Portland State University

PDXScholar

Electrical and Computer Engineering Faculty
Publications and Presentations

Electrical and Computer Engineering

1-1-1997

New Objective Classification System for Nuclear Opacification

Donald D. Duncan
Portland State University

O. B. Shukla

Sheila West

Oliver D. Schein

Follow this and additional works at: https://pdxscholar.library.pdx.edu/ece_fac

Part of the [Electrical and Computer Engineering Commons](#)

Let us know how access to this document benefits you.

Citation Details

D. D. Duncan, O. B. Shukla, S. K. West, and O. D. Schein, "New objective classification system for nuclear opacification," J. Opt. Soc. Am. A 14, 1197-1204 (1997)

This Article is brought to you for free and open access. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

New objective classification system for nuclear opacification

D. D. Duncan and O. B. Shukla

The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723-6099

S. K. West and O. D. Schein

The Johns Hopkins University Dana Center for Preventive Ophthalmology, Baltimore, Maryland 21205

Received February 26, 1996; revised manuscript received December 9, 1996; accepted December 17, 1996

We have developed an autonomous objective classification scheme for degree of nuclear opacification. The algorithm was developed by using a series of color 35-mm slides acquired with a Topcon photo slit-lamp microscope and use of standard camera settings. The photographs were digitized, and first, and second-order gray-level statistics were extracted from within circular regions of the nucleus. Classifications of severity were performed by using these features as input to a neural network. Training versus classification performance was tested by using photographs of different eyes, and test/retest classification reproducibility was evaluated by using paired photographs of the same eyes. We demonstrate good performance of the classifier against subjective assessments rendered by the Wilmer grading system [Invest. Ophthalmol. Visual Sci. **29**, 73 (1988)] and markedly better test/retest reproducibility. © 1997 Optical Society of America [S0740-3232(97)00406-7]

Key words: Cataract, nuclear, opacification, classification, standards.

1. INTRODUCTION

Cataract is the single leading cause of blindness worldwide.¹ In the United States, Medicare reimbursement for cataract extraction is 12% of the entire Medicare budget.² For these reasons the National Eye Institute has identified research on cataract as a major priority in the 1994–1998 National Plan for Vision Research.³ Specifically, the plan has identified the investigation of techniques that can detect small amounts of change in cataract development over time.⁴ Such research requires valid and reliable systems for classifying cataract type and severity, i.e., systems that show sensitivity and can detect progression but are robust against measurement error.

Many of the current classification systems in use are based on photographic documentation of various regions of the lens with use of different camera systems. Typically the lens nucleus image is captured by using a photo slit-lamp microscope. The presence and severity of the opacity are then determined by trained graders reading the photographs and comparing them with a series of standard photographs.^{5–9} A number of investigative teams have developed standard sets of photographs of increasing opacification of the lens nucleus.^{10–12} Each of these systems has been shown to be reproducible and valid, and a comparative study of two of these systems (Wilmer and LOCS II) suggests that they give essentially equivalent results.¹³

This approach to grading cataracts is dependent on observations of trained photograph graders and hence has been termed subjective. Systems of this type are simple and robust but are insensitive to cataractous changes on time scales of less than approximately one year. Fur-

thermore, they are labor intensive and liable to observer drift and systematic error over time unless rigorous and expensive checking procedures are built into the photographic reading protocols. To ensure accuracy, photographs often are read by two independent graders with adjudication; this is time-consuming and costly. Moreover, measurement error inherent in any subjective system can be large relative to other sources of error, such as taking and processing the photographs.¹⁴

As a result, research into refining ways to grade cataracts is progressing toward objective classification schemes.^{15,16} While such schemes may never completely supplant the current subjective systems, it is clear that they have features that complement these existing approaches. Here we describe our approach to classification of nuclear opacification based on computer analysis of photographic imagery for feature extraction and the use of neural network classifiers to render an opacification grade.

2. APPROACH

Two sets of data, each acquired under the same conditions, were used for this study. The nuclear photographs were of participants with 1/16 or less (fractional area of involvement) cortical or posterior subcapsular changes and represented a full range of severity of opalescence, from grade 0.2 to grade 4.0. The first set (set 1) consisted of 100 photographs of different eyes. The second set (set 2) consisted of 48 pairs of photographs taken of the same eye on the same day. Photographs for set 1 were chosen without regard to quality. In fact, there were several that were considered rather poor.

Fig. 1. Illustration of typical region of interest (ROI) within lens nucleus for gray-level calculations.

All photographs were graded with the Wilmer grading scheme by two trained graders and were adjudicated by a third if the grades differed by more than 0.3.¹⁴ Photographs were acquired with a Topcon SL-7D slit-lamp microscope and Kodak Ektachrome ISO 200 film, with a flash setting of 5 and a slit beam angle of 45°. Slit width and height were standardized at 0.3 and 9 mm, respectively, and magnification was set at 16. Although these conditions were standardized, the actual photographs were taken at eight different sites around the U.S., by different photographers, and were developed at different labs. The resulting slides were digitized with a Nikon Coolscan. Digitization was in color with an 8-bit resolution in each of the three colors (red, green, and blue). A composite 24-bit color rendition resulted in color levels of 0–255 in each of the three colors. Spatial scan resolution was approximately 50 μm . Image processing was performed with use of IPLab software from Signal Analytics Corporation. Use was also made of extensions to this software that were developed with MATLAB and coded in C. The neural network classifier was developed with the NeuralWorks software from NeuralWare. This software package allows the user to train and test a variety of neural network paradigms.

A. Extraction of Nuclear Opacification Features

Perhaps the most important step in developing a classification scheme is to identify which features are relevant: One needs to select features of the photographs that are representative of the opacification and, one hopes, the impact on vision. A major portion of the development effort was devoted to this. Subsequent to the feature identification was the development of procedures for accurately and reproducibly extracting these features from the imagery.

For calculation of gray-level statistics we place a 75-pixel-diameter region of interest (ROI) within the nucleus of the lens. A typical ROI is illustrated in Fig. 1. These gray-level assessments are performed on the Y-encoded¹⁷

portion of the color image, i.e., the luminance. The justification for this choice is that since performance of the classifier is to be compared with the performance of human photo interpreters, the image should be comparable to what the human observer sees. Although the interpreters are trained (for the assessment of nuclear opalescence) to ignore color, their perception of photographic density is still affected by the eye's spectrally weighted photopic sensitivity.¹⁸

Within the ROI in the nucleus we calculate a longitudinal profile by averaging over the rows. An example of such a profile is shown in Fig. 2. Also shown in this figure is a weighted second-order polynomial least-squares fit. This polynomial is given by

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2, \quad (1)$$

where the independent variable x is zero at the anterior point and unity at the posterior point. In performance of the least-squares fit, the relative weights are the number of rows over which the average is computed at each longitudinal position. At the central column of the ROI, the weight is proportional to 75 (the diameter of the ROI). At the edges of the ROI, the weights are proportional to 13. In this manner, the degree to which the fit follows the profile is proportional to the confidence in the value of the profile. This approach reflects our objective of concentrating interest in the center of the nucleus and of relaxing the sensitivity to the exact placement of the ROI. We found that a circular ROI as opposed to a square one with the same side dimensions provided gray-level statistics that were less sensitive to the ROI placement. Further, with a square ROI we found it difficult to encompass the nucleus of the lens while simultaneously avoiding the dark zone of discontinuity that separates the nucleus and the cortex.

Another example of a profile and resulting fit is shown in Fig. 3. For this case the cataract is quite severe ($N_s = 3.9$) as opposed to the case illustrated in Fig. 2 ($N_s = 2.3$). Parameters of the second-order fit illustrated here constitute the metrics that we are interested in. For a moderate cataract, the sulcus is still clearly discernible and the profile has a weak overall upward convexity. For a more severe cataract the sulcus is no longer visible and the profile develops a strong upward convexity.

Fig. 2. Example longitudinal profile and second-order fit for moderate cataract ($N_s = 2.3$).

Fig. 3. Example longitudinal profile and second-order fit for severe cataract ($N_s = 3.9$).

These general properties are reflected in the residual of the fit and the slope of the fit at the posterior point. In the case of mild cataract the residual is relatively large (the single inflection allowed by the second-order fit cannot describe the sulcus) and the slope of the fit at the posterior point is shallow. On the other hand, for severe cataract, the residual is quite small, and the slope of the fit at the posterior point is steep. Thus the residual of the fit provides information on the fine-scale spatial structure within the nucleus and the uniformity of the opacification. Further, the slope of the profile at the posterior point yields information on the total integrated scatter; consideration of the Lambert–Beer law¹⁹ shows that this slope is proportional to the path integral of the extinction (scatter plus absorption) cross section. In addition to these two parameters we use a third—the mean value of the fit. In terms of the variables of Eq. (1), these parameters are given by

$$\begin{aligned}\mu &= \int_0^1 \hat{y} \, dx \\ \beta_p &= \left. \frac{d\hat{y}}{dx} \right|_{x=1} \\ \sigma &= \left[\int_0^1 (\hat{y} - y)^2 dx \right]^{1/2},\end{aligned}\quad (2)$$

where y is the original profile and $\hat{y}$ is the least-squares estimate.

To gain a measure of insensitivity to overall photograph exposure, we chose to define these metrics in relative terms. The normalization factor was chosen to be the peak gray-level value, c , of the cornea along the optical axis. Thus the classification metrics were the normalized nuclear mean gray level, μ/c , the normalized slope at the posterior point of the profile, β_p/c , and the fractional residual of the second-order fit (residual normalized to the mean), σ/μ . We found these features most useful for classification of the degree of nuclear cataract. The latter two features incorporate information on the spatial structures within the nucleus that are responsible for scatter. As such, they are second-order statistics. Our earlier efforts concentrated on the use of first-order gray-level statistics for classification; experience showed

that such features are insufficient. Other researchers also have suggested that a measure of spatial variability or structure be incorporated in a classification scheme.^{20,21}

The size of the ROI was a compromise between the conflicting requirements of being large enough to smooth out image artifacts, yet not so large as to suppress detail. Often, if the patient has any cortical cataract, the effect seen in slit-lamp photography is that of crepuscular rays²² or “sunlight shining through the clouds.” This effect is highly variable and depends on details of the illumination beam and fixation. An ROI of this size is effective in smoothing out such an image artifact, making the resulting features less sensitive to details of the geometry. On the other hand, because the illumination is presented as a beam of finite width entering a scattering medium, the illumination level tends to decrease away from the beam axis. As a result, if the ROI is made too large in the transverse direction (the vertical direction in these photographs), this illumination variation tends to suppress the dynamic range of the average profile. Further, since the width of the beam entering the lens is determined by the pupil, it is sensitive to dilation. For an ROI of this size, the gray levels seldom varied in the transverse (vertical) direction by more than 10% of the mean.

The placement of the ROI can be automated because nuclear photographs all share several features (horizontal symmetry, relative positions of the cornea, lens, etc.) that serve as markers. Variations among nuclear photographs include the presence of optical artifacts, the mean value of the intensity, and the size and shape of the lens nucleus. The ROI placement algorithm first removes any optical artifacts that typically exist at the edge of the cornea. Then the optical axis of the lens is determined with use of the horizontal symmetry of the lens (equalizing the number of pixels above and below the optical axis). A profile of the lens taken along the optical axis contains a number of peaks corresponding to the cornea, anterior chamber, and nucleus boundaries. The location and extent of the lens nucleus is determined by detecting these peaks. The boundary of the ROI within the nucleus is computed and finally the pixels within the ROI are extracted to generate the profile.

B. Classification Algorithm

The three features discussed above were used as input to a neural network to determine a grade for degree of nuclear opacification. The architecture of the neural classifier was that of an input layer with three input nodes (one for each classification feature), one hidden layer with 26 nodes, and an output layer consisting of a single output node (for the objective grade). The input layer was fully connected to the hidden layer, which in turn was fully connected to the output node. Strengths of the connections between the nodes in each layer were established through a training algorithm. There are various means of training a neural network. The one that we found to display a good compromise between trainability and subsequent generalizability (ability to classify new data) was the reinforcement algorithm.^{23,24} Set 1 ($n = 100$) constituted the training set. Because of the subjective nature of the Wilmer grading scheme, this

training set was graded on five separate occasions. To minimize noise in the training set, we chose to train only on the median of the five gradings for each slide. To train the network, we used the set of input parameters (the three metrics for each photograph in set 1) matched with the desired set of output values (the median subjective grade for the corresponding photograph). The reinforcement algorithm strengthened those connections (weights), which yielded a network output approaching the desired output. The algorithm continued this process, repeatedly cycling through the entire set of input parameters until the difference between the neural network output and the desired output was acceptably small. This training process took approximately six hours on a Macintosh 9500, 132-MHz PowerPC.

To gain acceptance, any new classification scheme for cataract should show a favorable comparison against a clinical judgment of degree of opacification. As a result, we have chosen to train our network on and assess its performance against subjective grades of opacification on the basis of well-documented systems.^{7,8,10,13,14} While there is some variability in the subjective classification scheme, we feel this approach to the issue of validity would have the most meaning from a clinical perspective. Moreover, such subjective clinical assessments have indeed shown very good correlation with measured data on nuclear scatter obtained with use of the Lens Opacity Meter,^{25,26} with visual acuity and contrast sensitivity,²⁷ and with measured line-spread functions.²⁸

3. RESULTS

The issues that we deal with here are the trainability, generalizability, reproducibility, and sensitivity of the neural classifier. Our first objective is to evaluate the ability of the network to "learn." This ability of the classifier to learn is equivalent to its validity; unless a reasonable concordance with accepted grading schemes can be shown, the objective classification scheme will not be viewed as valid.

The ultimate objective of an autonomous classification scheme is to evaluate new photographs with use of a classifier that has been previously "trained." To do this we evaluated the capability of our previously trained algorithm to classify the paired photographs of set 2. Results are compared with the subjective assessment rendered by the Wilmer grading scheme.

An important objective in performance evaluation was to quantify reproducibility. Since our entire classification scheme is automated (from the ROI placement to the rendering of a grade), the major source of variability is in the photographs themselves. To assess our scheme's ability to cope with these variations, we used the paired photos of set 2. Performance of the algorithm was compared against the test/retest scores obtained with the Wilmer grading system. Note that by the use of the nomenclature "test" and "retest," we mean test and retest of *the patient*. Each of the paired photos received subjective and objective assessments.

Finally, we address the question of sensitivity. At issue here is the ability of the classifier to measure small changes in degree of opacification.

A. Training

In recognition of the fact that there is variability in the training set, the individual photos of set 1 were graded five times. These subjective grades constituted the training set. Shown in Fig. 4 is the grade/regrade performance of the Wilmer subjective grading scheme for these photos. For each member of the training set we show the five individual scores and the 95% confidence interval about the mean. The confidence intervals ran from ± 0 to ± 0.933 with a mean of ± 0.319 . Also shown in this figure are the ± 0.3 limits for adjudication. The figure illustrates the fact that some of the photographs were of rather poor quality. For these cases, we observed wide variations in the subjective scores. An estimate of the global variability in the scores for the training set was derived by calculating a mean for each slide, subtracting this mean from each of the five grades, and pooling the 500 differences. The standard deviation of these differences was found to be 0.254. To suppress some of the variability in the training set, we used the median grade for each photograph rather than the entire set of 500 grades.

Figure 5 illustrates the training of the neural network. The network had a three-layer architecture: an input layer of three nodes, a hidden layer of 26 nodes, and an output layer consisting of a single node. Training was conducted with a reinforcement algorithm. Agreement between the objective and the subjective grades was good. Also shown for comparison between the two is the linear regression fit with the subjective grade treated as the independent variable:

$$N_o = 0.253 + 0.867 N_s, \quad r = 0.932, \quad n = 100, \quad (3a)$$

where N_s and N_o are, respectively, the subjective and the objective grades. The intercept was small but significant. For the subjective-objective difference ($\Delta \equiv N_s - N_o$), the maximum, standard deviation, and 95% confidence interval (CI), respectively, were the following:

$$|\Delta| \leq 1.41, \quad \sigma_\Delta = 0.397, \quad 95\% \text{ CI} = 0.079. \quad (3b)$$

Fig. 4. Variability in subjective Wilmer grading system for photographs in set 1. Overall confidence interval (95%) of training set is ± 0.319 of an integer grade. Also shown are the ± 0.3 limits for adjudication.

Fig. 5. Results of training neural network on photograph set 1.

B. Classification

Figure 6(a) is a plot of the performance of the trained neural classifier against the Wilmer scores for the first half of photographic set 2 (test set). Performance of the neural classifier for the second half of photographic set 2 (retest set) is shown in Fig. 6(b). These figures represent the ability of the previously trained classifier to generalize to the classification of new photos. Statistics on the differences for, respectively, the first and second halves of set 2 were the following:

Test

$$N_o = 0.957 N_s$$

$$r = 0.933, \quad n = 48,$$

$$|\Delta| \leq 1.09, \quad \sigma_\Delta = 0.424, \quad 95\% \text{ CI} = 0.123.$$

Retest

$$N_o = 0.964 N_s,$$

$$r = 0.951, \quad n = 48,$$

$$|\Delta| \leq 1.14, \quad \sigma_\Delta = 0.354, \quad 95\% \text{ CI} = 0.123. \quad (4)$$

Although there are some minor disagreements between the objective and subjective scores, in general there is a high degree of correlation.

C. Reproducibility

The benchmark against which we compare the performance of our neural classifier is the test/retest performance of the subjective Wilmer scheme. Shown in Fig. 7(a) is the test/retest performance of the Wilmer subjective grading scheme for the paired photographs of set 2. Here, variability arises from the human evaluation and from variability between the photographs of the same eye. The corresponding results for the neural network are shown in Fig. 7(b). Relative test/retest performances of the two grading schemes were as follows:

Subjective

$$\text{slope} = 0.994,$$

$$r = 0.962, \quad n = 48,$$

$$|\Delta| \leq 0.8, \quad \sigma_\Delta = 0.323, \quad 95\% \text{ CI} = 0.094$$

Objective

$$\text{slope} = 1.003,$$

$$r = 0.990, \quad n = 48,$$

$$|\Delta| \leq 0.448, \quad \sigma_\Delta = 0.154, \quad 95\% \text{ CI} = 0.045. \quad (5)$$

Note that the classification performance of the objective classifier versus the subjective classification system (Fig. 6) is comparable to that of the subjective system against itself [Fig. 7(a)]. This is further evidence of the validity of the objective system. Relative performances of the subjective and objective classifiers are further described in Table 1.

Principal causes of the subjective test/retest variabilities are the grading process itself and variabilities in the photographs. The global standard deviation of subjective grading differences, 0.254, (see Subsection 3.A) represents the majority proportion of the test/retest standard deviation of 0.323, thus suggesting that the variability in the photographs is rather low. Objective test/retest variations are due to variabilities between the photos and to inconsistencies within the training set. The worst case is when one observes identical objective feature vectors for photographs that nevertheless have been given

Fig. 6. (a) Classification performance of reinforcement neural network on first half of photograph set 2 (test set). (b) Classification performance of reinforcement neural network on second half of photograph set 2 (retest set).

Fig. 7. (a) Test/retest reproducibility of Wilmer system for photographs of set 2. (b) Test/retest performance of neural network for photographs of set 2.

Table 1. Comparison of Confidence Intervals on Test-Retest Differences

Grade Interval	Subjective 95% CI (<i>n</i>)	Objective 95% CI (<i>n</i>)
(0,1]	0.195 (12)	0.104 (11)
(1,2]	0.215 (11)	0.115 (12)
(2,3]	0.268 (11)	0.094 (11)
(3,4]	0.145 (14)	0.086 (14)
(0,4]	0.094 (48)	0.045 (48)

different subjective grades. On the basis of this analysis, it is not possible to partition the objective test/retest difference standard deviation, 0.154, between these two sources.

D. Sensitivity

As illustrated in Figs. 7(a) and (b) and in Table 1, the global as well as within-integer-grade confidence intervals for the objective classifier are approximately half those of the subjective. While this global comparison suggests that the objective system provides better reproducibility, these results are incomplete. To proceed, we wish to explore the issue of sensitivity. To do this, we inspect the test/retest reproducibility at the within-integer grade

level (Fig. 8 and Table 2). In Fig. 8, each line represents a linear least-squares fit between the test and retest scores within an integer grade range. As seen in the results for subjective assessment, the within-integer-grade scores have sufficient noise that except for the highest interval (3,4], there is little relationship between the grades. For the objective classifier, however, all within-integer-grade correlations are high and the slopes are essentially unity. The subjective assessment system, except at the highest opacification levels, is less able to track changes that are smaller than approximately an integer grade; small changes are masked by assessment noise. On the other hand, the lower noise (better repro-

Fig. 8. (a) Within-integer-grade test/retest reproducibility of Wilmer system for photographs of set 2. (b) Within-integer-grade test/retest performance of neural network for photographs of set 2.

Table 2. Comparison of Within-Grade Performances

Interval	Subjective		Objective	
	<i>r</i> (<i>n</i>)	Slope	<i>r</i> (<i>n</i>)	Slope
(0,1]	0.103 (11)	0.133	0.653 (10)	1.066
(1,2]	0.512 (12)	0.584	0.949 (13)	0.995
(2,3]	0.121 (10)	0.231	0.928 (12)	1.004
(3,4]	0.721 (15)	1.001	0.843 (13)	1.001

ducibility) of the objective classifier suggests that many small changes can be detected.

4. DISCUSSION

There are a number of important issues in establishing the validity of our classifier; some are related to the network, but others are related to the subjective assessment scheme itself. Regarding the classifier, we need to establish first, the adequacy of the training set and second, the adequacy of the features extracted from the imagery.

For the subjective assessment scheme, which serves as the basis of network training, we note that there is no "zero" standard. Thus the photointerpreter must extrapolate beyond standard photograph "one" toward "zero" to a precision of 0.1. This may be one source of the lack of reproducibility at the lowest levels of opacification [see Fig. 8(a)]. Further, in the subjective scheme there is no possibility of a grade beyond 4. Clearly, it is possible for a given eye to display a greater degree of opacification than our standard photograph 4. Nevertheless, such a photograph is simply judged a 4. In a clinical setting this is not an important distinction, since a class 4 cataract has a severe impact on vision. However, it introduces a certain inconsistency into the training set, as there are no limitations placed on the possible values of the features that are used for classification.

Other aspects of the adequacy of the training set are whether the training set possesses enough diversity in terms of photographic quality, exposure level, focus, fixation, etc., to span the entire range of new photographs to be assessed.

Finally, we address the issue of adequacy of the features extracted from the imagery. As noted above, the features that we have chosen have relevance in terms of the scatter and absorption processes taking place inside the lens. One could take our approach further and fit a higher-order polynomial (and use the parameters of the fit) or even use the profile itself²⁹ to train a neural network. The trade-off here is between the size of the training set versus the return in terms of robustness and sensitivity. Obviously, use of the profile itself would require a much larger training set. Further, in the use of the profile itself, it becomes less obvious what aspect of the profile is a measure of the actual impact on vision. This is an important issue in establishing more direct links between objective measures of impact on vision and classification of severity.

Another feature that can be explored is that of color. It is well recognized that the different wavelengths are scattered and absorbed to varying degrees.^{30,31} For instance, red is more strongly scattered, while blue is more strongly absorbed. This is an important feature, for instance in quantifying degree of brunescence, but a more problematic one. The issue here is that color is more highly variable in the photographic process. Therefore one must pursue the use of some measure of relative color. For the purposes of this effort, we have avoided the issue of color because we wished to demonstrate the validity of an autonomous classifier in comparison with an established clinical grading scheme in which color is

given a separate grade. It is clear that additional features such as color are desirable.

Many objective classification systems for nuclear opacities³² developed thus far have been based on an analysis of the gray levels of a defined area in the nuclear lens. Early researchers incorporated a standard density device into the camera equipment, which was then imaged onto the film.³³ This is an ideal approach that provides the means of performing absolute densitometry, and it does obviate problems due to variable illumination levels and photographic processing. However, it is not a feature on many standard slit-lamp cameras. Further, the results presented herein suggest that features associated with the relative spatial structure may be more important for classification than is absolute density.

More recently, slit-lamp imagery has been digitized and subjected to densitometric analysis along a slice through the nucleus³⁴ or within a square or rectangular area.^{15,35} The distinguishing feature of our efforts is the use of second-order statistics extracted from these gray-level profiles and the close agreement with clinical assessment.

In summary, researchers in the area of cataract classification systems are moving toward development of objective methodologies that use digital imagery to enhance sensitivity to change and to improve costs and efficiencies in classification.³⁶ We have developed, tested, and used our subjective classification scheme in a number of research projects^{10,37,38} and are well aware of the limitations.¹⁴ Herein we have described the current state of our efforts toward the development of an objective classification system that corrects deficiencies in other approaches. We note that there are three major issues in developing an automated classifications system: validity, robustness, (reproducibility in the face of variable conditions—illumination, film processing, patient fixation, etc.) and sensitivity. A high correlation between subjective and objective grades as evidenced by the training results (Fig. 5) demonstrates validity. Performance analyses shown in Figs. 6–8 and Tables 1 and 2 demonstrate superior reproducibility of the objective classifier. To emphasize the significance of this point we distinguish between inherent and effective sensitivity. The subjective system has inherent sensitivity as a result of the human observer's ability to detect slight nuances in the images. However, this inherent sensitivity can be masked by the noise (lack of reproducibility) associated with rendering a grade. As a result, the effective sensitivity of the subjective assessment system can be rather low. On the other hand, the objective system's much better reproducibility suggests that the effective sensitivity would surpass that of the subjective system. These results are suggestive but not conclusive evidence that the objective classifier has superior sensitivity. Further tests to confirm this are currently under way.

5. ACKNOWLEDGMENTS

We are grateful to Beatriz Muñoz for reviewing this paper and providing helpful suggestions. This effort was supported by National Institutes of Health Grant

1R01EY10857-02. S. K. West is a Research to Prevent Blindness Senior Scientific Investigator.

Address correspondence to D. D. Duncan, The Johns Hopkins University Applied Physics Laboratory, Johns Hopkins Road, Laurel, Maryland 20723-6099. Tel: 301-953-6000, ext. 8548; fax: 301-953-6779; e-mail: donald.duncan@jhuapl.edu.

REFERENCES

1. B. Thylefors, "The World Health Organization's programme for the prevention of blindness," *Int. Ophthalmol.* **14**, 211–219 (1990).
2. E. P. Steinberg, J. C. Javitt, P. D. Sharkey, A. Zuckerman, M. W. Legro, G. F. Anderson, E. B. Bass, and D. O'Day, "The content and cost of cataract surgery," *Arch. Ophthalmol.* **111**, 1041–1049 (1993).
3. U. S. Department of Health Services, Public Health Services, National Institutes of Health, National Eye Institute, *Vision Research: A National Plan: 1994–1998, A Report of the National Advisory Eye Council* (National Institutes of Health, Bethesda, Md., 1993), pp. xvi–xvii.
4. U. S. Department of Health Services, Public Health Services, National Institutes of Health, National Eye Institute, *Vision Research: A National Plan: 1994–1998, A Report of the National Advisory Eye Council* (National Institutes of Health, Bethesda, Md., 1993), p. 195.
5. L. T. Chylack, Jr., M. C. Leske, D. McCarthy, P. Khu, T. Kashiwagi, and R. Sperduto, "Lens opacities classification system II (LOCS II)," *Arch. Ophthalmol.* **107**, 991–997 (1989).
6. B. E. K. Klein, Y. L. Magli, M. Neider, and R. Klein, *Wisconsin System for Classification of Cataracts from Photographs*, NTIS Accession No. PB90-138306. (Available from National Technical Information Services, 5285 Port Royal Rd., Springfield, Va. 22161.)
7. S. K. West, F. S. Rosenthal, H. S. Newland, and H. R. Taylor, "Use of photographic techniques to grade nuclear cataracts," *Invest. Ophthalmol. Visual Sci.* **29**, 73–77 (1988).
8. H. R. Taylor and S. K. West, "A simple system for clinical grading of lens opacities," *Lens Res.* **5**, 175–181 (1988).
9. N. A. P. Brown, A. J. Bron, W. Ayliffe, J. Sparrow, and A. R. Hill, "The objective assessment of cataract," *Eye* **1**, 234–246 (1987).
10. H. R. Taylor and S. K. West, "The clinical grading of lens opacities," *Austr. N. Z. J. Ophthalmol.* **17**, 81–86 (1989).
11. M. C. Leske, L. T. Chylack, Jr., R. Sperduto, P. Khu, S.-Y. Wu, D. McCarthy, and the Lens Opacities Case-Control Study Group, "Evaluation of a lens opacities classification system," *Arch. Ophthalmol.* **106**, 327–329 (1988).
12. B. E. K. Klein, R. Klein, K. L. P. Linton, Y. L. Magli, and M. W. Neider, "Assessment of cataracts from photographs in the Beaver Dam eye study," *Ophthalmology* **97**, 1428–1433 (1990).
13. H. R. Taylor, J. A. Lee, F. Wang, and B. Muñoz, "A comparison of two photographic systems for grading cataracts," *Invest. Ophthalmol. Visual Sci.* **32**, 529–532 (1991).
14. S. K. West, B. Muñoz, F. Wang, and H. R. Taylor, "Measuring progression of lens opacities for longitudinal studies," *Curr. Eye Res.* **12**, 123–132 (1992).
15. D. M. Singer, J. K. Wolfe, and L. T. Chylack, Jr., "Reproducibility of image analysis techniques used to assess cataract," *Invest. Ophthalmol. Visual Sci. (Suppl.)* **34**, 882 (1993).
16. L. T. Chylack, Jr., J. K. Wolfe, and D. M. Singer, "Relative power of subjective and objective methods of quantitating cataract *in vivo*," *Invest. Ophthalmol. Visual Sci. (Suppl.)* **34**, 881 (1993).
17. J. C. Russ, *The Image Processing Handbook*, 2nd ed. (CRC, Boca Raton, 1995), Chap. 1, 36–40.
18. P. Buser and M. Imbert, *Vision* (MIT Press, Cambridge, Mass., 1993), Chap. 2, pp. 53–55.
19. A. Ishimaru, *Wave Propagation and Scattering in Random Media, Vol. 1: Single Scattering and Transport Theory* (Academic, San Diego, Calif., 1978), Chap. 3, pp. 63–64.
20. K. Sasaki, K. Fujisawa, and Y. Sakamoto, "Quantitative evaluation of nuclear cataract using image analysis," *Ophthalmic Res. Suppl* **1**, 26–31 (1992).
21. P. M. Khu and T. Kashiwagi, "Quantitating nuclear opacification in color scheinplufug photographs," *Invest. Ophthalmol. Visual Sci.* **34**, 130–136 (1993).
22. R. Greenler, *Rainbows, Halos, and Glories* (Cambridge U. Press, Cambridge, 1989), Chap. 5, pp. 129–131.
23. J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, 1991), Chap. 7, pp. 188–196.
24. S. Haykin, *Neural Networks* (Macmillan, Englewood Cliffs, 1994), Chap. 2, pp. 59–65.
25. J. A. Lee and H. R. Taylor, "The lens opacity meter: a method of quantification of lens opacity by measurement of scattering of incident light," *Lens Toxicity Res.* **7**, 31–38 (1990).
26. P. Rouhiainen, H. Rouhiainen, I.-L. Notkola, and J. T. Salonen, "Comparison of the lens opacities classification system II and lensmeter 701," *Am. J. Ophthalmol.* **116**, 617–621 (1993).
27. G. Maraini, F. Rosmini, P. Graziosi, M. C. Tomba, M. Bonacini, R. Cotichini, P. Pasquini, R. D. Sperduto, and the Italian American Cataract Study Group, "The influence and severity of pure forms of age-related cataract on visual acuity and contrast sensitivity," *Invest. Ophthalmol. Visual Sci.* **35**, 262–267 (1993).
28. M. Karbassi, P. C. Magnante, J. K. Wolfe, and L. T. Chylack, Jr., "Objective line spread function measurements, Snellen acuity, and LOCS II classification in patients with cataract," *Optom. Vis. Sci.* **70**, 956–962 (1993).
29. P. A. Edwards, M. B. Datiles III, M. Unser, B. L. Trus, V. Freidlin, and K. Kashima, "Computerized cataract detection and classification," *Curr. Eye Res.* **9**, 517–525 (1990).
30. J. H. Seland, L. T. Chylack, Jr., and J. K. Wolfe, "Indirect spectral transmission ratio measurements of the aging crystalline lens nucleus," *Acta Ophthalmol.* **70**, 376–382 (1992).
31. J. Pokorny, V. C. Smith, and M. Lutze, "Aging of the human lens," *Appl. Opt.* **26**, 1437–1440 (1987).
32. M. B. Datiles, "Clinical Evaluation of Cataracts," in *Duanes Clinical Ophthalmology*, W. Tasmin and E. A. Jaeger, eds. (Lippincott, Philadelphia, Pa., 1992), Chap. 73B.
33. N. Brown, "Quantitative slit-lamp photography of the lens," *Trans. Ophthalmol. Soc. U.K.* **92**, 303–307 (1972).
34. P. A. Edwards, M. B. Datiles, and S. B. Green, "Reproducibility study on the Scheimpflug cataract video camera," *Curr. Eye Res.* **7**, 955–960 (1988).
35. L. T. Chylack, Jr., J. K. Wolfe, J. Friend, P. M. Khu, D. M. Singer, D. McCarthy, J. del Carmen, and B. Rosner, "Quantitating cataract brunescence, the Harvard and LOCS systems," *Optom. Vis. Sci.* **70**, 886–895 (1993).
36. M. B. Datiles III, B. V. Magno, and V. Freidlin, "Study of nuclear cataract progression using the National Eye Institute Scheimpflug System," *Br. J. Ophthalmol.* **79**, 527–534 (1995).
37. S. K. West, B. Muñoz, E. A. Emmett, and H. R. Taylor, "Cigarette smoking and risk of nuclear cataracts," *Arch. Ophthalmol.* **107**, 1166–1169 (1989).
38. S. Vitale, S. K. West, J. Hallfrisch, C. Alston, F. Wang, C. Moorman, D. Muller, V. Singh, and H. R. Taylor, "Plasma antioxidants and risk of cortical and nuclear cataract," *Epidemiology* **4**, 195–203 (1993).