



# HHS Public Access

Author manuscript

*Med Image Anal.* Author manuscript; available in PMC 2023 December 19.

Published in final edited form as:

*Med Image Anal.* 2012 January ; 16(1): 216–226. doi:10.1016/j.media.2011.07.004.

## Automatic Diabetic Macular Edema Detection in Fundus Images Using Publicly Available Datasets

**Luca Giancardo,**

University of Burgundy and the Oak Ridge National Laboratory

**Fabrice Meriaudeau,**

University of Burgundy.

**Thomas P. Karnowski,**

Oak Ridge National Laboratory.

**Yaqin Li,**

University of Tennessee Health Science Center.

**Seema Garg,**

University of North Carolina - School of Medicine.

**Kenneth W. Tobin Jr.,**

Oak Ridge National Laboratory.

**Edward Chaum**

University of Tennessee Health Science Center.

### Abstract

Diabetic macular edema (DME) is a common vision threatening complication of diabetic retinopathy. In a large scale screening environment DME can be assessed by detecting exudates (a type of bright lesions) in fundus images. In this work, we introduce a new methodology for diagnosis of DME using a novel set of features based on colour, wavelet decomposition and automatic lesion segmentation. These features are employed to train a classifier able to automatically diagnose DME. We present a new publicly available dataset with ground-truth data containing 169 patients from various ethnic groups and levels of DME. This and other two publicly available datasets are employed to evaluate our algorithm. We are able to achieve diagnosis performance comparable to retina experts on the MESSIDOR (an independently labelled dataset with 1200 images) with cross-dataset testing (e.g., the classifier was trained on an independent dataset and tested on MESSIDOR). Our algorithm is robust to segmentation uncertainties, does not need ground truth at lesion level, and is computationally efficient, as it generates a diagnosis on an average of 9.3 seconds per image on an 2.6 GHz platform with an unoptimised Matlab implementation.

## 1. Introduction

Diabetic retinopathy (DR) is the most common cause of preventable blindness in working age populations in the industrialized world. The World Diabetes Foundation estimates that there will be over 438 million people with diabetes worldwide by 2030. Diabetic macular edema (DME) is a complication of DR and is the most common cause of vision loss and blindness (Singer et al., 1992). DME is defined as swelling of the retina in diabetic patients due to leakage of fluid within the central macula from microaneurysms (dilated small blood vessels) that form as the result of chronic damage due to elevated blood sugar levels. The presence of clinically significant DME is an important indication for the initiation of laser treatment. Early diagnosis and treatment are essential to prevent vision loss, but the large and growing number of diabetic patients and costs of current office-based detection methods (by eye care specialists) are a barrier to achieving the recommended screening compliance in at-risk populations. Instead, an automated system which uses computer vision and machine learning to diagnose retina images for the detection of DR and DME is a potential solution to this problem. The image analysis research community has shown the potential of such systems and is actively working on their development as has been shown in recent large studies by Abramoff et al. (2008), Philip et al. (2007) and Fleming et al. (2009).

One particular characteristic of DME is thickening of the retina, which cannot be directly quantified from a single fundus image because of the lack of 3-D information. Instead, ophthalmologists can infer the presence of the fluid that causes the retina thickening from 2-D images, by the presence of accompanying lipid deposits called exudates, which appear as bright structures with well defined edges and variable shapes. In this paper we address the problem of automated detection of DME through three aspects: the *dataset*, *exudate segmentation* and the *DME diagnosis*.

The first aspect we address in this work is the dataset. In the medical imaging field, there are some publicly available annotated datasets of retinal images which have different goals, characteristics, and levels of completeness. We are aware only of six of them: STARE (Hoover and Goldbaum, 2003), DRIVE (Niemeijer et al., 2004), ARIA (ARIA, 2006), DIARETDB1 (Kauppi et al., 2007), MESSIDOR (Messidor, 2010) and ROC (Niemeijer et al., 2010). Their aims include vessel segmentation, DR diagnosis and microaneurysm localisation. Whenever a single common dataset has been employed by different research groups the advantages and disadvantages of different proposed methods were easily measurable and comparable as shown in Niemeijer et al. (2010). Unfortunately the majority of methods for exudate detection found in the literature were tested on independent datasets with various characteristics, and with very different evaluation methods. In addition, the only dataset containing manually segmented exudates is DIARETDB1, but this dataset suffers from a lack of ethnic variability (88 of its 89 images show the pigmentation typical of Caucasians, which represents only a small percentage of the population requiring mass screening throughout the world). This is problematic, especially considering that the pigmentation of the human retina is significantly variable, especially between different ethnic groups as well as from person to person. The MESSIDOR dataset is composed of 1200 images and it contains the diagnosis for diabetic retinopathy, including DME. Although it does not provide a reference standard for the exudate segmentation, it is an

important resource given the number of images available and the independent diagnosis. To contribute to the development of better image analysis methods, we introduce a new public dataset, the Hamilton Eye Institute Macular Edema Dataset (HEI-MED). This dataset contains high quality fundus images of patients from different ethnic backgrounds. In addition, a manually produced ground-truth lesion map and other meta-data are provided.

The second aspect we address is the automated detection of exudates. In the literature, the majority of approaches to diagnose DME are based solely on exudate segmentation. The confidence level of one or more lesions is employed for the diagnosis. The technique by Agurto et al. (2010) is one of the few exceptions where the lesion map is not employed directly, but rather inferred by a set of frequency domain based features that describe the image as a whole. However, the technique described is for the diagnosis of DR not of DME. The other approaches for exudation detection can be roughly divided into four different categories. *Thresholding methods* base the exudate identification on a global or adaptive grey level analysis. A first attempt was presented by Phillips et al. (1993) and recently a more sophisticated method based on image normalisation and distribution analysis was presented by Sanchez et al. (2009). *Region growing methods* segment the images using the spatial contiguity of grey levels; a standard region growing approach is used by Sinthanayothin et al. (2002), which is computationally expensive by being employed to the whole image. Li and Chutatape (2004) limit the computational issues by employing edge detection to decrease the size of regions. *Morphology methods* employ greyscale morphological operators to identify all structures with predictable shapes (such as vessels). These structures are removed from the image so that exudates can be identified (Walter et al., 2002; Sopharak et al., 2008). *Classification methods* build a feature vector for each pixel or pixel cluster, which are then classified by employing a machine learning approach into exudates or not exudates (Gardner et al., 1996; Osareh et al., 2003; Garcia et al., 2009) or additional types of bright lesions, such as drusen and cotton wool spots (Niemeijer et al., 2007; Fleming et al., 2009). In this work we discuss a method for exudate detection first described by Giancardo et al. (2011) which is summarised here for completeness.

Finally, we present the main contribution of this paper: a new method for the diagnosis of DME. We do not attempt to grade DME using a precise scale like the one presented by Gangnon et al. (2008) or the three levels of International Clinical DME Severity Scale, but rather provide an automated “DME/no DME” grading. The diagnosis simulates what a retina expert would normally attempt with a single fundus image, i.e. inferring the presence of DME from the exudates. We approach the detection of DME with a hybrid technique: first an exudate segmentation map is generated with a method that falls into the category of *thresholding methods*, then we compose a feature vector based on the image as whole by employing different colour spaces, wavelet analysis and the exudate segmentation likelihood. Such image based approach to feature vector composition accommodates the fact that the segmented lesions might have some errors. This approach to feature extraction seems to be applicable not only to this problem domain, but also to all classification tasks that are based on an uncertain segmentation map. Our algorithms can be trained uniquely with the diagnosis as ground truth, i.e. no lesions need to be manually segmented; this is possible thanks to the exudate segmentation algorithm. This technique does not use machine learning methods to classify candidate exudates into true positives and false positives, which

require extensive ground truth segmentations by experts which is problematic to attain. As an example, in images with many lesions often the ophthalmologist will omit some which can cause confusion to an automated lesion level classifier.

We would like to emphasize that our algorithm development was conducted on the HEI-MED dataset, but we also used two of the available public datasets, MESSIDOR and DIARETDB1 for comparison testing. We believe our experiments, as presented, show that our methods generalize well to other data sets which is a common problem in automatic retina image processing.

This paper begins with a description of the compiled dataset in Section 2. Then, Section 3 summarises the characteristics of the other datasets used for testing. Section 4 introduces the techniques employed for the automated DME diagnosis. Section 5 presents the final results for the diagnosis on the various datasets and the comparison with two retina experts. Finally, the paper concludes in Section 6 with some discussion of the results and approach.

## 2. The HEI-MED Dataset

### 2.1. Teleophthalmology Network

Since 2005, our group has been designing and developing a semi-automated, HIPAA-compliant, teleophthalmology network for the screening of diabetic retinopathy and related conditions in under served regions of the Mid-South of the United States of America. Currently, five clinics are provided with Zeiss Visucam PRO fundus cameras appropriately modified to automatically send pictures and patients' metadata to our central server for a diagnosis. The current telemedicine network is described by Li et al. (2009). In January 2010, the server stored 1907 images of which 75% were healthy and the remaining 25% had some form of retinal condition which required referral to an ophthalmologist. The images were acquired from 910 patients in 971 sessions. This wealth of images are of particular interest for research purposes because it provides an appropriate testbed for lesion segmentation and diagnosis stratification algorithms required for a fully automated eye screening system. We believe that the creation of a dataset based on this pool of data, obtained under true clinical conditions with a broad-based screening protocol, has several advantages over existing datasets as described below.

#### 2.1.1. Section removed: A Priori Data Distribution

**2.1.2. Diverse Ethnicity**—The clinics employing the telemedicine network are mainly located in the Mid-South of the USA, where the ethnic groups are heterogeneous and where retinal pigmentation covers the spectrum generally found in diverse populations. It is important to emphasize this aspect because the appearance of the retinal fundus varies greatly depending on the pigmentation of the retinal pigment epithelium, which is correlated to the ethnic group and eye colour. Depending on this pigmentation, lesions or other type of structures are more or less apparent which makes the development of lesion segmentation and diagnosis algorithms that work on a broad spectrum of patients more challenging as shown in Fig. 1.

**2.1.3. Image Consistency**—All the images are captured with the Zeiss Visucam PRO fundus camera, at a resolution of 2196×1958 pixels and with a 45° Field of View (FOV)<sup>1</sup>.

**2.1.4. Quality Assessment**—The image capturing process is vetted by an automatic quality assessment algorithm based on the Elliptical Local Vasculature Density features (ELVD) (Giancardo et al., 2010) that generates a numeric quality metric between 0 and 1. Every time an image is captured the algorithm is run and the result is shown to camera operators, giving them the chance of taking a new image if required. This process statistically increased the overall quality of the images submitted to the system as shown by Karnowski et al. (2009).

**2.1.5. Metadata**—In all the images, additional information is supplied. In addition to the ethnicity and quality metric, age of the patient at image capture, patient gender, type of diabetes, the machine segmented vasculature (employing the method of Zana and Klein (2001)) and a manually generated location of the Optic Nerve (ON). None of this information is employed in our technique, but it is an addendum which is likely to be useful in other type of experiments.

## 2.2. Public Dataset

From the pool of images collected on the server, we have randomly extracted a set of 169 images representative of various degree of DME. We have verified that all images are of sufficient quality, no patient is duplicated, and a reasonable mixture of ethnicity and disease stratification is represented. Table 1 shows the distributions of the ethnicity, DME diagnosis, ELVD quality, diabetes type and patients' age in the dataset.

Although our algorithm does not use machine learning methods to classify lesions into true positives and false positives, for the purposes of algorithm development each image of the dataset was manually segmented by one of the co-authors (E. Chaum, a practising Retina specialist). He identified all the exudation areas and other bright lesions such as cotton wool spots, drusen or clearly visible fluid occurring in fundus images. The manual segmentation and grading of each image was performed on a 17 inch tablet device that can be operated with a stylus, in order to increase the labelling precision and throughput. The software used to perform the task was internally developed with the look and feel of ordinary bitmap painting software. It allows segmentation of the different lesions, revision of already processed images by viewing or amending the existing segmentation, and zooming views. The zooming capability can be somewhat problematic, for while it allows the user to define the contour of the lesion very precisely, it might greatly burden the grader when many very small lesions are clustered together. Therefore, it was agreed to segment clusters of small lesions as a single one if the lesions are not discernible at the first zoom level (a complete view of the image at a resolution of 1280×1024 pixels).

We have decided to release the dataset to the research community on the website: <http://vibot.u-bourgogne.fr/luca/heimed.php>.

---

<sup>1</sup>Some clinics in our network use a 30° FOV. However, the images used to create the HEI-MED dataset do not contain any image of this type.

### 3. Other Public Datasets

Besides HEI-MED, the only two publicly available datasets that can be used to evaluate our algorithm are MESSIDOR and DIARETDB1. While the former provides the DR/DME diagnosis for each image only, the latter provides a lesion map generated by four different ophthalmologists but no explicit diagnosis. Hence, for the DIARETDB1 case, the DME diagnosis was generated by analysing the exudation ground truth map. Each image having more than 75% chance of having exudates (according to the four ophthalmologists) is diagnosed with DME (the labels are available in the electronic annexes of this journal). Table 1 shows also some details of these two datasets.

In all the datasets used in this work, all the images were captured with a 45° FOV using different fundus cameras (even throughout the same dataset). In the MESSIDOR dataset, the images were obtained by a variety of research groups based in France. The heterogeneity, the substantial number of images, the public availability and the independence make these datasets ideal for an unbiased validation of our algorithms.

### 4. Methods

The method proposed for the DME diagnosis is based on the classification of single feature vector generated for each image. The feature vector is based on three types of analysis: the *Exudate probability map* (which is computed using the background subtraction technique described in the *Preprocessing*), the *Colour Analysis* and the *Wavelet Analysis*. The rationale and approach for the feature computations is described in Section 4.5, which is followed by a description of the classifier used and the automatic techniques adopted to select the subsets of features that are employed in our experiments.

#### 4.1. Preprocessing

Our approach uses the green channel  $I_g$  of the image and the  $I_i$  channel from the HSI colour space. First, we resize the image to a height of 752px maintaining the original height/width ratio. We then estimate the background with a large median filter, whose size is  $\frac{1}{30}$  the height of the fundus image on  $I_i$ . This approach has been used previously (Niemeijer et al., 2005; Fleming et al., 2006) and has considerable computation performance advantages over other methods (Foracchia et al., 2005). In other median filtering normalisation techniques, the background is subtracted from the original image in order to obtain a normalised version. Instead, in our approach, we enhance the normalisation with morphological reconstruction (Vincent, 1993) as opposed to the more common practise of subtracting the median filtered result, which improves the removal of nerve fibre layer and other structures at the edges of the optic nerve (ON). This is because the shape of the estimated background is more adapted to the original image. The pseudocode of Algorithm 1 illustrates this step.

**Algorithm 1** Background Estimation (NOTE: fixed typo)

---

```

1: function MORPHBgEst( $I_i$ )
2:    $bgEst \leftarrow$  MEDIANFILTER( $I_i$ )
3:   INITIALISE( $bgMask$ ) ▷ set all the pixels to 0
4:   for  $y = 0$  to HEIGHT( $I_i$ )-1 do
5:     for  $x = 0$  to WIDTH( $I_i$ )-1 do
6:       if  $I_i(x, y) < bgEst(x, y)$  then
7:          $bgMask(x, y) \leftarrow bgEst(x, y)$ 
8:       else
9:          $bgMask(x, y) \leftarrow I_i(x, y)$ 
10:      end if
11:    end for
12:  end for
13:   $bgEst_2 \leftarrow$  MORPHRECONSTR( $bgEst, bgMask$ )
14:  return  $bgEst_2$ 
15: end function

```

---

Fig. 2(b) shows an example of the estimated background  $bgEst_2$ . The subsequent step is to normalise the image by subtracting  $bgEst_2$  from the original image with signed precision, obtaining  $I_{noBg}$  as shown in Fig. 2(c). In  $I_{noBg}$ , the highest peak of the histogram is always centred on zero regardless of the ethnicity of the patient or disease stratification (as shown in the second row of Fig. 3(a,b)). The histogram shows a clear distinction between *dark structures* and *bright structures*. The former represent the vasculature, macula, dark lesions and other structures and their distribution varies depending on the ethnicity or pigmentation of the patient. *Bright structures* are found in the positive side of the histogram and include bright lesions, nerve fibre layer reflections and other structures as shown in Fig. 3(c,d). The distribution is fairly constant across different ethnicities.

Because of the characteristics of the normalized image, we can select all the exudate candidates  $I_{cand}$  with a hard threshold  $th_{cand}$ . This has obvious computational advantages in comparison with model fitting operations which are also more sensitive to suboptimal background subtraction (Sanchez et al., 2009). In our case we use  $th_{cand} = 3$  simply by empirically choosing a value slightly above 0 in order to accommodate small background estimation errors. Fig. 2(e) shows an example of the candidates selected. Note that although all the lesions are identified, there are other lesion candidates corresponding to false positives such as nerve fibre layer reflections and other noisy background structures.

The area outside the viewing aperture (the black area around the fundus image) is identified with a fast method (16 milliseconds per image) based on region growing with four seeds placed at the corners of a downsampled version of the image. This approach is described in Giancardo et al. (2010) and it permits to identify the effective area occupied by the image, i.e. the circular area not masked out by the black FOV.

The pre-processing step of ON removal seems to be common to all methods in the literature. This is understandable given the potential colour resemblance between ON and exudates in fundus images. Tobin et al. (2007) described a Bayesian approach for the location of the centre point of ON and macula. The method uses three features of the segmented vascular tree (vessel thickness, orientation and density) and the intensity level of the green plane of the fundus image to create a 4D feature set. Regions of optic nerve and non-optic nerve are characterized by a Gaussian distribution in this feature space. We have trained this algorithm on internal set of 164 images which were not present in any of the public dataset and employed its output as a mean to remove the ON from  $I_{cand}$  by masking out a region slightly

larger than the average ON size. The size of this region was conservatively estimated to be 0.125 times the width and the height of effective retina area.

#### 4.2. Exudate probability map

From the exudate candidate map, an initial exudate detection is performed by assigning a score for each candidate. The exudate candidates are selected by running a 8-neighbour connected component analysis on  $I_{cand}$ .

We seek to take advantage of the outer edge values of exudates in comparison to non-exudate structures with Kirsch's Edges (Kirsch, 1971), which seeks to capture the external edges of the lesion candidates. This edge detector is based on the kernel  $k$  (shown below) evaluated at 8 different directions on  $I_g$ . The kernel outputs are combined together by selecting the maximum value found on each pixel output. The result is stored in the final  $I_{kirsch}$  image, an example is shown in Fig. 2(d).

$$k = \begin{bmatrix} \frac{5}{15} & -\frac{3}{15} & -\frac{3}{15} \\ \frac{5}{15} & 0 & -\frac{3}{15} \\ \frac{5}{15} & -\frac{3}{15} & -\frac{3}{15} \end{bmatrix} \quad (1)$$

The average edge outputs of  $I_{kirsch}$  under each lesion cluster of  $I_{cand}$  are calculated and assigned to the lesion in their entirety. The result is stored in a image  $I_{kCand}$ . In this way we measure the prominence of the candidate's border, which is highly correlated with the probability of having a true exudate (based on our assumption).  $I_{kCand}$  is scaled in order to create an exudate probability map  $I_{exProb}$  that represents  $P(isExudate|I)$ .

$$P(isExudate | I) = \begin{cases} 0, & \text{if } px < th_{low} \\ 1, & \text{if } px > th_{high} \\ \frac{px - th_{low}}{th_{high} - th_{low}}, & \text{otherwise} \end{cases} \quad (2)$$

where  $px$  is a pixel of  $I_{kCand}$ ,  $th_{low}$  and  $th_{high}$  are respectively: the lower boundary below which it is unlikely to identify any significant exudate, and the average upper boundary that can be identified in images with exudates. While  $th_{high}$  does not substantially influence the performance of the algorithm,  $th_{low}$  needs to be carefully selected in order to maintain most of the lesions without an excessive amount of false positives. In a recent publication (Giancardo et al., 2011), we have evaluated this exudate segmentation technique on the HEI-MED dataset and compared it to two other segmentation techniques (Sopharak et al., 2008; Sanchez et al., 2009). Based on these results we have picked a  $th_{low}$  of 4.5 that corresponds to a Sensitivity/Positive Predictive Value of 0.81/0.50 and a  $th_{high}$  of 30 which is where the Free-response Receiver Operating Characteristic (FROC) curve levels out. The Sensitivity was calculated as the average ratio of lesions found in an image, i.e. a Sensitivity



of 0.81 means that, on average, 81% of the lesions were identified on each image. Fig. 2(f) shows an example of the final exudate probability map  $I_{exProb}$

### 4.3. Colour Analysis

Cree et al. (2005) devised a method to significantly reduce the inter-patient colour variability. They assume that a background-less fundus image has colours normally distributed. Hence, the image can be characterized by the scalar mean ( $\mu$ ) and standard deviation ( $\sigma$ ) across the entire image. By taking a reference image and calculating these two parameters, it is possible to equalise the colours of the new image to the reference one in a more effective way than by simple histogram equalisation. The background is estimated by a large median filter, whose size is  $\frac{1}{30}$  the height of the fundus image (however, unlike the exudate candidate detection, no morphological reconstruction step is applied). The description of the process for a single colour plane follows.

$$\begin{aligned}
 I_{ref}^2 &= I_{ref} - \text{medianFilter}(I_{ref}) \\
 \mu_{ref} &= \text{mean}(I_{ref}^2) \\
 \sigma_{ref} &= \text{std}(I_{ref}^2) \\
 I_{new}^2 &= I_{new} - \text{medianFilter}(I_{new}) \\
 \mu_{new} &= \text{mean}(I_{new}^2) \\
 \sigma_{new} &= \text{std}(I_{new}^2) \\
 I_{new}^3 &= (I_{new}^2 - \mu_{new}) \div \sigma_{new} \\
 I_{new}^4 &= (I_{new}^3 \times \sigma_{ref}) + \mu_{ref}
 \end{aligned} \tag{3}$$

where  $I_{ref}$  is the reference image,  $I_{new}$  is the image to be equalised and  $I_{new}^4$  is the equalised image. We have applied this process to the three planes of the RGB colour space using a single good quality reference image for all the datasets. With this equalisation we aim to increase the reliability of colour based features which normally have a significant variability given the patient ethnicity, camera settings and image quality. Fig. 4 shows an example of the equalisation of an image with a given reference image. Even if the two initial images have different ethnic backgrounds and quality levels (Fig. 4(c) is slightly blurred), the resulting images (Fig. 4(b,d)) have very similar colours, particularly with respect to areas of exudation. Finally, we also generate features using the YCbCr and HSI colour spaces, using unequalised images.

### 4.4. Wavelet Analysis

Wavelet analysis is a powerful multi-resolution signal analysis technique that has many applications, such as denoising or compression. In comparison to a traditional Fourier analysis, a wavelet approach has many advantages, the most prominent being the fact that wavelet functions are localized in space and that their scale can vary. In order to analyse an image (or any other signal), a mother wavelet needs to be chosen. From this mother wavelet a series of scaling and wavelet functions are derived which are able to decompose the image at different scales. In 1-D, two approximation signals are generated for each scale level: one containing the details and the other the basis. In 2-D, apart from the basis, three different details are generated for each scale level: vertical, diagonal and horizontal. For more information, Mallat (1999) provides an introduction on the topic of wavelet analysis.

Quellec et al. (2008) proposed a method for the detection of retina microaneurysms in the wavelet space. We try to capture the strong peak at the centre of exudates by developing a method that employs a similar wavelet analysis, but that evaluates the results in image space. A stationary Haar wavelet analysis is performed up to the second level on  $I_I$  (I channel of the HSI colour space of the original image). The process is inverted maintaining the last vertical, diagonal and horizontal details only (see Fig. 5(a–c)), as these are the wavelet coefficients that seem to contain most of the foreground structures. By transforming back to the image space we obtain a background-less image  $I_{wav}$  (see Fig. 5(d)). It is interesting to notice that the distribution of  $I_{wav}$  has similar properties as the image obtained during the image normalisation phase (i.e. centred at 0 and with the exudates located on the positive side of the histogram). Hence, we set to 0 all the reconstructed pixels that correspond to the negative side of the histogram. As it can be seen in Fig. 5(e,f), we are able to enhance the response of exudates particularly at their central areas. This can be better appreciated by comparing Fig. 5(f) with Fig. 2(d) which are magnification of the same image area.

#### 4.5. DME Feature Vector

The selection of the appropriate feature vector to diagnose DME is a challenging problem. We want to characterise an image with a feature vector that has a fixed number of dimensions derived from a variable number of lesions, which might or might not be properly segmented. Other authors avoid this problem by providing a classification at lesion level, where a set of features is generated for each lesion which is then classified as false or true positive. The diagnosis can be subsequently generated by combining the likelihood of the lesions. This approach requires a set of training images labelled at a lesion level, which implies more work to the reader who has to manually draw exudates and other lesions/pigmentation changes for each image. This might be troublesome for a classifier susceptible to outliers or when the samples in the dataset are not enough to average out the human error, which is inevitable since it is not a straightforward to identify all lesions (when there are many) and precisely define their edges. Instead, we are able to describe the lesions segmented as a whole by analysing the exudate probability map, the colour and the wavelet properties of the detected lesion set previously described. Two approaches have been examined, in the first one, the exudate probability map is converted to two binary masks which are overlaid on the colour and wavelet analysis outputs; in the second one, the exudate probability map is used to weigh the analysis outputs at a pixel level.

Fig. 6 shows an example of two binary masks, one is generated by considering all the areas that have at least some chance of being exudative ( $P(isExudate|I) > 0$ ), the other is generated by considering the areas where  $P(isExudate|I) = 0$ . The masks are applied to the following image planes:

- *Wavelet*: The  $I_{wav}$  plane (see Sec. 4.4).
- *CreeRGB*: The three RGB channels after the retinal colour normalisation described in Sec. 4.3.
- *YCbCr*: The three channels of the YCbCr colour space.
- *HSI*: The saturation channel of the HSI colour space.

For each set of pixels extracted, the following statistics are calculated: mean, median, standard deviation, maximum and minimum. With these five measures we attempt to capture the correlation between the colour/wavelet analysis and the exudate probability map. This allows us to automatically identify problematic images which might have a high number of false positives because of consistent fibre layer reflection artefacts (which have a diffuse whitish colour and have different density in the wavelet analysis) or other unwanted false detections. By employing a mask where  $P(isExudate|I) = 0$ , we analyse also the areas not explicitly identified by the exudate segmentation. This appears to have a positive effect on the global classification, particularly on the images with a substandard exudate segmentation due to a noisy image or overexposed image.

The statistical measures described so far do not make full use of exudate probability map. To address this important aspect of the detection, we also use an approach where the same binary masks, image planes and statistical measures are employed, but each pixel in the image planes previously described is weighted on  $P(isExudate|I)$  before the statistics computation.

By combining the weighted and unweighted statistical measures, we obtain a total number of 80 features per image. The total number of features is pretty substantial with the risk of suffering of the “curse of dimensionality” during the classification phase. Hence, they underwent an automatic feature selection process. There are many techniques available to perform feature selection. In this work we utilize *Information Gain*, an approach that seem to be relatively independent of the classifier used and that show reproducible results.

Information theory (Yang and Pedersen, 1997) provides us a straightforward way to apply conditional entropy to evaluate the significance of a feature as follows: information gain =  $H(Class) - H(Class|Attribute)$ , where  $H(X)$  is the entropy and  $H(Y|X)$  is the conditional entropy.

$$\begin{aligned}
 H(X) &= - \sum_{i=0}^n p(v_i) \log_2[p(v_i)] \\
 IG(Y, X) &= \sum_{i=0}^n p(X = v_i) H(Y | X = v_i)
 \end{aligned}
 \tag{4}$$

where  $IG(Y, X)$  is the information gain,  $p(x)$  is the probability of  $x$  (i.e. the frequency),  $n$  is the number of samples and  $v_j$  is the value of a sample. We have employed this technique to reduce the full feature set (weighted and unweighted) to a number of 48 as shown in Table 2. This number was chosen by selecting only the features that had an average  $IG > 0$ .

In this section, the test has been run by splitting the HEI-MED dataset into three folds. The feature selection was performed on each of these three subsets. Each feature receives a final score that is the average between the three folds. This allows to select features that do not “overfit” the dataset, so that any test on the MESSIDOR or the DIARETDB1 dataset would not have any chance of being tainted by the feature selection process.

#### 4.6. DME Diagnosis Classification (This whole section has been completely modified)

In order to select the most appropriate classification strategy, we have run a series of tests with different combinations of classifiers/feature sets. The feature sets used are: all the unweighted features, all the weighted features, the combination of the two sets and the features selected with Information Gain. As far as classifiers are concerned, we decide to cover the three different classification families described by Jain et al. (2000): *probabilistic*, *geometric* and *tree-based*. For the *probabilistic* family we tested the Naive Bayes classifier with two ways of estimating the prior probabilities, by assuming a Gaussian distribution of the data and by employing the Parzen Window approach. For the *geometric* family, two Support Vector Machines (SVMs) were tested, one with a linear kernel and one with the radial basis function as implemented in libSVM (Chang and Lin, 2001). For the *tree-based* family, the Random Forest algorithm with 10 trees was chosen (Breiman, 2001). In order to have a baseline, we reported the result of a nearest neighbour classifier which is likely to be the simplest classification method available (Duda et al., 2001).

The classifiers comparison tests are based on a Receiver Operating Characteristic (ROC) analysis with a Hold-One-Out (HOO) approach uniquely on the HEI-MED dataset. In HOO, the classifier is trained once for each image, each time holding out a different image which is used for testing. We want to avoid any type of optimization on the other two databases (Messidor and DIARETDB1) which will be used in the following section for an unbiased ROC analysis. Table 3 contains all the results for the classifiers comparison. The best Area Under the ROC Curve (AUC) for each feature set is highlighted in bold.

4.6.1. Section "Feature selection" removed, part of the text moved to Section 4.5

### 5. Results (This whole section has been heavily modified)

We have evaluated our automatic diagnosis system using three datasets. The results are presented as ROC curves where a positive image is an image showing signs of DME, and a negative image does not present any sign of this disease (but it might have other conditions). Based on the results of Table 3 we picked the SVM with a linear kernel since it showed the best performance for three out of the four feature set evaluated for the HEI-MED dataset. The ROC curves are calculated by varying the threshold on the output positive diagnosis probability generated by the SVM classifier.

Fig. 7 shows the curves for the classification performed employing the two feature sets that had the best AUC in Table 3: the full unweighted set and the features selected with Information Gain (InfoGain). In each of the plots, three different tests are shown. Each test is represented by two ROC curves, one for each feature sets. The first test is a HOO, which should generally show the best performance since it is trained with images coming from the same dataset as the tested image. The other two are cross validation tests: the classifier is trained on each of the other dataset and tested on the current one. This type of test simulates very well the predicted performance of the system on a real environment because it is trained on datasets with a different number of images, captured by a different camera/operator and labelled by a different expert than the current test set.

In Fig. 7(a) the performance on the HEI-MED dataset are shown. While the HOO tests reach considerable AUCs for both feature sets (0.94 and 0.93), the cross validation tests follow not so closely behind (although they are still acceptable). This suggests that HEI-MED dataset, albeit small, has a representation of fundus appearance and conditions that might not be so obvious in the other datasets. This hypothesis is supported by the fact that in Fig. 7(b,c), the performance difference between HOO and cross validation tests is less significant (at least for the InfoGain features).

In Fig. 7(b) the performance on the Messidor dataset are shown. The InfoGain features perform consistently regardless of the type of test (AUC between 0.88 and 0.89). This is particularly encouraging in the validation of our technique because the HEI-MED and DIARETDB1 datasets have a significantly lower number of images than Messidor (169 and 89 vs. 1200). Interestingly, the weighted features have not performed as well: the AUC goes from 0.9 to 0.54. The reason is probably due to a subset of features that do not discriminate the same aspect of the disease in the different datasets. This stress the importance of feature selection and cross validation tests.

In Fig. 7(c) the performance on the DIARETDB1 dataset are shown. The performance of HOO and cross validation tests with the HEI-MED dataset are very satisfying, particularly with the InfoGain features (the AUC is 0.93 in both tests). The cross validation with the Messidor training is not as good. This is a constant throughout all the tests and it is probably due to some inconsistencies in the Messidor labelling which was confirmed by the two ophthalmologists co-authors of this paper. Fortunately, the Messidor dataset has a considerable amount of images, so we feel that these errors are averaged out.

The computational performance are evaluated on a Dual Core 2.6 GHz machine with 4 GB of RAM with an unoptimised Matlab implementation. The average time to generate the exudate probability map is ~1.9 seconds. ~4.9 seconds needs to be added for the localization of the optic nerve. The average time to compute the other analysis types and classify the image is on average ~2.5 seconds. Therefore, the total time to generate a diagnosis from a raw image (without the optic nerve location) is ~9.3 seconds. This time can be reduced further by optimizing the optic nerve localization algorithm, which does also calculate the macula location with some computationally expensive operations (Tobin et al., 2007).

### 5.1. Comparison with Experts (This whole section has been heavily modified)

Two retina specialists determined the presence or absence of exudates in order to diagnose DME on a random sample of 350 images of the MESSIDOR dataset, with 120 images exhibiting ME and 230 exhibiting no ME. We compared the performance of the automatic system by creating the ROC Curve and overlaid the Specificity/Sensitivity of the two retina specialists as shown in Fig. 8. The reference standard is provided by the MESSIDOR dataset. Two ROC curves are shown and represent tests run on the 350 images only. One is a HOO test on these image, the other a cross validation test trained on the HEI-MED dataset. It can be seen that even if the experts outperform the automatic system in an absolute sense, the system manages to obtain a comparable sensitivity with a slightly worse specificity.

To further quantify these tests, we used Kappa value/AC1-statistics as statistical concordance metrics. The Kappa value is a metric widely used in the literature, however Gwet (2002) underlines some of the pitfalls, and proposes the AC1-statistics to overcome them. In Table 4 and 5, we have compared the diagnosis between the experts, the MESSIDOR GT and the algorithm (trained on the HEI-MED dataset and tested on the Messidor subset). For the algorithm we have picked 0.22 as threshold, which simultaneously maximized both statistics. The diagnosis of the two experts can be downloaded from the electronic annexes of this journal.

## 6. Discussion and Conclusion

In this paper, we have presented a new automatic system for the detection of DME using non-stereo fundus images. The method is based on an algorithm able to detect exudates with some attached confidence level without the use of machine learning methods to separate false positives from true positives, on a colour space analysis and on new methods to characterise the lesions by the means of wavelet analysis. To our knowledge, our approach for the creation of the feature vector with a inner and outer lesion maps has never been attempted before and proved to be quite successful. We are confident that this type of approach can be applied to all the problem domains where a diagnosis (or other types of classifications) needs to be performed on the basis of an uncertain lesions (or other defects) segmentation.

We have tested our system against three different datasets, by HOO and cross-validation. By employing the automatically selected feature we have achieved an AUC between 0.88 and 0.93. We compared this results with the performance of two retina specialist obtaining comparable results with different test modalities (ROC, K-value and AC1-statistic). We have shown that our newly introduced multi-ethnic database is well suited to train algorithms that can be employed on real world images. Additionally, the good computational performance make this method suitable as a component for a complete retinal disease screening tool.

A system to diagnose diabetic retinopathy that was tested with the largest dataset (approximately 15 000 patients) is presented by Niemeijer et al. (2009). They report an AUC of 0.88. In another recent study, Agurto et al. (2010) diagnose diabetic retinopathy avoiding the usage of manually segmented lesion too, by employing AM-FM features. They have obtained an AUC of 0.84 by considering a subset of 400 images of the MESSIDOR dataset. These works prove the feasibility of a fully automated screening system for a large population. Although our technique is not directly comparable because it diagnoses DME and not diabetic retinopathy, the AUCs presented are comparable or above these two systems.

The effectiveness of this method in conjunction with other aspects of lesion detection and retina processing will be pursued in future work in order to attempt to create a competitive diabetic retinopathy screening system able to transparently diagnose the disease state.

## Acknowledgment

These studies were supported in part by grants from Oak Ridge National Laboratory, the National Eye Institute, (EY017065), by an unrestricted UTHSC Departmental grant from Research to Prevent Blindness (RPB), New York, NY, Fight for Sight, New York, NY, by The Plough Foundation, Memphis, TN and by the Regional Burgundy Council, France. Dr. Chaum is an RPB Senior Scientist.

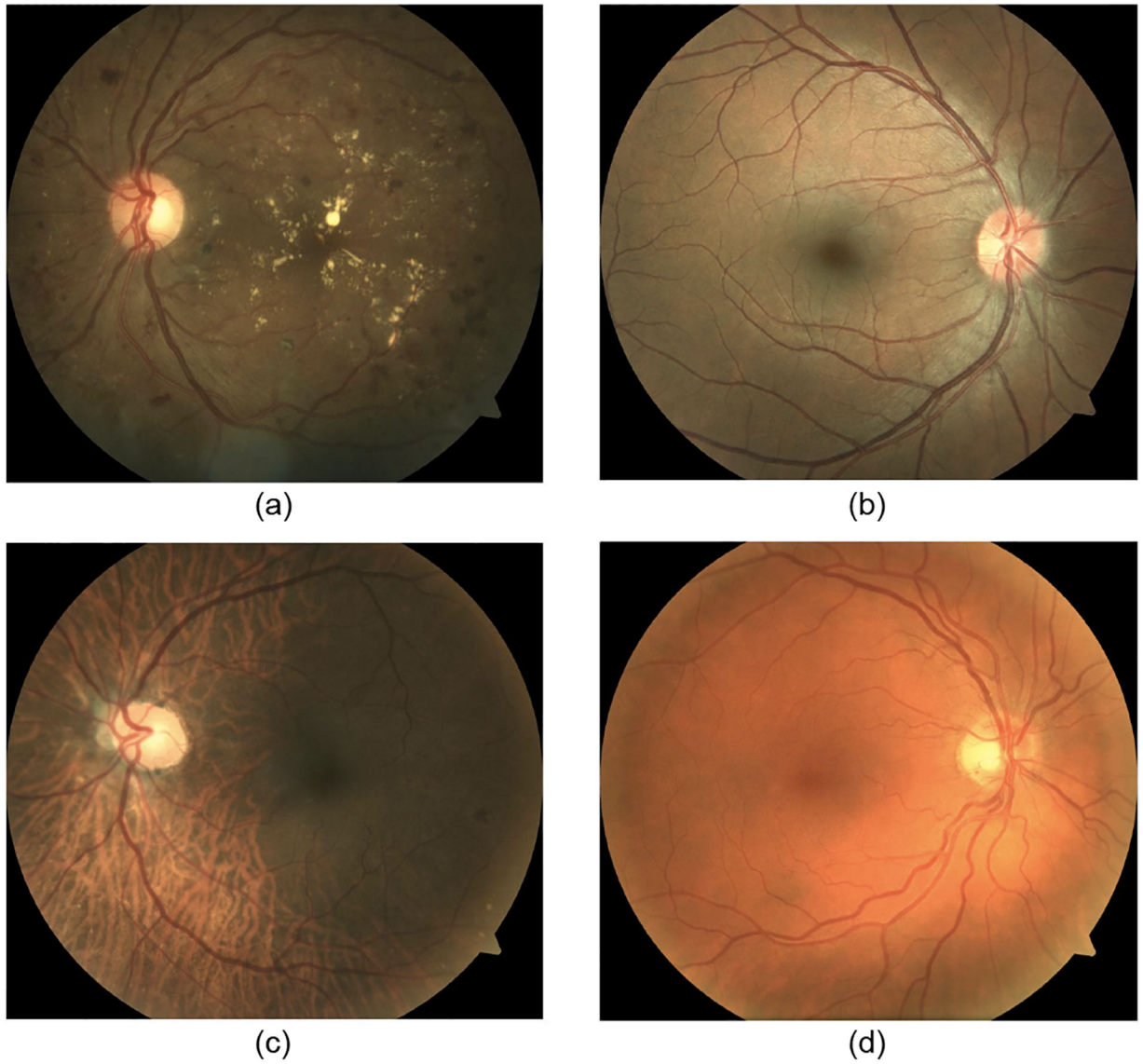
## References

- Abramoff MD, Niemeijer M, Suttorp-Schulten MSA, Viergever MA, Russell SR, van Ginneken B, 2008. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care* 31, 193–198. [PubMed: 18024852]
- Agurto C, Murray V, Barriga E, Murillo S, Pattichis M, Davis H, Russell S, Abramoff M, Soliz P, 2010. Multiscale am-fm methods for diabetic retinopathy lesion detection. *IEEE Transaction on Medical Imaging* 29, 502–512.
- ARIA, 2006. Retinal image archive. Online [<http://www.eyecharity.com>].
- Breiman L, 2001. Random forests. *Machine learning* 45, 5–32.
- Chang CC, Lin CJ, 2001. LIBSVM: a library for support vector machines.
- Cree MJ, Gamble E, Cornforth D, 2005. Colour normalisation to reduce interpatient and intra-patient variability in microaneurysm detection in colour retinal images, in: *WDIC2005 ARPS Workshop on Digital Image Computing*.
- Duda RO, Hart PE, Stork DG, 2001. *Pattern Classification*. Wiley-Interscience.
- Fleming AD, Goatman KA, Philip S, 2009. The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy. *British Journal of Ophthalmology*.
- Fleming AD, Philip S, Goatman KA, Olson JA, Sharp PF, 2006. Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE Transactions on Medical Imaging* 25, 1223–1232. [PubMed: 16967807]
- Foracchia M, Grisan E, Ruggeri A, 2005. Luminosity and contrast normalization in retinal images. *Medical Image Analysis* 9, 179–190. [PubMed: 15854840]
- Gangnon RE, Davis MD, Hubbard LD, Aiello LM, Chew EY, Ferris FL, Fisher MR, Group, E.T.D.R.S.R., 2008. A severity scale for diabetic macular edema developed from etdrs data. *Invest Ophthalmol Vis Sci* 49, 5041–5047. [PubMed: 18539929]
- Garcia M, Sanchez CI, Lopez MI, Abasolo D, Hornero R, 2009. Neural network based detection of hard exudates in retinal images. *Comput Methods Programs Biomed* 93, 9–19. [PubMed: 18778869]
- Gardner GG, Keating D, Williamson TH, Elliott AT, 1996. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *Br J Ophthalmol* 80, 940–944. [PubMed: 8976718]
- Giancardo L, Meriaudeau F, Karnowski TP, Chaum E, Tobin K, 2010. Quality Assessment of Retinal Fundus Images using ELVD. *IN-TECH. chapter New Developments in Biomedical Engineering*. pp. 201–224.
- Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Tobin K Jr, Chaum E, 2011. Automatic retina exudates segmentation without a manually labelled training set, in: *Proceedings of International Symposium on Biomedical Imaging*.
- Gwet K, 2002. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. *Statistical Methods For Inter-Rater Reliability Assessment Series 2*, 1–9.
- Hoover A, Goldbaum M, 2003. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transactions on Medical Imaging* 22, 951–958. [PubMed: 12906249]
- Jain AK, Duin RPW, Mao J, 2000. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 4–37.

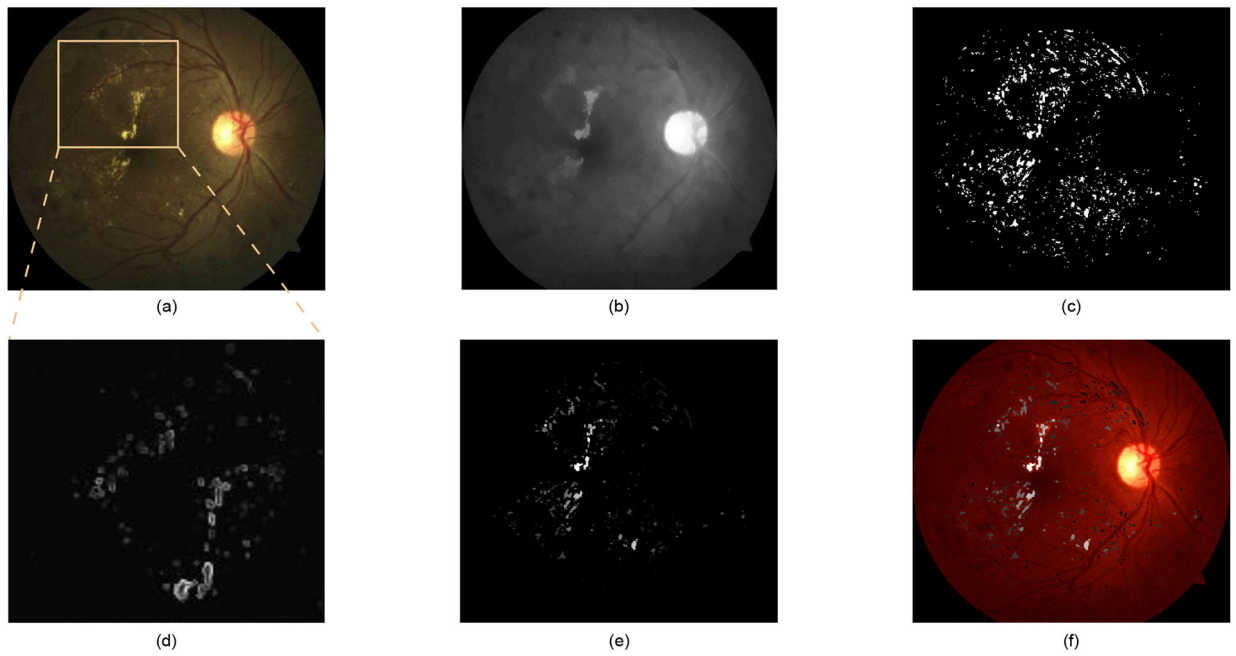
- Karnowski TP, Aykac D, Chaum E, Giancardo L, Li Y, Tobin KW, Abramoff MD, 2009. Practical considerations for optic nerve location in telemedicine, in: Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2009, pp. 6205–6209.
- Kauppi T, Kalesnykiene V, Kamarainen JK, L L, Sorri I, Uusitalo H, Pietila J, Kalviainen H, Uusitalo H, 2007. the diaretdb1 diabetic retinopathy database and evaluation protocol, in: Proceedings of British Machine Vision Conference.
- Kirsch RA, 1971. Computer determination of the constituent structure of biological images. *Computers and Biomedical Research* 4, 315–328. [PubMed: 5562571]
- Li H, Chutatape O, 2004. Automated feature extraction in color retinal images by a model based approach. *IEEE Transactions on Biomedical Engineering* 51, 246–254. [PubMed: 14765697]
- Li Y, Karnowski TP, Tobin KW, Giancardo L, Chaum E, 2009. A network infrastructure for automated diagnosis of diabetic retinopathy. Proc. American Telemedicine Association Annual Meeting. Online [<http://vibot.u-bourgogne.fr/luca/papers/Li2009.pdf>].
- Mallat S, 1999. *A wavelet tour of signal processing*. Academic press.
- Messidor, 2010. Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. Online [<http://messidor.crihan.fr>].
- Niemeijer M, Abramoff MD, van Ginneken B, 2009. Information fusion for diabetic retinopathy cad in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 28, 775–785. [PubMed: 19150786]
- Niemeijer M, van Ginneken B, Cree MJ, Mizutani A, Quelled G, Sanchez CI, Zhang B, Hornero R, Lamard M, Muramatsu C, Wu X, Cazuguel G, You J, Mayo A, Li Q, Hatanaka Y, Cochener B, Roux C, Karray F, Garcia M, Fujita H, Abramoff MD, 2010. Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 29, 185–195. [PubMed: 19822469]
- Niemeijer M, van Ginneken B, Staal J, Suttorp-Schulten MSA, Abramoff MD, 2005. Automatic detection of red lesions in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 24, 584–592. [PubMed: 15889546]
- Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD, 2004. Comparative study of retinal vessel segmentation methods on a new publicly available database. *Proceeding of SPIE Medical Imaging* 5370, 648–657.
- Niemeijer M, Van Ginneken B, Russell SR, Suttorp-Schulten MSA, Abramoff MD, 2007. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest Ophthalmol Vis Sci* 48, 2260–2267. [PubMed: 17460289]
- Osareh A, Mirmehdi M, Thomas B, Markham R, 2003. Automated identification of diabetic retinal exudates in digital colour images. *Br J Ophthalmol* 87, 1220–1223. [PubMed: 14507751]
- Philip S, Fleming AD, Goatman KA, Fonseca S, McNamee P, Scotland GS, Prescott GJ, Sharp PF, Olson JA, 2007. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol* 91, 1512–1517. [PubMed: 17504851]
- Phillips R, Forrester J, Sharp P, 1993. Automated detection and quantification of retinal exudates. *Graefes Arch Clin Exp Ophthalmol* 231, 90–94. [PubMed: 8444365]
- Quelled G, Lamard M, Josselin PM, Cazuguel G, Cochener B, Roux C, 2008. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Trans on Medical Imaging* 27, 1230–1241. [PubMed: 18779064]
- Sanchez CI, Garcia M, Mayo A, Lopez MI, Hornero R, 2009. Retinal image analysis based on mixture models to detect hard exudates. *Medical Image Analysis* 13, 650–658. [PubMed: 19539518]
- Singer DE, Nathan DM, Fogel HA, Schachat AP, 1992. Screening for diabetic retinopathy. *Ann Intern Med* 116, 660–671. [PubMed: 1546868]
- Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, Usher D, 2002. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Medicine* 19, 105–112. [PubMed: 11874425]
- Sopaharak A, Uyyanonvara B, Barman S, Williamson TH, 2008. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics* 32, 720–727. [PubMed: 18930631]



- Tobin KW, Chaum E, Govindasamy VP, Karnowski TP, 2007. Detection of anatomic structures in human retinal imagery. *IEEE Transactions on Medical Imaging* 26, 1729–1739. [PubMed: 18092741]
- Vincent L, 1993. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Journal of Image Processing* 2, 176–201.
- Walter T, Klein JK, Massin P, Erginay A, 2002. A contribution of image processing to the diagnosis of diabetic retinopathy—detection of exudates in color fundus images of the human retina. *IEEE Transactions on Medical Imaging* 21, 1236–1243. [PubMed: 12585705]
- Yang Y, Pedersen J, 1997. A comparative study on feature selection in text categorization, in: *Machine Learning international workshop 10th conference*, Citeseer. pp. 412–420.

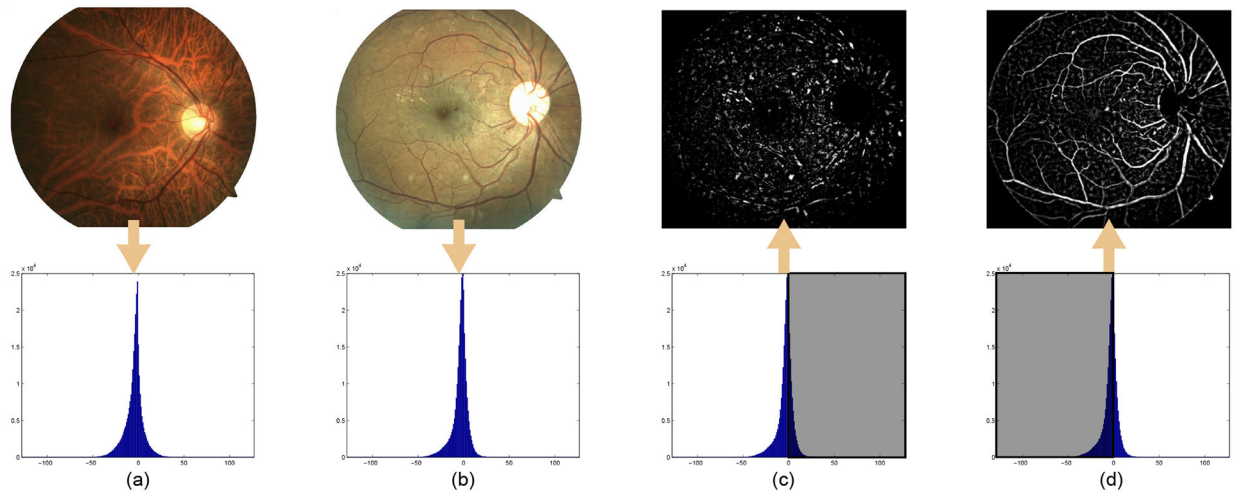


**Figure 1:**  
Examples of fundus images in our dataset. (a) African American patient showing clear exudates; (b) Hispanic patient without signs of DME; (c) African American patient that shows choroidal vessels under the pigment epithelium layer and some small exudates; (d) Caucasian patient without signs of DME.

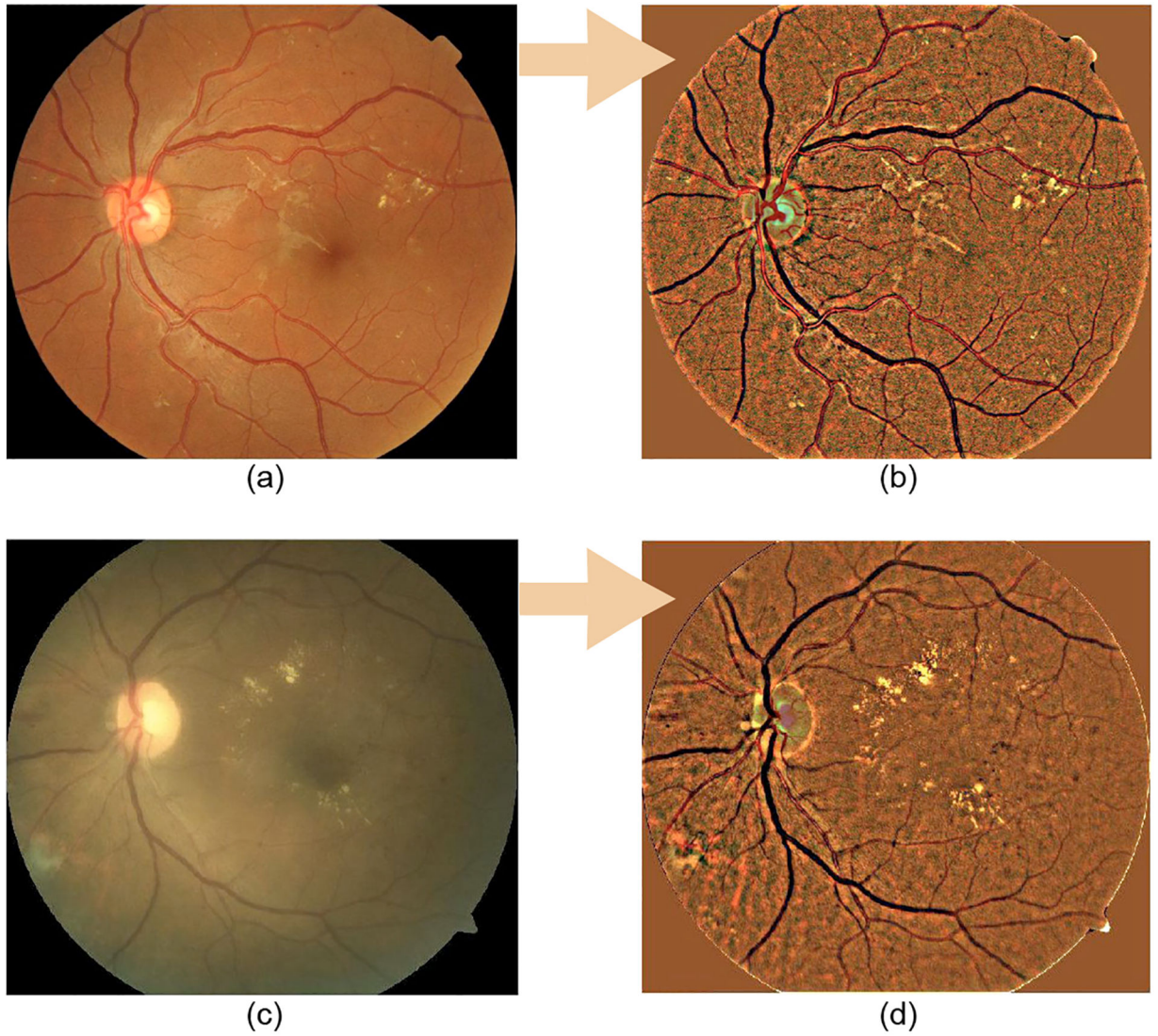


**Figure 2:**

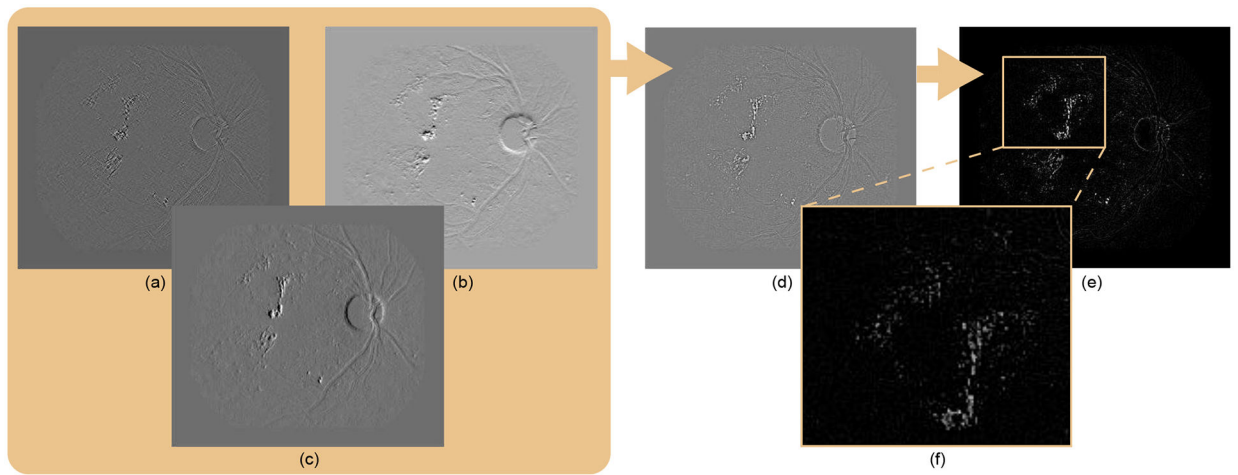
(a) Example of a fundus image of the dataset used. The square represents the area shown in (d); (b) Estimated background,  $bgEst_2$ ; (c) Image without background,  $I_{noBg}$ ; (d) Image detail of Kirsch's Edges Image Analysis; (e) Initial exudates candidates,  $I_{cand}$ ; (f) Exudates probability ( $I_{kirsch}$ ) overlaid on an enhanced version of the original image (to improve the contrast with the probabilities). White corresponds to a probability of 1.0 of being a true exudates, black to a probability of 0.



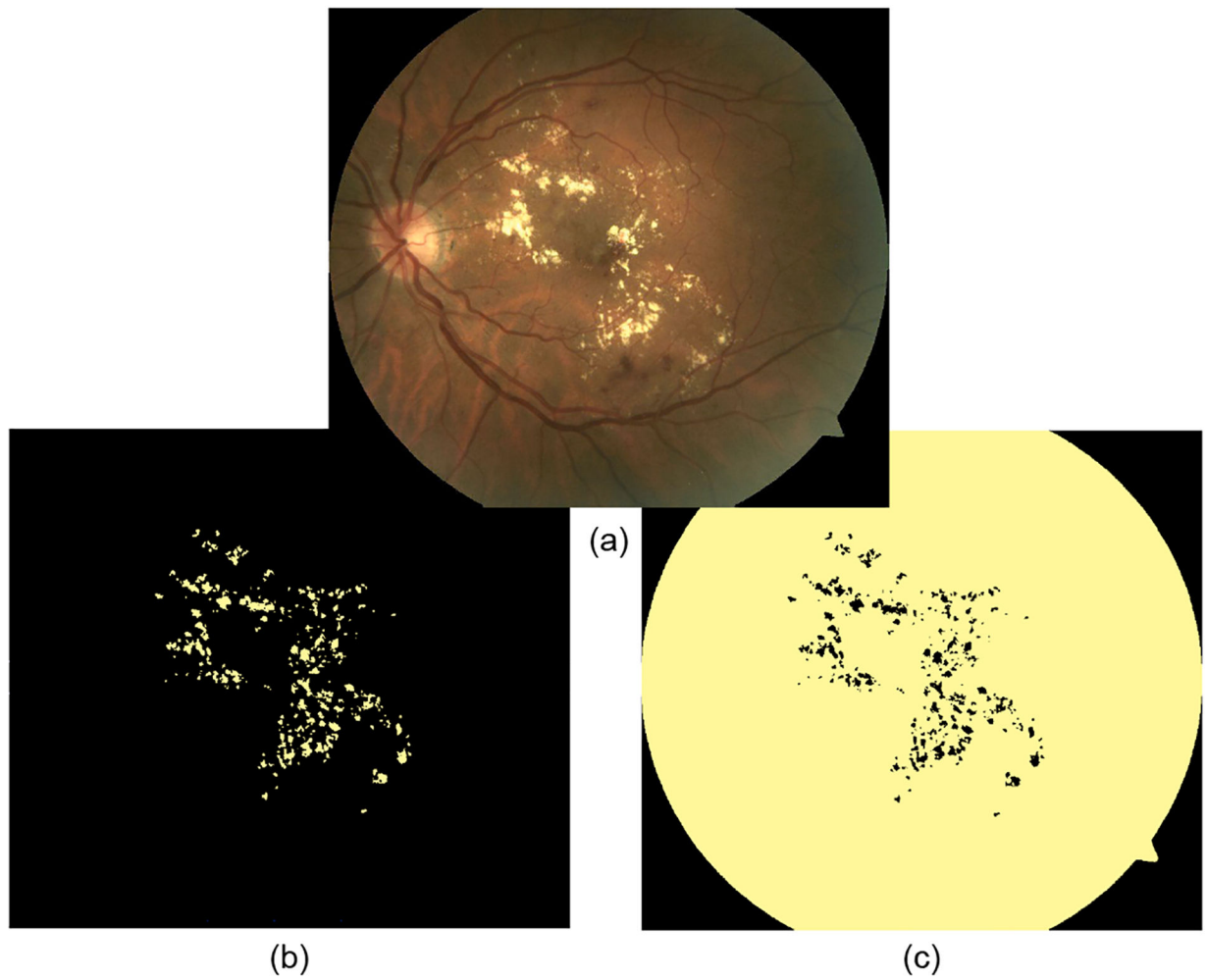
**Figure 3:**  
 (a,b) The first row contains original images showing different types of pigments and lesions;  
 the second row shows the image histograms after the normalisation process; (c) The bright  
 structures located on the positive side of the histogram; (d) The dark structures located on  
 the negative side of the histogram.



**Figure 4:**  
 (a) Reference image  $I_{ref}^1$ ; (b) Background-less reference image  $I_{ref}^2$ ; (c) Image to be equalised  $I_{new}^3$ ; (d) Equalised image  $I_{new}^4$ .

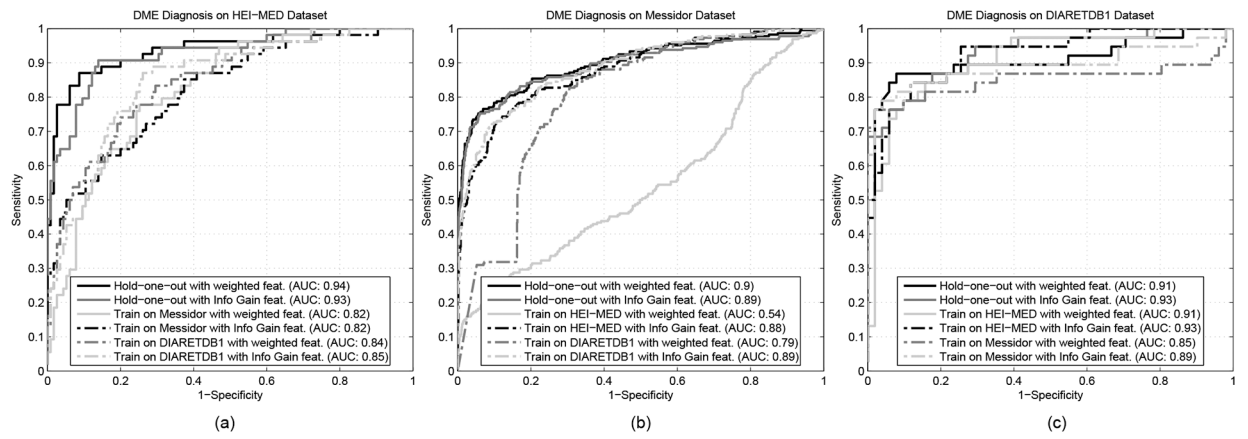


**Figure 5:** Haar stationary wavelet analysis of the image shown in Fig. 2(a). (a,b,c) Respectively diagonal, vertical and horizontal coefficients of the 2<sup>o</sup> level decomposition; (d) Image reconstructed using uniquely the coefficients of (a,b,c); (e) Exudate enhanced image by histogram thresholding; (f) Details of (e) highlighting an area showing exudation.



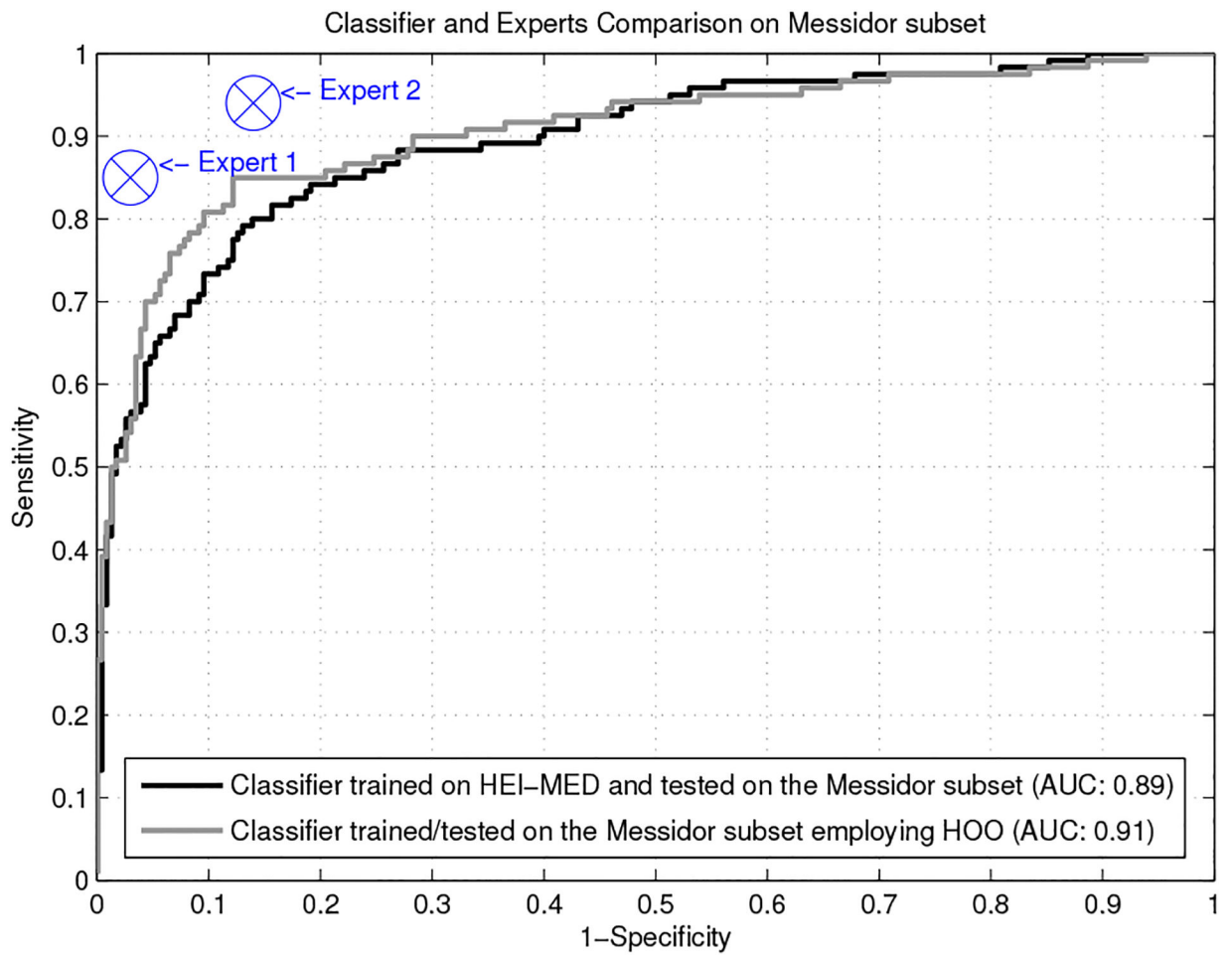
**Figure 6:**

It shows the inner and outer binary masks generated from the exudate probability map. In the binary masks the black colour corresponds to a pixel which is ignored, the yellow to a pixel which should be taken into consideration. (a) Original image; (b) Inner binary lesion map; (c) Outer binary lesion map.



**Figure 7:** ROC curves for the DME diagnosis. Cross-datasets and HOO testing are employed on the three public datasets with a SVM (linear) classifier.





**Figure 8:** ROC curves for the DME diagnosis applied on the image subset diagnosed by the two experts. The feature set selected by Information Gain with a SVM (linear) classifier is used in both cases.

**Table 1:**

Characteristics of the three datasets used in this work

	HEI-MED	Messidor	DIARETDB1
<b>DME Diagnosis</b>			
Negative	115 (68%)	974 (81%)	51 (57%)
Positive	54 (32%)	226 (19%)	38 (43%)
<b>ELVD Quality Metric</b>			
Poor (ELVD < 0.5)	14 (8%)	62 (5%)	3 (3%)
Good (0.5 ≤ ELVD < 0.8)	31 (18%)	356 (30%)	27 (30%)
Excellent (ELVD ≥ 0.8)	124 (74%)	782 (65%)	59 (67%)
<b>Ethnicity</b>			
African American	104 (62%)		
Caucasian	42 (25%)		88 (98%)
Hispanic	19 (11%)		
Unknown	4 (2%)		1 (2%)
<b>Diabetes Type</b>			
Type I	160 (95%)		
Type II	6 (3.5%)		
Unknown	3 (1.5%)		
<b>Patients' Age</b>			
age < 26	5 (3%)		
26 ≤ age < 43	20 (12%)		
43 ≤ age < 61	105 (62%)		
age ≥ 61	39 (23%)		

**Table 2:**

Feature set selected with Information Gain.

---

wavMed, wavAvg, wavStd, wavMax, cbStd, cbMax, cbMin, crMed, crAvg, crMin, crOutMed, rOutCreeAvg, gCreeStd, gOutCreeAvg, bCreeMax, bOutCreeAvg, wavMedW, wavAvgW, wavStdW, wavMaxW, wavOutAvgW, sMedW, sAvgW, sStdW, sMaxW, yStdW, cbMedW, cbAvgW, cbStdW, cbMaxW, crMedW, crAvgW, crStdW, crMaxW, rCreeMedW, rCreeAvgW, rCreeStdW, rCreeMaxW, rOutCreeAvgW, gCreeMedW, gCreeAvgW, gCreeStdW, gCreeMaxW, gOutCreeAvgW, bCreeAvgW, bCreeStdW, bCreeMaxW, bOutCreeAvgW

---

the feature names should be interpreted as follows:  
<analysis plane><In/Out mask><statistics used>[W for Weighted]

---

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

Cross validation with HEI-MED as testing target.

Classifier \ Feature Set	Unweighted	Weighted	Full	Information Gain
Nearest Neighbour	0.692	0.827	0.787	0.86
Naive Bayes (Gaussian)	0.78	0.837	0.864	0.897
Naive Bayes (Parzen Win.)	0.813	0.857	0.876	0.89
SVM (linear kernel)	0.914	<b>0.94</b>	<b>0.919</b>	<b>0.93</b>
SVM (radial kernel)	<b>0.923</b>	0.878	0.904	0.908
Random Forests	0.865	0.888	0.907	0.903

The results are expressed as AUC for the HOO tests on HEI-MED.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 4:**

Kappa-value Comparison.

	<b>MESSIDOR</b>	<b>Expert 1</b>	<b>Expert 2</b>	<b>Algorithm</b>
MESSIDOR	x	0.84	0.76	0.71
Expert 1	x	x	0.76	0.72
Expert 2	x	x	x	0.7
Algorithm	x	x	x	x

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5:**

AC1-statistics Comparison

	<b>MESSIDOR</b>	<b>Expert 1</b>	<b>Expert 2</b>	<b>Algorithm</b>
MESSIDOR	x	0.88	0.78	0.76
Expert 1	x	x	0.79	0.77
Expert 2	x	x	x	0.73
Algorithm	x	x	x	x

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript