Diabetic Retinopathy Image Quality Assessment, Detection, Screening and Referral

Ramon Pires[1], **Anderson Rocha**[1]

[1]Institute of Computing – University of Campinas (UNICAMP)
13.083-852 – Campinas – SP – Brazil

{pires.ramon,anderson}@ic.unicamp.br

Abstract. *Diabetic Retinopathy (DR), a common complication caused by diabetes, manifests through different lesions that can result in blindness if not be discovered in time. In this work, we present a general framework whose objective is to automate the eye-fundus image analysis. The work comprises four steps: image quality assessment, DR-related lesion detection, screening, and referral. In each step, we provide satisfactory results, comparable to the State of the Art and, in many cases, surpassing it, especially when dealing with hard-to-detect lesions. An important advance in our work is the validation protocol, the cross-dataset, which is closer to real situations. Furthermore, we proposed a Bag-of-Visual-Words representation highly suitable to retinal image analysis.*

1. Introduction

Diabetes mellitus (DM) is a chronic end-organ disease caused by a decrease in insulin sensitivity or a loss of pancreatic function, depending on the type of diabetes, both leading to an increase in the blood glucose level. An increased blood sugar level may lead to damage of blood vessels in all organ systems of the body. Currently, diabetes affects 366 million people worldwide or 8.3% of adults. It is estimated that this number will increase to approximately 552 million people (one adult in 10 worldwide will have diabetes)

The growing prevalence of diabetes creates an increasing prevalence of the complications related to the disease, including *Diabetic Retinopathy* (DR). DR occurs in approximately 2-4% of the population but is greater in indigenous populations. DR is the main cause of blindness in the 20 to 74 age group in developed countries, creating the need for systems that screen diabetic retinopathy in its early stages, so to allow an economically viable management of the disease.

In this context, we present herein a complete step solution for diabetic retinopathy image quality assessment, detection, screening and referral. In the first step, we apply characterization techniques to assess image quality by two criteria: field definition and blur detection. In the second step, we propose an approach for detection of any lesion, in which we explore several alternatives for low-level (dense and sparse extraction) and mid-level (coding/pooling techniques of bags of visual words) representations, aiming at the development of an effective set of individual DR-related lesion detectors. The scores derived from each individual DR-related lesion, taken for each image, represent a high-level description, fundamental point for the third and fourth steps. Given a dataset described in high-level (scores from the individual detectors), we propose, in the third step of the work, the use of machine learning fusion techniques aiming at the development of a multi-lesion detection method. The high-level description is also explored in the fourth

step for the development of an effective method for evaluating the necessity of referral of a patient to an ophthalmologist within one year.

2. Datasets

Two different datasets tagged by medical specialists, DR1 and DR2, were used to perform the experiments. In the experiments with cross-dataset protocol, DR1 is used for training while DR2 is the test set. The datasets were created by the Department of Ophthalmology, Federal University of Sao Paulo (UNIFESP).

DR1 has 1,077 retinal images with an average resolution of 640 $\times$ 480 pixels of which 595 images are normal and 482 images have at least one disease. Each image was manually annotated for DR-related lesion (presence/absence) by three medical specialists, and only the images for which the three specialists agree were kept in the final dataset.

DR2 comprises 520 images with 12.2 megapixels cropped to 867 $\times$ 575 pixels to increase the processing speed. Among the 520 images, 300 are normal and 149 have at least one lesion. Ignoring the specific lesion that can be present, 337 images have been manually categorized by two independent specialists as not requiring referral and 98 images require referral within one year. Both datasets are freely available through FigShare repository, under URL `http://dx.doi.org/10.6084/m9.figshare.953671`.

3. Quality Assessment

Image quality is an important aspect of automated image analysis and the factor that successful image analysis relies on. Although it is a common task in lesion detection projects, the manual quality assessment is expensive. Most of the works in the literature focus only on the *blur detection* and discard important factors such as *field definition*.

For this stage, a method was developed for analyzing image quality regarding motion blur and field definition [Pires et al. 2012]. Furthermore, alternative methods were also developed for blur detection [Jelinek et al. 2013].

3.1. Field Definition

In this problem, a good retinal image for DR analysis is one image centered on the macula.

The method we discuss herein operates based on the methodology of full-reference comparison. In this methodology, a reference image with assured quality is assumed to be known and quantitative measures of quality for any image are extracted by comparisons with the reference. Given that the macular region has a distinguishable contrast, and we are interested in the content of the center of retinal images, metrics of similarity have shown to be highly suitable for this objective. To characterize the retinal images we use the method known as Structural Similarity (SSIM).

3.2. Blur Detection

Although image quality analysis can have several ramifications before arbitrating on the quality of an image, we focus on two very common problems during image acquisition: blurring and out-of-focus capture. Our method involves a series of different blurring classifiers and classifier fusion to optimize the classification. Basically, we rely upon four descriptors: *vessel area*, *visual dictionaries*, *progressive blurring* and *progressive*

sharpening. We also explore combinations of them. For a description about the methods, see the complete dissertation and the paper [Pires et al. 2012]. Fig. 1 presents the ROC curves achieved under the cross-dataset validation.

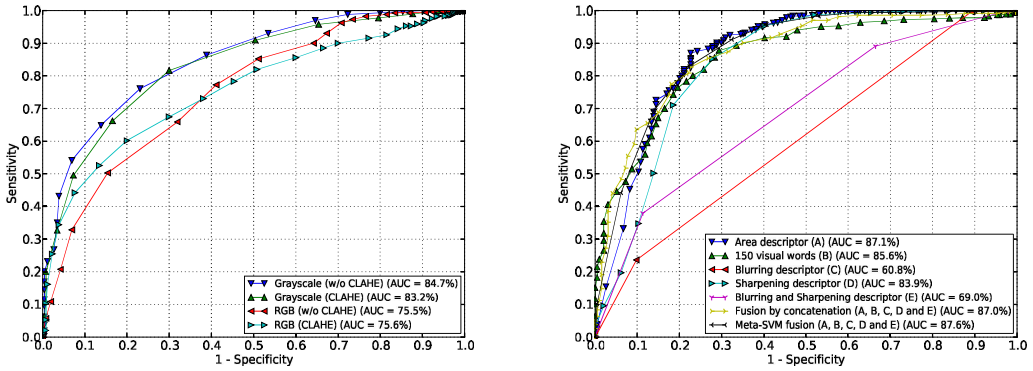

Figure 1. Cross-dataset validation for field definition (left) and blur detection (right) using DR1 as training and DR2 as testing sets.

4. DR-related Lesion Detectors

Due to several lesions related to DR and their diversified characteristics, there are several works which focus on the detection of individual lesions, exploiting particular pre- and post-processing methods for each disease. In this stage, it was developed a series of individual detectors for the most important DR-related lesions: *hard exudates*, *superficial hemorrhages*, *deep hemorrhages*, *cotton wool spots*, and *drusen*. An additional classifier able to detect both superficial and deep hemorrhages was also implemented: *red lesions*.

This section comprises a brief description of the experiments performed for the detection of individual DR-related lesions, as well as presents the experimental results for each anomaly [Pires et al.]. Appreciating the reproducibility, the source code is freely available through GitHub: `https://github.com/piresramon/pires.ramon.msc.git`.

In our work, we employ a different strategy. We use a unified methodology, based on bag-of-visual-words (BoVW) representations, associated to maximum-margin support-vector machine (SVM) classifiers. Such methodology has been widely explored for general-purpose image classification, and consists of the following steps: (i) extraction of low-level local features from the image; (ii) learning of a codebook using a training set of images; (iii) creation of the mid-level (BoVW) representations for the images based on that codebook; (iv) learning of a classification model for one particular lesion, using an annotated training set; (v) using the BoVW representation and the learned classification model to make decisions on whether or not a particular image has a lesion.

The mid-level BoVW features are based upon the low-level features, whose choice has great impact on performance. Two treatments are usual: **sparse features**, based upon the detection of salient regions, or points-of-interest; and **dense features**, sampled over dense grids of different scales. A challenging step, the codebook learning is usually performed by a k-means clustering over features chosen at random from a training set.

With the codebook in hand, the next steps are the BoVW operations: *coding* and *pooling*. For coding, besides the traditional **hard** assignment, we have tested the **soft** assignment, and proposed a new **semi-soft** assignment especially conceived for the DR-related lesion detection application. The semi-soft coding tries to combine the advantages of both hard and soft assignments, i.e., avoiding the boundary effects of the former, and the dense codes of the latter.

For the pooling step, we forgo the traditional **sum**-pooling and employ the more recent **max**-pooling. The pooling step is considered one of the most critical for the performance of BoVW representations, and max-pooling is considered an effective choice.

The detailed results are presented in Table 1, which shows the AUCs obtained for each lesion in the cross-dataset protocol (training with DR1 and test with DR2 dataset).

Table 1. AUCs in %, for Training with DR1, Testing with DR2

	Sparse features			Dense features		
	Hard	Semi-soft	Soft	Hard	Semi-soft	Soft
Hard Exhudates (HE)	93.1	**97.8**	95.5	94.5	95.6	95.6
Red Lesions (RL)	92.3	**93.5**	87.1	89.1	90.6	89.9
Cotton-wool Spots (CS)	82.1	**90.8**	84.9	84.5	90.4	90.3
Drusen (D)	66.5	82.8	62.6	**84.1**	82.5	75.5

5. Detector Fusion

Given a set of detectors of individual DR-related lesions developed with a method which provides satisfactory results for the definition of presence/absence of the most common lesions, this work involves the use of combining approaches aimed at pointing out whether an image is normal or has any lesion including possible ones not present during training.

The classifier fusion was explored for combination of the individual DR-related lesions [Jelinek et al. 2012]. Our main approach consists in investigating fusion of different detectors to identify the presence of DR. The work contains a set of classifiers that act in cooperation to solve a pattern recognition problem, followed by several methods for classifier fusion. This approach is intuitive since it imitates our nature to seek several opinions before making a decision.

For this step, we have a set of detectors for six individual DR-related lesions. The assignment approach explored for the development of the detectors is the semi-soft, explained in Section 4. Two fusion methods were evaluated: OR and meta-classification.

- **OR -** Labels as positive the data classified as positive in at least one classifier.
- **Meta-classification -** Employed in Section 3 for quality assessment, can be loosely defined as learning from information generated by different learners.

We performed the fusion in three different testing steps:

(1) images from DR2 for testing with at least one of the discussed lesions;
(2) images from DR2 for testing with any DR-related lesion (including neovascularization, increased vascular tortuosity, foveal atrophy, chorioretinitis scar, etc.);
(3) images from DR2 for testing which present signals of other anomalies (except the ones we trained for). This step was performed only for the OR fusion technique.

Figure 2 depicts the results achieved using the logic OR fusion and the meta-classification method. The results for meta-classification express the mean and the standard deviation of the AUCs obtained using the 5×2-fold cross-validation protocol.

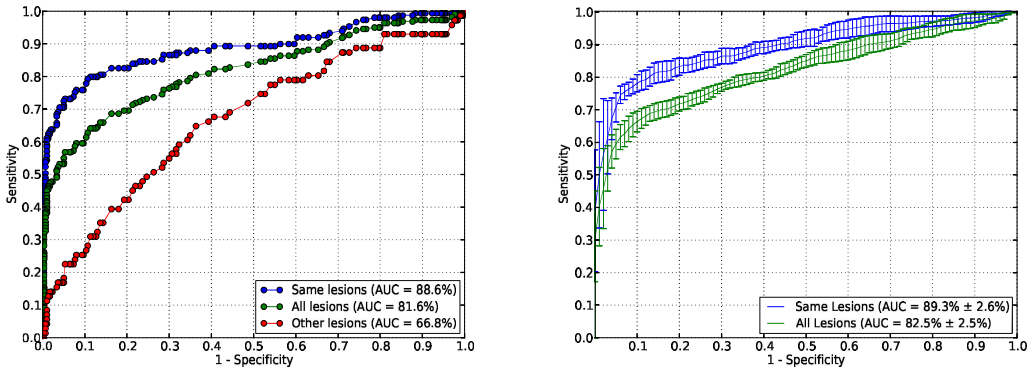

Figure 2. Cross-dataset validation for fusion by logic OR (left) and 5×2-fold cross-validation for fusion by meta-classification (right).

6. Referral

In order to achieve early detection of DR, helping to stop or slow down its progress, international guidelines recommend annual eye screening for all diabetic patients. However, the increasing number of diabetic patients and the decreasing number of ophthalmologists make this suggested annual examination difficult to be performed sufficiently. This factors tend to overwhelm the specialist even more during the next years.

Thus, aiming at referring to a specialist only the patients who really need a consultation, this work includes a stage for classifying retinal images as referable (to be referred to a specialist) or non-referable (not to be referred to a specialist) in the interval of one year [Pires et al. 2013].

Table 2 summarizes all the results obtained for referral. The experiments were performed without normalization, and normalizing with *z-scores* and *term-frequency*.

Table 2. AUCs for referral

Technique	Hard-sum	Soft-max
Without normalization	90.8%±3.1%	**93.4%±2.1%**
Term-frequency	82.5%±4.6%	83.4%±4.6%
Z-score	**91.7%±2.1%**	89.4%±3.0%

7. Conclusion

In this research, we proposed original solutions to deal with diabetic retinopathy related problems. The major results were published in international conferences [Pires et al. 2012, Jelinek et al. 2012, Jelinek et al. 2013] and in a top-tier international journal [Pires et al. 2013]. Furthermore, part of the achievements were combined with new discoveries in the doctorate (under way), resulting in a paper recently accepted for publishing in the top-tier journal PLOS ONE [Pires et al.].

The main breakthrough for the **quality analysis** step was the use of classifier fusion to optimize the classification. This tactic gave us an interesting result: to ensure that a satisfactory percentage of poor quality images will be discovered ($\simeq 90.0\%$), we can establish that only 10.0% of the enough quality images will be unnecessarily retaken. The quality assessment constitutes a key step of a robust DR-related lesion screening system because it helps preventing misdiagnosis and posterior retake.

The **detection of individual DR-related lesions** is one of the most important topics of this work. The development of detectors aim at facilitating the attendance in rural and remote communities. A considerable contribution of this step was the proposal of a new coding scheme called *semi-soft*, that outperforms the state of the art, mainly for hard-to-detect DR-related lesions, such as drusen and cotton-wool spots.

Based upon the scores associated to the detection of the most common DR-related lesions, we developed an accurate **multi-lesion detector** which showed to be effective for the detection of all the considered lesions. Taking as strategy the fusion of individual lesion detectors, the meta-classification method provided us the most satisfactory results.

For assessing the need for **referral**, our proposed method can be used especially in remote and rural areas. The method captures retinal images, evaluates them in real-time, and suggests whether or not the patient requires a review by an ophthalmic specialist within one year. We have achieved important results with this methodology. For example, for a sensitivity of 90.0%, we have a specificity of 85.0%, which means that the specialist time may be saved in 85.0% (only 15.0% of the attended patients will be normal).

In closing this work, we would like to emphasize that there is still important researches to be done in DR image analysis. For instance, identifying the precise location of the lesion, and defining the DR severity degree of a patient further classifying the images as related to DR cases in early, mild, moderate and severe stages. We intend also to explore more sophisticated methodologies for machine learning and image representation, such as *BossaNova*, *Fisher Vectors* and *Deep Leaning*.

References

Jelinek, H. F., Pires, R., Padilha, R., Goldenstein, S., Wainer, J., and Rocha, A. (2012). Data fusion for multi-lesion diabetic retinopathy detection. In *IEEE Intl. Computer-Based Medical Systems*, pages 1–4.

Jelinek, H. F., Pires, R., Padilha, R., Goldenstein, S., Wainer, J., and Rocha, A. (2013). Quality control and multi-lesion detection in automated retinopathy classification using a visual words dictionary. In *Intl. Conference of the IEEE Engineering in Medicine and Biology Society*.

Pires, R., Jelinek, H., Wainer, J., Valle, E., and Anderson, R. Advancing bag-of-visual-words representations for lesion classification in retinal images. *PLoS ONE*, (In Press).

Pires, R., Jelinek, H. F., Wainer, J., Goldenstein, S., Valle, E., and Rocha, A. (2013). Assessing the need for referral in automatic diabetic retinopathy detection. *IEEE Transactions on Biomedical Engineering*, 60(12):3391–3398.

Pires, R., Jelinek, H. F., Wainer, J., and Rocha, A. (2012). Retinal image quality analysis for automatic diabetic retinopathy detection. In *Intl. Conference on Graphics, Patterns and Images*, pages 229–236.