# A Novel Classification and Estimation Approach for Detecting Keratoconus Disease with Intelligent Systems

Murat Ucar[1], Baha Sen[2], Hasan Basri Cakmak[3]

[1]Karabuk University, Faculty of Engineering, Department of Computer Engineering, Karabuk, Turkey
mucar@meb.gov.tr
[2]Yıldırım Beyazıt University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Ankara, Turkey
bsen@ybu.edu.tr
[2]Yıldırım Beyazıt University, Faculty of Medicine, Ankara, Turkey
hbcakmak@ybu.edu.tr

## Abstract

**Keratoconus is an eye disease characterized by progressive thinning of cornea which is the front based transparent layer of the eye. In other words, it is a progressive distortion of corneal layer and at least getting conical shape that should be like a dome camber. The vision reduces more and more while cornea gets shape of cone which should be like a sphere normally. The aim of this study is to define a new classification method for detecting keratoconus based on statistical analysis and to realize the prediction of these classified data with intelligent systems. 301 eyes of 159 patients and 394 eyes of 265 refractive surgery candidates as the control group have been used for this study. Factor analysis, one of the multivariate statistical techniques, has been mainly used to find more meaningful, easy to understand, and independent factors amongst the others. Later, a new classification method has been defined using clustering analysis techniques on these factors and finally estimated by using artificial neural networks and support vector machines.**

## 1. Introduction

The word “Keratoconus” was consisted of words “Kerato” meaning cornea and “Conus” meaning settlement in Latin. A detailed description of Keratoconus as an illness and difference from other corneal diseases characterized by ectasia firstly defined in 1854 [1].

Patients are acquainted with the disease on their 20’s which generally occurs in adolescence. It progresses between 20s and 40s and after 40s enters the stationary term. Nowadays keratoconus is observed in one out of every 2000 in western societies. The incidence of Keratoconus increases with each passing year. Keratoconus disease, which shows symptoms such as progressive myopia and astigmatism, thinning and protrusion of cornea, could be diagnosed in early stages by special surveys. From the beginning of disease there are remissions and recurrings, but, it is difficult to predict at what severity and how long it will continue. Mostly it is observed that the disease stabilized on the $4^{th}$ term [2, 3].

Early topographic modification of keratoconus was first defined by Amsler in 1938. Introduction of computer aided videokeratoscopy was a revolution in the diagnosis and follow up of keratoconus in the early 1980s [3].

In order to increase the efficiency of keratoconus diagnosis and classification, several quantitative parameters and indexes, which are golden standards on diagnosing and following up keratoconus, have been formed using corneal topography device and some assistive devices such as aberrometer and pachymeter. Rabinowitz was te first one to try to differentiate the normal cornea from the cornea affected by keratoconus using three values consist of central keratometry (Central K), gradient asymmetry in inferior and superior cornea (I-S) and corneal refractive power difference between the left and right eye [4].

Rabinowitz and Mc Donnel developed a method for the classification of keratoconus using central keratometry value I-S [5].

In a study performed by Li et al., it is reported that central K value was not a good prediction value for classification of keratoconus [6].

Studies related to the wavefront aberrations in keratoconus from past to present showed some significant differences between keratoconic corneas and normal corneas. These differences are utilized for diagnosis and classification of keratoconus. In many studies conducted so far it is observed that high order aberrations especially coma and coma like aberrations on eyes with keratoconus is higher than normal eyes [7,8].

Lagana, Cox, and Potvin claimed that total ocular aberration measurement on eyes with keratoconus using videoaberrescopy was a sensitive and reliable method on diagnosis and monitoring the progression of keratoconus [9].

Barbero et al. claimed that both videoaberrescopy and videokeratoscopy were useful on diagnosis of keratoconus [10].

Mahmoud et al. developed an index to use on keratoconus diagnosis that would be used on all corneal front face maps of many topography devices. It can be said that CLMI (The Cone Location and Magnitude Index) finds the steepest area on topography map and compare this area to the rest of the map whether the steepest area is a settlement or not. The size of CLMI represents the difference between the average gradient of steepest area and the average gradient of the rest of the map. The possibility of having keratoconus for the patient is calculated using the CLMI size gathered from axial map (PPK: Percent Probability Keratoconus). PPK values below 20% has been assumed as normal while PPK between 20% to 45% has been assumed as suspicious keratoconus, and PPK above 45% has been assumed as keratoconus in this study [11].

## 2. Methodology

The data used in this study was obtained from Atatürk Education and Research Hospital. The data was consisted of 301 eyes of 159 keratoconus patients and 394 eyes of 265 refractive surgery candidates as the control group. Beside general information of patients such as sex and age, other information for diagnosis of the disease also gathered such as refractive defects, visual acuity, Sim K1, Sim K2, average and central keratometry (Avg K, Central K), pachymetry values consist of thinnest corneal thicknesses, corneal wavefront aberrations, total high order aberrations (HOA), coma and coma-like aberrations, vertical coma (3-1), horizontal coma (3,1), spherical aberration (4,0), and CLMI magnitude obtained from axial map. The list of variables used in this study is shown in Table 1.

**Table 1.** The list of independent variables used in this study

| Variable Name | Data Type | Description |
|---|---|---|
| Age | Number | Age of the patients |
| SimKFlat | Number | Measurement of keratometry |
| SimKSteep | Number | |
| CentralK | Number | |
| CLMIAxis | Number | Measurement of CLMI(The Cone Location Magnitude Index) |
| CLMIfar | Number | |
| CLMIindex | Number | |
| Ho RMS | Number | Aberrations |
| (3,1) | Number | |
| (3,-1) | Number | |
| (4,0) | Number | |
| Coma | Number | |
| Pach | Number | Pachymetry values |
| Spher | Number | Level of refraction |
| Cylin | Number | |
| LogBc | Number | Level of vision |

### 2.1. Statistical Analysis

The main focus of this study is to define a new classification method for detecting keratoconus based on statistical analysis. Therefore, when deciding which variables would be used, factor analysis was used before clustering analysis.

Factor analysis is a common statistical technique that converts many interrelated variables to few, more meaningful and independent factors from each other. Factor analysis helps determining representative variables among large number of variables [12].

After the factor analysis, clustering analysis was used for classifying the data. Clustering analysis one of multivariate statistical technique, is used for classification of ungrouped data with unknown group number according to the similarity. Clustering analysis is a technique that provides collection of data in discrete clusters in terms of similarity of units or variables. Clustering analysis shows similarity with discriminant analysis in terms of aiming collection of similar individuals in the same groups and shows similarity with factor analysis in terms of aiming collection of similar variables and has data reduction feature [13].

### 2.2. Prediction Methods

The second aim of this study is to realize the prediction of these classified data with intelligent systems. Therefore, two popular prediction/classification methods are used (and compared to each other): artificial neural networks and support vector machines. These prediction methods are selected because of their superior capability of modelling classification type prediction problems. What follows is a brief description of these modelling techniques.

Artificial neural networks are commonly known as biologically inspired mathematical techniques, capable of modelling extremely complex non-linear functions [14]. In this study, we used a popular NN architecture called multi-layer perceptron (MLP) with back-propagation type supervised-learning algorithm. Multilayer perception is a method which uses monitored learning method and is usually used for academic or experimental classification purposes. MLP is capable of generating both classification and prediction models. It is used in estimation problems very often because of its' capability of learning the complex non-linear relationships between the input and output variables [15]. The ANN architecture used in this study is shown in Fig.1.

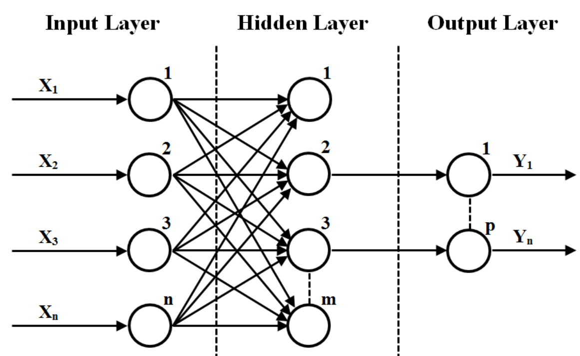


**Fig. 1.** A graphical representation of the ANN architecture used in this study

Support vector machine (SVM) is a novel learning machine introduced first by Vapnik [16]. Support vector machines (SVMs) belong to the family of generalized linear models which aims to achieve a prediction decision (classification or regression) based on a linear combination of features derived from the variables [17]. The aim of the SVMs for classification is to determine a hyper plane that optimally separates two classes. An optimum hyper plane is determined using train data sets and its generalization ability is verified using test data sets. SVMs have been successfully used in many fields [18,19,20].

## 3. Results

First of all, records with null values have been cleared in the application. Factor analysis has been applied on the variables using SPSS Statistics 20 and clustering analysis has been applied on variables of SimKFlat, CentralK, Coma, HoRMS, LogBc and CLMIindex which are most relevant to each other (Table 2).

**Table 2.** The results of the factor of the variables

| | Components | | | |
|---|---|---|---|---|
| | **1** | **2** | **3** | **4** |
| SimKsteep | -,177 | **,934** | -,039 | -,037 |
| Coma | **,900** | ,331 | -,015 | ,000 |
| Ho RMS | **,900** | ,317 | ,059 | -,028 |
| Central K | **,864** | -,317 | -,057 | -,089 |
| (3,-1) | -,860 | -,332 | **,188** | -,080 |
| SimKflat | **,836** | -,277 | ,065 | ,174 |
| LOGBC | **,711** | -,168 | ,314 | ,017 |
| CLMIindex | **,672** | ,489 | ,044 | -,070 |
| Cylin | -,566 | -,123 | ,206 | **,327** |
| Pach | -,509 | **-,038** | -,097 | -,275 |
| (4,0) | -,578 | **,701** | ,134 | ,137 |
| CLMIfar | -,424 | **,589** | ,526 | ,037 |
| CLMIaxis | ,051 | ,084 | -,766 | **,472** |
| Spher | -,494 | **,340** | -,571 | -,050 |
| (3,1) | ,211 | ,040 | ,052 | **,727** |
| Age | -,226 | -,233 | ,347 | **,461** |

Clustering analysis methods such as single linkage, complete linkage, average linkage, Wards and K-Means has been used in the study. As a result of these analyses, it has been observed in the application that grouping performed by Wards method is better and finally discriminant analysis has been used for determining the results of clustering analysis (Table 3). The number of cluster with the highest materiality has been decided according to Wilk's Lambda value which has been calculated by discriminant analysis that is performed according to different number of clusters.

**Table 3.** Clustering analysis results using Wards method

| **WARDS** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **2C** | 134 | 150 | | | | | | | | |
| **3C** | 120 | 150 | 14 | | | | | | | |
| **4C** | 55 | 150 | 65 | 14 | | | | | | |
| **5C** | 55 | 150 | 65 | 9 | 5 | | | | | |
| **6C** | 55 | 89 | 65 | 61 | 9 | 5 | | | | |
| **7C** | 55 | 89 | 64 | 61 | 9 | 5 | 1 | | | |
| **8C** | 52 | 89 | 64 | 61 | 9 | 5 | 3 | 1 | | |
| **9C** | 31 | 89 | 64 | 61 | 21 | 9 | 5 | 3 | 1 | |
| **10C** | 31 | 62 | 64 | 61 | 27 | 21 | 9 | 5 | 3 | 1 |

The lowest number of groups or clusters, in other words for the case that number of cluster is two, significance of discriminant function or say the number of clusters minus one has been checked. The same process has been repeated for three clusters or two discriminants function, four clusters or three discriminants function, five clusters or four discriminants function and six clusters or five discriminants function. As a result of discriminant analysis it is observed that the fifth function of six clusters was not significant (Table 4).

**Table 4.** Wilk's Lambda value related to six clusters

| **Wilks' Lambda** | | | | |
|---|---|---|---|---|
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Significance |
| 1 through 5 | ,031 | 962,199 | 30 | ,000 |
| 2 through 5 | ,345 | 295,052 | 20 | ,000 |
| 3 through 5 | ,669 | 111,184 | 12 | ,000 |
| 4 through 5 | ,904 | 28,105 | 6 | ,000 |
| 5 | ,983 | 4,873 | 2 | ,087 |

According to the significance result of six clusters, by the performed discriminant analysis it has been observed that using four discriminant functions for grouping is enough and it has been decided to put these functions into five groups. The analysis data consisting of 6 classes has been obtained by adding the data which belongs to non-patients to groups derived from the analysis of Wards.

In this study we used artificial neural networks and support vector machines for predicting. Software of neural network model and support vector machine model was developed in C# programming language in Visual Studio 2010 platform. Training results and the data used during training and testing are kept in ACCESS database.

As the output of the dataset consists of 6 classes, 15 SVM have been trained in the training process of support vector machines. Only two different classes of data have been taken in the each training phase. The most successful results have been achieved when linear function has been selected as the kernel function. Training and test phases have taken 1-1,5 seconds. The parameters and success rates as a result of training are as shown in Table 5.

**Table 5.** The error value and percentage of success according to the changed parameters (SVM)

| | Kernel Function | Multiple DVM | Percentage Of Success (%) | Train And Test Time |
|---|---|---|---|---|
| SVM | Polynomial | One to One | 96,591 | 1,388 |
| | Sigmoid | One to One | 64,773 | 1,513 |
| | Radial Basis | One to One | 94,886 | 1,435 |
| | Linear | One to One | 97,727 | 1,404 |

Training making with MLP is designed in an input layer, a hidden layer and an output layer and input layer has 6 neurons, hidden layer has 10 neurons and the output layer has 6 neurons. Back-propagation algorithm that has adjustable momentum and learning rate is used. At training phase, the best results are obtained when the learning rate is set to 0.2 and momentum is set to 0.07. Training process is completed in 1000 iterations.

Sigmoid activation function was chosen for all layers within the model. Training and testing phases are maintained approximately in 6-15 seconds. MLP model's error value and percentage of success according to the changed parameters are given in Table 6.

**Table 6.** The error value and percentage of success according to the changed parameters (MLP)

| | Learning Rate | Input Layer Activation function | 1th Hidden Layer Neuron | 1th Hidden Layer Activation Function | 2th Hidden Layer Neuron | 2th Hidden Layer Activation Function | Output Layer Activation Function | Error | Percentage Of Success (%) | Train And Test Time |
|---|---|---|---|---|---|---|---|---|---|---|
| MLP | 0,1 | Sigmoid | 20 | Sigmoid | - | --- | Sigmoid | 0,04 | 97,15 | 11,27 |
| | 0,2 | Sigmoid | 10 | Sigmoid | - | --- | Sigmoid | 0,03 | 98,29 | 6,28 |
| | 0,3 | Sigmoid | 10 | Sigmoid | 8 | Sigmoid | Sigmoid | 0,08 | 95,20 | 9,98 |
| | 0,4 | Sigmoid | 10 | Sigmoid | 8 | Sigmoid | Sigmoid | 0,08 | 94,31 | 9,42 |
| | 0,5 | Sigmoid | 10 | Sigmoid | 8 | Sigmoid. | Sigmoid | 0,07 | 95,38 | 9,07 |
| | 0,3 | Sigmoid | 30 | Sigmoid | - | --- | Sigmoid | 0,09 | 93,44 | 15,07 |
| | 0,4 | Sigmoid | 30 | Sigmoid | - | --- | Sigmoid | 0,07 | 95,45 | 15,24 |
| | 0,5 | Sigmoid | 30 | Sigmoid | - | --- | Sigmoid | 0,09 | 88,63 | 15,24 |
| | 0,2 | Tanjant | 30 | Tanjant | - | --- | Tanjant | 2,06 | 69,15 | 15,53 |

MLP and SVM methods success with respect to time chart is shown in Fig. 2.

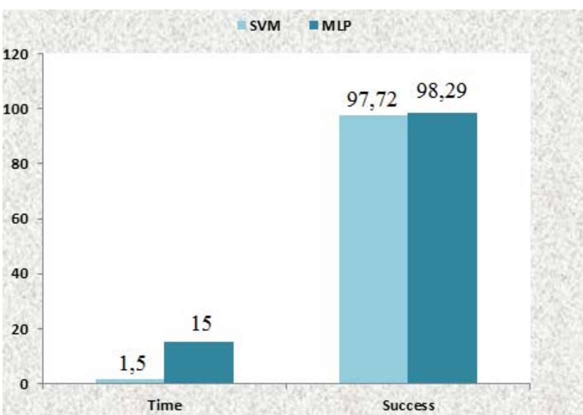


**Fig. 2.** MLP and SVM methods success with respect to time chart

## 4. Discussion and Conclusions

The findings have showed that sample size with 695 entries is enough and data is normally distributed. As a result of statistical analysis, data has been decided to be gathered in 5 groups.

Methods used for classification of keratoconus in previous works are not enough to classify the whole measured values. Because of increase and decrease of values, that determine the disease is not associated with each other, it has caused not to track the progress of disease well. Some researchers classified the disease into four groups by only looking at the keratometry values or by only eyesight while others classified the disease into two groups by only looking at the placido disk. In this study it is introduced that only looking at one variable is not enough for the classification of the disease. In addition, statistical data base classification has been used in this study to classify all of the measured values in order to follow up patient's progress in a healthier way.

Furthermore, looking at the accuracy rates of the prediction models used, the neural network model has produced the best result with a value of 98,29%. Neural network is followed by support vector machine models with a result value of 97,72%. But, when we look at the duration of training and testing phases, support vector machine exhibits better performance than neural network.

## 5. References


[1] Nottingham J. "Practical observations on conical cornea", *London: Churchill*, London;1984.pp.1-19.

[2] Rahmen W, Anwar S. "An unusual case of keratoconus", *J Pediatr Ophthalmol Strabismus*, 2006; 43:373-5.

[3] Rabinowitz YS. "Keratoconus*". Surv Ophthalmol*, 1998; 42: 297-319.

[4] Rabinowitz YS, Garbus J, McDonnell PJ. "Computer assisted corneal topography in family members of patients with keratoconus", *Arch Ophthalmol*, 1990; 108:365-371.

[5] Rabinowitz YS, McDonnell PJ. "Computer assisted corneal topography in keratoconus", *Refract Corneal Surg.* 1989;5: 400-8.

[6] Li X, Yang H, Rabinowitz YS. "Keratoconus: Classification scheme based on videokeratography and clinical signs*", J Cataract Refract Surg,* 2009;35(9):1597-603.

[7] Maeda N, Fujikado T, Kuroda T, Mihashi T, Hirohara Y, Nishida K, Tano Y. "Wavefront aberrations measured with Hartmann-Shack sensor in patients with keratoconus", *Ophthalmology* 2002;109(11):1996-2003.

[8] Alio JL, Shabayek MH. "Corneal higher order aberrations: a method to grade keratoconus", *J Refract Surg,* 2006;22:539-545.

[9] Lagana MA, Cox IG, Potvin RJ. "The effect of keratoconus on the wavefront aberration of the human eye", *Invest Ophthalmol Vis Sci* 2000;41(4).

[10] Barbero S, Marcos S, Merayo-Loves J, Moreno-Barriuso E. "Validation of the estimation of corneal aberrations from videokeratography in keratoconus", *J Refract Surg* 2002, 18:263-270.

[11] Mahmoud AM, Roberts CJ, Lembach RG, Twa MD, Herderick EE, McMahon TT. "CLMI:The cone location and magnitude index", *Cornea 2008*;27:480-7.

[12] Hair, J. F., Anderson, R. E., Tatham R.L., Black, W.C., "Multivariate Data Analysis", *Prentice Hall, New Jersey*, 1998, s:95-99.

[13] Çakmak, Z. (1999) "Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 1 (3), 187-205.

[14] Haykin, S. (2008). "Neural networks and learning machines (3rd ed.)", *New Jersey: Prentice Hall*.

[15] Hornik K, Stinchcombe M, White H. "Universal approximation of an unknown mapping and its derivatives using multilayer feed forward network", *Neural Networks* 1990;3: 359−366.

[16] V. Vapnik, "The Nature of Statistical Learning Theory", *Springer-Verlag*, New York, 1995.
[17] Pontil, M., Verri, A. (1998). "Support vector machines for 3D object recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 637–646.
[18] Turkan, M., Dulek, B., Onaran, I., Cetin, A.E., 2006. "Human face detection in video using edge projections", *In: Proceedings of SPIE-6246*, doi:10.1117/12.666704.
[19] Lau, K.W., Wu, Q.H., 2008. "Local prediction of non-linear time series using support vector regression", *Pattern Recognition, 41*, 1556–1564.
[20] Hong, J., Min, J., Cho, U., Cho, S., 2008. "Fingerprint classification using one-vs-all support vector machines dynamically ordered with naive Bayes classifiers", *Pattern Recognition 41*, 662–671.