Active Learning Approach for Detection of Hard Exudates, Cotton Wool Spots and Drusen in Retinal Images

Clara I. Sánchez[a], Meindert Niemeijer[b,c], Thessa Kockelkorn[a],Michael D. Abràmoff[b,c,d], Bram van Ginneken[a],

[a]Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands;
[b]Dept. of Electrical and Computer Engineering, and
[c]Dept. of Ophthalmology and Visual Sciences, The University of Iowa, Iowa City, IA, USA
[d]Veterans Affairs Medical Center, Iowa City, IA, USA

ABSTRACT

Computer-aided Diagnosis (CAD) systems for the automatic identification of abnormalities in retinal images are gaining importance in diabetic retinopathy screening programs. A huge amount of retinal images are collected during these programs and they provide a starting point for the design of machine learning algorithms. However, manual annotations of retinal images are scarce and expensive to obtain. This paper proposes a dynamic CAD system based on active learning for the automatic identification of hard exudates, cotton wool spots and drusen in retinal images. An uncertainty sampling method is applied to select samples that need to be labeled by an expert from an unlabeled set of 4000 retinal images. It reduces the number of training samples needed to obtain an optimum accuracy by dynamically selecting the most informative samples. Results show that the proposed method increases the classification accuracy compared to alternative techniques, achieving an area under the ROC curve of 0.87, 0.82 and 0.78 for the detection of hard exudates, cotton wool spots and drusen, respectively.

Keywords: active learning, retinal images, uncertainty sampling

1. INTRODUCTION

Diabetic Retinopathy is one of the most important causes of blindness in developed countries. Early detection and diagnosis is crucial for the prevention of visual loss and blindness in patients with diabetes.[1,2] In the onset of the disease, several visual retinal abnormalities appear in the retinal fundus, representing a visual indicator of changes in the eye. The automatic identification of ocular signs provides an effective way to obtain an early diagnosis of diabetic retinopathy and to prevent future complications.

In addition to red lesions, early diabetic retinopathy lesions may be also classified into "bright lesions", such as hard exudates and cotton wool spots.[3] The identification of these bright abnormalities involves differentiating them from drusen, lesions with similar appearance associated with age-related macular degeneration.[3] The design of a machine learning approach for the detection and differentiation of these abnormalities requires an initial training set with labeled samples for these different lesion types. However, acquiring manual annotations of these abnormalities in retinal images is a tedious and time consuming task, especially for large amounts of training data that are needed in order to ensure enough informative samples.

In active learning approaches, only the most informative examples are dynamically selected from the unlabeled data and their correct labels are requested from an expert.[4] Therefore, the number of labeled examples required to obtain an optimum classification accuracy is reduced, as well as the user's involvement in the labeling process. Compared to standard classification, where the goal is to minimizing classification error, active learning has an additional goal: minimizing the amount of samples to be labeled.

An active learning framework has been used in different applications, such as detection of mine-like objects in sonar imagery,[5] labeling video data[6] or content-based information retrieval.[7] In this paper, we propose a CAD system for the detection and differentiation of hard exudates, cotton wool spots and drusen in retinal images incorporating an active learner in the training process.

Send correspondence to Clara I. Sánchez: E-mail: clarisa@isi.uu.nl

Medical Imaging 2009: Computer-Aided Diagnosis, edited by Nico Karssemeijer, Maryellen L. Giger
Proc. of SPIE Vol. 7260, 72601I · © 2009 SPIE · CCC code: 1605-7422/09/$18 · doi: 10.1117/12.813679

2. MATERIALS

A total of 4000 unlabeled retinal images were randomly selected from EyeCheck project,[8] a online retinal screening program in the Netherlands. The image data was acquired at ten different sites. The image resolution varied from 768x576 to 2048x1536 pixels while the field of view coverage varied between 35 and 45 degrees. Three different camera types were employed; the Topcon NW 100, the Topcon NW 200 (Topcon, Tokyo, Japan) and the Canon CR5-45NM (Canon, Tokyo, Japan). All images were JPEG compressed.

From this data set, three images containing hard exudates, cotton wool spots and drusen were selected as the initial training set. A human observer, first author, performed manual annotations of hard exudates, cotton wool spots and drusen in these images. This initial set contained 12 hard exudates, 5 cotton wool spots, 30 drusen and 92 candidates not corresponding to a bright lesion (see the next section for a description of the candidate selection algorithm).

Ten unseen images were selected as the test set to evaluate the final algorithm performance. These images contained 114 hard exudates, 21 cotton wool spots and 48 drusen. The remaining images are used in the active learning process. The images contained 15.84 candidates on average. Around 2.5% of these images are marked as suspect. How many of these contain bright lesions is not annotated in the screening process.

3. METHODS

The algorithm is divided in two different steps: candidate extraction and region classification using active learning. A description of each step is given below.

3.1 Candidate extraction

In the candidate extraction step, the objects that are potential bright lesions (hard exudates, cotton wool spots or drusen) are extracted from the images using a previously described technique.[3] The green channel of the RGB image is convolved with 14 digital filters based on Gaussian derivatives. These filters are invariant to rotation and translation of the image and were selected from a larger set of second order irreducible invariants using a feature selection algorithm.[3]

As the combination of the filter responses in pixels inside lesions are different from the filter responses for non-lesions pixels, a classifier can be trained to extract bright lesions in the images. A k-Nearest Neighbor (kNN) classifier, with k=41, is used to classify the pixels on the basis of the filter responses.[3] After this classification, a lesion probability map is obtained that indicates the probability that each pixel is part of a bright lesion. Fig. 1 shows examples of the lesion probability maps. Thresholding this map at 60%, a set of bright lesion candidate clusters is obtained. Candidates that overlap with the optic disc are automatically removed using the optic disc detection algorithm described in Ref. 9.

3.2 Region classification using active learning

The next step is to classify each candidate as hard exudate, cotton wool spots or drusen incorporating an active learning system in the training process. This system is an iterative procedure where at each iteration the active learner is called to select an unlabeled sample from a pool of unlabeled data and an expert is asked for its label. The idea is to select efficiently a set of training samples from the unlabeled data in an active way to boost the performance of the classifier and reduce the number of samples that need to be labeled.

Assume that an initial training set X_t, a classifier c, an active or query function F and unlabeled data X_u are given. The query function F assigns a value to each unlabeled sample in X_u depending on how informative the sample is. These values permit ranking the unlabeled objects and selecting the most informative sample x^*. This sample is the one that, after revealing its label and adding to the training set, improves the classification performance the most.[4]

The general framework of the active learning system can be described as follows:[4]

1. Train classifier c on the current training set X_t.
2. Select an object x^* from the unlabeled data X_u according to the active query function F.

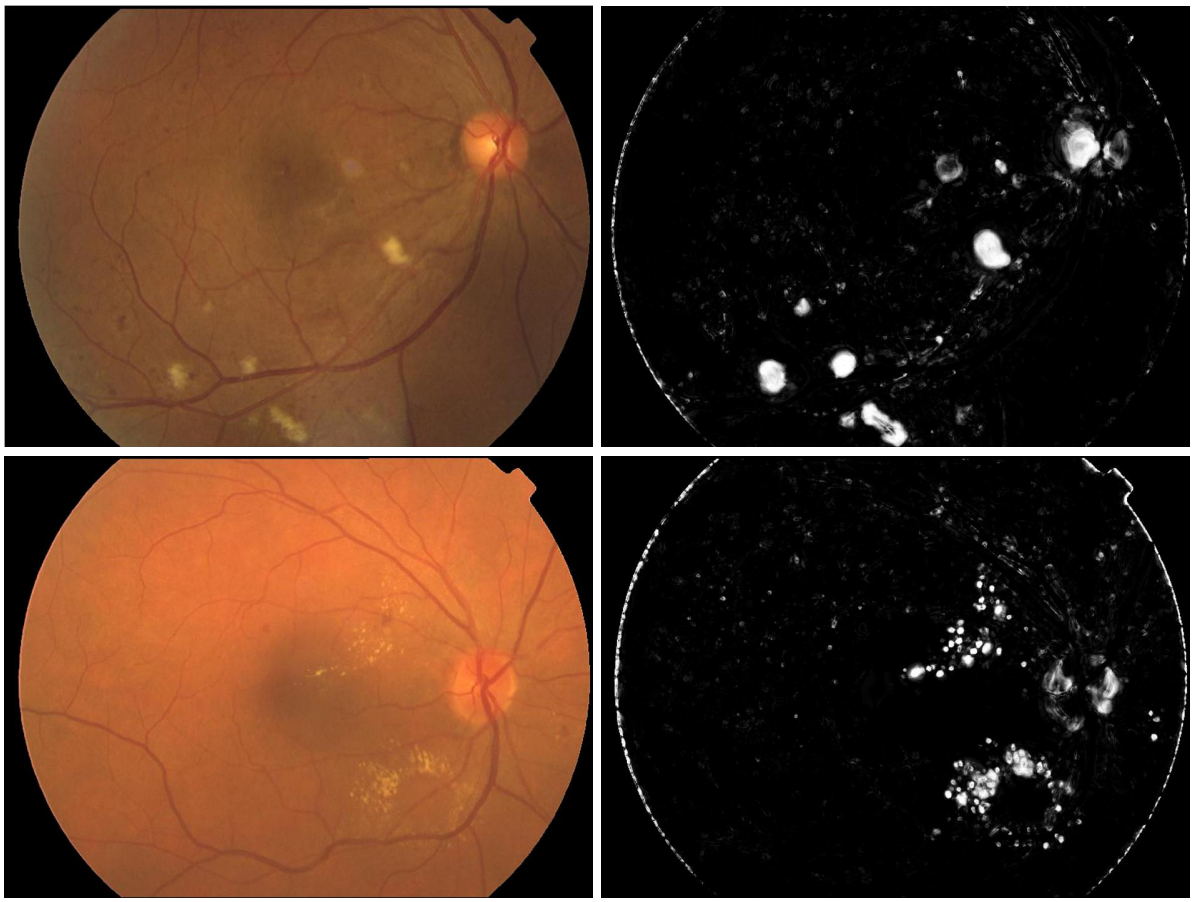


Figure 1. Original retinal images (a,c) and their lesion probability maps (b,d).

3. Ask an expert for the label of x^*. Enlarge the training set X_t and reduce X_u: $X_t = X_t \cup \{x^*\}$, $X_u = X_u \backslash \{x^*\}$.

4. Repeat steps (1)-(3) until a stopping criterion is reached.

The active function F determines the sampling selection, i.e., decides which sample in the unlabeled data X_u to query next. In this work, uncertainty sampling is used as the query function.[4] Uncertainty sampling is an heuristic query function and is one of the simplest active learning algorithms.[10] This method queries unlabeled objects for which the current classifier is most uncertainty about. The intuitive justification for this approach is to increase the performance of the classifier clarifying the position of the decision boundary. Therefore, this selection method picks samples lying near the vicinity of the decision boundary, which are the ones on which the classifier is least confident, and asks for their correct labels. This type of learning is commonly known as exploitation learning[4]

To measure uncertainty, the query function F can be defined as follows:

$$F \equiv x^* = argmin_{x_i \in X_u}[max(P(w^j|x_i), \bigvee j)]. \quad (1)$$

This approach picks the sample among the pool of unlabeled data with the smallest margin to the decision boundary.

In the retinal framework proposed in this work, the pool of unlabeled data consists of the group of candidates found in the unlabeled retinal images. From this pool, a set of training samples has to be actively selected using active learning. Each candidate is described using a set of 83 features[3] (see Table 1). These features represent visual concepts that the ophthalmologists use to differentiate among the retinal lesions. The features provides information about the size, shape, color and contrast of the candidates, as well as contextual information such as the proximity to the vessels and the proximity to the closest red lesion. To calculate these values, algorithms that perform red lesion classification[11] and vessel segmentation[12] were previously applied to the retinal images. No feature selection algorithms were applied in the process.

Table 1. Set of features extracted for each candidate object.

Features 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 the filter outputs.
81, 82	Measurements indicating the proximity of a vessel.
83	Distance to the closest red lesion.

For the classifier c, we choose a four-class linear discriminant classifier. This classifier sets linear decision boundaries between the different classes in order to separate hard exudates, cotton wool spots, drusen and non-lesions. As the dimensionality of the problem is high, the selection of a linear classifier permits obtaining adequate classification performance as well as reducing time complexity.

The final algorithm for the training process in the classification of candidates as hard exudates, cotton wool spots, drusen or non-lesions can be defined as:

1. Train the 4-class linear classifier c on the current training set X_t. This training set consists of samples of hard exudates, cotton wool spots, drusen and non-lesions characterized by its 83-feature vector.
2. Select a candidate x^* from the pool of candidates X_u in the unlabeled retinal images according to the active query function F.
3. Ask an expert if x^* is a hard exudate, a cotton wool spot, a drusen or a non-lesion. Enlarge the training set X_t and reduce X_u: $X_t = X_t \cup \{x^*\}$, $X_u = X_u \backslash \{x^*\}$.
4. Repeat steps (1)-(3) until a stopping criterion is reached.

After applying the active learning algorithm,a classifier dynamically trained for the identification of hard exudates, cotton wool spots and drusen is obtained.

4. EXPERIMENTS AND RESULTS

4.1 Experiments

The algorithm started with the initial training set $X_t^{(0)}$, which contained 12 hard exudates, 5 cotton wool spots, 30 drusen and 92 candidates not corresponding to a bright lesions. We performed 160 iteration of the active learning approach. No stopping criterion was performed. In each run, we included only one additional training sample to X_t.

Table 2. Area under the ROC curve after 160 iterations of the selection process for each type of lesions.

	Hard exudates	**Cotton wool spots**	**Drusen**
Active sampling	0.87	0.82	0.78
Random sampling	0.72	0.78	0.68

Table 3. Number of actively selected samples to obtain the same performance as the random sampling system after 160 iterations

	Number of samples	**Area under the ROC curve**
Hard exudates	24	0.72
Cotton wool spots	68	0.78
Drusen	1	0.68

For each iteration i of the active learning system, the updated training set $X_t^{(i)}$ was used to train the classifier $c^{(i)}$. The classification performance $P^{(i)}$ was evaluated on the test set calculating the sensitivity and specificity values and creating the Receiver Operating Characteristic (ROC) curve. We use the area under the ROC curve of the current classifier as our performance metric. This metric is a comprehensive measurement of the system performance and eliminates the need to set a threshold on the class posterior probabilities.

The results were compared with the performance obtained using random sampling. This method randomly selects the next sample from the pool of unlabeled data. It corresponds to a *passive* learning model where the training set is a random sampling of the data.

4.2 Results

The results for each type of lesion and for each type of sampling criterion (uncertainty sampling and random sampling) are shown in Fig. 2. The figures plots the algorithm performance values obtained at each iteration of the learning process. It is seen that active sampling outperforms random sampling in the identification of the three abnormalities. The areas under the ROC curve after the 160 iterations for each type of lesions using both sampling selection methods are shown in Table 2. Table 3 shows the number of samples selected by the active learning process needed to obtain the same performance as the random sampling system after including 160 samples. Fig. 3 shows the distribution of the training samples for the final training set using the two different sampling criteria.

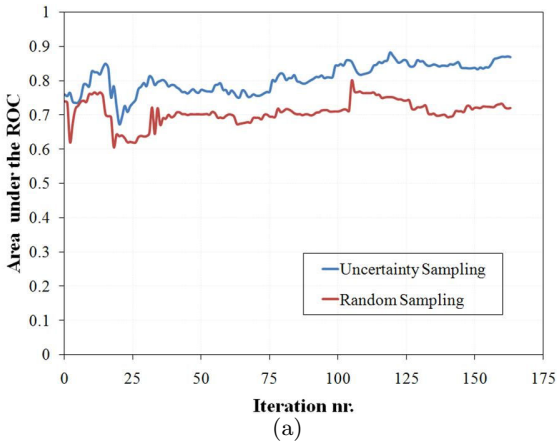
5. DISCUSSION

This paper presents a dynamic CAD system based on active learning for identifying hard exudates, cotton wool spots and drusen. The algorithm selected representative samples from a pool of unlabeled data. Therefore, this algorithm permitted using efficiently the expert knowledge because he/she only has to label those samples that increase the classifier performance.

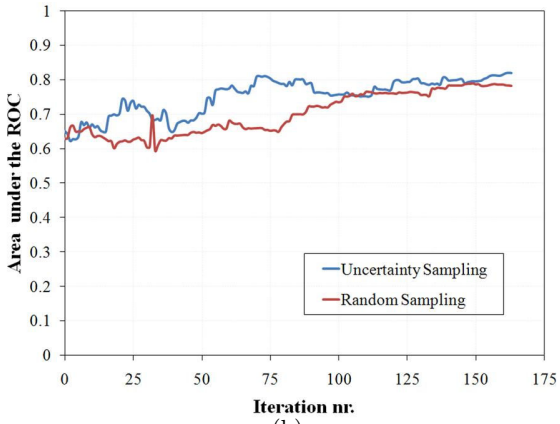
Fig. 2 shows that the active learning system outperformed random sampling in the identification of the three abnormalities because more informative samples were selected with the former algorithm that those that were picked randomly.

On the other hand, as shown in Table 3, to achieve the same classification performance, the active learning needed less labeled samples that using random sampling. Therefore, we reduced the user's involvement in the labeling process.

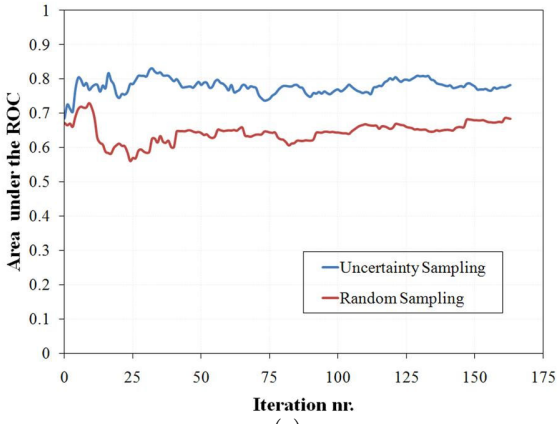
As we can see in Fig. 3, the active learning selected more samples of the different classes, obtaining a balanced training set. Thus, this training approach provides a high benefit when the classification problem presents infrequent classes because more representative samples from these classes can be picked from the pool of unlabeled data. Therefore, active learning can also be seen as a form of *data mining* on the training database.



(a)



(b)



(c)

Figure 2. Learning curves using active sampling and random sampling for the classification of (a) hard exudates; (b) cotton wool spots; and (c) drusen. The area under the ROC curve is measured in each iteration of the learning process.

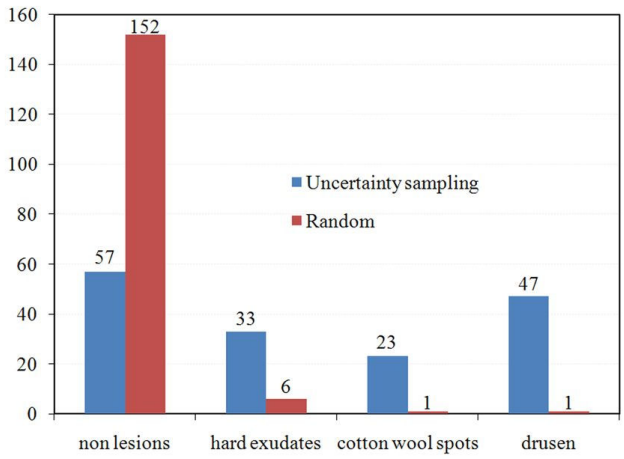


Figure 3. Number of samples selected by the uncertainty sampling function and the random sampling function.

Uncertainty sampling was used as the query function to select the samples that need to be query next. This exploitation learning samples the region in the vicinity of the decision boundary, where the classifier is least confident. Figure 4 depicts some examples of the selected objects. These are hard cases to classify, which supports the notion that the active learner selects samples close to its decision boundary. There exists other query functions, called explorative learning, which sample from regions where no labeled samples are presented.[4] It could be interesting to apply an approach using alternatively exploitation and explorative learning to sample in different regions of the sample space.

These are several topics that need to be dealt with in future works. As it is depicted in Fig. 2, the learning curves are quite noisy, mainly due to the reduced number of images used as the test set. Increasing this number the noise would be removed obtaining smoother curves. Additionally, different sizes of the initial training set could be studied to analyze its influence on the final classification performance. Finally, it is also interesting to experiment with different classifiers and feature selection during the active learning process.

6. CONCLUSIONS

An active learning approach for detection and differentiation of hard exudates, cotton wool spots and drusen was proposed. This permits the design of a classifier with automatically selected informative samples. Results show that the proposed method increases the classification accuracy compared to random sampling. Active learning reduces the number of labeled examples required and allows CAD systems to be trained with vast incompletely labeled databases that are now available in many application areas.

REFERENCES

[1] Kinyoun, J., Barton, F., Fisher, M., Hubbard, L., Aiello, L., and Ferris, F., "Detection of diabetic macular edema. Ophthalmoscopy versus photography–Early Treatment Diabetic Retinopathy Study Report Number 5. The ETDRS Research Group.," *Ophthalmology* **96**, 746–750 (1989).

[2] Early Treatment Diabetic Retinopathy Study Research Group, "Early Photocoagulation for Diabetic Retinopathy: ETDRS report 9," *Ophthalmology* **98**, 766–785 (1991).

[3] Niemeijer, M., van Ginneken, B., Russel, S., Suttorp-Schulten, M., and Abràmoff, M. D., "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *Investigative Ophthalmology & Visual Science* **48**, 2260–2267 (2007).

[4] Juszczak, P., *Learning to recognise*, PhD thesis, Delft University, the Netherlands (2006).

[5] Dura, E., Zhang, Y., Liao, X., Dobeck, G. J., and Carin, L., "Active learning for detection of mine-like objects in side-scan sonar imagery," *IEEE Journal of Oceanic Engineering* **30**(2), 360–371 (2005).
[6] Yan, R., Yang, J., and Hauptmann, A., "Automatically labeling video data using multi-class active learning," in [*Proc. Ninth IEEE International Conference on Computer Vision*], 516–523 (2003).
[7] Zhang, C. and Chen, T., "An active learning framework for content-based information retrieval," *IEEE Transactions on Multimedia* **4**(2), 260–268 (2002).
[8] Abràmoff, M. D. and Suttorp-Schulten, M. S. A., "Web-based screening for diabetic retinopathy in a primary care population: the EyeCheck project," *Telemedicine and e-Health* **11**(6), 668–674 (2005).
[9] Abràmoff, M. D. and Niemeijer, M., "The automatic detection of the optic disc location in retinal images using optic disc location regression," in [*Engineering in Medicine and Biology Society, 2006. EMBS '06.*], 4432–4435 (2006).
[10] Dima, C. and Hebert, M., "Active learning for outdoor obstacle detection," in [*Proceedings of Robotics: Science and Systems*], (2005).
[11] Niemeijer, M., van Ginneken, B., Staal, J., Suttorp-Schulten, M., and Abràmoff, M. D., "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on Medical Imaging* **24**(5), 584–592 (2005).
[12] Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., and Abràmoff, M. D., "Comparative study of retinal vessel segmentation methods on a new publicly available database," in [*Proceedings of SPIE: Medical Imaging*], **5370**, 648–656 (2004).

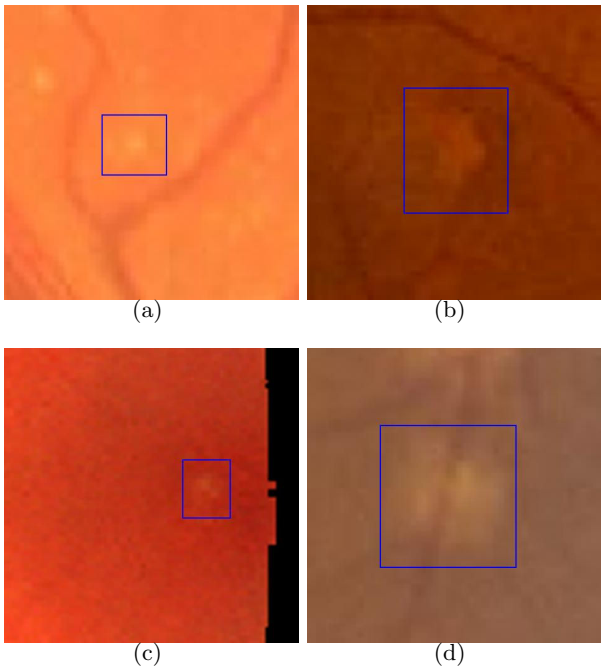


Figure 4. Examples of samples selected by the active learner.