A New Ensemble Learning Approach for Microcalcification Clusters Detection

Xinsheng Zhang
School of Management, Xi'AN University of Architecture and Technology, Xi'AN, China
Email: xinsheng.zh@hotmail.com

***Abstract*—A new microcalcification clusters (MCs) detection method in mammograms is proposed, which is based on a new ensemble learning method. In this paper, we propose a bagging with adaptive cost adjustment ensemble algorithm; and a new ensemble strategy, called boosting with relevance feedback, by embedding the relevance feedback technique into the heterogenous base learner training, and meanwhile carefully design an effectively systematical feedback scheme, which promise the preventing of overfitting. The ground truth of MCs is assumed to be known as a priori. In our algorithm, each MCs is enhanced by a well designed high-pass filter. Then the 116 dimentional image features are extracted by the feature extractor and fed to the ensemble decision model. In image feature domain, the MCs detection procedure is formulated as a supervised learning and classification problem, and the trained ensemble model is used as a classifier to decide the presence of MCs or not. Case study on microcalcification clusters detection for breast cancer diagnosis illustrates that the proposed algorithm is not only effective but also efficiency.**

***Index Terms*—feature, microcalcification clusters, bagging, bootstrap, boosting, ensemble learning**

I. Introduction

Breast cancer is the most common cancer in women. Mammogram is, at present, one of the most suitable methods for early detection of breast cancer. One of the important early signs of breast cancer is the appearance of microcalcification clusters (MCs) in mammogram. Calcifications in mammograms appears as relatively bright regions due to the higher X-ray attention coefficient (or density) of calcium as compared with normal breast tissue. Calcifications present within dense masses or superimposed by dense tissues in the process of acquisition of mammograms could present low gray-level differences or contrast with respect to their local background. On the other hand, calcifications present against a background of fat or low-density tissue would possess higher differences and contraset. Malignant calcifications tend to be numerous, clustered, small, varying in size and shape, angular, irregularly shaped, and branching in orientation. On the other hand, calcifications associated with benign diseases are generally larger, more rounded, smaller in number, more diffusely, and more homogeneous in size and shape.

Because of the importance in early breast cancer diagnosis, accurate detection of MCs has become a key problem. The detection and classification of microcalcification clusters has been extensively studied, with many authors reporting on several successful approaches to this task. A thorough review of various methods for MCs detection was made in [1]. Considered MCs detection as a classification problem, various classification methodologies have been proposed for the characterization of MCs, such as, fuzzy rule-based systems [2], [3-8] , [9-16], etc.

Most of the above methods are single models, which achieve the final decision task after training with the training samples. They can do well in some special cases, but, the result of classification accuracy is not good either when we change the features as inputs for classification sometimes. Ensemble learning techniques [17] have demonstrated powerful capacities to improve upon classification accuracy of a base learning algorithm. Ensemble methods are typically composed of multiple methods comprising different classification strategies or different classifiers with a unified objective function. The final predictions are chosen from the ensemble of methods by a learning rule. To improve the ensemble performance, in this paper a new ensemble algorithm, bagging with relative-feedback and adaptive cost (bracing), is proposed in this paper. By iteratively changing the training dataset with the relative feedback samples from the previous test procedure and the weights of the base learner, bracing is able to focus on the nearest error or true samples that the current base learning algorithm barely predicts right or wrong, on the other hand, our method attempts to pick a subset of classifiers from an existing ensemble to improve the effectiveness and efficiency of the ensemble.

To get a good performance of the available MCs detection algorithms, bracing are employed as a final decision model to distinguish the MCs from the other ROIs (region of interests). In the experiments, a database of more than 100 case clinical mammograms from DDSM is selected as test bed. Three subsets are randomly chosen from test bed, the first one is used for training, the other for validation and another for testing. Compared with other existing methods, the proposed approach

Manuscript received January 1, 2009; revised February 1, 2009; accepted February 20, 2009.
Corresponding author: xinsheng.zh@hotmail.com

doi:10.4304/jsw.4.9.1014- 1021

yields superior performance when evaluated using receiver operating characteristic (ROC) curves.

The rest of the paper is organized as follows. Image features and extraction method used in our paper is introduced in Section II. The proposed ensemble learning method with bracing is formulated in Section III. The MCs detection approach with bracing ensemble learning is proposed in section IV. Performance evaluation study and database used in the proposed approach are described in Section V, and the experimental results are given in Section VI. Finally, conclusions are drawn in Section VII.

II. FEATURE EXTRACTION

To get the best feature or combination of features and get the high classification rate for MCs classification is one of main aims of the proposed research.In our task, a set of 116 features, shown in Table I, is calculated for each suspicious area from the textural, spatial and transform domains in our research.

TABLE I.
IMAGE FEATURES USED IN OUR EXPERIMENTS

Feature groups	Type	#No.
Histogram based texture features	Mean Standard deviation Smoothness Third moment Uniformity Entropy	1 2 3 4 5 6
Multi-scale histogram features	Histograms with different number of bins 3,5,7,9	7~30(total 24)
Zernike Moments Features[18]	Zernike Moments	31~66 (total 36)
Tamura texture signatures[19]	Tamura features	67~72 (total 6)
Chebyshev transform histogram feature	Chebyshev histogram	73~104 (total 32)
Radon Transform Features	Radon features	105~116 (total 12)

Before training the classifier, we use the feature extractor discussed in Table 1 to extract MCs features in the feature domain. The 116-dimension feature vector will be calculated for each image block. When we get the image feature vector, feature normalization program should be user for normalizing the features to be real numbers in the range of 0-1. The normalization is accomplished by the following step: (1) change all the features to be positive by adding the magnitude of the largest minus value of this feature times 1.01^2, (2) divided by the maximum value of the same feature. The normalized features are used as the inputs of the proposed ensemble learning algorithm for training and classification.

(1) Intensity histogram based texture feature

A frequently used approach for texture analysis is based on statistical properties of the intensity histogram. One class of such measures is based on statistical moments. The expression for the nth moment about the mean is given by

$$\mu_n = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i) \tag{1}$$

where z_i is a random variable indicating intensity, $p(z)$ is the histogram of the intensity levels in a region, L is the number of possible intensity levels, and

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \tag{2}$$

is the mean(average)intensity. Table II lists the descriptors used in our experiments based on statistical moments and also on uniformity and entropy.

TABLE II.
DESCRIPTORS OF THE TEXTURE BASED ON THE INTENSITY HISTOGRAM OF A REGION

Moment	Formula	#No.
Mean	$m = \sum_{i=0}^{L-1} z_i p(z_i)$	1
Standard deviation	$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2}$	2
Smoothness	$R = 1 - 1/(1+\sigma^2)$	3
Third moment	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$	4
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$	5
Entropy	$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$	6

(2) Multiscale intensity histograms

We compute signatures based on "multi-scale histograms" idea. Idea of multi-scale histogram comes from the belief of a unique representation of an image through infinite series of histograms with sequentially increasing number of bins. Here we used 4 histograms with number of bins being 3,5,7,9 to calculate the image feature, so we get a 1*24 row vectors.

(3) Zernike moments

We use derived moments based on alternative complex polynomial functions, know as Zernike polynomials[18]. They form a complete orthogonal set over the interior of the unit circle $x^2 + y^2 = 1$ and are defined as

$$Z_{pq} = (p+1)/\pi \int_{x^2+y^2 \le 1} f(x,y) V(\rho,\theta) dxdy, \tag{3}$$

$$V_{pq}(x,y) = V_{pq}(\rho,\theta) = R_{pq}(\rho)\exp(jq\theta), \tag{4}$$

$$R_{pq}(\rho) = \sum_{s=0}^{(p-|q|)/2} \frac{(-1)^s [(p-s)!] \rho^{p-2s}}{s!\left(\frac{p+|q|}{2} - s\right)!\left(\frac{p-|q|}{2} - s\right)!}, \tag{5}$$

where p is a nonnegative integer, q is an integer subject to the constraint $p - |q| = even$ and $|q| \le p$, $\rho = \sqrt{x^2 + y^2}$ is the radius from (x, y) to the image centroid, $\theta = \tan^{-1}(y/x)$ is the angle between ρ and x-axis. The Zernike moment Z_{pq} is order p with repetition q. For a digital image, the respective Zernike moments are computed as

$$Z_{pq} = (p+1)/\pi \sum_i f(x_i, y_i) V(\rho,\theta) dxdy, x^2 + y^2 \le 1, \tag{6}$$

where i runs over all the image pixels. Zernike moments are used as the feature extractor where by the order is varied to achieve the optimal classification performance.

(4) Combined first four moment features

Signatures on the basis of first four moments (also known as mean, std, skewness, kurtosis) for data generated by vertical, horizontal, diagonal and alternative diagonal 'combs'. Each column of the comb results in 4 scalars [mean, std, skewness, kurtosis], we have as many of those [...] as 20. So, 20 go to a 3-bin histogram, producing 1x48 vectors.

(5) Tamura texture signatures

Tamura et al. [19] took the approach of devising texture features that correspond to human visual perception. Six textural features: coarseness, contrast, directionality, line-likeness, regularity and roughness, are defined for the image feature for object recognition.

(6) Transform domains feature

We computes signatures (32 bins histogram) from coefficients of 2D Chebyshev transform (10th order is used in our experiments) [20]. Also we used signatures based on the Radon transform as a kind of image features. Radon transform is the projection of the image intensity along a radial line (at a specified orientation angle), total 4 orientations are taken. Transformation n/2 vectors (for each rotation), go through 3-bin histogram and convolve into 1x12 vectors.

III. BRACING ENSEMBLE LEARNING

The important issue of ensemble learning is how to create a relatively small sized ensemble with a good bias and variance trade-off. Currently, many authors use the constructive building technology, such as bagging and boosting. Although the existing ensemble methods have proved to be powerful, each has its own drawbacks. To get rid of the drawbacks of previous methods, a new ensemble algorithm, bagging with relative-feedback and adaptive cost (bracing), is proposed in this paper. Suppose we have a set of classifiers $\{L_k \mid k = 1,..., N\}$ for a two-class problem and a dataset S . S is composed of two subsets, namely A and B , where each sample in A and B belongs to class +1 and -1, respectively. Each classifier L_k will give each sample X_i in the training set a classification label y_i^k . If y_i^k is classified as class +1 by L_k , then $y_i^k = 1$ and $y_i^k = -1$ otherwise. The final classification decision for the two-category case by the ensemble is decided by looking at the sign of the weighted average of the output of those single base learning algorithm. If this weighted average on a test point is positive, this point is classified as class +1 and vise versa. Thus the bracing procedure is shown in Fig.2 and Fig. 3.

Algorithm: bracing(S , L , t , $fraction$, $threshold$)

Input: $S = \{(X_1, y_1),...,(X_m, y_m)\}$, a dataset with m pair of examples and labels;
$L = \{L_k \mid k = 1,..., K\}$, base learning algorithms;
Integer t specifying the number of maximum training rounds;
$fraction$, fraction of S for training, and 1-fraction for validation;
$threshold$, threshold of the expected classification accuracy.

Process: Construct a bracing consisting K base learners.

$w_i \leftarrow 1/K, i = 1...K$

for $i \in \{1,..., K\}$ do

$accuracy \leftarrow 0$

$(S_i^1, S_i^2) \leftarrow \text{randselect}(S, fraction)$

for $j \in \{1,..., t\}$ do

if($accuracy \geq threshold$) break;

$M_{L_i} \leftarrow train(L_i, S_i^1)$;

$tst \leftarrow test(M_{L_i}, S_i^2)$;

$[TP, TN, FP, FN] \leftarrow caculate(tst)$;

$accuracy_j \leftarrow (TP + TN)/(TP + TN + FP + FN)$;

if ($accuracy_j > accuracy$ &&

$accuracy_j < threshold$)

$(S_i^{1}{}', S_i^{2}{}') \leftarrow \text{resel\_with\_RF}(tst, S_i^1, S_i^2)$

$S_i^1 \leftarrow S_i^{1}{}', S_i^2 \leftarrow S_i^{2}{}'$

$accuracy \leftarrow accuracy_j$

$w_i \leftarrow accuracy_j$

else if ($accuracy_j > threshold$)

break

end if

end for

$w_i \leftarrow w_k / \sum_{k=1}^{K} w_k, (k = 1,..., K)$

end for

Outputs: $ENS \leftarrow \{(M_{L_k}, w_k) \mid k = 1,.., K\}$

Testing: $\tilde{H} \leftarrow sign\left(\sum_{k=1}^{K} w_k M_{L_k}(S)\right)$

Figure 2. Pseudo code for bracing algorithm

Algorithm: $(S_1', S_2') \leftarrow \text{resel\_with\_RF}(tst, S_1, S_2)$

Input: tst, the test result of a learning model;
S_1 , the training dataset used by a learning model
S_2 , the validate dataset used by a learning model

Process: $(inx_1, inx_2) \leftarrow finderr(tst, S_1, S_2)$;
$(S_1', S_2') \leftarrow exchange(S_1, S_2, inx_1, inx_2)$;

Outputs: S_1', S_2'

Figure 3. Pseudo code for reselecting samples with relative feedback

Bracing is able to focus on the nearest error or true samples that the current base learning algorithm barely predicts right or wrong, by iteratively changing the training dataset with the relative feedback samples from the previous test procedure and the weights of the base learner, so that the margins on these ‘difficult’ points can be considered by the learner, on the other hand, our method attempts to pick a subset of classifiers from an existing ensemble to improve the effectiveness and efficiency of the ensemble.

Our approach is based on the observation that the generalization error of an ensemble base model could be improved if the base learners on which averaging is done disagree and if their fluctuations are uncorrelated[21]. The theoretical background of bracing is provided by diversity, and the bias/variance decomposition of the generalization error. As we known, many ensemble paradigms employ the same classification model, for example, a decision tree or a neural network, but there is no evidence that this strategy is better than using different models. We argue that an ensemble of heterogeneous models leads to a reduction of the ensemble variance and an increase of the ensemble diversity because the errors of the individual components have small correlation and thus the cross terms in the variance are small. To improve the ensemble accuracy, bracing employs different base learner to build an ensemble. As we known, the combination of the output of several base learners is only useful if they disagree on some inputs. In our paper we refer to the measure of disagreement as the diversity of the bracing ensemble. For regression problems, mean squared error is generally used to measure accuracy, and variance is used to measure variance. In this context, Krogh and Vedelsby [22] show that the generalization error, E, of the ensemble can be expressed as $E = \bar{E} - \bar{D}$, where $\bar{E}$ and $\bar{D}$ are the mean error and variance of the ensemble while maintaining the average error of ensemble members, should lead to a decrease in ensemble error. There is strong reason to believe that increasing diversity or decrease variance should decrease ensemble generalization error. In our task, we use the disagreement of an ensemble member (base learner) with the ensemble’s final decision as a measure of diversity. More precisely, if $L_i(X)$ is the prediction of the entire ensemble, then the diversity of the *ith* base learner on example X is given by

$$Div_i(x) = \begin{cases} 0: & if\ \left|L_i(x) - L^*(x)\right| \le T \\ 1: & otherwise \end{cases}, \tag{7}$$

where $L^*(x)$ is the ground truth of sample x, T is the error tolerant parameter, which is set as the variance of prediction accuracy of all the ensemble members. To compute the variance of an ensemble of size K, on a training set of size m, we average the above term:

$$Div_{ens}(X) = \frac{1}{mK}\sum_{i=1}^{K}\sum_{j=1}^{m} Div_i(x_j). \tag{8}$$

This measure estimates the probability that a classifier in an ensemble will disagree with the prediction of the ensemble as a whole. Our approach is to build ensembles that are consistent with the training data and that attempt to maximize this diversity term and minimize the variance term at the same time.

IV. MCs DETECTION BASED ON BRACING ENSEMBLE

In this paper, MCs detection is formulated as a binary classification problem. At each location in a mammogram, the proposed ensemble is applied to determine where a MCs object is present or not. We defined $X \in R^n$ as a pattern to be classified, and y as its class label (i.e., $y \in \{\pm 1\}$).

A. Mammogram preprocessing

Before MCs detection our task is to suppress the mammogram background of each image block. A high-pass filter is designed to preprocess each mammogram before extracting samples. With the Gaussian filter, we use a $n \times n$ window size Gaussian filter where $n = 4\sigma^2 + 1$, experimentally in the study $\sigma = 2$. The output of high-pass filter is denoted by $I_2(x,y) = I_1(x,y) - f(x,y) * I_1(x,y)$, where $*$ is linear convolution.

To enhance the spot-like characteristics (microcalcifications), the top-hat operation is performed, $I_3(x,y) = I_2(x,y) - (I_2(x,y) \exists B(x,y)) \oplus B(x,y)$, where $\exists$ is a morphological erosion operation, $\oplus$ is a morphological dilation operation, and $B(x,y)$ is structure element. After applied the filter and enhancement method each image blocks seem to be effective in reducing the inhomogeneity of the background and the microcalcifications are also enhanced in mammogram.

B. Input patterns for MCs detection

After the preprocessing stage, we extract image 116-dimension features X_i of $A^i_{m \times m}$. $A^i_{m \times m}$ is a small window of $m \times m$ pixels centered at a location that we concerned in a mammogram image. The choice of m should be large enough to include MCs (in our experiment, we take m =115). The task of the trained ensemble classifier is to decide whether the input window $A^i_{m \times m}$ at each location is a MCs pattern or not.

The procedure for extracting training data from the training mammograms is given as follows. For each MCs location in a training mammogram set, a window of $m \times m$ image pixels centered at its center of mass is extracted; the area is denoted by $A^i{}_{m \times m}$, with respect to $\mathbf{x}_i$ after subspace feature extraction, and then $\mathbf{x}_i$ is treated as an input pattern for the positive sample ($y_i = +1$). The negative samples are collected ($y_i = -1$) similarly, except that their locations are randomly selected from the non MCs locations in the training mammograms. In the procedure, no window in the training set is allowed to overlap with any other training window.

C. Base learning algorithms for bracing

In our task we choose a lot of models (base learner) as base learning algorithm to solve our problem. All models belong to the canonical collection of machine learning algorithms for classification tasks so details can be found in the textbooks like for instance Duda et al. [23]. Table

III shows the used base models or learners in our research work.

TABLE III.
MODELS USED IN ENSEMBLE LEARNING

Types	Models	#	Base Learner
Global Models	Linear Models	1	LDA
	Neural Networks	2	MLP
		3	BP
	Support Vector Machines	4	$\mathrm{SVM_{RBF}}(\sigma=2, C=1000)$
		5	$\mathrm{SVM_{RBF}}(\sigma=3, C=1000)$
		6	$\mathrm{SVM_{RBF}}(\sigma=4, C=1000)$
		7	$\mathrm{SVM_{RBF}}(\sigma=5, C=1000)$
		8	
Semi-global Models	Decision Trees	9	C4.5
		10	CART
Local Models	K-Nearest-Neighbors	11	KNN

D. Training ensemble with bracing algorithm

In the MCs detection context, we have formulated our task as a binary classification problem. For a given data set of input-output pairs $(X_i, y_i) \in R^D \times R, i = 1,...,m$, we aim to choose a ensemble model (function) $\hat{f}$ out of hypothesis space H which is as close to true dependency f as possible, where $f: R^D \mapsto \{+1, -1\}$. In order to improve the classification by combining classifiers trained on randomly generated subsets S_1, S_2 of the dataset S , we using bracing algorithm to reselected the subset S_1, S_2 by using the relative feedback of the previous test/validation procedure. In the feedback stage the nearest classification error samples of the previous test step in S_2 will replace the nearest true classification samples in S_1.

V. DATABASE AND PERFORMANCE EVALUATIONS

In this part, the digital database for screening mammography (DDSM) database [24] built by University of South Florida is used, which is available for research at [25]. In making the database, the optical density range of the scanner was 0-3.6 (OD). The 12bits digitizer was calibrated so that the gray values were linearly and inversely proportional to the optical density. In our experiments, all selected images are intensity images, digitized at 43.5 $\mu m / pixel$ and a *12*-bit gray scale. In the DDSM database, the boundaries for the suspicious regions are derived from markings made on the film by at least two experienced radiologists. Each boundary for the abnormality is specified as a chain code, which allows us to easily extraction ROIs for each of the suspicious areas in the image files.

To evaluate the proposed MCs detection method, a set of 267 images of clinical mammograms from the DDSM database were selected to form the evaluation database. In our experiments, the negative samples were automatically selected from the normal breast region. While the positive samples dataset were manually selected from the suspicious areas of each selected images, following the reported ROI selection method in [26]. A 115×115 window (approximately $5mm \times 5mm$) was chosen as the ROI size, since the microcalcification cluster was defined as a region containing three or more microcalcifications per $5mm \times 5mm$ area. Therefore, we need to select ROIs at a shorter interval so that the center of a microcalcification cluster will be at the center of one of the ROIs. Although we must select ROIs at intervals of 1 pixel (0.0435 mm) to analyze a mammogram in detail, there were no large differences between adjacent ROIs selected at intervals of 1 mm. Therefore, we selected the ROIs at intervals of 23 pixels (approximately 1 mm) so that one ROI would overlap with the adjacent ROIs. So we can get more positive ROIs because of the overlapping. In our experiments, we got 2,231 positive samples in our dataset, and the negative samples were not limited (e.g., we choose 8,364 negative samples) because we could get a lot more normal tissues than the suspicious areas.

To evaluate the performance of the trained bracing classifier, we used receiver operating characteristic (ROC) curves as an evaluation criteria [27]. Receiver operating characteristic (ROC) analysis, based on statistical decision theory, is a commonly used approach for classification performance evaluation. A ROC curve is a plotting of the classifier's true positive detection rate (TPR) (also known as sensitivity) as a function of the classifier's false positive detection rate (FPR) (1.0-TPR also known as specificity). The true positive rate is the probability of correctly classifying a target object, while the false positive rate is the probability of incorrectly classifying a target object. The Area under the ROC curve (Az) is an accepted way of comparing classifier performance. A perfect classifier should have a TPR rate of 1.0 (or 100%) and FPR rate of 0.0% and therefore an Az of 1.0. A higher Az would indicate greater discrimination capacity of a classifier. We constructed the ROC curves by the following procedure by applying the trained bracing classifier with varying thresholds to classify each test samples as positive sample ($+1$) or negative sample (-1).

VI. EXPERIMENTAL RESULTS

Up till now, we have shown our approach to MCs detection. In this section we evaluate the performance of our method by using the real mammogram data from DDSM. The data in the training, validation, and test sets was randomly selected from the preprocessed dataset. Each selected sample was covered by a 115x115 window whose center coincided with the center of mass of the suspected MCs. The database of blocks included 2231 with true MCs and 8364 with normal tissue. To simply our task, we only choose the same number ($2231 \times 50\% \approx 1116$) of positive samples (MCs) and negative samples (normal tissue) from the database during the training stage each time. In our experiments, 70% of the blocks were assigned to the training set, 15% to the validation set, and 15% to the test set. Table IV gives a summarization of the selected blocks in the three data sets.

TABLE IV.
DISTRIBUTION OF SELECTED DATA SET.

Data pool	Training data set(1)	Validation data set (2)	Test data set (3)	Total (1)+(2)+(3)
MCs(a)	1116*70%	1116*15%	1116*15%	1116
Normal tissue(b)	1116*70%	1116*15%	1116*15%	1116

In the experiments, we first run experiments with 11 base classifiers for 20 rounds to select the suitable base classifiers. The results are averaged over twenty five-fold cross-validations. Table 5 summarizes the comparison performance results of each single base learner on the same dataset.

TABLE V.
TEST RESULTS OF OVERALL ACCURACY(ACC), SENSITIVITY (SE), SPECIFICITY(SP) AND AREA UNDER ROC CURVE(AZ) ACROSS 100 EXPERIMENTS USING 5-FOLD CROSS-VALIDATIONS WITH THE SAME TRAINING AND TEST DATASET

#	Acc	Se	Sp	Az
1	**87.83±0.02**	**85.61±0.08**	**90.62±0.07**	**0.9304±0.0002**
2	87.73±0.02	85.50±0.08	89.62±0.07	0.9296±0.0002
3	**91.79±0.01**	**91.69±0.05**	**92.01±0.03**	**0.9605±0.0001**
4	91.75±0.01	91.93±0.04	91.71±0.04	0.9460±0.0002
5	91.79±0.01	92.14±0.04	91.58±0.04	0.9537±0.0001
6	92.37±0.01	92.78±0.04	92.08±0.03	0.9553±0.0001
7	**92.81±0.01**	**93.09±0.03**	**92.62±0.03**	**0.9574±0.0001**
8	**92.76±0.01**	**92.67±0.03**	**92.94±0.03**	**0.9579±0.0001**
9	**92.65±0.01**	**93.99±0.03**	**91.48±0.03**	**0.9764±0.0000**
10	86.30±0.02	86.36±0.08	86.46±0.06	0.9238±0.0002
11	**88.09±0.02**	**89.73±0.07**	**86.77±0.05**	**0.9271±0.0003**

According to the experimental results listed in Table 5, we find that some base classifiers are not suitable for our task, so we only choose the best diverse five as base classifiers in the bracing ensemble algorithm to evaluate our method. For constructing the bracing ensemble, we set the ensemble size to 5 in our experiments, because we only use 5 of the base learning algorithms in Table 5. Note that in the case of bracing we can specify a desired ensemble size, if we have enough base learners.

To evaluate the performance and observe the characteristics of bracing ensemble algorithm on the microcalcification clusters detection problem thoroughly, five-fold cross validation is carried out, where the labeled samples are sampled into five subsets with similar class distribution to that in the original labeled samples and in each partition samples are divided into two groups: one for training and the other for test. In each fold, classifiers are evaluated on the test set after being trained on the other training group in each partition. We first run experiments 20 times with 11 base classifiers for each rounds. The results are averaged over twenty five-fold cross-validations. According to the experimental results, we find that some base classifiers are not suitable for our task, so we only choose the best five as the base classifiers in the bracing ensemble algorithm to evaluate our method.

So in our following experiments, we set the ensemble size to 5, because we only use 5 base learning algorithms. Note that in the case of bracing we can specify a desired ensemble size, if we have enough base learners. In the bracing relevance feedback stage the algorithm terminates if the number of iterations exceeds the maximum limit or the desired result is obtained.

The performance of each base classifier was evaluated and compared with bracing by using $N(N=10,20,...,100)$ complete runs of five fold cross validation. Final experimental results are shown in Fig. 4.

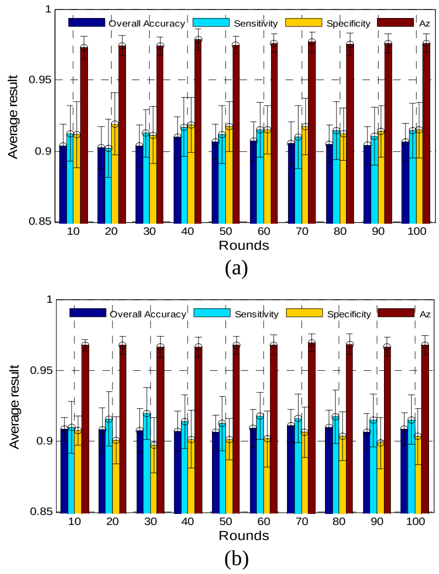

Figure 4. Average results of MCs detection under different training rounds, which from 10 to 100: average results of MCs detection by using (a) bracing ensemble and (b) bagging ensemble learning method.

Since bracing can be viewed as a modification of bagging, a comparison was made between bracing and bagging (shown in Fig. 5). In each test, the same bootstrap samples were obtained for both algorithms in each round. Compared the performance of bracing with base learning algorithms, we also drawn the ROC curve of each base learner in our experimental context with the same relevance feedback scheme. Results are shown in Fig. 6.

All of the above experiments, which compare this algorithm with other peer algorithms on the real datasets, demonstrate the capabilities of the new algorithms. Our results indicate that bracing outperforms bagging and it is also ahead of its peer algorithms in terms of accuracy. Its application in the classifier improves the effectiveness and robustness of the selected ensembles.

VII. CONCLUSION

In this paper, we proposed a new ensemble learning algorithm, named arcing, to detect microcalcification clusters (MCs) in digital mammograms. The proposed ensemble learning method hold the potential for providing improvements in MCs detection accuracy without resorting to the use of additional data. Compared with other ensemble methods, such as bagging, in terms of their ability to detection MCs in the breast cancer early stage, our proposed method can increase classification performance and reduce the false positive rate. Experimental results using a set of 267 mammograms indicate that bracing outperforms bagging, and it is also ahead of its peer algorithms in terms of accuracy. In our experiments, ROC curves also indicate that the proposed ensemble approach yielded the best performance compared with the traditional ensemble methods.

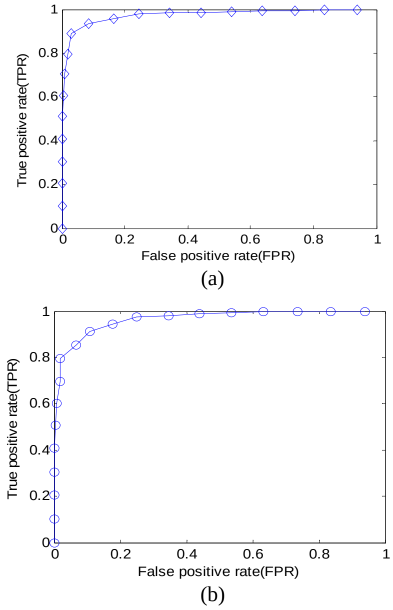

Figure 5. ROC curves of different ensemble method in MCs detection: (a) ROC curve of the bracing ensemble learning method with overall accuracy=92.60%, sensitivity=94.18% and specificity (1-FPR)=91.13%; (b) ROC curve of the bagging ensemble learning method with overall accuracy=89.29%, sensitivity=91.40% and specificity (1-FPR)=87.38%.

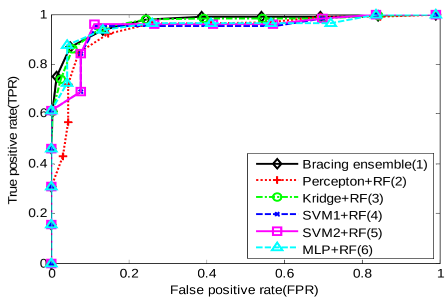

Figure 6. ROC curves of base classifiers and bracing ensemble classifier with the same training and test set to detect MCs.

ACKNOWLEDGMENT

This work presented in our paper is supported by the Natural Science Basic Research Plan in Shaanxi Province of China (No. 2007F$_2$48) and the youth scientific research foundation in Xi'an Univ. of Arc. & Tech. (No. QN0738). The authors are grateful for the anonymous reviewers who made constructive comments.

REFERENCES

[1] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, pp. 2967-2991, 2003.

[2] C. Heng-Da, L. Yui Man, and R. I. Freimanis, "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Trans. Med. Imag.*, vol. 17, pp. 442-450, 1998.

[3] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, pp. 141-150, 2005.

[4] W. Chao, J. Wei, and D. Xifeng, "Characterization of clustered microcalcifications in mammograms based on support vector machines with genetic algorithms," in *Proc. IEEE Int'l Conf. Biophotonics, Nanophotonics and Metamaterials*, 2006, pp. 114-117.

[5] C. D'Elia, C. Marrocco, M. Molinara, G. Poggi, G. Scarpa, and F. Tortorella, "Detection of microcalcifications clusters in mammograms through TS-MRF segmentation and SVM-based classification," in *Proc. IEEE Int'l Conf. Pattern Recognition(ICPR 2004)*, 2004, pp. 742-745 Vol.3.

[6] W. Liyang, Y. Yongyi, R. M. Nishikawa, M. N. Wernick, and A. Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications," *IEEE Trans. Med. Imag.*, vol. 24, pp. 1278-1285, 2005.

[7] W. Liyang, Y. Yongyi, R. M. Nishikawa, and J. Yulei, "A study on several Machine-learning methods for classification of Malignant and benign clustered microcalcifications," *IEEE Trans. Med. Imag.*, vol. 24, pp. 371-380, 2005.

[8] S. Sukhwinder, K. Vinod, H. K. Verma, and S. Dilbag, "SVM Based System for classification of Microcalcifications in Digital Mammograms," in *Proc. IEEE Int'l Conf. Engineering in Medicine and Biology Society(EMBS '06)*, 2006, pp. 4747-4750.

[9] S. Sehad, S. Desarnaud, and A. Strauss, "Artificial neural classification of clustered microcalcifications on digitized mammograms," in *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics*, 1997, pp. 4217-4222.

[10] S. Halkiotis, T. Botsis, and M. Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks," *Signal Processing*, vol. 87, pp. 1559-1568, 2007.

[11] A. C. Patrocinio and H. Schiabel, "Classifying clusters of microcalcification in digitized mammograms by artificial neural network," in *Proc. IEEE Int'l Conf. Computer Graphics and Image Processing*, 2001, pp. 266-272.

[12] G. Rezai-rad and S. Jamarani, "Detecting microcalcification clusters in digital mammograms using combination of wavelet and neural network," in *Proc. IEEE Int'l Conf. Computer Graphics, Imaging and Vision: New Trends*, 2005, pp. 197-201.

[13] L. Bocchi, G. Coppini, J. Nori, and G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," *Medical Engineering & Physics*, vol. 26, pp. 303-312, 2004.

[14] J. Wang, H. Peng, J.-S. Hu, and J. Zhang, "Ensemble learning for keyphrases extraction from scientific document," Chengdu, China, 2006, pp. 1267-1272.

[15] Z. S. H. Chan and N. Kasabov, "Fast neural network ensemble learning via negative-correlation data correction," *Neural Networks, IEEE Transactions on,* vol. 16, pp. 1707-1710, 2005.

[16] N. Ueda, "Optimal linear combination of neural networks for improving classification performance," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, pp. 207-215, 2000.

[17] L. Ming and Z. Zhi-Hua, "Improve Computer-Aided Diagnosis With Machine Learning Techniques Using Undiagnosed Samples," *IEEE Trans. Syst., Man, Cybern. A,* vol. 37, pp. 1088-1098, 2007.

[18] L. Tsong-Wuu and C. Yun-Feng, "A comparative study of Zernike moments," 2003, pp. 516-519.

[19] M. S. Tamura H., Yamawaki T., "Texture features corresponsing to visual perception," *Systems, Man and Cybernetics, IEEE Transactions. on,* vol. 1, pp. 460-472, 1978

[20] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*. North Chelmsfo: Courier Dover Publications, 2001.

[21] A. Krogh and P. Sollich, "Statistical mechanics of ensemble learning," *Physical Review E,* vol. 55, pp. 811-825, 1997.

[22] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation and active learning," *Advances in Neural Information Processing Systems-7,* pp. 231–238, 1995.

[23] P. E. H. Richard O. Duda, David G. Stork, *Pattern Classification*, Second Edition ed.: John Wiley & Sons, Inc., 2001.

[24] R. M. Nishikawa, "Mammographic Databases," *Breast Disease,* vol. 10, pp. 137-150, 1998.

[25]"http://marathon.csee.usf.edu/Mammography/Database.html."

[26] R. Nakayama, Y. Uchiyama, K. Yamamoto, R. Watanabe, and K. Namba, "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms," *IEEE Trans. Biomed. Eng.,* vol. 53, pp. 273-283, 2006.

[27] J. B. Tilbury, W. J. Van Eetvelt, J. M. Garibaldi, J. S. H. Curnsw, and E. C. Ifeachor, "Receiver operating characteristic analysis for intelligent medicalsystems-a new approach for finding confidence intervals," *IEEE Trans. Biomed. Eng.,* vol. 47, pp. 952-963, 2000.

Xinsheng Zhang is presently pursuing his PhD degree in school of Electronic Engineering, Xidian University. He received his M.S. Degree and the B.S. degree in system engineering from Xi’an University of Architecture & Technology, China, in 2004 and 2001, respectively.

His research interests include image processing and machine learning.