Automatic classification and detection of clinically-relevant images for diabetic retinopathy

Xinyu Xu, Baoxin Li
Dept. of Computer Science and Engineering, Arizona State University, Tempe, AZ USA 85281

ABSTRACT

We proposed a novel approach to automatic classification of Diabetic Retinopathy (DR) images and retrieval of clinically-relevant DR images from a database. Given a query image, our approach first classifies the image into one of the three categories: microaneurysm (MA), neovascularization (NV) and normal, and then it retrieves DR images that are clinically-relevant to the query image from an archival image database. In the classification stage, the query DR images are classified by the Multi-class Multiple-Instance Learning (McMIL) approach, where images are viewed as bags, each of which contains a number of instances corresponding to non-overlapping blocks, and each block is characterized by low-level features including color, texture, histogram of edge directions, and shape. McMIL first learns a collection of instance prototypes for each class that maximizes the Diverse Density function using Expectation-Maximization algorithm. A nonlinear mapping is then defined using the instance prototypes and maps every bag to a point in a new *multi-class bag feature space*. Finally a multi-class Support Vector Machine is trained in the multi-class bag feature space. In the retrieval stage, we retrieve images from the archival database who bear the same label with the query image, and who are the top K nearest neighbors of the query image in terms of similarity in the multi-class bag feature space. The classification approach achieves high classification accuracy, and the retrieval of clinically-relevant images not only facilitates utilization of the vast amount of hidden diagnostic knowledge in the database, but also improves the efficiency and accuracy of DR lesion diagnosis and assessment.

Keywords: Multiple-instance learning, diabetic retinopathy lesion detection and classification, image retrieval, computer-aided diagnosis.

1. INTRODUCTION

Diabetic retinopathy (DR) is a common cause of blindness in working populations [1]. Timely diagnosis and accurate grading are essential for effective treatment of the decease and the reduction of blindness risk due to DR. Nowadays, ophthalmologists use digital retinal fundus images routinely for the classification and grading of diabetic retinopathy. In this process, the ophthalmologist needs to visually examine a digital retinal image, and mentally or physically compare it with standard images like EDTRS [2] photograph, or images from archival DR image database, in order to clarify confusions or to interpret more accurately the image. This process is laborious, and is prone to error or review fatigue. On the other hand, a vast amount of past clinical records with proven pathology cannot be easily utilized by the ophthalmologist due to the fact that the ophthalmologist may not have immediate access to the relevant past clinical records that are most relevant to a new image to be diagnosed. The vast amount of diagnostic knowledge hidden in an archival database is thus wasted. Motivated by the above observations, we propose a method for automatic DR image classification based on lesions contained in the image, and for automatic retrieval of clinically-relevant DR images from an archival database based on the acquired image label. Images are considered as being clinically relevant if they contain the same types of lesions with similar visual appearance.

The proposed method provides three benefits for DR diagnosis and evaluation. Firstly, our proposed approach facilitates timely and accurate evaluation of DR images by automatic DR lesion classification and clinically relevant DR image retrieval from a database. Secondly, the proposed method provides the clinicians a novel way of exploiting the vast amount of expert knowledge that is otherwise hidden in the database, effectively enabling knowledge sharing/reutilization among experts. This is particularly useful as a training tool for medical students and novice doctors. Thirdly, the classification scheme reply on Multi-class Multiple Instance Learning (McMIL) which does not require the

clinicians to know in advance what type of lesion is present in the image, and McMIL is able to capture the clinical relevance based on the content and semantic concept of the image rather than the low-level features.

While computer diagnostic systems probably will not replace the current screening modalities, as suggested by the American Diabetes Association [3], they will provide a modality that greatly reduces subjective human variation and the cost of the current screening process and therefore facilitate a more objective and economical method of diagnosis. In addition, the technology offered by the computer system will likely be an aid to human experts to produce diagnostic results of higher quality and accuracy. The proposed approach attempts to achieve these benefits by designing new algorithms that not only perform automatic lesion detection but also provide a way for the clinician to utilize the diagnosis records from an archival database.

2. RELATED WORK

Digital retinal images are increasingly being used in onsite or remote diagnosis of diabetic retinopathy. In the research community, several research groups have developed content-based image retrieval (CBIR) systems for use in fields other than ophthalmology [13-16]. These CBIR systems mimic domain knowledge used by humans to extract image content and provide query methods for direct image matching using low-level features. Cai *et al.* [13] developed a prototype retrieving positron emission tomography images based on physiological kinetic features. Fast query using nearest neighbor search is addressed by Korn *et al* [14] who attempted to retrieve medical tumors that are similar to a given pattern. The system proposed by Nah and Sheu [15] used operational semantics to improve content-based retrieval of brain neuroscience images. In [16], the similarity that is used to guide the CBIR engine is learned from training examples provided by human observers.

Previous CBIR-related research in ophthalmology include the Structured Analysis of the Retina (STARE) project aimed at automatic diagnosis and comparison of images, including annotating image contents and searching for images similar in content [17, 4]. The similarity metric in this system is based on some basic image features (called primitives) for a particular class of images [5]. While having made significant progresses since the inception of the project, the lack of a unified and systematic solution or system that has been widely accepted in ophthalmology shows that the central problems have not been solved. In particular, when it comes to CBIR for retinal images, the retrieved images given in [5] appear to be largely similar in appearance but not similar in terms of clinical relevance, rendering the results of very limited clinical use.

A recently-awarded National Eye Institute project [18], "Automated Screening for Diabetic Retinopathy by Content", is investigating the feasibility of CBIR methods for accurate description and indexing of human retinal images in diabetic retinopathy. The aim is to demonstrate the feature-based indexing and retrieval process of CBIR and to verify that retinal pathology can be identified and quantified from visually similar retinal images. The key CBIR technique used in this project appear to rely on explicit coding of domain knowledge in terms of image features, which has obvious disadvantages. Among others, one well-known challenge is the difficulty in explicitly converting expert-knowledge to simple rules in building feature detectors.

3. METHODOLOGY

We propose an approach to automatically classifying DR images into different categories based on the types of lesions contained in the images and automatically detecting clinically relevant DR images from an archival database (or training data set). The proposed method includes two stages, as shown in Fig. 1. At the first stage, we train a multi-class classifier using labeled DR training images so that it will correctly classify a novel DR image into one of the three categories: microaneurysm (MA), neovascularization (NV) and normal (no DR lesion is present). The reason these three categories are chosen is that MA and NV are two milestone signs during the entire DR development cycle: MA is an early sign of diabetic retinopathy, and it has been found in patients with pre-diabetes [2]; NV is a growth of new blood vessels on the surface of the retina [2], it signifies that the disease has reached a vision-threatening phase, thus timely treatment of NV significantly reduces the risk of vision loss. In the second stage, given a novel query image, we first use the classifier to obtain its label, and then we retrieve images from the archival database who bear the same label with the query image, and who are the top K nearest neighbors of the query image in terms of similarity in a *multi-class bag feature space*. The

retrieved images are clinically relevant to the query image in the sense that they bear the same label (MA, or NV or normal) and they are visually similar in terms of image content. Note that, unlike previous approaches [4, 5] where the similarity metric is based on some low-level image features like color, texture, the similarity metric in our method is based on high-level image content defined by the Instance Prototypes (IPs) returned from McMIL.

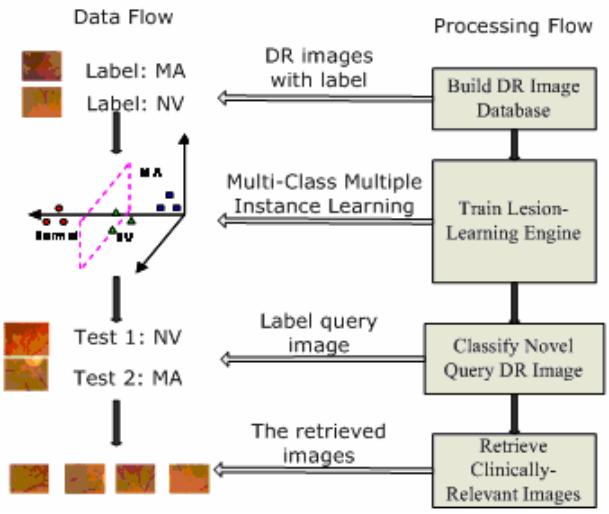

Fig. 1. The proposed framework.

3.1 Stage 1: DR image classification using McMIL

Multiple Instances Learning is a generalization of supervised classification in which training class labels are associated with the *bag*, i.e. a DR image, instead of individual instances. This is advantageous over instance-level classifiers since the labels are required only at a higher level. Let D be the labeled data which consists of a set of m images $B=\{B_1,B_2,...,B_m\}$ and their labels $L=\{l_1,l_2,...,l_m\}$, i.e. $D=\{<B_1,l_1>,...<B_m,l_m>\}$, where B_i, $i=1,...,m$ is called a *bag* (i.e. a DR image) in the MIL setting, and $l_i \in \{1,\ldots,c\}$ denotes the multi-class label. In our method, the bag B_i is partitioned into non-overlapping blocks of size MxM pixels, and we extract features for each block, i.e. $B_i=\{B_{ij}: j=1,..., N_i\}$,where B_{ij} represents the j^{th} instance in the i^{th} bag, and B_{ij} is a d-dimensional feature vector built for the j-*th* block. Our goal in the first stage is to induce a classifier from patterns to multi-class labels.

Fig. 2 shows the processing flow of the stage 1. We now present in detail each step of stage 1.

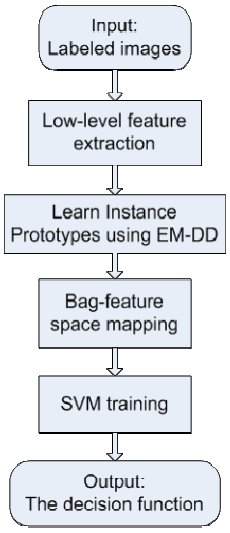

Fig. 2. The training stage of McMIL

3.1.1 Low-Level Feature Extraction

Each training image is partitioned into non-overlapping blocks of size MxM pixels. A feature vector is then extracted for each block. Each feature vector contains 13 features including color, texture, histogram of edge directions and shape. The computation of each feature component is described in details as follows. Table 1 summarizes the image features.

Color: Three of them are the average color components in a block. We use the HSV color space. The color feature is very useful in distinguishing lesion area from retina background such as optic disk and macula since optic disk (macular) appears brighter (darker) than MA/NV.

Texture: Three other features represent the square root of the second order moment of wavelet coefficients in high-frequency bands [6]. To obtain these moments, a Daubechies-4 wavelet transform is applied to the H component of the image. After a one-level wavelet transform, a 4x4 block is decomposed into four frequency bands: the LL, LH, HL, and HH bands. Each band contains 2x2 coefficients. As was done in [11], we suppose the coefficients in the HL band are $\{c_{k,l}, c_{k,l+1}, c_{k+1,l}, c_{k+1,l+1}\}$. One feature is

$$f = \left(\frac{1}{4} \sum_{i=0}^{1} \sum_{j=1}^{1} c_{k+i,l+j}^2 \right)^{1/2} \quad (1)$$

The other two features are computed similarly to the LH and HH bands. Moments of wavelet coefficients in various frequency bands are effective for representing texture. For example, the HL band shows activities in the horizontal direction. Thus, a DR image block with vertical retina vessels has high energy in the HL band and low energy in the LH band.

Histograms of edge directions: The edge histogram is translation invariant, and because the feature is local, it is robust to partial occlusion and local disturbance in the image. The histogram of edge directions are computed as follows. We first obtain the luminance image by transforming the RGB color image into intensity image; we then convolve the intensity image with horizontal and vertical Sobel operator (3x3) respectively to yield the horizontal and vertical gradient. The two gradients are then combined to get the final gradient. After that, we throw away the pixels whose gradient is less than a certain threshold (the threshold is set to 60% of the maximum gradient value). Next, we obtain the direction of the edges (angles in radians) by computing the arctangent of the gradient at each pixel. Then, the direction histograms are calculated by counting the edge pixels whose arctangent falls into the horizontal, the vertical, or the two diagonal directions, resulting in four features. At last, we normalize the histogram by the number of edge pixels. The histogram of edge directions is very powerful in distinguishing NVs from MAs in that the direction histogram of MA is uniform, but the direction histogram of NV usually biases toward a certain angle range, this is because MA typically assumes round shape (like a disk), while NV appears like vessels which grow toward a specific direction.

Shape: The shape is normalized inertia [7] of order 1, 2 and 3. For a block R_j in a DR image, the normalized inertia of order γ is given as

$$I(R_j, \gamma) = \frac{\sum_{r \in R_j} \left\| r - \hat{r} \right\|^{\gamma}}{V_j^{1+\frac{\gamma}{2}}} \quad (2)$$

where $\hat{r}$ is the centroid of block R_j, V_j is the number of pixels in block R_j. The minimum normalized inertia on a 2-D plane is achieved by circles. Denote the γ-th order normalized inertia of circles as I_γ. The shape features of region R_j is defined as

$$S_j = \left[\frac{I(R_j,1)}{I_1}, \frac{I(R_j,2)}{I_2}, \frac{I(R_j,3)}{I_3} \right] \quad (3)$$

Table 1. The low-level features for an instance (a block)

Feature Name	Description	Dimensions
Color	The average HSV color intensity within a block	3
Texture	The square root of energy in the high-frequency bands of the Daubechies-4 [6] wavelet transforms (HL, LH, and HH bands).	3
Histograms of edge directions	The histograms of edges in 4 directions: horizontal, vertical and two diagonals.	4
Shape	The normalized inertia [7] of order 1, 2, and 3.	3

3.1.2 Learning Instance Prototypes using EM-DD

Maron and Lozano-perez [8] first proposed Diverse-Density (DD) algorithm to solve MIL problems. Based on the DD framework, Zhang and Goldman [9] proposed the EM-DD algorithm. For completeness of this paper, we now briefly introduce the DD and EM-DD algorithms upon which our McMIL is based.

In the case of binary classification, the diverse density at a point ***p*** in the feature space is a probabilistic measure of both how many *different* positive bags have an instance near ***p***, and how far the negative instances are from ***p***. Intuitively, the diverse density of a hypothesis h is just the likelihood (with respect to the data) that h is the target. A high diverse density indicates a good candidate for a "true" concept. The diverse density of hypothesized target point h is defined as:

$$DD(h) = \Pr(h \mid D) = \frac{\Pr(D \mid h)\Pr(h)}{\Pr(D)} = \frac{\Pr(B, L \mid h)\Pr(h)}{\Pr(B, L)} \tag{1}$$

Assuming a uniform prior on the hypothesis space and independence of $<B_i, l_i>$ pairs given h, using Bayes' rule, the maximum likelihood hypothesis, h_{DD}, is defined as:

$$h_{DD} = \arg\max_{h \in H} \Pr(D \mid h) = \arg\max_{h \in H} \prod_{i=1}^{m} \Pr(B_i, l_i \mid h) = \arg\min_{h \in H} \sum_{i=1}^{m} \left(-\log \Pr(l_i \mid h, B_i)\right) \tag{2}$$

As in [19], $\Pr(l_i \mid h, B_i)$ is estimated as $1-|l_i-Label(B_i|h)|$ where $Label(B_i|h)$ is the label that would be given to B_i if h were the correct hypothesis. For most applications the influence each feature has on the label varies greatly. This variation is modeled in the DD algorithm by associating with each attribute a (unknown) scale factor. Hence the target concept really consists of two values per dimension, the ideal attribute value and the scale value. The following generative model was introduced by Maron and Lozano-perez [8] for estimating the label of bag B_i for hypothesis $h=\{h_1,...,h_n,s_1,...,s_n\}$:

$$Label(B_i \mid h) = \max_j \left\{ \exp\left[-\sum_{d=1}^{n} (s_d (B_{ijd} - h_d))^2 \right] \right\} \tag{3}$$

where s_d is a scale factor indicating the importance of feature d, h_d is the feature value for dimension d, and B_{ijd} is the feature value of instance B_{ij} on dimension d. Let $\mathrm{NLDD}(h, D) = \sum_{i=1}^{m} \left(-\log \Pr(l_i \mid h, B_i)\right)$, NLDD denotes the negative logarithm of DD. The DD algorithm of [8] uses a two-step gradient descent search to find a value of h that minimizes NLDD (and hence maximizes DD).

One reason why MI learning is so difficult is the ambiguity caused by not knowing which instance is responsible for the label associated with the bag. The basic idea behind EM-DD [9] is to view the knowledge of which instance corresponds to the label of the bag as a missing attribute which can be estimated using EM approach. EM-DD starts with some initial guess of a target point h obtained in the standard way by trying points from positive bags, and then repeatedly perform the following two steps that combine EM with DD to search for the maximum likelihood hypothesis. In the first step (*E*-step), the current hypothesis h is used to pick one instance from each bag which is most likely (given the generative model) to be the one responsible for the label given to the bag. In the second step (*M*-step), a two step gradient descent search (Quasi-Newton search) of the standard DD algorithm is used to find a new h' that minimizes NLDD(h). Once this maximization is completed, the proposed target h is reset to h' and return to the first step until the algorithm converges.

An instance prototype (IP) represents a class of instances that is more likely to appear in positive bags than in negative bags. In addition, IPs with the exactly reversed property (more likely to appear in negative images than in positive images) may be of equal importance [11], so we also record the local valley points of the DD function. We apply EM-DD [9] to locate the multiple IPs by using every instance in every positive (negative) bag with uniform weighs as the starting point of the Quasi-Newton search. Finally the IPs with larger DD values and are distinct from each other are selected to represent the target concepts.

3.1.3 Multi-class bag space feature mapping

Both DD and EM-DD were originally designed to solve binary MIL problems. For intrinsically multi-class MIL classification problems, these methods need to construct a set of one-versus-all binary classifiers, each is trained to separate one from all the remaining classes. There are a few drawbacks with binary MIL when applied to a multi-class classification problems [20]. We proposed Multi-class Multiple-Instance Learning (McMIL) [20] that bypasses the necessity of constructing a series of binary classifiers. In the McMIL, we first use EM-DD to learn the IPs, we then map every training image onto a multi-class bag feature space defined by the IPs of all the categories. Let $IP_r = \left\{ \left(I_r^1, I_r^2, ..., I_r^{n_r} \right) \right\}, r \in \{1, 2, ..., c\}$ be the set of instance prototypes learned by EM-DD for the r^{th} category. Each $I_r^k = \left[x_r^k, e_r^k \right]$, $k=1,2,...,n_r$ denotes the k^{th} IP in the r^{th} class, and n_r is the number of IPs in the r^{th} class. The learned IP consists of two parts: the ideal attribute value x and the weight factor e. To facilitate the multi-class classification, a

similarity mapping kernel ψ is defined which maps a bag of instances into a multi-class bag feature space based on a similarity metric function q. The mapping ψ is defined as

$$\psi(B_i)= \left[q(B_i, I_1^1),...,q(B_i, I_1^{n_1}),...,q(B_i, I_c^1),...,q(B_i, I_c^{n_c}) \right]^T \tag{4}$$

The similarity function q should be chosen properly such that it takes into account the prior knowledge or domain insights of the specific classification task. In our work, q is defined as

$$q(B_i, I_r^k) = \max_{j=1,...,|B_i|} \left\{ \exp\left[-\left\| B_{ij} - x_r^k \right\|^2_{\varepsilon_r^k} \right] \right\} \tag{5}$$
$$k = 1,...,n_r, r \in \{1,...,c\}$$

Each feature in $\psi(B_i)$ corresponds to the similarity between a bag and one IP, which is the similarity between I_r^k and the instance in B_i that is closest to I_r^k. Unlike [11], the mapping (Eq. 4) in our method simultaneously incorporate the similarity between a bag and each IP of *every* category into one feature matrix, while in binary MIL [10, 11], separate feature matrices are needed, each of which only considers the similarity between a bag and each IP of *one* category that is under consideration.

For a given training set of DR images with multi-class labels $D=\{<B_1,l_1>,...<B_m,l_m>\}$, applying the above mapping yields the following kernel similarity matrix **S**:

$$\mathbf{S}= \begin{bmatrix} q^1 \\ q^2 \\ \vdots \\ q^m \end{bmatrix} = \begin{bmatrix} q(B_1, I_1^1),....,q(B_1, I_1^{n_1}),....,q(B_1, I_c^1),....,q(B_1, I_c^{n_c}) \\ q(B_2, I_1^1),....,q(B_2, I_1^{n_1}),....,q(B_2, I_c^1),....,q(B_2, I_c^{n_c}) \\ ..... \\ q(B_m, I_1^1),...,q(B_m, I_1^{n_1}),...,q(B_m, I_c^1),...,q(B_m, I_c^{n_c}) \end{bmatrix} \tag{6}$$

3.1.4 Multi-Class SVM Training and Classification

Given the kernel similarity matrix **S**, the multi-class classification is achieved by multi-class SVM [12], which simultaneously allows the computation of a multi-class classifier by considering all the classes at once. The formulation is given by

$$\min_{w,b,\xi} \frac{1}{2} \sum_{r=1}^{c} w_r^T w_r + C \sum_{i=1}^{m} \sum_{r \neq l_i} \xi_i^r$$
$$s.t. \quad w_{l_i}^T \psi(B_i) + b_{l_i} \geq w_r^T \psi(B_i) + b_r + 2 - \xi_i^r, \tag{7}$$
$$\xi_i^r \geq 0, \quad i = 1,...,m, \quad r \in \{1,..,c\}/l_i .$$

where w is a vector orthogonal to the separating hyper-plane, $|b|/||w||$ is the perpendicular distance from the hyper-plane to the origin, C is a parameter to control the penalty due to misclassification. Nonnegative slack variables ξ_i are introduced in each constraint to account for the cost of the errors.

For a novel DR image B, we use Eq. 4 to project the image into the multi-class bag feature space □ defined by the IPs. The decision function is then given by

$$\arg\max_{r=1,...,c} \left(w_r^T \psi(B) + b_r \right) \tag{8}$$

Note that in multi-class SVM, there are c decision functions but all are obtained by solving the problem of Eq. (7).

3.2 Stage 2: retrieval of clinically-relevant images from database

In this stage, we develop algorithms to retrieve clinically-relevant images from a database, for a given query image. The processing flow of Stage 2 is depicted in Fig. 3. It includes two steps: classification and retrieval, as described in the following.

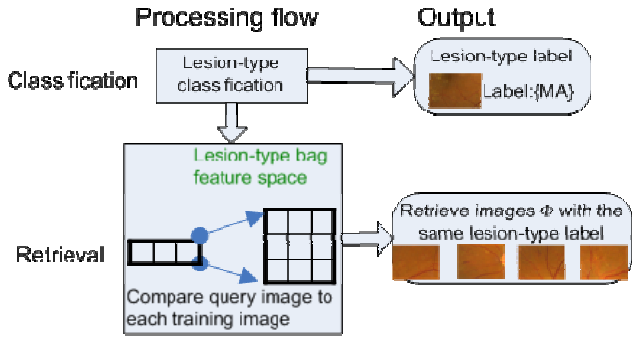

Fig. 3. The processing flow of Stage 2 and the outputs of the steps.

3.2.1 Classify a novel query image

Given a novel query DR image, we first extract low-level features, and then we map the query image into the bag feature space using Eq. (4-6). Finally, the label of the DR image is returned by SVM classification in the multi-class bag feature space using Eq. (8).

3.2.2 Retrieve clinically-relevant DR image

After we obtain the label of the novel query image, we will further retrieve the DR images from an archival database that are clinically relevant to the query image. This involves two steps. In the first step, we select the images from the database whose label is the same as the label of the query image. In the second step, we compare the feature vector of the query image, $\boldsymbol{\psi}(\mathbf{x})$, with the feature vector of each selected candidate image, $\boldsymbol{\psi}(\mathbf{x}_i)$, in the multi-class bag feature space (Eq. 4); the images in the database that are the top K nearest neighbors (in terms of Euclidean distance) of the query are declared as the clinically-relevant images. Note that the K-nearest neighbor search is carried out in the multi-class bag feature space defined by the instance prototypes, not the original low-level features, thus our method is able to capture the high-level concept of the image content.

4. RESULTS

We have built a small DR image database to test the proposed method. The DR image database is comprised of 321 DR images, among which 125 images are labeled as NV images, 88 are labeled as MA, and the remaining 108 images contain no lesion, thus are labeled as Normal. The images are in size 768x576, and they are taken from different field of view. The block size in each training image is 96x96. Images within each category were randomly partitioned in half to form a training set and a test set. We repeat the classification and retrieval experiment for 5 random splits, and report the classification accuracy over 5 random test sets.

The confusion matrices for the 5 random splits are shown in Table 2. The diagonal entries correspond to the classification accuracy for the respective class, and the off-diagonal entries represent the classification errors. The proposed McMIL approach is able to correctly classify 80% of NV and MA images, and all of the normal images have been correctly classified. Fig. 4 shows the lesion blocks in the training images that maximize the similarity between images and the instance prototypes. These blocks can be viewed as the target concepts of the lesions that distinguish one class from the other. We can see that the major target concept of the lesion have been captured by the Multiple-Instance Learning. Fig. 5 shows some query images and the corresponding top 4 clinically-relevant images whose labels are the same with that of the query image. We can see that the query image look very similar to the retrieved images in terms of visual appearance and clinical relevance.

5. CONCLUSIONS

In this paper, we proposed a novel approach to automatically classification of Diabetic Retinopathy (DR) images and retrieval of clinically-relevant DR images from a large-scale database. The DR image classification method is based on

McMIL approach, and it achieves high classification accuracy in classifying images with clinically important DR lesions including microaneurysm and neovascularization. The detection of clinically-relevant images from a large-scale database not only facilitates effective utilization of the vast amount of hidden diagnostic knowledge in the database, but also improves the efficiency and accuracy of DR lesion diagnosis and assessment.

Table 2. The confusion matrix (in %) of the DR image classification for the 5 random splits. 1: MA, 2: NV, 3: Normal

	The 1st split			The 2nd split			The 3rd split			The 4th split			The 5th split		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	91.9	8.1	0.0	85.5	14.5	0.0	64.5	35.5	0.0	85.5	14.5	0.0	98.4	1.6	0.0
2	27.3	72.7	0.0	15.9	84.1	0.0	15.9	84.1	0.0	29.5	70.5	0.0	31.8	68.2	0.0
3	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0

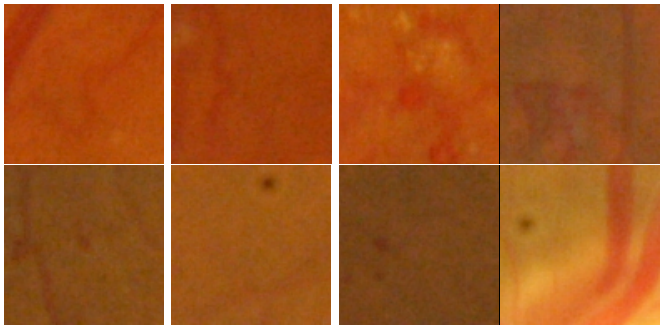

Fig. 4. Top row: the image blocks that contains NV. Bottom row: the image blocks that contains MA.

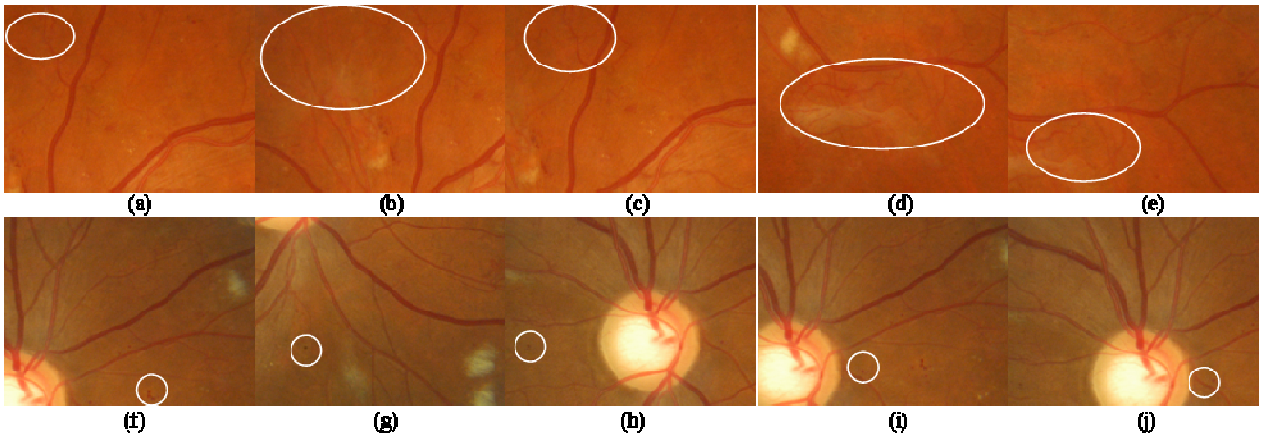

Fig. 5. (a) A query image which is correctly classified as NV. (b-e) The four retrieved images that also belong to NV and that are the top 4 nearest neighbors of the query image (a). The NV areas are marked by circles. (f) A query image which is correctly classified as MA. (g-j) The four retrieved images that also belong to MA and that are the top 4 nearest neighbors of the query image (f). The MA areas are marked by circles.

REFERENCES

[1] Centers for Disease Control and Prevention. Prevalence and incidence of diabetes mellitus in United States, 1980-1987. *MMWR Morb. Mortal. Wkly. Rep. 39* (1990), 809-812.

[2] Early Treatment and Detection of Diabetic Retinopathy Study Group. Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs - An Extension of the Modified Airlie House Classification. *EDTRS Report Nr. 10*, *Ophthalmology 98* (1991), 786-806.

[3] American Diabetes Association. Screening for diabetic retinopathy: clinical practice recommendations. *Diabetes Care **20*** (suppl 1):S1-S70 (1997).

[4] Hoover, A., Kouznetsoza, V., Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imag*., *19,* 3 (2000), 203–210.

[5] Gupta, A., Moezzi, S., Taylor, A., Chatterjee, S., Jain, R., Goldbaum, M., Burgess, S. Content-based retrieval of ophthalmological images. *ICIP* (1996).

[6] Daubechies, I. *Ten Lectures on Wavelets*. Capital City Press, 1992.

[7] Gersho, A. Asymptotically optimum block quantization. *IEEE Transactions on Information Theory*, *25*, 4 (1979) 373-380.

[8] Maron, O., Lozano-perez, T. A framework for multiple-instance learning. NIPS *10* (1998).

[9] Zhang, Q., Goldman, S. A.. EM-DD: An improved multiple-instance learning technique. In *NIPS 14* (2002), 1073–1080.

[10] Bi, J., Chen, Y., Wang, J. Z. A sparse support vector machine approach to region-based image categorization. *Proc. CVPR 1* (2005), 1121-1128.

[11] Chen, Y., Wang, J. Z. Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, *5* (2004), 913-939.

[12] V. N. Vapnik. *Statistical Learning Theory*. New York, NY: Wiley, 1998.

[13] Cai, W., Feng, D. and Fulton, R. Content-based retrieval of dynamic PET functional images. *IEEE Trans. Inf. Technol. Biomed*, *4*, 2 (2000), 152-158.

[14] Korn, F., Sidiropoulos, N., Faloutsos, C., Siegel, E. Protopapas, Z. Fast and effective retrieval of medical tumor shapes. *IEEE Trans. Knowl. Data Eng., 10*, 6 (1998), 889-904.

[15] Nah, Y. and Sheu, P. C.-Y. Image content modeling for neuroscience databases. In *Proc. Int. Software Engineering and Knowledge Engineering Conf* (Jul. 2002), 91–98.

[16] El-Naqa, I., Yang, Y., Galatsanos, N.P., Nishikawa, R.M., Wernick, M. N. A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Trans. Medical Imaging, 23*, 10 (Oct. 2004), 1233-1244.

[17] Goldbaum, M., Katz, N., Chaudhuri, S., Nelson, M. Image understanding for automated retinal diagnosis. In *Proc. IEEE*

Computer Soc. Symp. on Computer Aided Medical Care (SCAMC) (Washington D.C, Nov. 1989). 756-760.

[18] From http://crisp.cit.nih.gov.

[19] Amar, R. A., Dooly, D. R., Goldman , S.A. and Zhang, Q. Multiple-instance learning of real-valued data. *Proceedings 18th ICML* (2001), 3-10, CA: Morgan Kaufmann.

[20] Xu, X., Li, B., Multiple class multiple-instance learning and its application to image categorization. *International Journal of Image and Graphics (IJIG)* (2007) vol. 7, NO. 3, pp. 427-444.