Data fusion for multi-lesion diabetic retinopathy detection

[1]Herbert F. Jelinek, [2]Ramon Pires, [2]Rafael Padilha,
[2]Siome Goldenstein, [2]Jacques Wainer, [3]Terry Bossomaier, [2]Anderson Rocha

[1]*Centre for Research in Complex Systems and School of Community Health,*
Charles Sturt University, Australia
hjelinek@csu.edu.au
[2]*Institute of Computing, University of Campinas (Unicamp), Brazil*
{siome, wainer, rocha}@ic.unicamp.br, {pires.ramon, padilhax7}@gmail.com
[3]*Centre for Research in Complex Systems and School of Computing and Mathematics,*
Charles Sturt University, Australia
tbossomaier@csu.edu.au

Abstract

Screening of Diabetic Retinopathy (DR) with timely treatment prevents blindness. Several researchers have focused their work on the development of computer-aided lesion-specific detectors. Combining detectors is a complex task as frequently the detectors have different properties and constraints and are not designed under a unified framework. We extend our previous work for detecting DR lesions based on points of interest and visual words to include additional detectors for the most common DR lesions and investigate fusion techniques to combine different classifiers for classification of normal or signs of diabetic retinopathy. The combination methods show promising results and shed light on the possible advantages of combining complementary lesion detectors for the DR diagnosis problem.

1. Introduction

Early diagnosis of disease has been important for maintaining optimal health and reducing costs associated with treatment. Diabetic retinopathy (DR) if not treated in time, can lead to the loss of sight [1].

Diabetic retinopathy is characterized by the presence of red (microaneurysms and hemorrhages) and white (hard exudates) lesions as well as neovascularization. Cotton wool spots are also often observed in the retina, although these are not signs of diabetic retinopathy per se [2].

Developing a unified framework that can identify different retinal lesions has been published using a visual words dictionary model [3]. This model creates a large set of visual words, which increases with the number of lesions that are identified. Therefore lesion detectors need to be combined to optimize classification. Within the field of multi-lesion detection associated with DR, detector fusion methods have been applied [4] but require further development for the DR model.

The most common fusion methods can be classified into three levels: (1) abstract (each classifier outputs the class label for each input pattern); (2) rank (each classifier outputs a ranking list of possible classes for each input pattern); and (3) measurement (each classifier outputs a score, probability or confidence level for each input pattern) [5]. Majority voting, 'sum', 'product', 'max', 'min', 'average', 'median', 'OR' and 'AND' methods [6] are commonly employed.

This paper presents a unified visual words framework that is able to identify hard exudates, deep hemorrhages, superficial hemorrhages, drusen, and cotton wool spots.

Our main approach expands upon that of Acharya et al. [4] and our previous work. Points of Interest (PoIs) are combined into visual words and a visual dictionary [7] that identifies specific anomalies within the retina is created. A set of classifiers work in cooperation followed by methods for classifier fusion. Three fusion methods were evaluated: OR, majority voting and meta-SVM.

2. State-of-the-Art

A variety of different classifier fusion techniques can be used to augment the diagnostic performance of individual models in the context of diabetic retinopathy classification.

Niemeijer et al. [8] applied classifier fusion methods to combine several detectors for microaneurysms, using static combination rules and meta-classification. However, the results indicated, in some cases, that combining detectors for the same lesion does not necessarily result in better performance.

3. Methods

Detector fusion is applied to a multi-lesion detector algorithm based on visual words dictionary [7]. Detectors operate in parallel and are combined to obtain a result related to the presence of diabetic retinopathy.

Visual dictionaries constitute an approach in which each image is treated as a collection of regions. The only important information is the appearance of each region [9]. When creating a visual dictionary from a training set, the most representative regions for a given lesion are identified.

When creating a visual dictionary from a training set of examples, we want to learn the generative model that selects the most representative regions for a given problem. The number of selected regions must be large to distinguish relevant changes in the images, but not too large to the point where it starts distinguishing irrelevant variations such as noise.

3.1. Individual detectors

Speeded-Up Robust Features (SURF) algorithm was applied to find a set of points of interest [10] in the image and characterize their surrounding regions. [11].

Visual dictionaries contain a feature space by applying a clustering algorithm that finds the most discriminative features for the lesion of interest. Each region of PoIs becomes a visual 'word' of a 'dictionary'.

For the training set a region of interest within the retinal images is identified by the specialist and contains examples of the specific lesion. The training stage then locates the PoIs within the image under analysis, using SURF. Each image in the training set generates a series of PoIs, which are then filtered, retaining only the PoIs within the regions of interest for further processing to create the dictionary.

To perform the final classification, the Support Vector Machines [12] classifier was used. For training the classifier, the SVM is fed with the signatures of the training images containing positive and negative examples. For each image, the signature is computed mapping each PoI onto the closest (in Euclidean distance) word in the pre-computed dictionary.

To classify a new image the method computes the PoIs using SURF, quantizes the PoIs, generating a histogram of visual words, and feeds the trained classifier with the resulting signature.

3.2. Fusion of detectors

Currently, different pre- and post-processing procedures are required for each detector, making multi-lesion detection difficult and inaccurate [13], [14].

There are numerous methods for combining classifiers. The principal approach is classifier fusion, which considers that all classifiers contribute to the final decision, assuming competitive classifiers [15]. We investigated three methods of classifier fusion: OR (logical dichotomy between the output of different binary classifiers), majority voting (the class with more common outcomes is elected) and meta-SVM (a new binary classification layer fed with the margin distance outcomes of the SVM binary classifiers) and applied two post-normalization techniques: term frequency (on the individual detectors stage which basically divides each entry in the feature vector by the its sum) and *z*-score (on the fusion stage with meta-SVM which zero-center each entry in the feature-vector and divide it by its standard deviation).

4. Experimental results

The experiments were performed on the DR1 dataset, (Department of Ophthalmology, Federal University of São Paulo). The dataset comprises 7,137 images of which 4,293 are normal retinas, 234 images contain hard exudates, 146 contain deep hemorrhages, 102 contain superficial hemorrhages, 139 contain drusen and 73 contain cotton wool spots[1]. All images in this dataset were manually annotated by three medical specialists. The average resolution of the images is *640 x 480* pixels. The experiments have two parts:

- **Part #1.** Multi-lesion detection using individual detectors.
- **Part #2.** Investigation of three different fusion methods using the results from Part #1.

Results of Part #1 experiments are the average of a 5-fold cross-validation protocol shown as ROC curves.

4.1. Experiments – Part #1

Results for detection of hard exudates, deep hemorrhages, superficial hemorrhages, drusen, and

---

[1] Please visit http://www.recod.ic.unicamp.br/site/asdr/

cotton wool spots are presented. We utilized 500 visual words to describe each lesion in the dictionary (250 lesions-based words plus 250 normal-based words) [3] and SVM fed with the signatures (500 values followed by the class information) of the training images (Section 3.1).

Figures 1 and 2 show the results for the ROC curves for the individual lesion and multi-lesions present in the DR1 images. Table 1 summarizes the results for areas under the curve (AUC).

For single lesions (Fig. 1) our method performs best for classifying hard exudates (AUC=91.6%), whereas for multi-lesions (Fig. 2) an AUC of 88.3% was obtained.
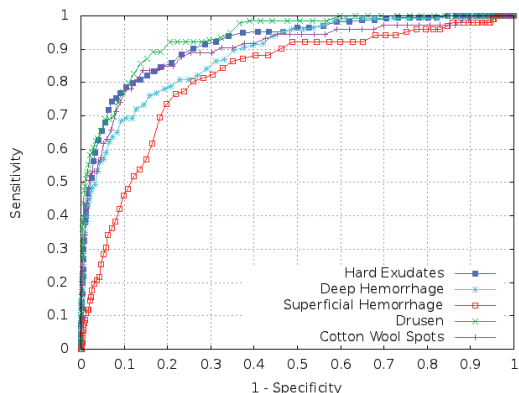
Figure 1. Results using visual dictionaries for detecting single lesions.
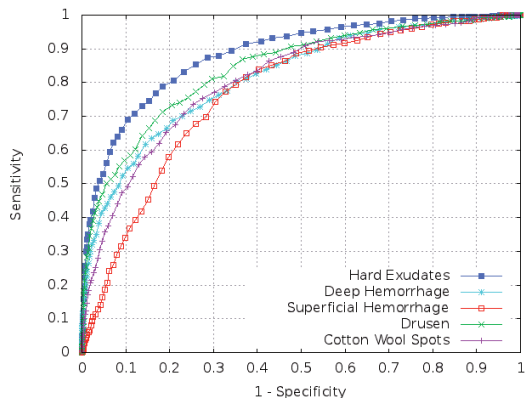
Figure 2. Results for detecting a specific lesion when more than one lesion is present in the set of retinal images.

4.1. Experiments – Part #2

To improve the multi-lesion classification, we investigated the effectiveness of three fusion techniques.

To perform the fusion experiments with the meta-SVM method, we used each raw SVM output and not only the class information (+ or −) by modeling the SVM algorithm to provide distances to the hyperplane as our criterion.

Table 1. Results expressed in AUC.

	Only the lesion in analysis for test	Any lesion for test
Hard Exudates	91.6%	88.3%
Deep Hemorrhage	88.7%	81.6%
Superficial Hemorrhage	81.5%	77.2%
Drusen	93.4%	84.3%
Cotton Wool Spots	89.6%	80.7%

For the meta-SVM fusion method, the signature format comprises the distances to the hyperplane obtained in each individual classifier followed by the class information. The meta-classification was performed with both the original signatures and applying z-norm as a post-recognition normalization technique over the distances (Section 3.2.4) to compare the respective results.

Figure 3 presents the ROC curves for the meta-SVM analysis. AUC using the normalization technique with z-norm was equal to 67%, whereas using the original distances a better accuracy (AUC = 83.7%) was achieved. Meta-SVM without normalization performs better because the difference and the division operations of z-norm method may lose the main characteristic that distinguishes normal and lesion retinas.
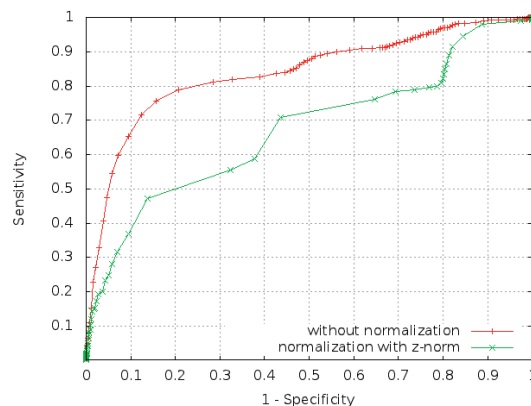
Figure 3. Results of meta-SVM fusion method.

Table 2 shows the experimental results for the fusion with logic OR, majority voting and meta-SVM. The majority voting fusion algorithm provided the highest accuracy but also the lowest sensitivity. A

comparison between OR and majority vote results establish that OR provides better DR classification due to its higher sensitivity.

Table 2. Results of the fusion methods.

	OR	Majority vote	Meta-SVM
Sensitivity	72%	30%	80%
Specificity	86%	99%	78%
Accuracy	84%	92%	78%

5. Conclusions

Single lesion detection combined with data fusion for single lesion performs best. All lesions are identified with accuracy greater than 80%. Only superficial hemorrhages are detected at below 80% accuracy for multi-lesion detection.

The best result was obtained with meta-SVM (AUC=83.7%) suggesting that the use of outputs of individual detectors in a new classification level can be defined as a powerful framework to identify the presence of DR. To improve this result, we are investigating possible fusion solutions such as [16] which maps the distances to a normalized probability space. Furthermore, we are investigating the effects of poor quality retinal images in a DR screening.

6. Acknowledgements

7. References

[1] Q. Mohamed, M. C. Gillies, and T. Y. Wong, "Management of diabetic retinopathy: A systematic review", *Journal of the American Medical Association*, vol. 298, no. 8, pp. 902–916, 2007.

[2] J.S. Kim, A.S. Maheshwary, D.U.G. Bartsch, L. Cheng, M.L. Gomez, K. Hartmann, and W.R. Freeman, "The Microperimetry of Resolved Cotton-Wool Spots in Eyes of Patients With Hypertension and Diabetes Mellitus", *Arch Ophthalmol.*, pp. 879–884, 2011.

[3] H.F. Jelinek, A. Rocha, T. Carvalho, S.K. Goldenstein, and J. Wainer, "Machine Learning and Pattern Classification in Identification of Indigenous Retinal Pathology", in *33rd Annual Intl. Conference of the IEEE EMBS*, 2011.

[4] U.R. Acharya, C.M. Lim, E.Y.K. Ng, C. Chee, and T. Tamura, "Computer-based detection of diabetes retinopathy stages using digital fundus images", *Journal of Engineering in Medicine*, pp. 545–553, 2009.

[5] L. Xu, A. Kryzak, and C.Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition", *IEEE Trans. Systems, Man and Cybernetics*, pp. 418–435, 1992.

[6] M.P. Ponti-Jr, "Combining Classifiers: from the creation of ensembles to the decision fusion", in *Conference on Graphics, Patterns and Images*, pp. 1–10, 2011.

[7] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos", *in IEEE Intl. Conference on Computer Vision*, 2003, pp. 1470–1477.

[8] M. Niemeijer, M. Loog, M.D. Abràmoff, M.A. Viergever, M. Prokop, and B. van Ginneken, "On combining computer-aided detection systems", *IEEE Transactions on Medical Imaging*, pp. 215–223, 2011.

[9] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russel, M. Abràmoff, and P. Soliz, "Multiscale AM-FM methods for diabetic retinopathy lesion detection", *IEEE Trans. on Medical Imaging*, vol. 29, pp. 502–512, 2010.

[10] J. Stoettinger, A. Hanbury, N. Sebe, and T. Gevers, "Do colour interest points improve image retrieval", in *IEEE Intl. Conference on Image Processing*, 2007, pp. 169–172.

[11] H. Bay, T. Tuytelaars, and L.V. Gool, "SURF: Speeded up robust features", in *European Conference on Computer Vision*, pp. 1–14, 2006.

[12] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. Springer, 2006.

[13] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis", *Investigative Ophthalmolology and Visual Science*, vol. 48, pp. 2260–2267, May 1, 2007.

[14] M. D. Abràmoff, M. Niemeijer, M. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. van Ginneken, "Evaluation of a System for Automatic Detection of Diabetic Retinopathy From Color Fundus Photographs in a Large Population of Patients With Diabetes. Response to Olson", *Diabetes Care*, vol. 31, p. e64, 2008.

[15] L. Kuncheva, "Combining pattern classifiers: methods and algorithms", *Wiley-Interscience*, 2004.

[16] W. Scheirer, A. Rocha, R. Michaels, T. Boult. "Robust Fusion: Extreme Value Theory for Recognition Score Normalization", in *European Conference on Computer Vision*, 2010.