

Decision Tree Classification of Spatial Data Patterns From Videokeratography Using Zernike Polynomials*

M. D. Twa* S. Parthasarathy† T. W. Raasch*
M. A. Bullimore*

†Department of Computer and Information Sciences, Ohio State University

*College of Optometry, Ohio State University

Contact: srini@cis.ohio-state.edu,

Abstract

Topological spatial data can be useful for the classification and analysis of biomedical data. Neural networks have been used previously to make diagnostic classifications of corneal disease using summary statistics as network inputs. This approach neglects global shape features (used by clinicians when they make their diagnosis) and produces results that are difficult to interpret clinically. In this study we propose the use of Zernike polynomials to model the global shape of the cornea and use the polynomial coefficients as features for a decision tree classifier. We use this model to classify a sample of normal patients and patients with corneal distortion caused by keratoconus. Extensive experimental results, including a detailed study on enhancing model performance via adaptive boosting and bootstrap aggregation leads us to conclude that the proposed method can be highly accurate and a useful tool for clinicians. Moreover, the resulting model is easy to interpret using visual cues.

Keywords: Application Case Study, Clinical Data Mining, Spatial Data Analysis

1 Introduction

Storage and retrieval of biomedical information in electronic format has become increasingly common and complex. Data repositories have grown due to an increase in the number of records stored as well as a proliferation of the number of features collected. There is growing need for methods of extracting useful information from such databases. Due to the richness of the information stored in these databases, there are many potential uses of this data including the conduct of biomedical research, patient care decision support, and health care resource management.

Simple database queries can fail to address specific informational needs for several reasons: the query may not retrieve information needed because of user bias, lack of skill or experience, or limitations of the query software or database platform. In addition, this data often represents extremely complex relationships that

may escape even the most experienced content expert working with a highly competent database developer. As a result, many industries have looked to data mining as a general approach for the automated discovery of knowledge from these databases. Data mining has been widely applied in other domains such as fraud detection and marketing and is now found increasingly useful in a variety of health care database environments including insurance claim processing, electronic medical records, epidemiological surveillance, and drug utilization.

In the current study, we propose a new application of data mining and explore its use for clinical decision support and research applications. It is our belief that there are significant limitations when adopting existing approaches to mine biomedical clinical data. First, the data itself is more complex than a typical business transactional database. Second, information in these databases contain data with more elaborate relationships that are difficult to model accurately (e.g. structure-function relationships). Third, there is often a language barrier when clinicians attempt to apply data mining techniques. Problems range from selecting appropriate techniques and associated parameters, to interpretation of the resulting patterns.

In this study we focus on analysis of clinical data relating to keratoconus, a progressive non-inflammatory corneal disease that frequently leads to corneal transplantation. We describe a mathematical transformation of spatial data from clinical images, evaluate the usefulness of this method for pattern classification by decision trees, and explore further data reduction to optimize this approach for accurate and efficient data mining of spatial (image) data.

The main contributions of this study are:

- The use of a well known mathematical global model (Zernike Polynomials), and the experimental determination of the polynomial order required, to capture the underlying spherical harmonics of

*Support—MDT: NIH Grant T32-EY13359, American Optometric Foundation Ocular Sciences Ezell Fellowship; SP: Ameritech faculty fellowship; MAB: R01-EY12952.

corneal shape and use as a basis for classification. A key result of using the Zernike representation is a massive reduction in number of data points associated with a particular elevation map.

- An in-depth experimental evaluation of a decision tree algorithm with real patient data. We also evaluated the use of bagging and boosting and the implication of these strategies on the resulting model. We find that the resulting classification model has an overall accuracy of 91% on our dataset.
- In addition to the quantitative results, the data transformation we use can be easily visualized so that the resulting classification model is simple to interpret. This leads us to believe that this overall approach is quite promising.

The rest of this paper is organized as follows. In Section 2 we describe the relevant clinical background and the work that is closely related to this study. In Section 3 we describe the data transformation we use to represent the shape of the cornea, our data set, and how we applied decision tree decision tree induction in our experiments. Section 4 documents the experimental results we use to support our claims. Finally, we conclude and identify some possible avenues of future research in Section 5.

2 Clinical Background

The product of videokeratography is a high-resolution spatial map of corneal shape. Videokeratography, also known as corneal topography has become an essential tool for managing patients with corneal diseases that affect corneal shape, as well as corneal refractive surgery, and contact lens fitting [18]. The most common method of videokeratography is performed by measuring the distortion of a series of calibrated concentric rings reflected from the convex surface of the cornea. The image of these reflected rings is captured and corneal shape is mathematically derived from this captured image, see Figure 1. The derived corneal shape is displayed as a false color three-dimensional map.

The need to interpret and classify these color maps has led to several qualitative classification schemes to identify corneal surface features associated with ocular disease. While false color maps have improved qualitative data visualization, they do not provide a basis for quantitative methods of analysis. The first quantitative summaries of videokeratography were introduced in 1989 to describe surface asymmetry [4]. Alternative methods soon followed that continue to discriminate abnormal corneal shape based upon platform dependent

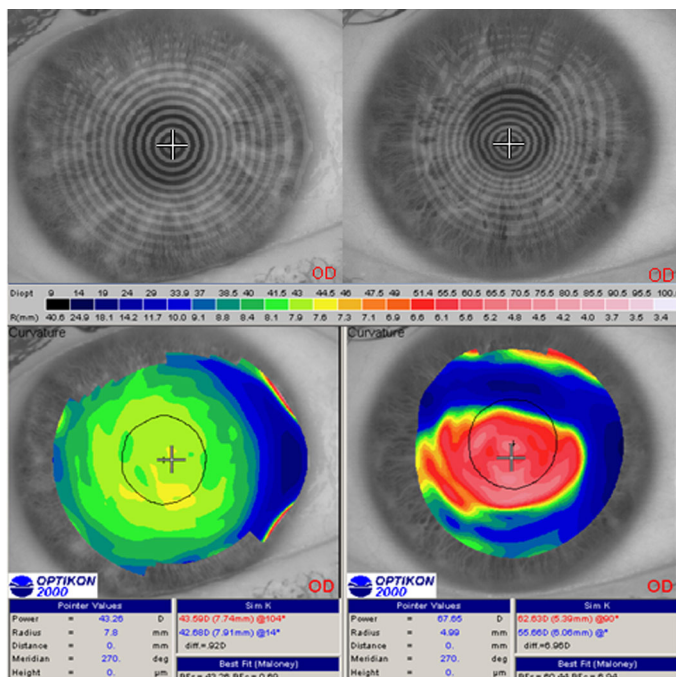


Figure 1: Videokeratography images. Top left, circular rings representing normal corneal shape. Top right, circular rings from cornea with keratoconus. Bottom left, false color elevation map representing normal cornea. Bottom right, false color elevation map representing cornea with keratoconus

indices [10, 26]. In their simplest form, these indices detect localized areas of increased curvature without regard to location, area, or other features that may be important to correct diagnostic distinctions. Maeda et al. constructed a more sophisticated classifier by combining multiple statistical summaries into a linear discriminant function [13]. Smolek et al. used multiple indices as input for a multilayer perceptron and compared this method with several other approaches [23]. Rabinowitz et al. describe an alternative classification method based upon a combination of four localized statistical indices (summaries) and report high accuracy for classification of keratoconus [17]. Published results for each of these methods demonstrate very good classification accuracy (85-95%) for their particular datasets, yet clinicians have been slow to embrace these techniques. Reasons for this limited acceptance include proprietary platform restrictions, poor interpretability, and questionable relevance with respect to other clinical data related to keratoconus.

In light of these results and limitations we seek to provide an alternative method of videokeratography classification that would provide high sensitivity and specificity for correct classification of keratoconus. Our design objectives were to: i) develop a classification model that was platform independent; ii) maintain high

sensitivity and specificity; iii) construct a classifier that was easily interpreted; and iv) develop a classification scheme that could be applied to individual or population data. We describe this alternative approach in the next section.

3 Overall Approach

Our approach to mining spatial patterns from videokeratography data was to transform our numeric instrument output to a Zernike polynomial representation of the corneal surface. We then used the coefficients from these orthogonal polynomials as inputs for a decision tree induction algorithm, C4.5 [14, 15].

3.1 Zernike Representation We selected Zernike polynomials to model the corneal shape because they can potentially provide a platform independent description of shape [25]. *Moreover they provide a precise mathematical model that captures global shape while preserving enough information for capturing local harmonics.* This is extremely important to detect keratoconus. Zernike polynomials possess several other attributes desirable for pattern recognition and have been applied across diverse domains including optical character recognition, optical encryption, and even tropical forest management[12, 9]. Also referred to as moments, Zernike polynomial coefficients each provide additional independent information to the reconstruction of an original object. We hypothesized that these orthogonal moments could be useful as descriptors of corneal shape and as attributes for induction of a decision tree classifier. Finally, the original dataset is in the form of an elevation map (typically 4000-7000 points), a lower order Zernike representation (say a 10th order one) requires very few coefficients (66) to represent this information. So it also has the additional advantage of data reduction (two orders of magnitude).

Zernike polynomials are an orthogonal series of basis functions normalized over a unit circle. These polynomials increase in complexity with increasing polynomial order[1]. In Figure 2, we present the basis functions that result from deconstruction of a 6th order fit to a corneal surface. Surface complexity increases by row (polynomial order), and radial location of the sinusoidal surface deviations become more peripheral away from the central column of Figure 2. Mean square error of the modeled surface is reduced as polynomial order is increased. An important attribute of the geometric representations of Zernike polynomials is that lower order polynomials approximate the global features of the anatomic shape of the eye very well, while the higher ordered polynomial terms capture local surface features. It is likely that the irregular sinusoidal surfaces of higher

ordered polynomial terms are not representative of normal corneal surfaces, however they may provide a useful basis for classification of normal and keratoconic corneal surfaces. A second important attribute of Zernike polynomials is that they have a direct relationship to optical function. These common optical and geometric properties make Zernike polynomials particularly useful for study of the cornea and the optical systems of the eye. A key aspect of our study is to try to identify the order and the number of Zernike terms required to faithfully represent keratoconic corneal surfaces and distinguish them from normal corneal surfaces.

Zernike polynomials consist of three elements[24]. The first element is a normalization coefficient. The second element is a radial polynomial component, and the third element is a sinusoidal angular component. The general form for Zernike polynomials is given by:

$$Z_n^{\pm m}(\rho, \theta) = \begin{cases} \sqrt{2(n+1)}R_n^m(\rho) \cos(m\theta), & \text{for } m > 0 \\ \sqrt{2(n+1)}R_n^m(\rho) \sin(|m|\theta), & \text{for } m < 0 \\ \sqrt{(n+1)}R_n^m(\rho), & \text{for } m = 0 \end{cases} \quad (3.1)$$

where n is the polynomial order and m represents sinusoidal frequency. The normalization coefficient is given by the square root term preceding the radial and sinusoidal components. The radial component of the Zernike polynomial, the second portion of the general formula, is given by:

$$R_n^m(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s! (\frac{n+|m|}{2} - s)! (\frac{n-|m|}{2} - s)!} \rho^{n-2s} \quad (3.2)$$

The final component is a frequency dependent sinusoidal term.

Following the method suggested by Schwiegerling et al. [21] we computed a matrix of Zernike height values for each pair of coordinates (ρ, θ) to a N -th order Zernike polynomial expansion using the equations above. We then performed a standard least-squares fit to the original data to derive the magnitude of the Zernike polynomial coefficients. Corneal shape is known to be mirror symmetric between right and left eyes and non-superimposable [22]. Thus, we mathematically transformed all left eyes to represent mirror symmetric right eyes by sign inversion of appropriate terms before analysis.

3.2 Classification using Decision Trees Decision tree induction has several advantages to other methods

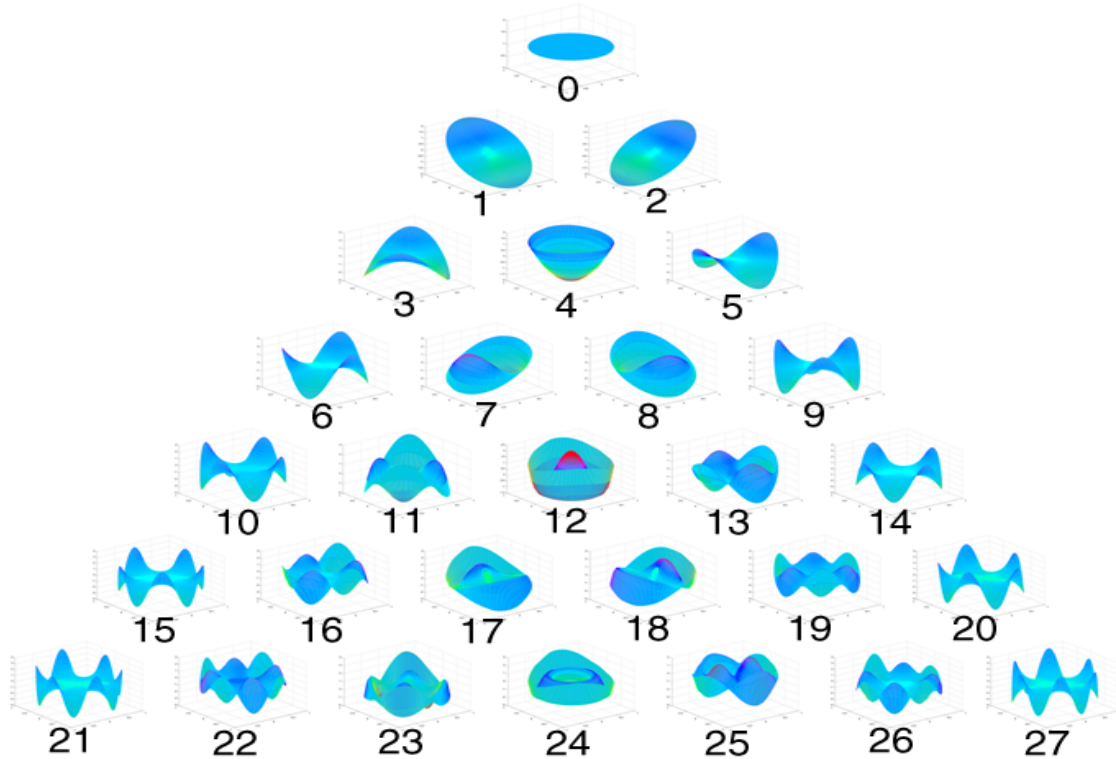


Figure 2: Geometric representation of 6th order Zernike polynomial series surface deconstruction plotted in polar coordinate space. Surfaces are labeled with single numeric indices using Optical Society of America conventions.

of classifier construction. Compared with neural networks, decision trees are much less computationally intensive, relevant attributes are selected automatically, and classifier output is a series of if-then conditional tests of the attributes that are easy to interpret. We selected C4.5 as the induction algorithm for our experiments [14, 15]. Freund, and Schapire describe the combination of bootstrap aggregation and adaptive boosting to the standard method of classifier induction with C4.5 [6, 7, 8, 19]. In principle, bootstrap aggregation is a compilation of standard classifiers constructed from separate samples drawn using bootstrap methods [5]. Adaptive boosting also results in a classifier constructed from multiple samples. Quinlan, and others have shown that each of these methods can enhance classification model performance with as few as ten iterations adding minimal computational expense[16]. Using the Zernike coefficients as input, we induced decision trees from our data to describe the structural differences in corneal shape between normal and keratoconic eyes.

3.3 Experimental Methodology and Dataset

Details Our classification data consisted of corneal height data from 182 eyes (82 normal control eyes, 100 keratoconic eyes) exported from a clinical videokeratog-

raphy system (Keratron, Optikon 2000, Rome, Italy). We fit data from the central 5 mm region to a N -th order Zernike polynomial using a least-squares fitting procedure described above. The resulting matrix of polynomial coefficients describes the weighted contribution of each orthogonal polynomial’s contribution to the overall shape of the original corneal surface. The videokeratography data files contained a three dimensional matrix of 7000 data points in polar coordinate space where height was specified by: $z = f(\rho, \theta)$ where height relative to the corneal apex is a function of radial distance, and angular deviation from the map origin. This corresponds to approximately 4100 data points for the central 5 mm radius.

3.3.1 Metrics One method of judging the performance of a classifier is to compare the accuracy of all classifications. This number is calculated by the following formula:

$$(3.3) \quad Accuracy = \frac{(TP) + (TN)}{TotalSample}$$

where TP is the total number of correct positive classifications and TN is the total number of correct negative classifications, or correct rejections. An other common

method of evaluating classifier performances is to look at the sensitivity of a classification model. Sensitivity is equivalent to the true positive rate, and is calculated as the number of true positive classifications divided by all positive classifications:

$$(3.4) \quad \textit{Sensitivity} = \frac{TP}{(TP) + (FN)}$$

Another common metric in biomedical literature is the specificity of a classification model. Specificity, also known as the correct rejection rate, is defined as the number of true negative classifications divided by all negative classifications

$$(3.5) \quad \textit{Specificity} = \frac{TN}{(TN) + (FP)}$$

As a classifier becomes more sensitive it will identify a greater proportion of true positive instances, however, the number of false negative classifications will consequently rise. Similarly, as a classification model becomes more specific, i.e. correctly rejecting greater proportion of true negative instances, then the number of false positive classifications will also rise. One way to visualize this relationship is to plot the true positive rate or sensitivity as a function of $(1 - \textit{specificity})$. This graphical illustration of classification accuracy is known as a receiver operator characteristic curve, or ROC curve. The diagonal line joining the origin $(0, 0)$ with the point $(1, 1)$ represents random classification performance. Classifiers with accuracy better than random chance are represented by a curve beginning at the origin and arcing above and away from the line representing random chance. The area beneath this curve describes the quality of classifier performance over a wide potential range of misclassification costs.

4 Experimental Results

We selected a Java implementation of C4.5 (Weka 3.2, The University of Waikato, Hamilton, New Zealand, <http://www.cs.waikato.ac.nz/~ml/>). This version of C4.5 implements Revision 8, the last public release of C4.5 [27]. We used the complete data set available for model construction and performed 10-fold stratified cross validation of the data for model selection and performance evaluation [11]. We evaluate the accuracy of our classifier as well as sensitivity and specificity of the model.

4.1 C4.5 Classification Results The purpose of our first experiment was to systematically explore the relationship between increasing polynomial order and classification accuracy. Our results show an increase

in model accuracy for our classifier as a function of the number of polynomial terms up to 7th order, see Figure 3 and Figure 4. Using a set of 36 Zernike polynomial terms to describe the seventh order fit to a corneal surface we were able to accurately classify 85% (154/182), and misclassified 15% (28/182) of instances in our data consisting of 15 false positive classifications as keratoconus and 13 false negatives. Similar results were observed for specificity, and sensitivity. Beyond 7th order, specificity remained constant while both accuracy and sensitivity were slightly worse.

We diagram the model generated from the 7th order pruned decision tree in Figure 6. This tree is relatively compact consisting of six rules with nine leaves, and is based upon 8 polynomial coefficients. This is pleasing since interpretability is one of the primary objectives of this study. Additionally, several of the attributes that form part of the decision tree (e.g. coefficients 4, 9, 14, and 24) have some consistency with clinical expectations of corneal shape in keratoconus. This is consistent with clinical experience and expectations of corneal shape with keratoconus.

4.2 C4.5 with Adaptive Boosting The adaptive boosting algorithm implemented in Weka is the Adaboost algorithm described by Freund and Schapire[6, 7, 8]. This was combined with the C4.5 induction algorithm in separate experiments as an attempt to enhance classifier performance. Boosting is an iterative method of model generation. Prior to our experiments, we evaluated the effect of increased boosting iterations from 1 to 50 on classifier performance to select the best number of boosting iterations. We found our error rate decayed to approximately 10% with as few as 10 iterations of boosting. Additional iterations did help slightly. However, results became erratic with increasing iterations similar to findings published by Quinlan[16]. We chose 10 boosting iterations for our experiments.

As with the standard induction method we generated classifiers using increasing numbers of attributes from the ordered polynomial coefficients. With adaptive boosting, we found improved classifier performance with fewer polynomial coefficients. Performance again degenerated beyond a 7th order surface representation. Best performance was achieved with a 4th order polynomial fit that resulted in correct classification of 91% of our instances and misclassification of 9% of our data instances consisting of 8 false positive classifications as keratoconus and 10 false negatives. We summarize performance statistics including accuracy, sensitivity, and specificity for this model also in Figure 3. Boosting provided approximately 10% accuracy improvement with fewer attributes. With seven orders of coefficients (36

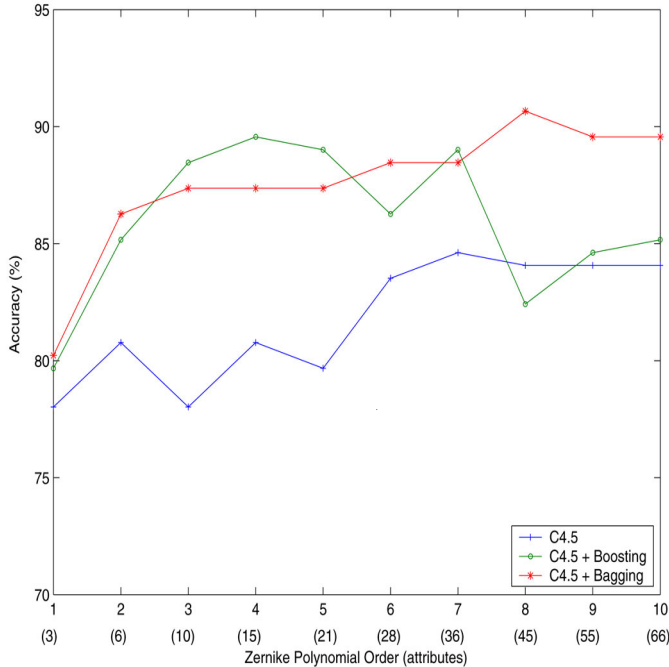


Figure 4: Accuracy comparison of C4.5 classification models with and without enhancements

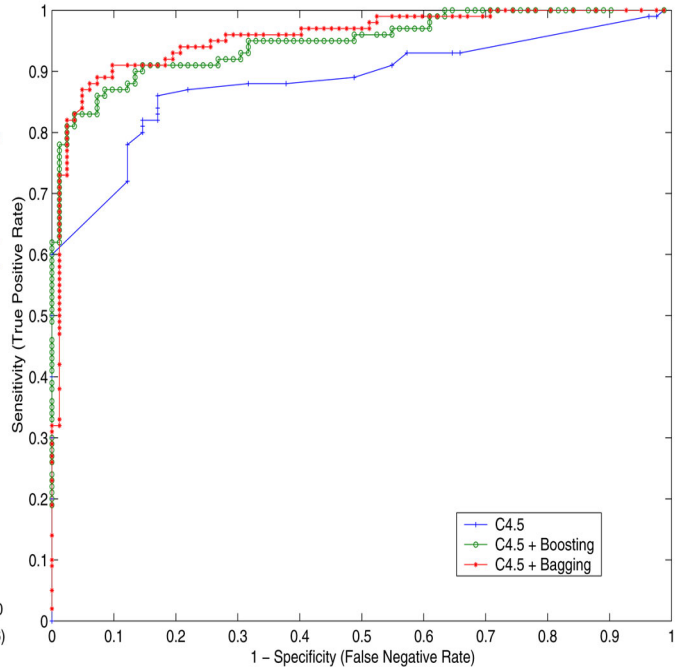


Figure 5: ROC comparison of C4.5 classification models with and without enhancements

attributes) the difference in accuracy with boosting was only 4% better than the standard classifier.

4.3 C4.5 with Bootstrap Aggregation Bootstrap aggregation, or bagging as described by Breiman, was applied in combination with the C4.5 induction algorithm[2]. As with boosting, we found that 10 bagging iterations provided nearly maximum improvements in classifier performance. Note that stratified ten-fold cross validation was applied in combination with these techniques to evaluate classifier performance.

Bagging provided the greatest improvement in classifier performance. The maximum accuracy resulted from eight orders of polynomial coefficients, forty-five attributes. The maximal performance values are highlighted in bold type in Figure 3, with accuracy of 91%, sensitivity of 87% and specificity of 95%. In Figure 4, we show a comparison of the standard classifier to the standard combined with bagging or boosting. The results show that the bagging was a useful enhancement to the standard classifier providing nearly 10% improvement in all model performance parameters evaluated as compared to the standard classifier. In Figure 5, we present the ROC curves for the best classifiers induced by each method-standard model, standard with bagging, and standard with boosting. There is mini-

mal difference between the models induced with boosting and bagging when compared by this method. The standard C4.5 classification model is less accurate than either enhanced model.

Bagging and boosting both produced considerable improvement compared to the standard classifier induction by C4.5. However, bagging was more stable compared to boosting performance, which suffered a loss of almost ten percent in accuracy as the number of polynomial attributes increased between seventh and eighth order. Best overall model performance was achieved using ten iterations of bagging in combination with C4.5.

The instability of boosting observed in these experiments is similar to those observed by others [16, 7]. This previously observed instability of boosting observed by others remains largely unexplained although there is some empirical evidence that suggests this may be a result of noise in the data. In theory, boosting performs best on data sets that are not significantly perturbed by subsampling. Boosting works by re-weighting misclassified instances forcing the algorithm to concentrate on marginal instances. With a small number of samples, this could overweight marginal instances and contribute to erratic classifier performance. By comparison, bagging does not reweigh instances, and is known to perform well with the datasets that are perturbed

C4.5	Polynomial order (attributes)									
	1 (3)	2 (6)	3 (10)	4 (15)	5 (21)	6 (28)	7 (36)	8 (45)	9 (55)	10 (66)
Performance (%)										
Accuracy	78	81	78	81	80	84	85	84	84	84
Sensitivity	74	83	76	84	86	86	87	86	86	86
Specificity	83	78	80	77	72	80	82	82	82	82
Precision	84	82	83	82	79	84	85	85	85	85
C4.5 + Adaptive Boosting										
Accuracy	80	85	88	90	89	86	89	82	85	85
Sensitivity	80	85	88	88	89	85	89	83	86	84
Specificity	79	85	89	91	89	88	89	82	83	87
Precision	82	88	91	93	91	89	91	85	86	88
C4.5 + Bootstrap Aggregation										
Accuracy	80	86	87	87	87	88	88	91	90	90
Sensitivity	79	85	86	86	85	85	84	87	86	86
Specificity	82	88	89	89	90	93	94	95	94	94
Precision	84	89	91	91	91	93	94	96	95	95

Figure 3: Performance estimation for different models. Number of coefficients associated with each order is given in parenthesis. Highest values are noted in **bold type**

significantly by subsampling.

4.4 Qualitative Results The results with bagging and boosting are important. Based on the above results we surmise that a 4th order polynomial is capable of discriminating between normal and abnormal corneas when adaptive boosting is added to the standard classifier. Additional improvement may be obtained by adding selected (not all) higher order polynomial terms (up to 7th order). When the standard classifier is combined with bagging, eight orders of coefficients are required to induce the best classifier and this gave the highest sensitivity and specificity of any model.

We believe that many of the higher order terms contribute noise to the model. The intuition for this is based on the results we observe and the fact that it is strongly believed that boosting is highly susceptible to noise[19, 3]. We believe that a 4th order representation has less overall noise (i.e. most all the terms contribute to the classification task). The improved results with a 7th order representation (without boosting) indicates that some terms at higher orders may improve the classifier performance further. This issue is part of ongoing and future work.

The resulting classification model is simple, and interpretable. More than 82% (82/100) of the total num-

ber of keratoconic eyes in this sample could be identified from the values of 4 coefficients. More than 90% of keratoconus cases could be correctly classified using the values of two additional coefficients. Interestingly, two rules (a subset of the total number) correctly classified 83% of the normal instances.

Decision trees avoid either bias by automating relevant feature selection based upon the quantitative separability of the training instances. In addition, the Zernike polynomial coefficients that became splitting attributes in the decision tree model appear to have strong empirical validity since the shape represented by the polynomial functions $j = 4, 5, 9, 14, 16$ and 24 (see Figures 2 and 6) correspond with either a greater parabolic shape or an inferior temporal shape irregularity. Several researchers and clinical practitioners have known about the relationship between keratoconus and features describing increased parabolic shape, oblique toricity, and high frequency peripheral irregularities. That the model automatically identified the particular Zernike coefficients that correspond to these features, and used them to discriminate diseased and normal eyes, is clear validation of the method. Moreover, unlike previous studies[20], our model has the advantage that since it uses coefficients that are associated with a central elevation, it would be more sensitive (and therefore

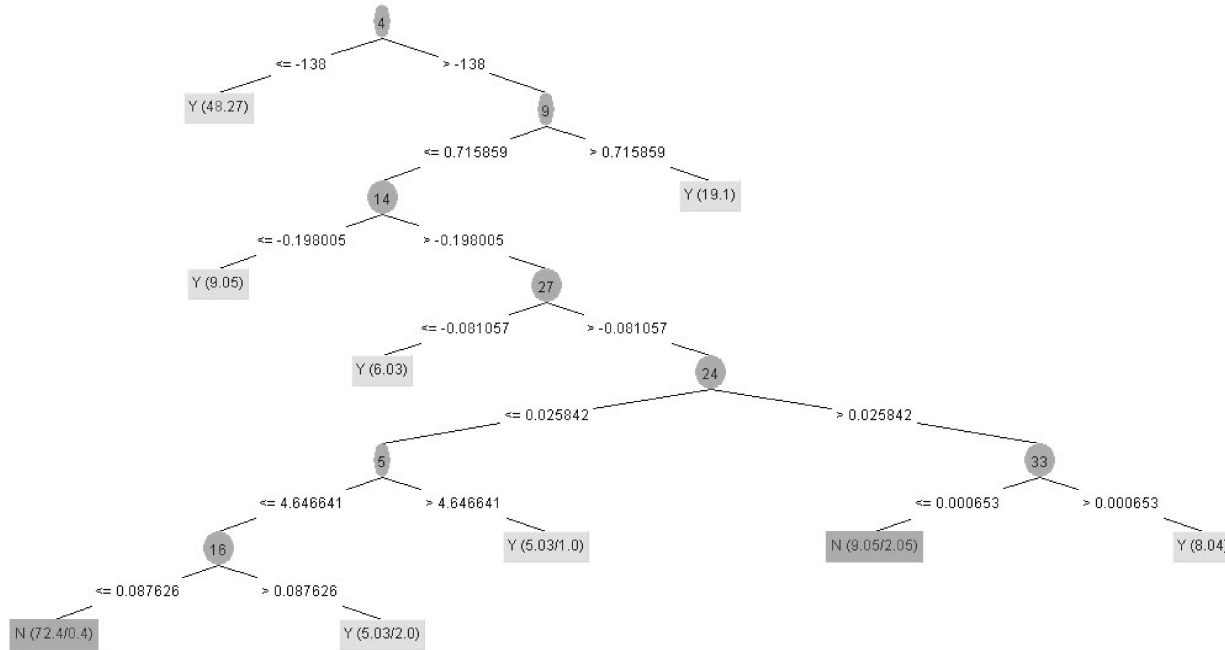


Figure 6: Decision Tree classification model from C4.5. Splitting attributes are labeled with Optical Society of America’s single index Zernike polynomial naming conventions. Leafs are labeled with classification and accuracy where Y = keratoconus, and N= normal; (percent correct / percent incorrect). This tree resulted from induction with 36 Zernike polynomial coefficients as attributes corresponding to a 7th Order Zernike polynomial surface representation.

be able to recognize) to centrally symmetric keratoconus which occurs frequently in patients.

Our attempt to reduce the dimensionality of data while maintaining interpretive ability using a Zernike polynomial data transformation was successful. By selecting the polynomial coefficients as a weighted representation of shape features, the initially large data set was reduced to 66 numeric attributes for a tenth order polynomial fit. Our results show that a subset of these polynomial coefficients could be enhanced classifier performance with further reduction in the dimensionality of the data representation.

A difference from this study and the work of others is the complexity and portability of the methods as well as the interpretability of the result. In addition, the ability to extend decision tree models to include other clinical data such as measures of visual function, corneal thickness, and even quality of life measures will permit classification and characterization of the severity of disease based upon other attributes. Note that direct comparisons between our results and those reported in other studies are difficult to interpret due to the differences in sample size, it’s sample selection procedures, dissimilarity of disease severity across studies, and other possible factors.

One limitation of the current study is the diame-

ter of data analyzed. While 5 mm is a reasonable size limit, it is possible that the area of analysis will influence feature selection in a significant way. Ideally, we conceive an automated feature selection method, which we may explore in the near future. Second, the sample size of this study is too limited to allow broad generalizations about the usefulness of this classification approach for all instances of keratoconus. Rather, we intend this study to show the benefits of spatial data transformation to medical image data and subsequent feature detection from the application of data mining techniques. We also believe that this approach to spatial data mining may prove useful for other domains such as cartography, optical inspection and other forms of biomedical image analysis.

5 Conclusions

In this study we propose the use of Zernike polynomials to model the global shape of the cornea and use the polynomial coefficients as features for decision tree classification of a sample of normal patients and patients with corneal distortion caused by keratoconus. Extensive experimental results, including a detailed study on enhancing model performance via adaptive boosting and bootstrap aggregation leads us to conclude that the proposed method can be highly accurate and a useful

tool for clinicians.

We feel there are a number of advantages to this approach. First, unlike neural networks decision trees allow some intuitive interpretation of the underlying structure to the data. Transforming videokeratography data with Zernike polynomials preserves important spatial information about corneal surface features, while utilizing objective and quantitative methods to recognize patterns from the data. Several of the lower order Zernike polynomial terms (i.e. 3-9) have inferred clinical significance. We are pleased to find that most of the coefficients identified as important for correct classifications by standard decision tree induction have some empirical clinical validity. Those that do not are interesting exceptions that may lead us to further exploration. Although boosting and bagging enhancements to the standard decision tree induction algorithm improved classification accuracy, the resulting model was less interpretable. We are currently working on improved visualization techniques for these more complex models.

We believe the approach presented in this study facilitates decision support, rather than decision making. Additionally, the method of analysis presented herein is portable and may be applied to data from virtually any clinical instrument platform. For example, the Ohio State University's Collaborative Longitudinal Evaluation of Keratoconus (CLEK), has years of data collected from over a thousand patients with multiple types of videokeratography instruments prior to standardization with a single instrument platform. Analysis of data from this study would benefit from our approach.

For vision scientists who may be interested in analysis of ocular surfaces this approach may yield new methods of analysis. Specifically, keratoconus research may benefit from such a quantitative method of analysis that would allow statistically based classification, and prediction. Additional data mining methods such as clustering and association rules could provide insights into the etiology, severity, and longitudinal progression of this disease, and these are directions we would like to explore and pursue.

References

- [1] M. Born and E. Wolf. Principles of optics : electromagnetic theory of propagation, interference and diffraction of light. In *7th expanded ed. Cambridge ; New York: Cambridge University Press*, 1999.
- [2] Leo Breiman. Bagging predictors. Technical Report 421, University of California at Berkeley, September 1994.
- [3] Thomas G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2):139–157, 2000.
- [4] S. A. Dingeldein, S. D. Klyce, and S. E. Wilson. Quantitative descriptors of corneal shape derived from computer- assisted analysis of photokeratographs. *Refract Corneal Surg*, 5(6):372–8., 1989.
- [5] Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap*. Monographs on statistics and applied probability ; 57. Chapman and Hall, Boca Raton, Fla., 1st crc press reprint. edition, 1998.
- [6] Y. Freund. Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2):256–285, 1995.
- [7] Y. Freund and R.E. Schapire. Decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.
- [8] Yoav Freund and Robert E. Schapire. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, 1999.
- [9] D.H. Hoekman and C. Varekamp. Observation of tropical rain forest trees by airborne high-resolution interferometric radar. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(3):584–594, 2001.
- [10] S. D. Klyce and S. E. Wilson. Methods of analysis of corneal topography. *Refract Corneal Surg*, 5(6):368–71., 1989.
- [11] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, pages 1137–1145, 1995.
- [12] U.V. Kulkarni, T.R. Sontakke, and G.D. Randale. Fuzzy hyperline segment neural network for rotation invariant handwritten character recognition. In *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on*, volume 4, pages 2918–2923 vol.4, 2001.
- [13] N. Maeda, S. D. Klyce, M. K. Smolek, and H. W. Thompson. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci*, 35(6):2749–57., 1994.
- [14] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.
- [15] J. R. Quinlan. *C4.5 : programs for machine learning*. The Morgan Kaufmann series in machine learning. Morgan Kaufmann Publishers, San Mateo, Calif., 1993.
- [16] J.R. Quinlan. Bagging, boosting, and c4.5. In *13th American Association for Artificial Intelligence National Conference on Artificial Intelligence*, pages 725–30, Portland, OR, 1996. AAAI/MIT Press.
- [17] Y. S. Rabinowitz and K. Rasheed. Kisaminimal topographic criteria for diagnosing keratoconus. *J Cataract Refract Surg*, 25(10):1327–35., 1999.
- [18] David J. Schanzlin and Jeffrey B. Robin. *Corneal topography : measuring and modifying the cornea*. Springer-Verlag, New York, 1992.
- [19] Robert E. Schapire. The boosting approach to machine learning: An overview. In *MSRI Workshop on*

Nonlinear Estimation and Classification, 2002.

- [20] J. Schwiegerling and J. E. Greivenkamp. Keratoconus detection based on videokeratoscopic height data. *Optom Vis Sci*, 73(12):721–8., 1996.
- [21] J. Schwiegerling, J. E. Greivenkamp, and J. M. Miller. Representation of videokeratoscopic height data with zernike polynomials. *J Opt Soc Am A*, 12(10):2105–13., 1995.
- [22] M. K. Smolek, S. D. Klyce, and E. J. Sarver. Inattention to nonsuperimposable midline symmetry causes wavefront analysis error. *Arch Ophthalmol*, 120(4):439–47., 2002.
- [23] M.K. Smolek and S.D. Klyce. Current keratoconus detection methods compared with a neural network approach. In *Investigative Ophthalmology Vis Sci 38(11):2290-9*, 1997.
- [24] Larry N. Thibos, Raymond A. Applegate, James T. Schwiegerling, and Robert Webb. Standards for reporting the optical aberrations of eyes. In *TOPS*, Santa Fe, New Mexico, 1999. OSA.
- [25] R.H. Webb. Zernike polynomial description of ophthalmic surfaces. In *Ophthalmic and visual optics topical Conference*, pp 38-41, 1992.
- [26] S. E. Wilson and S. D. Klyce. Quantitative descriptors of corneal topography. a clinical study. *Arch Ophthalmol*, 109(3):349–53., 1991.
- [27] I. H. Witten and Eibe Frank. *Data mining : practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann, San Francisco, CA, 2000.