Toward Automatic Phenotyping of Retinal Images from Genetically Determined Mono- and Dizygotic Twins Using Amplitude-Modulation Frequency-Modulation Methods

P Soliz* [a,b,c], B Davis[a], V Murray[c], M Pattichis[c], S Barriga[a], and S Russell[b]
[a]VisionQuest Biomedical, LLC, Albuquerque, NM 87109, [b]University of Iowa Department of Ophthalmology and Visual Sciences, Iowa City, IA 52242, and [c]University of New Mexico Department of Electrical and Computer Engineering, Albuquerque, NM 87131

ABSTRACT

This paper presents an image processing technique for automatically categorize age-related macular degeneration (AMD) phenotypes from retinal images. Ultimately, an automated approach will be much more precise and consistent in phenotyping of retinal diseases, such as AMD. We have applied the automated phenotyping to retina images from a cohort of mono- and dizygotic twins. The application of this technology will allow one to perform more quantitative studies that will lead to a better understanding of the genetic and environmental factors associated with diseases such as AMD. A method for classifying retinal images based on features derived from the application of amplitude-modulation frequency-modulation (AM-FM) methods is presented. Retinal images from identical and fraternal twins who presented with AMD were processed to determine whether AM-FM could be used to differentiate between the two types of twins. Results of the automatic classifier agreed with the findings of other researchers in explaining the variation of the disease between the related twins. AM-FM features classified 72% of the twins correctly. Visual grading found that genetics could explain between 46% and 71% of the variance.

Keywords: age-related macular degeneration, amplitude-modulation, frequency-modulation, phenotyping

1. INTRODUCTION

1.1 Age-related retinal disease

Age-related macular degeneration (AMD) is the most common cause of visual loss in the United States and is a growing public health problem. One third of Americans will develop AMD in their lifetimes. The estimated annual cost burden from AMD in the US is $30 billion (USD) or about 0.3% of gross domestic product.[1] According to the Center for Disease Control (CDC), the prevalence of AMD is expected to double over the next 25 years, due to the increase in the number of older Americans. The principal objective of this project was to develop a computer-based grading system establishing phenotypes of AMD that go beyond visually recognizable features observed and documented by studies, such as the simplified severity scale for AMD given in the Age-related Eye Disease Study (AREDS) report #18.[2] AREDS reports have described biomarkers that correlate to the 5-year risk of developing AMD. Our proposed computer-based approach will detect other features based more on mathematical representations. Because of its automation and objectivity it will eliminate inter- and intra-reader variability concerns, and most importantly will create a means for the discovery of new biomarkers with greater specificity in associating these markers with disease progression. The ultimate goal of this system is to provide biomarkers to identify patients at risk of developing advanced stages of AMD and to improve outcomes by identifying more specific interventions. In this manner one will be able to select the best treatment earlier in the disease, which will lead to improved therapeutic outcomes.

1.2 The nature of AMD phenotyping

AMD expresses an appearance on the fundus that is complex and variable both spatially and longitudinally. The most common feature associated with AMD is drusen deposits. These deposits appear as bright lesions which may be nearly round or confluent. Studies focused on the role of drusen have shown that AMD visual outcomes are related to these clinically visible features. For example, current manual or human-observer based phenotyping have used drusen size,

* psoliz@visionquest-bio.com; phone 1 505-508-1994; www.visionquest-bio.com

numbers, and area for classification. Additionally, these manual grading systems document the presence of pigmentary abnormalities (hyperpigmentation), and the occurrence of contralateral disease severity. Some visually identifiable features have been found to be more useful than others in phenotypic characterization, possibly due to the qualitative and somewhat variable nature of the human observer.

Research, such as that of Seddon et al., has aimed to discover the complex interplay of genetics and environment.[3] In their study of the larger set of these same mono- and dizygotic twins, they were able to estimate the relative contribution of genetic and environmental factors associated with age-related macular degeneration. Numerous investigators have reported on the complexity of the AMD disorder, presenting with various phenotypes. We suggest that a more objective and precise definition of the different AMD phenotypes will increase the likelihood of identifying genes and other factors that predispose an individual to the disease.

Other researchers have hypothesized that AMD is composed of phenotypic groups that can be defined through quantitative techniques based on image processing. As a result, we have been engaged on a project to apply AM-FM techniques to provide features to a process of finding the salient features and using them as inputs for classifying AMD phenotypes.[4,5] The specific application in this phase of the project is to study the AM-FM features in order to determine whether phenotypic features can be used to distinguish identical (monozygotic) from fraternal (dizygotic) twins based on the similarities of the presentation of the disease. These individuals represent a cohort of twins who have been diagnosed with AMD. These data are described by Seddon et al.[3] The methods section will describe each step in the processing that starts from the mathematically-derived image features (using AM-FM) and clustering them into phenotypic groups, then finally related those clusters to the genetic information, i.e. monozygotic (MZ) or dizygotic (DZ). The result of the image analysis and classification will be evaluated for accuracy based on correctly identifying zygosity as determined genetically, i.e. total correct / total samples. Additionally, we compare the results of the classification with the classification given by ophthalmologists.

2. THE DATA

2.1 The twin data

A subset of the Twin data set, captured in 35mm color slides, was digitized for this study. N = 21 monozygotic and N = 19 dizygotic twin pairs were randomly selected. Both eyes had been imaged and were digitized. The images were saved in TIF format at 1301 by 1953 pixel resolution. The image set was screened for image quality by trained ophthalmic technicians. N = 27 images (15 monozygotic and 12 dizygotic) were removed for unusable image quality. The remaining useable dataset consisted was divided into 18 cases for training and 18 cases for testing.

Rejection was based on poor alignment, e.g. the fovea was more than one optic disc diameter from the margin of the region of the retina presented on the image. Images with significant scares were identified for possible exclusion from the training set. The images were divided into two sets. Set 1 was unmasked to the image processing analyst and intended for use as training data. Set 2 (N=10 MZ and N = 8 DZ) was masked and treated as such for the purposes of testing the classification algorithm. The twin population is composed of WWII veterans, all over the age of 80.

As a basis for comparison of the computer-based results, three masked individuals were asked to make a determination as to the zygosity of the individuals based on their retinal images. Two of the individuals are board certified retinal specialists and one is an image processing expert.

2.2 Age-related Eye Disease Study (AREDS) standard images

To evaluate the feature extraction process, a preliminary test was performed on the standard high-resolution AREDS images. These images were downloaded from the Wisconsin Fundus Photo Reading Center's website.[6] Image size is 1200x1000 pixels in TIF format. A certified ophthalmic technician (reader) selected the retinal features. For these studies, we used only areas that contain retinal background, vessels, soft drusen, and hard drusen. After the reader selected the features, regions of interest (ROI) of 40x40 pixels were extracted from the image. The goal of this preliminary study was to evaluate the ability to correctly classify specific, commonly used features, such as drusen types, hard, soft, etc; and to differentiate these drusen features from other retinal structures.

3. METHODS

3.1 Mathematical characterization of retinal images

The process of computerized phenotyping is based on the characterization of an image using an image processing technique that we have previously reported on, amplitude modulation-frequency modulation (AM-FM).[5] Feature sets are mathematically-derived from the images utilizing AM-FM techniques.[7,8] The AM-FM features encapsulate the characteristics of each image, capturing the normal anatomical structures as well as the pathological lesions present. The hypothesis is that some select AM-FM features are associated with putative phenotypes of AMD, while others will encode normal anatomical structures. These features will subsequently be used to attempt to find "matching" features between the two individuals in a twin pair. The hypothesis is that the matching features will be closer for monozygotic twin pairs than for dizygotic twins.

3.2 AM-FM estimates

The AM-FM feature extraction methodology described in this document was first reported by Pattichis et al.[8] One of this groups' early applications was in characterizing histology specimens from biomicroscopy. Murray defined three estimates that are calculated using AM-FM processing: 1) instantaneous frequency magnitude (|IF|), 2) instantaneous amplitude (IA), and 3) the angle of the instantaneous frequencies (ψ).[9] In this section we define these parameters mathematically as well as conceptually in the context of the retinal images used in this study.

An image $I(x,y)$ is represented in terms of its AM-FM components using $I(x,y) \approx \sum_{n=1}^{M} a_n(x,y)\cos\varphi_n(x,y)$.

where M is the number of AM-FM components, $a_n(x,y)$ denotes instantaneous amplitude functions (IA) and $\varphi_n(x,y)$ denotes the instantaneous phase functions. Decomposing the image in this form allows one to perform an independent analysis of the characteristics of the image. For example, the amplitude modulation term $a_n(x,y)$ captures the contrast variations in the image while $\varphi_n(x,y)$ captures spatial phase information.

For each AM-FM component, the instantaneous frequency (IF) is defined in terms of the gradient of the phase φ, i.e. $\nabla\phi(x,y)$. Conceptually, IF is used to differentiate structures based on the gradient of the pixel-to-pixel phase. It allows one to model orientation variations or structures in an image region in terms of the direction of their gradient. In this study the estimates of the IF magnitude, |IF|, were used as one of the features. The angle between φ_x and φ_y is calculated to create other estimation for the extraction of features. This angle estimation is used to differentiate the inclination of the structure that could appear in the retinal image. To identify the structures of the retina, features vectors based on the histograms of each of the 3 estimates are created.

3.3 Frequency Scales and Filterbanks

One concept that is central to feature extraction is the idea of scales and filterbanks. Nine Combinations of Scales (CoS) are used to decompose the image using AM-FM. Though relatively new, the filterbank concept and techniques related to the AM-FM image processing methodology were introduced recently by Murray, Agurto et al., and Pattichis et al. and is repeated here as background necessary for understanding this new application.[9,4,8]

Figure 1 shows the separable filterbank supported in two quadrants. The frequency spectrum of an image is separated using 25 bandpass filters. Each scale, that contains 6 different bandpass filters, correspond to a range of frequencies. The high

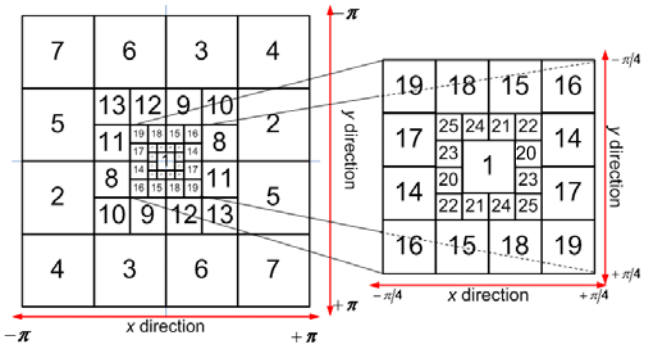


Figure 1. Filterbank and scales used to define features within images.

frequencies, filters from 2 to 7, correspond to the frequencies [π/2 π] for each direction (sign). Each additional scale corresponds to half of the range of the frequencies than the previous bandpass with higher frequencies than the first one.

At lower frequency scales, the magnitude values of the |IF| are small and the periodicity corresponds to image structures separated by several pixels. For example, the most appropriate scale for blood vessels is the one that captures frequencies with a period that is proportional to their width. On the other hand, the fine details within individual lesions, such as the small vessels in geographic atrophy structures, are captured by the higher-frequency scales. To capture the different scales, a separable filterbank that captures the whole frequency spectrum is applied.

The idea of scales and filterbanks also considers the size variability among structures such as drusen, hard drusen, confluent drusen, geographic atrophy, etc. A predominant characteristic of patients with AMD is that the lesion sizes will vary. The drusen sizes and other lesions will vary from 10s of micrometers to over a hundred micrometers. Other features, such as geographic atrophy can affect much larger areas of the retina. *Scales* are used to capture these features of different size. The different filters (within any given scale) also consider the orientation of the feature being encoded.

3.4 Evaluation of AM-FM feature for AMD application

The idea of scales and filterbanks also considers the size variability among structures such as drusen, hard drusen, confluent drusen, geographic atrophy, etc. A predominant characteristic of patients with diabetic retinopathy is that the lesion sizes will vary. The drusen sizes and other lesions will vary from tens of micometers to over a hundred micrometers. Other features, such as geographic atrophy can affect much larger areas of the retina. *Scales* are used to capture these features of different size. The different filters (within any given scale) also consider the orientation of the feature being encoded.

Table 1 relates the number of pixels and the frequencies ranges of each scale shown in Figure 1. Twenty-seven AM-FM estimates were computed corresponding to IA, |IF|, and ψ for nine combinations of scales. The filterbank used in this approach is composed of 25 filters that cover the whole spectrum.

Table 1. Bandpass filters from Figure 1.

Frequency band	Filters	Range in Pixels	Range in μm
Low-pass filter	1	22.6 to ∞	266 to ∞
Very low	20-25	11.3 to 32	113 to 32
Low	14-19	57 to 16	57 to 16
Medium	8-13	2.8 to 8	28 to 8
High	2-7	1.4 to 4	14 to 4

3.5 Classification of zygosity

For each combination of scales, three histograms are created, i.e. amplitude, frequency magnitude, and phase. This processing step will generate 11 histogram vectors with 96 elements each. In order to improve performance and generalization, the dimensionality of this problem must be reduced. Partial least squares was applied. A form of regression analysis, partial least squares (PLS), was used as the means for classifying the zygosity of the twin pairs based on the histogram features. Both principal component regression (PCR) and PLS share a similar structure based on reducing the matrix of independent variables to a lower dimensional sub-space. To better explain this, the regression problem is given by

$$y = X\beta + \varepsilon \tag{1}$$

where y is an $n \times 1$ vector of dependent variables, X is an $n \times p$ matrix of independent variables, β is a $px1$ vector of regression weights, and ε is an $nx1$ vector of residuals. The least squares solution to estimating β is given by the normal equations

$$\beta = (X'X)^{-1}X'y \tag{2}$$

When there are more variables than cases ($p>n$) or when the columns of X are highly correlated (multi-colinearity), both of which are the case in the analysis of FMRI activation maps, $X'X$ will be singular or nearly singular and a unique solution to the normal equations does not exist.

Both PCR and PLS are based on a reduction of X to a lower dimensional subspace ($k<<p$). The first step for both methods is to factor X as

$$X = T*L \tag{3}$$

where T is an orthogonal $n \times p$ matrix of T-scores and L is a $p \times p$ matrix of factor loadings. Such a factorization is not unique and the definitions of PCR and PLS are given by the criteria of the factorization. First, PCR factors X using Principal Component Analysis (PCA) which seeks to represent X with as few columns as possible by minimizing the mean square error between X and a subset of columns of T and rows of L:

$$X = T_1 L_1 + T_2 L_2 \approx T_1 L_1 \tag{4}$$

with the later approximation being a close as possible with as few columns of T as possible ($k<<p$). This then allows a re-framing of the regression equation as

$$y \approx T_1 L_1 + \varepsilon = T_1 \gamma + \varepsilon \tag{5}$$

Letting k be sufficiently small that $k<n$, and since the columns of T are orthogonal, the new normal equations are computationally easily solved.

Partial Least Squares is basically similar to PCR with two exceptions. First the factorization optimization criteria of PLS is different from that of PCA and second the dependent variables y can be a matrix (nxd) which will be useful in the analysis of FMRI activation maps. PCR relies on PCA which uses a spectral decomposition of $X'X$ to represent X as parsimoniously as possible. Unfortunately, the relationship of X to y which is the object of the analysis is not a part of the optimization. The matrix $X'X$ contains the correlation of the columns of X with each other and does not explain the relationship with y. One is somewhat serendipitously relying on the information about y to be in the first eigenvectors, but there is nothing to guarantee that this will be the case. Conversely, PLS relies on a spectral decomposition involving $X'y$ which contains the correlation of the independent variables X with the dependent variables y. So PLS tries to explain the relationship of X with y as parsimoniously as possible.

3.6 Application of AM-FM to twin data set for zygosity analysis

There were two different experiments conducted. The experiments used eyes from data set 1 for calibration and eyes from data set 2 for validation. For each twin pair a preferred set of two eyes, one from each twin, were identified in addition to the eyes rejected for quality. For all the experiments only the preferred pair of eyes was used in the validation data set to avoid a mixed prediction (one pair predicting dizygotic and the second pair predicting monozygotic). In the first experiment only the preferred set of eyes were used in the calibration data set while in the second experiment a modified bootstrap experiment was used where three of the calibration twin pairs (from data set 2) were held out at a time and the rest of data set 2 was added to the calibration data set.

The images were all analyzed using PLS to process the AM-FM generated histograms. For each image the center of the fovea was manually identified and a region of size 600x600 centered on the fovea was extracted for the analysis. The three AM-FM features IA, |IF| and Ψ were calculated for each of the 600x600 pixels and for each of the 11 CoSs. The features were then assembled for each CoS in the form of the three histograms with 32 bins resulting in 96 histogram bins for each of the 11 COSs. The dependent variables were coded as 0 for a dizygotic pair and 1 for a monozygotic pair.

In the first phase of the analysis the histogram bins for each CoS were analyzed separately. For each CoS a jackknife (leaving one twin pair out at a time) analysis was used to select the best model to predict zygosity. If the best model for that particular CoS included no more than 4 latent variables and an area under the ROC curve of at least 75% the CoS was retained for the second stage of the analysis. For the CoSs that were retained, a PLS model was calculated for the entire calibration data set and for the selected CoS. The selected latent variables were then retained for the second phase of the analysis as well as the loading matrix which would be used to calculate latent vectors in the validation data set.

The second phase of the analysis combined the prediction models from the selected CoSs. The latent variables were concatenated into a single matrix X_c which will constitute the independent variables for the second phase of the analysis

$$X_c = [T_1 : YT_2 : \ldots : T_k] \tag{6}$$

A jackknife (leave one twin pair out at a time) analysis was used to select the best model to predict zygosity based on the calibration input matrix X_c. The T matrices were calculated for the validation data set using the loading matrices obtained from the first phase of the analysis

$$T\text{-}hat = XL^* \tag{7}$$

The resulting final regression coefficients from the second phase of the analysis were then used with the estimated latent variables to predict zygosity in the calibration data set.

4. RESULTS

4.1 Application of AM-FM to AREDS standard Images

Attempting to develop an automated technique for defining AMD phenotypes was known to be a significant challenge. The retinal images are complex, presenting with highly variable normal and pathological structures. A much simpler problem is one of segmenting the lesions. This has been demonstrated by a number of investigators.[10] The next level of complexity is to classify structures based on mathematical features, such as those derived from the decomposition of the images using AM-FM. This experiment was conducted and showed that we can indeed classify these sample structures using only AM-FM features.

From the AREDS standard images, 36 regions were selected by a trained ophthalmic analyst to represent four types of retinal features: hard drusen, soft drusen, retinal vessels, and retinal background. Figure 2 shows the examples used for the training and testing of the AM-FM feature-based classification. Although, these were selected by a trained grader, it is clear that the two drusen classes have members which visually have characteristics that are similar. The original images from the database are in RGB format (Red-Green-Blue). As with our approach, most of the standard image processing methods for retinal images use the green layer only as input for the algorithms.

(a) hard drusen (b) soft drusen (c) retinal vessels (d) retinal background

Figure 2 – Samples of structures used in the AM-FM – based feature classification. (a) hard drusen, (b) soft drusen, (c) retinal vessels, and (d) retinal background

First, we compute the AM-FM estimates for each image using the 11 different CoS, as described in Table 1. For each ROI image (120 images total) and for each of the 11 CoS, we create a 96-bin feature vector per layer that contains the AM-FM histograms, i.e., |IF| and IA at each pixel. Next, for each CoS, we apply a linear regression using PLS to classify the images based on the AM-FM feature vectors.[11] PLS will use the output features from the AM-FM algorithms to classify the ROIs containing different structures. This linear regression method is based on linear transition from a large number of original descriptors to a new variable space based on a *small number of orthogonal factors (latent variables)*. In other words, factors of original descriptors are mutually uncorrelated (orthogonal) linear combinations.

After calculating the AM-FM histograms, we applied principal components analysis to the matrix containing histograms of the 120 ROIs (30 for each retinal structure) for three different AM-FM features: IA, |IF|, and IF Angle. PCA transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. For our experiments, we kept the principal components that accounted for 95% of the variability of the data. Finally, the Mahalanobis distances between the different retinal structures based on PCA decomposition are calculated. These distances are shown in Table 2.

Table 2. Mahalanobis distance between retinal features: Retinal Background (RB), Hard Drusen (DRH), Soft Drusen (DRS), and Vessels (VES).

	RB	DRH	DRS	VES
RB		3.37	5.56	5.34
DRH			2.82	4.42
DRS				4.5

The numbers in Table 2 represent the standard deviations separating the histograms of the retinal structures. The entries under DRH (hard drusen) and DRS (soft drusen) show that they are significantly differentiated from other structures in

the retina. A distance of 3 standard deviations represents a classification accuracy of greater than 90%. The most interesting entry is the Mahalanobis distance between the two drusen types (2.8 standard deviations). Though still considered a high classification rate (85%), this demonstrates the challenge not only to the algorithm, but for the grader in unequivocally assigning drusen to one class or another. Previously, Figure 2 showed the examples used for the training and testing of the AM-FM feature-based classification.

Figure 3 shows the plots of the cumulative density functions (cdf) for different retinal structures and combinations of scales. By definition, the cdfs completely describe the probability function of a random variable, in this case the variable being the AM-FM features. Figure 3(a) shows that instantaneous amplitude clearly differentiates the soft and hard drusen. The differences in their cdf are a result of the brighter cores of the soft drusen as compared to the hard drusen. (b) shows the cdf for the instantaneous frequency magnitude features that characterize the difference between hard drusen vessels. As one can see from the figures, the instantaneous amplitude is the AM-FM feature that best soft from hard drusen, while the instantaneous frequency magnitude best separates hard drusen and vessels. Similarly, (c) and (d) show that IA separates soft drusen from retinal background and the IF produces best results for separation between soft drusen and vessels.

One can hypothesize that since in the retinal background there is low pixel-to-pixel contrast variation, its IA has low values. Conversely, retinal structures that present with larger variations in contrast, such as drusen-background, will have higher values of IA. On the other hand, IA is not as effective at separating other structures with contrast changes such as retinal vessels. In this case, it is the instantaneous frequency magnitude that produces the largest difference in histograms because of the variations in frequencies between the structures.

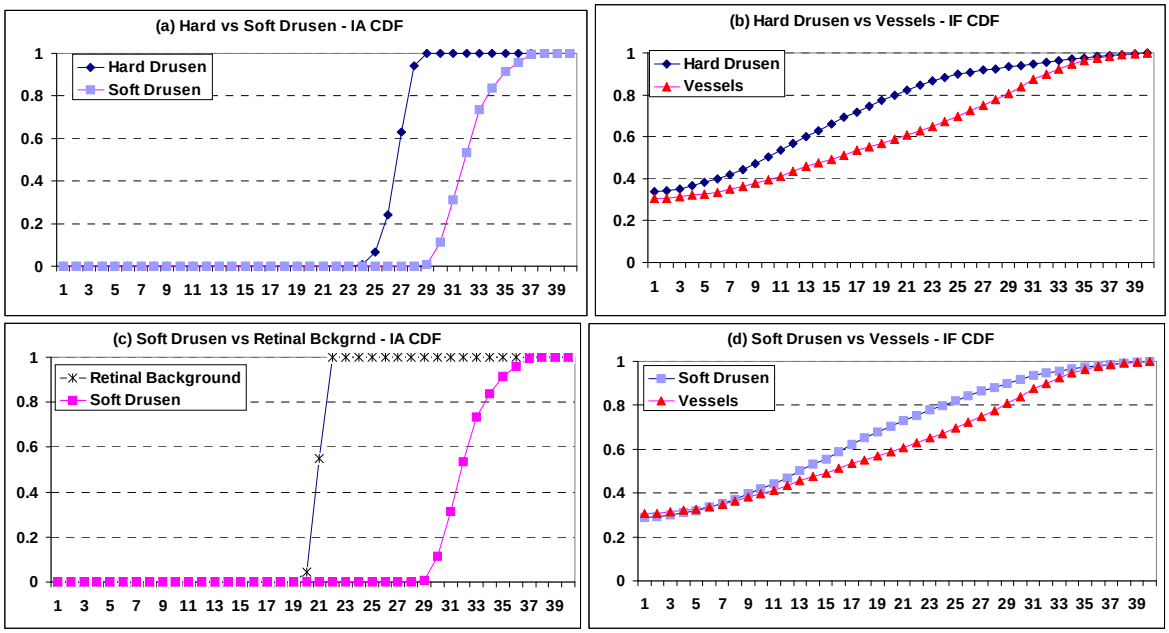


Figure 3. Cumulative Density Functions (cdfs) for drusen vs. other retinal structures. Instantaneous Amplitude (IA) cdfs clearly differenciate between hard and soft drusen (a) and between retinal background and drusen on (c). The instantaneous frequency magnitude (IF) is the feature that best differentiates drusen from retinal vessels on (b) and (d).

4.2 Twin Data Classification Results

In the training phase, a T-hat for each of the 56 sets was created for the calibration. Based on the training results a threshold between monozygotic (MZ) and dizygotic (DZ) was selected. In the validation, any twin set with a predicted value less than MZ-DZ threshold was classified as DZ, while greater than threshold was classified as MZ. The validation data used the best quality pair of images from each individual in a twin pair for data set 2. This preferred pair

was selected prior to performing the validation test. The accuracy was 72% in predicting DZ and MZ, where ground truth (MZ or DZ) was based on the genetic information for each twin set.

Eighteen twin pairs (72 images) were selected for testing. The analysis was performed using all acceptable pairs of eyes, i.e. not images rejected for quality or artifacts. There were up to 4 samples per twin pair yielding 72 possible image samples. A visual image quality evaluation determined that on 56 of the 72 pairs were of acceptable quality. Image quality was based on sufficient vessel sharpness and a view of 100% of the central region, i.e. macula/fovea. Figure 4 presents the $\hat{Y}$ for each of the 18 images in the test set. Based on these results a threshold between MZ and DZ was

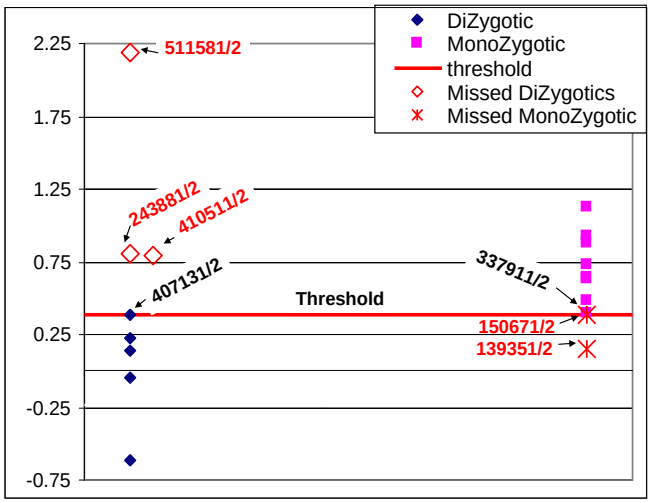


Figure 4 - Plot of the T-hat values for each of the validation images in Set 2. There three misclassified DZ, i.e T-hat valued greater than the threshold. There are two misclassified MZ, i.e. T-hat values less than the threshold. Two other values are very near the threshold, but are classified correctly.

Table 3. Summary of testing on Set 2 (mask, independent data set)

Zygosity	Correct	%
MZ	8 of 10	80.0%
DZ	5 of 8	62.5%
Total	13 of 18	72.2%

selected as 0.38 (marked as 'threshold' and a solid line in Figure 4. In the validation, any twin set with a predicted value less than 0.38 will be classified as DZ, while greater than 0.38 will be classified as a MZ.

The validation data is the preferred pair of eyes from each twin pair in data set 2. A preferred pair was visually selected prior to performing the validation test. The accuracy was 72% in predicting DZ and MZ based on the genetic information for each twin set.

5. DISCUSSION AND CONCLUSIONS

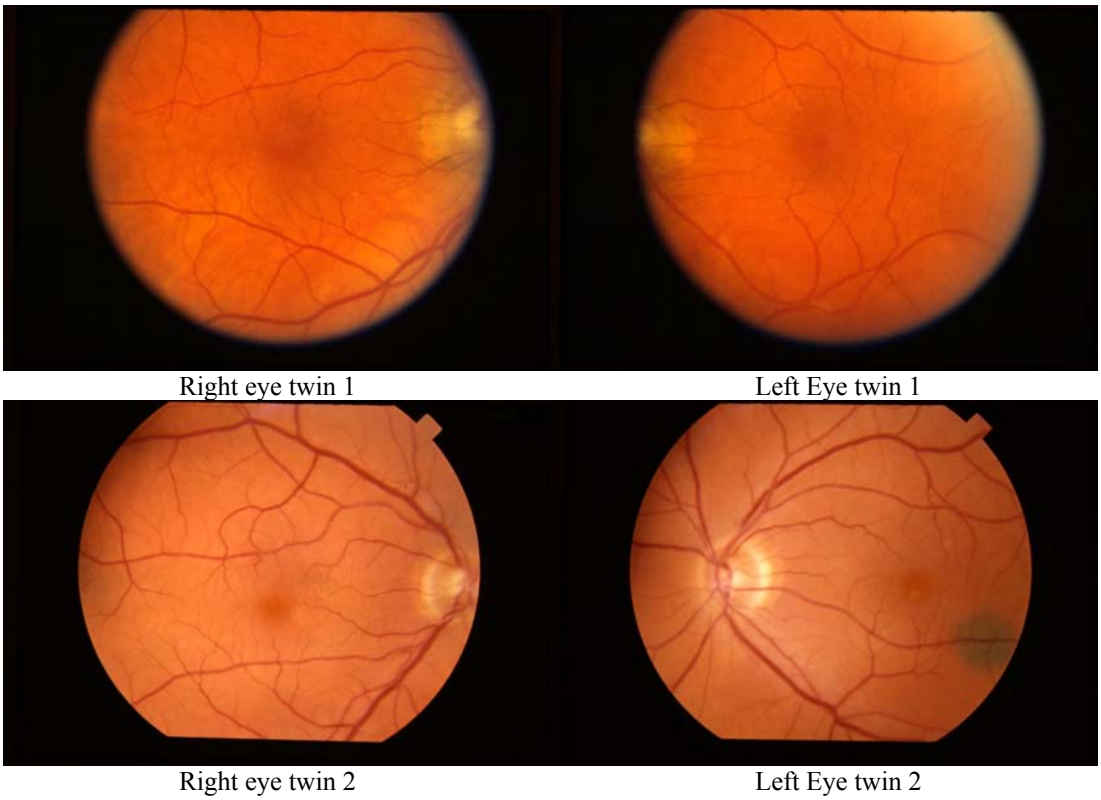
Our study has presented an entirely new method for automatic phenotyping. This approach can be extended to the automatic classification (grading) of retinal images from other large studies, such as the AREDS.[12] The result will be totally reproducible classification without the human grader variability issues present in today's studies.

In other applications the 72% accuracy in correctly classifying two categories would be unacceptable. However, in the case of classifying accuracy for MZ and DZ twin pairs, this may be within the limits of the data itself. Seddon et al. found that even human graders were unable to achieve classification accuracies greater than 65%.[3] They hypothesize that genetics, which is the driving factor, may only play the major role in determining the phenotype, but that environmental factors can alter how the genes alter the phenotype effect.

Additionally, from Figure 4, one can see that the separation of the DZ from the MZ cases at a threshold of 0.38 has cases at 0.24 that were classified correctly as DZ and 0.42 as MZ. The tails of the two distributions have a significant overlap. Nevertheless, the results are highly encouraging. It is expected that through the selection of new combination of filters that will better characterize the larger scale features of the drusen and their soft edges that results will improve. However, the accuracy will continue to be limited by what has been established through analysis of visual features; that is that environmental factors which contribute 25%-35% to the nature of the phenotypes and therefore may pose an upper bound to the best classification accuracy that is achievable.

5.1 Case study (243881/2)

The following two cases are dizygotic, but were misclassified as monozygotic. We start with case 243881/2 which had a $\hat{Y}$ error of 0.80, i.e. 0.42 above the DZ/MZ threshold of 0.38 (as determined by the calibration results). This is a case where although the twins are dizygotic, their disease presents in very much the same manner. Difficult to see at this printed resolution are a number of small drusen near the fovea (central dark region of the retina) in both twins. Except for differences in the images due to coloration due perhaps to film variations, the natural features and the pathology present in much the same manner. The most compelling evidence in the images similarity is in the observation that in the visual testing of the three individuals (two ophthalmologists and one image processing expert) who were masked from the genetically determined zygosity, all three experts mis-classified this twin pair as monozygotic.



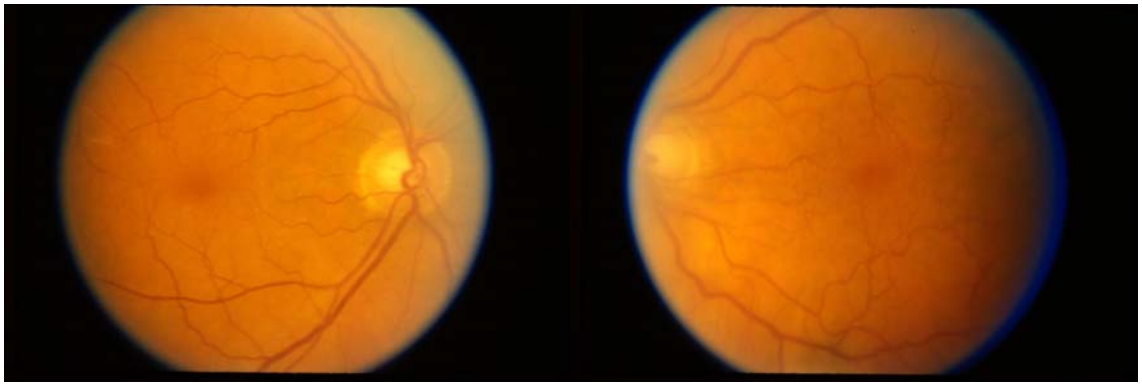
Right eye twin 1 Left Eye twin 1

Right eye twin 2 Left Eye twin 2

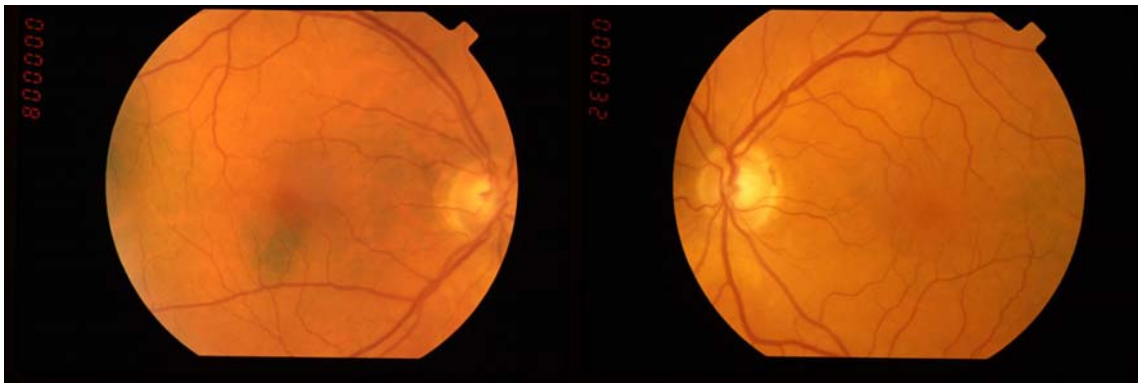
Figure 5 – Case 1. Twin pairs 243881/2 are DZ, but Mis-classified as MZ. $\hat{Y}$= 0.80

5.2 Case study (243881/2)

This is one case where the AM-FM features failed to detect the soft drusen in twin 1 (left eye temporal). The need for lower frequency filters are made apparent in this case. Because we have observed that AM-FM may not characterize soft edges, contrast enhancement may become necessary. However, to the untrained eye, like the previous case, this set of twins present normal and pathological features that are very similar. The AM-FM has not discovered new of sub-visual features that would classify this case as dizygotic.



Right eye twin 1 Left Eye twin 1



Right eye twin 2 Left Eye twin 2

Figure 6 - Twin pairs 410511/2 are DZ, but Mis-classified as MZ. $\hat{Y}$= 0.79

5.3 Discussion

The first obvious question is why the automated phenotyping produced a relatively low accuracy of 72%, when compared to the genetic information. Part of the answer lies in the nature of the experiment. Genetically inherited visual features of the AMD may explain only part of the disease characteristics found in the retinas. Seddon et al. concluded that significant percent of the variance in the staging of certain characteristics of the disease, such as drusen, could be statistically explained by genetics and attributed the remaining 19 to 37% to environmental factors.

There are other analytical reasons for misclassifying some of the images. The cases presented in Figures 5 and 6 show the retinal images from the two twins (a right and a left retina from each). In each case both of the twins presented with small and medium size drusen distributed throughout the macula. These are dizygotic twin pairs, which were consistently misclassified by ophthalmologists who were asked to visually classify the twin sets masked to the genetic information. In this case the AM-FM features for these two retinas were found to be similar, indicating a need for additional scales and larger sample regions for calculating the histograms. Further experimentation with additional scales found that this case could be correctly classified.

Acknowledgment: Academy of Sciences–National Research Council World War II Veteran Twin Registry, the participating physicians who performed the study examinations, and laboratory staff; J R Armstrong at the University of Wisconsin Reading Center fundus photograph grading expertise; and A M Rotunda, MD, Epidemiology Unit, Massachusetts Eye and Ear Infirmary, Boston; and J M Seddon, MD, Department of Ophthalmology, Harvard Medical School.

REFERENCES

[1] Brown MM. "Age-related macular degeneration: economic burden and value-based medicine analysis." Brown GC, Stein JD, Roth Z, Campanella J, and Beauchamp GR. Can J Ophthalmol 40, 277-287. 2005.

[2] "A Simplified Severity Scale for Age-Related Macular Degeneration. AREDS Report #18," Arch Ophthalmol, Vol 123, Nov 2005

[3] Seddon, JM; Cote J; Page WF; Aggen SH; Neale MC, "The US Twin Study of Age-Related Macular Degeneration: Relative Roles of Genetic and Environmental Influences," Arch Ophthalmol. 2005;123:321-327.

[4] Agurto, C., S Murillo, V Murray, M Pattichis, SR Russell, P Soliz, "Detection and Phenotyping of Retinal Disease using AM-FM Processing for Feature Extraction." Asilomar Conference on Signals, Systems, and Computers, Asilomar, Oct 2008.

[5] Murray, V.; Pattichis, M.; Soliz, P., "New AM-FM analysis methods for retinal image characterization," Signals, Systems and Computers, 2008 42nd Asilomar Conference on26-29 Oct. 2008 Page(s):664 – 668

[6] Available online at: http://eyephoto.ophth.wisc.edu/ResearchAreas/AREDS/AREDSstdPhotoIndex.htm.

[7] V. Murray, P. Rodriguez, and M. S. Pattichis, "Multi-scale AM-FM demodulation and reconstruction methods with improved accuracy," submitted to the IEEE Transactions on Image Processing, 2008.

[8] M. S. Pattichis, C.S. Pattichis, M. Avraam, A. Bovik, and Kyriacos Kyriacou, "AM-FM Texture Segmentation in Electron Microscopy Muscle Imaging," IEEE Transactions on Medical Imaging, Vol. 19, No. 12, Dcember 2000.

[9] V. Murray Herrera, "AM-FM methods for image and video processing," Ph.D. dissertation, University of New Mexico, 2008.

[10] K. Rapantzikos, M. Zervakis*, K. Balas, "D etection and segmentation of drusen deposits on human retina: Potential in the diagnosis of age-related macular degeneration," Medical Image Analysis 7 (2003) 95–108

[11] Wold, H., "Personal memories of the early PLS development," Chemometrics and Intelligent Laboratory Systems, 58:83–84, 2001. Wold, H., "Personal memories of the early PLS development," Chemometrics and Intelligent Laboratory Systems, 58:83–84, 2001.

[12] The Age-Related Eye Disease Study Research Group. "The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS Report No. 17." Arch Ophthalmol. 2005;123:1484-1498.