# IMPROVING HARD EXUDATE DETECTION IN RETINAL IMAGES THROUGH A COMBINATION OF LOCAL AND CONTEXTUAL INFORMATION

*C. I. Sánchez [1,2], M. Niemeijer [1], M. S. A. Suttorp Schulten [2], M. Abràmoff [3] and B. van Ginneken [1,2]*

[1]Image Sciences Institute, University Medical Centre Utrecht, Utrecht, The Netherlands

[2]Diagnostic Image Analysis Group, Department of Radiology, Radboud University Nijmegen Medical Centre, The Netherlands

[3]Ophthalmology Service, OLVG Hospital, Amsterdam, The Netherlands

[4]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, USA

## ABSTRACT

Contextual information is of paramount importance in medical image understanding to detect and differentiate pathologies, especially when interpreting difficult cases. Current computer-aided detection (CAD) systems typically employ only local information to classify candidates, without taking into account global image information or the relation of a candidate with neighboring structures. In this work, we improve the detection of hard exudates in retinal images incorporating contextual information in the CAD system. The context is described by means of high-level contextual-based features based on the spatial relation with surrounding anatomical landmarks and similar lesions. Results show that a contextual CAD system for hard exudate detection is superior to an approach that uses only local information, with a significant increase of the figure of merit of the Free Receiver Operating Characteristic (FROC) curve from 0.840 to 0.945.

***Index Terms*—** Computer-aided detection, contextual information, hard exudates, retinal images.

## 1. INTRODUCTION

The automatic detection of hard exudates (HEs), one of the most prevalent lesions during early stages of diabetic retinopathy (DR), has received considerable attention in recent years [1, 2, 3]. The automatic identification of HEs is important to obtain an early diagnosis of DR and to prevent blindness. Although promising results have been reported, computer-aided detection (CAD) systems need to be further improved to be useful in screening practice [4]. The identification of subtle HEs and the reduction of spurious detections have been identified as possible enhancements for CAD performance [4]. However, this topic has received little attention so far.

In the daily practice, the medical experts make use of context to interpret medical scenes and differentiate abnormalities, especially when dealing with cases that are difficult diagnose or that contain inconclusive local evidence. For example, knowing the fact that lesions often occur in clusters, a particular slightly suspect candidate has a higher chance of being a true lesion when another more obvious lesion is present nearby. Although contextual knowledge plays an important role in scene interpretation by humans, how to effectively exploit this information in CAD approaches has not been established yet. The vast majority of CAD systems described in the literature employ only local-based features such as color, shape, textures, and so on, for the classification of suspicious candidates, without taking into account information derived from the complete image or from the candidates relation with neighboring structures. Contextual information has been scarcely used in retinal image understanding. In [1], the presence of microaneurysms was used to assist in the detection of hard exudates. In [2], the distance to the closest red lesion, as well as the proximity of a vessel, was employed to improve the classification of retinal images with bright lesions.

In this paper, we study the benefit of including contextual information in a CAD system for the detection of hard exudates in retinal images. We propose a general group of high-level contextual-based features which try to emulate human cognitive processes in retinal image interpretation to classify and differentiate candidates. In contrast to previous works [1, 2], the proposed features take into account not only the presence of anatomical landmarks or other lesions but also the spatial relationship between objects of the same class, exploiting the knowledge that lesions may occur in clusters. We also present a human observer study to investigate to what extend retinal specialists make use of contextual information to infer the label of a potential bright lesion.

## 2. MATERIALS

A total of 144 color retinal images from anonymous patients with diabetes were selected. 69 images contained one or more types of bright lesions and 75 were determined by an expert to have no bright lesion (not containing any hard exudates, cotton wool spots, or drusen). Images were obtained from multiple types of fundus cameras, resolution varied from 768x576 to 2048x1536 pixels while the field of view coverage varied between 35 and 45 degrees. The images were automatically resized to have a field of view with a standardized diameter of 650 pixels. 72 images (38 abnormal) were used to train the system and 72 (37 abnormal) were used as a test set to perform the final algorithm evaluation. A retinal specialist performed annotations outlining hard exudates on all images. These annotations were taken as the reference standard. A second retinal specialist performed manual annotation on the test set in order to assess the inter-observer variability.

## 3. METHOD

### 3.1. HE candidate extraction

The objects that are potential bright lesions (HE, cotton wool spots or drusen) are extracted from the images using a previously described technique [2]. The green channel of the RGB image is convolved with 14 digital filters based on Gaussian derivatives [2]. They are selected from a larger set of Gaussian filterbank outputs up to and including second order derivatives ($G$, $G_x$, $G_y$, $G_{xx}$,$G_{xy}$,$G_{yy}$) at five different scales $\sigma = 1, 2, 4, 8, 16$ using a feature selection algorithm [2]. A $k$-Nearest Neighbor (kNN)classifier is then used to classify the pixels on the basis of the filter responses. After this classification, a lesion probability map is obtained that indicates the probability that each pixel is part of a bright lesion. Thresholding this map, a set of bright lesion candidate clusters is obtained [2]. Because their output is required for further processing, algorithms that perform red lesion classification, optic disc segmentation, and vessel segmentation are applied to the images as well. These algorithms were previously described by our group [4]. Candidates that overlap with the optic disc are automatically removed using the segmented optic disc. The detected $N$ objects per image represent the candidates $\chi = \{x_1, x_2, ..., x_j, ..., x_N\}$ to be bright lesions.

### 3.2. Contextual classification

After the candidate extraction step, the candidates $\chi$ need to be classified as hard exudate or non-exudate. For each candidate, two types of information are extracted: local information describing the candidate, and contextual information describing its spatial relation to other elements in the image. The contextual information is obtained by means of a group of high-level contextual-based features. These features are based on two different relationships: The relation between the candidate and pre-computed neighboring structures, such as blood vessels and red lesions; and the relation between the candidate and neighboring candidates belonging to the same class of the current candidate. In that way, to classify a candidate as hard exudate, we will study the proximity of red lesions or vessels but also if there exist other hard exudates in the neighborhood.

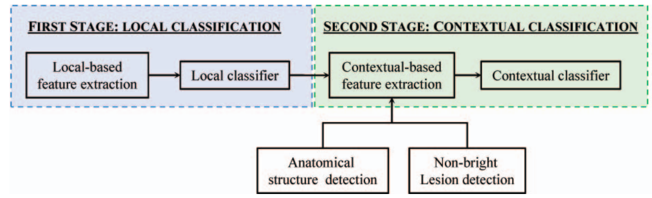



**Fig. 1**. Flowchart of the proposed approach. The problem is divided into two stages: local classification and contextual classification. The contextual features are calculated using the posterior probabilities calculated in the first stage.

**Table 1**. Set of local-based features for local classification

| Nr | Description |
|---|---|
| 1 | Area (size). |
| 2 | Length of the perimeter of the candidate. |
| 3 | Compactness of the candidate. |
| 4, 5 | Length and Width of the candidate. |
| 6-31 | Features measuring the contrast and color of the candidate. |
| 32-80 | Mean and standard deviation of Gaussian derivative filter outputs in the candidate. |

In this framework, a chicken-and-egg dilemma arises: we need the label of the candidates to calculate the contextual features and vice versa. In order to solve this problem, we propose a two-stage approach. In the first stage, only local information extracted from the segmented object is used to obtain preliminary posterior probabilities $p(exud|x_j)$ for each candidate $x_j$ in the image. The second stage is a refinement stage where the contextual information, along with the local one, are employed to improve the initial classification. The contextual features are calculated using the posterior probabilities extracted in the first stage. Figure 1 shows a flowchart of the proposed approach.

- *First stage: Local classification*: This first classification is needed in order to calculate contextual features in the second stage that depend on the relation with any surrounding lesions of the same class. The posterior probabilities of the candidates can be obtained using a group of local-based features, summarized in Table 1 and previously proposed in [2]. They only use the appearance or statistics of the candidate itself to get a classification irrespective of the neighbors' class labels.
- *Second stage: Contextual classification*: A group of contextual-based features are extracted using the posterior probabilities obtained in the previous stage and additional information of pre-computed structures whose proximity may also influence the decision about hard exudate candidates, namely the blood vessels and red lesions. These features measure the minimum distance to HEs, red lesions or vessels; the number of these elements in the neighborhood and the exudate probability distribution around the object. To calculate the dis-

tance to HEs or their quantity in the neighborhood, we threshold the posterior probabilities ($p(exud|x_j) > T$) to obtain a hard classification. The threshold $T$ is chosen to obtain a trade-off between the number of true and false positives in the training set. Given a neighborhood $\Omega(r)$ of radius $r$ around the candidate $x_j$, the 6 context-based features shown in Table 2 are defined for $x_j$.

After pilot experiments on the training data with several classifiers [5], a linear discriminant classifier was chosen for both stages, in favor of $k$ nearest neighbor classifier, quadratic discriminant classifier and support vector machine classifier. Feature selection is carried out for each classifier by Sequential Forward Floating Selection (SFFS) to establish the most discriminative features. The SFFS procedure uses leave-one-out training and testing on the training dataset only, with the area under the Receiver Operating Characteristic (ROC) Curve as the criterion to be optimized. The maximum number of features to be selected is set at 30. Before performing the feature selection, the feature values are transformed using the Box-Cox transformation [6] in order to approximate the feature distribution to a normal distribution.

## 4. RESULTS

The proposed algorithm was assessed on the test set of 72 images and compared to an approach that employed only local information, i.e., removing the second stage. The Free Response Operating Characteristic (FROC) curve was computed for the local approach and for the contextual approach using a radius $r = 64$ pixels to compare the performance of the systems. The radius $r = 64$ allows to obtain enough information about the candidate surroundings. The FROC curve plots the sensitivity of the system against the average number of false positive per image. Figure 2 depicts the FROC curves for the detection of hard exudates using the local and the contextual approaches. For comparison, the performance of an independent second human observer is also included. A Figure of Merit (FOM) of 0.840 with a 95% confidence interval of [0.778, 0.890] [7] was obtained for the local system, whereas a FOM of 0.945 and confidence interval of [0.904, 0.971] was achieved with the contextual system (significant difference, $p < 0.05$).

To study the use of context information by human observers, a group of 310 detected regions were selected from the test set of 72 images. This subset contained 100 hard exudates and 210 non-exudates. The regions were presented using a random order in three reading sessions. In the first session, a neighborhood with a radius of 16 pixels around the region was presented to the observer. The region was shown in the center of the display in order to exclude position information in the image. The neighborhood of 16 pixels radius is similar to the local neighborhood that the computer used to calculate local features. In the second session, a larger neighborhood (64 pixel radius) around the region was presented to the observer. In the last session, the whole image was displayed. In the three reading sessions, the observer was asked to classify the region as hard exudates or non-exudates. Figure 3 shows the ROC curve of the proposed system for hard exudate detection in the subset of 310 regions and the performance of the second observer in the three reading sessions.

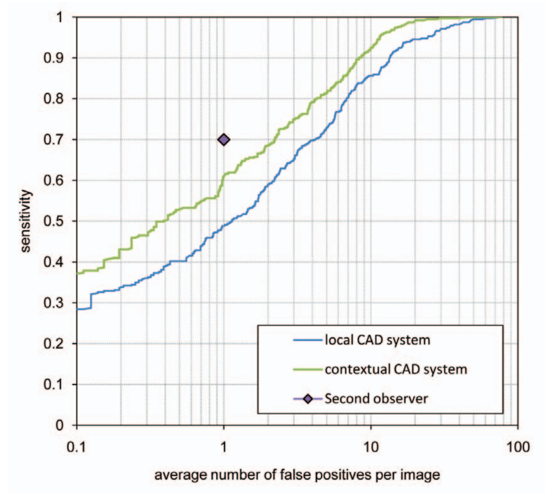



**Fig. 2**. FROC curves for the detection of hard exudates in the 72 test images using local and contextual CAD systems. The diamond represents the result of a second human observer for hard exudate detection in the test set. Note that the horizontal axis has a logarithmic scale.

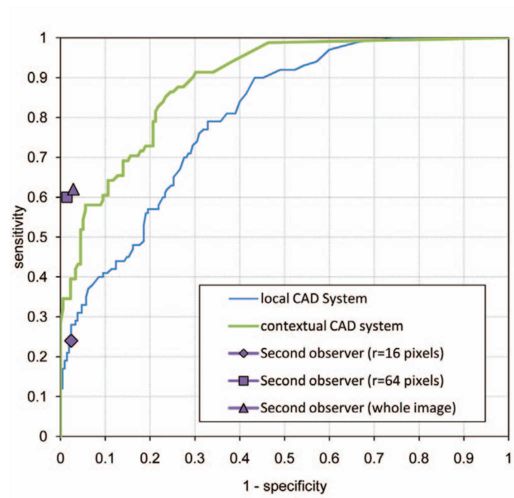


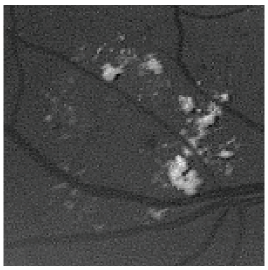

**Fig. 3**. ROC curves using the selected subset of 310 regions for the classification of hard exudates using local and contextual CAD systems. The points represent the human observer's performance in the reading sessions of the observer study.

## 5. DISCUSSION AND CONCLUSIONS

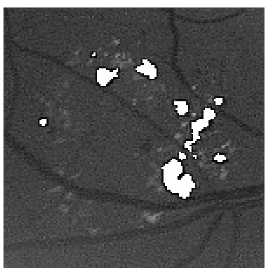
In this paper, context was introduced by means of a group of context-based features, which described spatial informa-

**Table 2**. Set of contextual-based features for contextual classification

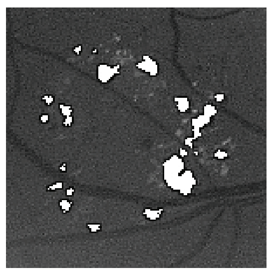
| Nr | Description | Formal expression |
|---|---|---|
| 81 | Total posterior probability of candidates in $\Omega(r)$ | $PP = \sum_{x_i \epsilon \Omega(r)} p(exud \| x_i)$ |
| 82 | Distance to the closest exudate in $\Omega(r)$ | $D(exud; T) = \min_{x_i} \|\|x_i - x_j\|\|^2, \, \forall x_i \epsilon \Omega(r) \, / \, p(exud \| x_i) > T$ |
| 83 | Number of exudates in $\Omega(r)$ | $N(exud; T) = \sum_{x_i \epsilon \Omega(r)} 1, \, \forall x_i \epsilon \Omega(r) \, / \, p(exud \| x_i) > T$ |
| 84 | Distance to the closest red lesion in $\Omega(r)$. | |
| 85 | Number of red lesions in $\Omega(r)$. | |
| 86 | Distance to the closest point on the vascular tree. | |



(a) Patch from an original image



(b) Local CAD system



(c) Contextual CAD system

**Fig. 4**. Example of hard exudate detection with local and contextual CAD systems. The green plane of the image is shown here. Detections are shown in white with a black outline.

tion of neighboring elements such as anatomical elements or other lesions. These features were among the first 10 most significant features, highlighting the importance of context information for inferring the correct label of candidates. The results show that contextual CAD system outperformed significantly a system which employed only local-based features in the identification of hard exudates. Figure 4 depicts an example of hard exudate detection. Only obvious lesions were detected by the local CAD system (Figure 4b), misclassifying subtle hard exudates as non-lesions. Exploiting contextual information, such as the proximity of red lesions or obvious hard exudates, those lesions were correctly detected by the contextual CAD system (Figure 4c). We hypothesize that our approach is generally helpful for the identification of abnormalities that appear in groups, inferring their class labels based on the presence of similar candidates in the neighborhood. Results of the observer study show that the performance of the observers diminished when the context information was limited. The performance of the proposed CAD system is similar to the human observers' behavior, obtaining better results when contextual information was exploited.

Although a significant improvement is obtained, there is still room for improvement. Including contextual features that describe the relation with other bright lesions, such as drusen or cotton wool spots, would be useful to differentiate hard exudates from these lesions and to increase the classification performance. We consider this a topic for future research.

## 6. REFERENCES


[1] A D Fleming, S Philip, K A Goatman, G J Williams, J A Olson, and P F Sharp, "Automated detection of exudates for diabetic retinopathy screening," *Physics in Medicine and Biology*, vol. 52, pp. 7385–7396, 2007.

[2] M. Niemeijer, B. van Ginneken, S. R. Russel, 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 Ophthalmology & Visual Science*, vol. 48, pp. 2260–2267, 2007.

[3] C I Sánchez, M García, A Mayo, M I López, and R Hornero, "Retinal image analysis based on mixture models to detect hard exudates.," *Medical Image Analysis*, vol. 13, pp. 650–658, 2009.

[4] M D Abràmoff, M Niemeijer, M S A 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," *Diabetes Care*, vol. 31, pp. 193–198, 2008.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley and Sons, New York, 2nd edition, 2001.

[6] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society, Series B*, vol. 26, pp. 211–252, 1964.

[7] Dev P Chakraborty, "Analysis of location specific observer performance data: validated extensions of the jackknife free-response (JAFROC) method.," *Academic Radiology*, vol. 13, pp. 1187–1193, 2006.