

Automatic Detection of Diabetic Retinopathy and Age-Related Macular Degeneration in Digital Fundus Images

Carla Agurto,^{1,2} E. Simon Barriga,^{1,2} Victor Murray,² Sheila Nemeth,¹ Robert Crammer,^{1,3} Wendall Bauman,⁴ Gilberto Zamora,¹ Marios S. Pattichis,² and Peter Soliz^{1,5}

PURPOSE. To describe and evaluate the performance of an algorithm that automatically classifies images with pathologic features commonly found in diabetic retinopathy (DR) and age-related macular degeneration (AMD).

METHODS. Retinal digital photographs ($N = 2247$) of three fields of view (FOV) were obtained of the eyes of 822 patients at two centers: The Retina Institute of South Texas (RIST, San Antonio, TX) and The University of Texas Health Science Center San Antonio (UTHSCSA). Ground truth was provided for the presence of pathologic conditions, including microaneurysms, hemorrhages, exudates, neovascularization in the optic disc and elsewhere, drusen, abnormal pigmentation, and geographic atrophy. The algorithm was used to report on the presence or absence of disease. A detection threshold was applied to obtain different values of sensitivity and specificity with respect to ground truth and to construct a receiver operating characteristic (ROC) curve.

RESULTS. The system achieved an average area under the ROC curve (AUC) of 0.89 for detection of DR and of 0.92 for detection of sight-threatening DR (STDR). With a fixed specificity of 0.50, the system's sensitivity ranged from 0.92 for all DR cases to 1.00 for clinically significant macular edema (CSME).

CONCLUSIONS. A computer-aided algorithm was trained to detect different types of pathologic retinal conditions. The cases of hard exudates within 1 disc diameter (DD) of the fovea (surrogate for CSME) were detected with very high accuracy (sensitivity = 1, specificity = 0.50), whereas mild nonproliferative DR was the most challenging condition (sensitivity = 0.92, specificity = 0.50). The algorithm was also tested on images with signs of AMD, achieving a performance of AUC of 0.84

(sensitivity = 0.94, specificity = 0.50). (*Invest Ophthalmol Vis Sci.* 2011;52:5862-5871) DOI:10.1167/iops.10-7075

Diabetic retinopathy (DR) is a disease that affects up to 80% of diabetics around the world. It is one of the leading causes of blindness in the United States, and it is the second leading cause of blindness in the Western world.¹ On the other hand, age-related macular degeneration (AMD) is the leading cause of blindness in people older than 65 years. More than 1.75 million people have AMD in the United States, and this number is expected to increase to 3 million by 2020.² Many studies have demonstrated that early treatment can reduce the amount of advanced DR and AMD cases mitigating the medical and economic impacts of the disease.³

Accurate, early detection of eye diseases is important because of its potential for reducing the number of cases of blindness around the world. Retinal photography for DR has been promoted for decades for both the screening of the disease and in landmark clinical research studies, such as the Early Treatment Diabetic Retinopathy Study (ETDRS).⁴ Although the ETDRS standard fields of view (FOVs) may be regarded as the current gold standard⁵ for diagnosis of retinal disease, studies have demonstrated that the information provided by two or three of these fields is sufficiently comprehensive to provide an accurate diagnosis of diabetic retinopathy and more than sufficient for screening.⁶

In recent years, several research centers have presented systems to detect pathologic conditions in retinas. Some notable ones have been presented by Larsen et al.,⁷ Niemeijer et al.,⁸ Chaum et al.,⁹ and Fleming et al.^{3,10} However, these approaches must apply specialized algorithms to detect a specific type of lesion in the retina. To detect multiple lesions, the previous systems generally implement more than one of these algorithms. Furthermore, some of these studies evaluate their algorithms on a single dataset, which avoids the problems, such as resolution, that are associated with the differences in fundus imaging devices.

These methodologies primarily employ a "bottom-up" approach in which the accurate segmentation of all the lesions in the retina is the basis for correct determination. A disadvantage of bottom-up approaches is that they rely on the accurate segmentation of all the lesions in the retina to measure performance. Yet, the development of specialized segmentation methods can be challenging. In such cases, lesion detection can suffer from the lack of effective segmentation methods. This is particularly problematic for advanced stages of DR, such as neovascularization.

A top-down approach, such as the one used in our study, does not depend on the segmentation of specific lesions. Thus, top-down methods can detect abnormalities not explicitly used in training.¹¹ Our objective was to show that this approach is

From ¹VisionQuest Biomedical LLC, Albuquerque, New Mexico; ²Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico; ³New Mexico VA Health Care System, Albuquerque, New Mexico; ⁴The Retinal Institute of South Texas, San Antonio, Texas; and the ⁵Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa.

Supported by National Institutes of Health, National Eye Institute Grants EY018280, EY020015, and RC3EY020749.

Submitted for publication December 16, 2010; revised April 24, 2011; accepted May 27, 2011.

Disclosure: **C. Agurto**, VisionQuest Biomedical, LLC (E); **E.S. Barriga**, VisionQuest Biomedical, LLC (E); **V. Murray**, None; **S. Nemeth**, VisionQuest Biomedical, LLC (E); **R. Crammer**, VisionQuest Biomedical, LLC (E); **W. Bauman**, VisionQuest Biomedical, LLC (C); **G. Zamora**, VisionQuest Biomedical, LLC (E); **M.S. Pattichis**, None; **P. Soliz**, VisionQuest Biomedical, LLC (I)

Corresponding author: Carla Agurto, Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131; capaagri@unm.edu.

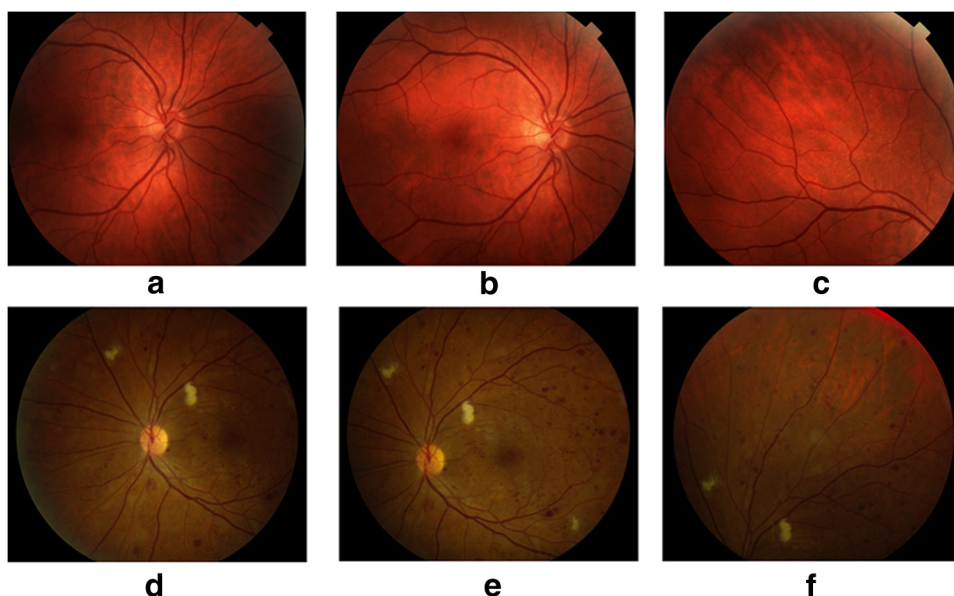


FIGURE 1. (a–c) FOVs 1, 2, and 3 of a normal retina from the RIST database; (d–f) FOVs 1, 2, and 3 of an abnormal retina from the UTHSCSA database.

a suitable implementation for eye disease detection with specific consideration to DR and AMD.

METHODS

Data Description

The retrospective images used to test our algorithm were obtained from the Retina Institute of South Texas (RIST, San Antonio, TX) and the University of Texas Health Science Center in San Antonio (UTHSCSA). Fundus images from 822 patients (378 and 444 patients from RIST and UTHSCSA, respectively) were collected retrospectively for this study. The use of the images was in accordance with the Declaration of Helsinki. The images were taken with a TRC 50EX camera (Topcon, Tokyo, Japan) at RIST and a CF-60uv camera (Canon, Tokyo, Japan) at UTHSCSA. Both centers captured 45° mydriatic images with no compression. The size of the RIST images was 1888 × 2224 pixels, and the size of the UTHSCSA images was 2048 × 2392 pixels. Both databases were collected in the south Texas area where, according to the U.S. Bureau of the Census, in 2009 the ethnicity distribution for this area was 58.3% Hispanic, 31.3% white (non-Hispanic), and 7.8% Afro-American. For the database provided by the UTHSCSA, no information about the age or sex of the patients was provided. In the case of the RIST database, the data were collected from July 2005 to February 2010. The distribution of patients was 50.8% female and 49.2% male. Ages were also obtained and were distributed as follows: 1.1% aged 0 to 24 years; 6.6%, 25 to 44 years; 26%, 45 to 64 years; and 66.3% d 65 years or older. All the images that presented with cataracts at their early stage, retinal sheen, or lighting artifacts were considered for this study. We excluded retinal images presenting advanced stages of cataracts, corneal and vitreous opacities, asteroid hyalosis, and significant eye lashes or eye lid artifacts. The number of images excluded in this study was 67, or 5.8% of the RIST database, and 57, or 5.2% of the UTHSCSA database.

Figure 1 shows examples of images from the three FOVs found in both databases. Figure 1a is centered on the optic disc (FOV1), Figure 1b is centered on the fovea (FOV2), and Figure 1c is focused on the superior temporal region of the retina (FOV3). Each image was graded independently into the following categories: normal, nonproliferative DR (NPDR), sight-threatening DR (STDR), and maculopathy. Table 1 shows the distribution of each subject’s eye in these categories. In addition, the 10 pathologic retinal conditions specified by the graders are shown in Tables 2 and 3. Seven were

related to DR: microaneurysms, hemorrhages, exudates less than 1 disc diameter (DD) away from the fovea, exudates elsewhere, intraretinal microvascular abnormalities (IRMA), neovascularization on the disc (NVD), and neovascularization elsewhere (NVE). The three pathologies related to AMD were drusen, abnormal pigmentation, and geographic atrophy (GA).

The graders assessed the image quality according to the criterion of clarity of vessels around the macula.¹² With this criterion, the quality of the images was classified as excellent, good, fair, or inadequate. Using the criterion, we removed 193 images, or 16.7% of the RIST database, and 111 images, or 10.2% of the UTHSCSA database. Figure 2 shows examples of images that were not considered for the study because of inadequate quality.

Image Processing

The detection process started with the extraction of features from the retinal images (see Fig. 3 for the complete procedure). Our algorithm uses a technique called amplitude-modulation frequency-modulation (AM-FM)^{11–14} to define the features and characterize normal and pathologic structures based on their pixel intensity, size, and geometry at different spatial and spectral scales. Please refer to Appendix A for a more detailed explanation of the AM-FM approach.

Since the result of AM-FM processing produces some features that may not be necessary for the accurate classification of images, we used informative outputs of a sequential backward elimination process in which the contributions of each feature were measured, and the ones that did not improve the classification performance were eliminated from our set. This process was applied independently for each of the pathologic features of interest, to obtain a better characterization of them.

To extract information from an image, AM-FM decomposes the images into different representations that reflect the intensity, geometry, and texture of the structures in the image. In addition to obtaining

TABLE 1. Distribution of the RIST and the UTHSCSA Databases

Database	Patients	Normal Eyes	NPDR Eyes	STDR Eyes	Maculopathy
RIST	378	64	486	158	174
UTHSCSA	444	116	418	292	207

Data are the number in each category.

TABLE 2. Distribution of DR Conditions among the RIST and UTHSCSA Images

Presence of Lesion	Microaneurysms	Hemorrhages	Exudates		IRMA	NVE	NVD
			Fovea	Elsewhere			
RIST images, <i>n</i>	378	511	174	248	80	30	58
UTHSCSA images, <i>n</i>	274	316	207	284	70	59	118

this information per image, filters were applied to obtain image representations in different bands of frequencies. For example, if a medium- or high-pass filter is applied to an image, the smaller retinal structures (e.g., microaneurysms, dot-blot hemorrhages, and exudates) microaneurysms are enhanced. This effect can be observed in Figures 4b and 4c, where the different types of red lesions, exudates, and thinner vessels present in the retinal region are captured. On the other hand, if a low-pass filter is applied, then larger structures are captured such as wider vessels, as shown in Figure 4e. By taking the difference of the two lowest scale representations, smaller vessels can also be captured, as seen in Figure 4f. These two ways of processing (AM-FM image representations and output of the filters) revealed more robust signatures of the different pathologies to be characterized. Thus, if we combine the representations of the different scales, we can obtain signatures for each structure, which allows us to detect and uniquely classify them.

To facilitate the characterization of early cases of retinopathy in which only a few small abnormalities are found, our process divides the images into regions of interest (ROIs). A sensitivity analysis of the size of the ROIs found that square regions of 140×140 pixels were adequate to represent features of small structures that can appear in the retina, such as microaneurysms or exudates. A total of 202 ROIs were necessary to cover the entire image. For classification, a feature vector was created with a concatenation of the following seven features from each region: (1) the first four statistical moments (mean, standard deviation, skewness, and kurtosis) and (2) the histogram percentiles (25th, 50th, and 75th).

A *k*-means clustering approach was performed to group the ROIs with similar features, by using the Euclidean distance between features. In this way, we avoid the time-consuming process of grading each region by use of an unsupervised algorithm. The resulting clusters become the representative feature vector per image. Once the feature vectors are extracted, we use them in the classification module (bottom right block in Fig. 3). This module used a partial least-squares (PLS) regression classifier to find the relevant features that classify images as normal or abnormal according to ground truth.

Experimental Design

The following paragraphs describe the experiments performed to assess the accuracy of the system in detecting the retinal pathologies listed in Tables 2 and 3. These pathologies are characteristic of either DR or AMD.

TABLE 3. Distribution of AMD Conditions among the RIST and UTHSCSA Images

Presence of Lesion	Drusen	Pigmentation	GA
RIST images, <i>n</i>	343	345	154
UTHSCSA images, <i>n</i>	188	86	54

In this section, we describe in detail the approaches taken for assessing the presence of these diseases in the retinal photographs.

DR Classification. For DR-related cases, the performance of the algorithm was measured by its ability to discriminate DR cases from normal cases. To do this, we created a mathematical model of the images by training the system on a subset of the data. This training set produced a model to which the testing images were compared. If the result of this comparison was greater than a predefined threshold, the image was considered abnormal (or suspect for DR). Images that fell below the threshold were labeled as normal.

In addition, the algorithm was tested separately on sight-threatening DR (STDR) cases, where STDR was defined as an image showing NVE, CSME, or NVD. In the following subsections, we detail the special properties that make the AM-FM representations ideally suited for the detection of CSME and NVD.

Clinically Significant Macular Edema. Previously, investigators have found an association between hard exudates near the fovea and CSME.^{15,16} Although hard exudates are one of the most common findings in macular edema, their presence is not always indicative of edema. Previous research has demonstrated that the sensitivity of exudates in predicting macular edema is 93.9%.¹⁷ Finding exudates near the fovea does not unequivocally ascertain edema's presence or absence. Our goal is simply to identify at-risk patients based on the presence of hard exudates. For the purposes of this study, the presence of exudates within 1 DD of the fovea was considered to be a surrogate for CSME.¹⁸ Figure 5 shows an example of how AM-FM highlights the presence of exudates while minimizing interference from blood vessels. In this figure, we see a normal retina (Fig. 5a) and one containing exudates within 1 DD of the fovea (Fig. 5b). Fig 5c shows the AM-FM decomposition of the normal retina for the high frequencies. The representation eliminates all the vessels from the image and shows only a dark background. In contrast, at the same high frequencies, the AM-FM decomposition for the abnormal retina clearly highlights the exudates while eliminating the vessels.

Neovascularization on the Optic Disc. NVD is defined as the growth of new vessels within 1 DD of the center of the optic disc.¹⁸ In Figure 6, we show how NVD is represented by AM-FM. Note that, when medium frequencies are applied (Figs. 6b, 6c), the vessels on the optic disc and NVD are highlighted, whereas the NVD is best represented by the high frequencies, as shown in Figure 6f. It is these combinations of AM-FM frequencies that were used by our classifier to determine the presence or absence of an abnormality. The noise present in Figures 6c and 6f is due to a bright nerve fiber layer, but its AM-FM representation has less intensity amplitude than the abnormal vessels.

AMD Classification. In addition to testing the images for the presence of DR, three different pathologies related to AMD were analyzed: drusen, abnormal pigmentation, and GA. Figure 7 shows an example of a retinal image with drusen and one of its corresponding AM-FM image representations. As seen in this example, drusen are noticeably highlighted by AM-FM. We then tested the system in the following scenarios:

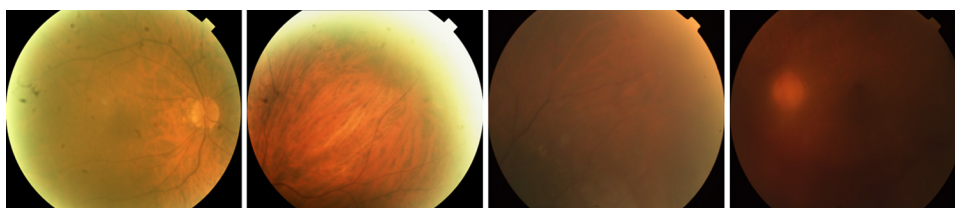


FIGURE 2. Examples of inadequate quality images that were not used by our algorithm.

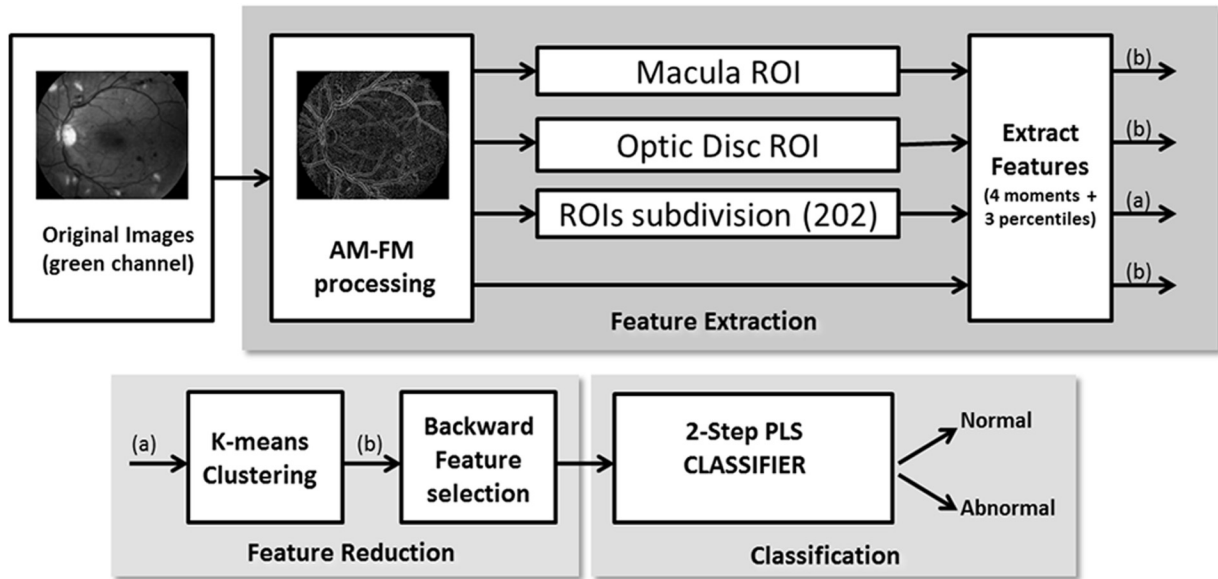


FIGURE 3. Procedure for classifying the retinal images. First, the green channel of the images is selected. Then, the images are processed by AM-FM to decompose them in their AM-FM estimates. Depending on the test, the images are subdivided in ROIs, the macula or the optic disc region is selected, or the entire image without the optic disc is entered into the block of the feature extraction. Features are obtained for each observation. If the image is represented in ROIs, the *k*-means method is applied; otherwise, the feature selection and the two-step PLS classifier is applied, to obtain the estimated class for each image.

normal versus drusen, normal versus abnormal pigmentation, normal versus GA, and normal versus all AMD pathologies. For the drusen experiment, all stages of the presence of drusen were categorized in the same group without distinction of severity (e.g., a few isolated drusen versus large, confluent drusen).

RESULTS

Interreader Variability

To analyze the consistency of the grading criteria, a randomly selected subset of 10% of the data from RIST and

UTHSCSA was given to two graders: an optometrist (grader 2) and an ophthalmologist (grader 3). This random selection process has been used by others such as Abramoff et al.¹⁹ who used 1.25% of their data (~40,000 images) to compare rates from three retinal specialists. Our new subset, the database described in Tables 2 and 3, was read by the three graders according to the original categories. The agreement between graders was calculated using the κ statistic. We calculated the κ statistic for three exclusive classes: normal retinas, abnormal retinas, and sight-threatening eye diseases, as reported in Table 4.

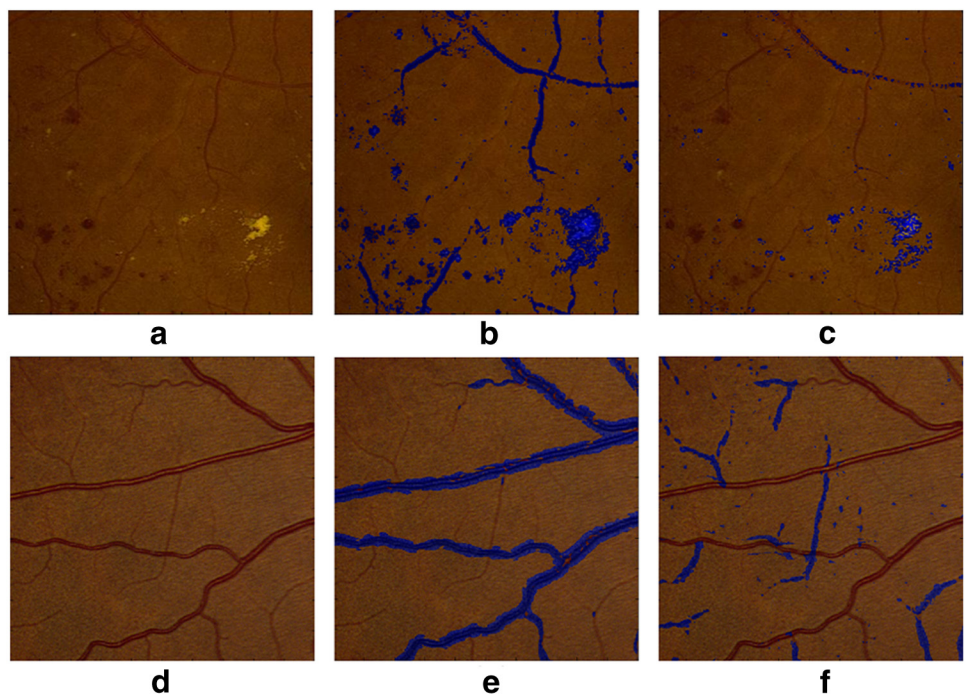


FIGURE 4. Structures in the retina captured by the AM-FM estimates using high values of the IA (*blue*). (a) Region of a retinal image with pathologies; (b) image representation using medium frequencies, which captures dark and bright lesions as well as vasculature; (c) image representation using high frequencies. Note that this image captures most of the bright lesions; (d) region of a retinal image with normal vessel structure; (e) image representation using a very low frequency filter; (f) Image representation of (d) obtained by taking the difference between the very low and the ultralow frequency scales, in this image the thinner vessels are better represented.

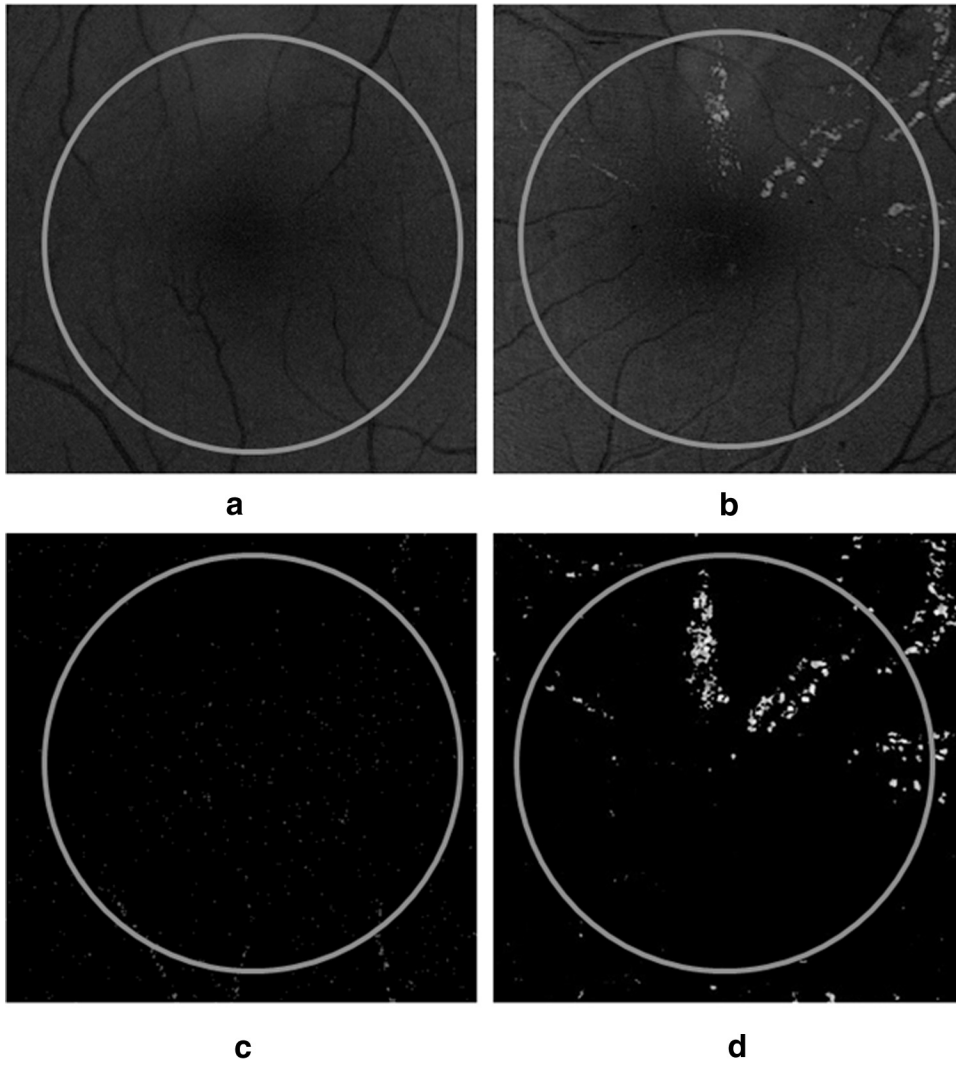


FIGURE 5. Examples of structures capture by the AM-FM estimates using high values of the IA in macular regions (the *circle* encloses an area equal to 1 DD from the fovea). (a) Normal macula, (b) macula with hard exudates, (c) normal retina at high frequency, and (d) retina with exudates at high frequency.

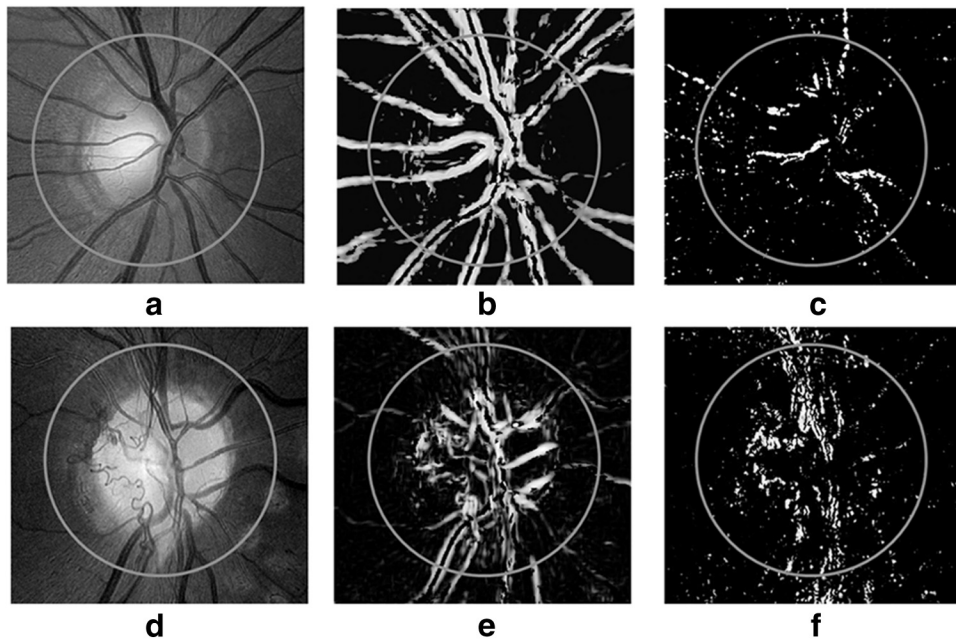


FIGURE 6. Examples of structures captured by the AM-FM estimates using high values of the IA for two different optic discs. (a) Normal optic disc, (d) NVD, (b, e) IA of the retinas in (a) and (d) at medium frequencies, (c, f) images of (a) and (d) at high frequencies.

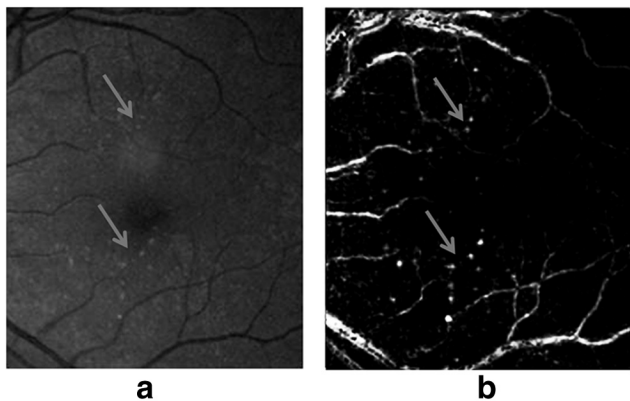


FIGURE 7. (a) Retinal region with drusen. (b) Structures captured by the AM-FM estimates using high values of the IA at low frequencies.

Automatic Detection Results

Cross validation was used to assess the performance of the algorithm. The ratio between training and testing data was selected so that 70% of the data were used for training and 30% were used for testing. A more robust classification model was estimated by randomly selecting the images in the training and testing sets. The average of 20 runs is presented in Tables 5 and 6. This procedure minimizes the possible bias incurred if the training and testing sets are fixed.²⁰ To compare our results with recently published algorithms, we fixed the specificity to two values: 0.50 and 0.60. These specificities have been used to report sensitivity

TABLE 4. Measurement of Agreement of Three Readers, According to Cohen’s κ for Normal Retinas, Abnormal Retinas, and Sight-Threatening Eye Diseases

Comparison	κ Class	κ	SE	95% CI
Grader 1 vs. grader 2	Unweighted	0.61	0.058	0.50-0.73
	Linear weighted	0.69	0.048	0.60-0.79
Grader 1 vs. grader 3	Unweighted	0.74	0.056	0.67-0.85
	Linear weighted	0.79	0.047	0.70-0.88
Grader 2 vs. grader 3	Unweighted	0.62	0.068	0.49-0.76
	Linear weighted	0.69	0.059	0.57-0.80

in two large studies.^{3,9} Figure 8 shows six ROC curves, three for each database for the following experiments: normal versus NPDR, normal versus STDR, and normal versus DR.

DISCUSSION

Interreader Variability

Table 4 shows the κ values obtained to measure the interreader variability. Based on the Landis and Koch interpretation of κ values, we obtained substantial agreement (κ values higher than 0.6) between all three graders. Further analysis was performed for grading the different sight-threatening categories obtaining κ values of 0.71 and 0.60 for CSME and NVD, respectively, whereas for the NVE and IRMA cases, a κ value of 0.55 (moderate agreement) was obtained. When we analyzed the differences between graders for this last category, we noted that most of the disagreements were present in the detection of IRMA. An illustration of this is shown in Figure 9. The two images shown in Figures 9a and 9b were presented to the graders again after applying local contrast enhancement, and they agreed that the pathologic condition was present. Lower image quality and blurring on some images are some of the factors that contributed to the disagreement between graders. In addition, we found that the presence of other pathologies masked the presence of IRMA, as shown in Figure 9a.

Automatic Algorithm Results

Although different databases were used, the results for both databases were consistent, in particular in patients with STDR. For both databases, the best results were obtained for CSME detection (hard exudates less than 1 DD away from the fovea), and the worst were obtained for the detection of images with drusen. It can be seen from Table 5 that in most of the experiments, especially those for DR, the results of the UTH-SCSA database (AUC = 0.89) were slightly better than those from the RIST database (AUC = 0.81). One of the main factors contributing to this difference was the higher quality images found in the UTHSCSA database, as assessed by the expert graders.

From Table 5 one can observe that the algorithm detected CSME cases (i.e., hard exudates) with high sensitivity in the range of (0.98-1.00) for 0.50 specificity. For detecting cases with nonproliferative DR, hemorrhages, and microaneurysms,

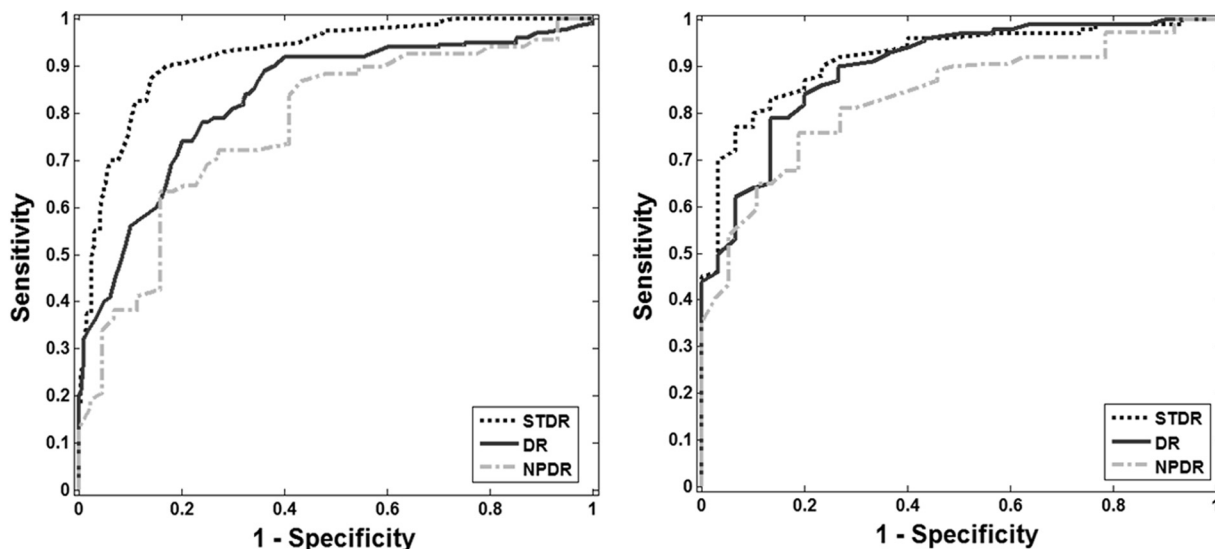


FIGURE 8. ROC curves for the classification of DR, STDR, and NPDR. (a) RIST (b) UTHSCSA databases.

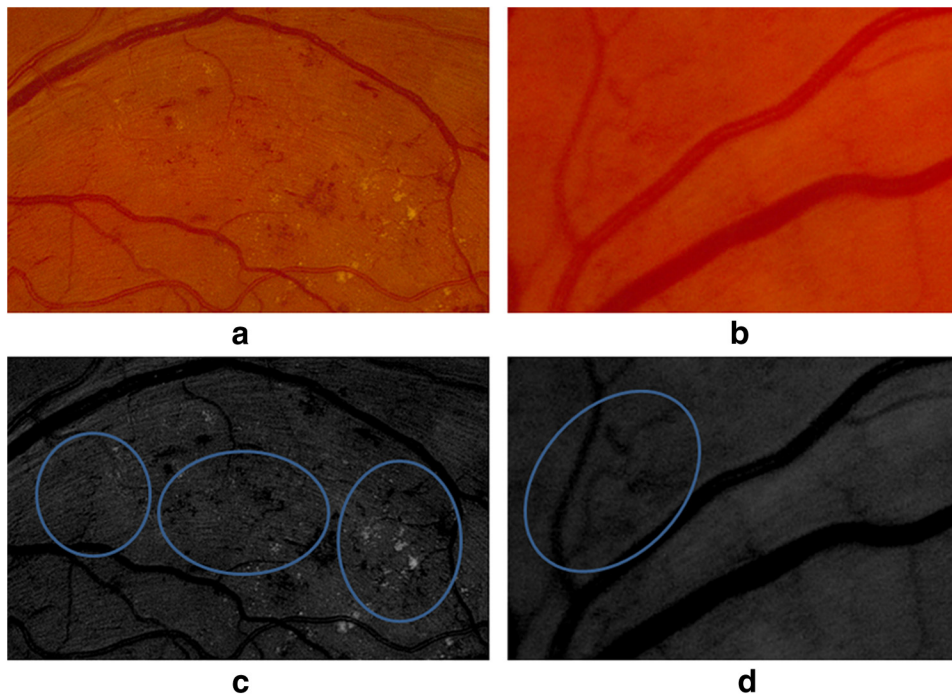


FIGURE 9. Retinal images with IRMA. The presence of IRMA was not detected in the image in (a) by the first grader or in the image in (b) by the second grader. (c, d) Images (a) and (b) with enhancement.

the algorithm produced results comparable or better than those reported in other large studies, such as Fleming et al.,¹⁰ who obtained sensitivity and specificity of 0.89 and 0.50 for mild DR.

For detection of NVD, the algorithm achieved AUCs of 0.88 and 0.91. This result highlights one of the advantages of our approach, as abnormal vessels can be discriminated from normal ones through analysis of the different image representations generated by AM-FM without the need for explicit segmentation of the vasculature.

In the detection of NVE and IRMA, the performance of the algorithm was 0.85 and 0.92 of sensitivity, respectively, for 0.50 specificity for both databases. We observed that one of the factors that affected the algorithm was the ground truth. When we compared the graders' measurements for these pathologies, we observed only moderate agreement ($\kappa = 0.55$). Furthermore, since ground truth is required to train the classifier, different graders can lead to different classification models due to inter-grader variability, even though the same set of images is used. To the best of our knowledge, this is the first published result on automatic detection of NVE.

One of the issues in testing our algorithm was the relatively low proportion of cases of early-stage DR. The low number was due to the nature of the centers from which the data were

collected, which tended to bias the samples to patients with advanced stages of retinal disease. We have reported other studies (Agurto C, et al. *IOVS* 2010;51:ARVO E-Abstract 1793)¹¹ with the available online database MESSIDOR, but this database was not useful in this study because it does not contain enough samples of advanced cases. In the future, we will train the system using a database that contains a greater number of DR stages, ranging from normal to NPDR, PDR, and maculopathy. In our experiments, we found that a robust training set is the most important aspect when improving the performance of the system. In fact, as the number of cases analyzed by the algorithm increased, so did its accuracy, evidenced by the improvements found in comparison with the results presented in our previous publications on the topic (Agurto C, et al. *IOVS* 2010;51:ARVO E-Abstract 1793).^{11,21-24}

An advantage of our top-down approach is clearly shown in the detection of abnormalities related to AMD. Although the system was not originally intended for those abnormalities, by adding AMD cases to the training database, we were able to detect these lesions with an accuracy of sensitivity/specificity of 0.94/0.50 and 0.90/0.50 for the RIST and UTHSCSA databases, respectively.

The results presented in this article are comparable with the ones published by other investigators. For example, Niemeijer

TABLE 5. Results of Performance Evaluation for DR Experiments and for Each Database

Diseases	RIST Database				UTHSCSA Database			
	Images, <i>n</i> *	AUC	Sens. for Spec. 0.60	Sens. for Spec. 0.50	Images, <i>n</i> *	AUC	Sens. for Spec. 0.60	Sens. for Spec. 0.50
DR	(419/144)	0.81	0.92	0.92	(437/136)	0.89	0.94	0.97
NPDR only	(226/144)	0.77	0.83	0.88	(124/136)	0.85	0.90	0.95
STDR only	(193/144)	0.92	0.95	0.98	(313/136)	0.92	0.96	0.96
CSME	(68/44)	0.98	1	1	(147/93)	0.97	0.98	0.99
IRMA and NVE	(95/144)	0.85	0.92	0.93	(137/136)	0.92	0.97	0.98
NVD	(28/50)	0.88	0.90	0.92	(74/94)	0.91	0.95	0.95

* The first term in the parentheses refers to the abnormal cases, and the second term refers to the number of normal cases used in each experiment. Sens., sensitivity; spec. specificity.

TABLE 6. Results of Performance Evaluation for AMD Experiments and for Each Database

Diseases	Rist Database				UTHSCSA Database			
	Images, n^*	Sens. for AUC	Sens. for Spec. 0.60	Sens. for Spec. 0.50	Images, n^*	Sens. for AUC	Sens. for Spec. 0.60	Sens. for Spec. 0.50
AMD only	(248/144)	0.84	0.90	0.94	(259/136)	0.77	0.90	0.90
Drusen	(91/98)	0.77	0.88	0.95	(143/136)	0.73	0.80	0.85
Pigmentation	(55/98)	0.80	0.90	0.90	(61/136)	0.81	0.87	0.90
GA	(100/98)	0.92	0.97	1	(76/136)	0.92	0.90	1

* The first term in the parentheses refers to the abnormal cases, and the second term refers to the number of normal cases used in each experiment. Sens., sensitivity; spec. specificity.

et al.⁸ tested their algorithm in 15,000 patients and obtained an AUC of 0.88 (sensitivity/specificity = 0.93/0.60) for the detection of diabetic retinopathy. In a study of 33,535 patients from the Scottish National DR screening program, Fleming et al.¹⁰ reported detection of background retinopathy with 0.84 sensitivity and 0.50 specificity and detection of maculopathy with a sensitivity of 0.99 for the same level of specificity. Both of those studies looked only at the detection of DR, in contrast to our study, which added cases of AMD. Chaum et al.⁹ conducted a study with 395 retinal images and reported a range of sensitivity of 0.75 to 1.00 in the detection of AMD and sensitivity of 0.75 to 0.947 in the detection of DR. In our approach, by using the information provided in Tables 5 and 6, we report detection of DR with sensitivity/specificity of 0.92/0.60 and 0.94/0.60, detection of CSME with sensitivity/specificity of 1/0.60 and 0.99/0.50, and detection of AMD with sensitivity/specificity of 0.94/0.50 and 0.90/0.60 for the RIST and UTHSCSA databases, respectively. As can be observed, this approach demonstrates an algorithm that has the capability of detecting the presence of pathologies associated with more than one eye disease.

By observing the ROC curves (Fig. 8), the performance of the algorithm on images from the RIST and UTHSCSA databases for the detection of STDR cases is very high, with sensitivities of 0.96 and 0.98, respectively, for a fixed specificity of 0.50. If we fix the specificity to 80%, the algorithm achieved sensitivities of 0.92 and 0.85 for the RIST and UTHSCSA databases, respectively. For the other two experiments, NPDR and DR, we achieved sensitivities in the range of (0.88–0.97) for 0.50 specificity.

In conclusion, this work presents a viable and efficient means of characterizing different retinal abnormalities and building binary classifiers for detection purposes. Although automatic detection of DR has been studied by different groups in the past decade, few studies have used a top-down approach like the one we propose. In addition to that, to our knowledge, automatic detection of STDR, as well as neovascularization, pigmentation, and GA, has not been concurrently addressed at the levels of performance presented in this work.

References

- Zhang X, Saaddine JB, Chou CF, et al. Prevalence of diabetic Retinopathy in the United States 2005–2008. *JAMA*. 2010;304(6):649–656.
- The Eye Diseases Prevalence Research Group. Prevalence of age-related macular degeneration in the United States. *Arch Ophthalmol*. 2004;122:564–572.
- Scotland GS, McNamee P, Fleming AD, et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. *Br J Ophthalmol*. 2010;94:712–719.
- Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the modified Airlie House classification. ETDRS report 10. 1991;98:786–806.
- Li HK, Hubbard LD, Danis RP. Digital versus film fundus photography for research grading of diabetic retinopathy severity. *Invest Ophthalmol Vis Sci*. 2010;51:5846–5852.
- Scanlon PH, Malhotra R, Thomas G. The effectiveness of screening for diabetic retinopathy by digital imaging photography and technician ophthalmoscopy. *Diabet Med*. 2003;20(6):467–474.
- Larsen N, Godt J, Grunkin M, Lund. Andersen H, Larsen M. Automated detection of diabetic retinopathy in a fundus photographic screening population. *Invest Ophthalmol Vis Sci*. 2003; 44(2):767–771.
- Niemeijer M, Abramoff MD, van Ginneken B. Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Transactions on Medical Imaging*. 2009;28(5):775–785.
- Chaum E, Karnowski TP, Govindasamy VP, Abdelrahman M, Tobin KW. Automated diagnosis of retinopathy by content-based image retrieval. *Retina* 2008;28(10):1463–1477.
- Fleming AD, Goatman KA, Philip S, Prescott GJ, Sharp PF, Olson JA. Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts. *Br J Ophthalmol*. 2010;94:1606–1610.
- Agurto C, Murray V, Barriga, et al. Multiscale AM-FM methods for diabetic retinopathy lesion detection. *IEEE Trans Med Imag*. 2010;29(2):502–512.
- Fleming AD, Philip S, Goatman K, Olson J, Sharp P. Automated assessment of diabetic retinal image quality based on clarity and field definition. *Invest Ophthalmol Vis Sci*. 2006;47:1120–1125.
- Murray V, Rodriguez P, Pattichis MS. Multi-scale AM-FM Demodulation and reconstruction methods with improved accuracy. *IEEE Trans Image Process*. 2010;19(5):1138–1152.
- Murray-Herrera VM. *AM-FM Methods for Image and Video Processing*. PhD dissertation. Albuquerque, NM: University of New Mexico; 2008.
- Kinyoun J, Barton F, Fisher M, Hubbard L, Aiello L, Ferris F. Detection of diabetic macular edema: ophthalmoscopy versus photography. Early Treatment Diabetic Retinopathy Study Report Number 5, The ETDRS Research Group. *Ophthalmology*. 1989;96:746–750.
- Welty CJ, Agarwal A, Merin LM, Chomsky A. Monoscopic versus stereoscopic photography in screening for clinically significant macular edema. *Ophthalmic Surg Lasers Imaging*. 2006;37:524–526.
- Rudnisky CJ, Tennant MT, de Leon AR, Hinz BJ, Greve MD. Benefits of stereopsis when identifying clinically significant macular edema via tele-ophthalmology. *Can J Ophthalmol*. 2006;41(6):727–732.
- Jelinek HF, Cree MJ, eds. *Automated Image Detection of Retinal Pathology*. Boca Raton: CRC Press; 2010.
- Abramoff MD, Niemeijer M, Russell SR. Automated detection of diabetic retinopathy: barriers to translation into clinical practice. *Expert Rev Med Devices*. 2010;7(2):287–296.
- Chen W, Gallas BD. Training variability in the evaluation of automated classifiers. *Proc. SPIE* 2010;7624:762404.
- Agurto C, Barriga S, Murray V, Pattichis M, Davis B, Soliz P. Effects of image compression and degradation on an automated diabetic retinopathy screening algorithm. *Proc. SPIE* 2010;7624:76240H.

22. Agurto C, Murillo S, Murray V, et al. Detection and phenotyping of retinal disease using AM-FM processing for feature extraction. *42nd IEEE Asilomar Conference on Signals, Systems and Computers, October 26-29, 2008*. 2008:659-663. Conference on.
23. Barriga ES, Murray V, Agurto C, et al. Multi-scale AM-FM for lesion phenotyping on age-related macular degeneration. *IEEE International Symposium on Computer-Based Medical Systems, August 2-5, 2009*. 2009:1-5.
24. Barriga ES, Murray V, Agurto C, et al. Automatic system for diabetic retinopathy screening based on AM-FM, partial least squares, and support vector machines. *IEEE Int Symposium on Biomedical Imaging: From Nano to Macro, August 14-17, 2010*; 2010: 1349-1352.

APPENDIX A

Amplitude-Modulation Frequency-Modulation and Feature Extraction

The image representations from which the features are generated are obtained using a technique called AM-FM. To extract information from an image, this technique decomposes the green channel of the images into different representations which reflect the intensity, geometry, and texture of the structures in the image. The AM-FM decomposition for an image $I(x, y)$ is given by:

$$I(x, y) \approx \sum_{n=1}^M a_n(x, y) \cos \varphi_n(x, y)$$

where M is the number of AM-FM components, $a_n(x, y)$ denotes the instantaneous amplitude (IA) estimate, and $\varphi_n(x, y)$ denotes the instantaneous phase. Using the latter, two AM-FM estimates are generated by extracting the magnitude and the angle of its gradient. These estimates are called instantaneous frequency magnitude (|IF|), and instantaneous frequency angle.

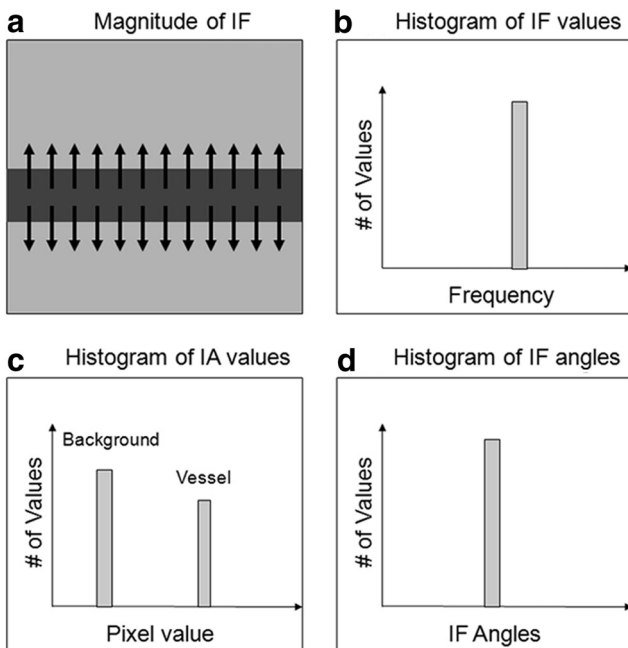


FIGURE A1. Conceptual AM-FM analysis for horizontally oriented blood vessel edge: (a) Instantaneous frequencies on top of a vessel-like structure. Histograms are shown for (b) instantaneous frequency, (c) instantaneous amplitude, and (d) instantaneous frequency angle.

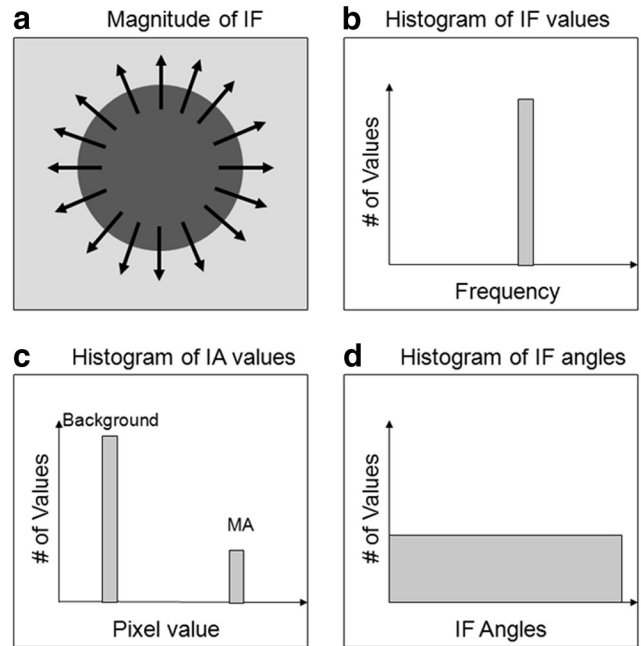


FIGURE A2. Conceptual AM-FM analysis for a round, dark lesion. (a) Instantaneous frequencies on top of the lesion. Histograms are shown for (b) instantaneous frequency, (c) instantaneous amplitude, and (d) instantaneous frequency angle.

In addition to obtaining this information per image, filters are applied to obtain image representations in different bands of frequencies. For example, if a medium- or high-pass filter is applied to an image, the smaller retinal structures (e.g., MAs, dot-blot hemorrhages, and exudates) are enhanced. Using these two ways of processing (AM-FM image representations and output of the filters), more robust signatures of the different pathologies can be characterized. At the end of this step, an image has 39 different representations that characterize the different pathologies found in the retina. A more extensive mathematical description of the AM-FM technique can be found in a previously published paper.¹¹ In this section we will describe conceptually the way AM-FM represents two structures commonly found in DR images: retinal vessels and round, dark lesions. The same analysis can be performed for bright lesions, large hemorrhages, and abnormal vessels, among other retinal features.

Figure A1 shows the way a horizontally oriented retinal vessel is represented by AM-FM, and the resulting histograms for the three different AM-FM estimates: IA, |IF|, and instantaneous frequency angle. The arrows in Figure A1a show the direction in which the frequency change is happening, meaning the way the pixel values are changing from dark (vessel) to bright (retinal background). The pixels in the background will have only slight changes in intensity, and therefore their frequencies are close to 0. The only areas generating a frequency response are those in the edge of the vessels, and they will have a very distinctive |IF|, as represented in Figure A1b. The IA will have high values for the areas with higher contrast, and therefore, in the ideal case, the histogram of the IA will have two distinctive peaks: one for the retinal background and one for the edge of the vessels, as seen in Figure A1c. One of the most distinctive features of vessel-like features is their directionality, which is captured by the IF angle. The direction of change will be roughly the same for an elongated structure like a vessel, and therefore the angle of the IF will generate a highly peaked histogram, as seen in Figure A1d.

Figure A2 shows the histogram of the AM-FM representation for a dark, round region such as MAs or dot-blot hemorrhages.

The lesion is characterized by the IF, with large values at the edge of the lesion and low values inside and outside the lesion, as depicted in Figure A2a. Just as in the case of the vessels, the resulting $|IF|$ histogram has a clear peak for the high-frequency values (Fig. A2b). The IA histogram contains two peaks: one for the contrast changes in the background and one for the contrast changes on the edges of the lesion, as seen in Figure A2c. This IA histogram is similar to the one for the vessel, but since MAs are smaller than vessels, the number of pixels with high contrast will be smaller, and therefore the histogram will have a smaller peak that represents the MAs. Finally, one of the biggest differences between vessels and MAs is seen on the IF angle. In the ideal case of a perfect circular shape where all the

angles of the IF are represented (as seen in Fig. A2a), the histogram for the angles is uniform (Fig. A2d), since all angles of the IF are represented.

These two examples illustrate conceptually the way in which AM-FM is obtained different signatures for each of the two analyzed structures. Any structure of any shape, color, and size can be characterized by combining the outputs of the three estimates. We are conscious that retinal images present additional information such as noise or blurring that is not considered in the ideal cases presented here, but by using appropriate statistical measurements to represent the AM-FM estimates, high classification accuracy can be obtained, as shown by the results presented in this article.