failure to objectively classify this complex spectrum further highlights the need for new models based on the nature of the data itself, independent of fixed threshold values and human bias.

Artificial intelligence algorithms are also widely used in ophthalmology, including corneal ectasia detection. However, many current automated keratoconus detection methods are supervised learning models based on clinically labelled data [29]. Whilst supervised learning approaches provide high accuracy in diagnosing distinct diseases, they require large amounts of labelled data, are susceptible to observer bias arising from manual classification, and force a continuous disease spectrum into categories using rigid threshold values [29]. In contrast, unsupervised machine learning offers the ability to objectively cluster data with similar features by detecting hidden groupings in unlabelled data, without human intervention or prior knowledge [30–32]. To overcome the limitations of supervised learning in the literature, which relies on fixed threshold values, and to objectively map the morphological spectrum ranging from healthy corneas to advanced ectasia, semi-supervised clustering models have recently been successfully developed [33]. However, there remains a need for end-to-end, fully unsupervised studies that do not require pre-labelled diagnostic groups at any stage of the process.

To our knowledge, no study has applied unsupervised clustering to the complete set of Zernike coefficients obtained from both corneal surfaces across the full clinical spectrum. In this study, we applied spherical k-means clustering to block-normalised dual-surface Zernike representations derived from 3,533 eyes. The fundamental question was: can an unsupervised machine learning algorithm group corneal aberrations into reproducible aberration clusters without any diagnostic labels? Can these candidate clusters be presented as a visual reference atlas?

Methods

Study design and ethical approval

This retrospective, cross-sectional, single-centre study utilised corneal tomography data obtained between September 2022 and October 2025 at Izmir City Hospital, a tertiary referral centre, using the Pentacam HR Scheimpflug imaging system (OCULUS, Wetzlar, Germany). Ethical approval for the secondary analysis of de-identified archival Pentacam records was obtained from the Clinical Research Ethics Committee of Izmir City Hospital (No. 2026/337) before data extraction for the present research analysis; the underlying clinical examinations had been accrued during routine care between September 2022 and October 2025. All procedures adhered to the tenets of the Declaration of Helsinki, and the

requirement for informed consent was waived because the study used retrospective anonymised data. The study complies with the STROBE guidelines for cross-sectional studies.

Cohort formation

Raw examination records were obtained from three standard Pentacam export files: Zernike wavefront analysis (ZERNIKE-WFA), the Belin/Ambrósio enhanced ectasia display (BADisplay-LOAD) and topometric indices (INDEX-LOAD). Primary study variables were derived from these device export data; no manual clinical chart review of diagnoses, progression status, or visual outcomes was undertaken. Sex was retrieved separately from hospital medical records for descriptive purposes because it is not included in the standard Pentacam export. A seven-step sequential exclusion pipeline was applied: (1) retention of examinations present in all three files; (2) exclusion of non-zero device error codes; (3) removal of missing Zernike coefficients; (4) exclusion of 8 eyes with missing BAD-D values because BAD-D was the prespecified post-hoc anchor for cluster interpretation; (5) removal of invalid pachymetry (CCT < 100 μm); (6) selection of the right eye only and retention of the earliest visit; (7) exclusion of participants under 18 years of age. Diagnostic labels were not available; the only disease-related classification used in the study was the BAD-D-based categorisation provided by the device (Normal < 1.6; Suspicious 1.6–2.6; Abnormal > 2.6). These three categories correspond to the manufacturer's built-in BAD-D output bands displayed on the Pentacam Belin/Ambrósio enhanced ectasia display and were used here solely as a post-hoc descriptive anchor. They are distinct from the single 1.88 cut-off reported by Villavicencio et al. [24] for binary ectasia screening, which was not used for any classification or analysis in the present study.The use of the right eye alone controls for the known enantiomorphic (mirror-symmetric) laterality effect in keratoconus corneas. Because this was a retrospective analysis of archival exports, no formal sample-size calculation was performed; all eligible examinations meeting the prespecified criteria were included. This process yielded a final cohort of 3,533 right eyes from 3,533 unique adult patients (Fig. 1).

Zernike wavefront extraction

Zernike coefficients defining the anterior (CF) and posterior (CB) corneal surface wavefronts were obtained according to the OSA standard [4] using a 6.0 mm analysis diameter (2nd-6th orders). A combined feature vector of 50 dimensions (25 CF + 25 CB) was constructed by excluding the piston and tilt terms. A Zernike expansion up to the sixth order was preferred; it has been shown that higher-order expansions do not significantly reduce fitting error in irregular corneas and that higher-order terms can even cancel out lower-order terms in an unpredictable manner [34]. A parallel high-order aberration (HOA) set (44 dimensions), limited to 3rd to 6th orders, was used in a dedicated sensitivity analysis to test whether the K = 4 over-segmentation pattern persisted in an independent feature subset.

Data standardisation

The magnitudes of the low-order Zernike terms are significantly greater than those of the high-order terms. To prevent these scale differences from dominating the clustering, each eye was partitioned into 10 surface-by-order blocks (CF and CB, orders 2–6). Each block was L2-normalised independently, meaning that it was rescaled to unit length so that larger numeric coefficients did not automatically dominate the comparison. A very small epsilon term was used only to keep the calculation stable when a block contained essentially no signal (EPS-stabilised division). The blocks were then concatenated and the combined vector was globally L2-normalised so that each eye entered the cosine-based analysis on the same overall scale. This preprocessing ensured equal weighting of both surfaces and all Zernike orders before downstream analyses by mapping each eye onto a unit hypersphere. The approach is conceptually aligned with spherical and robust functional-data strategies such as those described by Locantore et al. [35].

Discovery–replication framework

The cohort was split into a patient-level discovery set (70%; n = 2,473) and an internal replication set (30%; n = 1,060) using a fixed random seed (42). Cluster-number selection and severity-dependence analyses were performed in the discovery set. Internal split-sample reproducibility was then assessed in the held-out replication set by comparing independently fitted solutions with discovery-derived assignments after centroid matching. After this validation step, the selected K = 2 and clinically oriented K = 3 solutions were refit on the full cohort to generate the descriptive group sizes, tables, and atlas figures reported in the Results.

Dimension reduction and clustering

Within the discovery set, PCA was fit to the block-normalised vectors and the smallest number of components explaining at least 90% of the variance was retained, capped at 20 components. At the selected cap, 20 components captured 79.5% of the variance in the full dataset (Additional file 1: Figure S1). The discovery PCA transformation was then applied to the replication set, and the retained score vectors were clustered directly without additional post-PCA L2 re-normalisation for cluster-number comparison and split-sample reproducibility analyses. As in prior dimensionality-reduction studies by Yousefi et al. [30] and Rodríguez et al. [36], the aim was not exact reconstruction but retention of discriminative variation for clustering.

We used spherical k-means as the primary clustering algorithm (20 initialisations, 300 iterations, convergence 10^-6). For cluster-number comparison and split-sample reproducibility, spherical k-means was applied in the shared discovery PCA space. For the final descriptive atlases and tables, the selected K values were refit on the full block-normalised coefficient vectors. In both settings, the objective was cosine-based clustering by angular orientation rather than Euclidean distance. In practical terms, this means that eyes were grouped by the overall shape of the wavefront pattern rather than simply by how large the aberration values were.

Cluster number selection

We compared cluster numbers across K = 2 to K = 6 using a composite summary of four prespecified quantities: cosine silhouette, bootstrap stability, severity predictability, and maximum pairwise centroid similarity. Because this composite included an interpretive severity term, K = 2 is best described as the optimum of the prespecified study composite rather than as a purely internal statistical optimum. Severity predictability was included to quantify how strongly a partition collapsed onto scalar aberration magnitude. On that basis, K = 3 was retained as a prespecified clinically interpretable secondary solution to examine whether additional pattern-level granularity could be gained without a major loss of stability. K = 4 was tested as a deliberate over-segmentation probe (Additional file 1: Table S1; Additional file 1: Figure S2).

Post-hoc clinical characterisation

To avoid circularity, standard Pentacam indices (ISV, IVA, KI, CKI, IHA, IHD, BAD-D, Kmax, CCT) were excluded from clustering and used only to describe the resulting clusters. BAD-D was chosen as the main post-hoc anchor because it is a widely validated composite screening measure [37, 38]. Global differences were assessed using the Kruskal–Wallis H test and eta-squared effect sizes.

Validation of independence from severity

To test whether cluster assignment reflected wavefront shape rather than overall aberration size, we trained multinomial logistic-regression models to predict cluster labels from total RMS or HOA RMS alone. These analyses were performed in the discovery set using fivefold stratified cross-validation, and overall accuracy, balanced accuracy, macro-F1, and AUC were reported as descriptive probes rather than formal validation end points. Additional sensitivity analyses in Additional file 1 address cluster-number robustness, discovery-replication concordance, and cosine-versus-Euclidean silhouette behaviour.

Statistical significance was set at $p < 0.05$ (two-tailed). For the three-cluster solution, pairwise post hoc comparisons of tomographic indices were performed using Dunn's test with Bonferroni correction following a significant Kruskal–Wallis H test. Because these inferential comparisons were rank-based, pairwise Cohen's d values were reported only as standardised descriptive summaries of separation rather than as parametric inferential statistics. Sex distribution across clusters was compared using the chi-squared test with Cramer's V as the effect-size measure. The composite score for cluster-number selection weighted the four constituent metrics equally after min–max normalisation; the exact formula and robustness to alternative weighting schemes are detailed in Additional file 1: Table S7. Bootstrap stability was assessed by resampling with replacement (B = 100) and computing the mean Adjusted Rand Index between bootstrap and original cluster assignments.

Statistical software

All analyses were performed in Python 3.12 using NumPy 2.2.6, pandas 2.3.3, scikit-learn 1.7.1, SciPy 1.17.0, and Matplotlib 3.10.5.

Results

Study cohort

Following sequential filtering of 19,181 matched examinations, 3,533 right eyes were obtained. The median age was 32.10 years [IQR 24.41–47.41]. The median total RMS was 3.85 μm [IQR 2.27–6.93] and the median BAD-D was 2.02 [IQR 1.22–5.09; range − 1.15 to 28.87], reflecting a heterogeneous cohort ranging from normal corneas to advanced ectasia. Overall, 1,346 eyes (38.1%) were normal, 788 (22.3%) were suspicious, and 1,399 (39.6%) were abnormal by BAD-D. Sex data were obtained from hospital medical records, as they are not included in the standard Pentacam export (Table 1).

Table 1
Demographic and clinical characteristics of the study cohort (N = 3,533)

Variable	N	Mean ± SD	Median [IQR]	Range
Sex, male, n (%)	1,789 (50.6%)	—	—	—
Sex, female, n (%)	1,744 (49.4%)	—	—	—
Age (years)	3,533	37.62 ± 16.58	32.10 [24.41–47.41]	18.00–96.44
Total RMS (μm)	3,533	5.57 ± 4.84	3.85 [2.27–6.93]	0.63–35.36
HOA RMS (μm)	3,533	1.29 ± 1.28	0.66 [0.48–1.68]	0.17–8.61
BAD-D	3,533	3.69 ± 3.89	2.02 [1.22–5.09]	−1.15–28.87
Kmax (D)	3,533	48.54 ± 5.77	46.70 [44.80–50.22]	39.03–80.62
CCT (μm)	3,533	509.46 ± 57.29	515.00 [474.00–549.00]	178.00–733.00
ISV	3,533	42.01 ± 31.97	31.00 [20.00–53.00]	6.00–228.00
IVA	3,533	0.36 ± 0.36	0.19 [0.12–0.48]	0.02–2.38
KI	3,533	1.07 ± 0.11	1.03 [1.00–1.10]	0.72–1.76
CKI	3,533	1.02 ± 0.04	1.01 [1.00–1.02]	0.80–1.32
IHA	3,533	14.87 ± 17.46	8.50 [3.40–18.70]	0.00–122.70
IHD	3,533	0.04 ± 0.05	0.02 [0.01–0.05]	0.00–0.36

Primary finding: two-cluster partition

The composite metric identified two clusters as the best solution (cosine silhouette 0.251; bootstrap stability 0.989). The normal-to-mild cluster (n = 2,472, 70.0%) had low aberration levels (median total

RMS 2.91 µm, HOA RMS 0.54 µm; median BAD-D 1.52, Kmax 45.70 D, CCT 532 µm). Its average wavefront was regular and symmetrical, dominated by defocus (Fig. 2). The ectatic-severe cluster (n = 1,061, 30.0%) had high aberration levels (median total RMS 9.60 µm, HOA RMS 2.30 µm; median BAD-D 7.25, Kmax 53.27 D, CCT 456 µm). Its average wavefront showed asymmetry, dominant coma, and inferior decentration. Total RMS alone predicted cluster membership with 84.6% accuracy, confirming that this split largely reflects severity. The cosine silhouette of 0.251 reflects the overlap expected in a continuous disease spectrum, consistent with earlier corneal clustering studies [30, 33].

Table 2
Clinical summary of the two-cluster solution (K = 2). Values are reported as median [IQR], n (%), or percentages, as appropriate.

Variable	Cluster A (n = 2,472)	Cluster B (n = 1,061)	P-value
BAD-D	1.52 [1.01–2.19]	7.25 [5.03–10.31]	< 0.001
Kmax (D)	45.70 [44.29–47.22]	53.27 [49.51–58.00]	< 0.001
CCT (µm)	532 [507–559]	456 [425–487]	< 0.001
Total RMS (µm)	2.91 [1.95–4.31]	9.60 [6.31–13.55]	< 0.001
Normal (%)	53.3	2.7	—
Suspicious (%)	30.5	3.1	—
Abnormal (%)	16.2	94.2	—
Sex, male, n (%)	1,216 (49.2%)	573 (54.0%)	0.010
Sex, female, n (%)	1,256 (50.8%)	488 (46.0%)	—

Extended resolution: three-cluster partition

To test whether splitting further reveals clinically meaningful structure, we evaluated a three-cluster solution (Fig. 3): normal-mild cluster (n = 1,752, 49.6%; median BAD-D 1.46; 56.8% normal), intermediate cluster (n = 915, 25.9%; median BAD-D 2.12; 36.5% normal, 23.7% suspicious, 39.8% abnormal) and ectatic-severe cluster (n = 866, 24.5%; median BAD-D 8.07; 95.8% abnormal). All three pairwise comparisons were statistically significant: normal-mild vs intermediate mean |Cohen's d| = 0.579; normal-mild vs ectatic = 2.167; intermediate vs ectatic = 1.375. Global between-cluster effect sizes across the main tomographic indices were large (eta-squared range 0.25–0.58). The intermediate cluster had a median BAD-D in the suspect range and showed wavefront asymmetry with early inferior decentration. Whether this cluster represents a distinct population or a division within a continuum cannot be determined from cross-sectional data.

Table 3
Clinical summary of the three-cluster solution (K = 3). Values are reported as median [IQR], n (%), or percentages, as appropriate.

Variable	Cluster A (n = 1,752)	Cluster B (n = 915)	Cluster C (n = 866)	P-value
BAD-D	1.46 [0.99–2.05]	2.12 [1.52–4.27]	8.07 [7.07–12.29]	< 0.001
Kmax (D)	45.56 [44.22–47.06]	46.54 [44.78–48.64]	54.33 [50.47–59.42]	< 0.001
CCT (μm)	537 [511–561]	514 [485–544]	451 [420–480]	< 0.001
Total RMS (μm)	2.91 [2.05–4.00]	3.85 [2.50–6.50]	11.70 [8.45–15.39]	< 0.001
Normal (%)	56.8%	36.5%	0.2%	—
Suspicious (%)	32.0%	23.7%	4.0%	—
Abnormal (%)	11.2%	39.8%	95.8%	—
Sex, male, n (%)	874 (49.9%)	456 (49.8%)	459 (53.0%)	0.277
Sex, female, n (%)	878 (50.1%)	459 (50.2%)	407 (47.0%)	—

Resolution limit: four-cluster over-segmentation

We examined a four-cluster solution to test the upper limit of meaningful resolution. At K = 4, one pair (P1-P2) was not distinguishable on the post hoc tomographic indices: six of the nine variables were statistically non-significant, with a mean |Cohen's d| = 0.228 (negligible-small). In contrast, all other cluster pairs showed an average |d| of 0.998–2.757. The same pattern of over-segmentation was confirmed in a parallel HOA-only analysis using an independent set of features (average |d| = 0.229). Together with the optimum of the prespecified study composite at K = 2 and the added clinical granularity observed at K = 3, this supports a three-cluster solution as the highest clinically interpretable resolution among the examined partitions in this feature space. Details for K = 4 are provided in Additional file 1 (Figures S3-S4; Tables S2-S5).

Relationship with severity

Total RMS alone predicted discovery-set cluster membership with 84.6% accuracy, 80.3% balanced accuracy, and AUC 0.890 at K = 2, versus 63.2% accuracy, 56.3% balanced accuracy, macro-F1 0.492, and macro-averaged one-versus-rest AUC 0.766 at K = 3 (Fig. 4). High predictability at K = 2 indicates that the main partition largely reflects severity. The lower balanced metrics at K = 3, together with the still acceptable silhouette and stability values, suggest that the intermediate cluster captures information partly independent of severity, although this requires longitudinal validation.

Reproducibility

Cluster proportions were similar between the discovery and replication sets. Internal split-sample concordance was highest for the two-cluster solution (mean matched centroid cosine similarity 0.995;

ARI 0.970 between projected and independently refit replication labels) and more moderate for the three-cluster solution (0.886 and 0.607, respectively). These findings indicate that the clinically interpretable three-cluster solution remained reproducible, but less so than the more stable two-cluster partition.

Sex distribution across clusters

Sex distribution did not differ significantly across the three-cluster solution ($\chi^2 = 2.57$, $p = 0.277$, Cramer's $V = 0.027$). Cluster A contained 874 males (49.9%), Cluster B 456 males (49.8%), and Cluster C 459 males (53.0%) (Tables 2–3). In the two-cluster solution, a modest but statistically significant difference was observed ($\chi^2 = 6.69$, $p = 0.010$, Cramer's $V = 0.044$), with a slightly higher male proportion in Cluster B (54.0% vs. 49.2%); however, the negligible effect size ($V = 0.044$) indicates that sex had no meaningful influence on cluster assignment.

Discussion

This study set out to map the corneal ectasia spectrum without predefined diagnostic labels or scalar thresholds. Using surface-by-order block normalisation, PCA-based discovery-set model selection, and spherical k-means, we identified reproducible aberration-cluster structure from the native geometry of the data rather than from externally imposed categories. This distinction is important: after block-based L2 normalisation, clustering was driven primarily by angular pattern similarity rather than overall aberration magnitude. In that sense, the present framework is conceptually aligned with the spherical and robust functional-data perspective described by Locantore and colleagues [35] while adapting that intuition to dual-surface corneal aberration clustering.

The central interpretive finding is the divergence between the optimum of the prespecified study composite and clinical usefulness. K = 2 was the strongest solution on that composite and also showed the highest internal split-sample reproducibility, but total RMS predicted that partition with 84.6% accuracy, 80.3% balanced accuracy, and AUC 0.890, indicating that it largely recapitulates a familiar severity axis. By contrast, moving to K = 3 incurred only a modest loss in silhouette and stability whilst reducing RMS predictability to 63.2% accuracy, 56.3% balanced accuracy, macro-F1 0.492, and macro-averaged one-versus-rest AUC 0.766. That trade-off suggests that the third cluster contributes information beyond scalar magnitude. Consistent with this interpretation, the intermediate cluster comprised 915 eyes, had a median BAD-D of 2.12, and showed a diagnostic composition positioned between the normal-mild and ectatic-severe clusters rather than simple enrichment of a single BAD-D category. In other words, Cluster B was not merely a numeric midpoint on a severity scale; it represented a recurring suspect-range wavefront pattern that reappeared across eyes. Bootstrap stability remained high at K = 3, but internal split-sample concordance was more moderate than for K = 2 (mean centroid cosine similarity 0.886; ARI 0.607), supporting K = 3 as a clinically informative secondary solution rather than the most stable partition.

The K = 4 analysis was informative not because it improved resolution, but because it exposed the resolution limit of this feature space. Forcing a four-cluster solution split one cluster pair into

subclusters that were not distinguishable on the post hoc tomographic indices, and the same over-segmentation pattern reappeared in the independent HOA-only feature set. Taken together with the optimum of the prespecified study composite at K = 2 and the added granularity at K = 3, these findings support K = 3 as the most clinically informative secondary partition examined here. By contrast, K = 4 primarily introduces artificial subdivisions without a clear morphological or clinical counterpart.

These findings speak to a persistent limitation of current ectasia classification systems: most remain organised around scalar severity scores. Amsler-Krumeich, coma-based grading, and even the more comprehensive ABCD system summarise disease burden but do not resolve how aberration modes combine across the anterior and posterior cornea [21–25]. Prior machine-learning studies have shown that unsupervised and semi-supervised approaches can reveal structure in corneal data [30, 31, 33], and PCA-based work has clarified broad modes of corneal variation [36]. Our contribution is to combine dual-surface Zernike representation, block-normalised spherical geometry, discovery-set PCA compression, and label-free clustering in a single workflow that yields candidate aberration clusters that are internally reproducible and visually interpretable.

The potential clinical value of this framework lies less in replacing existing diagnostic indices than in complementing them. The intermediate cluster may help characterise eyes that occupy diagnostically ambiguous territory on scalar metrics yet share a coherent wavefront pattern. Clinically, an eye with a borderline or suspect BAD-D but a Cluster B-like spatial pattern may warrant caution rather than reassurance: in refractive surgery work-up, it could justify repeat tomography, adjunctive biomechanical or epithelial-thickness assessment, and a lower threshold to defer corneal laser ablation if other risk markers align. A similar logic may apply in cataract surgery planning, where a cornea that appears only moderately abnormal on scalar indices but falls into this pattern-based intermediate cluster may be a less suitable candidate for multifocal or other premium IOL strategies that are sensitive to irregular higher-order aberrations. These examples are hypothesis-generating rather than outcome-validated, but they illustrate how the atlas could change decision-making precisely in the grey zone where scalar summaries are most equivocal. More broadly, the atlas provides a label-free morphological scaffold for future studies linking these candidate clusters to biomechanics, genetics, and longitudinal progression.

Several limitations should temper interpretation. The cross-sectional design does not establish whether eyes within the intermediate cluster will progress over time, so longitudinal validation is essential. The data were derived from a single tertiary centre and a single device platform (Pentacam HR), which may limit generalisability and may enrich the cohort for more advanced ectasia than a population-based sample. The discovery-replication framework demonstrated internal split-sample reproducibility, not external validation across centres or devices. Because no independent clinical ground truth was available, cluster interpretation remained post hoc and relied on BAD-D and related tomographic indices; although these variables were excluded from clustering, they arise from the same imaging session and are not fully independent of the underlying corneal geometry. Excluding the 8 eyes with missing BAD-D values to preserve complete-case post hoc anchoring may also have introduced limited selection bias. Functional correlates such as best-corrected visual acuity and refraction were not available. This limits

direct structure-function inference, but the present study was designed primarily as a morphological atlas of aberration geometry rather than as a functional outcome study. The current atlas should therefore be viewed as a structural scaffold for prospective work linking cluster membership to visual quality, refraction, biomechanics, and postoperative outcomes. In addition, PCA retained 79.5% of the variance, so some high-frequency local irregularities may not have been preserved, and clustering on PCA scores does not reproduce the original hyperspherical geometry exactly. The RMS-based classifier analyses were intended as descriptive probes of severity dependence; we therefore report balanced accuracy and AUC alongside simple accuracy, but these probes do not establish causal independence from severity. Restricting the analysis to right eyes prevented assessment of inter-eye asymmetry, and Zernike polynomials themselves remain an imperfect basis for modelling highly irregular ectatic corneas.

Conclusions

This study shows that dual-surface Zernike coefficients contain reproducible unlabeled structure. Among the examined partitions, K = 2 was the most stable and reproducible solution and largely reflected aberration severity, whereas K = 3 provided a less stable but clinically interpretable secondary partition with an intermediate cluster centred in the BAD-D suspect range. That intermediate cluster may be most relevant in gray-zone eyes, where scalar indices alone do not fully describe the spatial pattern of irregularity. The failure of K = 4 to yield additional distinguishable structure indicates that this framework is best interpreted as a descriptive, hypothesis-generating atlas of candidate aberration clusters rather than a definitive phenotype taxonomy. External and longitudinal validation, including links to visual outcomes and surgical decision-making, will be required before any cluster is treated as a phenotype class.

Abbreviations

ARI
Adjusted Rand Index
AUC
Area Under the Receiver Operating Characteristic Curve
BAD-D
Belin/Ambrósio Enhanced Ectasia Display – D value
CB
Posterior corneal surface coefficients
CCT
Central Corneal Thickness
CF
Anterior corneal surface coefficients
CKI
Centre Keratoconus Index

HOA
Higher-Order Aberrations
IHA
Index of Height Asymmetry
IHD
Index of Height Decentration
IQR
Interquartile Range
ISV
Index of Surface Variance
IVA
Index of Vertical Asymmetry
KI
Keratoconus Index
Kmax
Maximum Keratometry
LR
Logistic Regression
OSA/ANSI
Optical Society of America / American National Standards Institute
PCA
Principal Component Analysis
RMS
Root Mean Square
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology

Declarations

Ethics approval and consent to participate

This study was approved by the Clinical Research Ethics Committee of Izmir City Hospital (No. 2026/337). Approval for the secondary analysis of anonymised archival data was obtained before extraction for the present research analysis, although the underlying clinical examinations had accrued between September 2022 and October 2025. The study adhered to the tenets of the Declaration of Helsinki, and the requirement for written informed consent was waived because of the retrospective design.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request, subject to institutional and ethical regulations regarding patient data confidentiality. The analysis code supporting this study, including analysis scripts, input specifications, environment requirements, and representative outputs, has been deposited on Zenodo and is publicly available at https://doi.org/10.5281/zenodo.19582389

Competing interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

OK: Conceptualisation, Data curation, Formal analysis, Software, Visualisation, Writing — original draft. EA, İÇ, AAD, HA: Investigation, Writing — review & editing. TK: Supervision, Methodology, Writing — review & editing. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Robert Iskander, D., Collins, M. J. & Davis, B. *Optimal Modeling of Corneal Surfaces with Zernike Polynomials*. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING* vol. 48 (2001).
2. Lombardo, M. & Lombardo, G. Wave aberration of human eyes and new descriptors of image optical quality and visual performance. *J. Cataract Refract. Surg.* **36**, 313–331 (2010).
3. Applegate, R. A., Ballentine, C., Gross, H., Sarver, E. J. & Sarver, C. A. *Visual Acuity as a Function of Zernike Mode and Level of Root Mean Square Error*. (2003).

4. Thibos LN, Applegate RA, Schwiegerling JT, Webb R; VSIA Standards Taskforce Members. Standards for reporting the optical aberrations of eyes. J Refract Surg. 2002;18(5):S652–S660. doi:10.3928/1081-597X-20020901-30.
5. Maeda, N. Clinical applications of wavefront aberrometry - A review. *Clinical and Experimental Ophthalmology* vol. 37 118–129 Preprint at https://doi.org/10.1111/j.1442-9071.2009.02005.x (2009).
6. Oliveira, C. M., Ferreira, A. & Franco, S. Wavefront analysis and Zernike polynomial decomposition for evaluation of corneal optical quality. *Journal of Cataract and Refractive Surgery* vol. 38 343–356 Preprint at https://doi.org/10.1016/j.jcrs.2011.11.016 (2012).
7. Bühren, J., Kühne, C. & Kohnen, T. Defining Subclinical Keratoconus Using Corneal First-Surface Higher-Order Aberrations. *Am. J. Ophthalmol.* **143**, (2007).
8. Marsack, J. D., Thibos, L. N. & Applegate, R. A. Metrics of optical quality derived from wave aberrations predict visual performance. *J. Vis.* **4**, 322–328 (2004).
9. Oie, Y. *et al.* Characteristics of ocular higher-order aberrations in patients with pellucid marginal corneal degeneration. *J. Cataract Refract. Surg.* **34**, 1928–1934 (2008).
10. CAMPBELL, C. E. A New Method for Describing the Aberrations of the Eye Using Zernike Polynomials. *Optometry and Vision Science* **80**, 79–83 (2003).
11. Kosaki, R. *et al.* Magnitude and Orientation of Zernike Terms in Patients with Keratoconus. *Investigative Opthalmology & Visual Science* **48**, 3062 (2007).
12. Charman, W. N. Wavefront technology: Past, present and future. *Contact Lens and Anterior Eye* **28**, 75–92 (2005).
13. Marcos, S., Barbero, S. & Jiménez-Alfaro, I. Optical Quality and Depth-of-field of Eyes Implanted With Spherical and Aspheric Intraocular Lenses. *Journal of Refractive Surgery* **21**, 223–235 (2005).
14. Mester, U., Dillinger, P. & Anterist, N. Impact of a modified optic design on visual function: Clinical comparative study. *J. Cataract Refract. Surg.* **29**, 652–660 (2003).
15. Chen, M. & Yoon, G. Posterior corneal aberrations and their compensation effects on anterior corneal aberrations in keratoconic eyes. *Invest. Ophthalmol. Vis. Sci.* **49**, 5645–5652 (2008).
16. Artal, P., Guirao, A., Berrio, E. & Williams, D. R. Compensation of corneal aberrations by the internal optics in the human eye. *J. Vis.* **1**, 1–8 (2001).
17. Elkady, B., Alió, J. L., Ortiz, D. & Montalbán, R. Corneal aberrations after microincision cataract surgery. *J. Cataract Refract. Surg.* **34**, 40–45 (2008).
18. Nakagawa, T. *et al.* Higher-order aberrations due to the posterior corneal surface in patients with keratoconus. *Invest. Ophthalmol. Vis. Sci.* **50**, 2660–2665 (2009).
19. Bühren, J., Kook, D., Yoon, G. & Kohnen, T. Detection of subclinical keratoconus by using corneal anterior and posterior surface aberrations and thickness spatial profiles. *Invest. Ophthalmol. Vis. Sci.* **51**, 3424–3432 (2010).

20. Salman, A. *et al.* Evaluation of Anterior and Posterior Corneal Higher Order Aberrations for the Detection of Keratoconus and Suspect Keratoconus. *Tomography* **8**, 2864–2873 (2022).
21. Belin, M. W., Villavicencio, O. F. & Ambrósio, R. R. Tomographic Parameters for the detection of keratoconus: Suggestions for screening and treatment parameters. *Eye and Contact Lens* vol. 40 326–330 Preprint at https://doi.org/10.1097/ICL.0000000000000077 (2014).
22. Alió, J. L. & Shabayek, M. H. Corneal higher order aberrations: A method to grade keratoconus. *Journal of Refractive Surgery* **22**, 539–545 (2006).
23. Belin, M. W. & Duncan, J. K. Keratoconus: The ABCD Grading System. *Klin. Monbl. Augenheilkd.* **233**, 701–707 (2016).
24. Villavicencio, O. F., Gilani, F., Henriquez, M. A., Izquierdo, L. & Ambrósio, R. R. Independent Population Validation of the Belin/Ambrósio Enhanced Ectasia Display: Implications for Keratoconus Studies and Screening. *Int. J. Keratoconus Ectatic Corneal Dis.* **3**, 1–8 (2014).
25. Gomes, J. A. P. *et al. Global Consensus on Keratoconus and Ectatic Diseases the Group of Panelists for the Global Delphi Panel of Keratoconus and Ectatic Diseases Background: Despite Extensive Knowledge Regarding the Diagnosis*. www.corneajrnl.com (2015).
26. McCormick, G. J., Porter, J., Cox, I. G. & MacRae, S. Higher-order aberrations in eyes with irregular corneas after laser refractive surgery. *Ophthalmology* **112**, 1699–1709 (2005).
27. Villa, C., Gutiérrez, R., Jiménez, J. R. & González-Méijome, J. M. Night vision disturbances after successful LASIK surgery. *British Journal of Ophthalmology* **91**, 1031–1037 (2007).
28. Chalita, M. R., Chavala, S., Xu, M. & Krueger, R. R. Wavefront analysis in post-LASIK eyes and its correlation with visual symptoms, refraction, and topography. *Ophthalmology* **111**, 447–453 (2004).
29. Hammoud, B., Wehbi, Z., Assaf, J. F., Roberts, C. J. & Awwad, S. T. From CLMI.X to CLMIX-AI: A Machine Learning–Based Upgrade of the Cone Location and Magnitude Index Expanded to Detect Keratoconus Suspects. *Ophthalmology Science* **5**, (2025).
30. Yousefı, S. *et al.* Keratoconus severity identification using unsupervised machine learning. *PLoS One* **13**, (2018).
31. Zéboulon, P., Debellemanière, G. & Gatinel, D. Unsupervised learning for large-scale corneal topography clustering. *Sci. Rep.* **10**, (2020).
32. Rokach, L. & Maimon, O. *DATA MINING WITH DECISION TREES*. http://www.worldscientific.com/series/smpai.
33. Kandakji, L. *et al.* Data-Driven Detection of Subclinical Keratoconus via Semi-Supervised Clustering of Multidimensional Corneal Biomarkers. *Ophthalmology Science* **6**, (2026).
34. Smolek, M. K. & Klyce, S. D. Zernike Polynomial Fitting Fails to Represent All Visually Significant Corneal Aberrations. *Invest. Ophthalmol. Vis. Sci.* **44**, 4676–4681 (2003).
35. Locantore, N. *et al. Robust Principal Component Analysis for Functional Data*. vol. 8 (1999).
36. Rodríguez, P., Navarro, R. & Rozema, J. J. Eigencorneas: Application of principal component analysis to corneal topography. *Ophthalmic and Physiological Optics* **34**, 667–677 (2014).

37. Shetty, R. *et al.* Keratoconus Screening Indices and Their Diagnostic Ability to Distinguish Normal From Ectatic Corneas. *Am. J. Ophthalmol.* **181**, 140–148 (2017).
38. Faria-Correia, F., Ramos, I., Lopes, B., Salomão, M. Q., Luz, A., Correa, R. O., Belin, M. W. & Ambrósio, R. Topometric and Tomographic Indices for the Diagnosis of Keratoconus. *Int. J. Keratoconus Ectatic Corneal Dis.* **1**, 92–99 (2012).

Figures

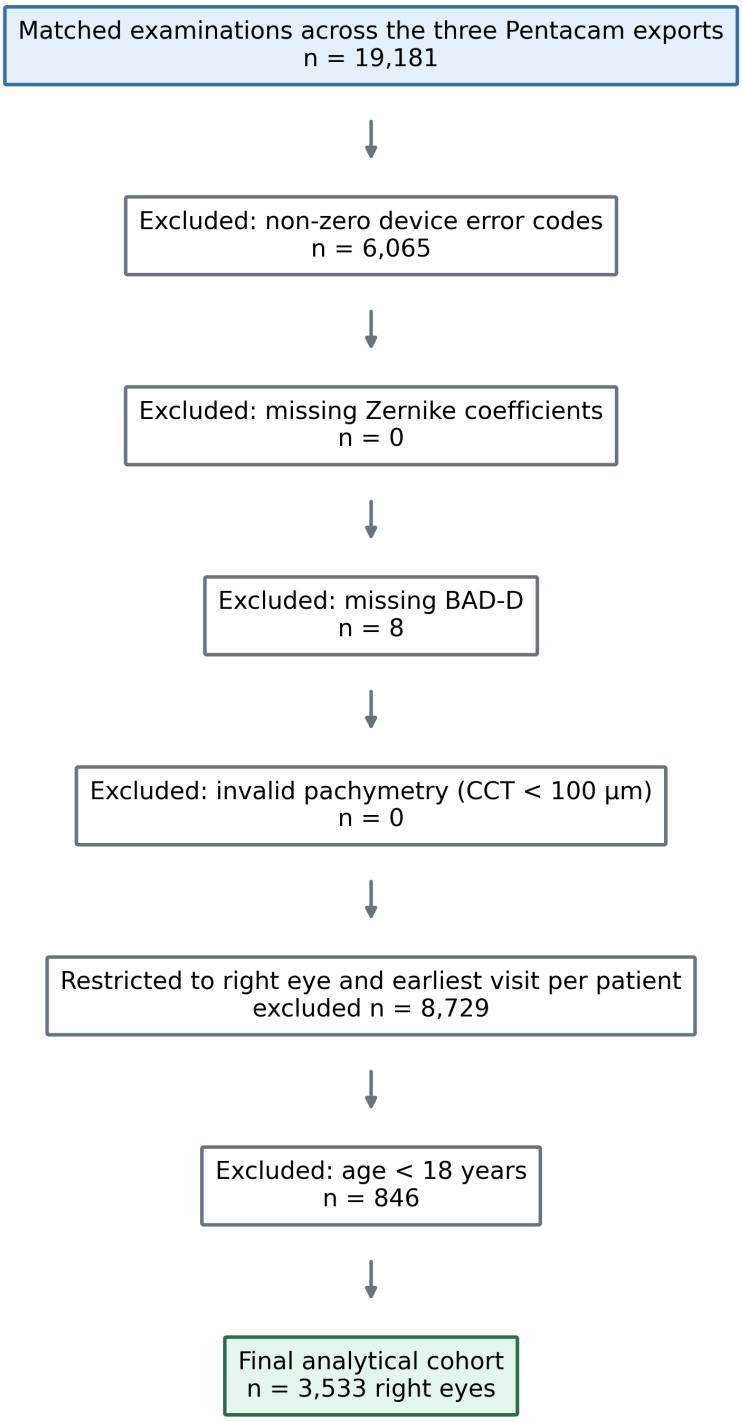


Figure 1

Cohort formation flowchart. A sequential exclusion pipeline leading from 19,181 matched Pentacam examinations to a final analytical cohort of 3,533 right eyes.

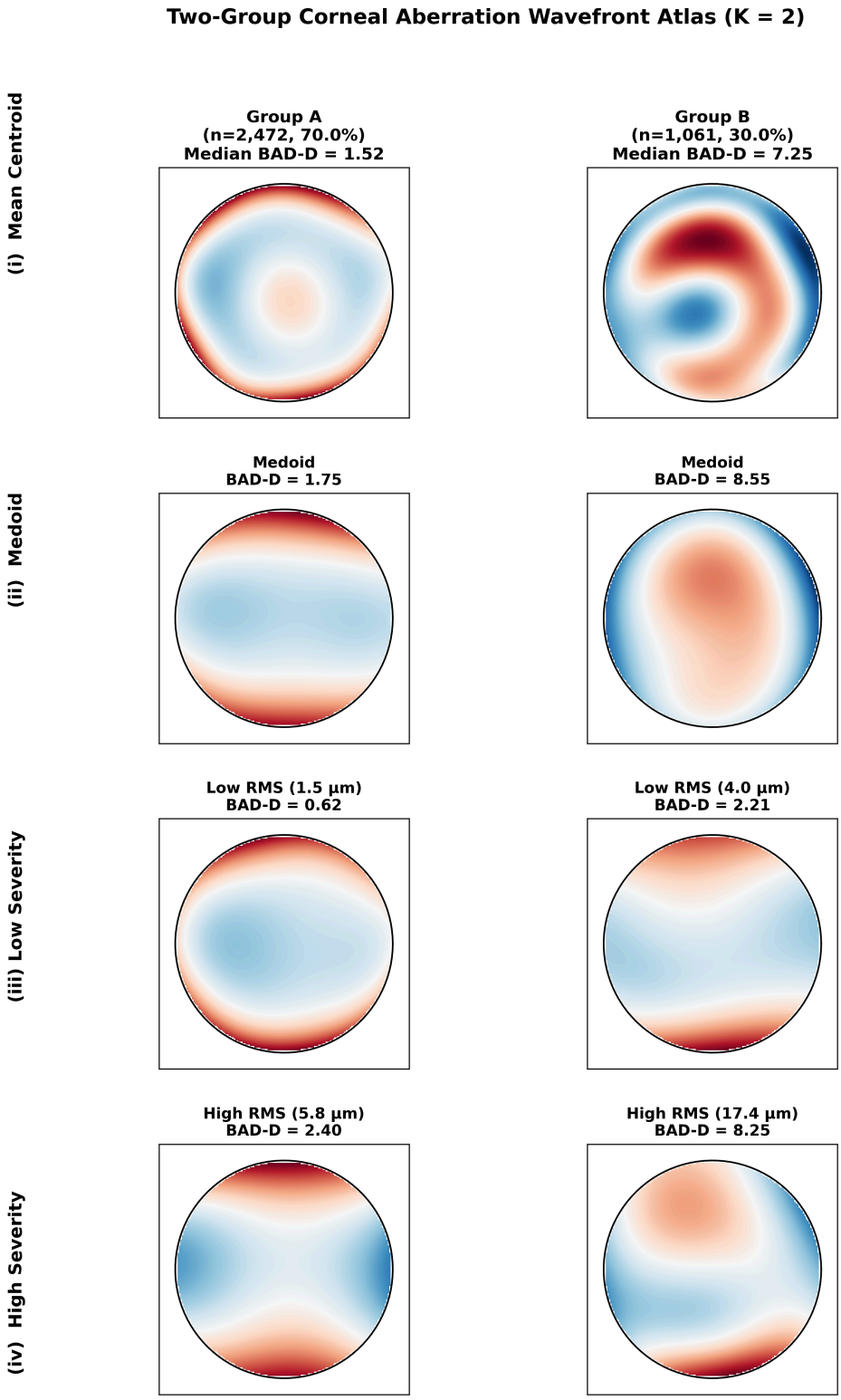


Figure 2

Two-cluster corneal aberration wavefront atlas (K = 2). Each column represents a cluster: Cluster A (n = 2,472, 70.0%; median BAD-D 1.52) and Cluster B (n = 1,061, 30.0%; median BAD-D 7.25). Row (i): centroid wavefront reconstructed from the cluster centre. Row (ii): medoid (the cluster member with the highest cosine similarity to the centre). Row (iii): low-RMS exemplar (approximately the 10th percentile of within-cluster total RMS). Row (iv): high-RMS exemplar (approximately the 90th percentile of within-cluster total

RMS). Colour scale: red–blue dichromatic map; warm colours indicate positive wavefront deviation, cool colours indicate negative wavefront deviation.

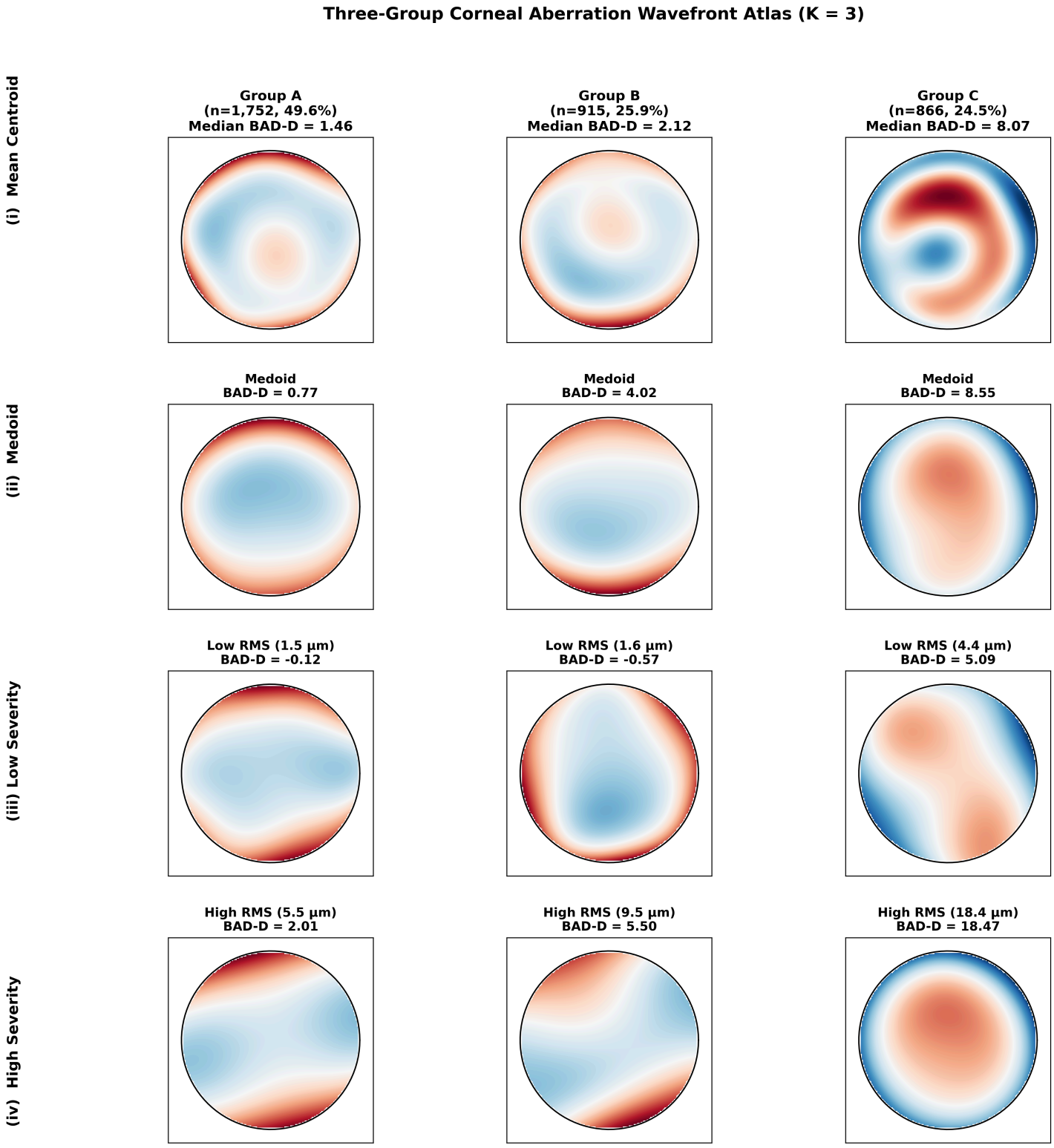


Figure 3

Three-cluster corneal aberration wavefront atlas (K = 3). Three columns: Cluster A (n = 1,752, 49.6%; median BAD-D 1.46), Cluster B (n = 915, 25.9%; median BAD-D 2.12) and Cluster C (n = 866, 24.5%;

median BAD-D 8.07). Row (i): centroid wavefront reconstructed from the cluster centre. Row (ii): medoid (the cluster member with the highest cosine similarity to the centre). Row (iii): low-RMS exemplar (approximately the 10th percentile of within-cluster total RMS). Row (iv): high-RMS exemplar (approximately the 90th percentile of within-cluster total RMS). Cluster B exhibits wavefront asymmetry and early inferior decentration. Cosine silhouette = 0.182; bootstrap stability = 0.921.

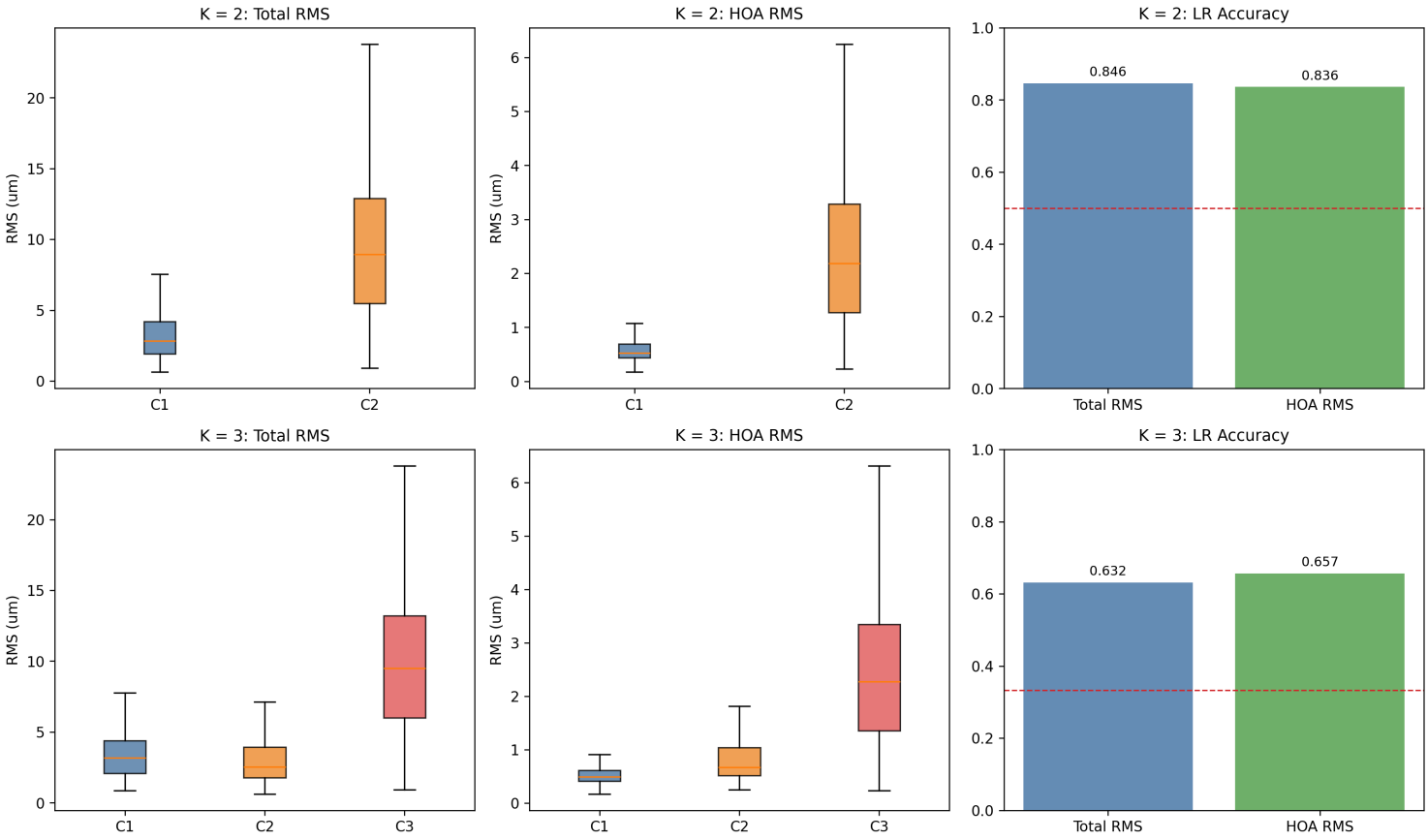


Figure 4

Relationship between discovery-set cluster assignment and scalar aberration magnitude (K = 2 and K = 3). Top row: two-cluster partitioning (K = 2). Bottom row: three-cluster partitioning (K = 3). Left panels: box plots of total corneal RMS (μm) by discovery-set cluster. Middle panels: box plots of HOA RMS (μm) by discovery-set cluster. Right panels: fivefold stratified cross-validated logistic-regression accuracy for predicting discovery-set cluster labels from total RMS or HOA RMS alone. Dashed red line = chance level. At K = 2, total RMS predicted cluster membership with 84.6% accuracy; at K = 3, accuracy fell to 63.2%.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfileslegends.docx
- Additionalfile1SupplementaryMaterialv3.docx