STATISTICAL VALIDATION OF AN AUTOMATIC ALGORITHM FOR DIABETIC RETINOPATHY SCREENING

#4105

ES Barriga[1,2], V Murray[2], C Agurto[1,2], G Zamora[1], H Yu[1], J Wigdahl[1], S Nemeth[1], W Baumann[3], and P Soliz[1,4];
[1]1VisionQuest Biomedical, LLC, Albuquerque, NM, [2]University of New Mexico, Department of Engineering and Computer Engineering,
[3]Retina Institute of South Texas, [4]University of Iowa, Department of Ophthalmology and Visual Sciences

Purpose

Objective

- To present and update the results of a population-based statistical validation of an automatic computer system that screens retinal photographs for the presence of lesions related to diabetic retinopathy (DR).

Background

- Studies have shown that less than 50% of diabetics in the U.S. receive their recommended yearly eye exams.
- Automatic systems for DR screening could increase the amount of yearly examinations by boosting the throughput of screening centers and bringing screening to remote and underserved populations.
- In this work, we present results of the statistical validation of a DR screening algorithm previously tested in retrospective databases[1].

Dataset

- Images for this study were collected at VisionQuest's DR screening sites in New Mexico and San Antonio, TX.
- The dataset is comprised of 487 cases collected with a non-mydriatic Canon CR1 Mark II camera.

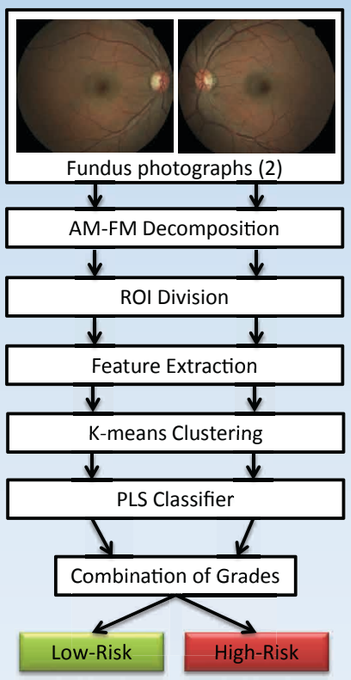

Figure 1. Block diagram of screening algorithm function

1. C. Agurto, E.S. Barriga, V. Murray, S. Nemeth, R. Crammer, W. Bauman, G. Zamora, M.S. Pattichis and P. Soliz, " Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images." Invest. Ophthalmol. Vis. Sci. June 10, 2011 iovs.10-7075.

Methods

- A block diagram of the screening algorithm is shown in Figure 1. The main elements of the system are AM-FM decomposition, K-means clustering, and the PLS classifier.
- AM-FM is a mathematical technique used to represent signals or images in multiple frequency decompositions. This technique has been used previously for detecting features in retinal images.
- K-means clustering is used for dimensionality reduction and for clustering retinal regions into self-similar groups.
- Partial Least Squares (PLS) is a regression technique that can be used as a fast and efficient classifier.

Training and Testing

- 388 cases were used for the training and testing of the system.
- The dataset is randomly divided, using 70% of the data for training and 30% for testing.
- Training was repeated 20 times to reduce the bias from the random selection of training and testing sets.
- We obtained a per-image classification area under the ROC curve (AUC) of 0.87.
- Figure 2 shows the ROC curves for three of the 20 runs. These three curves depict the best, worst, and median performing models.

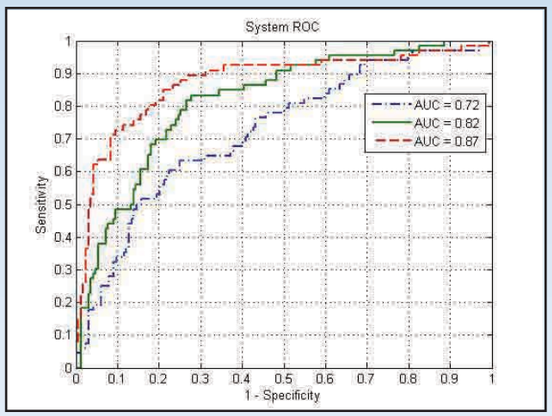

Figure 2. ROC curves for per-image performance of the DR screening system.

Results

- Of the 20 models obtained we selected the best performing to be our operating system.
- During the training/testing phase, this model had an AUC of 0.87 with sensitivity (sens) = 0.92 and specificity (spec) = 0.76.
- The model was tested on 99 images not previously seen by the system.
  - --Normal controls: 62
  - --DR: 24
  - --Sight Threatening DR (STDR): 7
  - --DR and Age-related Macular Degeneration (AMD): 5
- The system achieved sens = 0.95 and spec = 0.61.
- All cases with STDR were correctly classified.
- Figures 3 and 4 show examples of false negative images. Only images with small microaneurysms (MAs) were missed by the system.

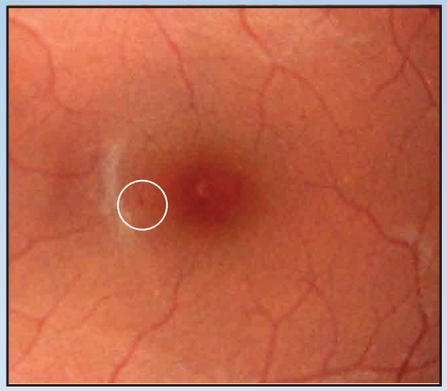

Figure 3. False negative (FN) image zoomed in to show the presence of an MA in the macula (circled).

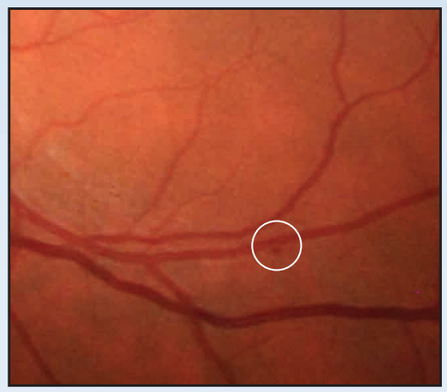

Figure 4. . False negative (FN) image zoomed in to show the presence of an MA next to a major retinal vessel (circled).

Results

Statistical Validation Design

- Independent statistical validation will be completed in Q3 2012.
- We need to reject the area of the null hypothesis where:

$$H_0 = \{sens < 0.80,\ spec < 0.50\}$$

- Our target performance is sens > 0.90 and spec > 0.60.
- Figure 5 illustrates our current ROC curves and the areas of the null hypothesis (red) and of rejection of the null hypothesis (green).

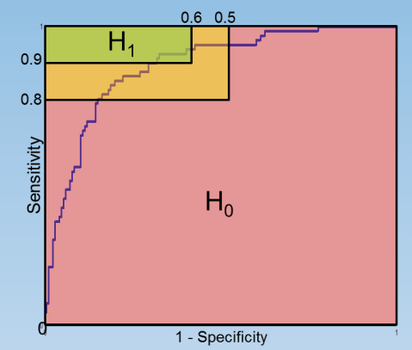

Figure 5. ROC curve (in blue) and areas of null hypothesis (H0) and of rejection of the null hypothesis (H1).

Conclusions

- Improved AM-FM methodology and a combination of individual grades have resulted in improvements in the algorithm performance.
- We validated this algorithm in an independent set of images, obtaining sensitivity and specificity values of 0.95 and 0.61, respectively.
- A final statistical validation with a population-based sample will be performed to demonstrate the statistical significance of the diagnostic test.

Acknowledgements

- This work was funded by NEI grants EY018280 and RC3EY020749.
- Commercial Relationship: Barriga, Agurto, Zamora, Yu, Wigdahl, VisionQuest Biomedical E; Murray, University of New Mexico, F; Baumann, Retina Institute of South Texas, I; Soliz, VisionQuest Biomedical, I.

Contact: Simon Barriga, Ph.D.
(505) 508-1994
sbarriga@visionquest-bio.com
VisionQuest Biomedical LLC
2501 Yale Blvd. SE Ste. 301
Albuquerque, NM 87106