

# Automated, computerized, feature-based phenotype analysis of slit lamp images of the mouse lens

Jenny Yuen<sup>a</sup>, Yi Li<sup>a</sup>, Linda G. Shapiro<sup>a</sup>, John I. Clark<sup>b,c,d,\*</sup>, Ernest Arnett<sup>b</sup>,  
E. Helene Sage<sup>b,c</sup>, James F. Brinkley<sup>a,b,e</sup>

<sup>a</sup> Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA

<sup>b</sup> Department of Biological Structure, University of Washington, Seattle WA 98195, USA

<sup>c</sup> Hope Heart Program, Benaroya Research Institute at Virginia Mason, 1201 9th Avenue, Seattle, WA 98101, USA

<sup>d</sup> Department of Ophthalmology, University of Washington, Seattle, WA 98195-2350, USA

<sup>e</sup> Department of Medical Education and Biomedical Informatics, University of Washington, Seattle, WA 98195-2350, USA

Received 19 July 2007; accepted in revised form 28 November 2007

Available online 4 March 2008

## Abstract

Longitudinal studies of a variety of transgenic mouse models for lens development can create substantial challenges in database management and analysis. We report a novel, automated, feature-based informatics approach to screening lens phenotypes in a large database of slit lamp images. Digital slit lamp images of normal and abnormal lenses in eyes of wild type (wt), SC1 null and SPARC null transgenic mice were recorded for quantitative evaluation of their structural phenotype. The images were processed to improve the contrast of structural features that corresponded to rings of opacity and fluctuations in scattering intensity in the lenses. Measurable attributes were assigned to the features in the lens images and given as an output vector of 46 dimensions. Characteristic patterns were correlated with the structural phenotype of each mutant and wt lens and a statistical fit for each phenotype was defined. The genotype was identified correctly in nearly 85% of the slit lamp images on the basis of an automated computer analysis of the lens structural phenotype. The automated computer algorithm has the potential to evaluate a large database of slit lamp images and distinguish mouse genotypes on the basis of lens phenotypes objectively using a neural network analysis of the structural features observed in the slit lamp images. The neural network approach is a promising technology for objective evaluation of genotype/phenotype relationships based on structural features and light scattering in lenses. Further improvements in the automated method can be expected to simplify and increase the accuracy and efficiency of the feature based analysis of structural phenotypes linked to genetic variation.

© 2008 Elsevier Ltd. All rights reserved.

**Keywords:** informatics; phenotype; genotype; slit lamp; imaging; lens

## 1. Introduction

Genetic variations in the eye are well known and often present structural phenotypes that can be mapped to specific, sometimes multiple, sites on chromosomes. In contrast, the ability to

evaluate variations in structural phenotype observed in a large database of images can be complicated by difficult and subtle differences in the structural appearance of the eye that can depend on exposure conditions, lighting and the angle of imaging. This report describes the results of an effective, automated method for the quantitative comparison of a large database of slit lamp images using a neural network approach to the computerized analysis of variations in structure of three different lens phenotypes and their correlated genotype.

The lens is an excellent model for development of an objective analysis of structural phenotype because the lens structure is highly ordered and readily observed and recorded in living

\* Corresponding author. Department of Biological Structure and Ophthalmology, University of Washington, Box 357420, Seattle, WA 98195-2350, USA. Tel.: +1 206 685 0950; fax: +1 206 543 1524.

E-mail addresses: [jenny@cs.washington.edu](mailto:jenny@cs.washington.edu) (J. Yuen), [yi@cs.washington.edu](mailto:yi@cs.washington.edu) (Y. Li), [shapiro@cs.washington.edu](mailto:shapiro@cs.washington.edu) (L.G. Shapiro), [clarkji@u.washington.edu](mailto:clarkji@u.washington.edu) (J.I. Clark), [earnett@u.washington.edu](mailto:earnett@u.washington.edu) (E. Arnett).

animals using a slit lamp. The lens is a sensitive indicator of hereditary and genetic disorders. A normal mammalian lens consists of symmetric layers of transparent cells containing a condensed protein solution that increases the refractive index to permit the lens to function as an optical element in the visual pathway. A complex network of special interactions between the concentrated cellular proteins favors short range, glass-like order and transparency across layers of lens cells (Benedek, 1983; Clark, 2004). Loss of transparency results when interactions between proteins are altered to produce inhomogeneities and large fluctuations in the index of refraction that scatter light. The opacity or cloudiness of the crystalline lens increases progressively with age (Harding, 1991; Kashima et al., 1993a; Thurston et al., 1997; Chylack and Khu, 2001; Thylefors et al., 2002). The spatial distribution and intensity of opacities within the lens varies with the molecular or genetic basis for the cataract (Phelps-Brown and Bron, 1996; Kojima et al., 2002; Fan et al., 2003; Seeberger et al., 2004).

Several methods for classification have been used effectively in clinical and epidemiological studies of cataract formation (Chylack et al., 1993a; Leske et al., 1996; Chylack and Khu, 2001; Klein et al., 2005) and it is widely recognized that specific patterns of opacification can be linked with hereditary and metabolic diseases including Alport's, Lowe's, Marfan's, and Down's syndromes, as well as myotonic dystrophy, coralliform cataract, and chromosomal mutations (Phelps-Brown and Bron, 1996; Alizadeh et al., 2004; Graw, 2004). A distinctive form of opacity in Alzheimer's patients was identified using slit lamp imaging, suggesting a relationship between structural phenotype and genotype (Goldstein et al., 2003). Mutant and transgenic mice have been linked to variations in the structural patterns of light scattering in lenses and in many cases, resemble human lens opacities (Clark and Carper, 1987; Rodriguez et al., 1992; Bettelheim et al., 1997; Graw, 1999; Norose et al., 2000; Tumminia et al., 2001; Yan et al., 2002). The success of the subjective classification of slit lamp images encourages the development of an automated computer based analysis of structural phenotypes that can be correlated with genetic mutations when the number of categories as well as the number of images increases in a very large database, as in a pre-clinical trial or genetic screen (Datiles et al., 1987; Chylack et al., 1995; Chylack and Khu, 2001; Seeberger et al., 2004; Wolf et al., 2005).

This report presents a novel method for computerized and quantitative evaluation of the progressive variations in lens structural features resulting from various mechanisms of opacification. No previous methods for the correlation of genotype with patterns of structural features in slit lamp images of mouse lenses have been reported. A Dynamic Light Scattering (DLS) fiber-optic probe that measured particle size distributions in animal lenses involves a different technology (Chenault et al., 2002; Datiles et al., 2002; Ansari et al., 2003; Simpanya et al., 2005) based on earlier quasielastic light scattering (QLS) instrumentation for lens opacification (Nishio et al., 1984; Benedek et al., 1987; Thurston et al., 1997). The DLS was used in the analysis of the distribution of particles of different sizes in the eyes of desert rodents, two groups of mice,

guinea pigs and in a rat model for cataract formation (Ansari et al., 2003; Simpanya et al., 2005; Chenault et al., 2002). The DLS method is not automated and is limited in its ability to consider the spatial distribution of the scattering features in lens with respect to phenotype.

The density of opacities in different areas of the eye lens (posterior capsular, posterior cortical, nuclear, anterior cortical, and anterior capsular) and in different age ranges has been studied (Edwards et al., 1990; Chylack et al., 1993b; Kashima et al., 1993b). The minimum number of views required was studied in patients to avoid subjecting patient eyes to the repeated and extended exposure of intense light used for camera recordings and to reduce the number of images stored. Light exposure and photographic storage were problems that could limit the number of images per patient. Current imaging methods are significantly advanced with regard to light sensitivity for image acquisition using digital video imaging, and short exposure time for patients. The advancement in modern compression technologies allows storage of more images with higher resolution than was possible ten years ago.

A number of classification methods are in use to define clinically relevant features obtained from densitometric analyses that are correlated with opacity but are not for general phenotype analysis (Dimock et al., 2000; Seeberger et al., 2004). Some of these studies were automated for cataract classification of nuclear sclerosis from slit-lamp photographs (Fan et al., 2003) and gray scale processing with predefined landmarks was useful for classification of clinically important cataracts. The purpose of previous studies was different from the current approach which concentrates on patterns of subtle variations in structural features resulting from genetic mutation of mouse lenses at the earliest possible age.

Two mouse models that have abnormal lens phenotypes on the 129Sv background, the SC1/Hevin null and the SPARC (osteonectin, BM40) null mouse, were selected for the evaluation using the phenotype analysis (Brekken and Sage, 2001; Sullivan and Sage, 2004; Sullivan et al., 2006). The selection of the mouse models was based on the slow temporal progression of the opacification pattern in lens with age which involves zones of discontinuity characteristic of ageing in many mammalian lenses. The patterns are subtle without distinct or unique structural features. SC1/Hevin and SPARC are matricellular proteins that modulate interactions between the lens basement membrane or capsule and the epithelial cells. The absence of SC1/Hevin or SPARC alters mobility of differentiating fibers at the basal attachments to the lens capsule which results in a progressive loss of transparency in layers of lens fibers (Bassuk et al., 1999; Norose et al., 2000). The microscopic structure responsible for the slit lamp images remains to be determined using direct, systematic histological methods (Kuszak and Costello, 2004; Kuszak et al., 2006; Blankenship et al., 2007). In the current study, an innovative computerized approach quantified collectively the patterns and features of the structural phenotype observed in images of transgenic mouse eyes. Structural features in the slit lamp image were assigned attributes that were concentrated into a vector of 46 dimensions for each image. The neural network assigned

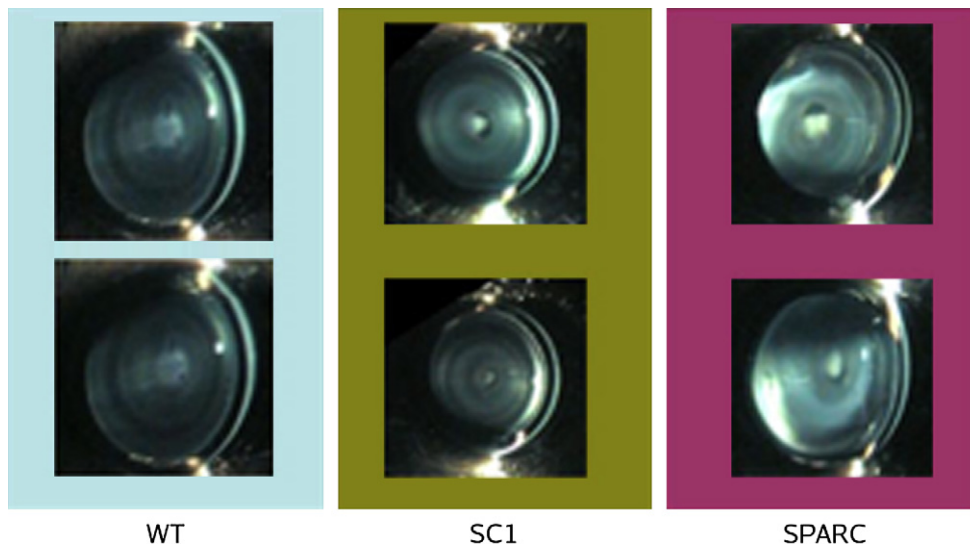


Fig. 1. Lens phenotype for: wild type (WT), synaptic cleft 1 protein null (SC1), and secreted protein acidic and rich in cysteine (SPARC) null.

each input image to a set representing its genotype and labeled it accordingly. On the basis of the structural features alone the algorithm separated a testing set of slit lamp images into their three genetic categories, WT, SC1 null or SPARC null, with a high degree of accuracy. The results provide baseline data for improved informatics approaches to phenotype/genotype relationships in living animals. In a broader study the molecular basis for each phenotype will need to be investigated.

## 2. Materials and methods

### 2.1. Slit-lamp eye image capture

We are given a set of  $k$  known cataract classes  $S = \{C_1, C_2, \dots, C_k\}$ , a training set of slit lamp mouse eye images,  $T_S$ , where each image  $t \in T_S$  is labeled as corresponding to a category  $C_i \in S$ , and a test image set  $T_T$  such that  $T_S \cap T_T = \emptyset$ . We want to assign  $C_i \in S$  to which each image  $Q \in T_T$  belongs. The wild type animals were the same back-crossed strain, 129Sv, as the mutants and the same animals were used as the wild type group for comparison with both mutant groups. The age range of the mice was 3 to 25 months. The transgenic mouse models were examined and imaged using a Nikon FS-2 slit lamp ophthalmoscope and a Canon Op-tura 20 digital video camera. Frames from each examination were selected from video recordings, using Adobe Premiere, saved in TIF file format, and then cropped and oriented using Adobe Photoshop. The total image database numbered over 100 different eyes in different mice. Duplicate images obtained under the same conditions were identical.

The set  $S$  of lenses used in this work is either: the WT (wild-type which has no laboratory-induced opacity), the Secreted Protein Acidic and Rich in Cysteine (SPARC) null (a matricellular protein), and the Synaptic Cleft (SC1) null.

$$S = \{\text{WT}, \text{SCI}, \text{SPARC}\}$$

In both the SPARC-null and SC1-null animals, the gene encoding the specific protein was truncated or replaced so that the functional protein is no longer expressed. Fig. 1 shows typical slit lamp images for each class. The original high-resolution images were resized to 300 by 300 pixel versions. The properties of the images in the datasets  $T_S$  and  $T_T$  are:

- The eyes in the images are approximately the same size and are approximately centered at the same location in each image.
- Each image is of a single eye in an individual mouse.
- The illumination during the image capture process varies among the different experiments (Fig. 2).
- There are artifacts caused by the illumination that are independent from the cataracts (Fig. 3).
- The pattern in the center of the lens is directly related to the cataract.
- A ring pattern can be observed in the WT class. Depending on the cataract, partial or total modification of this ring pattern can be observed.

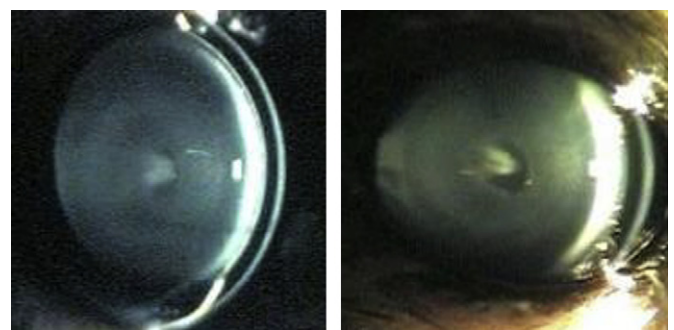


Fig. 2. The shape of the eye in an image varies depending on both the mouse and the angle of incidence of the slit-lamp light source on the surface of the eye. Illumination was not controlled.



Fig. 3. Artifacts caused by the imaging process are marked by yellow ellipses.

- Because of the manner in which the images are taken, the ring pattern is not circular but elliptical. This makes detection of the rings more difficult (Fig. 2).
- The images contain structures such as eyelashes, which are not of interest.

The opacity or cloudiness of the lens changes the perception of the layers in the lens. The pattern of the rings or the relative colors between rings may vary, making this a useful feature in the characterization of the pattern of opacity in a lens. The angle of incidence of the slit-lamp light on the cornea accounts for the elliptical shape of the ring pattern shown in the image. Fitting circular arcs to the rings in the images was a good approximation to the layers of opacification. To include the largest amount of information, images with ring patterns as close to a circle as possible were preferred. Noise factors in the image were non-trivial because control of the imaging conditions was limited in the initial testing. Constant illumination and shape were not assumed (Fig. 2).

The analysis involved feature extraction and learning. Quantifiable features characteristic of particular patterns were defined. A defined set of features was extracted from each slit lamp image  $t \in T_S$ , and these features were used to

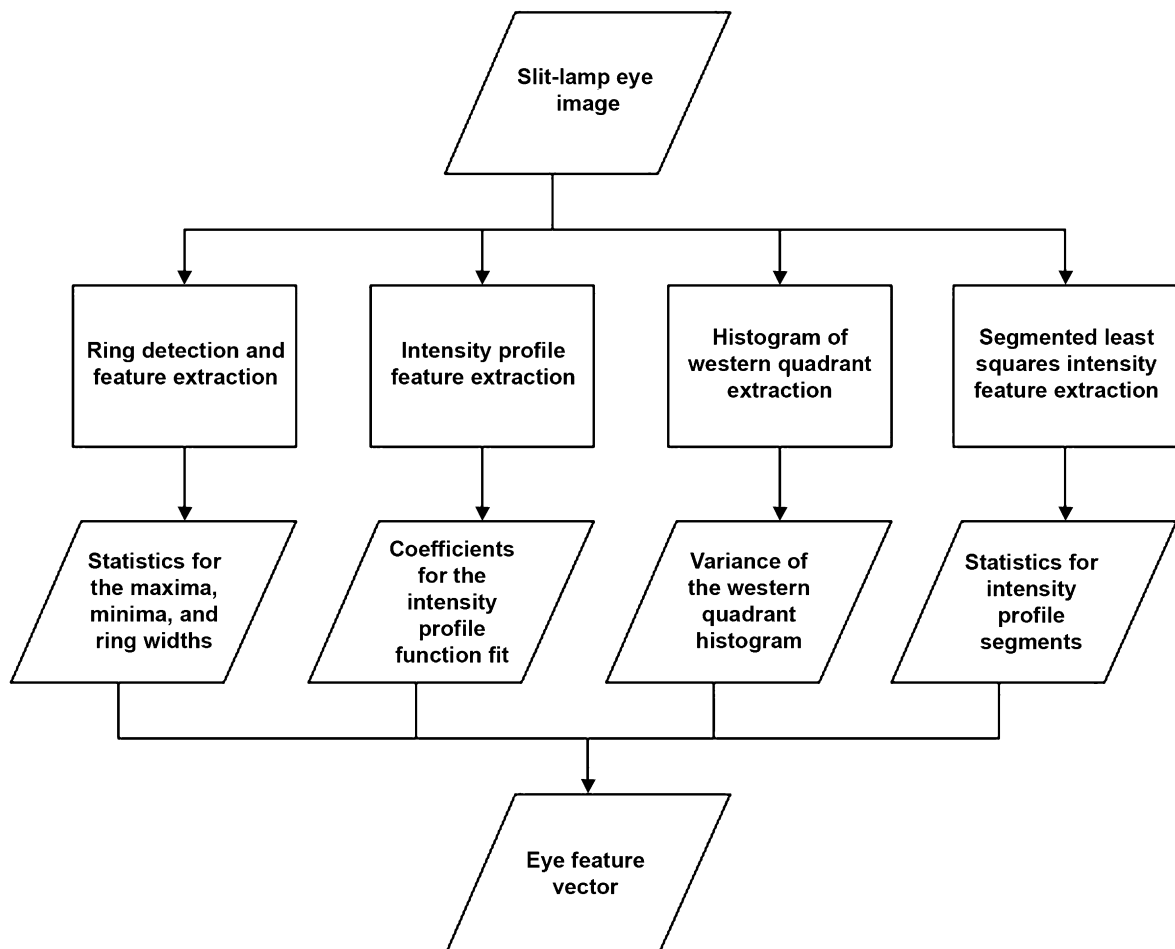


Fig. 4. Feature extraction process.



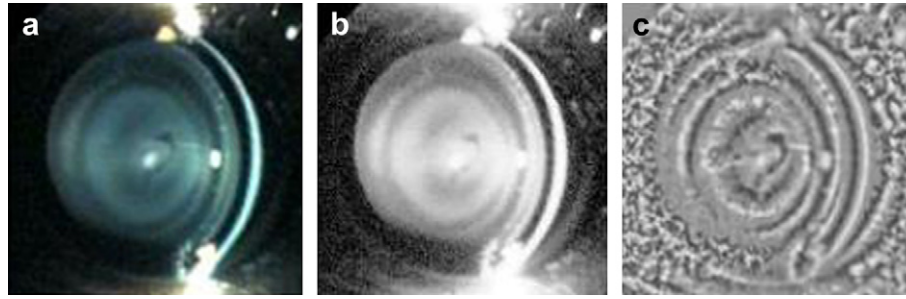


Fig. 5. The original image (a), the original image after a histogram equalization (b), and the contrast enhancement (c).

train a neural network (Russell, 1995). The training or learning process produced a model using the information from the training set. This model was used in the final analysis of each image.

## 2.2. Feature extraction

A set of image features with measurable attributes was developed to discriminate between the different known genotypes. The features were rings, intensity profile, and histogram. For each slit-lamp image input into the feature extraction process, a numeric feature vector of 46 dimensions was output. The process is summarized in Fig. 4.

## 2.3. Ring detection

Rings, or elliptical layers of varying transparency, are present in the WT, mutant SC1 null, and SPARC null lenses. Visualization of the layers of lens cells depends on magnification and contrast. In the WT lens, the contrast between adjacent cells or layers of cells is small when observed in a slit lamp biomicroscope. The contrast increases when

opacity occurs in some layers and not in adjacent layers. The increased contrast allows the rings to be observed in the slit-lamp. The pattern of rings can be correlated with the mutation. The proposed approach in identifying and quantifying these characteristics of the rings consists of a five-step process:

1. Ring enhancement
2. Isolation of an elliptical sector of the lens
3. Transformation of the elliptical sector into a rectangular image containing only pixels corresponding to the lens
4. Compression of the rectangular image representation into a 1-dimensional array of mean intensities
5. Extraction of feature vector values from the 1-dimensional array produced in the previous step

This process contributes 23 numerical values from the raw input image that will be used for the training phase.

### 2.3.1. Ring enhancement

As shown in Fig. 5, the contrast between the rings is small. A standard solution for this problem is histogram equalization transform, which modifies the intensity values of image pixels to achieve nearly equal numbers of the distinct intensity values; this tends to stretch out the gray tones and increase the contrast. Fig. 5b illustrates the histogram equalization transform applied to the image of Fig. 5a. Although there is some enhancement, the rings are not yet distinct. To solve this problem, a local histogram equalization transform that operates on local windows instead of the whole image is applied to the image. Each pixel value in the locally-equalized image is the result of applying histogram equalization to the sub-image formed by a window of fixed size centered at a pixel. The window size was 5 pixels squared. Fig. 5c shows the results of the local histogram equalization transform, which strongly enhanced the rings. The locally-equalized image was used for ring feature extraction.

### 2.3.2. Isolation of an elliptical sector and transformation to a rectangular image

Once the rings were enhanced, the parameters of the ring pattern were extracted. Let the coordinates of the center of the eye be  $(c_x, c_y)$ . Consider the region that corresponds to the elliptical sector centered at  $(c_x, c_y)$  and between the angles

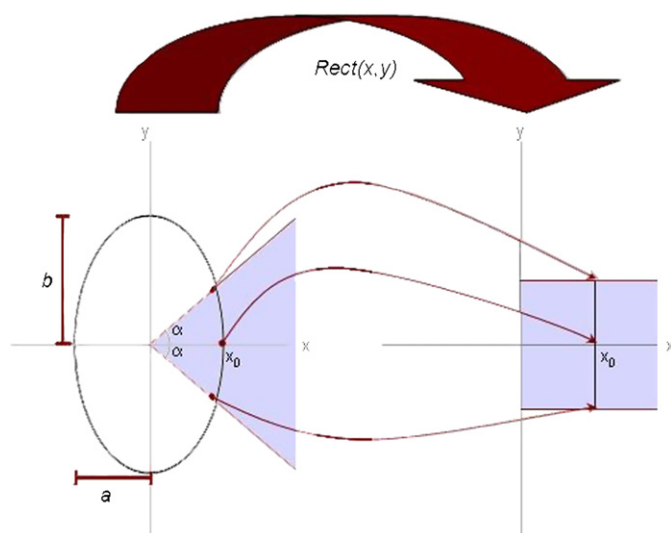


Fig. 6. The region inside an elliptical sector is mapped to a rectangular sector using the  $\text{Rect}(x, y) = (cx + a, y)$  transformation.

$\alpha$  and  $-\alpha$ , where  $0 < -\alpha < \pi/2$ . A ring was modeled as an elliptical arc with axes of length  $a$  and  $b$  parallel to the  $x$  and  $y$  axes respectively and centered at  $(c_x, c_y)$ . Every point  $(x, y)$  in this elliptical arc was mapped to a point on a vertical line by the transformation  $Rect(x, y) = (c_x + a, y)$  as shown in Fig. 6.

The  $Rect$  transformation was applied to every ellipse centered at  $(c_x, c_y)$  with a fixed  $c = a/b$  and  $0 < a < w - c_x$ , where  $w$  and  $h$  are the width and height of the image, respectively, to form a rectangular version of the elliptical eye. The resulting image of the  $Rect$  mapping was cropped to remove the cornea and the area outside the lens. The elliptical-to-rectangular transformation had three degrees of freedom:  $c_x$ ,  $c_y$ , and  $c$ . The value of  $c_y$  was restricted to  $h/2$ , working with two degrees of freedom for each image.

### 2.3.3. Compression of the rectangular eye representation to a 1-dimensional array of data

The  $Rect$  transformation converts the elliptical rings into vertical lines for easier and more accurate analysis. It is important to choose the correct center  $(c_x, c_y)$  and  $c$  value for the lens (Fig. 7). The previous step generates a large number of potential fits to the data. To evaluate the goodness of a chosen set of parameters, two vectors are constructed:

- a vector  $M(c_x, c)$  of values containing the mean intensity value for each column in the  $Rect(c_x, c)$  image, and
- a vector  $V(c_x, c)$  containing the variance of each column in  $Rect(c_x, c)$ .

The score associated with each pair of values  $(c_x, c)$  is defined as:

$$\text{EllipseFit}_{\text{score}}(c_x, c) = - \sum_{v \in V(c_x, c)} v$$

Using this score, a pair of  $(c_x, c)$  values that will be called  $(\text{best}_{c_x}, \text{best}_c)$  is chosen. This pair has the property that  $\forall c'_x, c'_y$  such that:

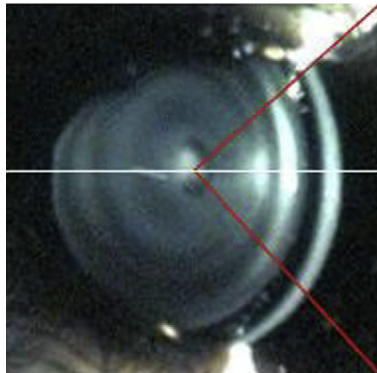


Fig. 7. An iterative process selected a center point, isolated an elliptical sector of the lens which was mapped to a rectangle. The variance at each column of the rectangular map was computed, and the case where the sum of the column variances was the lowest was selected as the best elliptical sector isolation; its center corresponded to the best center of the lens.

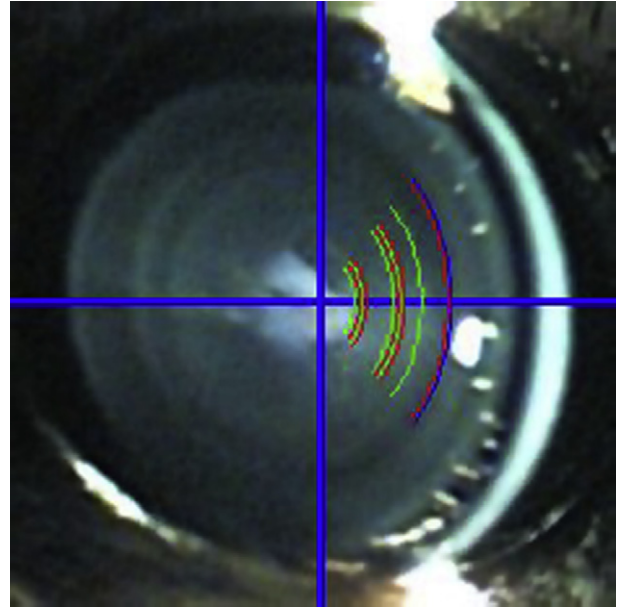


Fig. 8. The ring extraction process allowed detection of the lens center and characterization of the lens rings as observed in the fitted elliptical arcs.

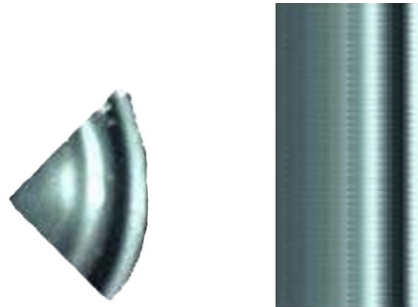
$$\left| \frac{w}{2} - c'_x \right| < \varepsilon$$

and

$$\left| \frac{h}{2} - c'_y \right| < \varepsilon$$

$$\text{EllipseFit}_{\text{score}}(\text{best}_{c_x}, \text{best}_{c_y}) = \text{EllipseFit}_{\text{score}}(c_x, c_y)$$

for some  $\varepsilon > 0$ . Note that only the vector of variances  $V$  is used for parameter evaluation. The vector of means  $M$  was a simpler way of characterizing the ring information by exploiting the property of homogeneous intensity of every ring. The result of the mapping by  $M$  will be used for the extraction of numerical features in the next step.



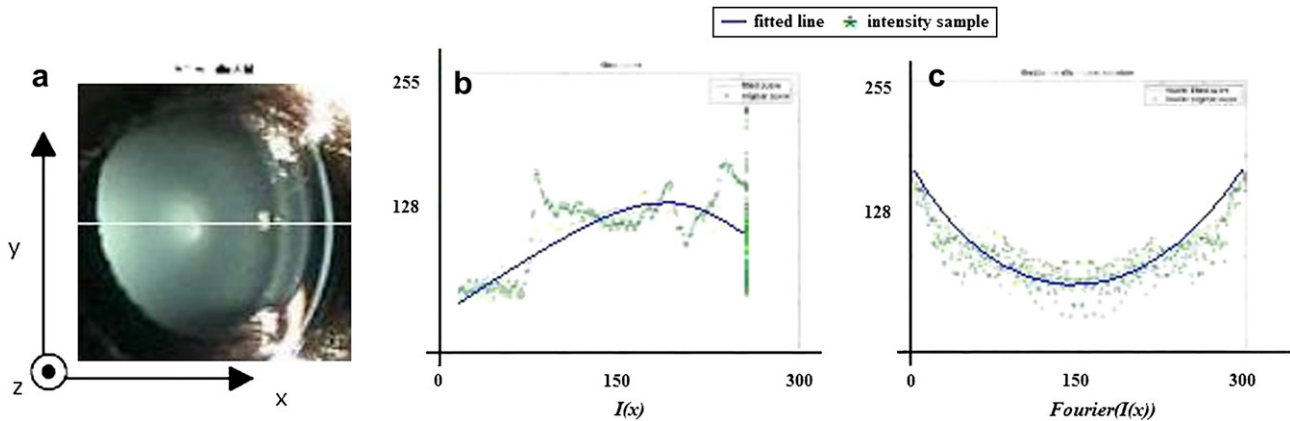


Fig. 9. The white line in the left image (a) indicates the row of pixels that was considered for the intensity profile feature. The middle image (b) shows a polynomial fit of degree 5 on the intensity profile. The right image (c) shows a polynomial fit on the output of the Fast Fourier Transform of the intensity profile.

#### 2.4. Ring feature extraction

The last transformation by  $M$  provided an easier method for identifying the boundaries of the rings, their average intensities, and other relationships that were useful features for the recognition of cataract patterns.

The mean vector  $M$  was analyzed to provide a feature vector of numeric attributes that was used for pattern recognition. The vector we used contained the following 23 ring attributes:

- number, maximum, minimum, mean, and variance of maxima of the function;
- number, maximum, minimum, mean, and variance of minima of the function;

- maximum, minimum, mean, and variance of the distance between pairs of consecutive maxima;
- maximum, minimum, mean, and variance of the distance between pairs of consecutive minima;
- maximum, minimum, mean, and variance of the difference between each pair of consecutive maxima and a minima;
- number of consecutive pairs of maxima and minima (regions formed by peaks and valleys).

These attributes captured several features of the rings. Globally, the attributes included the number of rings and the distribution of their properties such as mean intensity, width, and others. Locally, the attributes described the width and opacity of each ring and provided a comparison of these characteristics with the characteristics of other rings in the same image (Fig. 8).

#### 2.5. Intensity profile

A fast Fourier transform, FFT, was used to provide a parameter representing the spatial intensity fluctuations observed in a line scan across the diameter of slit lamp images of subtle opacities (Fig. 9a). Intensity profiles were normalized to the corneal opacity for each mouse being examined, and duplicate images from the same eyes appear the same. For the function  $I(x)$ , where the domain is each column in the image, and the range  $\{0, 255\}$  quantified the pixel values for intensity  $I(x)$ , a polynomial was a poor fit to the spatial fluctuations in the noisy intensity profile across the center of the slit lamp image (Fig. 9b). Instead, the entire intensity profile (Fig. 9b) was the input for a fast Fourier transform,  $Fourier(I(x))$ , which provided a power spectrum (Fig. 9c) containing the same information as the intensity profile in a broad, simplified reciprocal plot. The result is a much better polynomial fit. 12 coefficients of this polynomial (6 real and 6 imaginary) were concatenated to form a feature vector.

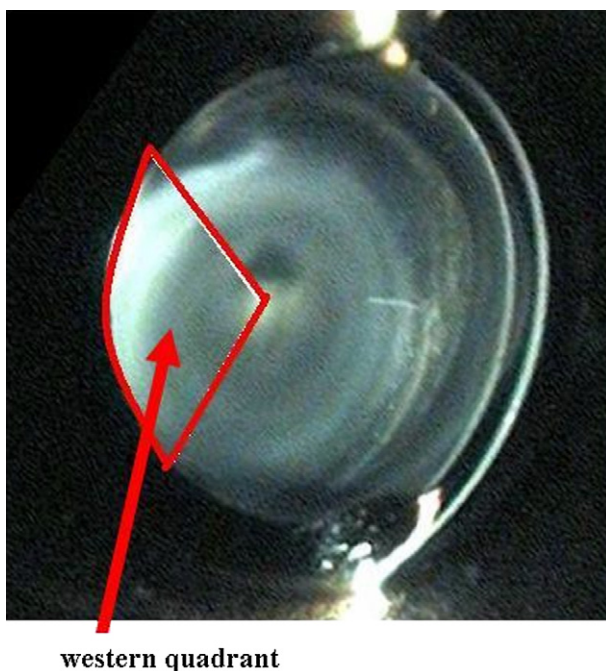


Fig. 10. Western quadrant of the lens.

## 2.6. Histogram of western quadrant of the lens

There was a characteristic pattern of rings for each lens class. A histogram of the western quadrant (Fig. 10) of the lens was created and was fit to a 1-dimensional Gaussian using Maximum Likelihood. The variance of the Gaussian was included in the feature vector. Because of the variations in illumination caused by external sources in the image, only the variance, which was related to the distribution of intensities, and not the mean, which was directly associated with the change in external illuminations, was considered. Fig. 11

shows the histogram of the images corresponding to each of the three categories of images.

## 2.7. Segmented least squares fitting on the intensity profile

While the intensity profile feature characterized some lenses, the function was assumed to be continuous and non-continuous fluctuations were approximated as smooth curves. Some classes such as SPARC null and WT lenses had similar polynomial fits but differed in the number of non-continuous

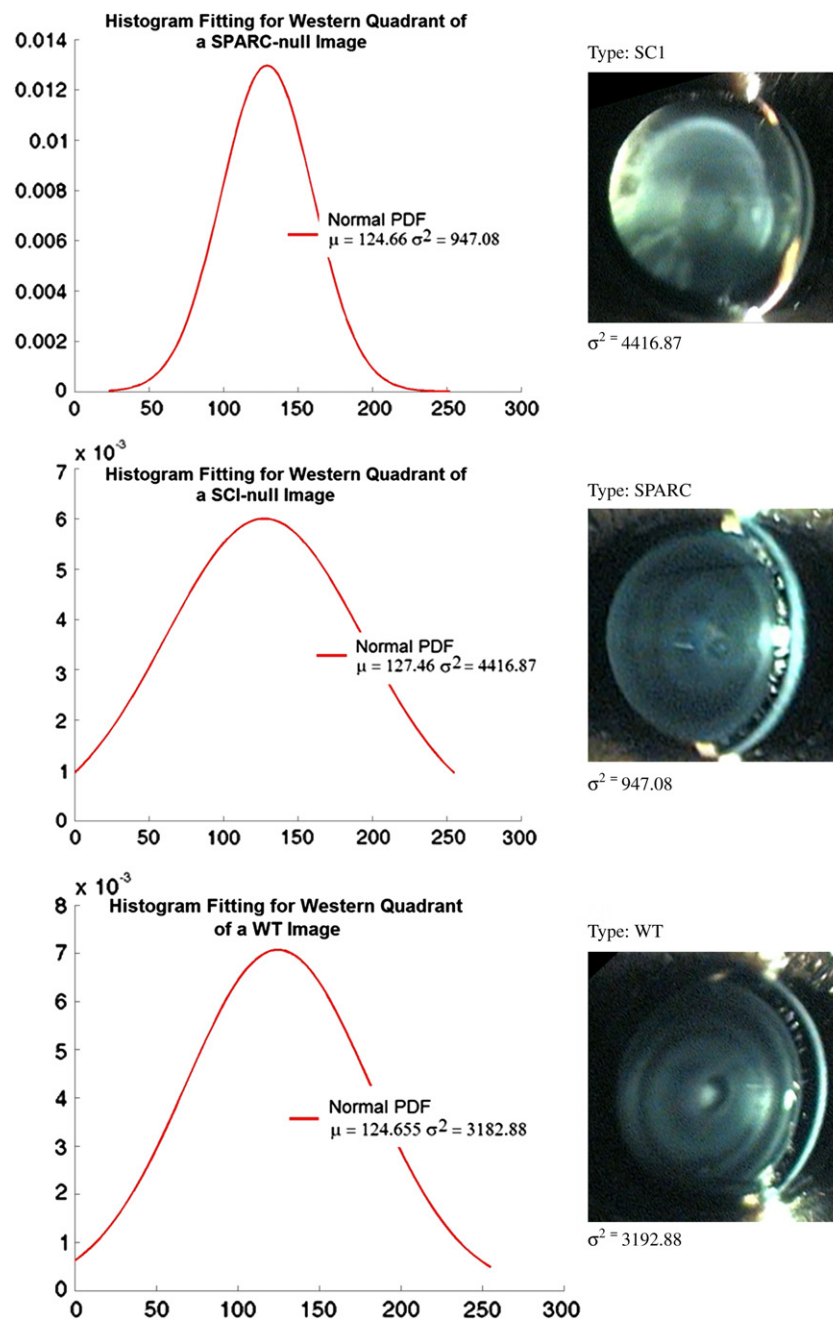


Fig. 11. The histogram of the western quadrant of the images of three different lens mutants. The variance of pixel intensities is smaller in SPARC null images and larger in SCI null images, while the WT type images had a variance value in between.



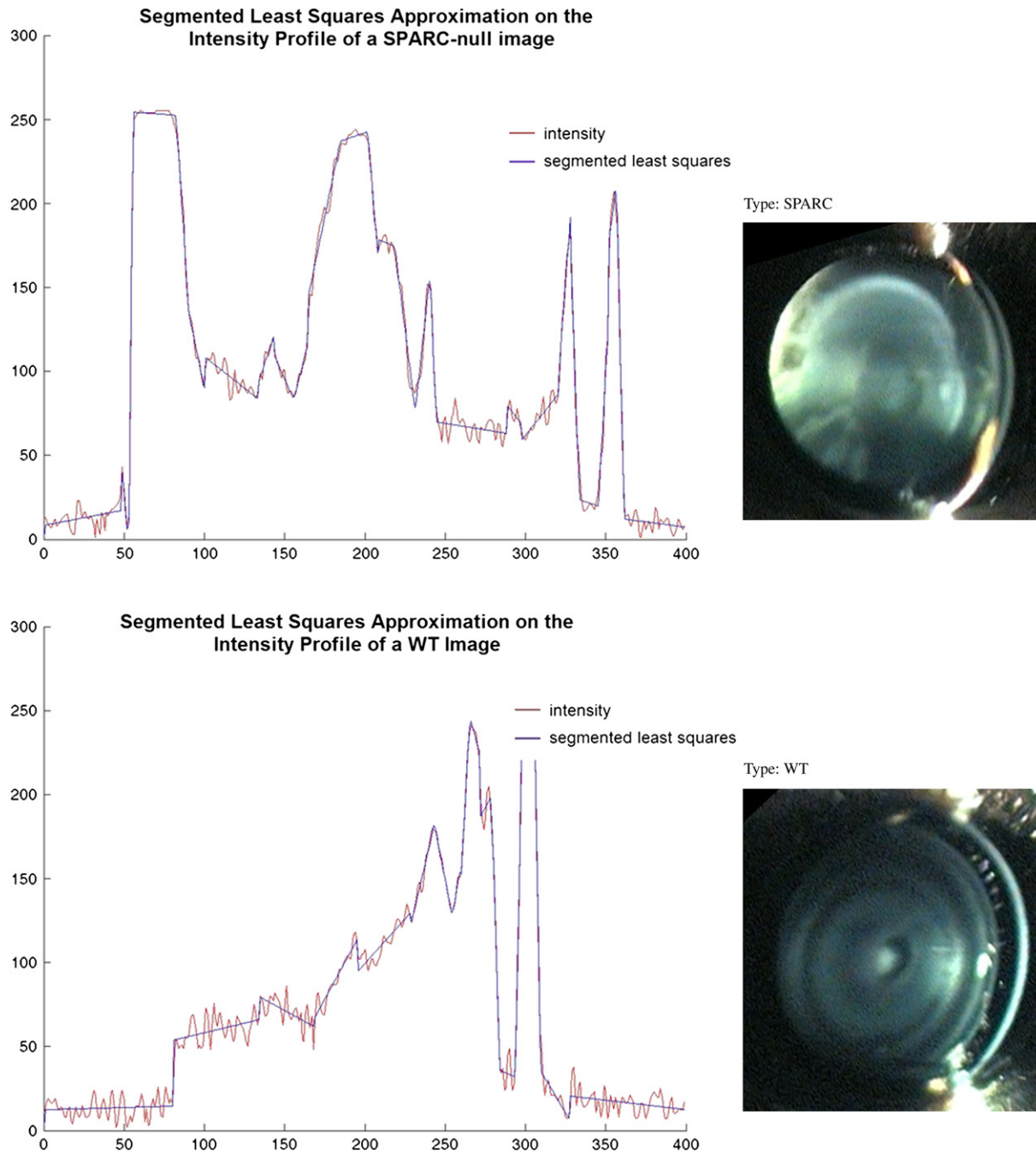


Fig. 12. The fluctuations in the intensity profile were not well characterized when fit to a continuous function. The upper graph was the intensity profile for the SPARC null lens (top right) and the bottom graph was the intensity profile for the WT (bottom right). The intensity fluctuations were observed as sharp edges in the intensity profile that were characteristic in SPARC null, while their presence was not as prominent in the WT class.

changes. To encapsulate the “continuity,” a piecewise linear function was fit on  $I(x)$ .

The standard least squares linear fitting algorithm fits a line to a set of  $n$  2-dimensional points  $P = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ . The line with the minimum error is  $y = ax + b$ , where

$$a = \frac{n \sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

$$b = \frac{\sum_i y_i - a \sum_i x_i}{n}$$

The error of the fit is:

$$\text{Error}(L, P) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

The objective of the segmented least squares approximation is to find the partition of consecutive points that minimizes the error of the fit, which corresponds to the sum of the errors

for each linear segment used in the fit with an additional cost  $C$  for each segment used as defined by:

$$E_{\text{fit}} = \sum_{p_i \in P} \text{Error}(l_i, p_i) + C$$

where  $p_i$  is a set of consecutive points in  $I(x)$ ,  $l_i$  is the least squares linear fit on this set, and  $C$  is the cost for each extra segment. We used values of  $C$  on the order of 500.

After fitting the piecewise linear function to  $I(x)$ , only the left side of the lens was considered for this feature extraction. This was done by considering the segments up to the middle of the image and discarding the first segments that had a very small value for their slope and intersection at the origin. These segments corresponded to the dark section of the image, not a part of the mouse lens. The features extracted (Fig. 12) were:

- number of segments;
- means of the slope and the intersection at the origin ( $a, b$ ) of each of the segments;
- variances of the slope and the intersection at the origin ( $a, b$ ) of each of the segments.

## 2.8. Learning

The analysis uses an inference engine mapping a feature space to a discrete set of labels for lens mutants. Each set is distinguished by the way it learns or obtains information to output decisions generalizing the training data presented to unseen situations. Supervised learning uses labeled training data as a knowledge base. Unsupervised learning determines patterns from unlabeled training input. Reinforcement learning involves a reward when the system output is correct and a penalty when the output is incorrect with the objective of identifying the output on the basis of the maximum reward. For the current problem supervised learning is expected to be most effective given the features of the images and their respective classes or labels.

### 2.8.1. Neural networks

The information processing capacity of the human brain is thought to emerge from networks of neurons. This was the theoretical basis for the creation of artificial neural networks as inference engines. Neural networks are composed of nodes or units connected by directed links. A link from unit  $j$  to unit  $i$  serves to propagate the activation  $a_j$  from  $j$  to  $i$ . Each link also has a numeric weight  $W_{j,i}$  associated with it, which determines the strength and sign of the connection. A network of units or a neural network is a structure with a set of units connected in a particular way and a set of defined weights for each unit. A network can be feed-forward, or recurrent. Feed-forward networks represent a function of their inputs, while recurrent networks feed their output back to their inputs. Feed-forward networks are arranged with layers of units such that each unit in a layer receives input from only the units in its preceding layer. A neural network represents the creation of a separator in an  $n$ -dimensional space (where  $n$  is the size of

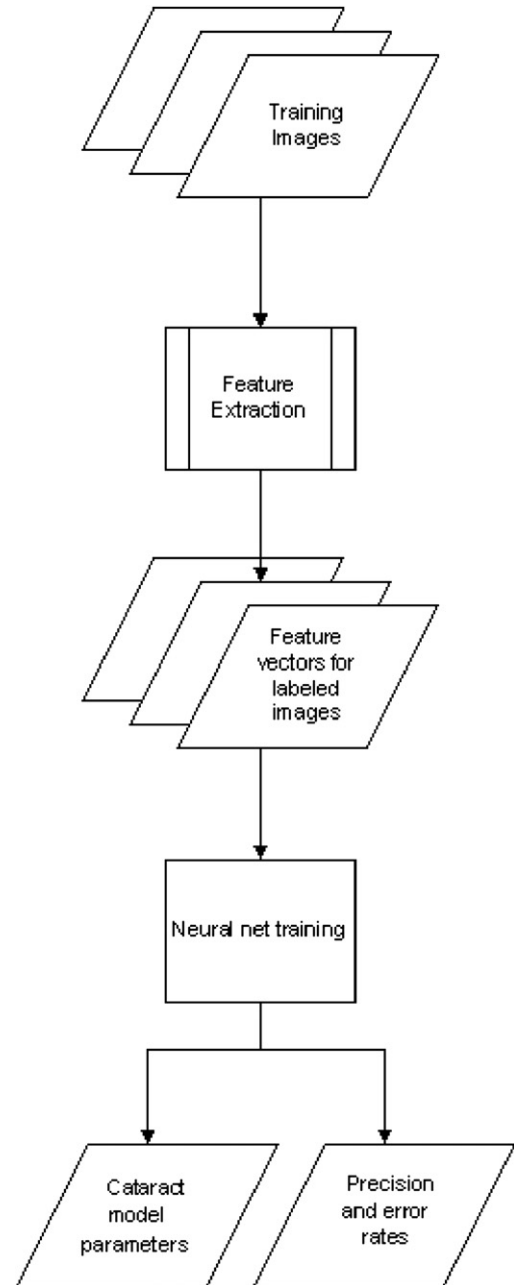


Fig. 13. Learning process.

the feature vector or the number of inputs to the neural network). Neural networks can learn (adjust the weights of their units) using different techniques. Backpropagation is an iterative technique where an example is applied to the network (originally set with random weights) and the weights are readjusted based on the quality of the output. The value is then propagated backwards through the network and changes are made to the weights of the nodes in each layer. This process is repeated until the overall error is below some threshold.

### 2.8.2. Use of neural networks

The numeric attributes from the four types of features in a lens images can be concentrated into a 46-dimensional feature vector  $F = \{f_1, f_2, \dots, f_{46}\}$ . For each training image  $I_T \in T_S$ ,

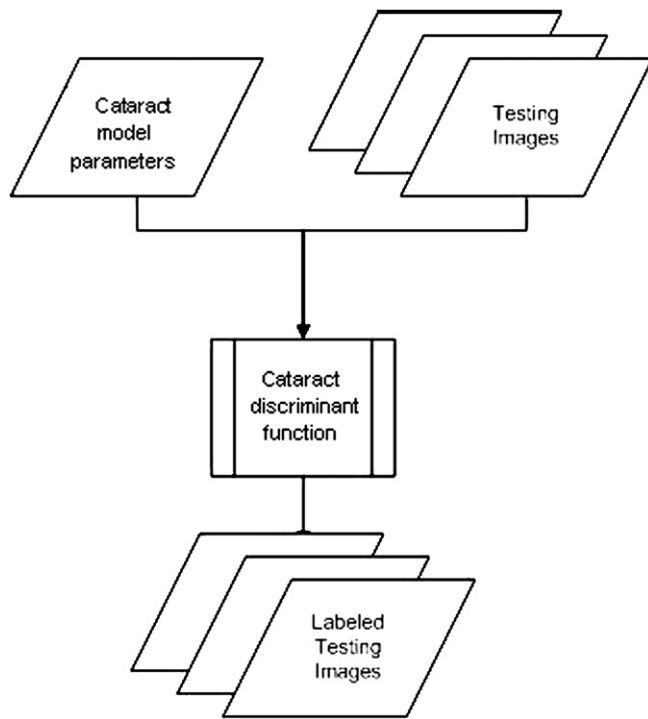


Fig. 14. Use of the final categorization.

a feature vector  $F(I_T)$  can be extracted and  $T = (F, c)$  can be constructed, where  $c$  indicates the known training vector  $F$ . The algorithm can be trained to discriminate between the different phenotypes and to become a decision function that, when given a vector  $U$  of attribute-values from a previously unseen lens image, will produce a value,  $c$ , that indicates the most likely genotype for the new image. The training process for obtaining the automated parameters is summarized in Fig. 13. The current analysis was implemented as a neural network with 15 hidden units in one hidden layer for training using the data mining software package Weka (Witten and Eibe, 2005). The model for the final system was trained with the whole dataset.

### 3. Results

The testing phase (Fig. 14) used the parameters obtained from the training phase and accepted slit-lamp images outside the training dataset for the analysis. The resulting output was compared to the correct genotypes and an error was calculated. The analysis was tested using three image sets: SC1 null, SPARC null, and WT lenses. The testing was performed with a 10-fold cross-validation. This method consisted of dividing each dataset into 10 subsets and using each of these

Table 2

Confusion matrix for the results trained using only intensity profile features

	SCI	SPARC	WT
SCI	103	1	3
SPARC	1	109	45
WT	0	60	61

subsets in one of the 10 iterations as the testing set while the rest were used as the training set. The dataset consisted of 107 images of SC1 null, 155 of SPARC null, and 121 of WT. The values listed in the following tables are the average for the ten individual experiments.

#### 3.1. Using only ring features

The performance of the trained analysis on the ring feature alone was shown in Table 1, expressed as a confusion matrix. The diagonal elements of the matrix (SC1, SC1), (SPARC, SPARC), and (WT, WT) indicated the number of instances in which an image was identified correctly. More than 69% of the images were identified correctly showing that the ring pattern feature correlated highly with the genotype. The off-diagonal elements show incorrect genotyping. Columns denote the underlying (real) categories and rows denote the inferred categories. For example, the entry in row SC1, column WT indicates that 17 SC1 images were wrongly categorized as WT.

#### 3.2. Using only intensity profile features

The performance of a trained analysis on the intensity profile feature alone is shown in Table 2. A significant improvement can be observed in the characterization of the SC1 null class in comparison to the ring feature. However, the analysis showed some confusion differentiating between SPARC null and WT. The percentage of correct identifications was 71%.

#### 3.3. Using both intensity profile and ring features

The performance for a combined ring-profile feature is shown in Table 3. Using both features not only reduced the confusion between the WT and SPARC null mutants, but also increased the accuracy for the SC1 null mutant. This analysis achieved an overall accuracy of 77%.

#### 3.4. Using ring, intensity profile, and western quadrant histogram features

Integrating the histogram of one of the quadrants of the lens, a higher accuracy was achieved and more instances of

Table 1

Confusion matrix for results using only ring features

	SCI	SPARC	WT
SCI	81	9	17
SPARC	8	110	37
WT	12	35	74

Table 3

Confusion matrix for trained using ring and intensity profile features

	SCI	SPARC	WT
SCI	99	6	2
SPARC	3	111	41
WT	1	34	86

Table 4

Confusion matrix for the results using ring, intensity profile, and western quadrant histogram features

	SCI	SPARC	WT
SCI	97	6	4
SPARC	2	118	35
WT	1	33	87

the classes SPARC null and WT were identified correctly. Less accuracy was achieved with SC1 null in comparison to the previous setup. The overall accuracy is 79% and the confusion matrix is shown in Table 4.

### 3.5. Using ring, intensity profile, western quadrant histogram, and segmented least squares fitting features

The introduction of the segmented least squares fitting feature increased the number of correct identifications and reduced the confusion observed using only the ring and intensity curve features. The accuracy using this set of features was of 83% and the confusion matrix is shown in Table 5.

## 4. Discussion

A novel neural network approach for analysis of mouse lenses with different genotypes was developed using four types of features: the ring pattern, the intensity profile, the western quadrant histogram, and segmented least squares. The first two features were of major importance while the last two were additions that improved the accuracy of the analysis. The natural layering of transparent fibers in the lens varies with the genotype of the mice. The intensity profile is a feature that captured information from the rings, and the FFT analyzed the pattern formed by all the rings collectively not locally. This novel analysis permits categorization of large data bases of various types and stages of opacification on the basis of spatial and temporal differences in the pattern of opacities of the lens. Image analysis that accurately and consistently uses recognition of ring and intensity patterns in the identification of the genotype of mice can be expected to add a simple, new, automated and standardized method for computerized evaluation of patterns of early stage opacity and to establish phenotype–genotype relationships between factors contributing to lens structure. The preliminary experimental data from Tables 1–5 established the success of the new approach in distinguishing patterns of opacity in the SC1 null class from patterns of opacity in the SPARC null and WT classes to nearly 99% accuracy

Table 5

Confusion matrix for the results using ring, intensity profile, western quadrant, and segmented least squares fitting features

	SCI	SPARC	WT
SCI	104	1	2
SPARC	1	122	32
WT	0	30	91

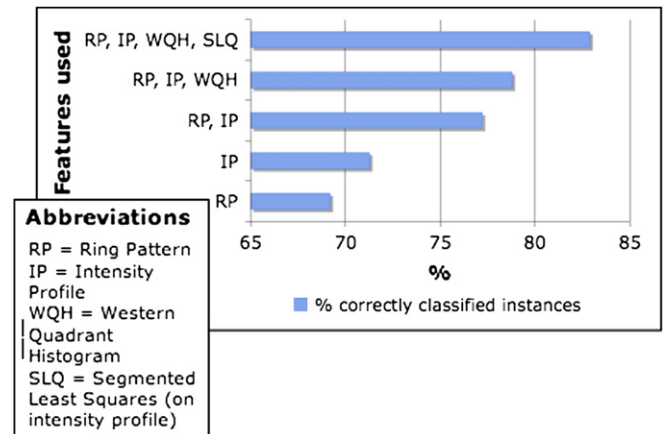


Fig. 15. Percentage of correct identification using different combinations of features.

when four features were used. It was reasonably successful in distinguishing between the SPARC null and WT classes with a 77% accuracy. Overall, the analysis achieved a 83% accuracy (Fig. 15). Repeated application of the algorithm to the same database produced the same results. The result was an impressive demonstration of a very promising method for an automated computer analysis of lens images that distinguished between selective categories of genetic mutation. The success of the algorithm is remarkable when considering the fact that key factors including age and cellular microstructure were not controlled, which suggests fundamental structural patterns remain at all ages even when differences are subtle (Fig. 16). It needs to be appreciated that slit lamp images are low resolution. The underlying cellular basis for the slit lamp phenotype can only be determined by detailed and systematic microscopic analyses (Kuszak and Costello, 2004; Kuszak et al., 2006; Blankenship et al., 2007). The standard deviation in phenotype identification rates for the 10-fold cross-validation using all the features was 0.015 and shows a robust system with the current training samples. The results provide baseline data for improved informatics approaches for establishing phenotype/genotype relationships using slit lamp images of eyes in living animals and encourage continued research on automated and quantitative screening of phenotypes in large image databases.

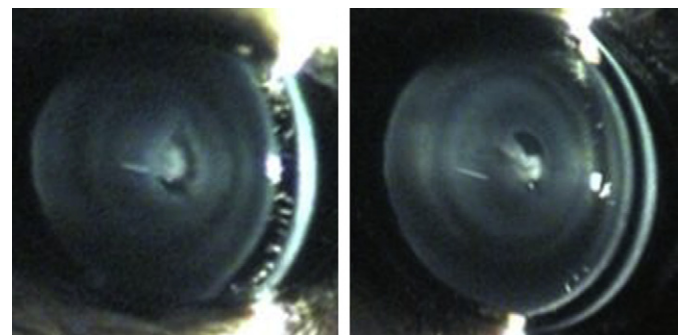


Fig. 16. A WT (left) and a SPARC null (right) image.



## Acknowledgments

Support is gratefully acknowledged from LM000714 to J.F.B. and L.G.S., EY04542, EY13180 to J.I.C. from the NIH; and IIS0097329 to L.G.S., Y.L. and J.Y. from the NSF.

## References

- Alizadeh, A., Clark, J., Seeberger, T., Hess, J., Blankenship, T., Fitzgerald, P.G., 2004. Characterization of a mutation in the lens-specific CP49 in the 129 strain of mouse. *Investigative Ophthalmology and Visual Science* 45, 884–891.
- Ansari, R.R., King, J.F., Seeberger, T., Clark, J.I., 2003. Early detection of cataract and response to pantethine therapy with non-invasive static and dynamic light scattering. *Proceedings of SPIE, Ophthalmic Technology XII* 4951, 168–178.
- Bassuk, J.A., Birkebak, T., Rothmier, J.D., Clark, J.M., Bradshaw, A., Muchowski, P.J., Howe, C.C., Clark, J.I., Sage, E.H., 1999. Disruption of the Sparc locus in mice alters the differentiation of lenticular epithelial cells and leads to cataract formation. *Experimental Eye Research* 68, 321–331.
- Benedek, G., 1983. Why the eye lens is transparent. *Nature* 302, 383–384.
- Benedek, G.B., Chylack Jr., L.T., Libondi, T., Magnante, P., Pennett, M., 1987. Quantitative detection of the molecular changes associated with early cataractogenesis in the living human lens using quasielastic light scattering. *Current Eye Research* 6, 1421–1432.
- Bettelheim, F.A., Churchill, A.C., Zigler Jr., J.S., 1997. On the nature of hereditary cataract in strain 13/N guinea pigs. *Current Eye Research* 16, 917–924.
- Blankenship, T., Bradshaw, L., Shibata, B., Fitzgerald, P., 2007. Structural specializations emerging late in mouse lens fiber cell differentiation. *Investigative Ophthalmology and Visual Science* 48, 3269–3276.
- Brekken, R.A., Sage, E.H., 2001. SPARC, a matricellular protein: at the crossroads of cell-matrix communication. *Matrix Biology* 19, 816–827.
- Chenault, V.M., Ediger, M.N., Ansari, R.R., 2002. In vivo assessment of diabetic lenses using dynamic light scattering. *Diabetes Technology and Therapeutics* 4, 651–659.
- Chylack Jr., L.T., Khu, P.M., 2001. Subjective classification and objective quantitation of human cataract. In: Albert, D.F.J. (Ed.), *Principles and Practice of Ophthalmology*. W.B. Saunders, Philadelphia, PA, pp. 1449–1463.
- Chylack Jr., L.T., Wolfe, J.K., Friend, J., Khu, P.M., Singer, D.M., McCarthy, D., del Carmen, J., Rosner, B., 1993a. Quantitating cataract and nuclear brunescence, the Harvard and LOCS systems. *Optometry and Vision Science* 70, 886–895.
- Chylack Jr., L.T., Wolfe, J.K., Singer, D.M., Leske, M.C., Bullimore, M.A., Bailey, I.L., Friend, J., McCarthy, D., Wu, S.Y., 1993b. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Archives of Ophthalmology* 111, 831–836.
- Chylack Jr., L.T., Wolfe, J.K., Friend, J., Tung, W., Singer, D.M., Brown, N.P., Hurst, M.A., Kopcke, W., Schalch, W., 1995. Validation of methods for the assessment of cataract progression in the Roche European-American Anticataract Trial (REACT). *Ophthalmic Epidemiology* 2, 59–75.
- Clark, J.I., 2004. Order and disorder in the transparent media of the eye. *Experimental Eye Research* 78, 427–432.
- Clark, J.I., Carper, D., 1987. Phase separation in lens cytoplasm is genetically linked to cataract formation in the Philly mouse. *Proceedings of the National Academy of Sciences of the United States of America* 84, 122–125.
- Datiles, M.B., Edwards, P.A., Trus, B.L., Green, S.B., 1987. In vivo studies on cataracts using the Scheimpflug slit lamp camera. *Investigative Ophthalmology and Visual Science* 28, 1707–1710.
- Datiles 3rd, M.B., Ansari, R.R., Reed, G.F., 2002. A clinical study of the human lens with a dynamic light scattering device. *Experimental Eye Research* 74, 93–102.
- Dimock, J., Robman, L.D., McCarty, C.A., Taylor, H.R., 2000. Digital cataract photography. *The Journal of Audiovisual Media in Medicine* 23, 7–11.
- Edwards, P.A., Datiles, M.B., Unser, M., Trus, B.L., Freidlin, V., Kashima, K., 1990. Computerized cataract detection and classification. *Current Eye Research* 9, 517–524.
- Fan, S., Dyer, C.R., Hubbard, L., Klein, B., 2003. An automatic system for classification of nuclear sclerosis from slit-lamp photographs. In: Ellis, R.E., Peters, T.M. (Eds.), *MICCAI 2003*, Vol. LNCS 2878. Springer-Verlag, pp. 592–601.
- Goldstein, L.E., Muffat, J.A., Cherny, R.A., Moir, R.D., Ericsson, M.H., Huang, X., Mavros, C., Coccia, J.A., Faget, K.Y., Fitch, K.A., Masters, C.L., Tanzi, R.E., Chylack Jr., L.T., Bush, A.I., 2003. Cytosolic beta-amyloid deposition and supranuclear cataracts in lenses from people with Alzheimer's disease. *Lancet* 361, 1258–1265.
- Graw, J., 1999. Mouse models of congenital cataract. *Eye (London, England)* 13 (Pt 3b), 438–444.
- Graw, J., 2004. Congenital hereditary cataracts. *The International Journal of Developmental Biology* 48, 1031–1044.
- Harding, J., 1991. *Cataract: Biochemistry, epidemiology, and pharmacology*. Chapman and Hall, London, UK.
- Kashima, K., Trus, B.L., Unser, M., Edwards, P.A., Datiles, M.B., 1993a. Aging studies on normal lens using the Scheimpflug slit-lamp camera. *Investigative Ophthalmology and Visual Science* 34, 263–269.
- Kashima, K., Unser, M., Datiles, M.B., Trus, B.L., Edwards, P.A., 1993b. Minimum views required to characterize cataracts when using the Scheimpflug camera. *Graefes's Archive for Clinical and Experimental Ophthalmology (Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie)* 231, 687–691.
- Klein, B.E., Hubbard, L., Ferrier, N.J., Klein, R., Klein, D.J., Lee, K.E., Ewen, A., Jensen, K., Evans, M.D., 2005. Detecting progression of nuclear sclerosis by using human grading versus semiautomated computer grading. *Investigative Ophthalmology and Visual Science* 46, 1155–1162.
- Kojima, M., Okuno, T., Miyakoshi, M., Sasaki, K., Takahashi, N., 2002. Environmental temperature and cataract progression in experimental rat cataract models. *Developments in Ophthalmology* 35, 125–134.
- Kuszak, J.R., Costello, M.J., 2004. The structure of the vertebrate lens. In: Lovicu, F.J., Robinson, J.L. (Eds.), *Development of the Ocular Lens*. Cambridge University Press, Cambridge, UK, pp. 71–118.
- Kuszak, J.R., Mazurkiewicz, M., Zoltoski, R., 2006. Computer modeling of secondary fiber development and growth: I. Nonprimate lenses. *Molecular Vision*, 251–270.
- Leske, M.C., Chylack Jr., L.T., Wu, S.Y., Schoenfeld, E., He, Q., Friend, J., Wolfe, J., 1996. Incidence and progression of nuclear opacities in the Longitudinal Study of Cataract. *Ophthalmology* 103, 705–712.
- Nishio, I., Weiss, J.N., Tanaka, T., Clark, J.I., Giblin, F.J., Reddy, V.N., Benedek, G.B., 1984. In vivo observation of lens protein diffusivity in normal and X-irradiated rabbit lenses. *Experimental Eye Research* 39, 61–68.
- Norose, K., Lo, W.K., Clark, J.I., Sage, E.H., Howe, C.C., 2000. Lenses of SPARC-null mice exhibit an abnormal cell surface-basement membrane interface. *Experimental Eye Research* 71, 295–307.
- Phelps-Brown, N.A., Bron, A.J., 1996. *Lens Disorders: a Clinical Manual of Cataract Diagnosis*. Butterworth Heinemann, Oxford, UK.
- Rodriguez, I.R., Gonzalez, P., Zigler Jr., J.S., Borrás, T., 1992. A guinea-pig hereditary cataract contains a splice-site deletion in a crystallin gene. *Biochimica et Biophysica Acta* 1180, 44–52.
- Russell, S.P.N., 1995. *Artificial Intelligence: A Modern Approach*, Prentice Hall Series on Artificial Intelligence. Prentice Hall, Englewood Cliffs, NJ.
- Seeberger, T.M., Matsumoto, Y., Alizadeh, A., Fitzgerald, P.G., Clark, J.I., 2004. Digital image capture and quantification of subtle lens opacities in rodents. *Journal of Biomedical Optics* 9, 116–120.
- Simpanya, M.F., Ansari, R.R., Suh, K.I., Leverenz, V.R., Giblin, F.J., 2005. Aggregation of lens crystallins in an in vivo hyperbaric oxygen guinea pig model of nuclear cataract: dynamic light-scattering and HPLC analysis. *Investigative Ophthalmology and Visual Science* 46, 4641–4651.
- Sullivan, M.M., Sage, E.H., 2004. Hevin/SC1, a matricellular glycoprotein and potential tumor-suppressor of the SPARC/BM-40/Osteonectin family. *The International Journal of Biochemistry and Cell Biology* 36, 991–996.

- Sullivan, M.M., Barker, T.H., Funk, S.E., Karchin, A., Seo, N.S., Hook, M., Sanders, J., Starcher, B., Wight, T.N., Puolakkainen, P.A., Sage, E.H., 2006. Matricellular hevin regulates decorin production and collagen assembly. *Journal of Biological Chemistry* 281, 27621–27632.
- Thurston, G.M., Hayden, D.L., Burrows, P., Clark, J.I., Taret, V.G., Kandel, J., Courogen, M., Peetermans, J.A., Bowen, M.S., Miller, D., Sullivan, K.M., Storb, R., Stern, H., Benedek, G.B., 1997. Quasielastic light scattering study of the living human lens as a function of age. *Current Eye Research* 16, 197–207.
- Thylefors, B., Chylack Jr., L.T., Konyama, K., Sasaki, K., Sperduto, R., Taylor, H.R., West, S., 2002. A simplified cataract grading system. *Ophthalmic Epidemiology* 9, 83–95.
- Tumminia, S.J., Clark, J.I., Richiert, D.M., Mitton, K.P., Douglas-Tabor, Y., Kowalak, J.A., Garland, D.L., Russell, P., 2001. Three distinct stages of lens opacification in transgenic mice expressing the HIV-1 protease. *Experimental Eye Research* 72, 115–121.
- Wolf, N., Penn, P., Pendergrass, W., Van Remmen, H., Bartke, A., Rabinovitch, P., Martin, G.M., 2005. Age-related cataract progression in five mouse models for anti-oxidant protection or hormonal influence. *Experimental Eye Research* 81, 276–285.
- Yan, Q., Clark, J.I., Wight, T.N., Sage, E.H., 2002. Alterations in the lens capsule contribute to cataractogenesis in SPARC-null mice. *Journal of Cell Science* 115, 2747–2756.
- Witten, I.H., Eibe, F., 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, CA.