35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Retina Image Analysis and Ocular Telehealth: The Oak Ridge National Laboratory-Hamilton Eye Institute Case Study

Thomas P. Karnowski, Ph.D., *Member, IEEE,* Luca Giancardo, Ph.D., *Member, IEEE,* Yaqin Li, Ph.D., *Member, IEEE,* Kenneth W. Tobin, Jr., Ph.D., *Fellow, IEEE,* Edward Chaum, M.D., Ph.D., *Member, IEEE*

Abstract— **Automated retina image analysis has reached a high level of maturity in recent years, and thus the question of how validation is performed in these systems is beginning to grow in importance. One application of retina image analysis is in telemedicine, where an automated system could enable the automated detection of diabetic retinopathy and other eye diseases as a low-cost method for broad-based screening. In this work, we discuss our experiences in developing a telemedical network for retina image analysis, including our progression from a manual diagnosis network to a more fully automated one. We pay special attention to how validations of our algorithm steps are performed, both using data from the telemedicine network and other public databases.**

I. Introduction

Diabetes afflicts more than 25 million people in the United States of America, with a projected increase to 115 million in the next four decades [1]. Diabetic Retinopathy (DR), the leading cause of blindness in the industrialized world, is one of many complications that can arise from diabetes. Thus there is a need for achieving inexpensive, broad-based screening for DR, which has led to a variety of image processing and pattern recognition approaches. For the purposes of this paper, we refer to this work as "Retina Image Analysis" (RIA) and it includes work in automated screening as well as image processing based tools to enhance measurements of physiology. During the past several years, these algorithms and the systems that use them have begun to reach high levels of maturity, and much of the published literature in RIA has shifted from algorithms that detect the signs of eye disease or eye features to algorithms that generate measurements of ocular conditions or estimate the probability of eye disease [2].

Many efforts in this field have originated as a collaborative effort between an ophthalmologist or group of ophthalmologists and a team specializing in machine or computer vision. This is certainly true of our case, as our collaboration involves researchers at Oak Ridge National Laboratory (ORNL) and the University of Tennessee Health Sciences Center Hamilton Eye Institute (HEI). Collaborations such as ours have led to interesting and innovative approaches, but it has also led to a large body of "private" data that cannot be shared between research groups for a variety of reasons. This has led to problems, ranging from the ever-present issue in machine-based detections (the fact that there is never enough good quality, ground-truth data) to more fundamental questions as to how repeatable or transferrable different algorithmic approaches are between different patient populations and imaging devices. Recognition of this situation has led to a consensus paper published by 14 international RIA research groups [3]. "Validation" in this context refers to determining the true effectiveness of the systems that are used to perform RIA.

There are a large number of applications in computer vision and machine learning where public data sets exist with a goal toward furthering the state-of-the-art. This includes data sets of hand-written images [4] and facial images [5], among others. These data sets have advanced these fields by providing common standards to evaluate algorithms and approaches. There are also other "Grand Challenge" type problems such as multi-modal biometrics [6] and visual object recognition [7] among others. Standardized training and testing data are created, then a set of secret validation data are generated for testing purposes. Participants must embed their algorithms within a framework for ease of testing by the challenge coordinators. These problems also often have annual revisions where new data and new annotations are introduced. This type of approach would serve many domains well, provided there is a sufficient body of research groups to warrant the efforts required to generate, tag, and organize the data sets. But such an approach may not be feasible for highly specialized domains such as ophthalmology, where it is not always obvious to the layman that there is actually an issue with a particular retina. By comparison, for example, virtually anyone could properly ground-truth a data set of face images and non-face images. Currently the closest RIA equivalent is the Retinopathy Online Challenge from the University of Iowa [8].

In this paper, we describe our RIA application, which is a telemedical network designed to detect retina disease. We then describe the individual components of automation and our experience with the problem of validating our work, both in terms of making an effective detection system and in terms of proving our work to the research community.

This work supported in part by the National Eye Institute of the National Institutes of Health (EY017065), Research to Prevent Blindness, New York, NY, and the Plough Foundation, Memphis TN. This paper was prepared by OAK RIDGE NATIONAL LABORATORY, Oak Ridge, TN, USA, 37831-6285, operated by UT-BATTELLE, LLC for the US DEPARTMENT OF ENERGY under contract DE-AC05-00OR22725.

T.P. Karnowski and K.W. Tobin, are with ORNL, Oak Ridge, TN 37831 USA; E. Chaum and Y. Li are with the University of Tennessee Health Science Center, 930 Madison Avenue, Suite 731, Memphis TN 38163 where E. Chaum is the Plough Foundation Professor of Retinal Diseases and an RPB Senior Scientist; and L. Giancardo is with the Istituto Italiano di Tecnologia, Genova, Italy. K.W. Tobin and E. Chaum have ownership of Hubble Telemedicine, which is commercializing the TRIAD technology.

II. Background

Our objective in this work has focused on an application for ophthalmic screening: the Telemedical Retinal Image Analysis and Diagnosis (TRIAD) Network, based in the Mid-South region of the United States [9]. TRIAD provides retina screening to diabetic patients in a primary care clinic environment in communities with disparities in healthcare. TRIAD has been a functional telemedicine network since Feb 2009. A supervisory ophthalmologist (SO) reviews each patient eye submitted. Fundus images of diabetic patients are submitted to a workflow process through TRIAD's secure, HIPAA-compliant architecture and stored in a database. A notification is sent to the SO for the clinic, who prepares a recommendation delivered via secure protocols to the general practitioner. There is a single automated component in this configuration, the quality assessment (QA) of the captured images [10]. Over time, greater levels of automation have been introduced: anatomical feature localization [11], lesion detection [12-14], and diagnosis [15]. The very nature of the network creates a useful validation mechanism, as any algorithms which are developed on a set of data prior to a particular date, that utilize training data from before that date, can be validated by using test data collected AFTER the date of interest. This method should work to improve performance over time, especially when using on-line supervised learning approaches as more data becomes available.

III. Feature and Anomaly Detection

The different automated components of the TRIAD system each have different levels of validation needs and issues due to both our approach and the complexity of the task (both in terms of technical difficulty and the level of domain knowledge needed to perform validation). In this section we discuss the validation process utilized for image quality assessment, vascular detection, optic nerve (ON) and macula localization, lesion detection and disease stratification. We conclude the section with some special issues regarding confidence metrics and temporal or longitudinal studies.

A. Image Quality Assessment

A diabetic patient who receives services at a primary care clinic served by the TRIAD Network can have their retinas scanned by an operator using a state-of-the-art non-mydriatic fundus camera. However, the quality of this image must be verified, as even these relatively easy-to-use cameras can produce images that are out of focus or non-uniformly illuminated. The method [10] consists of vascular segmentation [16], then a variety of measurements in localized neighborhoods in the retina image. These measurements were used to train a support vector machine to map features to manually assigned quality rankings. During the development hold-one-out validation was used for training and testing. An additional set of images, which were completely independent of the algorithm development set, was utilized to investigate the effectiveness of the quality assessment (QA) method. The QA module output is compared to a simple threshold and the image is graded passing or failing. Any images that fail QA trigger a re-acquisition request. However, the operator can decide that the image is as good as possible, which also allows for cases where the QA may not be accurate, the threshold level for the image may be too conservative. Also, while the QA module maps the retina to a continuous functional ranking of quality, comparison to the threshold value is the main criteria and thus we simply have to make a yes/no decision on the image quality, which simplifies the validation process. Post-development, QA is tracked by the SO. We have found that over time the values used have proven effective in triggering the re-imaging of patients in the clinics, and images above the threshold have unfailingly been sufficient quality for diagnosis. Indeed, in some cases we have found that a clinic that begins to deliver lower quality images has had a personnel change that requires additional training. With respect to validating the method, we rely on the continuous acquisition and QA of images received post-development.

B. Vascular Detection

A large number of retina research groups have performed interesting work in the detection, tracing, and measurements of vascular networks in the retina. In the TRIAD system, however, to date the vascular segmentation has been a means of evaluating the image quality and locating the optic nerve and macula. Consequently, the validation process for vascular detection accuracy has largely been unstudied by our group.

C. Optic Nerve and Macula Detection

The ON and macula detections establish a coordinate system for the retina and eliminate a potential source of false positives. The ON and macula localization is described in [11]. Four neighborhood features are measured from the image intensity and vascular segmentation on a pixel-by-pixel basis. A Gaussian model is used to estimate the probability of a pixel residing in the ON, after training with manually labeled ON centers. The manually labeled ON centers are also used to form an *a priori* estimate of the probability density function (PDF), so that the Gaussian parameters and PDF are used to compute a likelihood ratio function using maximum *a posteriori* (MAP) estimation. The pixel of the highest likelihood is declared the optic nerve center, then by modeling the vascular tree as a parabola the macula is found by assuming a fixed distance from the ON at the angle indicated by the parabolic fit. Note that the *a priori* estimate is enhanced by the fact that modern fundus cameras label the image as right and left-eye.

In our early development we focused on a custom dataset from an ophthalmology practice, which was not representative of images from a screening environment. As a measure of success for ON location, we used histograms of Euclidean distance between the ground-truth position and the estimated position, often normalized by the mean ON radius; thus any values below 1 indicated the estimated position resided on the ON. In our experience, establishing the ground-truth locations for the ON and macula is straightforward and can be performed satisfactorily by non-clinicians. Once we were satisfied with our approach, we

used the publicly available STARE dataset [17] for further testing as discussed in [11]. However, we note that this set is somewhat challenging and also does not represent the data seen in a typical screening environment, where the algorithm performs much better (greater than 98% success). We have tested the algorithm on TRIAD data post-development [18] and we have also used [14] the MESSIDOR [19] data set with ground-truth established by our team. Thus in this measurement we have relied on public data sets and our own post-development data sets for validation.

D. Lesion / Anomaly Detection

The detection of lesions and anomalies is the most important phase of this activity. Our approach has been to develop custom detectors for the most prominent lesion types, which are then optionally post-processed to reduce false positives, and amalgamated into a set of lesion population features which are used to generate a vector descriptor of the image for supervised learning.

In early work, we relied on the image set from an ophthalmic practice, with ground-truth by E. Chaum. Our initial goals were to test the feasibility of a set of descriptors of lesions and their amalgamation into the population vector, in a content-based image retrieval (CBIR) method [15], with manual lesion segmentation. As our work advanced, we began to explore automated methods for lesion detection and relied on image sets from the TRIAD network. Some initial ground-truth was created for a small number of these images (137 images), and we have relied on those ground-truth sets for our further development. We have found several issues with ground-truth data in this context. First, it is impossible to perfectly duplicate a human segmentation (and ground-truth is often inconsistent between experts). Instead, a measure of success or true positive has been defined as simply detecting the anomaly. Second, in many images where a lesion type is identified, particularly a case where a patient has many examples of the lesion, it is common for the ground-truth to miss lesions. This causes confusion in the training set because a machine-learning algorithm could identify a true lesion that the clinician missed. To circumvent this problem, we used ground-truth lesions as examples of "true positives" and then obtained examples of "true negatives" from images with a normal state (no signs of DR, AMD, or other diseases) that were confirmed as having no lesions present. A third and final issue is simply the shear difficulty of obtaining ground-truth on retina lesions. Ophthalmologists are an extremely limited resource and perform extremely important tasks regarding the health and quality of life of people daily, so finding and making time to perform ground-truth can be difficult to achieve. It is also fair to say that the medical profession in general has not embraced the possibility of automated screening or even telemedicine, and this too can cause issues as there has not been a dedicated effort or push to produce the type of data needed in machine learning applications. Despite this difficulty, we have pressed on with the ground truth data available and steered our efforts towards detectors that use threshold methods, and thus have not relied on post-processing to remove false-positives. We describe each detection method we have used in some detail below.

1) Microaneurysm and Hemorrhages

Microaneurysms – dilated blood vessels that appear as small dark spots in retina images – and blood dot hemorrhages are the primary indicators of DR. In the TRIAD data set, roughly 15% of patient eyes contain microaneurysms based on the diagnosis of the entire retina. Our research in detection methods for these lesions has utilized a radon transform approach [12] with a supplemental supervised-learning method applied adaptively and interactively, minimizing the need for clinician ground-truth. We have also explored the use of semi-supervised learning [20] for detection. For validation, we have relied on the ground-truth data available with our TRIAD data, and also have used the Retinopathy Online Challenge [8], which is an extremely important example of a public dataset for lesion detection validation.

2) Exudates

Diabetic Macular Edema (DME) is a complication of DR; it is a swelling of the retina due to fluid leakage from chronic damage due to elevated blood sugar levels. The primary means of detecting DME in monocular retina images is inference through the presence of exudates, lipid deposits that appear as bright lesions with well-defined edges in the retina. In the TRIAD data set, roughly 4.5% of patient eyes contain exudates based on the diagnosis of the entire retina. Our method for exudate detection utilizes a sequence of specialized image processing steps as described in [13, 14]. Additional research into the mapping of exudate detection into a feature vector capable of identifying the overall condition of DME has also been performed [14]. Our team has used our ground-truth data for exudates, along with some additional normal data sets, and created the HEI-MED data set which is publicly available for download [21]. This set of 169 images features ground-truth data, including the quality scores using our QA method, the ground-truth ON and macula locations, clinician exudate segmentation, and overall disease diagnosis. Validation of the DME detection processing has been performed using three sets: the HEI-MED set, MESSIDOR, and DIARETDB1 [22]. The latter two sets were useful for this validation, because they include diagnosis results for the image, although they do not have the ethnic variability of TRIAD and HEI-MED. The DME detection processing was also tested with the MESSIDOR data set using inner-rater statistical measures (Kappa and AC1) with two ophthalmologist collaborators[14].

3) Drusen

Drusen are subretinal pigment epithelial deposits that resemble yellowish blobs in retina images. They can be characteristic of age-related macular degeneration (AMD), but can be found in images diagnosed as "normal". [24]. Drusen tend to be less common in TRIAD and images diagnosed with age-related macular degeneration (AMD) are rare, composing only 0.5% of the images. A substantial number (4%) of normal retinas show some signs of drusen. We have performed some initial research in drusen detection [23]. This work used data from the second Age Related Eye Disease Study (AREDS) for drusen progression examples, and the TRIAD network for examples of non-drusen images

and non-drusen progression image sequences. The detection method was based on color transformations and statistical models of drusen and retina wall structures. Although we believe our preliminary work in this area shows promise, the validation methods are limited.

4) Other Anomalies

Other retina conditions fall outside the previously described anomalies. In TRIAD, approximately 2.5% of the images show signs of other disease types that require some means of detection. Validation of these types of detectors is problematic, given the limited number of data.

F. Other System Issues

The diagnosis of the disease state is the ultimate goal of the system, and to this point the validation has relied on standard hold-one-out methods and post-development data from the TRIAD method, with the exception of the DME estimation discussed previously. As discussed in [2], for safety issues systems such as TRIAD in automated mode should minimize false negatives, with a trade-off of more false positives, which could be screened by the SO. Thus, we have investigated means of assigning a confidence to the various measures made by the automated components. These include a confidence for disease stratification [15] based on Poisson statistics, and a confidence for optic nerve location using complementary methods [18].

Finally, we address the issue of temporal changes for recurring patients, which we envision undergoing a comparison with their previous visits. The data obtained up to September 2012 shows 18% of patients have returned at least one time, however very few showed disease progression. Handling temporal change is therefore a particularly difficult exercise in validation. One possible method for creating test and training data for such cases is the introduction of artificial lesions into the subsequent retinas of returning patients. This is a rather imperfect approach, however it may well be the only possible validation method without acquiring more data and finding actual cases of disease progression.

IV. CONCLUSION

In this work we have reviewed the approach taken by the ORNL-UTHSC team in developing a telemedical network for the automated detection of DR and other eye diseases. With respect to validation, our work has shown the importance of public databases to develop algorithms and verify their performance. In our experience, due to its nature TRIAD could be developed without public data sets, the effectiveness of the algorithms is hard to compare with other approaches if only private data is considered. The difficulty of generating useful public data, however, is understandable given the application domain and other factors, such as privacy of protected health information (HIPAA compliance) and intellectual property concerns. In addition to computer vision researchers, ophthalmic medical groups should be engaged as well to bring about effective solutions.

ACKNOWLEDGMENT

The authors would like to acknowledge team members Karen Matthews from the Delta Health Alliance, Scott Morris and Beth Sparrow from the Church Health Center, Memphis TN, Seema Garg of the University of North Carolina; from ORNL, T. Nichols, Vincent Paquit, Hector Santos-Villalobos, Deniz Aykac, Bobby Whitus and Priya Govindasamy, and from the University of Burgundy: Kedir Adel, Shar Mohammed, Desire Sidibe, and Fabrice Meriaudeau.

REFERENCES

[1] J. Boyle et. al., "Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence", *Population Health Metrics*, v. 8:, p 29 2010

[2] M.D. Abramoff M. Niemeijer, and S.R. Russell, "Automated detection of diabetic retinopathy: barriers to translation into clinical practice", *Expert Rev. Medical Devices*, v. 7, pp 287-296, 2010

[3] E. Trucco, et al, "Validating retinal fundus image analysis algorithms: issues and a proposal", (Accepted for publication), J. IOVS 2013

[4] Y. LeCun et al., "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, v. 86, pp. 2278-2324, 1998

[5] http://www.nist.gov/humanid/feret/

[6] http://www.nist.gov/itl/iad/ig/mbgc.cfm

[7] http://pascallin.ecs.soton.ac.uk/challenges/VOC/

[8] http://roc.healthcare.uiowa.edu/index.php

[9] Y. Li, et.al., "TRIAD: A HIPPAA-Compliant Ocular Telehealth Network for the Remote Diagnosis and Management of Diabetic Retinopathy", *Telemedicine and E-Health* Oct;17(8):627-34. doi: 10.1089/tmj.2011.0004, 2011

[10] L. Giancardo, L., et.al., "Quality assessment of retinal fundus images using ELVD", in *New Developments in Biomedical Engineering*, D. Campolo, Ed. INTECH, 2010, pp 201-224

[11] KW Tobin, et al, "Detection of anatomic structures in human retinal imagery", *IEEE Trans Med Imaging* 26:1729-39, 2007

[12] L. Giancardo, et. al, "Microaneurysms detection with the Radon Cliff operator in retinal fundus images", *Proc. SPIE Medical Imaging*, 7623:7623, 2010

[13] L. Giancardo, et. al,, "Automatic retina exudates segmentation without a manually labeled training set", *Proc. IEEE International Symposium on Biomedical Imaging*, 2011

[14] L Giancardo, et al, "Exudate-based Diabetic Macular Edema Detection in Fundus Images Using Publicly Available Datasets", *Medical Image Analysis* 16(1), 216—226, January 2012.

[15] E. Chaum, E. "Automated diagnosis of retinopathy by content-based image retrieval". *Retina* 28:1463-1477, 2008

[16] F. Zana and J. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation", *IEEE Trans. Image Processing*, 10:1010-1019, 2001

[17] A. Hoover and M Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels", *IEEE Trans Med Imaging* 22:951-958, 2003

[18] Karnowski, T., et al "Automatic detection of retina disease: robustness to image quality and localization of anatomy structure", *33rd Annual International Conf. of the IEEE EMBS* Boston, MA, 2011

[19] http://messidor.crihan.fr/description-en.php

[20] K Adal et al, "Automated detection of microaneurysms using robust blob descriptors", *Proc. SPIE Medical Imaging* 8670, 2013

[21] http://vibot.u-bourgogne.fr/luca/heimed.php

[22] http://www2.it.lut.fi/project/imageret/diaretdb1/

[23] H. Santos-Villalobos, et al, "Statistical Characterization and Segmentation of Drusen in Fundus Images," *33rd Annual International IEEE EMBS Conference*, 2011.

[24] WR Green, PJ McDonnell, JH Yeo, "Pathologic features of senile macular degeneration", *Ophthalmology,* 92(5):615-627, 1985