

Journal of Biomedical Optics

SPIDigitalLibrary.org/jbo

Identification of suitable fundus images using automated quality assessment methods

Uğur Şevik
Cemal Köse
Tolga Berber
Hidayet Erdöl

Identification of suitable fundus images using automated quality assessment methods

Uğur Şevik,^{a,*} Cemal Köse,^b Tolga Berber,^a and Hidayet Erdöl^c

^aKaradeniz Technical University, Department of Statistics and Computer Science, Faculty of Science, Trabzon 61080, Turkey

^bKaradeniz Technical University, Department of Computer Engineering, Faculty of Engineering, Trabzon 61080, Turkey

^cKaradeniz Technical University, Department of Ophthalmology, Faculty of Medicine, Trabzon 61080, Turkey

Abstract. Retinal image quality assessment (IQA) is a crucial process for automated retinal image analysis systems to obtain an accurate and successful diagnosis of retinal diseases. Consequently, the first step in a good retinal image analysis system is measuring the quality of the input image. We present an approach for finding medically suitable retinal images for retinal diagnosis. We used a three-class grading system that consists of good, bad, and outlier classes. We created a retinal image quality dataset with a total of 216 consecutive images called the Diabetic Retinopathy Image Database. We identified the suitable images within the good images for automatic retinal image analysis systems using a novel method. Subsequently, we evaluated our retinal image suitability approach using the Digital Retinal Images for Vessel Extraction and Standard Diabetic Retinopathy Database Calibration level 1 public datasets. The results were measured through the F1 metric, which is a harmonic mean of precision and recall metrics. The highest F1 scores of the IQA tests were 99.60%, 96.50%, and 85.00% for good, bad, and outlier classes, respectively. Additionally, the accuracy of our suitable image detection approach was 98.08%. Our approach can be integrated into any automatic retinal analysis system with sufficient performance scores. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.19.4.046006]

Keywords: computer-assisted image analysis; automated image quality assessment; suitable retinal image frame; retinal image analysis; machine learning.

Paper 130804RR received Nov. 8, 2013; revised manuscript received Mar. 13, 2014; accepted for publication Mar. 14, 2014; published online Apr. 9, 2014.

1 Introduction

Medical imaging is an essential phase of automated medical diagnosis and decision support systems. In particular, ophthalmologists use medical imaging systems to evaluate the retinal diseases intensively.¹ Retinal imaging systems are based on fundus images that are acquired using a predefined field-of-view (FOV). However, some fundus images are of a medically unsatisfactory quality caused by insufficient contrast rate, blurring, incorrect focus, and frame inputs. Automated systems are unable to evaluate such images medically and cause ophthalmologists to waste precious time attempting to interpret the images of poor quality. Hence, the image quality becomes a crucial property of retinal images and must be assessed initially in a retinal imaging system.

Studies on image quality assessment (IQA) can be categorized into two groups: segmentation-based retinal IQA and histogram-based retinal IQA. The goal of segmentation-based IQA methods is to identify the high-quality images that have more separable lesions and anatomical structures than the low-quality images.² For instance, Fleming et al.³ proposed a segmentation-based IQA method based on the detection of the capillary vessel structure in the macula, whereas Giancardo et al.⁴ measured the vessel distinctive in different retinal image regions for IQA. Histogram-based IQA methods aim to detect nonuniform gray level histograms of retinal images because the gray level histograms of the low-quality images are more skewed than the high-quality images. For example,

Lalonde et al.⁵ used the gradient magnitude of image histograms and local histogram information to grade image quality. Lee and Wang⁶ defined a quality index based on the convolution of a high-quality image histogram and evaluated the image histogram. Further, Niemeijer et al.⁷ classified retinal image quality into five classes using image structure clustering and the distribution of image brightness.

Although each method obtains the great performance scores, it is extremely difficult to compare the results of one method with the results of another. The reason is that each method might use different datasets or quality metrics to measure its performance. Furthermore, there is no widely accepted retinal image quality scale that exists in the literature. Additionally, each method uses either segmentation of retinal structures or retinal image histograms. Both approaches can increase the false positive rate because the segmentation-based approaches can be error-prone and histogram-based approaches discard spatial information.²

In this paper, we present an approach for finding medically suitable retinal images (MSRIs) that should contain the optic disc (OD), the macula, and the vascular arch to identify the exact position of lesions caused by any retinal disease on the posterior pole. Our approach aims to combine the segmentation method with global image information in order to eliminate the drawbacks of segmentation-based image quality methods. In other words, we propose a hybrid retinal IQA approach to detect the ideal retinal images to be used for the retinal diagnostic

*Address all correspondence to: Uğur Şevik, E-mail: usevik@ktu.edu.tr

process. In addition, we have developed a new public retinal IQA dataset.

2 Materials and Methods

We require a retinal image quality dataset to measure the performance of our IQA approach. Although several retinal imaging datasets exist in the literature, we were required to create our own retinal image quality dataset because of the lack of quality information in the existing datasets. We created a new retinal image dataset called Diabetic Retinopathy Image Database (DRIMDB) in order to evaluate the performance of our retinal image quality approach. DRIMDB has three image quality classes: good, bad, and outlier. An expert annotated a total of 216 images into those quality classes. The total number of retinal images in each quality class is listed in Table 1. DRIMDB is publicly available at <http://isbb.ktu.edu.tr/multimedia/drimdb>.

All the images used in DRIMDB are obtained from the Retina Department of Ophthalmology, Medical Faculty, Karadeniz Technical University. Furthermore, all images were obtained with a Canon CF-60UVi Fundus Camera using 60 deg FOV and stored in JPEG files at 570×760 pixels resolution.

We designed our quality scale considering its simplicity and efficiency. In particular, we included an outlier class to identify the nonretinal images that could have been obtained for several reasons, such as wrong focus or patient absence. Then, we divided the images into two groups, which are good- and bad-quality images. It is important to note that retinal images with good quality in DRIMDB are medically suitable image candidates.

We developed a two-step IQA approach to find MSRIs. The first step of our approach aims to identify the good retinal images, which are worthy of further analysis. The second step of our approach identifies medically suitable images among the good quality retinal images. In the first step, we used DRIMDB to measure the classification performance of our approach. In addition, we used two additional public retinal image datasets [Digital Retinal Images for Vessel Extraction (DRIVE) and Standard Diabetic Retinopathy Database Calibration level 1 (DIARETDB1)^{8,9}] in our experiments to identify MSRIs. Given that medical experts utilize these public datasets, we assume that each image is a medically suitable image for our experimental setups.

The details of our approach are shown in Fig. 1 and are described in detail in the subsequent sections. Briefly, our approach consists of six phases. Initially, we enhanced an image's visual information using image processing techniques to increase the visibility of crucial anatomical structures. This was followed by the feature extraction phase. Then, the image features were classified using machine learning methods to grade retinal images according to our quality scale. Finally, MSRIs were identified within the images classified as being of good quality.

Table 1 Image distribution of DRIMDB.

Class name	Total number of images
Good	125
Bad	69
Outlier	22
Total	216

2.1 Preprocessing

The visibility and clarity of blood vessels play an important role in assessing the retinal image quality. We applied image processing techniques to retinal images in order to enhance the visibility of blood vessels. Retinal images are generally stored in JPEG files using the RGB color space. According to many works in this area,¹⁰ the green channel of the retinal images contains the most valuable data because of its informative contrast rate. Hence, we extracted the green channel of the retinal image initially [Fig. 2(a)]. We inverted the gray level values of the green channel data to improve the visibility of blood vessels. Then, we stretched the contrast of the inverted green channel image. Given that the concavity of the posterior pole causes intensity variations in retinal images, we applied contrast-limited adaptive histogram equalization (CLAHE)¹¹ to adjust the intensity distribution uniformity of the retinal image [Fig. 2(b)].

2.2 Segmentation of Retinal Blood Vessels

The preprocessing goal is to improve the visibility of blood vessels in the segmentation process. We preferred to use Zana and Klein's vessel segmentation approach¹² because of its simplicity and robustness. The method has a high accuracy and efficiency

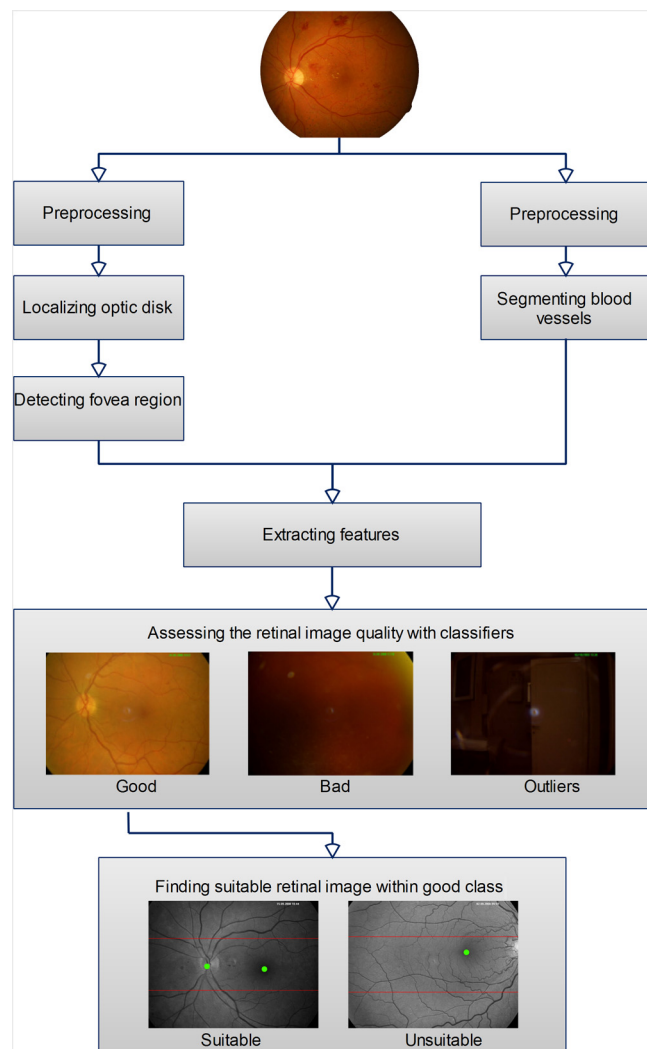


Fig. 1 Block diagram of our approach.

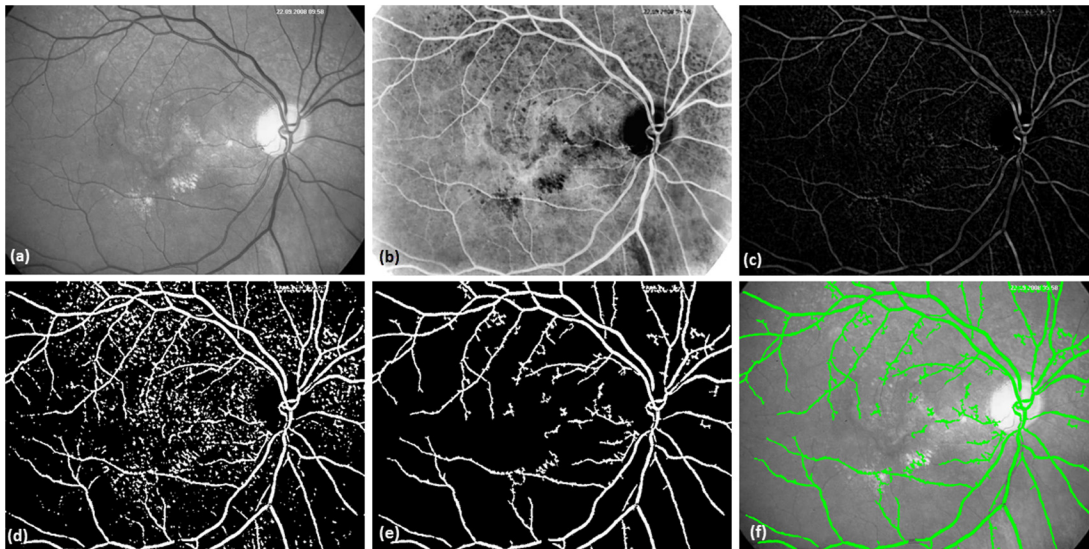


Fig. 2 Blood vessel segmentation phases: (a) original green channel retinal image, (b) applying contrast-limited adaptive histogram equalization transform to inverse of green channel retinal image, (c) applying the morphological methods, (d) binarization of (c) with a triangle automatic threshold method, (e) removal of small region without vessels, and (f) overlaying the segmented vessels on the original green channel retinal image.

on the vessel segmentation process in retinal images. In detail, the method uses simple morphological operators to extract blood vessels of a retinal image.

2.3 OD Detection

One of the most important steps of our approach is the OD detection step because it plays an important role in both retinal IQA and identification of MSRI processes.^{13,14} First, we used the visibility of the OD to identify the good quality images because these images should contain the OD properly. Afterward, we determined that the location of the OD should be within a specific boundary in order to diagnose the retinal diseases on the posterior pole.

The OD can be detected easily in a retinal image because of its unique features, such as high-intensity value and circularity. However, retinal images might contain such abnormalities that could be extremely similar to the OD. A simple threshold segmentation method might produce inaccurate results in this case.

In our method, we used the vertical edge information to detect the OD because the OD is the vertical junction of vessels in a regular retinal image [Fig. 3(a)]. Therefore, it is expected that there should be several vertical edges around the OD. We used a vertical Sobel operator to extract the vertical edge information of preprocessed retinal images [Fig. 3(b)]. We calculated the vertical edge histogram using the following formula

$$VEH_j = \sum_{i=0}^n VI(i, j), \quad (1)$$

where VEH denotes vertical edge histogram, VI is the result of the vertical Sobel operator, n denotes the total number of rows in VI, and i and j denote the row and column numbers of this histogram, respectively. The maximum value index of this histogram is an approximation of the horizontal center of the location of the OD. Therefore, we were able to reduce our OD search area from the entire retinal image to a neighborhood of two OD diameters of maximum value index. In other words, the

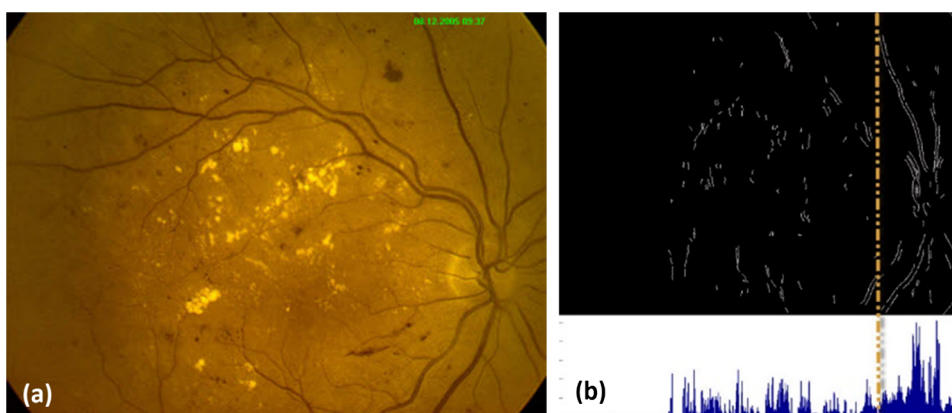


Fig. 3 Detecting candidate area of the optic disk: (a) original image, (b) vertical edge histogram, and candidate optic disk (OD) region of the image.

horizontal center of the OD should remain within $[\min_x, \max_x]$, where $0 < \min_x < \max_x$, $\min_x = \operatorname{argmax}_j \{ \text{VEH}_j \} - 2 \times r(\text{OD})$, and $\max_x = \operatorname{argmax}_j \{ \text{VEH}_j \} + 2 \times r(\text{OD})$ [Fig. 3(b)].

Once the candidate region was determined, we extracted the candidate region image from the green channel image and adjusted the gray scale values of the target region [Figs. 4(a) and 4(b)]. We used the Canny edge detection operator to find edges in the target region [Fig. 4(c)]. Successively, we used the Hough transform to find circles in the edge image [Fig. 4(d)]. The Hough transform is an accumulation method defined as

$$H(x_c, y_c, r) = \sum_{i=0}^n \sum_{j=0}^m g(i, j, x_c, y_c, r), \quad (2)$$

where (x_c, y_c) is the center point of the circle being tested, r is the radius of circle, n and m are the total number of rows and the total number of columns of the edge image, respectively, and $g(i, j, x_c, y_c, r)$ is a binary function defined as

$$g(i, j, x_c, y_c, r) = \begin{cases} |\text{EI}(i, j)|, & r^2 = (i - x_c)^2 + (j - y_c)^2 \\ 1, & \text{otherwise} \end{cases}, \quad (3)$$

where $\text{EI}(i, j)$ is the pixel value of the edge image at the (i, j) point.¹⁵ We selected the maximum accumulator value of the Hough transform as the OD.¹⁶ The result of our OD detection approach is depicted in Fig. 4(e).

2.4 Fovea Detection

The fovea, which is the darkest part of a normal retinal image, is located at the center of the macula. In a formal manner, the location of the fovea remains within a ring between the two circles that have the same center as the OD and that have the radius of two and three OD diameters, respectively.¹⁷ Hence, we created a ring-shaped mask to locate the fovea that is depicted in Fig. 5(b). We used morphological closing to eliminate the side effects of retinal diseases and to increase the separability of the fovea. The best approximation of the fovea center is the location of the darkest point in the mask. Figure 5(c) shows the identified fovea and its neighborhood in the mask. Finally, we were

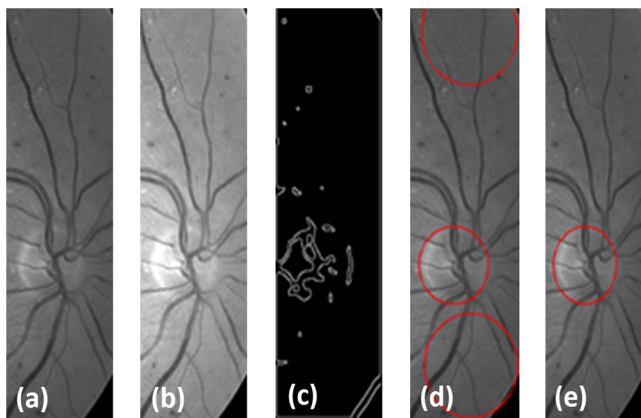


Fig. 4 Optic disc detection steps: (a) candidate region image of OD, (b) adjusted image, (c) edges of the region, (d) possible circles, and (e) detected OD.

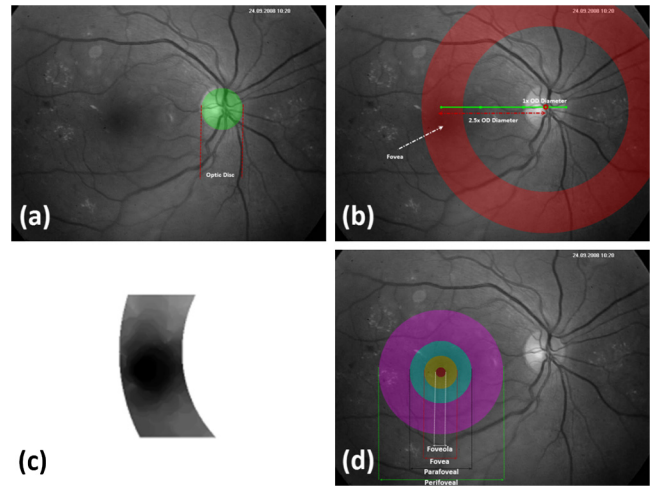


Fig. 5 Fovea detection: (a) localized optic disc, (b) finding the candidate fovea area, (c) localization of fovea region, and (d) finding the macula components.

able to locate the rest of the fovea components as shown in Fig. 5(d).¹⁸

2.5 Feature Extraction for Retinal IQA

In this work, we extracted a total of 177 features from a retinal image in order to determine the quality of an image. Our features can be categorized into three groups: shape, texture, and intensity.

Shape features aim to represent the shape information of a segmented binary image. Given that we previously segmented a retinal image, we can use the shape features of the segmented image to express the vessel structure of the retina. We chose two shape features: one is called the inferior, superior, nasal, and temporal (ISNT) quadrant features¹⁹ and the other is the Zernike moments,²⁰ which are expected to have high descriptive power. Moreover, the ISNT quadrant features and the Zernike moments describe the local and global properties of the vessel structure, respectively.

Our goal is to identify the retinal images, as well as the non-retinal images obtained from fundus cameras. It is obvious that the texture of both types of images should be different from each other. Hence, we used texture features to distinguish retinal images from nonretinal images. Texture features model the repeating patterns of a local pixel intensity variation in an image. In this work, we used Haralick texture features, which aim to describe the texture globally,²¹ and the edge histogram, which expresses the texture locally.²²

Finally, we used two color-related features to represent the color information of the images. The first feature is concerned with the pixel intensity distribution properties of a retinal image. It is expected that the histogram distribution of the high quality retinal image should be similar to a normal distribution. Therefore, we calculated the high-order statistical properties of the gray level histogram to represent the intensity distribution of a retinal image. The second feature represents the dominant color of the image in both the RGB and gray scale color spaces.

2.5.1 Regional features of ISNT quadrants

The ISNT quadrants features represent the vessel structure of a retinal image in the four ISNT regions.¹⁹ The vessel information

at each region is represented by shape descriptors. Hence, we extracted localized shape properties of vessel distribution in a retinal image with the help of these features.

After segmenting the blood vessels, we created four masks to extract the ISNT regions. We calculated the area, major and minor axis lengths, eccentricity, orientation, Euler number, circle diameter, solidity, extent, and the perimeter properties of each region. In addition, high-quality retinal images should have a uniform vessel distribution among these quadrants. Consequently, we calculated two additional features from the ISNT quadrants, which are given in the following formulas:

$$\begin{aligned}
 R_{IS} &= \frac{A(I)}{A(S)}, \\
 A(I) &= \sum_{(x,y) \in I} I(x,y), \\
 A(S) &= \sum_{(x,y) \in S} S(x,y),
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 R_{ISNT} &= \frac{A(I) + A(S)}{A(N) + A(T)}, \\
 S(N) &= \sum_{(x,y) \in N} N(x,y), \\
 A(T) &= \sum_{(x,y) \in T} T(x,y).
 \end{aligned} \tag{5}$$

Figure 6 shows a graphical representation of the ISNT features.

2.5.2 Zernike moments

Zernike polynomials²⁰ are used to extract the global shape information of an image. They provide robust shape information that is invariant to rotation and scale. In mathematics, the Zernike moments are orthogonal moments that use the unit vector representation of an image. They are denoted as in the following formula:

$$\begin{aligned}
 A_{nm} &= \frac{n+1}{\pi} \sum_x \sum_y I(x,y) V_{nm}^*(\rho, \theta), \quad n > 0, \\
 n - |m| &\text{ is even, } |m| < n,
 \end{aligned} \tag{6}$$

where $|\cdot|$ denotes the absolute value of a real number, ρ is the length of the vector from the origin to point (x, y) , θ is the angle from the x -axis to the vector, and $A_{nm}^* = A_{n,-m}$. Here, $V_{nm}^*(\rho, \theta)$ are the Zernike polynomials and are denoted as in the following formula:

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jm\theta}, \tag{7}$$

where

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s}. \tag{8}$$

We used the first 15 Zernike polynomials to assess the global quality of retinal images in this work because they provide higher accuracy rates for shape-based classification tasks.

2.5.3 Gray level co-occurrence matrix features

The features of the gray level co-occurrence matrix (GLCM) were proposed by Haralick et al.²¹ It is intensively used in many texture-related studies. These features describe the texture using spatial intensity dependencies. In the original work, texture was defined by calculating the statistical properties of GLCM that represent intensity value co-occurrences at d -distance and θ angle. As a result, GLCM is a square matrix whose rows and columns represent gray level intensity values. In other words, the size of the GLCM matrix depends on the number of gray levels in the image. The GLCM matrix is calculated using the following formula:

$$\begin{aligned}
 C_{\Delta x, \Delta y}(i, j) &= \sum_{p=1}^m \sum_{q=1}^n \\
 &\times \begin{cases} 1, & I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases},
 \end{aligned} \tag{9}$$

where $I(x, y)$ is the image function, m and n are the image width and height, and Δx and Δy are the distance parameters defined by d and θ , respectively.

After calculating the co-occurrence matrix, each element of the matrix is normalized using the following formula to create a normalized co-occurrence matrix $[p(i, j)]$

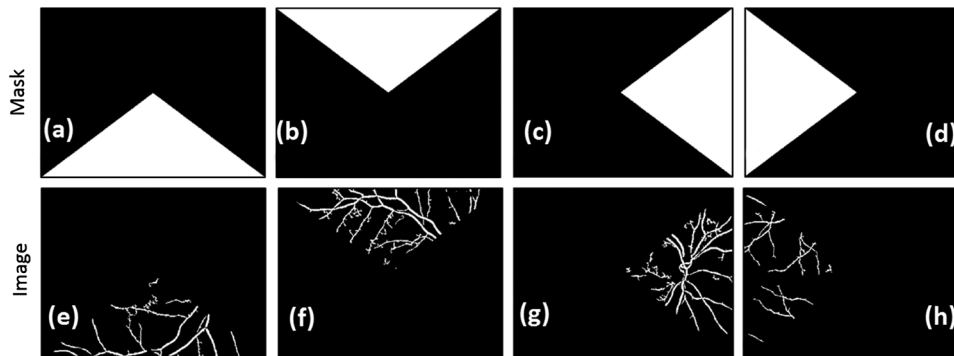


Fig. 6 Masks and images of ISNT quadrants: (a) inferior quadrant mask, (b) superior quadrant mask, (c) nasal quadrant mask, (d) temporal quadrant mask, (e) inferior vessel information, (f) superior vessel information, (g) nasal vessel information, and (h) temporal vessel information.

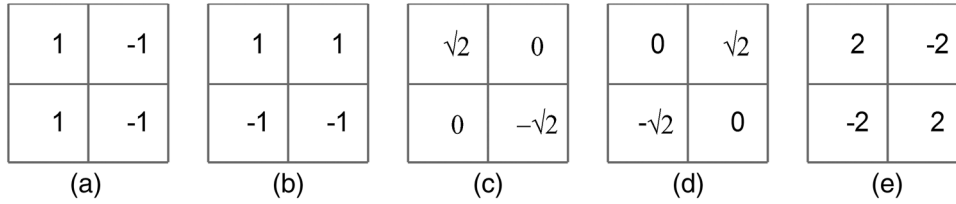


Fig. 7 Filters used to determine edges (a) horizontal edge filter, (b) vertical edge filter, (c) 45-deg edge filter, (d) 135-deg edge filter, and (e) nondirectional edge filter.

$$p(i, j) = \frac{C(i, j)}{\sum_{k=1}^L \sum_{l=1}^L C(k, l)}, \quad (10)$$

where L denotes the number of gray levels in the image and $p(i, j)$, which is also denoted as p_{ij} , is the (i, j) 'th element of the normalized co-occurrence matrix.

We used seven Haralick feature functions: contrast, correlation, sum of average, entropy, homogeneity, inverse difference moment, and maximum probability; they are computed at 0, 45, 90, and 135 deg angles for distance value one.

2.5.4 Edge histogram descriptor

Edge information in an image can describe both shape and texture. We used the edge histogram descriptor, which is an element of the MPEG-7 standard, to extract the primitive information of the edge distribution.^{23,24} First, the retinal image was divided into 16 (4×4) subimages. A histogram of five standardized edge directions, shown in Fig. 7, was computed for each subimage.

The edge histogram was calculated using the edge filter responses at each 2×2 nonoverlapping blocks of each subimage. A block was assigned to the edge whenever the filter response was a maximum value. Hence, the number of elements in the edge histogram feature is 80 (the union of five bin histograms of each 16 subimage).

2.6 Finding a Suitable Retinal Image

In this paper, we propose a method to determine MSRIs that include the OD, fovea, and superior and inferior arcades. MSRIs provide valuable information about the retina and are preferred by ophthalmologists. We call a retinal image an MSRI when it has both the OD and the fovea at a specific location

called the expected region of OD (EROD). An example of EROD is depicted in Fig. 8.

The EROD upper and lower bounds are calculated using the following formulas:

$$L_{EROD} = y_c + r(OD), \quad (11)$$

$$U_{EROD} = y_c - r(OD), \quad (12)$$

where y_c is vertical centerline of the retinal image, $r(OD)$ is the diameter of the OD, and L_{EROD} and U_{EROD} denote the lower and upper bounds of EROD, respectively. We called an image MSRI if it has both the OD and the fovea within EROD. Furthermore, we defined a new suitability metric (S) for retinal images, which is calculated using the following formula:

$$S = |y_c - y_{OD}|, \quad (13)$$

where y_{OD} is the y component of the OD position. Hence, our metric decreases when the center of the OD approaches the image's centerline.

2.7 Performance Measurement of Retinal IQA

We implemented our feature extraction code using the MATLAB software. We executed the DRIMDB feature extraction on two different computer configurations (PC1 and PC2). PC1 has Intel Core i7 3770 quad-core CPU with 16 GB of RAM, and PC2 has Intel Core i7 3930K six-core CPU with 16 GB of RAM. We repeated feature extraction experiments 10 times to eliminate the cache effect using serial and parallel implementations. Moreover, we calculated the speed-up and efficiency scores for parallel implementation of the feature

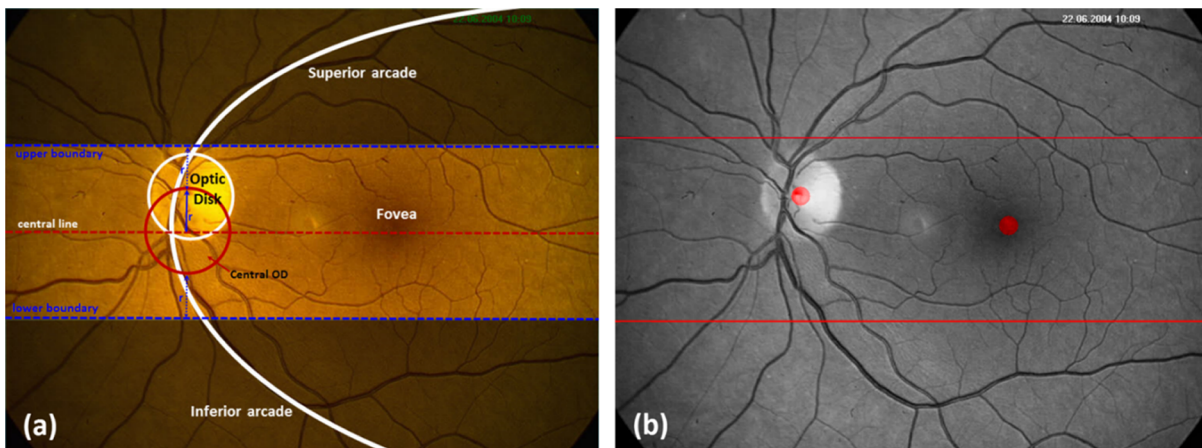


Fig. 8 Examples of EROD (a) optimal boundaries for EROD (b) an example of retinal image having OD and fovea in EROD.

Table 2 Time analysis for feature extraction of images in DRIMDB.

	PC1	PC2
Serial implementation	385.83 s	304.32 s
Parallel implementation	145.60 s	56.51 s
Speed-up	265.00%	538.49%
Efficiency	33.12%	44.87%

extraction process. The results of our experiments are listed in Table 2.

In this study, we measured the classification performance of the features described in Sec. 2.5 on the IQA task with the help of a data mining software.²⁵ In addition, we conducted experiments to identify the best machine classifier for the retinal IQA task among five well-known machine classifiers: the linear discriminant analyze (LDA) with principal component analysis (PCA),^{26,27} the k-Nearest Neighbor²⁸ (k-NN), the Neural Network (NN),²⁹ the Naïve Bayes,³⁰ and the support vector machines (SVMs).³¹ Some of these classifiers require parameters to be adjusted. We tested several parameter configurations to find the optimum parameter values. Furthermore, we used 10-fold cross-validation to ensure the objectivity of each experiment.

We used PCA to manage the singularity problem of the LDA classifier. However, we had to adjust the PCA component count of the LDA+PCA classifier. Our experiments showed that the optimum parameter value of the PCA component count is 77. It is also noticeable that the best *k* value for the k-NN classifier is found at one. We used one hidden layer for the NN classifier and calculated the hidden layer size using the input vector length and class count. We used Laplace correction to prevent the high influence of zero probabilities on Naïve Bayes experiments. The SVM classifier was the only binary classifier in our experiments. We used one-versus-all approach on experiments related with the SVM classifier, and our experiments show that the best *C* value for the SVM classifier is -1.0.

All our initial tests were performed on the combined vector of all the features explained in this paper. Classification performance can be degraded by feature vector dimension. This negative effect is known as the “curse of dimensionality.” There are two common methods to overcome this issue: selecting the most valuable features from the original subset and transforming the feature space into an effective low-dimensional one. These two methods are called feature selection and dimensionality reduction, respectively. Because we are considering a vector whose size is 177, we must ensure that our experiments are not affected by the curse of dimensionality side effect. We performed two feature selection approaches: a genetic algorithm³² and a forward selection³³ method. All the experiments were executed on PC1. The time comparison of the methods is listed in Table 3.

Most of the classification-related studies are concerned with assessing samples in two distinct classes called the binary classification systems. There are state-of-the-art evaluation metrics available in the literature to measure the performance of binary classification systems, such as receiver operating characteristics analysis, sensitivity, and specificity. However, our task requires evaluation metrics capable of measuring the performance of multiclass classifications. The most preferred metrics used for

Table 3 Time comparison of feature selection methods.

Classifier	All features test time (s)	Genetic algorithm		Forward selection	
		Feature vector size	Test time (s)	Feature vector size	Test time (s)
LDA	<1	85	<1	8	<1
k-NN (<i>k</i> = 1)	<1	177	<1	6	<1
Neural network	<1	98	<1	4	<1
Bayes	<1	84	<1	4	<1
SVM	<2	95	<1	10	<1

this task are precision and recall. Precision determines the rate of correctly predicted samples among classification results and recall determines the rate of correctly classified samples among actual samples (in some studies, recall is referred to as sensitivity).³⁴ However, these two metrics can show different characteristics in some experiments. Therefore, we chose the F1 measure, which is the harmonic mean of precision and recall, to measure the performance of our experiments.³⁴ As a result, we are able to identify the experimental setups obtaining high classification accuracy, as well as high predictive power. The formal definition of the F1 metrics is given in the following formula:

$$F1 = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

3 Results

Our initial experimental results are presented in Fig. 9. According to the results, the SVM classifier slightly outperformed other classifiers at experiments on good and outlier classes with 99.60% and 84.21% F1 scores, respectively, whereas the LDA classifier outperformed other classifiers at experiments on the bad class with 97.78% F1 score. Although

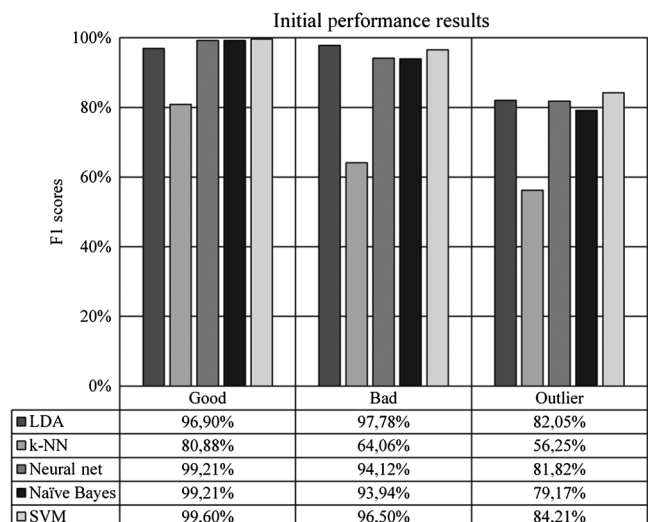


Fig. 9 Results of initial experiments.

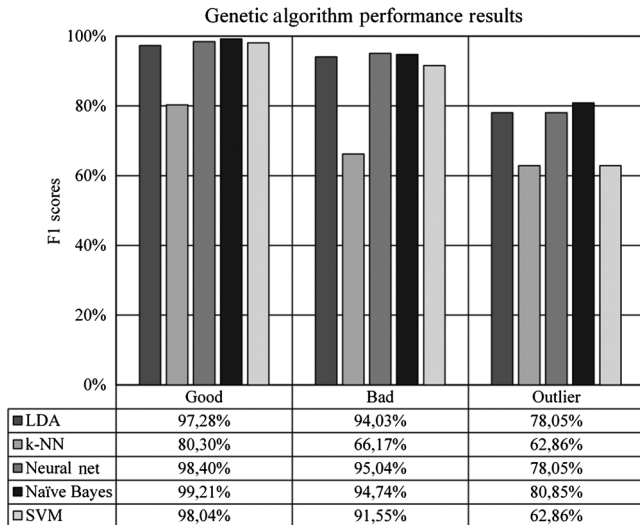


Fig. 10 Results of genetic feature selection experiments.

two different classifiers presented different characteristics on different classes, we suggest the use of the SVM classifier because of its overall classification success rate.

Furthermore, we tested the performance effect of the dimensionality reduction approaches on the IQA task. Figures 10 and 11 depict the classification results of the genetic and forward feature selection approaches, respectively. According to the results, the feature selection methods increase the success rates of probabilistic classifiers. For example, the classification performance of the Naïve Bayes classifier increases with the help of the feature selection, as does the LDA classifier. It is also noticeable that the performance of the k-NN classifier increases considerably using the forward selection approach. However, the performance of some classifiers degraded for specific classes, according to our results. In particular, the classification performance of the experiments on the outlier class had the worst performance scores at feature selection tests of the SVM and NN classifiers. The main cause of this issue is closely related to the number of samples that belong to the outlier class because it has the smallest sample size compared to the

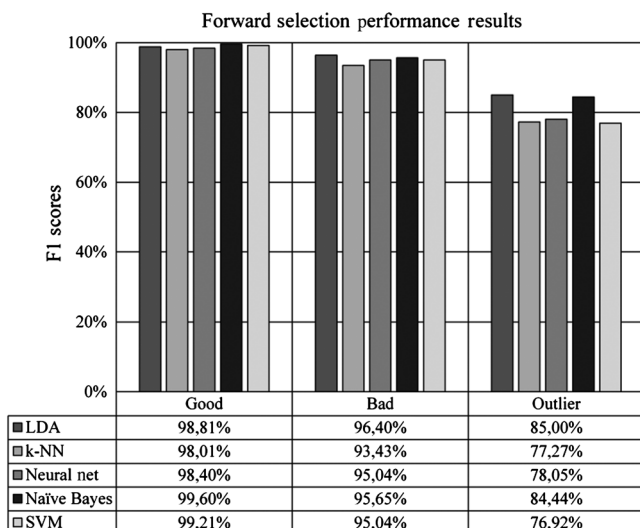


Fig. 11 Results of forward selection experiments.

other quality classes. Reducing the dimension might remove information, which is crucial for outlier detection. In addition, the feature selection approach increased the overall performance rates of the LDA classifier noticeably among other classifiers. The classification performance of the LDA classifier measured at 98.81%, 96.40%, and 85.00% for the good, bad, and outlier classes, respectively. Although individual classification performances of classifiers varied with the quality classes, the SVM classifier obtained the optimum classification performance scores without any feature selection method considering the overall classification performances.

Moreover, we performed additional tests to determine whether the image fits our MSRI condition. Hence, after assessing the quality of an image, we performed further experiments to measure the performance of our MSRI detection approach on DRIMDB, DRIVE, and DIARETDB1 and obtained 97.60%, 100.0%, and 96.63% of detection rates, respectively. The performance scores of our approach are listed in Table 4.

Figure 12 shows the successful and failed sample images for DRIMDB and DIARETDB1, respectively. Each image includes red dots that represent our OD and fovea detection results. Moreover, the EROD boundaries are presented in each image with red lines. Given that our approach depends on an accurate OD position, any image transformation related to the OD position will have a negative impact on our results. For example, the image frame of the DRIMDB failed sample includes an incomplete OD. Hence, our approach found the OD location incorrectly. In addition, our approach was unable to find fovea. Furthermore, the DIARETDB1 failed sample has extremely unbalanced spatial intensity distribution. So, our approach only found the location of the OD correctly. Besides, our approach could successfully find the exact location of the OD and fovea on the images given in Figs. 12(a) and 12(c).

Our approach requires good-quality retinal images to detect the fovea. We inspected each image in the DRIMDB for their qualities. However, we used each image in the DIARETDB1 directly in our MSRI detection approach. Because the DIARETDB1 has no quality information, our approach fails to detect the exact location of the fovea in images in the DIARETDB1 where there is an insufficient contrast rate.

4 Discussion

The most important anatomical structures of the retina are the vessels, OD, macula, and vascular arches. MSRIs should contain all the important retinal structures as much as possible and should ensure that those structures are as clear as possible. Successful retinal diagnosis depends on the clarity of the retinal structures. However, the clarity of the retinal image is insufficient without a proper image frame. In this paper, we presented a novel approach to identifying the retinal images with a good

Table 4 Performance scores of MSRI detection approach.

Databases	Total number of MSRIs	Total number of correctly classifier MSRIs	True MSRI detection rate (%)
DRIMDB	125	122	97.60
DRIVE	40	40	100.00
DIARETDB1	89	86	96.63

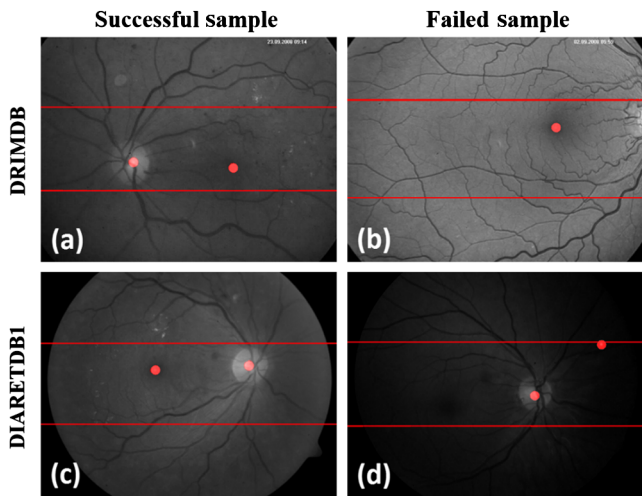


Fig. 12 Successful and failed samples from DRIMDB and DIARETDB1 (a) DRIMDB successful sample, (b) DRIMDB failed sample, (c) DIARETDB1 successful sample, and (d) DIARETDB1 failed sample.

clarity and a proper frame. Hence, automatic retinal image analysis systems can acquire high accuracy rates using these images.

In fact, several retinal images acquired in clinical systems have insufficient information to diagnose the retinal diseases. These images are useless for automated image analysis systems and occupy redundant disc space. This issue can be resolved by assessing the image quality with the help of automated IQA approaches, before such images are saved into automatic retinal image analysis systems. Therefore, the valuable time of ophthalmologists can be saved by eliminating the inadequate retinal images. In other words, IQA will improve both the quality and the time efficiency of the retinal image diagnostic process.

Most of the current studies measure the retinal image quality with the help of vessel densities or intensity variations.²⁻⁷ Additionally, some studies discard nonretinal image identification and are concerned only with retinal image quality. However, nonretinal images can have a negative impact on automatic retinal analysis systems and increase the rate of false positives. For instance, Fleming et al.³ and Paulus et al.² discarded outlier images and categorized retinal image quality. Paulus et al.² used Haralick et al.²¹ texture features and reported 92.7% of an area under curve (AUC) rate. Wen et al.³⁵ proposed an IQA approach using blood vessel information. However, they ignored the contrast and color saturation knowledge of retinal images. Fleming et al.³ proposed a complex metric table for IQA. Although they presented more quality classes than we do in this paper, Fleming et al.³ propose one bad class and more than one good classes. In other words, there is no outlier class that contains nonretinal images in their grading system. On the other hand, some works attempt to detect nonretinal images. For instance, Giancardo et al.⁴ include the outlier images in their study, but the success rate of their outlier detection is 80%, caused by detecting false vessel-like structures in the outlier images. Moreover, some retinal images were labeled as outlier because of their insufficient diagnostic data. Lalonde et al.⁵ propose an outlier detection system based on local pixel intensity distribution. The approach is open to errors for images that possess similar intensity distributions. In our paper, we proposed a two-phased MSRI detection approach. First, we graded the quality

of retinal images; then, we identified MSRI images among good-quality images. Our approach provides a better identification scheme for images that can be used in the retinal diagnostic process.

There are several reports on the success rates of visual features for different medical imaging systems in the literature.³⁶ However, our current works on retinal imaging generally involve new features instead of existing ones.^{3,4,18} Specific visual features for problems are thought to be more efficient, although the existing features have high success rates for different applications.³⁷⁻³⁹ Moreover, the existing features can be integrated easily into any visual analysis system. Hence, the performance evaluation of the existing visual features will provide more valuable information for the literature. Another aspect of our work is to find the most suitable visual features for the retinal IQA task. In addition, we proposed two novel simple and robust features for measuring the accuracy of the image frame. We selected our visual feature to represent both local and global properties of the retinal images in terms of texture and shape. In other words, we analyzed the shape and texture information of retinal images at the local and global scales. Finally, we presented our experimental results, and our feature-set obtained extremely high accuracy rates.

In this work, we proposed a new MSRI metric, which is exceedingly simple and robust. According to the experimental results, our MSRI metric obtained at least a 96% true detection rate. Hence, our approach identifies the MSRI images with high sensitivity scores.

Our approach can be integrated into any retinal image diagnostic system easily. Furthermore, our approach can increase both the efficiency and accuracy of the retinal image diagnostic process. In conclusion, we can say that the robustness of automatic retinal image analysis systems can increase with the help of our approach.

Acknowledgments

The authors would like to thank the Ophthalmology Department of Karadeniz Technical University, the Image Sciences Institute of the University Medical Center Utrecht, and the Machine Vision and Pattern Recognition Laboratory of the Lappeenranta University of Technology for allowing us to use their retinal image databases.

References

1. C. Köse et al., "Simple methods for segmentation and measurement of diabetic retinopathy lesions in retinal fundus images," *Comput. Methods Programs Biomed.* **107**(2), 274–293 (2012).
2. J. Paulus et al., "Automated quality assessment of retinal fundus photos," *Int. J. Comput. Assist. Radiol. Surg.* **5**(6), 557–564 (2010).
3. A. D. Fleming et al., "Automated assessment of diabetic retinal image quality based on clarity and field definition," *Invest. Ophthalmol. Vis. Sci.* **47**(3), 1120–1125 (2006).
4. L. Giancardo et al., "Elliptical local vessel density: a fast and robust quality metric for retinal images," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Vol. 2008, pp. 3534–3537, IEEE, Vancouver, BC (2008).
5. M. Lalonde, L. Gagnon, and M. Boucher, "Automatic visual quality assessment in optical fundus images," in *Proc. of Vision Interface*, pp. 259–264, Canadian Image Processing and Pattern Recognition Society, Ottawa, Ontario, Canada (2001).
6. S. C. Lee and Y. Wang, "Automatic retinal image quality assessment and enhancement," *Proc. SPIE* **3661**, 1581–1590 (1999).
7. M. Niemeijer, M. D. Abràmoff, and B. van Ginneken, "Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening," *Med. Image Anal.* **10**(6), 888–898 (2006).

8. J. Staal et al., "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004).
9. T. Kauppi, "Eye fundus image analysis for automatic detection of diabetic retinopathy," Ph.D. Thesis, Diss. Lappeenranta University of Technology (2010).
10. Y. Wang, W. Tan, and S. C. Lee, "Illumination normalization of retinal images using sampling and interpolation," *Proc. SPIE* **4322**, 500–507 (2001).
11. A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process. Signal, Image Video Technol.* **38**(1), 35–44 (2004).
12. F. Zana and J. C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Trans. Image Process.* **10**(7), 1010–1019 (2001).
13. S. Sekhar, W. Al-Nuaimy, and A. K. Nandi, "Automated localisation of retinal optic disk using Hough transform," in *2008 5th IEEE Int. Symp. Biomed. Imaging From Nano to Macro*, pp. 1577–1580, IEEE, Paris, France (2008).
14. M. U. Akram et al., "Retinal images: optic disk localization," in *Int. Conf. Image Anal. Recognit.*, pp. 40–49, Springer Berlin Heidelberg, Povo de Varzim, Portugal (2010).
15. S. Liangwongsan and B. Marungsri, "Extracted circle hough transform and circle defect detection algorithm," *World Acad. Sci. Eng. Technol.* **5**(12), 316–321 (2011).
16. C. Köse, U. Sevik, and O. Gençaliöglü, "Automatic segmentation of age-related macular degeneration in retinal fundus images," *Comput. Biol. Med.* **38**(5), 611–619 (2008).
17. C. Sinthanayothin et al., "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Br. J. Ophthalmol.* **83**(8), 902–910 (1999).
18. J. Nayak, P. S. Bhat, and U. R. Acharya, "Automatic identification of diabetic maculopathy stages using fundus images," *J. Med. Eng. Technol.* **33**(2), 119–129 (2009).
19. J. Nayak et al., "Automated diagnosis of glaucoma using digital fundus images," *J. Med. Syst.* **33**(5), 337–346 (2009).
20. S.-K. Hwang and W.-Y. Kim, "A novel approach to the fast computation of Zernike moments," *Pattern Recognit. Lett.* **39**(11), 2065–2076 (2006).
21. R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973).
22. B. S. Manjunath et al., "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.* **11**(6), 703–715 (2001).
23. D. K. Park, Y. S. Jeon, and C. S. Won, "Efficient use of local edge histogram descriptor," in *Proc. 2000 ACM Work. Multimed.*, pp. 51–54, ACM Press, New York (2000).
24. C. S. Won, "Feature extraction and evaluation using edge histogram descriptor in MPEG-7," in *Proc. Adv. Multimed. Inf. Process.* Vol. 3333, pp. 583–590, Springer Berlin Heidelberg, Tokyo, Japan (2004).
25. I. Mierswa et al., "YALE: rapid prototyping for complex data mining tasks," in *Proc. 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 935–940, ACM Press, New York (2006).
26. S. Mika et al., "Fisher discriminant analysis with kernels," in *Proc. 1999 IEEE Signal Process. Soc. Work.*, pp. 41–48, IEEE, Madison, WI (1999).
27. J. Yang and J. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognit. Lett.* **36**(2), 563–566 (2003).
28. D. Bremner et al., "Output-sensitive algorithms for computing nearest-neighbour decision boundaries," *Discrete Comput. Geom.* **33**(4), 593–604 (2005).
29. H.K.D.H. Bhadeshia, "Neural networks in materials science," *ISIJ Int.* **39**(10), 966–979 (1999).
30. H. Zhang, "The optimality of Naive Bayes," in *Proc. 17th Int. FLAIRS Conf.*, Vol. 1(2), pp. 562–567, The AAAI Press, Menlo Park, CA (2004).
31. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
32. W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognit. Lett.* **10**(5), 335–347 (1989).
33. P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.* **15**(11), 1119–1125 (1994).
34. D.M.W. Powers, "Evaluation: from precision, recall and f-measure to roc., informedness, markedness & correlation," *J. Mach. Learn. Technol.* **2**(1), 37–63 (2011).
35. Y.-H. Wen, A. Bainbridge-Smith, and A. B. Morris, "Automated assessment of diabetic retinal image quality based on blood vessel detection," in *Proc. Image Vis. Comput.*, Vol. 36(1), pp. 132–136, Hamilton, New Zealand (2007).
36. H. Müller et al., "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions," *Int. J. Med. Inform.* **73**(1), 1–23 (2004).
37. N. A. Rosa et al., "Using relevance feedback to reduce the semantic gap in content-based image retrieval of mammographic masses," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Vol. 2008, pp. 406–409, IEEE, Vancouver, BC (2008).
38. B. Zheng, "Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future perspectives," *Algorithms* **2**(2), 828–849 (2009).
39. D. D. Feng, "Automatic medical image categorization and annotation using LBP and MPEG-7 edge histograms," in *2008 Int. Conf. Technology & Biomedical Application*, pp. 51–53, IEEE, Shenzhen, China (2008).

Uğur Şevik received a BS degree in statistics and computer science in 2003, a MS degree in 2007, and is a PhD candidate in computer engineering at the Karadeniz Technical University, Trabzon, Turkey. He has been a research assistant at the Department of Statistics and Computer Science, Karadeniz Technical University, Trabzon, Turkey, since 2005. His research interests include image processing, pattern recognition, statistical learning, data mining, retinal image analysis, and forestry management systems.

Cemal Köse received his BS and MS degrees in the Department of Electrical and Electronic Engineering from Karadeniz Technical University, Trabzon, Turkey, in 1986 and 1990, respectively. He received a PhD in the Department of Computer Science from the University of Bristol, Bristol, UK, in 1997. He is currently an associate professor in the Department of Computer Engineering, Karadeniz Technical University, Trabzon, Turkey. His research interests include parallel processing, pattern recognition, natural language processing, and computer graphics.

Tolga Berber received his BS degree in statistics and computer science from Karadeniz Technical University in 2002 and a PhD degree in computer engineering from Dokuz Eylül University in 2013. He has been working at the Department of Statistics and Computer Science, Karadeniz Technical University, since 2013. His current research interests include information retrieval, machine learning, image analysis, text mining, pattern recognition, and multimedia databases.

Hidayet Erdöl graduated from Istanbul University, Faculty of Medicine, Turkey, in 1988. He was a resident physician in the Department of Ophthalmology, Karadeniz Technical University from 1990 to 1995. He has been a professor of medicine in the same department since 2010. He has been the head of the Ophthalmology Department since 2007. His research interests include glaucoma, diabetic retinopathy, age-related macular degeneration, and retinal diseases.