Computerized Medical Imaging and Graphics 55 (2017) 54–67

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm

Fulong Ren[a,b], Peng Cao[a,b,*], Wei Li[b], Dazhe Zhao[a,b], Osmar Zaiane[c]

[a] College of Computer Science and Engineering, Northeastern University, Shenyang, China
[b] Key Laboratory of Medical Image Computing of Ministry of Education, Northeastern University, Shenyang, China
[c] Computing Science, University of Alberta, Edmonton, Alberta, Canada

ARTICLE INFO

Article history:
Received 31 January 2016
Received in revised form 17 June 2016
Accepted 29 July 2016

Keywords:
Microaneurysm detection
Classification
False positive reduction
Imbalanced data learning
Ensemble learning

ABSTRACT

Diabetic retinopathy (DR) is a progressive disease, and its detection at an early stage is crucial for saving a patient's vision. An automated screening system for DR can help in reduce the chances of complete blindness due to DR along with lowering the work load on ophthalmologists. Among the earliest signs of DR are microaneurysms (MAs). However, current schemes for MA detection appear to report many false positives because detection algorithms have high sensitivity. Inevitably some non-MAs structures are labeled as MAs in the initial MAs identification step. This is a typical "class imbalance problem". Class imbalanced data has detrimental effects on the performance of conventional classifiers. In this work, we propose an ensemble based adaptive over-sampling algorithm for overcoming the class imbalance problem in the false positive reduction, and we use Boosting, Bagging, Random subspace as the ensemble framework to improve microaneurysm detection. The ensemble based over-sampling methods we proposed combine the strength of adaptive over-sampling and ensemble. The objective of the amalgamation of ensemble and adaptive over-sampling is to reduce the induction biases introduced from imbalanced data and to enhance the generalization classification performance of extreme learning machines (ELM). Experimental results show that our ASOBoost method has higher area under the ROC curve (AUC) and G-mean values than many existing class imbalance learning methods.

1. Introduction

Diabetic retinopathy (DR) is a chronic progressive disease of the retinal microvasculature which is the most common cause of blindness in the past 50 years. Due to the burden of diabetes over the past decades, the prevalence of DR is expected to grow exponentially and affect over 366 millions people worldwide by 2030 (Erbas et al., 2011). Despite the high risk factor, it has been established that early detection and timely treatment can reduce the development of severe vision loss in 60% of cases. In order to prevent the damage of this severe complication to patients' vision, it is very important to diagnose diabetic retinopathy and provide appropriate treatment to minimize further deterioration as early as possible.

Since microaneurysms (MAs) are regarded as early signs of DR and are caused by the focal dilatations of thin blood vessels, the detection of MAs in the fundus of the eye is essential. Moreover, the grading performance of computer-aided DR screening systems highly depends on MA detection. Therefore, there has been extensive research on effective detection and localization of these abnormalities in retinal images (Antal and Hajdu, 2013). MAs appear as small circular dark spots on the surface of the retina. The variation in size, shape, intensity, and presence of other retina structures further complicate the problem of detecting MAs from fundus images. Fig. 1 shows a retinal image containing the signs of DR-specific lesions, such as MAs, hard exudates (HE), hemorrhages (H) and soft exudates/cotton wool spots (SE/CWS).

Automatic methods have been developed to help reduce the burden on specialists (Dupas et al., 2010a). Current computer aided diagnosis (CAD) schemes for microaneurysms have achieved high sensitivity levels, whereas current schemes for microaneurysms detection appear to report many false positives (Mrinal, 2015; Akram et al., 2013; Walter et al., 2007; Mizutani et al., 2009).

Many methods have been developed specifically to deal with the problem of false positive detections. For example, in Mizutani et al. (2009), the numbers of false positives per image (FPI) decreases from 183.38 (initial detection) to 27.04 (false positive reduction using ANN); in Walter et al. (2007), the segmentation algorithm

* Corresponding author at: College of Computer Science and Engineering, Northeastern University, Shenyang, China.
E-mail address: caopeng@ise.neu.edu.cn (P. Cao).

http://dx.doi.org/10.1016/j.compmedimag.2016.07.011
0895-6111/

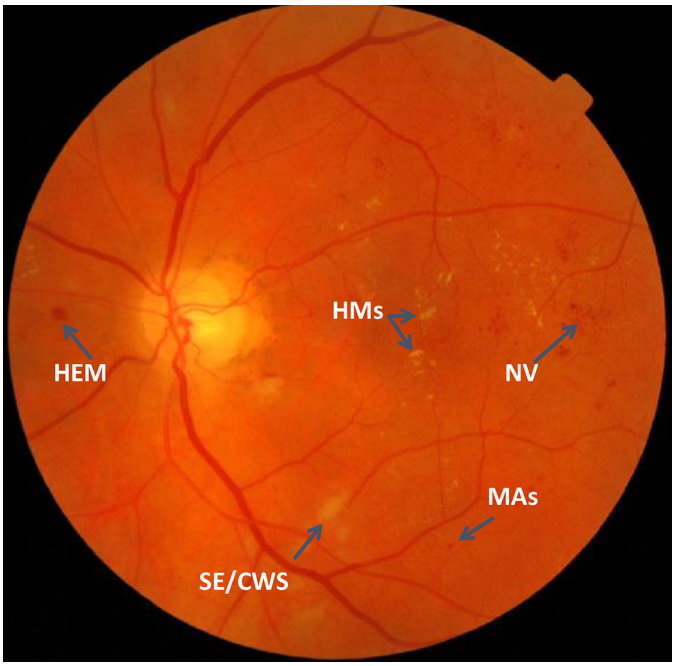


Fig. 1. Retinal image and MAs.

gave 8837 candidates, containing 373 MA, 92 doubtful objects (which were discarded) and 8372 false positives, false positive reduction yields a sensitivity of 89.0% and a FPI value of 2.92.

It is highly desirable to eliminate these false positives (FPs) because of two reasons: (1) The initial MA detection algorithms have high sensitivity that some non-microaneury structures are erroneously labeled as microaneurysms in the initial microaneurysms identification step, resulting in low values for specificity. Massive MA candidates bring much work and burden for radiologists, since the radiologists must examine each identified object. (2) The accurate grading of diabetic retinopathy (DR) depends on the number of lesions involving MAs, haemorrhages or exudates. The excessive false positives (capillary blood vessels and noise regions) of MAs negatively influence the accurate diagnosis of diabetic retinopathy.

Therefore, a fine level algorithm (usually a supervised classifier) is critical to remove the potentially false detections based on some assumptions about MAs after a coarse level initial microaneurysms identification. It is a binary classification between the MAs and non-MAs. In machine learning, the aim of classification is to learn a system capable of the prediction of the unknown output class of a previously unseen instance with a good generalization ability. The false-positive reduction step, or classification step, is a critical part in the microaneurysms detection system.

However, the two classes are skewed in the classification and have extremely unequal misclassification costs, which is a typical class imbalance problem (Chawla et al., 2004; He and Garcia, 2009). Class imbalanced data has detrimental effects on the performance of conventional classifiers. Typically classifiers attempt to reduce global error rate without taking the data distribution into consideration. As a result, all instances are misclassified as negative for high classification accuracy. However, in the potential MA detection and classification, this problem has attracted less attention.

Extreme learning machines (ELM) (Huang et al., 2006), as an effective and efficient machine learning technique, has attracted tremendous attention from various fields in recent years. ELM have been extensively studied and have shown remarkable success in many applications. However the success of ELM is very limited when it is applied to the problem of learning from imbalanced datasets. Much work has been done in addressing the class imbalance problem. The proposed methods can be grouped in two categories: the data perspective and the algorithm perspective. The re-sampling technique is the most straightforward and effective method for dealing with imbalance, since it is not dependent on the classifier and is simple to implement. In addition to re-sampling methods, ensemble methods (Galar et al., 2012) have also been used to improve performance on imbalanced datasets. They combine the power of multiple classifiers trained on similar datasets to provide accurate predictions for future instances. In order to improve microaneurysm detection, we propose an adaptive SMOTE (Synthetic Minority Over-sampling TEchnique), and incorporate the over-sampling technique with different ensemble frameworks to acquire better classification performance and generalization capability. We propose three different ensemble methods combined with adaptive SMOTE, to overcome the drawback of ELM on the imbalanced data learning and improve the performance of false positive reduction of MA candidates.

Specifically, the contributions of this paper are highlighted as follows:

1. Based on the ideas of the adaptive weighting (He et al., 2008) and probability function estimating (Cao et al., 2014), we extend SMOTE to generate appropriate synthetic instances for imbalanced data learning.
2. In this paper, we propose three different ensemble classifiers (boost, bagging and random subspace) combined with an adaptive over-sampling algorithm. In addition, ELM classifier is employed as base classifier and incorporates the evaluation measures (AUC and G-mean) of imbalanced data directly into the optimization of the intrinsic parameters to improve the performance of the classification.
3. Comprehensive experiments have been conducted to demonstrate the effect of over-sampling and ensemble learning for false positive reduction of MA. Moreover, we empirically compare the proposed three ensemble framework with adaptive SMOTE with existing state-of-the-art algorithms for imbalanced data learning algorithms and false positive reduction of MA candidates.
4. We implement a complete automated systems for the detection of MAs. The workflow of our system is shown in Fig. 2. The system involves the steps of initial detection of MAs candidate, feature extraction and classification.
5. The proposed methods have been tested on three publicly available Messidor DiaretDB1 and ROC datasets. The results verify the effectiveness of our automated microaneurysm detection system.

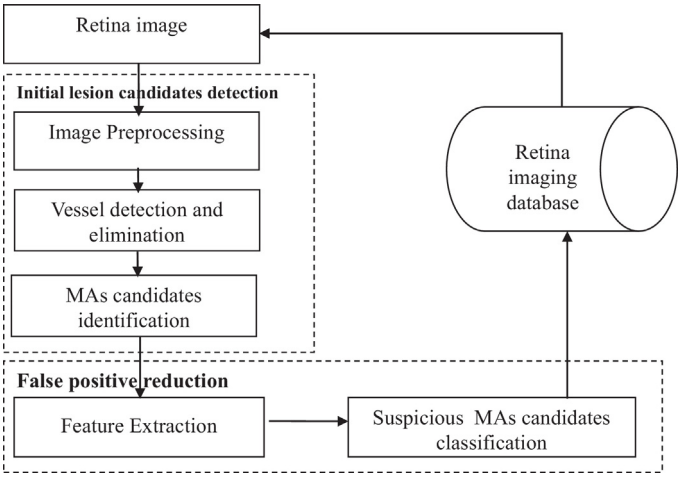


Fig. 2. The stage of MA detection.

The remainder of the paper is organized as follows: in Section 2 we review current state-of-the-art techniques for tackling the candidate MA classification problem as well as the imbalanced data learning. In Section 3, we present the preliminaries involving the algorithm of SMOTE and ELM. In Section 4, we introduce the proposed methods, adaptive SMOTE combined with Adaboost, Bagging and Ransom subspace ensemble. In Section 5 we present experimental results and draw our conclusion in Section 6.

2. Related work

2.1. Computer aided detection of MAs

Adal et al. (2014) proposed a semi-supervised based learning approach to train a classifier which can detect true MAs, which can be built from few supervised information (manually labeled training data) together with a large number of unsupervised information (unlabeled data). Zhang et al. (2012) propose new methods to detect MA based on Dictionary Learning (DL) with Sparse Representation Classifier (SRC), and extract retinal blood vessels using SRC. Akram et al. (2013) present a hybrid classifier which combines the Gaussian mixture model (GMM), support vector machine (SVM) and an extension of multimodel mediod based modeling approach in an ensemble to improve the accuracy of classification between MA and non-MA regions. Walter et al. (2007) propose a method based on diameter closing and kernel density estimation for automatic classification. All candidate MAs are detected by diameter closing and thresholding. Then 15 features are extracted for classification that relies on kernel density estimation with variable bandwidth. Dupas et al. (2010b) use a diameter-closing to segment MA candidate regions and *K*-nearest neighbors (kNN) to classify MA. Zhang et al. (2010) use multi-scale correlation coefficients (MSCF). They detect pixels which are candidates of the MA using MSCF and fine MA using features classification. Niemeijer et al. (2005) extract 68 features, and use a *K*-nearest neighbor (kNN) classifier for the classification of the candidate objects in the final step. In Gardner et al. (1996), an artificial neural network (ANN) was used to automatically classify microaneurysms. Giancardo et al. (2013) detect microaneurysms from morphological heuristics, and then apply a standard SVM to predict the disease status. Roberto et al. (2015a) proposed a hierarchical system of classifiers which consists of a structure of two levels of classification, so as to distinguish the MA and non-MA. Zhang et al. (2012) located all possible MA candidates with Multi-scale Gaussian Correlation Filtering (MSCF), and then classified these candidates with Dictionary Learning (DL) via Sparse Representation Classifier (SRC).

2.2. Imbalanced data learning

A dataset is unbalanced when the class of interest (minority class) is much smaller or rarer than a normal class (majority class). Traditional classification algorithms in general suffer when the data is skewed towards one class. Not only in the medical lesion detection domain, many real-world applications, such as Spam filtering, text classification and fraud detection in business transactions, have problems when learning from imbalanced data sets. In recent years, the imbalanced data learning problem has drawn a significant amount of interest from academia, industry, and government funding agencies. Much work has been done in addressing the class imbalance problem. These methods can be grouped in two categories: the data perspective and the algorithm perspective (He and Garcia, 2009). The methods with the data perspective rebalance the class distribution by re-sampling the data space, either over-sampling instances of the minority class or under-sampling instances of the majority class. The re-sampling techniques try to balance out the dataset either randomly or deterministically. A widely used over-sampling technique is called SMOTE (Synthetic Minority Over-sampling TEchnique), which creates synthetic samples between each positive sample and one of its neighbours (Chawla et al., 2002). SMOTE is effective to increase the significance of the positive class in the decision region. The methods with the algorithm perspective adapt existing common classifier learning algorithms to bias towards the small class, such as one-class learning and cost sensitive learning. Cost-sensitive learning is one of the most important topics in machine learning and data mining, and has attracted high attention in recent years. It takes misclassification costs into account during the model construction, and does not modify the imbalanced data distribution.

3. Preliminaries

3.1. SMOTE

A popular and effective over-sampling method is the Synthetic Minority Over-sampling TEchnique (SMOTE) (Chawla et al., 2002).The SMOTE algorithm creates artificial data based on the feature space similarities between existing minority instances. Specifically, for minority class dataset C_{min}, consider the K nearest neighbors for each instance $x_i \in C_{min}$. To create a synthetic sample, randomly select one of the K nearest neighbors, then multiply the corresponding feature vector difference with a random number between [0,1], and finally, add this vector to the x_i:

$$x_{new} = x_i + (x_i^k - x_i) \times \delta \tag{1}$$

where x_i^k is one of the K-nearest neighbors for x_i, and $\delta \in [0, 1]$ is a random number. Therefore, the resulting synthetic instance is a point along the line segment joining x_i under consideration and the randomly selected K nearest neighbor x_i^k.

Fig. 3 shows an example of the SMOTE procedure.

3.2. ELM

Extreme learning machine (ELM) is a competitive machine learning technique, which is simple in theory and fast in implementation. In contrast with those conventional training algorithms, ELM is generally more robust as it places emphasis on achieving both the small norm of output weights and the least training errors.

ELM can achieve better generalization performance than other conventional classification algorithms at an extremely fast learning speed. Besides, ELM is also less sensitive to user-specified

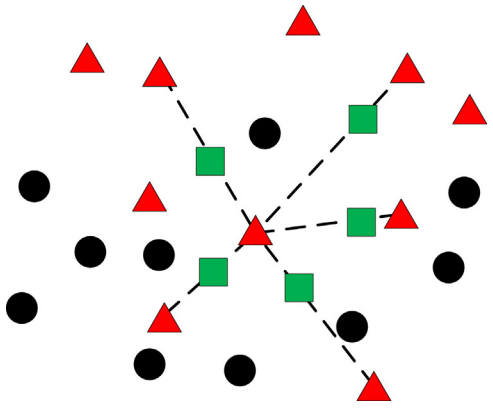


Fig. 3. An example of SMOTE algorithm. The black circle, red triangle and green rectangle point shapes represent the original majority data, original minority data, and the generated synthetic minority data, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parameters, and can be deployed faster and more conveniently. According to the ELM theory (Huang et al., 2006, 2015) ELM aims to simultaneously minimize the training errors and the norm of output weights.

This objective function can be expressed as follows: Let x_i be a training instance, and the training set is denoted as $\boldsymbol{X} = (x_1, \ldots, x_N) \in \mathbb{R}^{D\times N}$, where N and D are the number of training instances and dimensionality of x_i, respectively. Let $Y = (y_1, \ldots, y_N)^T$ be the label vector with $y_i \in \{+1, -1\}$ being the class label of the ith training instance.

The primal objective function of ELM is:

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \frac{1}{2}\left\|\boldsymbol{\beta}\right\|^2 + C\sum_{i=1}^{N}\left\|\xi_i\right\|^2 \\ \text{s.t.} \quad & h(x_i)\boldsymbol{\beta} = t_i^T - \xi_i^T \qquad \forall i \end{aligned} \tag{2}$$

where $\boldsymbol{w}$ is the normal of the separating hyperplane, b is the bias term, $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_N]$ is the vector of slack variables, and C is the regularization parameter.

The least square solution of the output weight β is analytically calculated

$$\beta = H^T\left(\frac{1}{C} + HH^T\right)Y \tag{3}$$

where H is the hidden layer output matrix, and HH^T is also called ELM kernel matrix $\Omega = HH^T : \Omega_{ij} = h(x_i)h(x_j) = K(x_i, x_j)$.

Then, the output function of the ELM learning algorithm is

$$f(x) = h(x)H^T\left(\frac{1}{C} + \Omega\right)^{-1}Y = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_n) \end{bmatrix}\left(\frac{1}{C} + \Omega\right)^{-1}Y \tag{4}$$

where $K(x_i, x_j)$ is the kernel function. In our work, the Radial Basis Function (RBF kernel) is chosen as the kernel function. The optimal C and the kernel function parameter γ can be obtained by cross-validation.

4. Proposed method: measure oriented balanced extreme learning ensemble classifier

4.1. Adaptive SMOTE, ASMOTE

SMOTE has been shown to be very successful in dealing with imbalance data. In SMOTE, all instances from the minority class are considered equally and uniformly. However, the importance and confidence of each instance are different for further learning classifiers, uniform over-sampling ratio for all instances from the minority class can generate redundant instances if the seed instances are noisy. Experimal results have empirically demonstrated that the samples closer to the decision boundary are given higher weights than others for neural network classifiers (Chen et al., 2010).

We utilize an adaptive scheme to assign weights to the minority class instead of uniform sampling weight. The adaptive scheme can automatically decide the number of synthetic samples that need to be generated for each minority instance. We apply the ideas of the adaptive weighting (He et al., 2008) and probability function estimating (Cao et al., 2014) on SMOTE to improve the sample selection scheme.

Step 1: Calculate the oversampling weight according to the local distribution of each minority instance.

Gaussian mixture models (GMM) are generative probabilistic models of several Gaussian distributions for density estimation in machine learning applications. A Gaussian mixture can be constructed to acceptably approximate any given density. After estimating the probability distribution from the training positive class, the probability of each instance x_i is calculated and normalized:

$$\hat{p}_i = p_i / \sum_{j=1}^{N_{\min}} p_j \tag{5}$$

where N_{min} is the size of the positive class.

Step 2: Calculate the number of synthetic data for each minority instance.

Then, the amount of new instances for each instance x_i is obtained according to:

$$n_i = N_{\min} \times \hat{p}_i \tag{6}$$

Step 3: Apply SMOTE with adaptive over-sampling weight.

Generate the synthetic data example according to Eq. (1) based on each positive class instance with the adaptive number of synthetic data examples.

4.2. Integration of adaptive SMOTE and ensemble

Ensembles are designed to increase the accuracy of a single classifier by training several different classifiers and combining their decisions to output a single class label. In this paper, we focus on data variation-based ensembles, which consist in the manipulation of the training examples in such a way that each classifier is trained with a different training set. Particularly, Boosting (Freund and Schapire, 1996), Bagging (Breiman, 1996) and Random subspace (Ho, 1998) are the most popular ensemble learning methods in the literature. We combined the different ensemble learning methods with over-sampling algorithm for class imbalance learning. The different mechanism in the three ensemble framework determine the chance of each minority example for generating the synthetic instances.

4.2.1. Integration of adaptive SMOTE and AdaBoost, ASOBoost

AdaBoost (Freund and Schapire, 1996) is the most well-known algorithm in the Boosting family. It builds classifiers sequentially with subsequent classifiers focusing on training examples that are misclassified by earlier ones. It builds base learners sequentially and emphasizes the hardest examples.

ASOBoost reduces the bias inherent in the learning procedure due to the class imbalance, and increases the sampling weights for the minority class by introducing adaptive SMOTE in each round of boosting. Algorithm 1 presents the pseudo-code of ASOBoost.

Algorithm 1. ASMOTE-boost

Require:
Training Dataset, D_{train}
Ensemble size, T
Over-sampling ratio, R

Ensure:
Ensemble classifier, H_f

1: Initialize the distribution D_1 over the D_{train}, such that $D_1(i) = \frac{1}{|D_{train}|}$.
2: **for** $t = 1$ to T **do**
3: Modify distribution D_t by creating synthetic examples from minority class using the adaptive SMOTE algorithm according to R
4: Train a ELM classifier h_t using the balanced distribution D_t
5: $e_t = P_{D_t}[h_t(x) \neq y]$, and $\alpha_t = 1/2 \log \frac{1-e_t}{e_t}$
6: Update D_t: $D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_t h_t(x_i))}{Z_t}$
7: **end for**
8: The final classifier is: $H_f(x) = \text{sign}(\sum_t \alpha_t h_t(x))$

The proposed ASOBoost algorithm proceeds in a series of T rounds. In every round, the distribution D_t is modified by creating synthetic examples from the minority class using the adaptive SMOTE algorithm, then a ELM classifier h_t is trained on a different

balanced dataset which emphasizes particular training instances. In such a manner, these methods alter and bias the weight distribution used to train the next classifier toward the minority class in every iteration. At the end, the different hypotheses are combined into a final hypothesis H_f.

4.2.2. Integration of adaptive SMOTE and Bagging, ASOBagging

The Bagging algorithm (Breiman, 1996), the acronym of Bootstrap AGGregatING, creates diversity by bootstrapping different training subsets. The ASOBagging samples several subsets independently from the whole training set. For each subset, the adaptive SMOTE process is carried out to balance the class distribution. Then, these subsets are used to train classifiers separately, and combine the trained classifiers with relatively balanced training subsets. The final ensemble of classifiers are generated from multiple oversampled training subsets. The final prediction is uniformly voted by those sub-classifiers. Hence, diversity is obtained with the resampling procedure by the usage of different data subsets. A pseudo code of ASOBagging construction is shown in Algorithm 2.

Algorithm 2. ASOBagging

Require:
Training Dataset, D_{train}
Ratio of bootstrap samples, R_s
Ensemble size, T
Over-sampling ratio, R
Ensure:
Ensemble classifier, H_f
1: **for** $t = 1$ to T **do**
2: A bootstrap sample B_t selected with replacement from D_{train} with R_s
3: Creating synthetic examples from minority class in data subset B_t om using the adaptive SMOTE algorithm according to R
4: Train a ELM classifier h_t on the balanced training subset
5: $H_f = H_f \cup h_t$
6: **end for**

4.2.3. Integration of adaptive SMOTE and RSM, ASORSM

The use of different spaces for ensemble construction has been extensively explained in recent research. Ho showed that the random subspace method was able to improve the generalization error (Ho, 1998). In the random subspace method, individual classifiers are built by randomly projecting the original data into feature subspaces and training a proper base-learner on these subspaces to capture possible patterns that are informative on classification. The ASORSM algorithm is described in Algorithm 3.

Algorithm 3. ASORSM

Require:
Training Dataset, D_{train}
Ratio of feature subspace, R_f
Ensemble size, T
Over-sampling ratio, R
Ensure:
Ensemble classifier, H_f
1: **for** $t = 1$ to T **do**
2: Select an random subspace with R_f from D_{train}, and construct a new data subset DR_t
3: Creating synthetic examples from minority class in DR_t with the subspace using the adaptive SMOTE algorithm according to R
4: Train a ELM classifier h_t on the balanced training subset
5: $H_f = H_f \cup h_t$
6: **end for**

4.2.4. Measure oriented extreme learning machine

Evaluation measures play a crucial role in both assessing the classification performance and guiding the classifier building a model. For imbalanced datasets, the evaluation metric should take into account the imbalance. The average accuracy is not an appropriate evaluation metric for the case of imbalanced data (Cao et al., 2013a). We used the G-mean and AUC to evaluate the cost sensitive classifiers. These two different evaluations reflect different aspect of the classifier. The AUC concerns the ranking ability more and the G-mean is concerned by the two accuracies of both classes at the same time.

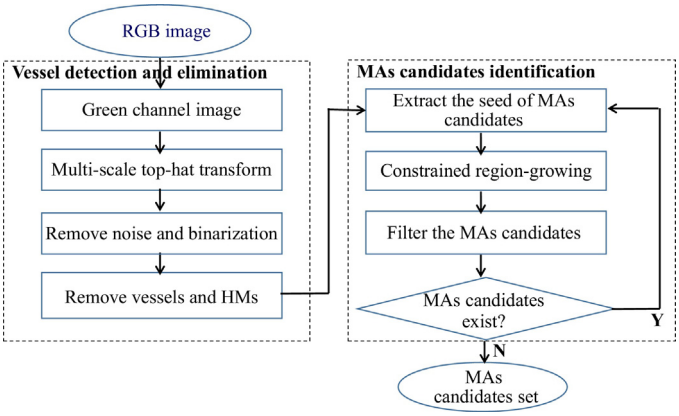


Fig. 4. Flow diagram for the MA candidates detection.

$$G\text{-}mean = \sqrt{Sensitivity \times Specificity} \tag{7}$$

where

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

The G-mean is the geometric mean of specificity and sensitivity, which is commonly utilized when performance of both classes is concerned and expected to be high simultaneously. It is a good indicator on overall performance, and has been used by several researchers for evaluating classifiers on imbalanced datasets.

ROC analysis (abbr. of receiver operating characteristic) has been introduced to evaluate machine learning algorithm. ROC curves measure the separating ability of a classier between two classes. It depicts all possible trade-offs between TP rate and FP rate. Closely related to ROC, AUC represents a ROC curve as a single scalar value by estimating the area under the curve, varying between 0 and 1. The AUC measures the performance of ranking a randomly chosen positive example higher than a randomly chosen negative example. In this case, it represents the performance of ranking an instance from the minority class higher than instances in the majority class.

In this article, the purpose of our ELM based classifier is to get a best AUC or G-mean evaluation metric. We train the cost sensitive learning using performance measures as the objective functions directly. Through the optimization of the regularization C and kernel parameter γ of ELM classifier with measure based objective functions, we can discover the best intrinsic parameters of ELM in terms of the different evaluation.

5. Experimental study

5.1. Initial MA detection

The database used in this work is the e-ophtha,[1] which is a color fundus image database provided by the Assistance Publique

[1] http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html.

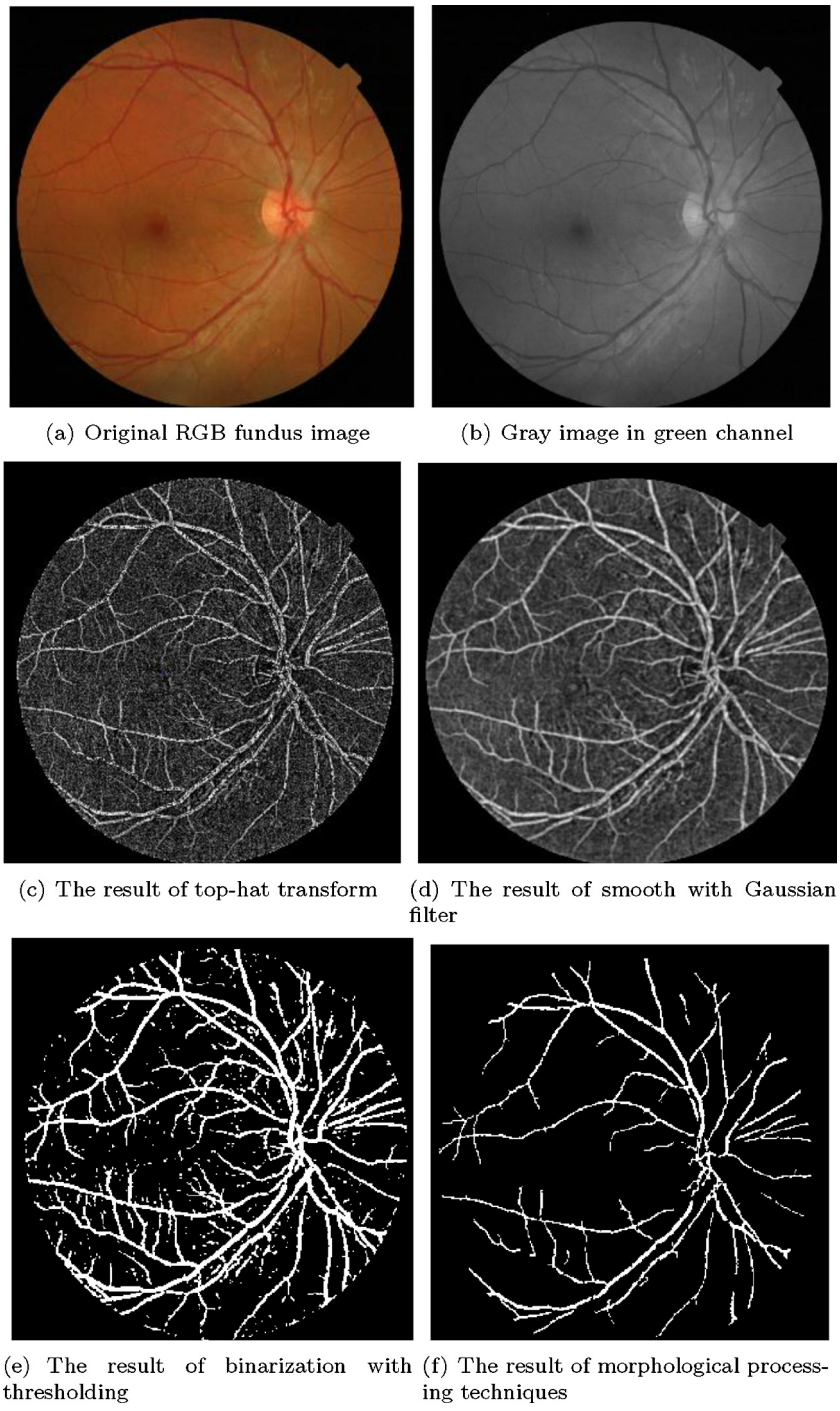


Fig. 5. Iterative procedure detection of MA candidates.

Hôpitaux de Paris. It contains 148 images with microaneurysms or small hemorrhages. The e-ophtha is a database of color fundus images especially designed for scientific research in Diabetic Retinopathy. They were captured within the OPHDIAT telemedical network for DR screening. In this database, ophthalmologists drew outlines for microaneurysms or small hemorrhages.

Fig. 4 shows the flow diagram for the MA candidate detection. MA size estimation is the first step of the algorithm, in which values for several important variables used in the subsequent algorithm will be calculated. The MA area can be estimated by the field of view (FOV) of the camera, image resolution and the diameter of the average human MA. The imaging area of MA ($A_M^i A$) of each image can be computed as Eq. (10):

$$\frac{A_{\mathrm{FOV}}}{A_{\mathrm{FOV}}^i} = \frac{A_{\mathrm{MA}}}{A_{\mathrm{MA}}^i} \tag{10}$$

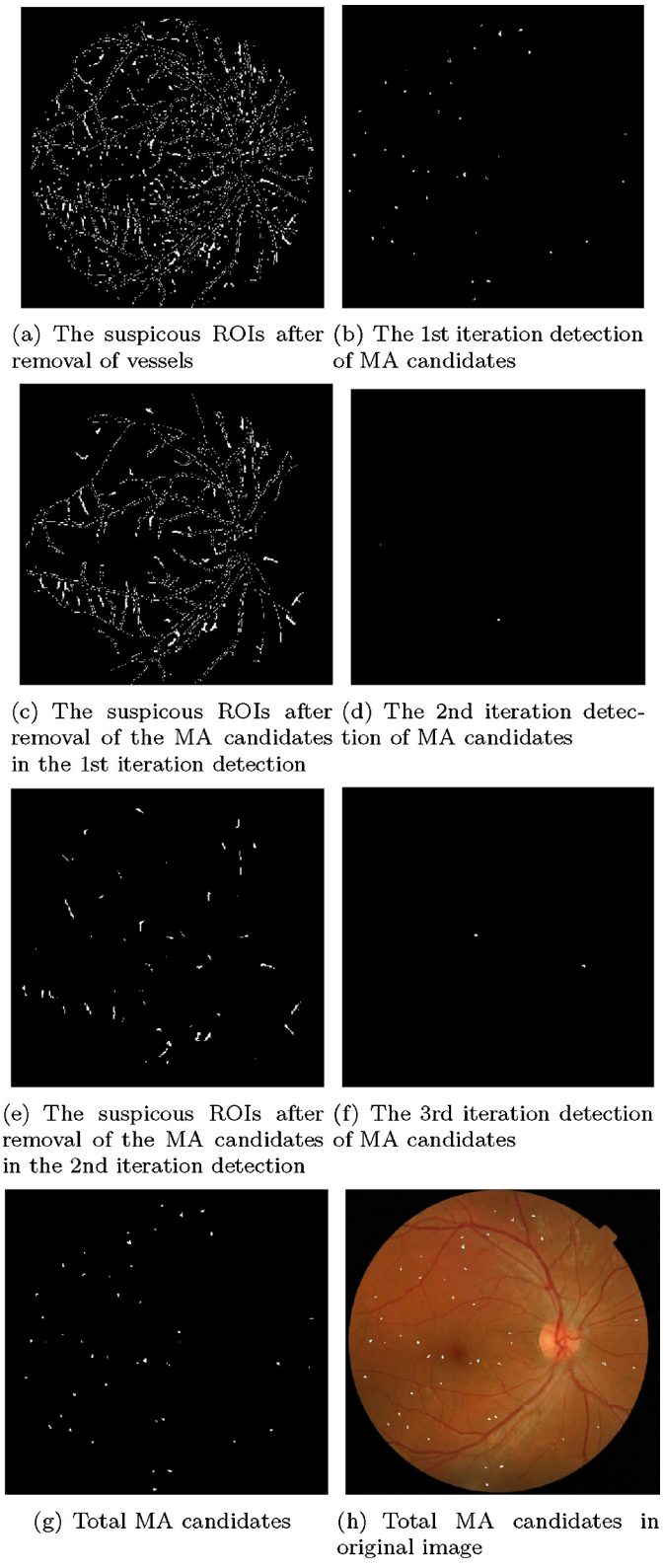
where $A_{\mathrm{MA}} = \pi(D_{\mathrm{MA}}/2)^2$, and D_{MA} is the average diameter of MA. The diameter of the average human OD has been reported to be approximately 0.0125 mm.

Fig. 5 shows the overall flow chart of the proposed CAD design for automated MA detection. Fig. 5(a) is the original RGB image. The

green channel of the original image in RGB color space is used since red lesions (MAs and blood vessels) have the highest contrast with the background in this color space. The grey image in the green channel of the original RGB image is shown in Fig. 5(b). The most important stage is to extract and remove vessel; since MAs and vessels both appear in the reddish color. The aim of the removal of vessel is to reduce the number of false positives caused by the similar appearance of vessel segments and microaneurysms. The morphological top hat transform is the most common morphological operation used to detect MAs. This is accomplished by removing all vessel objects and leaving only MA candidates. In our work, multi-scale top-hat transform is used to suppress large connected dark regions and vessels from further processing. The transformed result is shown in Fig. 5(c). The selected structural element in top-hat transform is a disc whose radius increases for each iteration and the scales to be considered for the top-hat transform will be in the range $[W, 2W]$, where W is the largest vessels width inside the FOV, and is approximately 15% of the optic disc (OD) diameter. Afterwards, a Gaussian filter with 7*7 arithmetic kernel is employed to remove the noise in Fig. 5(d). We then binarize the enhanced images with a fixed, empirically selected value by which the area approximately be the 20% of the whole FOV. The binarization image is shown in Fig. 5(e). The binarization image is then processed by morphological processing to extract the main vessels and Hemorrhages in Fig. 5(f).

An iterative procedure of MA candidates is conducted based on the result of detection blood vessels. Since MAs are of small size, the detection procedure is conducted on all suspicious (regions of interest) ROIs in Fig. 6(a), which is generated by subtracting from Fig. 5(e) and (f). The difference image removes the majority vessels and big hemorrhages. A connected-component labeling technique was then employed for identifying all of the isolated objects. The labeling algorithm identified many small objects that were mainly due to noise and other small non-MAs structures. Therefore, objects with an effective area smaller than A^i_{MA} were considered to be non-MAs and were eliminated hereafter. The remaining candidate microaneurysms do not represent the true microaneurysms size; therefore, the region of each initial MAs candidate segmented was employed as a seed region for a region growing in the original grey images. Moreover, we remove the ROIs which does not meet the requirements of area and compactness. The 1st iteration of detection of MA candidates is shown in Fig. 6(b). The 2nd (Fig. 6(c) and (d)) and 3rd (Fig. 6(e) and (f)) iteration results of detection of MAs are shown until no new MA candidates appear. The final set of MA candidates (Fig. 6(g) and (h)) are the sum of MA candidates generated by iteration of detection.

We obtained the appropriate candidate microaneurysm samples objectively using the detection algorithm, which successfully identifies 865 true MAs as positive class and 4690 non-MAs as negative class from all the cases with MAs in the database; the class imbalance ratio is 1:5. The imbalance level is not extremely high, but the misclassification costs of each class are very different. The misclassification of true MAs will influence the diagnosis and grading of DR, resulting in delayed or inappropriate treatment.



(a) The suspicous ROIs after removal of vessels

(b) The 1st iteration detection of MA candidates

(c) The suspicous ROIs after removal of the MA candidates in the 1st iteration detection

(d) The 2nd iteration detection of MA candidates

(e) The suspicous ROIs after removal of the MA candidates in the 2nd iteration detection

(f) The 3rd iteration detection of MA candidates

(g) Total MA candidates

(h) Total MA candidates in original image

Fig. 6. Iterative procedure detection of MA candidates.

5.2. Feature extraction

Feature extraction plays an important role in classification of suspicious MAs. However, there is not a single outstanding feature that can discriminate the MAs from non-MAs completely. This is due to the fact that the MAs vary enormously in volume, shape, and appearance, and the sources of false positives are different. The majority of false positives are mainly caused by blood vessels and other normal anatomic structures. Some of the false positives can be easily distinguished from true MAs, however, a large portion of them are difficult to distinguish. Therefore, for getting a high classification accuracy in candidate MAs classification, we should extract more features from many aspects, such as intensity, shape and gradient distribution. Our feature extraction process generated 26 image features from region of interest (ROI) for each potential MA object. Using these features, we construct the input space for

Table 1
Description of features for classification.

Notation	Name	Description
f_{1-4}	RGB_{mean}, RGB_{std}, RGB_{min}, RGB_{max}	The mean,standard deviation,minimum, maximum of intensity value from the RGB image
f_{5-8}	$Green_{mean}$, $Green_{std}$, $Green_{min}$, $Green_{max}$	The mean,standard deviation,minimum,maximum of intensity value from the green channel
f_{9-12}	Red_{mean}, Red_{std}, Red_{min}, Red_{max}	The mean,standard deviation,minimum,maximum of intensity value from the red channel
f_{13-16}	$CIElab_{mean}$, $CIElab_{std}$, $CIElab_{min}$, $CIElab_{max}$	The mean,standard deviation,minimum,maximum of intensity value in CIElab color space
f_{17-20}	$Grad_{mean}$, $Grad_{std}$, $Grad_{min}$, $Grad_{max}$	The mean,standard deviation,minimum,maximum of the gradient image
f_{21}	Area	The sum of pixels in possible candidate region
f_{22}	Perimeter	The count of boundary pixels
f_{23}	AspectRatio	The ratio of major axis length to minor axis length of the candidate region
f_{24}	Solidity	The ratio of the region area and the minimum bounding rectangle area of the region
f_{25}	Circularity	$Circularity = P^2/4 * Pi * A$,where A and P are the area and perimeter of the candidate region
f_{26}	Entropy	Calculated by including all pixels in bounding square around candidate region

our classifiers. This section gives a brief introduction to the features we have collected for analysis and selection.

Table 1

5.3. Evaluating the effectiveness of our proposed method

The candidate extraction step results in a set of candidate objects. The purpose of the candidate classification system is to classify each of these objects as either a MA or a non-MA (normal region). After obtaining the candidate MA instances, we calculate the features for each MA candidate from intensity, shape and gradient aspects. Then, we reduce the false positive MA instances with our proposed classification model. All the experiments are conducted by 10-fold stratified cross-validation.

5.3.1. Experiment I: The effect of data imbalance on classification performance of ELM

In this paper, we investigate the effect of data imbalance on the performance of the classification of MA candidates. The performance of three different measure optimized ELM are shown in the Table 2. The highest score among the methods is shown in bold.

The results in Table 2 show that the phenomenon of class imbalanced distribution in MA candidate data affects the classification performance of ELM. On imbalanced data, the decision boundary of ELM tends to be pushed towards the region of the minority class. The ACC-ELM, GM-ELM and AUC-ELM indicate different measures (accuracy, G-mean and AUC) for optimization of intrinsic parameters of ELM. From the results in Table 2, we found that simultaneously optimizing the parameters of regularization and kernel function with imbalanced data metrics (G-mean and AUC) help ELM learn on the imbalanced MA candidates data set. However, they still produce an undesirable model that is biased towards the majority class and has a low performance on the minority class. Moreover, we found that G-mean measure oriented ELM (GM-ELM) obtained a better performance. We believe that employing the G-mean evaluation measure as optimization objective could lead to more generalized performances.

5.3.2. Experiment II: The comparison between different over-sampling algorithms

In this experiment, we evaluate the effectiveness of our proposed adaptive SMOTE algorithm.

Based on the conclusion from the first Experiment, we use the measure of G-mean as the optimization objective of ELM. We compared our over-sampling and the other state-of-the-art re-sampling algorithm, such as the random under-sampling (RUS), random over-sampling (ROS), SMOTE over-sampling and Tomek link. The different re-sampling methods correct this skewness by over-sampling or under-sampling. For the random under-sampling (RUS) algorithm, the majority class is under-sampled by randomly removing samples from the majority class until it has the same number of instances as the minority class. The random over-sampling (ROS) method balances the ratio of the classes by copying

Table 2
Experimental results between all the ELM methods with different measure as optimization.

Method	ACC	Sensitivity	Specificity	G-mean	AUC
ACC-ELM	0.870 ± 0.015	0.264 ± 0.118	0.993 ± 0.005	0.501 ± 0.119	0.946 ± 0.019
GM-ELM	**0.888 ± 0.017**	**0.352 ± 0.099**	**0.997 ± 0.002**	**0.588 ± 0.086**	**0.953 ± 0.015**
AUC-ELM	0.882 ± 0.029	0.320 ± 0.180	0.994 ± 0.003	0.546 ± 0.158	0.947 ± 0.008

Table 3
The comparison among different re-sampling methods.

Method	ACC	Sensitivity	Specificity	G-mean	AUC
ROS	0.915 ± 0.024	0.762 ± 0.074	0.946 ± 0.018	0.849 ± 0.047	0.942 ± 0.017
RUS	0.903 ± 0.036	0.718 ± 0.083	0.942 ± 0.029	0.822 ± 0.059	0.935 ± 0.030
SMOTE	**0.916 ± 0.023**	0.720 ± 0.085	0.964 ± 0.017	0.832 ± 0.051	0.942 ± 0.025
Tomek link	0.891 ± 0.024	0.378 ± 0.158	**0.995 ± 0.001**	0.601 ± 0.134	0.927 ± 0.006
ASMOTE	0.895 ± 0.017	**0.908 ± 0.024**	0.854 ± 0.017	**0.880 ± 0.019**	**0.946 ± 0.022**

Table 4
The G-mean of different re-sampling methods in terms of varying multiple over-sampling ratio.

Method	100%	200%	300%	400%	500%
ROS	0.712 ± 0.101	0.802 ± 0.083	0.830 ± 0.046	0.849 ± 0.047	0.837 ± 0.019
SMOTE	**0.747 ± 0.063**	0.772 ± 0.044	0.804 ± 0.060	0.832 ± 0.051	0.830 ± 0.054
ASMOTE	0.744 ± 0.119	**0.846 ± 0.025**	**0.873 ± 0.023**	**0.881 ± 0.019**	**0.877 ± 0.019**

Table 5
The AUC of different re-sampling methods in terms of varying multiple over-sampling ratio.

Method	100%	200%	300%	400%	500%
ROS	0.942 ± 0.026	0.945 ± 0.025	**0.947 ± 0.018**	0.942 ± 0.017	0.932 ± 0.034
SMOTE	**0.949 ± 0.018**	0.941 ± 0.018	0.938 ± 0.026	0.942 ± 0.025	0.935 ± 0.026
ASMOTE	0.942 ± 0.029	**0.947 ± 0.020**	0.941 ± 0.024	**0.946 ± 0.022**	**0.942 ± 0.023**

Table 6
The G-mean of different ensemble based over-sampling in terms of varying multiple amount of component classifier.

Method	10%	30%	50%
ASOBoost	**0.889 ± 0.015**	**0.921 ± 0.009**	**0.890 ± 0.019**
ASOBagging	0.869 ± 0.031	0.894 ± 0.015	0.857 ± 0.003
ASORSM	0.878 ± 0.017	0.885 ± 0.023	0.869 ± 0.008

instances from the minority class randomly until the two classes have the same number of instances. Both SMOTE and adaptive SMOTE over-sample the minority class into a completely balanced training set. Tomek links (Tomek, 1976) is a data cleaning technique to filter the overlapping data points. If two data examples from different classes are the 1 nearest neighbors to each other, they form a Tomek Link. In the experiment, Tomek Link is employed to remove the instances from the majority class in a Tomek Link.

All the re-sampling methods improve the classification of ELM on the imbalanced data. From Table 3, it is apparent that adaptive SMOTE (ASMOTE) achieved higher G-mean and AUC values than the other contender methods, which demonstrates the appropriateness of weighting scheme of over-sampling for imbalanced data learning.

In adaptive SMOTE, the optimal over-sampling ratio may be unknown, and the parameter plays a vital role for the performance of over-sampling on the imbalanced data learning. It is desirable for an over-sampling method to be robust with respect to the different over-sampling ratio. We systematically analyze various sampling techniques by examining the effectiveness of different rates. The over-sampling rate R is varied in the range of [100%, 500%]. We compared our method ASMOTE with the popular over-sampling algorithm, ROS and SMOTE, and observe the influence of the re-sampling ratio in the different over-sampling on the classification performance. The comparative results are shown in Tables 4 and 5.

From Tables 4 and 5, it is apparent that ASMOTE achieved higher G-mean and AUC value than the other contender methods with respect to multiple over-sampling ratios (except 100%). Moreover, we found that the G-mean and AUC achieve the best in our case when R is 400% for the three over-sampling methods.

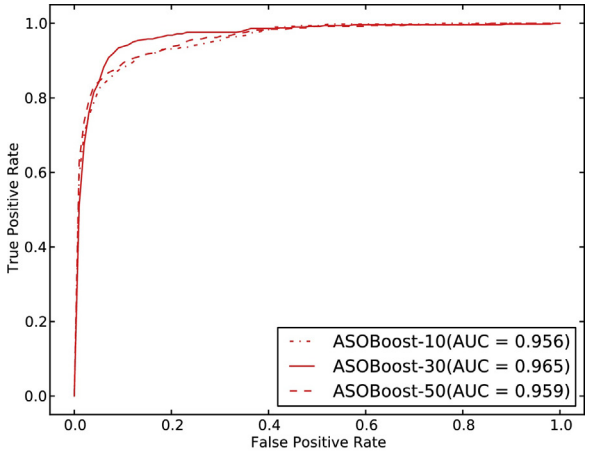
5.3.3. *Experiment III: Evaluating the effectiveness of the ensemble based adaptive over-sampling algorithm*

We evaluate the different ensemble classifiers' performance using a variety of amount of base ELM classifier. The results in terms of G-mean and AUC are shown in Tables 6 and 7.

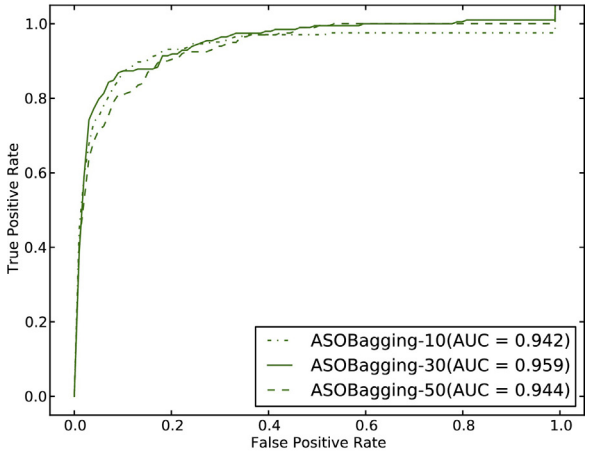
We show that ASOBoost, incorporating adaptive SMOTE with Adaboost, is more effective than the other combinations of ensemble and sampling methods. The success of the proposed algorithm lies in its Boosting mechanism forcing part of ASOBoost component

Table 7
The AUC of different ensemble based over-sampling in terms of varying multiple amount of component classifier.

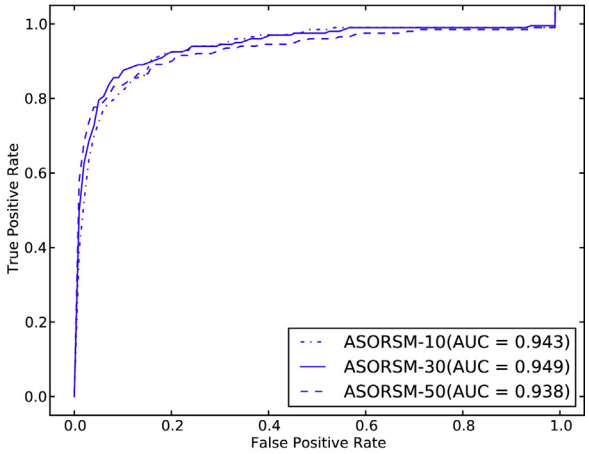
Method	10%	30%	50%
ASOBoost	**0.956 ± 0.013**	**0.965 ± 0.010**	**0.959 ± 0.010**
ASOBagging	0.942 ± 0.009	0.959 ± 0.014	0.944 ± 0.006
ASORSM	0.943 ± 0.011	0.949 ± 0.009	0.938 ± 0.016



(a) ASOBoost

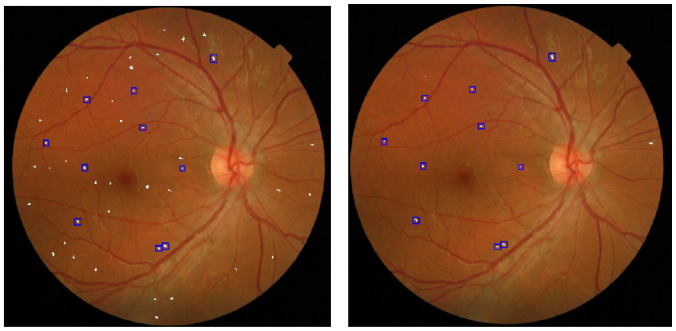


(b) ASOBagging

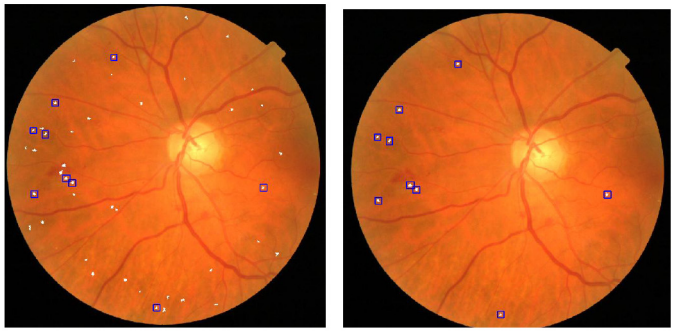


(c) ASORSM

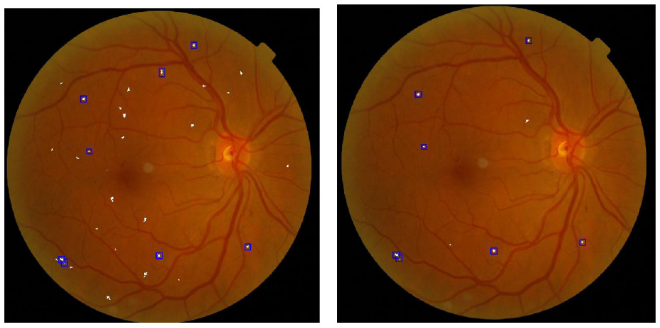
Fig. 7. ROC curves of three ensemble based adaptive SMOTE with different amount of component classifier.



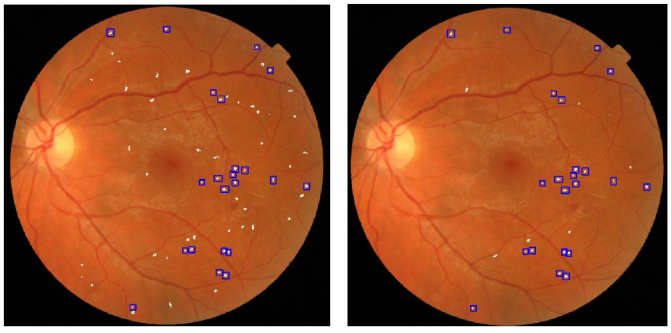
(a) Case 1: The detection result (b) The FPR result of (a)



(c) Case 2: The detection result (d) The FPR result of (c)



(e) Case 3: The detection result (f) The FPR result of (e)



(g) Case 5: The detection result (h) The FPR result of (i)

Fig. 8. The MA detection result of before and after false positive reduction with ASOBoost.

classifiers to focus on the misclassified instances in the minority class, which can prevent the minority class from being wrongly recognized as a noise of the majority class and classified into it, while ASOBagging and ASORSM are trained in parallel.

The ROC curves of the three ensemble approaches with different amount of base ELM classifier are depicted in Fig. 7. This can empirically show that the combination of Adaboost and adaptive SMOTE is a more appropriate strategy for imbalanced MA candidate data

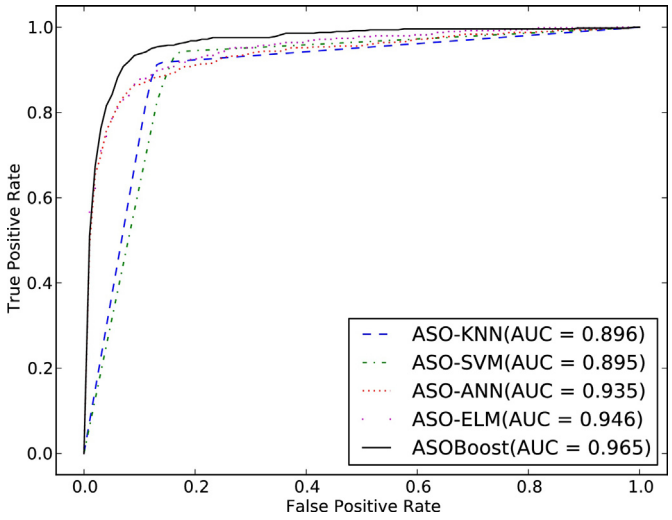


Fig. 9. ROC curves for the comparative classifiers.

learning. Fig. 8 shows some examples of classification results with our proposed ASOBoost (Fig. 9).

5.3.4. Experiment IV: The comparison with the state-of-the-art methods for false positive reduction of MA candidates

Most previous machine learning research for potential use in diabetic retinopathy is based on support vector machines (SVM) (Akram et al., 2013), artificial neural network (ANN) (Gardner et al., 1996) or *K*-nearest neighbor(KNN) (Niemeijer et al., 2005). To exhibit the performance of our proposed approaches, the comparison is conducted between our three methods and the commonly used classifier in MAs classification, such as SVM, ANN and KNN in order to validate the effectiveness of the ensemble based oversampling algorithm experimentally. Besides, we also compare our over-sampling based methods with a cost sensitive learning method. Cost-sensitive learning is one of the most important topics in machine learning and data mining, and has attracted high attention in recent years. It takes misclassification costs into account during the model construction, and does not modify the imbalanced data distribution directly. Cost sensitive ELM is proposed to target the imbalanced data problemby setting a high cost to the misclassification of a minority class sample (Zong et al., 2013). For CS-ELM, the misclassification cost matrix is generated in accordance with the class distribution (inversely proportional to the number of samples in the class). All the sizes of ensemble methods are set to 30. All the evaluation results of the compared methods are summarized in Table 8.

Experiment results (Table 8) demonstrate the benefit of ensemble-based ASMOTE in terms of AUC and G-mean, particularly ASOBoost compared to other commonly used methods for false positive reduction in MA detection. The results can confirm the advantages of our approach for false positive reduction of MA candidates. In addition, we find our oversampling based methods outperform the cost sensitive learning method. The reason is that oversampling adds new instances into the minority class, so as to magnify the distribution of the minority class and strengthen the representation of the minority class, while cost sensitive learning does not add new knowledge into minority class. Moreover, an important issue of applying the cost-sensitive learning algorithm to the imbalanced data is that the cost matrix is often unavailable for a problem domain, and it needs to be searched by some heuristic methods (Cao et al., 2013b; Nguyen et al., 2010). It is often not correct to set the cost ratio to the inverse of the imbalance size ratio (the number of majority instances divided by the number of minority instances) (Cao et al., 2013c; Nguyen et al., 2010).

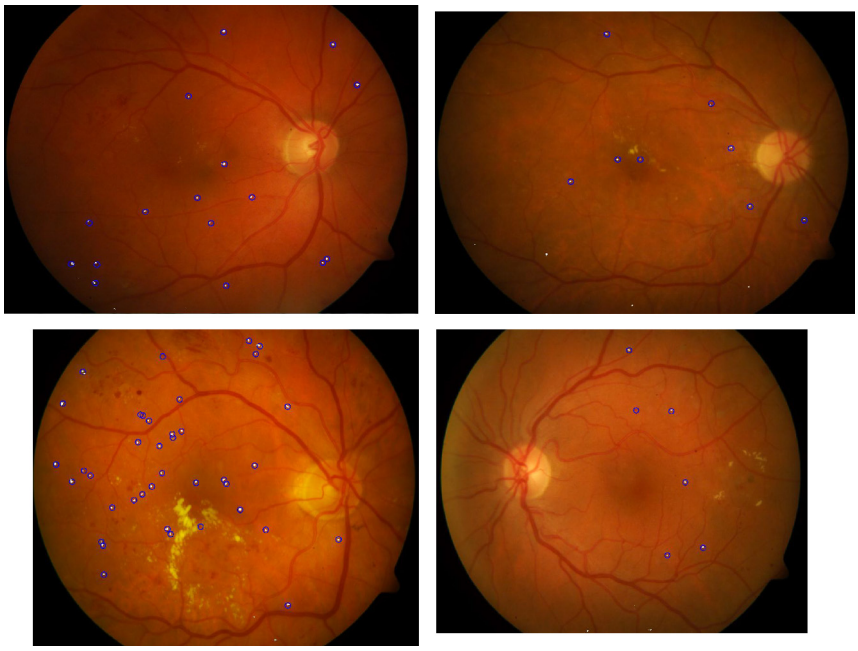


Fig. 10. Final detection results of MAs by ASOBoost on the DiaretDB1 dataset.

5.3.5. Experiment V: Evaluation on DiaretDB1 database

In order to evaluate our methods, we have evaluated the MA detection capabilities of the proposed method on two other public datasets: DiaretDB1 (DIAbetic RETinopathy DataBase)(standard diabetic retinopathy database, http://www.it.lut.fi/project/imageret/diaretdb1) (Kauppi et al., 2007) and ROC (Retinopathy Online Challenge, http://roc.healthcare.uiowa.edu) (Niemeijer et al., 2010).

DiaretDB1 (DIAbetic RETinopathy DataBase) is a database which is designed to evaluate automated lesion detection algorithms (Kauppi et al., 2007). It contains 89 retinal images with different retinal abnormalities, where 84 images contain different signs of DR and 5 represent healthy retina. Our initial MA detection method is applied on the 89 cases obtained from the DiaretDB1 database.

All the ROI data are split randomly into 90% for training and the remaining 10% for testing. The process is repeated ten times in a cross-validation procedure in order to generate unbiased results. The average results on the ten runs for each of the four classifiers are reported. Fig. 10 shows some examples of final detection results by ASOBoost. Moreover, Table 9 and Fig. 11 present the experimental results on our adaptive SMOTE based ensemble classifiers on the DiaretDB1 database.

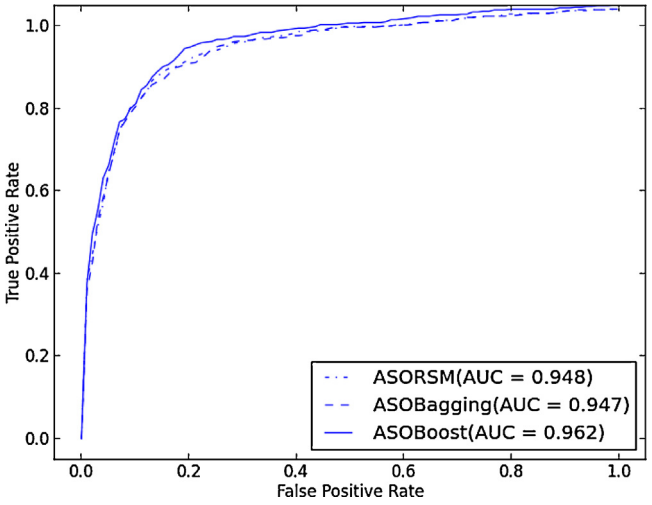


Fig. 11. ROC curves for our proposed ensemble classifiers combined with ASMOTE on the DiaretDB1 dataset.

Table 8
The comparison among different classifier combined with adaptive SMOTE.

Method	ACC	Sensitivity	Specificity	G-mean	AUC
ASOBoost	**0.918 ± 0.007**	0.924 ± 0.015	0.917 ± 0.007	**0.921 ± 0.009**	**0.965 ± 0.010**
ASOBagging	0.874 ± 0.006	0.922 ± 0.028	0.864 ± 0.017	0.894 ± 0.015	0.959 ± 0.014
ASORSM	0.886 ± 0.006	0.885 ± 0.049	0.886 ± 0.013	0.885 ± 0.023	0.949 ± 0.009
CS-ELM	0.876 ± 0.014	0.908 ± 0.029	0.870 ± 0.013	0.889 ± 0.019	0.956 ± 0.009
ASO-KNN	0.882 ± 0.006	0.916 ± 0.025	0.878 ± 0.010	0.895 ± 0.009	0.896 ± 0.010
ASO-SVM	0.864 ± 0.017	**0.942 ± 0.026**	0.848 ± 0.018	0.894 ± 0.018	0.895 ± 0.015
ASO-ANN	0.918 ± 0.011	0.828 ± 0.050	**0.936 ± 0.013**	0.880 ± 0.025	0.936 ± 0.021

Table 9
The comparison among different classifiers combined with adaptive SMOTE on DiaretDB1 dataset.

Method	ACC	Sensitivity	Specificity	G-mean	AUC
ASOBoost	**0.882 ± 0.017**	**0.961 ± 0.020**	**0.821 ± 0.027**	**0.888 ± 0.021**	**0.962 ± 0.022**
ASOBagging	0.867 ± 0.015	0.911 ± 0.023	0.811 ± 0.018	0.860 ± 0.015	0.947 ± 0.017
ASORSM	0.873 ± 0.021	0.924 ± 0.015	0.815 ± 0.029	0.868 ± 0.017	0.948 ± 0.015

Table 10
The comparison of different MA detection methods on DiaretDB1 dataset.

Authors	Method	ACC	Sensitivity	Specificity	FPI	AUC
Proposed method	ASOBoost	0.882	**0.961**	0.821	3.404	**0.962**
Proposed method	ASOBagging	0.867	0.911	0.811	3.348	0.947
Proposed method	ASORSM	0.873	0.924	0.815	**3.281**	0.948
Roberto et al. (2015b)	Bottom-hat transform combined with PCA	Not reported	0.923	0.939	Not reported	Not reported
Rahim et al. (2015)	Circular Hough transform combined with KNN	**0.909**	0.875	**1**	Not reported	Not reported
Oliveira (Oliveira et al., 2013)	Radon transform combined with SVM	0.842	0.895	0.895	Not reported	0.83

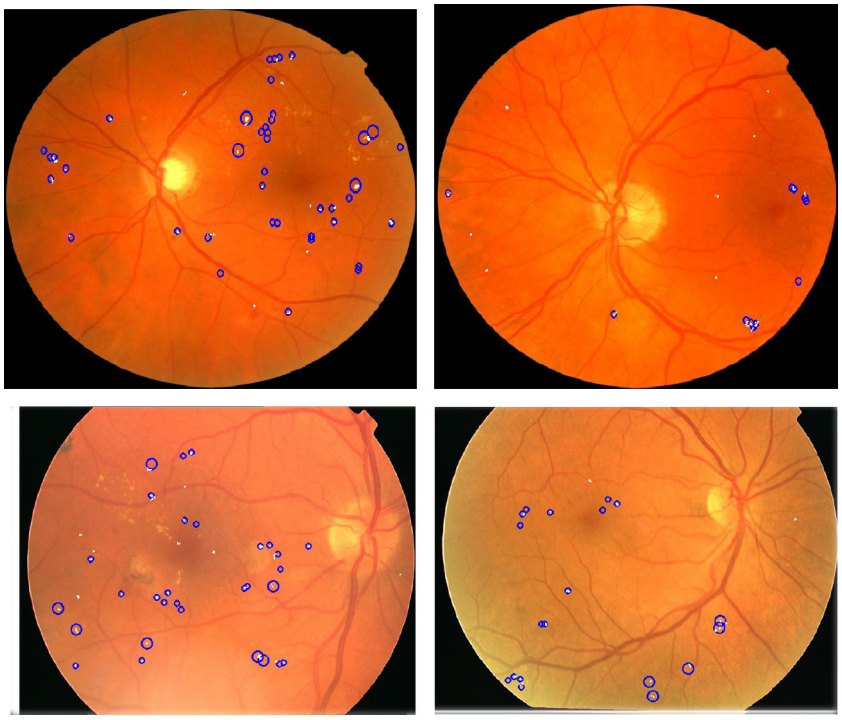


Fig. 12. Final detection results of MAs by ASOBoost on ROC dataset.

A comparisons of this method with existing MA detection algorithms is shown in Table 10. The evaluation results on the DiaretDB1 dataset show that our ASOBoost outperforms previously published methods with respect to sensitivity and AUC.

5.3.6. *Experiment V: Evaluation on ROC database*

The retinopathy online challenge (ROC) database is a worldwide competition dedicated to evaluate the accuracy of different proposed microaneurysm detection methods (Niemeijer et al., 2010). The ROC database comprises 50 training images and 50 test images. In the training images, the gold standard for MA determined by four ophthalmologists has been given. However, in the test images, the gold standard was not provided. Hence, we use the 50 labeled images to train and test with cross validation strategy. Fig. 12 shows some examples of the results of false positive reduction by ASOBoost. Table 11 shows the results of the proposed ensemble classifiers on the ROC Dataset (Our result only on the training data, since test data label is not available).

In addition, we also use the curve of ROC and the free-response receiver operating characteristic (FROC) to evaluate and investigate our methods in Figs. 13 and 14, respectively. The accuracy, sensitivity, specificity, FPI and AUC of the individual existing MA detection system on the ROC dataset are presented in Table 12. Although this comparison is not done on the same proportion of the database, our ASOBoost algorithm reaches a sensitivity of 42.5% at an average of 5.7 false positives (FPs) per image which is competitive against published methods. The best sensitivity (0.506) is obtained using double-ring filter combined with rule classifier in Mizutani et al. (2009), however its FPI is 105, which higher than 5.7 obtained by our ASOBoost.

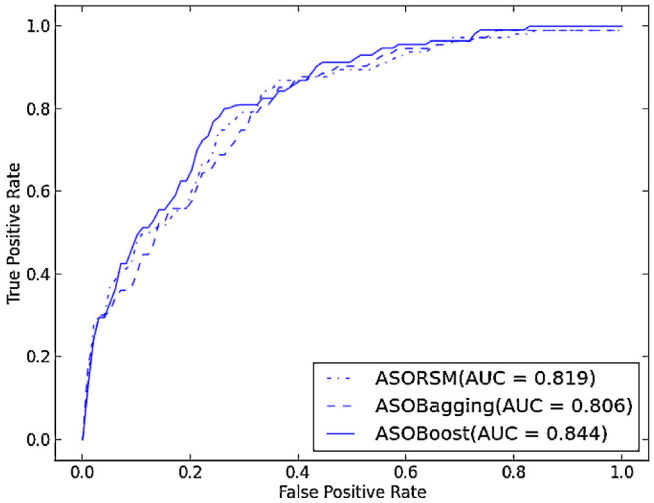


Fig. 13. ROC curves for our proposed ensemble classifiers combined with ASMOTE on ROC dataset.

Table 11
The comparison among proposed ensemble classifiers in false positive reduction.

Method	ACC	Sensitivity	Specificity	G-mean	AUC
ASOBoost	**0.758 ± 0.041**	**0.801 ± 0.024**	**0.762 ± 0.056**	**0.781 ± 0.028**	**0.844 ± 0.018**
ASOBagging	0.731 ± 0.039	0.730 ± 0.113	0.731 ± 0.077	0.726 ± 0.034	0.806 ± 0.052
ASORSM	0.743 ± 0.051	0.713 ± 0.090	0.753 ± 0.077	0.723 ± 0.043	0.819 ± 0.044

Table 12
The comparison among different classifiers with previous methods in the overall detection of MAs.

Authors	Method	Number of images	Sensitivity	FPI
Proposed method	ASOBoost	50(cross-validation)	0.425	5.700
Proposed method	ASOBagging	50(cross-validation)	0.405	**5.100**
Proposed method	ASORSM	50(cross-validation)	0.419	5.340
Adal et al. (2013)	Hessian operator combined with SVM classifier	50	0.446	35.2
Mizutani et al. (2009)	Double-ring filter combined with rule classifier	50	**0.506**	105
Mizutani et al. (2009)	Double-ring filter combined with ANN classifier	50	0.449	27
Inoue et al. (2013)	Eigenvalue analysis using a Hessian matrix	25	0.410	8
Zhang et al. (2010)	multi-scale correlation filtering combined with dynamic thresholding	50	0.33	328.3

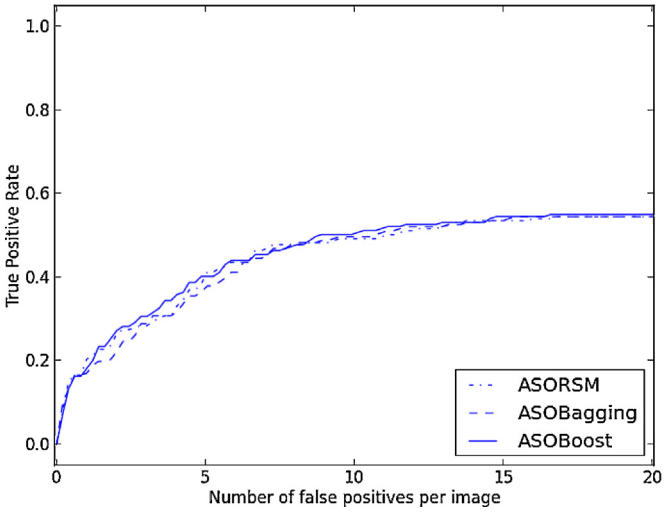


Fig. 14. FROC curves for our proposed ensemble classifiers combined with ASMOTE on ROC dataset.

6. Conclusion

The accurate detection of microaneurysms (MAs) is a critical step for early detection of diabetic retinopathy because they appear as the first sign of the disease. Classification plays an important role in the reduction of false positives. The data samples in CAD products exhibit large class imbalance, and an accurate identification of MAs (minority) from the non-MAs (majority) is a challenge. The main contribution of this work is the introduction of imbalanced data learning methods to the detection of MAs. We propose three ensemble based adaptive SMOTE algorithms for overcoming the imbalanced data problems. We compare the performance of our algorithms against the state-of-the-art approaches in the class imbalance classification and false positive reduction. Experimental results show the unique feature of our algorithm for overcoming the challenge and demonstrate a promising effectiveness. The main outcome of the study is that the ASOBoost method acquire best classification performance and generalization capability.

Acknowledgments

This research was supported by the National High Technology Research and Development Program (863 Program) of China (2015AA020106), the National Natural Science Foundation of China (61502091), the Fundamental Research Funds for the Central Universities under Grant N140403004, and the Postdoctoral Science Foundation of China 2015M570254.

References

Adal, K., Ali, S., Sidibé, D., Karnowski, T., Chaum, T., Mériaudeau, F., 2013. Automated detection of microaneurysms using robust blob descriptors. In: SPIE Medical Imaging.
Adal, K.M., Sidibé, D., Ali, S., Chaum, E., Karnowski, T.P., Mériaudeau, F., 2014. Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning. Comput. Methods Programs Biomed. 114 (1), 1–10.
Akram, M.U., Khalid, S., Khan, S.A., 2013. Identification and classification of microaneurysms for early detection of diabetic retinopathy. Pattern Recogn. 46, 107–116.
Antal, B., Hajdu, A., 2013. Improving microaneurysm detection in color fundus images by using context-aware approaches. Comput. Med. Imaging Graph. 37 (5), 403–408.
Breiman, L., 1996. Bagging predictors. Mach. Learning 24 (2), 123–140.
Cao, P., Zhao, D.Z., Zaiane, O., 2013a. An optimized cost-sensitive SVM for imbalanced data learning. In: Advances in Knowledge Discovery and Data Mining, pp. 280–292.
Cao, P., Zhao, D., Zaiane, O., 2013b. An optimized cost-sensitive SVM for imbalanced data learning. In: Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2013), pp. 280–292.
Cao, P., Zhao, D.Z., Zaiane, O., 2013c. An optimized cost-sensitive SVM for imbalanced data learning. In: Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 280–292.
Cao, P., Yang, J.Z., Li, W., Zhao, D.Z., Zaiane, O., 2014. Ensemble-based hybrid probabilistic sampling for imbalanced data learning in lung nodule. CAD Comput. Med. Imaging Graph. 38 (3), 137–150.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16 (1), 321–357.
Chawla, N.V., Japkowicz, N., Kotcz, A., 2004. Editorial: special issue on learning from imbalanced data sets. ACM SIGKDD Expl. Newslett. 6 (1), 1–6.
Chen, S., He, H., Garcia, E.A., 2010. RAMOBoost: ranked minority oversampling in boosting. IEEE Trans. Neural Netw. 21 (10), 1624–1642.
Dupas, B., Walter, T., Erginay, A., Ordonez, R., Deb-Joardar, N., Gain, P., Klein, J.C., Massin, P., 2010a. Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a computer-assisted diagnostic system for grading diabetic retinopathy. Diabetes Metab. 36 (3), 213–220.
Dupas, B., Walter, T., Erginay, A., Richard, O., Joardar, N.D., Gain, P., Klein, J.C., Massin, P., 2010b. Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a computer-assisted diagnostic system for grading diabetic retinopathy. Diabetes Metab. 36 (3), 213–220.
Erbas, T., Ertas, M., Yucel, A., Keskinaslan, A., Senocak, M., 2011. Prevalence of peripheral neuropathy and painful peripheral neuropathy in Turkish diabetic patients. J. Clin. Neurophysiol. 28, 51–55.
Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Proceedings of the International Conference on Machine Learning, pp. 148–156.
Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., et al., 2012. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. Part C 42, 463–484.

Gardner, G.G., Keating, D., Williamson, T.H., Li, Q., Elliott, A.T., 1996. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. Br. J. Ophthalmol. 803 (11), 940–944.
Giancardo, L., Meriaudeau, F., Karnowski, T.P., 2013. Validation of microaneurysm-based diabetic retinopathy screening across retina fundus datasets. In: 2013 IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS), pp. 125–130.
He, H., Garcia, E., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284.
He, H., Bai, Y., Garcia, E., Li, S., 2008. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: IEEE International Joint Conference on Neural Networks, pp. 1322–1328.
Ho, T., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20 (8), 832–844.
Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70 (1), 489–501.
Huang, G.B., Huang, G.B., Song, S., You, K., 2015. Extreme learning machine: theory and applications. Neural Netw. 61, 32–48.
Inoue, T., Hatanaka, Y., Okumura, S., Muramatsu, C., Fujita, H., 2013. Automated microaneurysm detection method based on eigenvalue analysis using Hessian matrix in retinal fundus images. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5873–5876.
Kauppi, T., Kalesnykiene, V., Kamarainen, J.K., Lensu, L., Sorri, I., Raninen, A., et al., 2007. DIARETDB1 diabetic retinopathy database and evaluation protocol. In: Medical Image Understanding and Analysis 2007.
Mizutani, A., Muramatsu, C., Hatanaka, Y., Suemori, S., Hara, T., Fujita, H., 2009. Automated microaneurysm detection method based on double ring filter in retinal fundus images. In: SPIE Medical Imaging, pp. 308–313.
Mrinal, H., 2015. Improved Microaneurysm Detection using Deep Neural Networks. arXiv:1505.04424.
Nguyen, T.N., Zeno, G., Lars, S.T., 2010. Cost-sensitive learning methods for imbalanced data. In: The 2010 International Joint Conference on Neural Networks, pp. 1–8.
Niemeijer, N., Staal, J., Abramoff, M.D., Suttorp-Schulten, M.A., Maria, S.A., van Ginneken, B., 2005. Automatic detection of red lesions in digital color fundus photographs. IEEE Trans. Med. Imaging 24 (5), 584–592.
Niemeijer, M., Van Ginneken, B., Cree, M.J., Mizutani, A., et al., 2010. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans. Med. Imaging 29, 185–195.
Oliveira, J., Minas, G., Silva, C., 2013. Automatic detection of microaneurysm based on the slant stacking. In: 2013 IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS), pp. 308–313.
Rahim, S.S., Jayne, C., Palade, V., Shuttleworth, J., 2015. Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening. Neural Comput. Appl. 44, 1–16.
Roberto, R.R., Jorge, J.M., Jonathan, H.C., Khan, S.A., Laura, U.V., 2015a. A method to assist in the diagnosis of early diabetic retinopathy: image processing applied to detection of microaneurysms in fundus images. Comput. Med. Imaging Graph. 44, 41–53.
Roberto, R.R., Jorge, M.C., Jonathan, H.C., Laura, J.U.V., 2015b. A method to assist in the diagnosis of early diabetic retinopathy: Image processing applied to detection of microaneurysms in fundus images. Comput. Med. Imaging Graph. 44, 41–53.
Tomek, I., 1976. Two modifications of CNN. IEEE Trans. Syst. Man Cybern. 6, 769–772.
Walter, T., Massin, P., Erginay, A., Richard, O., Jeulin, C., Klein, J.C., 2007. Automatic detection of microaneurysms in color fundus images. Med. Image Anal. 11 (6), 555–566.
Zhang, B., Wu, X., You, J., Li, Q., Karray, F., 2010. Detection of microaneurysms using multi-scale correlation coefficients. Pattern Recogn. 43 (6), 2237–2248.
Zhang, B.U., Karray, F., Li, Q., Zhang, L., 2012. Sparse representation classifier for microaneurysm detection and retinal blood vessel extraction. Inform. Sci. 200, 78–90.
Zong, W., Huang, G.B., Chen, Y., 2013. Weighted extreme learning machine for imbalance learning. Neurocomputing 101, 229–242.